hexagon:2610.00376v1

Slow Equilibration and Persistent Stationary Correlations in Spherical Spin Glasses

P. M. Aronow

Patrick Lopatto

AUTHORS' NOTE

With the exception of this note, the entire document is machine-written. We have carefully supervised its preparation to ensure that the proofs are readable and that previous literature is appropriately cited. We are releasing this manuscript on Hexagon to disseminate the result as quickly as possible because it may be of interest to the mathematical community. We welcome suggestions concerning mathematical corrections, omitted citations, and attribution of ideas.

P. L. was partially supported by NSF grant DMS-2450004.

ABSTRACT

We study the Langevin dynamics of spherical mixed p-spin glasses at fixed temperature, on time scales exponential in the dimension N. Below the static transition temperature, the dynamics started from the uniform distribution on the sphere stays at total variation distance close to one from the Gibbs measure up to a time exponential in N, with probability tending to one. We prove this for pure models of degree at least four at every such temperature, and for mixtures of minimum degree at least four that satisfy an explicit condition on a function of one variable. It extends to initial distributions that are independent of the disorder or depend weakly on it. Between the dynamical and static transition temperatures, we show for general finite mixtures that, with probability exponentially close to one, the generator has exponentially many exponentially small eigenvalues; a fixed part of the centered spin lies in the corresponding slow modes; the stationary spin correlation stays above every level below q_*^2 up to an exponential time, where q_* is an explicit root of the plateau equation of the physics literature; the dynamics started from most points of the Gibbs measure stays far from equilibrium; and the two endpoints of the stationary dynamics share mutual information at least linear in N. With probability tending to one, the correlation stays above every level below q_*. Near the dynamical transition, exponentially many eigenvalues stay below e^{-cN}, where c is proportional to an explicit power of the distance to the transition.

Contents

1 Introduction 1
- 1.1 The model and the two thresholds . . . 2
- 1.2 The uniform initial law and Gibbs-typical starting points . . . 5
- 1.3 Main results . . . 6
- 1.4 Background from physics . . . 10
- 1.5 Mathematical background . . . 13
- 1.6 Ideas of the proofs . . . 18
- 1.7 Organization of the paper . . . 22

2 Proof overview 24
- 2.1 Two kinds of slowness . . . 24
- 2.2 The condensed phase . . . 25
- 2.3 Below the static threshold . . . 36
- 2.4 Dependencies among the main steps . . . 41

PART I Preliminaries 45

3 Preliminaries 46
- 3.1 Conventions . . . 46
- 3.2 The model . . . 47
- 3.3 Langevin dynamics on the sphere . . . 50
- 3.4 Markov semigroups and relative entropy . . . 55
- 3.5 The Crisanti–Sommers functional and the static thresholds . . . 59
- 3.6 Two inputs from Huang, Montanari, and Pham . . . 64

PART II The condensed phase 66

4 Total variation lower bounds from Gibbs concentration and hitting estimates 67
- 4.1 Setting and notation . . . 67
- 4.2 Measurability and rotations . . . 69
- 4.3 Separation from Gibbs concentration, counts, and hitting bounds . . . 72
- 4.4 Concentration on the bands of a fixed window from a coarse cover . . . 74
- 4.5 Initial laws other than the uniform law . . . 77

5 Statics of the condensed phase 82
5.1 Ground states and the generalized TAP functional. 84
5.2 Multisamplable overlaps and the contact set. 86
5.3 Strict one-step replica symmetry breaking. 91
5.4 Value, radial derivative, and Hessian at radius $\sqrt{Q}$ 95
5.5 Continuation of the 1RSB solution and the derivative of the free energy . . . 99
5.6 Classes of strictly 1RSB covariances 102

6 The directing measure 110
6.1 Overlap arrays and directing measures 111
6.2 Pairs of samples near overlap one. 113
6.3 Replica exclusion and the norms of directing vectors 115
6.4 A Ghirlanda–Guerra identity for the covariance 116
6.5 Proof of Theorem 6.3 119

7 Determinant moments and the Kac–Rice formula 121
7.1 Standing notation. 122
7.2 The one-point law and the conditioned laws. 124
7.3 Gaussian concentration and determinant moments 127
7.4 The semicircle law and bordered determinants 132
7.5 The Kac–Rice formula with field-dependent weights. 139
7.6 Hitting probabilities as weights 145
7.7 Determinant-weighted transfer 145

8 Cap hitting estimates under critical-point conditioning 149
8.1 The main estimate 150
8.2 The mechanism and the origin of the errors 152
8.3 Harmonic coefficients and regularity of the field 154
8.4 Exact regression of the whole field 158
8.5 Latitude concentration and moments of the density ratio 160
8.6 An entropy–resolvent comparison. 164
8.7 Killing-time bound for the truncated process 168
8.8 The cutoff in the gradient variable 171
8.9 Proof of the main estimate 172

9 Pure models: critical-point counts and Gibbs concentration on bands 175
9.1 Parameters and statements 176
9.2 The exponent of the latitude integral 179
9.3 The conditional field and the density of critical values. 181
9.4 Localization at critical maxima 187
9.5 Empirical means of fixed replica groups. 189
9.6 A coarse cover from the directing measure 193
9.7 Band partition functions 196

9.8 The annealed scale and the partition function . . . 199
9.9 Concentration on microscopic bands . . . 201

10 Mixed models: one-point estimates and exclusion of high-energy points at positive overlap 204
10.1 The inputs . . . 206
10.2 The one-point intensity . . . 208
10.3 The marked complexity . . . 210
10.4 Centering and expected counts in fixed energy windows . . . 212
10.5 Conditional regularity and localization . . . 217
10.6 Exclusion of high-energy points at positive overlap . . . 220
10.7 Pure models . . . 227
10.8 Summary of one-point estimates . . . 231

11 Mixed models: pair estimates and counts in fixed energy windows 233
11.1 Setting, inputs, and parameters . . . 234
11.2 The two-point Gaussian structure . . . 237
11.3 Determinants under two-point conditioning . . . 242
11.4 The local pair intensity . . . 249
11.5 Counts in fixed energy windows . . . 253

12 Mixed models: Gibbs concentration on bands 257
12.1 Notation and inputs . . . 258
12.2 Regularity and regression at a center . . . 263
12.3 Fixed replicas and the coarse cover . . . 266
12.4 Determinant prefactors and the quadratic component . . . 272
12.5 The Palm expectation of the band partition function . . . 274
12.6 The intensity and the bands below the window . . . 279
12.7 The lower tail of the band partition function . . . 280
12.8 Gibbs concentration on bands in a fixed window . . . 283

13 Proofs of the condensed-phase theorems 286
13.1 A common time horizon for all fixed energy windows . . . 286
13.2 Proof of the main condensed-phase theorem . . . 288
13.3 Proofs of the corollary and of the theorem on initial laws . . . 290

PART III Below the static threshold 293

14 Constrained pressures 294
14.1 The planted and Gibbs-reference laws . . . 295
14.2 The cross-section formula . . . 297
14.3 The trial pressure and the equality at q_* . . . 302
14.4 Caps, core bands, and collars around reference points . . . 306

14.5 Replica concentration and the Gibbs mass of caps 309

15 Counting slow modes and the overlap-weighted Gibbs mean 311
15.1 Standing notation. 312
15.2 Counting with nonnegative trial functions. 313
15.3 Many slow modes from one aligned observable 314
15.4 Correlation and spectral projection estimates 317
15.5 The overlap-weighted Gibbs mean 318
15.6 Slow equilibration from Gibbs-typical points 320

16 Proof of the main theorem below the static threshold 322
16.1 Parameters 323
16.2 Proof of the theorem 323
16.3 A direct route to slow modes 327

17 Correlation bounds at every level below q_* 329
17.1 Setting and parameters 331
17.2 Reduction to the norm of the overlap-weighted Gibbs mean 331
17.3 The planted law and the transverse decomposition 334
17.4 The limiting pressure 338
17.5 Two amplitude derivatives 341
17.6 Transfer to the original disorder. 346

18 Eigenvalues near the dynamical threshold 347
18.1 Infinite mixtures 347
18.2 Counting slow modes. 362
18.3 Eigenvalues near the dynamical threshold. 363

A Exact rational bounds for the nonconvex example 369

References 372

SECTION 1

Introduction

Spherical spin glasses are mean-field models of disordered systems, in which the energy is a random Gaussian function on a high-dimensional sphere. Following the work of Kirkpatrick, Thirumalai, and Wolynes [100–102], p-spin models have served as mean-field models of the glass transition, and spherical models are among those that can be solved: Crisanti and Sommers computed their statics [52], and Crisanti, Horner, and Sommers their dynamics [54]. For pure spherical p-spin models with $p \geq 3$, the standard picture involves a dynamical transition temperature T_d and a lower static transition temperature T_s. Between T_s and T_d, the Langevin dynamics is predicted to fall out of equilibrium, although the free energy keeps its annealed value: equilibrium dynamics is predicted to remain for very long times in one of exponentially many states that together carry the Gibbs measure [19, 54, 73]. Below T_s, the Gibbs measure itself condenses onto few states. Started instead from a random configuration, the dynamics below T_d is predicted to age and to stay away from the states that carry the Gibbs measure, on all time scales that survive the limit of infinite dimension [56]. Mixtures can behave differently: the two transitions may coincide (Example 18.10), and the long-time behavior can depend on the preparation [70].

In this article, we prove versions of these predictions in finite dimension N, on time scales exponential in N. We study the Langevin dynamics of spherical mixed p-spin glasses at a fixed temperature and ask two questions. The first concerns temperatures below T_s: does the dynamics started from the uniform law on the sphere, which does not depend on the disorder, reach equilibrium, and on what time scale? The second concerns the interval $T_s < T < T_d$: what does the equilibrium dynamics remember of its starting point? For the first question, we show that the law of the dynamics started from the uniform law stays at total variation distance tending to one from the Gibbs measure up to a time exponential in N. This holds at every temperature below T_s for pure models of degree at least four, and for the mixtures in a class defined by an explicit scalar condition (Theorem 1.2 and Corollary 1.3), and it extends to every initial law chosen independently of the disorder (Theorem 1.4). For the second question, we show for general finite mixtures that the generator has exponentially many exponentially small eigenvalues, that a fixed part of the spin lies in the corresponding slow modes, that the stationary spin correlation stays above a plateau up to an exponential time, that the dynamics started from most points of the Gibbs measure stays far from equilibrium,

and that the two endpoints of the stationary dynamics share mutual information at least linear in N (Theorems 1.5 to 1.7).

1.1 The model and the two thresholds

Let $S_N = \{x \in \mathbb{R}^N : |x|^2 = N\}$, and let ν_N be the uniform probability measure on S_N. For $x, y \in \mathbb{R}^N$, we write $R(x,y) = x \cdot y/N$ for the overlap and $U(x) = x/\sqrt{N}$ for the normalized spin. A *finite mixture* is a nonzero polynomial

$$\xi(r) = \sum_{p=2}^{P} c_p r^p, \qquad c_p \geq 0.$$

Its minimum degree d is the smallest p with $c_p > 0$, and ξ is *pure* if exactly one coefficient is nonzero. The field H_N is the centered Gaussian process on S_N with covariance

$$\mathbb{E} H_N(x) H_N(y) = N\xi\big(R(x,y)\big).$$

For $\xi(r) = r^p$, we may take $H_N(x) = N^{-(p-1)/2} \sum_{i_1,\dots,i_p} g_{i_1 \cdots i_p} x_{i_1} \cdots x_{i_p}$ with independent standard Gaussian coefficients, and a mixture is a sum of independent fields of this form with weights $\sqrt{c_p}$. We allow a quadratic term and odd degrees. We refer to the randomness of H_N as the disorder. For an inverse temperature $\beta \geq 0$, the Gibbs measure and the free energy are

$$\pi(\mathrm{d}x) = \frac{e^{\beta H_N(x)}}{Z_{N,\beta}} \nu_N(\mathrm{d}x), \qquad Z_{N,\beta} = \int_{S_N} e^{\beta H_N} \, \mathrm{d}\nu_N, \qquad F(\beta) = \lim_{N \to \infty} \frac{1}{N} \mathbb{E} \log Z_{N,\beta}.$$

The Gibbs measure favors large values of H_N; in the physics literature, the energy is usually $-H_N$. The limit $F(\beta)$ exists by the Parisi formula, which we recall below.

The Langevin dynamics is the diffusion on S_N with generator

$$L = \Delta_{S_N} + \beta \nabla_{S_N} H_N \cdot \nabla_{S_N},$$

that is, the solution of $\mathrm{d}X_t = \sqrt{2}\, P_{X_t} \, \mathrm{d}W_t + \beta \nabla_{S_N} H_N(X_t) \, \mathrm{d}t - (1 - 1/N) X_t \, \mathrm{d}t$, where W is a standard Brownian motion in $\mathbb{R}^N$ and P_x is the orthogonal projection onto the tangent space of S_N at x. It is reversible with respect to π, and its Dirichlet form is $\mathcal{E}(f,f) = \int |\nabla_{S_N} f|^2 \, \mathrm{d}\pi$. The operator $-L$ has discrete spectrum $0 = \lambda_0 < \lambda_1 \leq \lambda_2 \leq \cdots$, where eigenvalues are counted with multiplicity. We write $P_t = e^{tL}$, $P_t(x, \cdot)$ for its transition kernel, and μP_t for the law at time t of the diffusion with initial law μ. All these objects depend on the realization of H_N, which we suppress from the notation. The total variation distance is $\|\mu - \nu\|_{\mathrm{TV}} = \sup_A |\mu(A) - \nu(A)|$. Physicists often write the generator as $T\Delta - \nabla E \cdot \nabla$ with $E = -H_N$; this changes the time scale by a constant factor, which does not matter on the exponential time scales considered here.

The two thresholds are defined through the functions

$$A(r) = \frac{(1-r)\xi'(r)}{r}, \qquad \Phi_\beta(r) = \beta^2\xi(r) + r + \log(1-r), \qquad 0 < r < 1.$$

The *dynamical threshold* $\beta_{\mathrm{d}} = \beta_{\mathrm{d}}(\xi)$ and the *static threshold* $\beta_{\mathrm{s}} = \beta_{\mathrm{s}}(\xi)$ are given by

$$\beta_{\mathrm{d}}^{-2} = \sup_{0<r<1} A(r), \qquad \beta_{\mathrm{s}}^2 = \inf_{0<r<1} \frac{-r - \log(1-r)}{\xi(r)}. \tag{1.1}$$

For a pure model, we write $\beta_{\mathrm{d}}(p)$ and $\beta_{\mathrm{s}}(p)$ for the thresholds of $\xi(r) = r^p$.

The static threshold is the point where the free energy leaves its annealed value. By Jensen's inequality, we have $F(\beta) \le \lim_N N^{-1} \log \mathbb{E} Z_{N,\beta} = \beta^2\xi(1)/2$, and equality holds if and only if $\Phi_\beta \le 0$ on $(0,1)$, that is, if and only if $\beta \le \beta_{\mathrm{s}}$. For even mixtures, Talagrand proved that the corresponding strict inequality is sufficient for this equality and that the non-strict inequality is necessary [146, Proposition 2.3]. The equivalence follows from the optimality criterion [146, Proposition 2.1] and the Parisi formula, which Chen proved for general mixtures [46] (see Lemma 3.18). Following Huang and Sellke [90, Definition 3], we say that (ξ, β) is *strictly replica symmetric* if $\Phi_\beta < 0$ on $(0,1)$ (Definition 3.16). For every $\beta < \beta_{\mathrm{s}}$, the pair is strictly replica symmetric and $\beta^2\xi''(0) < 1$ (Lemma 3.18). We call the range $\beta > \beta_{\mathrm{s}}$ the *condensed phase*.

The dynamical threshold enters through the derivative

$$\Phi_\beta'(r) = \frac{r}{1-r}\big(\beta^2 A(r) - 1\big).$$

If $\beta \le \beta_{\mathrm{d}}$, then this identity implies that Φ_β is nonincreasing, hence $\Phi_\beta \le 0$. We conclude that $\beta_{\mathrm{d}} \le \beta_{\mathrm{s}}$. For $\beta > \beta_{\mathrm{d}}$, the set $\{r \in (0,1) : \beta^2 A(r) > 1\}$ is nonempty and its closure does not contain 1, since $A(1^-) = 0$. We define its right endpoint

$$q_* = \sup\big\{r \in (0,1) : \beta^2(1-r)\xi'(r) > r\big\}. \tag{1.2}$$

By continuity, q_* solves the *plateau equation*

$$\beta^2(1-q)\xi'(q) = q. \tag{1.3}$$

It is the largest solution, unless $\beta^2 A$ touches the value 1 at some point to the right of q_* without exceeding it. In the physics literature, for $\beta_{\mathrm{d}} < \beta < \beta_{\mathrm{s}}$, the largest solution of (1.3) is the predicted long-time limit of the equilibrium spin correlation (see Section 1.4); below the static transition, the equilibrium plateau involves the parameters of the condensed phase instead. The threshold β_{d} is the infimum of the inverse temperatures at which (1.3) has a solution in $(0,1)$. For $\xi(r) = r^p$ with $p \ge 3$, the function A is maximized at $q_d = (p-2)/(p-1)$, and $\beta_{\mathrm{d}}(p)^{-2} = p(p-2)^{p-2}/(p-1)^{p-1}$; for $p = 2$, the supremum of A is approached as $r \downarrow 0$, and $\beta_{\mathrm{d}}(2) = \beta_{\mathrm{s}}(2) = 1/\sqrt{2}$. In the normalization $\xi(r) = r^p/2$,

which is common in the physics literature, this is the classical value of the dynamical temperature [40, Section 4.5]. The interval $\beta_{\mathrm{d}} < \beta < \beta_{\mathrm{s}}$ between the dynamical and static thresholds is nonempty for pure models with $p \geq 3$, and more generally whenever $c_2 = 0$ (see Remark 18.7). It can be empty: for $\xi(r) = r^2/2 + r^4$, we have $\beta_{\mathrm{d}} = \beta_{\mathrm{s}} = 1$ (Example 18.10).

To describe the condensed phase, we recall the Crisanti–Sommers formula. For a probability measure ζ on $[0,1]$, let $\hat{x}_\zeta(t) = \int (1 - \max\{s,t\})\,\zeta(\mathrm{d}s)$. The Parisi formula expresses $F(\beta)$ as the minimum over ζ of an explicit functional of $\hat{x}_\zeta$, and its minimizer is the Parisi measure (Definition 3.13 and Theorem 3.14). This formula was derived by Crisanti and Sommers [52] and proved by Talagrand for even mixtures [146, Theorem 1.1] and by Chen in general [46, Theorem 1]; Huang and Sellke gave a constructive proof [90]. The Parisi measure is δ_0 exactly when $\beta \leq \beta_{\mathrm{s}}$. A minimizer ζ is characterized by its first variation: the function

$$V_\zeta(r) = \int_0^r \frac{r - v}{\hat{x}_\zeta(v)^2}\,\mathrm{d}v - \beta^2 \xi(r), \qquad 0 \leq r < 1,$$

is nonnegative and vanishes on the support of ζ [146, Proposition 2.1], [90, Lemma 1.2]. For the Parisi measure ζ, we call V_ζ the *first-variation function* and its zero set in $[0,1)$ the *contact set*. The contact set contains the support of ζ, and it can be larger [90, Remark 1.4]. We consider measures with two atoms, $\zeta = x\delta_0 + (1-x)\delta_Q$, which describe one-step replica symmetry breaking (1RSB). For such ζ, $\hat{x}_\zeta$ is the function $\hat{x}$ of the next definition, and V_ζ is the function V.

Definition 1.1 (Strict one-step replica symmetry breaking)**.** Let ξ be a finite mixture with minimum degree at least four, and let $\beta > 0$. For $x, Q \in (0,1)$, let $s = 1 - Q$, $C = s + xQ$, and

$$\hat{x}(v) = \begin{cases} C - xv, & 0 \leq v \leq Q, \\ 1 - v, & Q \leq v < 1, \end{cases} \qquad V(r) = \int_0^r \frac{r - v}{\hat{x}(v)^2}\,\mathrm{d}v - \beta^2 \xi(r), \quad 0 \leq r < 1.$$

The pair (ξ, β) is *strictly 1RSB* if there exist $x, Q \in (0,1)$ such that its Parisi measure is $x\delta_0 + (1-x)\delta_Q$ and its contact set $\{r \in [0,1) : V(r) = 0\}$ is exactly $\{0, Q\}$. The contact at Q is *nondegenerate* if $V''(Q) > 0$. Equivalently (Section 5.3), (ξ, β) is strictly 1RSB with a nondegenerate contact at Q if and only if there exist $x, Q \in (0,1)$ such that

$$V(Q) = V'(Q) = 0, \qquad V(r) > 0 \quad \text{for } r \in (0,1) \setminus \{Q\}, \qquad V''(Q) > 0. \tag{1.4}$$

Note that $V(0) = V'(0) = 0$ holds automatically. The conditions (1.4) involve one explicit function of one variable, and they can be checked without the Parisi formula; we call them the *1RSB assumptions*. They imply that $x\delta_0 + (1-x)\delta_Q$ is the Parisi measure, so that x and Q are determined by (ξ, β) (Proposition 5.11). Since $\hat{x}(Q) = 1 - Q$, the

condition $V''(Q) > 0$ is equivalent to $\beta^2(1-Q)^2\xi''(Q) < 1$, that is, to the positivity of the replicon eigenvalue $(1-Q)^{-2} - \beta^2\xi''(Q)$ at Q [25, Section 1.3]. In the language of physics, the Gibbs measure is one-step replica symmetry breaking: Q is the overlap of two samples from the same state, and x is the mass of the atom of the Parisi measure at zero. Our terminology follows Huang and Sellke, who call a model strictly replica symmetric if its contact set is $\{0\}$ [90, Definition 3] and, at zero temperature, strictly 1RSB if the corresponding contact set is $\{0, 1\}$ [90, Definition 4]; Definition 1.1 is the analogous notion at positive temperature. Every pair that satisfies it lies in the condensed phase, $\beta > \beta_s$ (Lemma 5.10).

1.2 The uniform initial law and Gibbs-typical starting points

The two questions of this paper concern two different initial laws, and the distinction is essential. Three ways of measuring the distance to equilibrium at time t are

$$\sup_{x \in S_N} \|P_t(x, \cdot) - \pi\|_{\mathrm{TV}}, \qquad \|\nu_N P_t - \pi\|_{\mathrm{TV}}, \qquad \int \|P_t(x, \cdot) - \pi\|_{\mathrm{TV}} \, \pi(\mathrm{d}x).$$

An exponentially small spectral gap λ_1 means that the first quantity decays slowly: some observable relaxes slowly from some starting points. It does not show that the dynamics started from ν_N excites this observable, since the second quantity depends on the projection of the density $\mathrm{d}\nu_N / \mathrm{d}\pi$ onto the slow modes. Gheissari, Grant, and Yu gave an example of this phenomenon [82, Theorem 1.2]. In a spiked spherical matrix model at low temperature, the worst-case mixing time is exponential in N, while the dynamics started from any initial law that is symmetric in the coordinate along the top eigenvector, including the uniform law, mixes in time of order $\log N$. The exponentially small spectral gaps that are known for spherical spin glasses (see Section 1.5) therefore do not answer the first question.

The uniform law ν_N is the Gibbs measure at infinite temperature, and it does not depend on the disorder. The dynamics started from ν_N describes a quench from infinite temperature to the temperature $1/\beta$. This is the protocol of Cugliandolo and Kurchan [56], whose initial configuration is chosen independently of the disorder. For brevity, we call ν_N the *uniform start*.

If instead the initial point is drawn from π, then $\pi P_t = \pi$ for every t, and the law of X_t is in equilibrium at all times. What the equilibrium dynamics can remember lies in the joint law of X_0 and X_t. It is visible in the conditional law $P_t(X_0, \cdot)$ of the dynamics started from a point X_0 drawn from π and then held fixed, in the stationary correlation $\mathbb{E}_\pi R(X_0, X_t)$, and in the mutual information between X_0 and X_t. We speak of *Gibbs-typical starting points*. They depend on the disorder, and the example $\mu = \pi$ shows that no separation result for the uniform start can hold for all initial laws that depend on the disorder. Theorem 1.4 below says how much dependence is allowed.

1.3 Main results

Throughout, the covariance ξ and the inverse temperature β are fixed before $N \to \infty$, and probabilities refer to the disorder. The conventions for constants and for statements that hold with high probability are fixed in Section 3.1.

The condensed phase

Our first result concerns the dynamics started from the uniform law in the condensed phase.

Theorem 1.2 (Slow equilibration from the uniform law)**.** *Let ξ be a finite mixture with minimum degree $d \geq 4$, and let $\beta > 0$ be such that (ξ, β) is strictly 1RSB with a nondegenerate contact at Q. Then there exists $\gamma = \gamma(\xi, \beta) > 0$ such that*

$$\inf_{0 \leq t \leq e^{\gamma N}} \|\nu_N P_t - \pi\|_{\mathrm{TV}} \longrightarrow 1 \qquad \text{in probability as } N \to \infty. \tag{1.5}$$

We make no assumption on the number of degrees with $c_p > 0$, on their parity, or on the sizes of the coefficients. By the contraction of total variation along a Markov semigroup (Lemma 3.8), the infimum in (1.5) is the distance at the final time $e^{\gamma N}$. In fact, the proof shows that one set works for all times. For every $\varepsilon > 0$ and all sufficiently large N, with probability at least $1 - \varepsilon$, there exists a set B, depending on the disorder, such that $\pi(B) \geq 1 - \varepsilon$ and $(\nu_N P_t)(B) \leq \varepsilon$ for all $t \leq e^{\gamma N}$. One may take for B the set $B_{N,M}$ of Definition 4.1: the union of the bands of a small fixed width around the critical points u of the center field $H_Q(u) = H_N(\sqrt{Q}\, u)$ with $H_Q(u) \in [m_N - M, m_N + M]$ and with normalized radial derivative $K_Q(u)/N$ in a fixed interval, where $K_Q(u)$ is the derivative of $t \mapsto H_N(t\sqrt{Q}\, u)$ at $t = 1$, m_N is a deterministic centering, and M depends on ε. With probability close to one, these values lie within a bounded distance of the maximum of H_Q (see Section 1.6).

The 1RSB assumptions can be checked in many cases. The following corollary lists the classes for which we verify them.

Corollary 1.3 (Strictly 1RSB models)**.** *The conclusion of Theorem 1.2 holds in each of the following cases.*

- *(a) The pure model $\xi(r) = r^p$, for every $p \geq 4$ and every $\beta > \beta_{\mathrm{s}}(p)$.*
- *(b) Every finite mixture ξ with minimum degree at least four such that $r \mapsto \xi''(r)^{-1/2}$ is strictly convex on $(0, 1]$, and every $\beta > \beta_{\mathrm{s}}(\xi)$. This holds in particular if the smallest and the largest degrees d and P with $c_p > 0$ satisfy $6(d-2)(P-2) > (d+P-5)^2$; for example, if all degrees with $c_p > 0$ lie in $\{4, 5, \dots, 11\}$.*
- *(c) The covariances $\xi(r) = r^d + \varepsilon \sum_{d < p \leq P,\ p \text{ even}} b_p r^p$, where $d \geq 4$ is even, $P > d$, $\beta > \beta_{\mathrm{s}}(d)$, the numbers $b_p \geq 0$ satisfy $\sum_p b_p = 1$, and $0 < \varepsilon < \varepsilon_0$, for some $\varepsilon_0 = \varepsilon_0(d, P, \beta) > 0$ that does not depend on (b_p).*

(d) The covariance $\xi(r) = r^4 + br^{12}$ at the inverse temperature β, where

$$t = \frac{45 - 24\log 6}{72\log 6 - 85}, \qquad b = t\Big(\frac{10}{9}\Big)^8, \qquad \beta^2 = \frac{15000}{729(4+12t)}.$$

Numerically, $b \approx 0.10546$ and $\beta \approx 2.12778$ in (d), and the 1RSB assumptions hold there with $Q = 9/10$ and $x = 5/9$. In case (d), the function $\xi''^{-1/2}$ is not convex on $(0,1]$, so (d) is not contained in (b) (Proposition 5.24). In case (b), the degrees may have either parity, as in (a). The pure cubic model and models with a quadratic or cubic component are not covered; we return to this point in Section 1.3.

Next, we transfer Theorem 1.2 to other initial laws, with the same exponent γ.

Theorem 1.4 (Other initial laws)**.** *Let ξ and β satisfy the hypotheses of Theorem 1.2, let γ be the exponent given there, and let $T_N = e^{\gamma N}$.*

(i) For every N, every Borel probability measure μ on S_N, and every $T \geq 0$,

$$\mathbb{E}\big[1 - \|\mu P_T - \pi\|_{\mathrm{TV}}\big] \leq \mathbb{E}\big[1 - \|\nu_N P_T - \pi\|_{\mathrm{TV}}\big]. \tag{1.6}$$

In particular, for every $\varepsilon > 0$,

$$\sup_{\mu} \mathbb{P}\Big(\inf_{0\leq t\leq T_N} \|\mu P_t - \pi\|_{\mathrm{TV}} < 1 - \varepsilon\Big) \longrightarrow 0,$$

where the supremum is over all Borel probability measures μ on S_N. The same convergence holds for every random initial law that is independent of H_N, in the joint probability of the disorder and the random initial law.

(ii) Let $0 \leq \kappa < (d-2)/2$ and $C_0 > 0$. For each N, let $\mu_{N,1}, \dots, \mu_{N,K_N}$ be Borel probability measures on S_N that do not depend on H_N, with $K_N \leq C_0 N^\kappa$. Then

$$\inf\Big\{ \Big\| \Big(\sum_{j=1}^{K_N} w_j \mu_{N,j}\Big) P_t - \pi \Big\|_{\mathrm{TV}} : \ w_j \geq 0, \ \sum_{j=1}^{K_N} w_j = 1, \ 0 \leq t \leq T_N \Big\} \longrightarrow 1$$

in probability. If the collection is random and independent of H_N, the same convergence holds in the joint probability.

(iii) Let $H \mapsto \mu_N^H$ be a Borel probability kernel from fields to S_N such that $D(\mu_N^{H_N} \| \nu_N) / \log N \to 0$ in probability. Then

$$\inf_{0\leq t\leq T_N} \Big\| \mu_N^{H_N} P_t - \pi \Big\|_{\mathrm{TV}} \longrightarrow 1 \qquad \textit{in probability.}$$

In part (i), we may take any deterministic sequence of initial laws, for instance point masses $\mu_N = \delta_{x_N}$, and the bound on the probability of failure is uniform in the initial law. The supremum is outside the probability, and it cannot be moved inside: the initial law

$\mu = \pi$ depends on the disorder and is at distance zero from π at all times. In part (ii), the infimum over the weights allows them to depend on the disorder. For example, among N^κ prescribed points, one may choose, after observing the disorder, the point from which the dynamics comes closest to equilibrium. Such a choice can carry information of order $\kappa \log N$ about the disorder. In part (iii), the initial law may depend on the disorder in an arbitrary measurable way, provided that its relative entropy with respect to ν_N is $o(\log N)$; for comparison, the relative entropy of π with respect to ν_N is of order N.

Below the static threshold

We now turn to the interval $\beta_{\mathrm{d}} < \beta < \beta_{\mathrm{s}}$ and to the equilibrium dynamics. Recall the plateau value q_* from (1.2). The centered spin $Z_N(x) = U(x) - \pi(U)$ is an $\mathbb{R}^N$-valued function on S_N, and operators act on it componentwise.

Theorem 1.5 (Slow relaxation of the equilibrium dynamics)**.** *Let ξ be a finite mixture with $\beta_{\mathrm{d}}(\xi) < \beta_{\mathrm{s}}(\xi)$, let $\beta \in (\beta_{\mathrm{d}}, \beta_{\mathrm{s}})$, and let $q_* = \sup\{r \in (0,1) : \beta^2(1-r)\xi'(r) > r\}$. For every $0 < b < q_*^2$, there exist $\kappa, \gamma, a, c, C > 0$, depending on ξ, β, and b, such that for all sufficiently large N, with probability at least $1 - Ce^{-aN}$, the following hold.*

(i) $\lambda_{\lfloor e^{\kappa N} \rfloor} \le e^{-2\gamma N}$.

(ii) With $Z_N = U - \pi(U)$ and $\Pi_N = \mathbf{1}_{(0, e^{-2\gamma N}]}(-L)$, acting componentwise, we have $\|\Pi_N Z_N\|^2_{L^2(\pi;\mathbb{R}^N)} \ge b$.

(iii) For the stationary diffusion, with $X_0 \sim \pi$, we have $\mathbb{E}_\pi R(X_0, X_t) - |\pi(U)|^2 \ge b$ for all $0 \le t \le e^{\gamma N}$.

(iv) There exists a Borel set $\mathcal{X}_N \subset S_N$ with $\pi(\mathcal{X}_N) \ge 1 - Ce^{-aN}$ such that $\|P_t(x,\cdot) - \pi\|_{\mathrm{TV}} \ge 1 - Ce^{-aN}$ for all $x \in \mathcal{X}_N$ and $0 \le t \le e^{\gamma N}$.

(v) For the stationary diffusion in the given field, the mutual information satisfies $I(X_0; X_t) \ge cN$ for all $0 < t \le e^{\gamma N}$.

We allow a quadratic term and odd degrees, and we make no assumption on the shape of A: the endpoint q_* may be a degenerate root of (1.3), and A may have several local maxima. All five conclusions hold on one event. In words, the law of the stationary dynamics is π at every time, but its endpoints stay strongly dependent for an exponential time. We comment on the conclusions in turn.

Conclusion (i) asserts that exponentially many eigenvalues are exponentially small, and (ii) shows that a fixed part of the centered spin lies in the span of the corresponding eigenfunctions, which we call slow modes. Since π is invariant, we have $\mathbb{E}_\pi R(X_0, X_t) - |\pi(U)|^2 = \langle Z_N, P_t Z_N \rangle_{L^2(\pi;\mathbb{R}^N)}$, and by the spectral theorem this quantity is nonnegative and nonincreasing in t. Conclusion (iii) states that this connected correlation stays above b up to time $e^{\gamma N}$; up to a factor $e^{-e^{-\gamma N}}$, it follows from (ii). Conclusion (iv) concerns the dynamics started from Gibbs-typical points: for most starting points x under π, the law

at every time up to $e^{\gamma N}$ is at total variation distance close to one from π. It concerns starting points that depend on the disorder, and it says nothing about the uniform start. Conclusion (v) states that the endpoint X_t of the stationary dynamics carries information at least cN about the starting point; by the chain identity of Lemma 3.12, $I(X_0; X_t) = \int D(P_t(x,\cdot)\|\pi)\,\pi(\mathrm{d}x)$.

The plateau level in (ii) and (iii) is q_*^2, while the physics literature predicts that the stationary correlation stays close to the largest solution of (1.3), which is q_* except at the exceptional temperatures noted after (1.3). In the next theorem, we reach every level below q_*, at the cost of a weaker probability bound for (ii) and (iii).

Theorem 1.6 (Correlation bounds at every level below q_*)**.** *Let ξ, β, and q_* be as in Theorem 1.5. For every $0 < b < q_*$, there exist $\kappa, \gamma, a, c, C > 0$, depending on ξ, β, and b, such that conclusions (ii) and (iii) of Theorem 1.5 hold with probability tending to one, and conclusions (i), (iv), and (v) hold for all sufficiently large N with probability at least $1 - Ce^{-aN}$.*

We do not prove a matching upper bound: the theorem does not show that the correlation stays close to q_*, only that it does not fall below any fixed level smaller than q_*.

Our last main result describes the slow modes as β decreases to β_{d}. It holds for infinite mixtures $\xi(r) = \sum_{p\geq 2} c_p r^p$ with $c_p \geq 0$ that satisfy

$$\sum_{p\geq 2} p^{5/2}\sqrt{c_p} < \infty. \tag{1.7}$$

For such ξ, the field H_N is twice continuously differentiable, and the dynamics and its spectrum are defined as before (Lemmas 18.2 and 18.3). The thresholds and A are defined by the same formulas, and A extends continuously to $[0,1]$ with $A(0) = \xi''(0)$ and $A(1) = 0$.

Theorem 1.7 (Small eigenvalues near the dynamical threshold)**.** *Let ξ satisfy* (1.7) *and $\beta_{\mathrm{d}}(\xi) < \beta_{\mathrm{s}}(\xi)$, and let $A_{\max} = \sup_{(0,1)} A = \beta_{\mathrm{d}}^{-2}$. Then the set of global maximizers of A on $[0,1]$ is a finite subset of $(0,1)$. Let q_d be its largest element. There exist an integer $m \geq 1$ and $a > 0$ such that*

$$A(q_d + y) = A_{\max}\bigl[1 - ay^{2m} + O(y^{2m+1})\bigr] \qquad (y \to 0).$$

For $\beta > \beta_{\mathrm{d}}$, let $\varepsilon = \beta^2 A_{\max} - 1$, and set

$$C_d = \frac{2m\,q_d}{(2m+1)(1-q_d)}\,a^{-1/(2m)}.$$

There exists $\kappa_* > 0$ such that for every $\vartheta \in (0,1)$, there exists $\varepsilon_0(\vartheta) > 0$ such that for every fixed $\varepsilon \in (0, \varepsilon_0(\vartheta))$, there exist $a', C > 0$ such that, for all sufficiently large N, with probability at least $1 - Ce^{-a'N}$,

$$\lambda_{\lfloor e^{\kappa_* N}\rfloor} \leq \exp\bigl[-(1-\vartheta)C_d\,\varepsilon^{1+1/(2m)}N\bigr].$$

The number of slow modes stays exponential in N with an exponent κ_* that does not depend on ε, while the exponent of the eigenvalue bound tends to zero with ε. The maximizer q_d may be degenerate, and it need not be unique; the order $2m$ of its contact with the maximum is determined by ξ. When the maximum is nondegenerate, $m = 1$ and the exponent is $3/2$. The quantity $C_d\varepsilon^{1+1/(2m)}$ is the leading order of the height of a barrier in the explicit trial pressure of Lemma 14.9, which bounds from below the corresponding barrier of the constrained pressure of Definition 14.4, the potential of Franz and Parisi (see Section 1.4). For the pure cubic model, $q_d = 1/2$, $A_{\max} = 3/4$, $m = 1$, and $C_d = 1/3$ (Example 18.9). The constants a' and C and the threshold for N may depend on ε and ϑ.

Scope

We record here what the theorems do not assert. The convergence in Theorem 1.2 is in probability; we do not show that the probability of failure is exponentially small. The covariance and the temperature are fixed before $N \to \infty$. We do not assert that the exponent γ of Theorem 1.2 is uniform near β_{s}; the uniformity in the temperature is that stated in Theorem 1.7. The proof of Theorem 1.2 requires minimum degree at least four: for a component of degree two or three, an error term in our hitting estimate for the caps around critical points does not tend to zero (Remark 13.2). In particular, the pure cubic model is not covered in the condensed phase, although it is covered by Theorems 1.5 to 1.7. In Theorem 1.4(ii), the endpoint $\kappa = (d-2)/2$ is not covered. We do not study the dynamics started from the uniform law for $\beta_{\mathrm{d}} < \beta \le \beta_{\mathrm{s}}$, where, for pure spherical models, it is also predicted to be slow [26, arXiv v1, Section 2.5, Conjecture 6]; see Section 1.5. In Theorem 1.6, the probability in (ii) and (iii) tends to one without a rate. Theorem 1.7 bounds eigenvalues only. Finally, the exponent κ in Theorem 1.5(i) is a lower bound for the growth rate of the number of slow modes; we do not identify it with the complexity of metastable states.

1.4 Background from physics

The distinction between a temperature of dynamical arrest and a lower static, or Kauzmann, temperature goes back to Kirkpatrick and Thirumalai, who connected the dynamical equations of p-spin models with theories of the structural glass transition [100], and to Kirkpatrick and Wolynes, who explained freezing above the thermodynamic transition through metastable states and their configurational entropy [101]. Kirkpatrick, Thirumalai, and Wolynes developed the corresponding activated picture for liquids in finite dimension [102]. For the spherical p-spin model, Crisanti and Sommers computed the statics and found a one-step replica symmetry breaking transition [52], and Crisanti, Horner, and Sommers analyzed the dynamics and found a dynamical transition above

the static one [54]. In our notation, these transitions occur at β_{d} and β_{s}. A pedagogical account of statics, dynamics, and complexity is given by Castellani and Cavagna [40]. Between the two transitions, the free energy has its annealed value, but the dynamics is not expected to decorrelate: the replica symmetry of the Gibbs measure does not imply fast relaxation of the equilibrium dynamics.

Between the two transitions, the correlation of the equilibrium dynamics, computed in the limit $N \to \infty$, tends to a nonzero plateau at large times. In our normalization, the plateau solves (1.3) [54]. For the pure model, this equation appears in the work of Barrat, Burioni, and Mézard, who select its largest solution [19, Equation (14)], and for mixtures in the work of Folena, Franz, and Ricci-Tersenghi [70, Equation (A2)]. Below T_s, the equilibrium plateau involves the parameters of the condensed phase instead; in our notation, Q solves $\beta^2 \xi'(Q) = Q/((1-Q)(1-Q+xQ))$ (Proposition 5.11). The literature selects the largest solution of (1.3), the nonergodicity parameter of the stationary dynamics, which is often called the dynamical Edwards–Anderson parameter. Our q_* agrees with this solution unless $\beta^2 A$ touches the value 1 to the right of q_*; at such exceptional temperatures, we use q_* as the reference value for our lower bounds. The stationary plateau should be distinguished from the plateau of the aging dynamics started from a random configuration, which satisfies the marginality condition $\beta^2(1-q)^2\xi''(q) = 1$ [73, Equation (8)]. When q_d is an interior maximizer of A, the two conditions agree at $\beta = \beta_{\mathrm{d}}$ and $q = q_d$, but they differ for $\beta > \beta_{\mathrm{d}}$.

Cugliandolo and Kurchan solved the dynamics started from a configuration chosen independently of the disorder, in the limit where $N \to \infty$ before the time tends to infinity [56]. Below T_d, this dynamics ages: correlations depend on both times, time translation invariance is lost, and the energy approaches the energy of the marginal threshold TAP states instead of the equilibrium energy. The associated threshold of the Hamiltonian at zero temperature is $2\sqrt{(p-1)/p}$ per site for $\xi(r) = r^p$ in our normalization; the asymptotic energy at positive temperature depends on the temperature. Barrat, Burioni, and Mézard contrasted this behavior with the dynamics started from equilibrium, which remains in the state in which it starts [19]. Kurchan and Laloux explained the slowness of aging through the geometry of phase space, without trapping in stable metastable states [107]. For mixed models under gradient descent at zero temperature after a preparation at positive temperature, with extensions to small thermal noise, Folena, Franz, and Ricci-Tersenghi found that the asymptotic energy depends on the preparation and that the dynamics keeps a memory of its initial condition, in contrast with the picture of a single attracting threshold in pure models [70]. Folena and Zamponi subsequently gave evidence against weak ergodicity breaking in classes of mixed models [71]. Citro and Ricci-Tersenghi obtained numerical evidence for persistent memory during aging in the spherical $3+4$ model [51], and Lang, Sachdev, and Diehl studied a transition between strong and weak ergodicity breaking as the preparation temperature varies [109].

These works concern the dynamical mean-field equations and numerical evidence, rather than finite-N probability bounds up to exponential times. Lang, Citro, Leuzzi, and Ricci-Tersenghi studied mixed spherical $(p+s)$-spin models quenched from uniformly random configurations to zero temperature. They solved the asymptotic equations of Cugliandolo and Kurchan under the assumptions of ultrametricity between time scales and local equilibration on each time scale, found aging states with one, two, or continuously many effective temperatures, and obtained the asymptotic energies analytically [110].

Franz and Parisi introduced the free energy of configurations constrained to have a prescribed overlap with a reference configuration drawn from equilibrium [73]; see also [74]. In the replica-symmetric phase, their potential is $(F(\beta) - F_{\mathrm{FP}}(r))/\beta$, where F_{FP} is the constrained pressure of Definition 14.4, so that a local maximum of F_{FP} at a positive overlap corresponds to metastable states. Monasson counted metastable states by weakly coupling replicas [119], and Kurchan, Parisi, and Virasoro interpreted saddle points of the replica free energy as metastable states and barriers [108]. Cavagna, Giardina, and Parisi studied the structure of states, the barriers between them, and the stationary points of the TAP free energy [41–43]. Barrat, Burioni, and Mézard also followed an equilibrium state through a change of temperature [19]. Barrat, Franz, and Parisi developed this following of states, including bifurcations [20], and Sun, Crisanti, Krzakala, Leuzzi, and Zdeborová revisited it for spherical $s+p$ models with planted configurations [145]. Quiet planting, introduced by Krzakala and Zdeborová for random constraint satisfaction problems [105], represents a typical equilibrium configuration through a planted law of the disorder; we use a planted law in the same way in Section 14.

The clustering and condensation transitions were formulated for random constraint satisfaction problems by Krzakala, Montanari, Ricci-Tersenghi, Semerjian, and Zdeborová [106]. Mézard and Montanari related the dynamical transition to reconstruction on trees [117], and Montanari and Semerjian proved inequalities between point-to-set correlation lengths and relaxation times, with exponential lower bounds on relaxation times for sparse models [120]. Biroli and Kurchan defined metastable states and their configurational entropy through the low-lying spectrum of the Fokker–Planck operator [33], and Tănase-Nicola and Kurchan made this construction explicit [148]; related ideas appear in the work of Gaveau and Schulman [79]. Biroli derived dynamical TAP equations [32]. Lopatin and Ioffe used dynamical instantons to study barriers between metastable states in the spherical p-spin model [114]. Ros, Biroli, and Cammarota computed the complexity of barriers near a reference minimum [132], and combined Kac–Rice analysis with dynamical mean-field theory to construct activated trajectories through index-one saddles [133]. Rizzo computed the probabilities of activated transitions between states by path integrals [130]. Charbonneau, Folena, Malatesta, Rizzo, and Zamponi studied rare trajectories from equilibrium through a dynamical potential at prescribed final overlap, emphasizing that crossing a typical saddle need not decorrelate the dynamics [44]. Near

T_d, Franz described the constrained potential by the cubic normal form of a spinodal [72, Section 5.2.1], following the classical theory of nucleation near a spinodal [150]. In this scalar normal form, the height of the barrier between the local minimum and the saddle point scales as $|T - T_d|^{3/2}$. This is the height of the scalar potential barrier; the spatial droplet barrier studied in [72, Section 5.2.1] also includes a dimension-dependent factor.

We now compare our results with these predictions. Theorem 1.2 is consistent with the aging picture of [56], in which the dynamics from a disorder-independent start does not reach the states that carry the Gibbs measure. It is a statement of a different kind: it holds in finite dimension up to times $e^{\gamma N}$, but only below the static transition, and it says nothing about the energy reached by the dynamics. Conclusions (iii) of Theorems 1.5 and 1.6 are lower bounds for the stationary correlation at levels below q_*, which is the plateau predicted in [19, 54, 70] except at the exceptional temperatures noted above; we do not prove that the correlation is close to q_*. Conclusion (iv) is a finite-dimensional form of the prediction that a configuration drawn from equilibrium remains in its state [19, 73]. Conclusion (i) is close to the spectral description of metastable states in [33, 148], but we do not identify the exponent κ with the configurational entropy. Conclusion (v) is an information-theoretic expression of the persistence of the state of the dynamics, related to the configurational entropy [119] and to the interpretation of reconstruction in [117]. We did not locate an earlier explicit statement of a lower bound, linear in N, on the stationary endpoint information up to exponential times in our regime; in the regime of [65], such a bound follows from the persistence of clusters (see Section 1.5). For $m = 1$, the exponent $3/2$ in Theorem 1.7 matches the barrier exponent of the cubic normal form [72], and $1 + 1/(2m)$ is the same computation for a contact of order $2m$. The number of slow modes stays exponential as $\beta \downarrow \beta_{\mathrm{d}}$, which is consistent with the picture in which the complexity of states need not vanish at T_d.

1.5 Mathematical background

Statics and the geometry of the Gibbs measure

The free energy of spherical models was computed by Crisanti and Sommers [52]. Talagrand proved their formula for even mixtures [146], Chen extended it to general mixtures [46], and Huang and Sellke gave a constructive proof of the lower bound [90]. The ground state energy was computed by Chen and Sen [48] and by Jagannath and Tobasco [96], and properties of Parisi measures and contact sets of spherical models were studied in [9, 10, 97]. The TAP approach was introduced by Thouless, Anderson, and Palmer [149], and it was applied to spherical pure models by Kurchan, Parisi, and Virasoro [108] and by Crisanti and Sommers [53]. Subag developed a generalized TAP approach for spherical models [143]. He introduced multisamplable overlaps, values q such that many independent samples from the Gibbs measure can have pairwise overlaps close

to q, and characterized them by equality in a TAP representation of the free energy [143, Theorem 5]. For Ising spins, the generalized TAP free energy was developed by Chen, Panchenko, and Subag [49, 50]; see also [47]. Subag computed the free energy of pure models at all temperatures rigorously from the generalized TAP approach, without the Parisi formula [142]; the resulting formula had been predicted in [53, 108].

The expected number of critical points of spherical spin glasses was computed with the Kac–Rice formula by Auffinger, Ben Arous, and Černý for pure models [12] and by Auffinger and Ben Arous for mixtures [8]. For pure models, Subag proved that the number of critical points at low energies concentrates, by a second moment computation [140], and Subag and Zeitouni identified the extremal process of critical values [144]. In [141], Subag showed that for pure models at sufficiently low temperature, the Gibbs measure concentrates on thin bands around the deepest critical points. Ben Arous, Subag, and Zeitouni obtained a similar description for mixtures close to pure models at low temperature [30]. Dembo and Subag proved a decomposition of this kind for generic mixtures that are strictly k-step replica symmetry breaking at some $\beta > \beta_{\mathrm{s}}$. For $k = 1$, almost all of the Gibbs measure lies in narrow bands around critical points of H_N on the sphere of radius $\sqrt{qN}$, where q is the positive atom of the Parisi measure, and the energies of these critical points are maximal up to $o(N)$ [59, Proposition 3.1]. The Dovbysh–Sudakov representation describes the limits of overlap arrays [63, 123]. Under hypotheses that ensure the full family of Ghirlanda–Guerra identities [83, 124], such as generic mixtures or suitable generic perturbations, these limits are ultrametric [126]; see the monograph [125]. Related stochastic stability identities were derived by Aizenman and Contucci [4].

Dynamics

The rigorous study of Langevin dynamics for spin glasses was initiated by Ben Arous and Guionnet, who proved large deviations for the empirical measure of paths and derived self-consistent limiting equations for soft spins [23, 24]. Ben Arous, Dembo, and Guionnet proved aging for the spherical 2-spin model [27] and derived the Cugliandolo–Kurchan equations for mixed p-spin models with a soft spherical constraint and an initial law independent of the disorder [28]. Dembo, Guionnet, and Mazza treated the hard spherical constraint and analyzed the limiting equations at high temperature [60]. For asymmetric soft-spin dynamics of Sherrington–Kirkpatrick type, Dembo, Lubetzky, and Zeitouni proved universality of the empirical path law at fixed times under independence assumptions on the couplings [61]. Dembo and Subag derived the limiting dynamics for initial conditions spread uniformly over a band of prescribed overlap around a critical point [57] and for initial conditions drawn from the Gibbs measure, in replica-symmetric models and generic strictly one-step replica symmetry breaking models, in the limit $N \to \infty$ followed by a limit of hard constraint [58]. For $\beta_{\mathrm{d}} < \beta < \beta_{\mathrm{s}}$, where the model

is replica symmetric, the limiting correlation of the stationary dynamics has a nonzero plateau at a root of the plateau equation [58, Proposition 1.14 and Corollary 1.18]. These results concern the limit $N \to \infty$ at fixed times, and they do not control the dynamics at times exponential in N. Jagannath surveyed the dynamics on short and long time scales [95].

In finite dimension, Ben Arous, Gheissari, and Jagannath proved differential inequalities for the energy and the gradient along the dynamics, uniformly over initial points [29]. Sellke studied pure models at sufficiently low temperature. From every initial point, the energy of the dynamics rises above the threshold energy minus any fixed $\eta > 0$ and stays above this level up to an exponential time [136, Theorem 1.1]. For an initial law independent of the disorder, the energy stays below the threshold minus a positive $\delta(p, \beta_0)$ on every fixed bounded time interval, uniformly over $\beta \le \beta_0$, with exponentially high probability [136, Theorem 1.7]. Obstructions based on the overlap gap property were proved for approximate message passing [77] and for Langevin dynamics started independently of the disorder, on the time scales considered there [78]. Huang and Sellke proved sharp hardness results for Lipschitz optimization algorithms, with applications to Langevin dynamics on bounded time intervals [91]. The energy bounds from independent starts and the hardness results cited here concern bounded time intervals; Sellke's lower bound on the energy from arbitrary starts persists to exponential times.

Exponentially slow relaxation has been studied through the spectral gap. Gheissari and Jagannath proved for pure models with $p \ge 3$ that the spectral gap is bounded below at sufficiently high temperature and exponentially small at sufficiently low temperature [81, Theorem 1]. Ben Arous and Jagannath showed that a free-energy barrier in the overlap implies an exponentially small spectral gap, and verified such a barrier for classes of mixtures that include pure even models with $p \ge 4$ [25, Theorems 1.24–1.26]. In [26], for pure spherical models with $p \ge 4$ at temperatures below a threshold T_{BBM}, they constructed exponentially many bands around critical points such that the dynamics started from the Gibbs measure conditioned on a band has exponentially small probability of escaping before an exponential time. They also formulated a weak notion of shattering and proved it for pure models with $p \ge 4$ in an interval of temperatures just above the static transition. Such bands may carry a small part of the Gibbs measure. On the high-temperature side, Kim ruled out shattering in this fixed-overlap critical-point-band sense for all overlaps in the pure cubic model when $\beta \le \beta_{\mathrm{d}}(3)$, and for specified overlap ranges when $p \ge 4$ [99, Theorem 1.1 and Corollary 1.2]. This concerns the definition of Ben Arous and Jagannath, rather than the stronger cluster decomposition below; it does not establish rapid mixing. El Alaoui, Montanari, and Sellke proved shattering of pure models in a strong sense, with consequences for the spectrum and the equilibrium correlation [65], which we discuss in Section 1.5. Auffinger, El Alaoui, and Sellke studied mixtures with a discontinuous replica symmetry breaking transition and with $\beta_{\mathrm{s}} < \xi''(0)^{-1/2}$. For all β

in an interval immediately below β_s, they proved that the constrained pressure is strictly increasing on an interval of positive overlaps [13, Theorem 1.1], and they deduced an exponentially small spectral gap [13, Remark 1.2]. Sellke studied spherical models near their ground states. For every mixture, he showed that at all approximate ground states, the gradient is close to an explicit multiple of the point and the bulk of the spectrum of the Hessian has explicit edges, and for even mixtures, that the largest eigenvalue of the Hessian is at most the upper edge, up to a small error [138, Proposition 1.2 and Theorem 1.3]. For even mixtures whose approximate ground states are not marginally stable (in his terms, without full RSB endpoint behavior), he showed by the argument of [65, Corollary 2.9] that at sufficiently large β, the stationary dynamics started from a Gibbs sample keeps overlap at least $1 - \varepsilon$ with its starting point up to time e^{cN}, with probability at least $1 - e^{-cN}$, and he noted that this implies exponentially slow mixing [138, Corollary 1.6]. We compare his estimates near the ground state with ours before Proposition 10.10. For finite even mixtures with minimum degree at least four, under a strict separation between the ground-state energy and the energies at which the annealed complexities of critical points of index one and two vanish, Badiei Khuzani proved one-sided lower bounds on the relaxation time, in probability over the disorder: a sequential Arrhenius bound, in which $\beta \to \infty$ at fixed N before $N \to \infty$; a bound at every fixed β at which the free energy exceeds β times the first of these two energies; and an aggregate Eyring–Kramers bound at every sufficiently large fixed β [15, Definition 1 and Theorems 8–10]. The current version of that paper notes that these bounds do not show slow relaxation from the uniform law, since the slow test function in the bound at fixed temperature is odd under $x \mapsto -x$, while the uniform law, the Gibbs measure, and the generator are even [15, Section 4.3]; compare Section 1.2. For Ising models, exponentially slow worst-case mixing at low temperature was proved by Sellke for the Sherrington–Kirkpatrick model [137] and by Kouraich and Warzel for p-spin models [103].

At high temperature, sampling and fast mixing are now well understood. Huang, Montanari, and Pham gave a polynomial-time algorithm, based on stochastic localization, that samples from the Gibbs measure with vanishing error in total variation when $\beta^2 \xi''(q) < (1-q)^{-2}$ for all $q \in [0,1)$ [93]. Huang, Mohanty, Rajaraman, and Wu proved that simulated annealing with Langevin dynamics samples from the Gibbs measure up to the same stochastic localization threshold [94]. Their algorithm starts from the uniform law and runs through a sequence of temperatures, and weak Poincaré inequalities control its warm starts. In a different regime, Huang and Sellke proved mixing in logarithmic time at sufficiently low temperature for strongly topologically trivial landscapes with a unique local maximum, which include suitable external fields [92]. For Ising spins, spectral gaps and fast mixing at high temperature were proved in [2, 5, 21, 67].

The relation between metastable sets and small eigenvalues of reversible dynamics has a long history. It goes back to the exit and cycle theory of Freidlin and Wentzell [75] and to Witten's spectral approach to Morse theory [153] and the semiclassical analysis of multiwell operators [87]. It was developed in the potential-theoretic approach of Bovier, Eckhoff, Gayrard, and Klein [35–39] and in the work of Helffer, Klein, and Nier [88]; see the monograph [34]. The works on reversible diffusions and the semiclassical works give sharp asymptotics as a noise parameter tends to zero, while the works on Markov chains and disordered mean-field models also treat growing systems. Our results concern spherical spin glasses in high dimension, with exponentially many slow modes.

Comparison with the closest results

We begin with the dynamics started from the uniform law. For pure spherical p-spin models, Ben Arous and Jagannath conjectured that at every temperature below the dynamical transition, the dynamics started from the uniform law takes exponential time to thermalize, with probability tending to one, and that it thermalizes in bounded time above the dynamical transition [26, arXiv v1, Section 2.5, Conjecture 6]. Their thermalization time is the first time t at which $\max_{\|f\|_\infty \le 1} \int (P_t f - \pi(f))^2 \, \mathrm{d}\nu_N \le 1/e$. Taking $f = 2\mathbf{1}_A - 1$ and applying the Cauchy–Schwarz inequality, we see that this maximum is at least $4 \, \|\nu_N P_t - \pi\|_{\mathrm{TV}}^2$. Under the hypotheses of Theorem 1.2, this thermalization time is therefore at least $e^{\gamma N}$ with probability tending to one, in the normalization of [26] in which the dynamics is reversible with respect to the Gibbs measure, after the corresponding constant rescaling of time. (The generator displayed in [26, p. 2 of arXiv v1] and the stated Gibbs measure differ by a factor of two in the drift; we compare temperatures through the inverse temperature of the Gibbs measure.) In particular, for pure models with $p \ge 4$, the lower bound of the conjecture holds at every temperature below the static transition. For the mixtures that satisfy the 1RSB assumptions of Definition 1.1, the same lower bound follows, although the conjecture was stated for pure models. Theorem 1.2 does not resolve the conjecture: the conjecture also concerns the interval $\beta_{\mathrm{d}} < \beta \le \beta_{\mathrm{s}}$ and the case $p = 3$, where our methods do not apply to the uniform start, and it predicts thermalization in bounded time above the dynamical transition. On bounded time intervals, and at temperatures low enough that the Gibbs measure concentrates above the threshold energy, the energy bound of Sellke [136, Theorem 1.7] separates the dynamics of pure models, started independently of the disorder, from equilibrium. To our knowledge, Theorem 1.2 is the first result that separates the uniform start from equilibrium in total variation up to times exponential in N throughout the condensed phase, for pure spherical models with $p \ge 4$ at every $\beta > \beta_{\mathrm{s}}$, and for the mixtures that satisfy the 1RSB assumptions. The spectral gap estimates above concern worst-case starting points, and, as the example of [82] shows, they do not imply such a statement.

The closest antecedent of Theorem 1.5 is the work of El Alaoui, Montanari, and Sellke [65]. For pure models, they proved that with probability at least $1 - e^{-cN}$, all but an exponentially small part of the Gibbs measure is covered by exponentially many clusters of small diameter and exponentially small mass, separated by bottlenecks. This holds for sufficiently large p and every β in an interval $(C, \beta_{\mathrm{s}}(p))$, where C is an absolute constant [65, Theorem 2.1], and for $p \geq 3$ whenever the constrained pressure is strictly increasing on an interval of overlaps above 0.9999 [65, Theorem 2.4]. They deduced that the first e^{cN} eigenvalues of $-L$ are at most e^{-cN} [65, Corollary 2.8], and that, with probability at least $1 - e^{-cN}$ jointly over the disorder, the stationary initial point, and the dynamics, $\inf_{0 \leq t \leq e^{cN}} R(X_0, X_t) \geq 1 - \Delta^2$, where Δ is defined from the scales of the clusters [65, Corollary 2.9]; in particular, their annealed correlation is at least $1 - 2\Delta^2$ on this interval. Exponentially many slow modes and an exponential-time plateau were therefore known for pure models in these regimes. By the same argument, Sellke obtained such a plateau at sufficiently large β for even mixtures whose approximate ground states are not marginally stable [138, Corollary 1.6]. Theorem 1.5 extends them in three ways. First, it applies to every finite mixture, including a quadratic term and odd degrees, and to every $\beta \in (\beta_{\mathrm{d}}, \beta_{\mathrm{s}})$. Second, the plateau level is given in terms of q_*: every level below q_*^2 with probability at least $1 - Ce^{-aN}$, and every level below q_* with probability tending to one (Theorem 1.6). Third, it adds the conclusions on Gibbs-typical starting points and on stationary information. For finite and infinite mixtures, Theorem 1.7 gives eigenvalue bounds close to β_{d}. On the other hand, where the results of [65] apply, they contain more: an explicit decomposition into clusters and a pathwise plateau. In their regime, trapping from Gibbs-typical points and information at least linear in N can also be derived from the persistence of clusters. We construct no clusters, and our argument is different (see Section 1.6).

In the limit $N \to \infty$ at fixed times, followed by a limit of hard constraint, the plateau of the stationary correlation for $\beta_{\mathrm{d}} < \beta < \beta_{\mathrm{s}}$, where the model is replica symmetric, was identified by Dembo and Subag [58, Proposition 1.14 and Corollary 1.18]. Theorems 1.5 and 1.6 hold in finite dimension up to exponential times, but they are only lower bounds. In their regimes, [13, 25] show that one eigenvalue is exponentially small; in Theorem 1.5, we obtain exponentially many for every $\beta \in (\beta_{\mathrm{d}}, \beta_{\mathrm{s}})$, together with the plateau. The bands of [26] trap the dynamics started from the Gibbs measure conditioned on a band; since they may carry a small part of the Gibbs measure, they do not lead to a plateau of the equilibrium correlation, as noted in [65].

1.6 Ideas of the proofs

A detailed overview of the proofs is given in Section 2. Here we describe the main steps and indicate where the tools come from.

The condensed phase. Let (ξ, β) be strictly 1RSB with parameters x, Q and a nondegenerate contact at Q. The proof of Theorem 1.2 rests on a description of the Gibbs measure that goes back to the work of Subag on pure models at low temperature [141]. Consider the center field $H_Q(u) = H_N(\sqrt{Q}\,u)$ on S_N. Its critical points with value in a window of fixed width around a deterministic centering $m_N = NE_0 + O(\log N)$, and with normalized radial derivative $K_Q(u)/N$ close to an explicit value K_0, have bounded expected number, and the maximum of H_Q lies in this window with probability close to one when the width is large. The bands $\{x \in S_N : |R(x,u) - \sqrt{Q}| \le b\}$ around these critical points u carry almost all of the Gibbs measure. On the other hand, a point drawn from ν_N has overlap of order $N^{-1/2}$ with each of these points. We show that before time $e^{\gamma N}$, the diffusion does not reach the caps $\{x : |R(x,u)| \ge a\}$, which contain the bands when $a < \sqrt{Q} - b$. The union of the bands is then a set of Gibbs measure close to one that the diffusion does not visit. In Section 4, we turn these three properties into Theorem 1.2, using the contraction of total variation and a first moment computation with the Kac–Rice formula. We choose the exponent γ before the width of the window. The inputs are proved as follows.

Statics (Section 5). We identify the Parisi measure under the 1RSB assumptions through the Crisanti–Sommers optimality criterion [146, Proposition 2.1], [90, Lemma 1.2], and the stationarity equations for two atoms are the classical one-step equations of Crisanti and Sommers [52]. When $\xi''^{-1/2}$ is strictly convex, we verify the 1RSB assumptions at every $\beta > \beta_s$ by the argument of Crisanti and Sommers that bounds the number of contacts, as presented by Talagrand [146, Proposition 2.2]; its version at zero temperature is [90, Lemma 3.21]. Using the generalized TAP representation of Subag [143], we show for every finite mixture and every β that an overlap is multisamplable if and only if it belongs to the contact set, that is, it is a zero of the first-variation function of the Parisi measure (Theorem 5.7). Subag proved that every zero is multisamplable, and the converse for generic mixtures [143, Theorem 10]. Under the 1RSB assumptions, the multisamplable overlaps are 0 and Q. The ground state of H_Q is then strictly 1RSB at zero temperature in the sense of [90, Definition 4], by the restriction and rescaling of minimizers at a contact point [143, Proposition 11] and the formula for the ground state energy [48, 96]. This gives the ground-state energy E_0 of H_Q, which determines the centering m_N, and the value K_0 of the normalized radial derivative.

The directing measure (Section 6). Every limit of the overlap array of independent Gibbs samples is generated by a random measure on the unit ball of a Hilbert space, the directing measure [63, 123]. We show that it is carried by vectors of squared norm Q (Theorem 6.3). We exclude other positive squared norms with the characterization of multisamplable overlaps, and an atom at the origin with a single identity of Ghirlanda–Guerra type, which we derive from the self-averaging argument of Panchenko [124]. This argument uses no ultrametricity and no Poisson–Dirichlet structure.

Counts (Sections 7 and 9 to 11). We use the Kac–Rice formula [3, 14, 116, 139] with weights that are bounded Borel functions of the whole field, such as hitting probabilities of the diffusion. The other inputs are the law of the field at a point of a smaller sphere [30] and bounds on all moments of the Hessian determinant. For pure models, the bound on the counts follows from a first moment computation and the tightness of the centered maximum [144]. For mixtures, we follow the conditional trial argument of Huang and Sellke [90, Section 3.3] to show that a critical point near the ground state has no other high-energy points at intermediate overlap. We then bound the second moment of the number of critical points in a fixed window, in the spirit of [140]. The proof uses a local factorization of the two-point Kac–Rice integrand and an elementary bound for spherical codes in a cap [18, 129].

Concentration on bands (Sections 9 and 12). Using the estimates of Subag for replicated bands with a fixed number of replicas [143, Lemmas 17–19 and Proposition 24], together with the directing measure, we place a Gibbs sample in a band around a critical point of nearly maximal value. We then pass to critical points in a fixed window, with Laplace estimates for the partition functions of bands and the lower tail of normalized partition functions [93, Lemma 3.2 and Remark 3.3]. For pure models at sufficiently low temperature, the concentration of the Gibbs measure on bands around the deepest critical points was proved by Subag [141, Theorem 1], and for generic mixtures that are strictly one-step replica symmetry breaking, the decomposition of Dembo and Subag places almost all of the Gibbs mass in narrow bands around critical points (see Section 1.5). We prove that the bands around the critical points in a fixed window carry almost all of the Gibbs mass, for pure models with $p \geq 4$ at every $\beta > \beta_{\mathrm{s}}$ and for all mixtures that satisfy the 1RSB assumptions. Compared with [59, Proposition 3.1], our mixtures are finite, whereas generic mixtures have infinitely many nonzero coefficients, and the centers lie in an energy window of fixed width, which contains a bounded expected number of critical points, instead of a window of width $o(N)$. On the other hand, their bands are narrower, and their decomposition also describes the Gibbs weights of the bands.

The hitting estimate (Section 8). Fix a point u and condition the field to have a critical point of H_Q at u with prescribed value and radial derivative. Using Gaussian regression, we write the conditioned field as H_N plus a correction that is of order N near u but vanishes to order $|R(x,u)|^d$ near the equator of u. The diffusion is the original Langevin diffusion in the conditioned field. We compare it, for an axis u that is uniform on S_N, with the diffusion in the unconditioned field, for which the Gibbs mass of the caps around u is exponentially small by rotation invariance. For the averages over latitudes in this comparison, we use the logarithmic Sobolev inequality on the sphere, through the curvature criterion of Bakry and Émery [16], and the Funk–Hecke formula [7]. The main obstacle is the length $e^{\gamma N}$ of the time interval, since a comparison of path laws costs an amount proportional to the time. To overcome it, we compare resolvents instead, at a

cost controlled by the relative entropy of ν_N with respect to the Gibbs measure, which is of order N, and not by the time (Lemma 8.11). The resulting error $N^{1-d/2} + N^{3-d}$ tends to zero because $d \geq 4$, and this is where the hypothesis on the minimum degree enters.

Determinant weights and other initial laws (Sections 4, 7, and 13). The hitting estimate for the caps under critical-point conditioning concerns ordinary Gaussian conditioning, while the Kac–Rice formula weights each critical point by the absolute determinant of its Hessian. We transfer the estimate with Hölder's inequality and the moment bounds for the determinant, as in the conditional Cauchy–Schwarz argument of [90, Corollary 3.16]. The transfer turns the bound $N^{-(d-2)/2}$ into $N^{-(d-2)(1-1/r)/2}$ for a fixed $r > 1$, and this is why the endpoint $\kappa = (d-2)/2$ is excluded in Theorem 1.4(ii). Part (i) of Theorem 1.4 is an instance of the classical symmetrization over a compact group, the same averaging that underlies the Hunt–Stein theorem in statistics [113, Section 8.5]: the expected distance to equilibrium is convex and rotation invariant as a function of the initial law, and ν_N is the rotation average of every initial law, so ν_N minimizes the expected distance. For parts (ii) and (iii), we combine a polynomial bound on the mass of the bands at time $e^{\gamma N}$ with a union bound and with data processing for relative entropy [127, Theorem 2.17].

Below the static threshold. The proof of Theorem 1.5 has a probabilistic part and a deterministic part. In the probabilistic part, we study the constrained pressure $F_{\mathrm{FP}}(r)$, the free energy per site of the configurations whose overlap with a reference point drawn from π is close to r, which is the potential of Franz and Parisi [73] written as a pressure. We compute it through a planted model, in which the reference point is drawn first and the disorder is tilted toward it, following El Alaoui, Montanari, and Sellke, who treated pure models [65, Propositions 3.7 and 3.8]; planted models go back to [1, 105]. We transfer exponentially rare events from the planted law to the joint law of the disorder and a reference point drawn from the Gibbs measure, as in [65, Lemma 3.5] and [13, Lemma 2.3]. Dembo and Subag also obtained, through Gaussian conditioning, a formula for the constrained pressure of general spherical models in the replica-symmetric phase $\beta \leq \beta_{\mathrm{s}}$, and its extension to generic strictly one-step replica symmetry breaking models in the condensed phase $\beta > \beta_{\mathrm{s}}$ [59, Corollary 2.1 and Theorem 2.3]. We bound F_{FP} by an explicit trial pressure, as in [65, Proposition 3.9], and show that equality holds at q_* for every finite mixture, using the monotonicity of the ratio $r\xi''(r)/\xi'(r)$. Seen from a Gibbs-typical reference point x, the Gibbs measure of a collar $\{u : q_1 \leq R(x,u) \leq q_2\}$ below q_* is then exponentially smaller than that of the cap $\{u : R(x,u) > q_2\}$. Further, every cap of fixed latitude has exponentially small Gibbs mass, by the replica concentration of Huang, Montanari, and Pham [93, Lemma 7.7].

In the deterministic part, we use one observable, the overlap-weighted Gibbs mean B_N (Definition 15.6). For $x \in S_N$, $B_N(x)$ is the mean of U under the Gibbs measure reweighted by a cutoff that suppresses the points u with $R(x,u) < q_1$. Its norm is at most one, its Dirichlet energy is exponentially small because the collar has exponentially

smaller mass than the cap, and $U \cdot B_N \geq q_1 - o(1)$ on most of the Gibbs measure. We count slow modes with the averaged variational principle of Harrell and Stubbe [86, Theorem 3.1] and of El Soufi, Harrell, Ilias, and Stubbe [66, Theorem 2.1], for which Kröger's averaged trial functions are an antecedent [104]. We apply it to translated bumps of B_N, whose overlaps are controlled by the Gibbs mass of caps. In [65], the eigenvalues are counted with smoothed indicators of clusters with disjoint supports; the averaged principle needs no disjointness. By spectral calculus, we then obtain the projection and the correlation bounds (ii) and (iii) at levels below q_1^2, and q_1 can be taken close to q_*. We deduce (iv) from a bound on the mean-square displacement of B_N along the dynamics, Markov's inequality, and binary data processing, and (v) from the chain identity for mutual information [127, Theorem 4.1].

For Theorem 1.6, we note that $|B_N(x)|^2$ is the mean overlap of two independent samples from the reweighted Gibbs measure. We show that this overlap concentrates at q_* by differentiating a transverse pressure with respect to the amplitudes of its covariance, a classical device in the theory of mean-field spin glasses [122], [13, Proposition 2.2], combined with the envelope theorem [118] and the convergence of derivatives of convex functions [131, Theorem 24.5]. The transfer from the planted law uses contiguity [93, Lemma 3.2], and this step has no rate. For Theorem 1.7, we expand A near q_d to compute the barrier of the trial pressure. We also use a bound on the Gibbs mass of caps whose exponent is uniform in the temperature. For infinite mixtures, we use bounds on the derivatives of homogeneous polynomials that follow from Bernstein's inequality [128].

The tools above are classical or published. Our proofs establish or combine the following ingredients: the hitting estimate for the caps under critical-point conditioning and its comparison of resolvents; the concentration of the Gibbs measure on bands around critical points in a fixed window, for pure models with $p \geq 4$ at every $\beta > \beta_s$ and for mixtures that satisfy the 1RSB assumptions, together with the fact that the directing measure is carried by vectors of squared norm Q; the identification of the multisamplable overlaps with the zeros of the first-variation function for every finite mixture; the equality between the constrained pressure and the trial pressure at q_* for general mixtures; the overlap-weighted Gibbs mean and its use with the averaged variational principle; and the proof of Theorem 1.6 through derivatives in the amplitudes.

1.7 Organization of the paper

Section 2 gives an overview of the proofs. Part I consists of Section 3, which fixes conventions for constants, for statements that hold with high probability, and for the order in which parameters are chosen (Section 3.1), and collects facts on the model, the dynamics, Markov semigroups, and statics that both halves of the paper use. The two halves are independent of each other; a reader interested in one of them can skip the

other.

Part II proves Theorems 1.2 and 1.4 and Corollary 1.3. Section 4 reduces separation to three properties of critical points and bands, and treats other initial laws. Section 5 derives the static consequences of the 1RSB assumptions and verifies these assumptions for the classes of Corollary 1.3, with finite computations for example (d) in Section A. Section 6 proves that the directing measure is carried by vectors of squared norm Q. Section 7 contains the Kac–Rice formula and moment bounds for determinants, and Section 8 proves the conditioned hitting estimate. Section 9 proves bounds on counts and the concentration of the Gibbs measure on bands for pure models, and Sections 10 to 12 do so for mixtures. Section 13 assembles the proofs.

Part III proves Theorems 1.5 to 1.7. Section 14 studies the constrained pressure, Section 15 contains deterministic estimates for counting slow modes and for the dynamics started from Gibbs-typical points, together with the overlap-weighted Gibbs mean, and Section 16 proves Theorem 1.5. Section 17 proves Theorem 1.6, Section 18 develops the infinite-mixture inputs and proves Theorem 1.7. Notation specific to a section is introduced at its start.

SECTION 2

Proof overview

In this section, we explain the main steps of the proofs and the ideas behind them. We do not restate the theorems. For each step, we point to the result where it is proved. In Section 2.1, we compare the two kinds of slowness that the two halves of the paper address. We explain the proofs of Theorems 1.2 and 1.4 in Section 2.2 and the proofs of Theorems 1.5 to 1.7 in Section 2.3; these two parts can be read in either order. Section 2.4 lists the dependencies among the main propositions.

Both proofs have the same structure. We first show, by a short deterministic argument, that a few properties of the Gibbs measure, or of the critical points of the field, imply the dynamical conclusion. We then prove these properties in separate sections, with tools from the statics of spherical spin glasses. In the condensed phase, the deterministic argument is the reduction of Section 4, and most of the work goes into its inputs. Below the static threshold, the probabilistic input is a set of estimates on the constrained pressure (Section 14), and most of the new arguments are deterministic (Section 15).

We write $\approx$ in the displays that are heuristics. Such a display indicates the size of a quantity to leading exponential order, or the origin of a formula; it is not used in the proofs. All other displays are exact statements, which we prove where indicated.

2.1 Two kinds of slowness

We can measure the distance to equilibrium from a worst-case starting point, from the uniform law ν_N, or from a starting point drawn from the Gibbs measure (Section 1.2). The two halves of the paper concern the second and the third of these. We use different mechanisms for them.

We first explain why an exponentially small spectral gap does not by itself slow down the dynamics started from ν_N. Let $f = \mathrm{d}\nu_N / \mathrm{d}\pi = Z_{N,\beta} e^{-\beta H_N}$. By reversibility, the density of $\nu_N P_t$ with respect to π is $P_t f$. Let $(\phi_j)_{j \geq 0}$ be an orthonormal basis of eigenfunctions of $-L$ in $L^2(\pi)$ with $\phi_0 = 1$ (Lemma 3.5). Expanding $P_t f - 1$ in this basis and using the Cauchy–Schwarz inequality, we find

$$\|\nu_N P_t - \pi\|_{\mathrm{TV}} = \frac{1}{2} \int |P_t f - 1| \, \mathrm{d}\pi \leq \frac{1}{2} \Big(\sum_{j \geq 1} e^{-2\lambda_j t} \langle f, \phi_j \rangle^2 \Big)^{1/2}. \tag{2.1}$$

We see that a small eigenvalue λ_j delays the relaxation from ν_N only through the coefficient $\langle f, \phi_j \rangle$. Small coefficients on the slow modes remove that particular obstruction to relaxation; a fast-mixing bound also requires control of the remaining terms in (2.1). The example of [82] recalled in Section 1.2 shows that a small spectral gap can coexist with fast mixing from the uniform law. To bound the distance from below, we need a different argument.

We use the variational form of the total variation distance. Since $\|\mu - \pi\|_{\mathrm{TV}} = \sup_A (\pi(A) - \mu(A))$, the distance between $\nu_N P_t$ and π is close to one if and only if there exists a set A such that $\pi(A)$ is close to one and $(\nu_N P_t)(A)$ is close to zero. Thus separation at a given time is equivalent to small endpoint mass, for the dynamics started from ν_N, on a set carrying almost all of the Gibbs measure. Note that this set must depend on the disorder: for a deterministic set A, rotation invariance shows that $\pi(A)$ and $(\nu_N P_t)(A)$ both have mean $\nu_N(A)$ over the disorder. By the contraction of total variation (Lemma 3.8), it suffices to exhibit such a set at the final time. In the condensed phase, we take for A a union of thin bands around a bounded number of critical points of a field derived from H_N. In the proof, we establish the stronger property that the path avoids the caps around these critical points in each fixed window. Below the static threshold, the Gibbs measure is predicted to be spread over exponentially many states, each of exponentially small mass, and a set that carries nearly all of the Gibbs measure must meet exponentially many of them. Our conditioned hitting estimate has a polynomial error for each state, which cannot be summed over exponentially many states. This is one reason why our results on the dynamics started from ν_N are restricted to the condensed phase.

For Gibbs-typical starting points, the law of X_t is π at all times, and the memory of the dynamics lies in the joint law of X_0 and X_t. Here we use a slowly varying observable. If $B\colon S_N \to \mathbb{R}^N$ has small vector energy, then, for the stationary diffusion,

$$\mathbb{E}_\pi \, |B(X_t) - B(X_0)|^2 = 2\langle B, (1 - P_t)B \rangle \le 2t \, \mathcal{E}_{\mathrm{vec}}(B) \tag{2.2}$$

(Lemma 15.8). For every fixed $\rho > 0$ and each time t, Markov's inequality implies that $\mathbb{P}_\pi\{|B(X_t) - B(X_0)| > \rho\} \le 2t \, \mathcal{E}_{\mathrm{vec}}(B)/\rho^2$. The endpoints are close in probability when $t \, \mathcal{E}_{\mathrm{vec}}(B) \to 0$; this is not a bound on the supremum along a trajectory. If, in addition, $B(x)$ is aligned with the direction of x, in the sense that $U(x) \cdot B(x) \ge q > 0$ for most x, then $B(X_t)$ records where the dynamics started. In Part III, we construct one such observable with exponentially small energy and derive all five conclusions of Theorem 1.5 from it.

2.2 The condensed phase

Let (ξ, β) be strictly 1RSB with parameters x, Q and a nondegenerate contact at Q (Definition 1.1). Recall the center field $H_Q(u) = H_N(\sqrt{Q}\, u)$ and the radial derivative $K_Q(u)$,

which is the derivative of $t \mapsto H_N(t\sqrt{Q}\,u)$ at $t = 1$, for $u \in S_N$. The proof of Theorem 1.2 rests on a picture of the Gibbs measure that Subag established for pure models at sufficiently low temperature [141]. In this picture, the Gibbs measure is carried by the bands $B_u = \{x \in S_N : \left|R(x,u) - \sqrt{Q}\right| \le b\}$ around the critical points u of H_Q whose values lie within a bounded distance of the maximum of H_Q. We call these critical points centers.

We can see the reason in the static results of Section 5. By Proposition 5.11(5), we have

$$F(\beta) = \beta E_0 + \frac{1}{2}\log(1-Q) + \frac{\beta^2}{2}\psi_Q(1), \tag{2.3}$$

where E_0 is the limit of $N^{-1}\max_{S_N} H_Q$ and $\psi_Q(t) = \xi(Q + (1-Q)t) - \xi(Q) - (1-Q)\xi'(Q)t$ is the residual covariance. Each term has a meaning for a single band. Conditionally on a critical point u with $H_Q(u) = NE$, the field at the points of latitude $\sqrt{Q}$ around u is NE plus a residual Gaussian field with covariance $N\psi_Q$ (Lemma 12.4). The coarea density at latitude $\sqrt{Q}$ contributes the entropy term $\log(1-Q)/2$ per site, and a latitude interval of width $N^{-1/2}$ around $\sqrt{Q}$ has Haar mass $(1-Q)^{N/2+o(N)}$. For the fixed-width band, the weighted latitude integral must also be controlled by the saddle estimates. The residual field is strictly replica symmetric (Proposition 5.11(2)), which suggests that its partition function is close to its mean $e^{N\beta^2\psi_Q(1)/2}$. Heuristically, we find

$$\int_{B_u} e^{\beta H_N}\,\mathrm{d}\nu_N \approx e^{\beta H_Q(u)}(1-Q)^{N/2}e^{N\beta^2\psi_Q(1)/2} \approx e^{NF(\beta)}e^{\beta(H_Q(u)-NE_0)}. \tag{2.4}$$

Comparing with (2.3), we see that a single band around a center of nearly maximal value carries a Gibbs weight comparable to the whole partition function, to leading exponential order. We control the number of centers with the Kac–Rice formula. For non-pure mixtures, Theorem 10.6 shows that the expected number of local maxima of H_Q, with normalized radial derivative in its admissible interval U and value above $m_N + t$, tends to $e^{-c_0 t}$ for each fixed t, where $c_0 = \beta x$. Here m_N is chosen in that theorem by requiring the expected number of these local maxima above m_N to equal one, and $m_N = NE_0 - (\log N)/(2c_0) + O(1)$. In this heuristic, the bands of the centers u with energy deficit $m_N - H_Q(u)$ close to v have total weight of order $e^{c_0 v}e^{-\beta v}$, which is summable in v because $x < 1$. Note that the ratio $c_0/\beta = x$ is the mass of the atom at zero of the Parisi measure. We expect that a finite number of bands carries most of the Gibbs measure, and that the bands of centers with a larger energy deficit carry a part that decays exponentially in the deficit.

On the other hand, a point drawn from ν_N has overlap of order $N^{-1/2}$ with each center. Fix $a < \sqrt{Q} - b$, and let $C_a(u) = \{x : |R(x,u)| \ge a\}$ be the union of the two caps of latitude a around $\pm u$. Every band B_u lies in $C_a(u)$, and the diffusion started from ν_N must enter $C_a(u)$ before it reaches B_u. The heart of the proof is that, before time $e^{\gamma N}$, it does not enter $C_a(u)$ for any center.

The separation criterion

For $M > 0$, let $\mathcal{C}_{N,M}$ be the set of critical points of H_Q with value in $[m_N - M, m_N + M]$ and normalized radial derivative in a small interval U around K_0, and let $B_{N,M}$ be the union of their bands (Definition 4.1). For a time T_N, we write $\theta_N(H,u) = \mathbb{P}^H_{\nu_N}\{X_t \in C_a(u) \text{ for some } t \le T_N\}$ for the probability that the diffusion started from ν_N enters $C_a(u)$ before T_N. Testing $B_{N,M}$ in the definition of the total variation distance and taking a union bound over centers, we have, for every realization of the field,

$$\inf_{0\le t\le T_N} \|\nu_N P_t - \pi\|_{\mathrm{TV}} \ge 1 - \pi\big(B^c_{N,M}\big) - \sum_{u\in\mathcal{C}_{N,M}} \theta_N(H_N,u) \tag{2.5}$$

(Proposition 4.5). It suffices to prove three properties. The first is that the bands carry almost all of the Gibbs mass, in the sense that $\mathbb{E}\pi(B^c_{N,M})$ tends to zero as $N \to \infty$ and then $M \to \infty$. The second is that the expected number of centers in each fixed window is bounded. The third is that the expected sum in (2.5) tends to zero for each fixed M.

The third property concerns an average over critical points of a random field. We compute it with the Kac–Rice formula, which expresses it through the law of the field conditioned to have a critical point at a fixed u with prescribed marks, that is, prescribed value and radial derivative. Let $D_u = \left|\det \nabla^2_{S_N} H_Q(u)\right|$, and let $\mathbb{E}_{h,k,u}$ denote the expectation with respect to the field conditioned on $H_Q(u) = Nh$, $K_Q(u) = Nk$, and $\nabla_{S_N} H_Q(u) = 0$ (Definition 7.3). Using Theorem 7.17 and rotation invariance, we can write this average as

$$\mathbb{E}\sum_{u\in\mathcal{C}_{N,M}} \theta_N(H_N,u) = c_N \int_{\mathcal{W}_{N,M}} p_N(h,k)\, \mathbb{E}_{h,k,u}\big[D_u\, \theta_N(H_N,u)\big]\, \mathrm{d}h\, \mathrm{d}k, \tag{2.6}$$

where $\mathcal{W}_{N,M}$ is the set of normalized marks of the window, p_N is their density at a point, and c_N collects the volume of S_N and the density of the gradient at zero. For pure models, we have $K_Q = pH_Q$, and the integral is over the value alone. We meet two difficulties here. The first is that the weight θ_N is the probability of a trajectory event. It is a bounded Borel function of the whole field. We do not assume continuity or dependence only on the value, radial derivative, gradient, and Hessian at u. The marked formulas cited in Section 7 impose continuity or tameness hypotheses. We prove the required formula for Borel weights in the coefficient space of the polynomial field, with an explicit version of the conditional law (Proposition 7.16 and Theorem 7.17).

The second difficulty is the determinant. The conditioned hitting estimate, which we describe in Section 2.2, bounds $q_N = \sup_{h,k,u} \mathbb{E}_{h,k,u}\theta_N$, without the weight D_u. We cannot simply drop the weight, since D_u and the trajectory depend on the same field. To overcome this difficulty, we apply Hölder's inequality to the normalized determinant. If the normalized radial derivative k lies a fixed distance above the edge $2\sqrt{b_Q}$ of the semi-

circle law, where $b_Q = Q^2\xi''(Q)$, then every fixed moment of $D_u/\mathbb{E}_{h,k,u}D_u$ is bounded uniformly in N (Proposition 7.7). For $r > 1$, this implies that

$$\mathbb{E}_{h,k,u}\big[D_u\theta_N\big] \le \mathsf{K}_r^{1/r} q_N^{1-1/r}\, \mathbb{E}_{h,k,u}D_u \tag{2.7}$$

(Lemma 7.22). We insert (2.7) into (2.6) and apply (2.6) once more with $\theta_N = 1$. We find that the average is at most $\mathsf{K}_r^{1/r} q_N^{1-1/r}\mathbb{E}\,|\mathcal{C}_{N,M}|$ (Theorem 7.23). This is a form of the conditional Cauchy–Schwarz argument of Huang and Sellke [90, Corollary 3.16]. Since $q_N \le CN^{-(d-2)/2}$, it is at most $C_{M,r}N^{-(d-2)(1-1/r)/2}$ (Proposition 13.1).

The order in which we choose the parameters is part of the argument. The concentration property lets M tend to infinity after N. We must choose the time $T_N = e^{\gamma N}$ before M, and the bound on the expected sum must hold for every fixed M with the same γ; a hitting estimate whose exponent depended on M would not suffice. We prove the hitting estimate uniformly over a compact rectangle $\mathcal{K} = [E_0 - 1, E_0 + 1] \times U$ of normalized marks, which contains the marks of every fixed window once N is large (Corollary 8.3). We fix the interval U, the band width b, and the rectangle first, then the cap latitude a and the exponent γ, and only then the tolerance and the window half-width M. Only the constants and the threshold for N depend on M (Table 13.2).

Statics

We now explain where Q, E_0, K_0, and c_0 come from. Recall that the Crisanti–Sommers functional has a unique minimizer ζ, the Parisi measure, and that ζ is characterized by its first variation: the function V_ζ of Section 1.1 is nonnegative and vanishes on the support of ζ (Lemma 3.15). The 1RSB assumptions of Definition 1.1 ask that the two-atom measure $x\delta_0 + (1-x)\delta_Q$ pass this test, that its contact set $\{V = 0\}$ be exactly $\{0, Q\}$, and that the contact at Q be nondegenerate, $V''(Q) > 0$. They concern one explicit function of one variable. We verify them for the classes of Corollary 1.3 in Section 5.6 and Section A.

The first variation also describes the geometry of the Gibbs measure. Following Subag [143], we say that an overlap q is multisamplable if, for every fixed k, the probability that k independent samples from π have pairwise overlaps close to q is not exponentially small. Let V_β be the first-variation function of the Parisi measure, whose zero set is the contact set. In Theorem 5.7, we show for every finite mixture and every β that q is multisamplable if and only if $V_\beta(q) = 0$, that is, if and only if q belongs to the contact set. The proof uses Subag's generalized TAP functional $T_\beta(q)$, which is the sum of β times the ground energy at radius $\sqrt{q}$, the entropy term $\log(1-q)/2$, and the free energy of the residual mixture ψ_q. Subag proved that $T_\beta(q) \le F(\beta)$, with equality exactly when q is multisamplable [143, Theorem 5]. The ground energy and the residual free energy are the values of two variational problems. We show in (5.9) that $F(\beta)$ is the same sum with both problems evaluated at trial points built from the Parisi measure, and this gives an upper bound for $T_\beta(q)$. If $V_\beta(q) > 0$, then a variation of the ground profile in an explicit direction

lowers the zero-temperature Crisanti–Sommers functional, and this leads to the bound $T_\beta(q) \le F(\beta) - V_\beta(q)^2/(8\beta^3 M_q)$, with an explicit $M_q > 0$ (Theorem 5.7(3)). If $V_\beta(q) = 0$, we show that both trial points are optimal. That every zero of V_β is multisamplable is due to Subag, who also proved the converse for generic mixtures [143, Theorem 10].

We collect the static consequences of the 1RSB assumptions in Proposition 5.11. The multisamplable overlaps are 0 and Q, and the residual model at Q is strictly replica symmetric. The ground state of H_Q is strictly one-step replica symmetry breaking at zero temperature, with energy E_0 and normalized radial derivative K_0 given by explicit formulas, and $c_0 = \beta x < \beta$. Finally, K_0 lies strictly above the edge of the semicircle law: with $y = Q/(\beta(1-Q))$, we have

$$K_0 - 2\sqrt{b_Q} = \frac{(y - \sqrt{b_Q})^2}{y} > 0. \tag{2.8}$$

This gap matters for the following reason. Conditionally on $H_Q(u) = Nh$, $K_Q(u) = Nk$, and $\nabla_{S_N} H_Q(u) = 0$, the Hessian $\nabla^2_{S_N} H_Q(u)$ is a scaled GOE matrix with spectrum near $[-2\sqrt{b_Q}, 2\sqrt{b_Q}]$, shifted by $-k$ (Lemma 5.14). For a deterministic u and prescribed marks with k near K_0, the Hessian is bounded above by a fixed negative multiple of the identity with conditional probability $1 - Ce^{-cN}$ (Corollary 7.6). The statements about all high critical points are established separately in Proposition 10.10 and Lemma 9.10. The normalized determinant moments used in (2.7) follow from Proposition 7.7.

The directing measure

The concentration argument of Section 2.2 places a Gibbs sample in the band of a center by finding additional independent replicas whose overlaps with each other and with the given sample are all close to Q. Such replicas exist if the overlap structure of the Gibbs measure has no other scale. Recall the theorem of Dovbysh and Sudakov (Theorem 6.2): every limit in distribution of the overlap array of independent Gibbs samples is generated by drawing a random measure G on the unit ball of a Hilbert space, and then independent vectors $v_1, v_2, \dots$ with law G, and setting $R_{ij} = v_i \cdot v_j$ for $i \neq j$. We show in Theorem 6.3 that, under the 1RSB assumptions, G is carried by the vectors of squared norm Q. The vectors in a small ball around v_1 then have mutual inner products close to Q, and they produce these replicas.

We have to exclude two possibilities. Directing vectors of a squared norm $q \in (0,1) \setminus \{Q\}$ would make q multisamplable, and the exponential bound for replicas at overlaps outside the contact set rules them out (Lemmas 5.8 and 6.5); we exclude vectors of norm one with an entropy estimate (Lemma 6.4). An atom of G at the origin would describe a positive fraction of samples that are asymptotically orthogonal to all other samples. To exclude it, we use a single identity of Ghirlanda–Guerra type, with the test function $\xi(R)$, which we derive from Gaussian integration by parts and the self-averaging

of the energy (Lemma 6.7). It states that, for every limiting overlap law, every $n \geq 1$, and every bounded function f of $(R_{ij})_{1 \leq i < j \leq n}$,

$$\mathbb{E}\big[f\xi(R_{1,n+1})\big] = \frac{(1-x)\xi(Q)}{n}\mathbb{E}f + \frac{1}{n}\sum_{j=2}^{n}\mathbb{E}\big[f\xi(R_{1j})\big]. \tag{2.9}$$

Let a_n be the probability that $\xi(R_{1j}) = 0$ for $2 \leq j \leq n$, and take for f the indicator of this event. The sum on the right vanishes. After the first step, we know that $\big|R_{ij}\big| \leq Q$ for $i \neq j$, which implies that $\xi(R_{1,n+1}) \leq \xi(Q)$, and the left side is at most $\xi(Q)(a_n - a_{n+1})$. We find that $a_{n+1} \leq (1 - (1-x)/n)a_n$, and hence $a_n \leq n^{-(1-x)} \to 0$. A sample with directing vector zero would satisfy $\xi(R_{1j}) = 0$ for all j. We conclude that G has no atom at the origin. The argument uses no ultrametricity and no description of the atoms or the weights of G.

Critical points near the maximum

We prove the second property of Section 2.2, bounded counts in fixed windows, in Section 9 for pure models and in Sections 10 and 11 for non-pure mixtures, together with the existence of at least one center, which the concentration argument needs. The following limits for the complexity in two variables and for the normalized intensity concern non-pure mixtures. For pure models, $K_Q = pH_Q$, and Proposition 9.1 and Lemma 9.8 provide the count bound in the energy variable alone and the tightness of the centered maximum. The first moment is a one-point Kac–Rice computation. Ben Arous, Subag, and Zeitouni computed the exponential rate of the expected number of critical points of H_Q with normalized value e and normalized radial derivative k, which is the marked complexity [30]. We show that, for each e, it is strongly concave in k, that its maximizer at $e = E_0$ is K_0, and that its derivative in e at (E_0, K_0) is $-c_0$. A Laplace computation with bounded prefactors then shows that, for all fixed $a < b$,

$$\mathbb{E}\#\big\{u : \nabla_{S_N} H_Q(u) = 0,\ K_Q(u)/N \in U,\ H_Q(u) - m_N \in [a,b]\big\} \longrightarrow e^{-c_0 a} - e^{-c_0 b} \tag{2.10}$$

(Theorem 10.6). In particular, the expected count in every fixed window is bounded, and the probability that the maximum of H_Q exceeds $m_N + L$ is at most $e^{-c_0 L}$ in the limit.

For existence, we need a second moment. For pure models, existence follows from the tightness of the centered maximum, which is a consequence of the work of Subag and Zeitouni [144] (Lemma 9.8 and Proposition 9.1). For non-pure mixtures, we bound the second moment of the number $C_N(L)$ of critical points in the window $[m_N - L, m_N + L]$ that satisfy a selection condition described below. The two-point Kac–Rice integrand is singular as the overlap r of the two points tends to ± 1. We split the pairs into three ranges.

- *Overlaps $r \geq \varepsilon$.* We use the exclusion of high-energy points at positive overlap, Theorem 10.13, which is proved at the ground-state parameters (E_0, K_0) and follows

the conditional trial argument of Huang and Sellke [90, Section 3.3]. We show that, conditionally on a critical point at u with marks near (E_0, K_0), with probability $1 - e^{-cN}$ no point at overlap between ε and $1 - \eta$ with u has value larger than NE_0 minus a fixed multiple of N, and no other critical point has overlap at least $1 - \eta$ with u. The local maxima with this property are called retained maxima (Definition 10.14). Distinct retained maxima have overlap smaller than ε, and the selection removes only an exponentially small fraction of the expected count, also after weighting by the determinant (Proposition 10.15).

- *Overlaps $r \le -\delta$.* For even mixtures, these pairs are reflections of pairs with $r \ge \delta$. In general, an elementary bound for spherical codes in a cap suffices [18, 129]. Suppose that some unit vectors have pairwise inner products at most $\varepsilon < \delta^2$, and that $v_1, \dots, v_k$ among them have inner product at most $-\delta$ with a fixed v_0. Then $k\delta \le |v_1 + \cdots + v_k| \le (k + k(k-1)\varepsilon)^{1/2}$, which implies that each retained maximum has at most $(1 - \varepsilon)/(\delta^2 - \varepsilon)$ such points, deterministically (Lemma 11.12).
- *Overlaps $|r| < \delta$.* We show that, for $|r| \le N^{-\alpha}$ with $1/4 < \alpha < 1/2$, the two-point integrand is the product of the two one-point integrands and the density of the overlap of two independent uniform points, up to a factor $1 + O(N^{1-4\alpha} + N^{-1/2})$ (Theorem 11.10). On $|r| \le N^{-\alpha}$, the overlap density satisfies $f_N(r) = (1 + o(1))\sqrt{N/(2\pi)}\, e^{-Nr^2/2}$, uniformly in r. On $N^{-\alpha} \le |r| < \delta$, Theorem 11.10(c) bounds the full two-point integrand by $C\sqrt{N}\, e^{-Nr^2/4}$ times the product of the one-point integrands, using the overlap density together with the regression and determinant estimates. For the factorization, we need the expected product of the two Hessian determinants under two-point conditioning. We write each Hessian as a bulk block bordered by one row and a corner entry and compare the bulk determinants with a maximal correlation inequality for Gaussian vectors, whose constant does not depend on the dimension.

Combining the three ranges, we find that $\mathbb{E}C_N(L)^2 \le (1 + o(1))(\mathbb{E}C_N(L))^2 + (1 + K)\mathbb{E}C_N(L)$, with a constant K that does not depend on L (Theorem 11.13). Since $\mathbb{E}C_N(L) \to e^{c_0 L} - e^{-c_0 L}$, Chebyshev's inequality implies that the window contains a retained maximum with probability at least $1 - (1 + K)/(e^{c_0 L} - e^{-c_0 L}) - o(1)$ (Theorem 11.14).

Gibbs concentration on bands

We prove the concentration property in Section 9 for pure models (Theorem 9.2) and in Section 12 for non-pure mixtures (Theorem 12.22). For pure models at sufficiently low temperature, this concentration was proved by Subag [141, Theorem 1]; we need it at every temperature of the condensed phase and in a fixed window. For non-pure mixtures, the three inputs below verify Proposition 4.9. The pure proof instead starts from a coarse cap cover, lets the energy deficit per spin tend to zero, removes the caps around critical

points with a large energy deficit, and passes from the caps of a fixed window to bands of width of order $N^{-1/2}$ (Lemma 9.22 and Section 9.9).

The first step is a coarse cover: for every fixed energy tolerance $\tau > 0$, the caps (in pure models, Proposition 9.16) or the bands (in non-pure mixtures, Proposition 12.10) around the critical points of H_Q with value at least $N(E_0 - \tau)$ carry almost all of the Gibbs mass. By Theorem 6.3, among a fixed number of additional independent replicas, a Gibbs sample typically has n whose mutual overlaps and overlaps with it are close to Q. Using Subag's estimates for the free energy of replicated bands with a fixed number of replicas [143, Lemmas 17–19 and Proposition 24], we show that the normalized empirical mean of such a group has value close to NE_0 under H_Q (Proposition 9.15 and Lemma 12.7). Writing $\bar{\sigma} = n^{-1}\sum_i \sigma^i$ and $\bar{q} = |\bar{\sigma}|^2/N$, we have $\bar{q}$ close to $Q + (1-Q)/n$, since the diagonal of the overlap matrix enters the mean. The point $u = \bar{\sigma}/\sqrt{\bar{q}}$ lies on S_N, and $H_Q(u) = H_N(\sqrt{Q/\bar{q}}\,\bar{\sigma})$. We therefore choose n large, so that $\bar{q}$ is close to Q, but fixed before $N \to \infty$. A localization lemma then moves u to a nearby critical point of H_Q, and the given sample lies in its band (for pure models, in its cap of a fixed latitude). In the pure proof, compactness gives a number of additional replicas that is uniform over the limiting overlap laws. In the mixed proof, this number is chosen after passing to a limiting overlap law and may depend on that subsequence; the subsequence argument then proves the full-sequence coarse-cover limit. In both cases, it is fixed before the dimension tends to infinity.

For non-pure mixtures, the coarse band family can contain exponentially many critical points, and the next two inputs pass to a fixed window. Let $Z_H(D) = \int_D e^{\beta H}\,\mathrm{d}\nu_N$. Suppose that a center u in the window $[m_N - L, m_N + L]$ satisfies $Z_H(B_u) \geq \delta D_N e^{\beta(H_Q(u)-m_N)}$, for a deterministic scale D_N. Then $Z_{N,\beta} \geq \delta e^{-\beta L} D_N$, and on the event that $\max H_Q \leq m_N + M$, we have

$$\pi\big(B_{N,M}^c\big) \leq \pi\Big(\Big[\bigcup_{u \text{ coarse}} B_u\Big]^c\Big) + \frac{e^{\beta L}}{\delta D_N} \sum_{\substack{u \text{ coarse} \\ H_Q(u) < m_N - M}} Z_H(B_u) \tag{2.11}$$

(Remark 12.23). In the second step, we bound the expected sum. In the third step, we show that light bands, those with $Z_H(B_u) < \delta D_N e^{\beta(H_Q(u)-m_N)}$, are rare, so that a center as above exists.

For the bands of centers below the window, we use the Kac–Rice formula with the band partition function as a weight. We show that the Palm expectation of $Z_H(B_u)$ at a center of value E, that is, its conditional mean weighted by the determinant, is at most $C_1 D_N e^{\beta(E-m_N)+C_2 N(E/N-E_0)^2}$ (Proposition 12.15). The density of centers at value $m_N - v$ is at most $Ce^{c_+ v}$, with $c_0 < c_+ < \beta$ (Lemma 12.16). We choose the coarse energy width $0 < \tau \leq \tau_0 = \min\{\varepsilon_0, (\beta - c_+)/(4C_2)\}$, as in Corollary 12.17. This absorbs the quadratic correction into the decay in the energy deficit, and the expected sum in (2.11) is at most $CD_N e^{-a_0 M}$, with $a_0 = (\beta - c_+)/2 > 0$. This is the rigorous form of the heuristic

$e^{c_0 v} e^{-\beta v}$ above, and it is the only place where $c_0 < \beta$ enters the concentration argument for non-pure mixtures. The difficulty is that the upper bound must hold without any power of N, since M is fixed while $N \to \infty$. We write the Palm expectation as the ratio of a Laplace integral over the radial derivative and the latitude to a Laplace integral over the radial derivative alone. In the ratio, one latitude width $N^{-1/2}$ remains, and it cancels the growth of the constant $(N/2\pi)^{1/2}$ in the coarea formula over latitudes. The determinant prefactors are bounded (Lemmas 12.11 and 12.14).

The bound on light bands rests on the structure of the field on the band. At latitude exactly $\sqrt{Q}$, the field is the prescribed value plus a residual field with covariance $N\psi_Q$, whose law does not depend on the marks. Since the residual model is strictly replica symmetric with $\beta^2 \psi_Q''(0) < 1$, the theorem of Huang, Montanari, and Pham [93, Lemma 3.2 and Remark 3.3] implies that its normalized partition function is tight away from zero (Lemma 3.19). We pass from this latitude to the band with Jensen's inequality over a symmetric latitude interval of width $N^{-1/2}$ (Lemma 12.19). The Hessian determinant is correlated with the field on the band. We transfer the bound to the Palm law with the normalized second moment of the determinant, which needs no information on this correlation (Proposition 12.20). Then the expected number of light bands in each fixed window is small (Corollary 12.21), which is the remaining hypothesis of Proposition 4.9.

For pure models, the radial derivative is a multiple of the value, $K_Q = pH_Q$, and the conditioning involves the value and the gradient only. We show that the band partition functions at a critical point have normalized second moments bounded by a constant and a lower tail of order $1/T$ at level e^{-T}. The proof is an interpolation argument that needs an exponential moment of the covariance of the field on the band (Lemma 9.18 and Proposition 9.19). This moment is finite because the residual coefficient satisfies $\beta^2 \psi_Q(1) < 1/2$ for every $p \geq 4$ and $\beta > \beta_s(p)$ (Lemma 5.22). The contribution of the bands around critical points with a large energy deficit decays exponentially in the deficit, at a rate that is positive because $x < 1$ and the cap latitude is chosen close to $\sqrt{Q}$ (Lemma 9.3 and the proof of Theorem 9.2).

The conditioned hitting estimate

The hitting estimate for the caps under critical-point conditioning, Theorem 8.2, is the main dynamical input of Part II. For a finite mixture of minimum degree $d \geq 4$, fixed $Q \in (0,1]$, $\beta \geq 0$, and fixed bounds $\bar{h}, \bar{k}$ on the marks, there exists $a_* > 0$ such that for every $a \in (0, a_*]$ there are $\gamma, c, C > 0$ for which, for all sufficiently large N,

$$\mathbb{E}_{h,k,u}\Big[\mathbb{P}_{\nu_N}^{H_N}\Big(\sup_{0 \leq t \leq e^{\gamma N}} |R(X_t, u)| \geq a\Big)\Big] \leq C\big(N^{1-d/2} + N^{3-d}\big) + e^{-cN}, \tag{2.12}$$

uniformly in u and admissible marks (h,k), meaning $|h| \leq \bar{h}$, $|k| \leq \bar{k}$, and $k = ph$ when the mixture is pure of degree p. Here the diffusion is the original Langevin diffusion,

run in the conditioned field. Section 8.2 explains the proof in detail; here we describe its main points.

We begin with the unconditioned field. For an axis u that is uniform on S_N and independent of the field, the mean Gibbs mass of $C_a(u)$ is $\int \nu_N\{u : |R(x,u)| \geq a\}\, \pi(\mathrm{d}x)$. The rotation invariance of ν_N implies that this equals $\nu_N(C_a(v))$ for any fixed $v \in S_N$, which is exponentially small, for every realization of the field. The conditioning changes the whole field. By Gaussian regression, the conditioned field is $H_N + \delta_{h,k,u}$, where the correction is explicit and of order N near u (Proposition 8.7). We cannot treat it as a small perturbation. However, it vanishes to high order at the equator of u. The following correction bound is on the regularity event Ω_N of Lemma 8.6, whose complement has probability at most e^{-N}: with $r = R(x,u)$ and $V_x(u) = N^{-1/2}\nabla_{S_N} H_Q(u) \cdot x$, we have

$$|\delta_{h,k,u}(x)| \leq CN\,|r|^d + C\sqrt{N}\,|r|^{d-1}\,|V_x(u)|\,. \tag{2.13}$$

A fixed point x has overlap of order $N^{-1/2}$ with a uniform axis, and $V_x(u)$ is of order one. We see that the correction is of order $N^{1-d/2}$, and this is the source of both polynomial errors in (2.12). Fix the field in Ω_N. For a sufficiently small fixed $\alpha > 0$, let χ be an even smooth cutoff equal to one on $[-1,1]$ and zero outside $[-2,2]$, set $s = \nabla_{S_N} H_Q(u) \cdot x/N$, and define $\delta_\alpha(u,x) = \chi(r/\alpha)\chi(s/\alpha)\delta_{h,k,u}(x)$ and $w(u,x) = \beta\delta_\alpha(u,x)$. We compare the dynamics in this truncated field, at a uniform axis, with the unconditioned dynamics. Only two averaged moments enter, and they hold uniformly in x, the field in Ω_N, and the admissible marks (Proposition 8.10):

$$\eta = \mathbb{E}_u\,|e^w - 1| = O\big(N^{1-d/2}\big), \qquad \epsilon = \mathbb{E}_u\big[e^w(1-e^{-w})^2\big] = O\big(N^{2-d}\big). \tag{2.14}$$

There is an apparent difficulty here, since the point x in these averages may depend on the field, and so does the gradient $\nabla_{S_N} H_Q(u)$ in $V_x(u)$. The way out is to fix the field and x and to average u over each latitude $\{u : R(x,u) = r\}$. On a latitude, V_x concentrates by the logarithmic Sobolev inequality on the sphere [16], and the Funk–Hecke formula [7] gives its mean exactly (Lemmas 8.8 and 8.9 and Proposition 8.10).

The main obstacle is the length $e^{\gamma N}$ of the time interval, since a comparison of path laws, for instance by Girsanov's formula, costs an amount proportional to the time. To overcome this obstacle, we compare resolvents instead. Let τ be the killing time for a killing rate κ supported near the caps. The entropy–resolvent comparison of Lemma 8.11 bounds the Laplace transform $\mathbb{E}e^{-\lambda\tau}$ for the perturbed process by the integrated killing rate against a resolvent of the unperturbed process, plus a cost

$$\eta + \sqrt{\epsilon} + \epsilon\,(1 + \mathcal{H}_0) = O\big(N^{1-d/2} + N^{3-d}\big), \tag{2.15}$$

where $\mathcal{H}_0$ is an entropy of the density of ν_N with respect to the Gibbs measure. Note that $\mathcal{H}_0$ is of order N, since $|H_N| \leq KN$ with overwhelming probability. The point is that

the cost is not multiplied by the time: with $\lambda = e^{-\gamma N}$, the factor $e^{\gamma N}$ multiplies only the integrated killing rate, which is exponentially small when γ is small. The comparison cost other than the killing term depends on the initial density f_0 through $\mathcal{H}_0$; the killing term also involves its unperturbed resolvent $\lambda(\lambda + A_0)^{-1} f_0$, where A_0 is the generator of the unperturbed dynamics. A commutator identity in the proof cancels a cross term that would otherwise be of order one. For $d = 4$, both terms in (2.15) are of order N^{-1}. For $d = 3$, the second term is of order one, and this is where we need the hypothesis on the minimum degree.

Two further steps return to the true process. Recall that we have truncated the correction at a fixed scale α in both r and s, since the comparison requires a bounded perturbation. The truncated correction is at most $CN\alpha^d$, while the cap of latitude $\alpha/4$ has mass $e^{-cN\alpha^2}$. Since $d > 2$, we can choose α small, and then γ small. Except with exponentially small probability, a hit of the cap forces the smooth killing rate to act during a time interval of fixed length, and this converts the killing time into a hitting time (Proposition 8.12). The true and truncated processes coincide until $|r|$ or $|s|$ reaches α. Reaching $|r| = \alpha$ requires crossing the cap of latitude $a = \alpha/2$ first. Reaching $|s| = \alpha$ is exponentially unlikely, because the gradient $\nabla_{S_N} H_Q(u)$ is exactly the Gaussian vector that the conditioning removes by regression. In the unrestricted regression coupling, conditionally on u, this tangent gradient is independent of the conditioned field and its whole trajectory. Conditionally on u and the trajectory, s_t is centered Gaussian with variance at most a_Q/N at each fixed time. The resulting exponential bound for the gradient cutoff also holds after restricting to Ω_N, by bounding the unrestricted probability and dividing by $\mathbb{P}(\Omega_N)$ (Lemma 8.13). Finally, we pass from a uniform axis to every deterministic axis by rotation invariance.

Other initial laws

We deduce Theorem 1.4 from Theorem 1.2 and the polynomial bound of Proposition 13.1, by three arguments that use only the rotation invariance of the disorder (Section 4.5). For part (i), let μ be a deterministic initial law. The expected distance $\mu \mapsto \mathbb{E}\,\|\mu P_T - \pi\|_{\mathrm{TV}}$ is convex, and it is invariant under rotations of μ, because we can compensate a rotation of the initial law by a rotation of the field, whose law is invariant. The rotation average of μ is ν_N, and convexity implies that

$$\mathbb{E}\big[1 - \|\mu P_T - \pi\|_{\mathrm{TV}}\big] \le \mathbb{E}\big[1 - \|\nu_N P_T - \pi\|_{\mathrm{TV}}\big] \tag{2.16}$$

(Theorem 4.11). This is an instance of the classical symmetrization over a compact group, the same averaging that underlies the Hunt–Stein theorem [113, Section 8.5]. In particular, ν_N is the hardest case among the initial laws chosen independently of the disorder.

For parts (ii) and (iii), the same invariance shows that the expected mass of the band union at time T_N is the same from every starting point (Lemma 4.13). Proposi-

tion 13.1 implies that this mass is at most $q_{N,M} \le C_{M,r} N^{-(d-2)(1-1/r)/2}$. We handle a field-dependent choice among $K_N \le C_0 N^{\kappa}$ prescribed laws by a union bound, since $K_N q_{N,M} \to 0$ when $\kappa < (d-2)/2$ and r is large (Corollary 4.14). The loss of the factor $1 - 1/r$ in the Hölder step is why we exclude the endpoint $\kappa = (d-2)/2$. We handle a field-dependent law μ with small relative entropy with respect to ν_N by binary data processing. Let q be the mass of the band union under $\nu_N P_{T_N}$. Then its mass under μP_{T_N} is at most $(D(\mu\|\nu_N) + \log 2)/\log(1/q)$ (Lemma 3.9 and Corollary 4.15). For any fixed $r > 1$ and $0 < \tau < (d-2)(1-1/r)/2$, we have $q \le N^{-\tau}$ with probability tending to one, by Markov's inequality. The entropy bound then tends to zero in disorder probability when $D(\mu\|\nu_N)/\log N \to 0$ in that probability mode.

2.3 Below the static threshold

Let $\beta_{\mathrm{d}} < \beta < \beta_{\mathrm{s}}$, and recall the plateau value q_* of (1.2). The proof of Theorem 1.5 has a probabilistic part, in which we describe the Gibbs measure as seen from a Gibbs-typical point, and a deterministic part, in which we build one observable from this description and derive the five conclusions from it.

The constrained pressure and the equality at q_*

For a reference point x and an overlap r, consider the Gibbs mass of the configurations u with $R(x,u)$ close to r. For a reference point x drawn from π, the constrained pressure F_{FP} of Definition 14.4 is the Franz–Parisi potential written as a pressure. With $F = \beta^2\xi(1)/2$, for a fixed $h > 0$ such that $[r-h, r+h] \subset (-1,1)$, the leading exponential form under the Gibbs-reference law is

$$\pi\big\{u : R(x,u) \in [r-h, r+h]\big\} \approx \exp\Big\{N\Big[\max_{v \in [r-h,r+h]} F_{\mathrm{FP}}(v) - F\Big]\Big\}. \tag{2.17}$$

The midpoint exponent $F_{\mathrm{FP}}(r) - F$ is recovered by sending h to zero after the limit in N (Proposition 14.7). A direct computation is difficult, because x depends on the disorder. Following El Alaoui, Montanari, and Sellke [65], who treated pure models, we work under a planted law, in which x is uniform on S_N and the field is a Gaussian field plus the drift $\beta N \xi(R(x,\cdot))$. Below the static threshold, events of exponentially small probability under the planted law have exponentially small probability under the joint law of the field and a reference point drawn from π (Lemma 14.3). Let $\eta_r(t) = \xi(r^2 + (1-r^2)t) - \xi(r^2)$. Under the planted law, with $n = N-1$, the field on the section $\{u : R(x,u) = r\}$ has the law of $\beta N\xi(r) + W + Y$, where W is centered Gaussian with variance $N\xi(r^2)$ and is independent of the transverse spherical field Y, whose covariance is $N\eta_r(R_n) = n(N/n)\eta_r(R_n)$ (Lemma 14.5). The Parisi formula then gives

$$F_{\mathrm{FP}}(r) = \beta^2\xi(r) + \frac{1}{2}\log(1-r^2) + \mathcal{P}_\beta(\eta_r) \tag{2.18}$$

(Proposition 14.7). For $r \neq 0$, the mixture η_r has a positive term of degree one, even though ξ has none; at $r = 0$, we have $\eta_0 = \xi$. The Parisi formula of Chen [46] allows the degree-one term.

The infimum $\mathcal{P}_\beta(\eta_r)$ is not explicit, but an upper bound suffices. For the trial bound, we restrict to $r \in [0, 1)$. Evaluating the Crisanti–Sommers functional of η_r at the point mass $\delta_{r/(1+r)}$, we obtain $F_{\mathrm{FP}}(r) \le \Psi(r)$ (Lemma 14.9), where

$$\Psi(r) = F + \frac{1}{2}\Phi_\beta(r), \qquad \Psi'(r) = \frac{r}{2(1-r)}\bigl(\beta^2 A(r) - 1\bigr). \tag{2.19}$$

For pure models, this is the trial bound of [65, Proposition 3.9]. By the definition of q_*, the trial pressure increases on an interval $(q_* - \varepsilon_0, q_*)$ and does not increase on $[q_*, 1)$. The new point is that $F_{\mathrm{FP}}(q_*) = \Psi(q_*)$ for every finite mixture (Proposition 14.11). This amounts to showing that δ_{r_0}, with $r_0 = q_*/(1 + q_*)$, minimizes the Crisanti–Sommers functional of η_{q_*}, which we check with the Crisanti–Sommers optimality criterion. Above r_0, the derivative of the first variation has the sign of $\beta^2 A - 1$ to the right of q_*, which is nonpositive. Below r_0, we need that $\xi'(v)/(v - q_*^2)$ is nonincreasing on $(q_*^2, q_*]$. We deduce this from the monotonicity of the ratio $\chi(v) = v\xi''(v)/\xi'(v)$, whose derivative is the variance of $p - 1$ under the weights $pc_p v^{p-1}/\xi'(v)$, divided by v, together with $A'(q_*) \le 0$. We allow q_* to be a degenerate root of the plateau equation.

Fix an admissible collar $0 < q_1 < q_2 < q_*$, so that $\beta^2 A(r) > 1$ on (q_1, q_*). Since Ψ increases on $[q_1, q_*]$ and does not increase on $[q_*, 1)$, the equality at q_* implies two bounds. On the cap $[q_1, 1)$, the constrained pressure is at most $F - J$, with equality at q_*, where $J = -\Phi_\beta(q_*)/2 > 0$. On the collar $[q_1, q_2]$, it is at most $F - J - G$, where $G = \Psi(q_*) - \Psi(q_2) > 0$. After fixing the collar, choose $0 < \delta < \min\{J, G\}/8$. Using the planted law and likelihood transfer, we turn these limits into quenched estimates (Proposition 14.13). With probability at least $1 - Ce^{-aN}$, there exists a set $\mathcal{G}$ of good reference points with $\pi(\mathcal{G}^c) \le e^{-aN}$ such that, for every $x \in \mathcal{G}$,

$$\pi\{u : R(x,u) \ge q_2\} \ge e^{-N(J+2\delta)}, \qquad \pi\{u : q_1 \le R(x,u) \le q_2\} \le e^{-N(J+G-2\delta)}. \tag{2.20}$$

In words, around a Gibbs-typical point there is a region of Gibbs mass about e^{-NJ} at overlap close to q_*, and it is separated from the rest of the Gibbs measure by a collar of exponentially smaller mass.

The second input is a uniform bound on the Gibbs masses of caps. By the replica concentration of Huang, Montanari, and Pham [93, Lemma 7.7], for each fixed $s > 0$, two independent samples from π have overlap larger than s in absolute value with Gibbs probability at most $2e^{-cN}$, on an event of probability at least $1 - e^{-cN}$ (Proposition 14.14). We convert this into a bound for every cap, uniformly over its axis, with an elementary barycenter estimate (Lemma 14.15):

$$\sup_{v \in S_N} \pi\{x : R(x,v) \ge s\} \le \Bigl[\frac{2 - s^2}{s^2}\,(\pi \otimes \pi)\bigl\{R > s^2/2\bigr\}\Bigr]^{1/2}. \tag{2.21}$$

The axis may depend on the field, and we need no union bound over axes. In the same way, we find that $|\pi(U)|^2 \le s + e^{-cN}$ (Corollary 14.16).

The overlap-weighted Gibbs mean

The estimates (2.20) suggest an observable that records the region around x. Let $h\colon [-1,1] \to [0,K]$ be smooth and nonincreasing, with $h = K$ on $[-1,q_1]$ and $h = 0$ on $[q_2,1]$, and let $\psi_u(x) = e^{-Nh(R(x,u))}$. The overlap-weighted Gibbs mean (Definition 15.6) is

$$B_N(x) = \int U(u)\, w_x(\,\mathrm{d}u), \qquad w_x(\,\mathrm{d}u) = \frac{\psi_u(x)^2\, \pi(\,\mathrm{d}u)}{\int \psi_{u'}(x)^2\, \pi(\,\mathrm{d}u')}. \tag{2.22}$$

In words, $B_N(x)$ is the mean of U under the Gibbs measure reweighted by a cutoff that suppresses the points u with $R(x,u) < q_1$. We call w_x the overlap-reweighted Gibbs measure seen from x. Note that $|B_N| \le 1$.

We show that B_N has two properties at the points $x \in \mathcal{G}$, and both follow from (2.20) (Lemma 15.7). The first is that its gradient is small. Differentiating (2.22), we find that $\nabla B_N(x)$ is twice the w_x-mean of $(U(u) - B_N(x)) \otimes \nabla \log \psi_u(x)$. The factor $\nabla \log \psi_u(x)$ vanishes unless u lies in the collar of x, and only the w_x-mass of the collar counts. With $d_N(x)$ the denominator in (2.22), we obtain

$$\|\nabla B_N(x)\|_{\mathrm{HS}}^2 \le 4N \left\|h'\right\|_\infty^2 \frac{\pi\{u : q_1 \le R(x,u) \le q_2\}}{d_N(x)} \le 4N \left\|h'\right\|_\infty^2 e^{-N(G-4\delta)}, \tag{2.23}$$

where we used $d_N(x) \ge \pi\{u : R(x,u) \ge q_2\}$ and (2.20). The second property is alignment. The references with $R(x,u) \le q_1$ carry w_x-mass at most $e^{-2KN}/d_N(x)$, and with $2K = J + 2\delta + 1$ we find that $U(x) \cdot B_N(x) \ge q_1 - (1+q_1)e^{-N}$ (Section 16.2).

The definition averages over all references u, including those outside $\mathcal{G}$, at which the estimates (2.20) may fail. The function B_N is then smooth on all of S_N, and $\|\nabla B_N\|_{\mathrm{HS}}^2 \le 4N \|h'\|_\infty^2$ at every point, because $|\nabla \psi_u|^2 \le N \|h'\|_\infty^2 \psi_u^2$ and w_x is a probability measure (Lemma 15.7). The points x outside $\mathcal{G}$ therefore enter the vector energy only through their Gibbs mass:

$$\mathcal{E}_{\mathrm{vec}}(B_N) \le 4N \left\|h'\right\|_\infty^2 \left(e^{-N(G-4\delta)} + e^{-aN}\right). \tag{2.24}$$

From one observable to five conclusions

On the event of Proposition 14.13 and Corollary 14.16, the rest of the proof is deterministic (Sections 15 and 16). We list in Table 2.1 the tool that we use for each conclusion of Theorem 1.5.

For (i), we use the averaged variational principle of Harrell and Stubbe [86, Theorem 3.1] in the following form (Lemma 15.1). Suppose that a family of nonnegative trial functions f_z, indexed by a measure space, has total squared $L^2(\pi)$ mass

Conclusion	Tool	Result
(i) many slow modes	averaged variational principle applied to translated bumps of B_N; bound on caps	Lemmas 15.1 and 15.2
(ii) projection onto slow modes	spectral projection of Z_N against $B_N - \pi(B_N)$; small magnetization	Lemma 15.5
(iii) correlation plateau	spectral calculus, from (ii)	Lemma 15.5
(iv) Gibbs-typical starts	displacement bound and Markov's inequality; bound on caps	Lemma 15.8
(v) endpoint information	binary data processing and the chain identity, from (iv)	Lemmas 3.9 and 3.12

Table 2.1. The five conclusions of Theorem 1.5 and the deterministic results that derive them from the energy and the alignment of the overlap-weighted Gibbs mean.

$0 < \mathsf{T} = \int \|f_z\|_2^2 \, \mu(\mathrm{d}z) < \infty$ and averaged energy at most $r\mathsf{T}$, that each function satisfies $\pi(f_z) \le \mathsf{R}$, and that the family covers each point with total weight at most C. Then the number of eigenvalues below $\Lambda > r$ is at least $(1 - r/\Lambda)\mathsf{T}/(\mathsf{RC})$. We apply this to the bumps $f_z(x) = \alpha(N[1 - |B_N(x) - z|^2/\rho^2])$, with z averaged over a ball of $\mathbb{R}^N$ (Lemma 15.2). A bump is supported where B_N is within ρ of z. Since B_N is aligned with U on $\mathcal{G}$, the part of this set in $\mathcal{G}$ lies in a cap of a fixed latitude, and (2.21) bounds its Gibbs mass by e^{-cN}. With the index measure of Lemma 15.2, we have $\int f_z(x) \, \mu(\mathrm{d}z) \le 1$ at every x and $1/2 \le \int \|f_z\|_2^2 \, \mu(\mathrm{d}z) \le 1$. The averaged energy of the bumps is at most $CN^2\rho^{-2}\mathcal{E}_{\mathrm{vec}}(B_N)$, which is exponentially small by (2.24). This shows that at least e^{cN} eigenvalues lie below $e^{-2\gamma N}$. Unlike the smoothed indicators of disjoint clusters used in [65], the bumps overlap, and the averaged principle needs only the bounds R and C.

For (ii) and (iii), let $Z_N = U - \pi(U)$, $\overline{B}_N = B_N - \pi(B_N)$, and $\Pi_\Lambda = \mathbf{1}_{(0,\Lambda]}(-L)$. The part of $\overline{B}_N$ above Λ has squared norm at most $\mathcal{E}_{\mathrm{vec}}(B_N)/\Lambda$, and we find that

$$\|\Pi_\Lambda Z_N\|_2 \, \|\overline{B}_N\|_2 \ge \langle Z_N, \overline{B}_N \rangle - \big(\mathcal{E}_{\mathrm{vec}}(B_N)/\Lambda\big)^{1/2} \tag{2.25}$$

(Lemma 15.5). The alignment implies that $\langle U, B_N \rangle \ge q_1 - o(1)$, and the magnetization $\pi(U)$ is small by Corollary 14.16. Together, they give $\langle Z_N, \overline{B}_N \rangle \ge q_0$ for a fixed $q_0 < q_1$ close to q_1, while $\|\overline{B}_N\|_2 \le 1$. With $\Lambda = e^{-2\gamma N}$, we obtain $\|\Pi_\Lambda Z_N\|_2^2 \ge q_0^2 - o(1)$. For every $b < q_*^2$, we can choose the collar and q_0 so that $q_0^2 > b$, and this proves (ii). The connected correlation satisfies $\langle Z_N, P_t Z_N \rangle \ge e^{-\Lambda t} \|\Pi_\Lambda Z_N\|_2^2$, which implies (iii) up to time $e^{\gamma N}$.

Conclusion (iv) follows from (2.2) and Markov's inequality. Let $T = e^{\gamma N}$. For all x outside a set of Gibbs mass e^{-cN}, we find that the law $P_T(x, \cdot)$ gives mass at least $1 - e^{-\gamma N}$ to the set $D_x = \{y : |B_N(y) - B_N(x)| \le \rho\}$. For $x \in \mathcal{G}$, the set D_x lies in the

union of $\mathcal{G}^c$ and a cap around the direction of $B_N(x)$, and its Gibbs mass is at most $w = e^{-cN}$. Testing D_x at time T, we obtain $\|P_T(x,\cdot) - \pi\|_{\mathrm{TV}} \geq 1 - e^{-\gamma N} - w$, and the contraction of total variation gives the same lower bound for every $0 \leq t \leq T$. For (v), by binary data processing at time T, we have $D(P_T(x,\cdot)\|\pi) \geq (1 - e^{-\gamma N}) \log(1/w) - \log 2$. The contraction of relative entropy gives this bound for every $0 < t \leq T$, and it is at least cN for all sufficiently large N (Lemma 15.8). We conclude with the chain identity $I(X_0; X_t) = \int D(P_t(x,\cdot)\|\pi)\, \pi(\mathrm{d}x)$ of Lemma 3.12, which goes back to the golden formula [127, Theorem 4.1].

Correlation bounds at every level below q_*

The proof above gives levels below q_*^2. The physics prediction uses the largest positive solution of the plateau equation, which equals q_* except when $\beta^2 A$ touches one from below at an isolated point farther to the right (see Section 1.1). Theorem 1.6 gives lower bounds approaching q_*. The loss of the square occurs in (2.25), where we used $\|\overline{B}_N\|_2 \leq 1$. Recall from (2.22) that $|B_N(x)|^2$ is the mean overlap of two independent samples from w_x, which concentrate at overlap about q_* with x. If their mutual overlap also concentrates at q_*, then $\|\overline{B}_N\|_2^2 \leq q_* + o(1)$. Combining this with (2.25), for a fixed sufficiently small $0 < \vartheta < \min\{1, q_1/3\}$ and a time exponent γ chosen as in Section 17.2, the norm, magnetization, and energy estimates give, with probability tending to one,

$$\|\Pi_{\Lambda_N} Z_N\|_2^2 \geq \frac{(q_1 - 3\vartheta)^2}{q_* + \vartheta}, \qquad \Lambda_N = e^{-2\gamma N}. \tag{2.26}$$

Choosing q_1 close to q_* and then ϑ small makes this lower bound exceed any fixed $b < q_*$. The main step of Section 17 is therefore the estimate $\int |B_N|^2 \, \mathrm{d}\pi \to q_*$ in probability (Proposition 17.1).

We prove it under the planted law. We write each u through its latitude $r = R(x,u)$ and its transverse part y, a point of the sphere in $x^\perp$. For two such points, we have $R(u,u') = rr' + ((1-r^2)(1-r'^2))^{1/2} R_n(y,y')$. The latitude concentrates at q_* under w_x, by the constrained pressure. It remains to show that the transverse overlap $R_n(y,y')$ concentrates at $t_0 = q_*/(1+q_*)$, since $q_*^2 + (1-q_*^2)t_0 = q_*$. By the optimality of the replica-symmetric trial at q_*, the Parisi measure of the transverse spin glass at latitude q_* is the point mass at t_0 (Lemma 17.5). Note that the transverse covariance η_{q_*} contains the monomials t and t^2 with positive coefficients. We multiply two Gaussian components of the field that contribute to these monomials by amplitudes s_1 and s_2 and differentiate the expected logarithm of the reweighted partition function in these amplitudes. Using Gaussian integration by parts, we find that the derivative in s_k at $s = (1,1)$ is, once the latitude has concentrated, an affine function of the k-th moment of the transverse overlap (Lemma 17.9). We compute its limit from the Crisanti–Sommers formula and the envelope theorem [118]: it equals $\beta^2 a_k^2 (1 - t_0^k)$, where $a_k^2 > 0$ is the coefficient of t^k contributed by

the k-th component (Lemma 17.8). Convexity in the amplitudes lets us pass from the convergence of the pressures to the convergence of their derivatives [131, Theorem 24.5]. We find that the first two moments of the transverse overlap tend to t_0 and t_0^2, and that its variance tends to zero. This use of derivatives in the coefficients of a covariance is classical in the theory of mean-field spin glasses [122]; see [13, Proposition 2.2] for spherical mixtures. Finally, since (ξ, β) is strictly replica symmetric with $\beta^2 \xi''(0) < 1$, the planted law and the Gibbs-reference law are mutually contiguous [93, Lemma 3.2, Remark 3.3, and Corollary 3.5], which transfers the estimate to the original disorder. Neither the convex derivative convergence nor the likelihood transfer supplies a rate. The derivative argument does not give exponentially small planted failure probabilities, so the exponential transfer of Lemma 14.3 cannot be applied; this is why (ii) and (iii) of Theorem 1.6 are proved only with probability tending to one.

Eigenvalues near the dynamical threshold

The proof of Theorem 1.7 reuses the estimates above.

For Theorem 1.7, we count slow modes with smooth cap cutoffs $f_u(x) = \varphi(R(x,u))$ as trial functions, where φ is smooth, vanishes on $(-\infty, q_1]$, and equals one on $[q_2, \infty)$, indexed by good references with measure $\mathbf{1}_{\mathcal{G}}(u)\pi(\mathrm{d}u)$ (Remark 16.1 and Lemma 18.8). The exponent of the eigenvalue bound can then be any number below the collar gap G, which we compute as $\beta \downarrow \beta_{\mathrm{d}}$. Near the largest maximizer q_d of A, we write $\varepsilon = \beta^2 A_{\max} - 1$ and $r = q_d + hz$ with $h = (\varepsilon/a)^{1/(2m)}$. On the component of $\{\beta^2 A > 1\}$ that ends at q_*, (2.19) shows that, heuristically, $\Psi'(r) \approx q_d \varepsilon (1 - z^{2m})/(2(1 - q_d))$. Integrating over $z \in [-1, 1]$, we find

$$\Psi(q_*) - \Psi(q_-) \approx \frac{q_d}{2(1-q_d)}\, h\varepsilon \int_{-1}^{1} (1 - z^{2m})\, \mathrm{d}z = C_d\, \varepsilon^{1+1/(2m)}, \tag{2.27}$$

where q_- is the left end of this component. The precise forms are (18.22) and, for the collar gap, (18.23). In the counting bound, the core exponent J enters both T and R and cancels, and the covering constant C is the largest Gibbs mass of a cap. We use a cap bound whose exponent is uniform in the temperature (Lemma 18.5). The number of slow modes then stays $e^{\kappa_* N}$, with κ_* independent of ε, while the exponent of the eigenvalue bound tends to zero. Here $\varepsilon > 0$ is fixed before $N \to \infty$; the probability exponent and the threshold for N may depend on ε and ϑ. The computed barrier is that of the trial pressure, and it yields a lower bound for the constrained-pressure gap, rather than an asymptotic formula for that gap or for the eigenvalues.

2.4 Dependencies among the main steps

We list in Tables 2.2 and 2.3 the principal proof inputs for each main result, including selected intermediate results, in the order discussed above. We suppress most general

preliminary facts, but retain the Crisanti–Sommers optimality criterion where it is a central variational input.

Result	Content	Uses
Theorem 1.2	separation from the uniform law	Proposition 13.1, Proposition 5.11 and Lemma 5.10; counts and concentration Proposition 9.1 and Theorem 9.2 in pure models, or Theorem 12.22 with the counts of Theorem 11.14 and Lemma 12.2 in non-pure mixtures
Theorem 1.4	other initial laws	Theorem 1.2, Proposition 13.1, Theorem 4.11, Corollary 4.14, Corollary 4.15
Proposition 13.1	a common time horizon for all fixed windows	Proposition 4.5, Corollary 8.3, Theorem 7.23
Theorem 7.23	determinant-weighted transfer	Theorem 7.17, Proposition 7.7, Lemma 7.22
Theorem 8.2	conditioned hitting estimate	Lemma 8.6, Proposition 8.7, Proposition 8.10, Lemma 8.11, Proposition 8.12, Lemma 8.13, Lemma 4.2
Proposition 5.11	static consequences of strict 1RSB	Lemma 3.15, Theorem 5.7, Theorem 5.5
Theorem 6.3	directing vectors have squared norm Q	Proposition 5.11, Lemma 5.8, Lemma 6.4, Lemma 6.5, Lemma 6.7
Theorem 11.14	counts and existence, non-pure mixtures	Theorem 10.6; exclusion Theorem 10.13 and Proposition 10.15; pair bound Theorem 11.10 and Corollary 11.11; Lemma 11.12 and Theorem 11.13; finite counts Theorem 7.17 and Corollary 7.18
Proposition 9.16, Proposition 12.10	coarse cap cover in pure models; coarse band cover in non-pure mixtures	Theorem 6.3; replica means Proposition 9.15 or Lemma 12.7; localization Proposition 9.12 or Lemma 12.8, with Proposition 10.10 and Lemma 12.2
Theorem 12.22	concentration, non-pure mixtures	Proposition 4.9, Lemma 12.2, Proposition 12.10, Corollary 12.21, Corollary 12.17; measurability and finite critical sets Lemma 12.9 and Theorem 7.17
Theorem 9.2	concentration, pure models	Proposition 9.16 and Lemma 9.22, Lemma 9.9, Proposition 9.19, Proposition 9.20, Proposition 9.21, Lemma 9.8

Table 2.2. Dependencies in Part II.

Result	Content	Uses
Theorem 1.5	slow relaxation of the equilibrium dynamics	Proposition 14.11, Proposition 14.13, Corollary 14.16, Lemma 15.7, Lemma 15.2, Lemma 15.5, Lemma 15.8
Proposition 14.13	bounds on caps, core bands, and collars around good reference points	Lemma 14.6, Proposition 14.7, Proposition 14.11, Lemma 14.3
Proposition 14.11	the equality at q_*	Lemma 14.9, Lemma 3.15
Corollary 14.16	uniform bound on the Gibbs masses of caps	Proposition 14.14, Lemma 14.15
Lemma 15.2	many slow modes from one aligned observable	Lemma 15.1
Theorem 1.6	correlation bounds at every level below q_*	Theorem 1.5, Proposition 17.1, Lemma 17.2, Lemma 15.5, Corollary 14.16
Proposition 17.1	the norm of the overlap-weighted Gibbs mean	Proposition 17.3, Lemma 17.6, Lemma 17.8, Lemma 17.9; contiguity (Section 17.6)
Theorem 1.7	eigenvalues near the dynamical threshold	Lemma 18.2, Lemma 18.3, Lemma 18.4, Lemma 18.6, Lemma 18.8, Lemma 18.5

Table 2.3. Dependencies in Part III.

PART I

Preliminaries

SECTION 3

Preliminaries

This section collects the definitions and general facts that are used in both halves of the paper. Section 3.1 fixes our conventions for constants, probabilities, and the order in which parameters are chosen. Section 3.2 defines the model and records two standard Gaussian facts: the regularity of the field and of its derivatives (Lemma 3.2) and the concentration of free energies over deterministic domains (Lemma 3.3). Section 3.3 introduces the Langevin dynamics on the sphere (Definition 3.4 and Lemma 3.5) and the spectral calculus of its generator (Lemma 3.6). Section 3.4 contains facts that hold for every Markov semigroup with an invariant law. These are the contraction of total variation and relative entropy (Lemma 3.8), binary data processing (Lemma 3.9), an endpoint criterion for separation in total variation (Proposition 3.10), and the chain identity for the mutual information between the two endpoints (Lemma 3.12). Section 3.5 recalls the Crisanti–Sommers functional, the Parisi formula, and the first-variation characterization of the minimizer (Definition 3.13, Theorem 3.14, and Lemma 3.15), and it defines the thresholds β_{d} and β_{s} (Definition 3.16 and Lemma 3.18). Finally, Section 3.6 states two results of Huang, Montanari, and Pham in our notation (Lemmas 3.19 and 3.20).

The facts in Sections 3.3 and 3.4 concern a fixed field, and no randomness enters. They are used in Sections 4 and 13 for the condensed phase and in Sections 15 and 16 below the static threshold. The field estimates of Section 3.2 are used in Part II, in the Kac–Rice formula and in the proof that the Gibbs measure concentrates on bands, and in Section 14. The static facts of Section 3.5 enter Sections 5, 14, and 18. Lemma 3.19 is used in Sections 12 and 17, and Lemma 3.20 in Section 14. We prove the elementary statements and give precise citations for the others. Facts that are used in a single section are stated there.

3.1 Conventions

Throughout, $N \geq 2$ is the dimension. The covariance ξ and the inverse temperature β are fixed at the start of every statement, and all limits are taken as $N \to \infty$ with them fixed. We write C for a large positive constant and c and a for small positive constants. In estimates taken as $N \to \infty$, these constants depend only on ξ, β, and the other parameters fixed in the statement at hand, never on N or on the realization of the field. Their values may change from line to line, and the same convention applies to the constants implicit in

the notation $O(\cdot)$ and $o(\cdot)$. Constants used only in deterministic analytic arguments for a fixed N and a fixed field H may depend on N and H; we then indicate this dependence, as in $C_{r,N,H}$. Logarithms are natural.

Probabilities. We refer to the randomness of the field as the disorder and write $\mathbb{P}$ and $\mathbb{E}$ for its law and expectation. We say that a statement holds *for all sufficiently large* N if there exists N_0, depending only on the fixed parameters, such that it holds for all $N \geq N_0$. We say that an event, defined for each N in terms of the field, holds *with probability at least* $1 - Ce^{-aN}$ if there exist constants $C, a > 0$ such that its probability is at least $1 - Ce^{-aN}$ for all sufficiently large N. It holds *with probability tending to one* if its probability tends to one as $N \to \infty$. A statement of the form "with probability at least $1 - Ce^{-aN}$, the following holds" is quenched. This means that on the event, every listed property holds for the realized field, simultaneously for all times, starting points, and other variables quantified inside the statement, unless the statement says otherwise.

Order of parameters. When a statement or proof involves several parameters, they are chosen in a fixed order. First ξ and β are fixed. Then the accuracy parameters of the statement (for example a plateau level, a tolerance, or the half-width of an energy window) are chosen in the order given in the proof. Then the constants C, c, a, γ, κ are chosen, and they may depend on all of the above. Last, N_0 is chosen. No statement is uniform as β approaches β_{d} or β_{s} unless this is said explicitly. Each proof with more than two parameters states their order.

Measures and information. For a fixed field, we write $\mathbb{P}_\mu$ and $\mathbb{E}_\mu$ for the law and expectation of the diffusion of Definition 3.4 with initial law μ, and $\mathbb{P}_x$, $\mathbb{E}_x$ when $\mu = \delta_x$. For probability measures μ and ν on a measurable space, the total variation distance is $\|\mu - \nu\|_{\mathrm{TV}} = \sup_A |\mu(A) - \nu(A)|$. The relative entropy is $D(\mu\|\nu) = \int \log(\,\mathrm{d}\mu / \,\mathrm{d}\nu)\,\mathrm{d}\mu$ if $\mu \ll \nu$, and $D(\mu\|\nu) = \infty$ otherwise. The mutual information of two random variables X and Y is the relative entropy of their joint law with respect to the product of the marginal laws. For $p, a \in [0,1]$, the binary relative entropy and the binary entropy are

$$d(p\|a) = p\log\frac{p}{a} + (1-p)\log\frac{1-p}{1-a}, \qquad h(p) = -p\log p - (1-p)\log(1-p),$$

with $0\log 0 = 0$ and $d(p\|a) = \infty$ if $a \in \{0,1\}$ and $p \neq a$. We also use $0\log(0/a) = 0$ for $a \in [0,1]$ and $0 \cdot \infty = 0$, while $p\log(p/0) = \infty$ for $p > 0$. All state spaces are standard Borel spaces.

3.2 The model

Definition 3.1 (Model)**.** Let $S_N = \{x \in \mathbb{R}^N : |x|^2 = N\}$, let ν_N be the uniform probability measure on S_N, and for $x, y \in \mathbb{R}^N$ let $R(x,y) = x \cdot y/N$ and $U(x) = x/\sqrt{N}$. A *finite*

mixture is a nonzero polynomial

$$\xi(r) = \sum_{p=2}^{P} c_p r^p, \qquad c_p \geq 0 \quad (2 \leq p \leq P).$$

Its *minimum degree* is $d = \min\{p : c_p > 0\}$, and ξ is *pure* if exactly one coefficient is nonzero and *non-pure* if at least two coefficients are nonzero. The *field* H_N is the centered Gaussian process on S_N with covariance $\mathbb{E}H_N(x)H_N(y) = N\xi(R(x,y))$, realized as in Lemma 3.2. For $\beta \geq 0$, the *Gibbs measure* $\pi = \pi_{N,\beta}^H$ and the *partition function* $Z_{N,\beta}$ are

$$\pi(\,\mathrm{d}x) = \frac{e^{\beta H_N(x)}}{Z_{N,\beta}}\,\nu_N(\,\mathrm{d}x), \qquad Z_{N,\beta} = \int_{S_N} e^{\beta H_N}\,\mathrm{d}\nu_N.$$

The *free energy* is $F_N(\beta) = N^{-1}\mathbb{E}\log Z_{N,\beta}$, and $F(\beta) = \lim_{N\to\infty} F_N(\beta)$. The limit exists by Theorem 3.14.

We allow odd degrees and a quadratic term. Note that the Gibbs measure favors large values of H_N. For a function f on S_N, we write $\pi(f) = \int f\,\mathrm{d}\pi$; in particular, $\pi(U) \in \mathbb{R}^N$ is the Gibbs mean of the normalized spin. For $x, y \in S_N$, we have $|U(x)| = 1$, $R(x,x) = 1$, and $|R(x,y)| \leq 1$. For each $x \in S_N$, the variable $H_N(x)$ is centered Gaussian with variance $N\xi(1)$. By Fubini's theorem, we have

$$\mathbb{E}Z_{N,\beta} = \int_{S_N} \mathbb{E}e^{\beta H_N(x)}\,\nu_N(\,\mathrm{d}x) = e^{N\beta^2\xi(1)/2}. \tag{3.1}$$

Lemma 3.2 (Representation, rotation invariance, and regularity of the field)**.** *Let ξ be a finite mixture. Let $g_{i_1\dots i_p}^{(p)}$, for $2 \leq p \leq P$ and $1 \leq i_1, \dots, i_p \leq N$, be independent standard Gaussian variables, and define the random polynomial*

$$H_N(x) = \sum_{p=2}^{P} \sqrt{c_p}\,N^{-(p-1)/2} \sum_{i_1,\dots,i_p=1}^{N} g_{i_1\dots i_p}^{(p)}\,x_{i_1}\cdots x_{i_p}, \qquad x \in \mathbb{R}^N. \tag{3.2}$$

1. *For all $x, y \in \mathbb{R}^N$, we have $\mathbb{E}H_N(x)H_N(y) = N\xi(R(x,y))$. In particular, the restriction of H_N to S_N is the field of Definition 3.1.*
2. *For every $g \in O(N)$, the random polynomials $H_N \circ g^{-1}$ and H_N have the same law.*
3. *There exist $K, c > 0$, depending only on ξ, such that for every $N \geq 2$, with probability at least $1 - e^{-cN}$,*

$$\begin{aligned} &\sup_{|x|\leq\sqrt{N}} |H_N(x)| \leq KN, && \sup_{|x|\leq\sqrt{N}} |\nabla H_N(x)| \leq K\sqrt{N}, \\ &\sup_{|x|\leq\sqrt{N}} \left\|\nabla^2 H_N(x)\right\|_{\mathrm{op}} \leq K, && \sup_{|x|\leq\sqrt{N}} \left\|\nabla^3 H_N(x)\right\|_{\mathrm{op}} \leq K/\sqrt{N}, \end{aligned} \tag{3.3}$$

and, on the same event,

$$\sup_{x\in S_N} \left|\nabla_{S_N} H_N(x)\right| \le K\sqrt{N}, \qquad \sup_{x\in S_N} \left\|\nabla^2_{S_N} H_N(x)\right\|_{\mathrm{op}} \le 2K. \tag{3.4}$$

Here ∇^k is the k-th Euclidean derivative, $\|T\|_{\mathrm{op}} = \sup\{|T[v_1,\dots,v_k]| : |v_1|,\dots,|v_k| \le 1\}$ for a k-linear form T, and ∇_{S_N} and $\nabla^2_{S_N}$ are the Riemannian gradient and Hessian on S_N.

Proof. For (1), since the variables $g^{(p)}_{i_1\dots i_p}$ are independent and centered, we have

$$\begin{aligned}\mathbb{E}H_N(x)H_N(y) &= \sum_{p=2}^{P} c_p N^{-(p-1)} \sum_{i_1,\dots,i_p=1}^{N} x_{i_1}y_{i_1}\cdots x_{i_p}y_{i_p} \\ &= \sum_{p=2}^{P} c_p N^{1-p}(x\cdot y)^p = N\xi(R(x,y)).\end{aligned}$$

For (2), note that g^{-1} preserves inner products. By (1), the process $H_N \circ g^{-1}$ is then a centered Gaussian process on $\mathbb{R}^N$ with the same covariance as H_N. Both are random elements of the finite-dimensional space of polynomials of degree at most P on $\mathbb{R}^N$. Recall that such a polynomial is determined by its values on the finite set of points of $\{0,1,\dots,P\}^N$ whose coordinates sum to at most P. The law of a random polynomial of degree at most P is therefore determined by the joint law of its values at these points. For $H_N \circ g^{-1}$ and H_N, these joint laws are centered Gaussian with the same covariance, which proves (2).

For (3), we use [93, Proposition 3.6], which is quoted there from [91]. Their Hamiltonian [93, (1.1)] is (3.2) with $\gamma_p^2 = c_p$, and their summability hypothesis $\sum_p 2^p\gamma_p^2 < \infty$ holds because only finitely many c_p are nonzero. For a k-tensor A, their normalized operator norm is $\|A\|_{\mathrm{op},N} = N^{(k-2)/2}\|A\|_{\mathrm{op}}$, and their supremum runs over $|x| \le \sqrt{N}$. In our notation, the proposition states that there exist constants C_k and $c > 0$ such that, with probability at least $1 - e^{-cN}$, we have $\sup_{|x|\le\sqrt{N}} N^{(k-2)/2}\left\|\nabla^k H_N(x)\right\|_{\mathrm{op}} \le C_k$ for all $k \ge 0$. Taking $k = 0,1,2,3$, we obtain (3.3) with $K = \max(C_0, C_1, C_2, C_3)$.

For (3.4), let $x \in S_N$, and let $P_x = I - xx^{\mathsf{T}}/N$ be the orthogonal projection onto the tangent space at x. Then $\nabla_{S_N} H_N(x) = P_x \nabla H_N(x)$. Differentiating twice along the great circle through x in the direction of a tangent vector, we find that

$$\nabla^2_{S_N} H_N(x)[v,w] = \nabla^2 H_N(x)[v,w] - N^{-1}\big(x\cdot\nabla H_N(x)\big)\, v\cdot w$$

for tangent vectors v and w. On the event in (3.3), it follows that $\left|\nabla_{S_N} H_N(x)\right| \le K\sqrt{N}$ and $\left\|\nabla^2_{S_N} H_N(x)\right\|_{\mathrm{op}} \le K + N^{-1}|x|\,|\nabla H_N(x)| \le 2K$. □

We always use the extension (3.2) when Euclidean derivatives of H_N are taken. The next lemma is the classical Gaussian concentration inequality for Lipschitz functions, due

to Tsirelson, Ibragimov, and Sudakov, applied to free energies. We use it in the form [111, Proposition 4.1 and (4.1)]; see also [112].

Lemma 3.3 (Concentration of free energies)**.** *Let ξ be a finite mixture and $\beta > 0$. Let $B \subset S_N$ be a Borel set with $\nu_N(B) > 0$ and let $h\colon S_N \to \mathbb{R}$ be a bounded Borel function, both deterministic. Then*

$$Y = \log \int_B e^{\beta H_N + h} \, \mathrm{d}\nu_N$$

is integrable and, for every $t > 0$,

$$\mathbb{P}\big(|Y - \mathbb{E}Y| \ge tN \big) \le 2 \exp\Big(- \frac{t^2 N}{2\beta^2 \xi(1)} \Big). \tag{3.5}$$

Proof. Let g be the vector of all the variables $g^{(p)}_{i_1 \dots i_p}$ in (3.2). Then $H_N(x) = g \cdot \Psi(x)$, where $\Psi(x)$ is the vector with entries $\sqrt{c_p}\, N^{-(p-1)/2} x_{i_1} \cdots x_{i_p}$. By the computation in the proof of Lemma 3.2, we have $|\Psi(x)|^2 = N\xi(1)$ for $x \in S_N$. The function $g \mapsto Y$ is smooth, and its gradient is β times the average of Ψ under the probability measure on B with density proportional to $e^{\beta H_N + h}$. Since this average has norm at most $\sqrt{N\xi(1)}$, the function Y is $\beta\sqrt{N\xi(1)}$-Lipschitz in the standard Gaussian vector g. The bound $|Y - \log \nu_N(B)| \le \|h\|_\infty + \beta\,|g|\,\sqrt{N\xi(1)}$ shows that Y is integrable. Recall that an L-Lipschitz function Y of a standard Gaussian vector satisfies $\mathbb{P}(|Y - \mathbb{E}Y| \ge s) \le 2e^{-s^2/(2L^2)}$. Taking $s = tN$ and $L^2 = \beta^2 N \xi(1)$ gives (3.5). □

The bounded function h accommodates a deterministic shift of the field, such as the one that appears under the planted law in Section 14. The same proof applies to a centered Gaussian field with covariance $N\eta(R(x,y))$ for any nonzero polynomial η with nonnegative coefficients, including a term of degree one, with $\xi(1)$ replaced by $\eta(1)$. If $\eta \equiv 0$, the free energy is deterministic and its deviation probability is zero. The lemma does not apply to domains B that depend on the field.

3.3 Langevin dynamics on the sphere

We equip S_N with the Riemannian metric induced by $\mathbb{R}^N$ and write ∇_{S_N} and Δ_{S_N} for its gradient and Laplace–Beltrami operator. For $y \in \mathbb{R}^N$, let $P_y = I - yy^{\mathsf{T}}/N$. For $y \in S_N$, this is the orthogonal projection onto the tangent space at y.

Definition 3.4 (Langevin dynamics)**.** Let $\beta \ge 0$ and let H be a smooth function on S_N; in our applications, H is a realization of the field H_N. Let π be the probability measure with density proportional to $e^{\beta H}$ with respect to ν_N. The *generator* $L = L_{\beta,H}$ and the *Dirichlet form* $\mathcal{E}$ act on smooth functions f, g on S_N by

$$Lf = \Delta_{S_N} f + \beta \nabla_{S_N} H \cdot \nabla_{S_N} f, \qquad \mathcal{E}(f,g) = \int_{S_N} \nabla_{S_N} f \cdot \nabla_{S_N} g \, \mathrm{d}\pi,$$

and we write $\mathcal{E}(f) = \mathcal{E}(f, f)$. We use the same letters for the closed form, with domain $\mathcal{D}(\mathcal{E})$, and for the self-adjoint operator constructed in Lemma 3.5. The *semigroup* is $P_t = e^{tL}$, $t \geq 0$. We list the eigenvalues of $-L$ in nondecreasing order with multiplicity as $\lambda_0 \leq \lambda_1 \leq \lambda_2 \leq \cdots$. For a vector-valued function $B = (B_1, \dots, B_m)$ with components in $\mathcal{D}(\mathcal{E})$, its Dirichlet energy, which we call the *vector energy*, is $\mathcal{E}_{\mathrm{vec}}(B) = \sum_{i=1}^m \mathcal{E}(B_i)$. The *Langevin diffusion* $(X_t)_{t\geq 0}$ is the diffusion process on S_N described in Lemma 3.5(5).

We work in the real Hilbert space $L^2(\pi)$ of real-valued square-integrable functions, with inner product $\langle f, g\rangle = \int fg \, \mathrm{d}\pi$ and norm $\|f\|_2$. For vector-valued functions, inner products, norms, and operators are taken componentwise and summed. For example, $\|B\|_2^2 = \int |B|^2 \, \mathrm{d}\pi$. We write $H^1(S_N)$ for the Sobolev space of functions on S_N with one weak derivative in $L^2(\nu_N)$.

Lemma 3.5 (Basic properties of the dynamics)**.** *Let $\beta \geq 0$ and let H be a smooth function on S_N, with π, L, and $\mathcal{E}$ as in Definition 3.4.*

1. Integration by parts. *For smooth f and g, we have $\int fLg \, \mathrm{d}\pi = -\mathcal{E}(f, g) = \int gLf \, \mathrm{d}\pi$.*
2. Self-adjoint generator. *The form $\mathcal{E}$ on smooth functions is closable in $L^2(\pi)$, and its closure has domain $\mathcal{D}(\mathcal{E}) = H^1(S_N)$. The associated nonnegative self-adjoint operator A on $L^2(\pi)$ satisfies $\mathcal{D}(A^{1/2}) = \mathcal{D}(\mathcal{E})$ and $\mathcal{E}(f) = \left\|A^{1/2} f\right\|_2^2$. The operator $-L$ on smooth functions is essentially self-adjoint, and its closure is A.*
3. Discrete spectrum. *The operator A has compact resolvent. There exist an orthonormal basis $(\phi_j)_{j\geq 0}$ of $L^2(\pi)$ consisting of real smooth functions and numbers $0 = \lambda_0 < \lambda_1 \leq \lambda_2 \leq \cdots$ with $\lambda_j \to \infty$ such that $A\phi_j = \lambda_j \phi_j$ for all j. The kernel of A consists of the constant functions, and $\phi_0 = 1$. For $f \in L^2(\pi)$, let $\hat{f}_j = \langle f, \phi_j\rangle$. Then $\|f\|_2^2 = \sum_j \hat{f}_j^2$. Moreover, $f \in \mathcal{D}(\mathcal{E})$ if and only if $\sum_j \lambda_j \hat{f}_j^2 < \infty$, and in that case $\mathcal{E}(f) = \sum_j \lambda_j \hat{f}_j^2$.*
4. Semigroup. *The semigroup $P_t = e^{-tA}$ is given by a transition kernel with a smooth density. Precisely, there exists a smooth function $(t, x, y) \mapsto p_t(x, y)$ on $(0, \infty) \times S_N \times S_N$, with $p_t \geq 0$, $p_t(x, y) = p_t(y, x)$, and $\int p_t(x, y) \, \pi(\mathrm{d}y) = 1$, with the following property. The kernels $P_t(x, \mathrm{d}y) = p_t(x, y) \, \pi(\mathrm{d}y)$ for $t > 0$ and $P_0(x, \cdot) = \delta_x$ satisfy $P_t f(x) = \int f(y) P_t(x, \mathrm{d}y)$ for every $f \in L^2(\pi)$, every $x \in S_N$, and every $t > 0$. These kernels form a Markov semigroup that maps continuous functions to continuous functions, and π is invariant and reversible: $\pi(\mathrm{d}x) P_t(x, \mathrm{d}y) = \pi(\mathrm{d}y) P_t(y, \mathrm{d}x)$.*
5. The diffusion. *Let $(W_t)_{t\geq 0}$ be a standard Brownian motion in $\mathbb{R}^N$. For every $x \in S_N$, the equation*

$$\mathrm{d}X_t = \sqrt{2}\, P_{X_t} \, \mathrm{d}W_t + \beta \nabla_{S_N} H(X_t) \, \mathrm{d}t - \frac{N-1}{N} X_t \, \mathrm{d}t, \qquad X_0 = x, \tag{3.6}$$

with coefficients extended smoothly and with compact support off a neighborhood of S_N, has a unique strong solution. It stays on S_N, its law does not depend on the extension, and

$\mathbb{P}_x(X_t \in \cdot) = P_t(x, \cdot)$ *for all* $t \geq 0$. *For every smooth* f *on* S_N, *the process* $f(X_t) - f(x) - \int_0^t Lf(X_s)\,\mathrm{d}s$ *is a martingale.*

All of these facts are standard for a smooth elliptic operator on a compact connected manifold that is symmetric with respect to a smooth positive density. We refer to [17, Section 1.7] for the correspondence between closed forms, self-adjoint operators, and spectral decompositions. We give the short arguments below. The diffusion is the spherical Langevin dynamics studied, up to normalization, in [25, 65, 81]. The factor $\sqrt{2}$ in front of the noise makes L, without a factor $1/2$, its generator.

Proof. In this proof, $\nabla = \nabla_{S_N}$ and $\Delta = \Delta_{S_N}$. We prove (1)–(3) with the standard correspondence between closed forms and self-adjoint operators. For (4) and (5), we expand in eigenfunctions and identify the resulting semigroup with the transition function of (3.6). We use three classical facts. The first is the Rellich–Kondrachov theorem: the inclusion $H^1(S_N) \subset L^2(\nu_N)$ is compact [64, Theorem 15.10]. The second is elliptic regularity for the second-order elliptic operator $1 - L$. A weak solution of the eigenvalue equation is smooth [64, Theorem 15.1], and iterating the elliptic estimate [64, Theorem 15.7], we find that for every integer $r \geq 0$ and every eigenfunction ϕ of A with eigenvalue λ and $\|\phi\|_2 = 1$,

$$\|\phi\|_{H^{2r}(S_N)} \leq C_{r,N,H}(1 + \lambda)^r.$$

Together with Sobolev embedding in local charts [64, Theorem 12.16], applied with $2r > k + (N-1)/2$, this gives $\|\phi\|_{C^k} \leq C_{k,N,H}(1+\lambda)^{m_k}$ for some integer m_k. The third is strong existence and uniqueness for stochastic differential equations with bounded Lipschitz coefficients [151, Theorem 4.1]. Note that the density $\rho = \mathrm{d}\pi / \mathrm{d}\nu_N$ is smooth, that $\nabla\rho = \beta\rho\nabla H$, and that $e^{-\beta \operatorname{osc} H} \leq \rho \leq e^{\beta \operatorname{osc} H}$, where $\operatorname{osc} H = \max H - \min H$.

Integration by parts. By the divergence theorem on the closed manifold S_N, we have

$$\int f \Delta g \,\rho \,\mathrm{d}\nu_N = -\int \nabla(f\rho) \cdot \nabla g \,\mathrm{d}\nu_N = -\int \nabla f \cdot \nabla g \,\rho \,\mathrm{d}\nu_N - \beta \int f \,\nabla H \cdot \nabla g \,\rho \,\mathrm{d}\nu_N,$$

where in the last step we used $\nabla\rho = \beta\rho\nabla H$. Rearranging, we have $\int fLg \,\mathrm{d}\pi = -\mathcal{E}(f,g)$. The second identity in (1) follows from the symmetry of $\mathcal{E}$.

The form and the operator. By the bounds on ρ, the norm $(\mathcal{E}(f) + \|f\|_2^2)^{1/2}$ on smooth functions is equivalent to the norm of $H^1(S_N)$. Since $H^1(S_N)$ is the completion of the smooth functions in that norm, $\mathcal{E}$ is closable and $\mathcal{D}(\mathcal{E}) = H^1(S_N)$. Let A be the nonnegative self-adjoint operator associated with the closed form. Recall that $g \in \mathcal{D}(A)$ and $Ag = k$ if and only if $g \in \mathcal{D}(\mathcal{E})$ and $\mathcal{E}(f,g) = \langle f, k\rangle$ for all $f \in \mathcal{D}(\mathcal{E})$, and that $\mathcal{D}(A^{1/2}) = \mathcal{D}(\mathcal{E})$ and $\mathcal{E}(f) = \|A^{1/2}f\|_2^2$. Using (1) and the density of smooth functions in $\mathcal{D}(\mathcal{E})$, we find that every smooth g belongs to $\mathcal{D}(A)$ and that $Ag = -Lg$. We prove the essential self-adjointness after (3).

The spectrum. If $Ag + g = k$, then $\mathcal{E}(g) + \|g\|_2^2 = \langle g, k\rangle \le \|g\|_2 \|k\|_2$. This shows that $(A+1)^{-1}$ maps bounded subsets of $L^2(\pi)$ to bounded subsets of $H^1(S_N)$. The Rellich–Kondrachov theorem and the equivalence of $L^2(\pi)$ and $L^2(\nu_N)$ then imply that the operator $(A+1)^{-1}$ is compact. The spectral theorem for compact self-adjoint operators on a real Hilbert space now provides an orthonormal basis of eigenfunctions of A. The eigenvalues $\lambda_j \ge 0$ have finite multiplicity and tend to infinity, and by elliptic regularity the eigenfunctions are smooth. Suppose that $Af = 0$. Then $\mathcal{E}(f) = \langle f, Af\rangle = 0$, and the weak gradient of f vanishes almost everywhere. As S_N is connected, f is almost everywhere constant. We conclude that the kernel of A consists of the constants, that $\lambda_0 = 0$ is simple, and that $\phi_0 = 1$ and $\lambda_1 > 0$. The identities for $\|f\|_2^2$ and $\mathcal{E}(f)$ follow from the expansion $A^{1/2} f = \sum_j \lambda_j^{1/2} \hat{f}_j \phi_j$ for $f \in \mathcal{D}(A^{1/2})$. Finally, the finite linear combinations of eigenfunctions are smooth. They form a core for A, since the truncations of the expansion of $f \in \mathcal{D}(A)$ converge to f in the graph norm. As $-L$ on smooth functions is a restriction of the closed operator A, its closure is A. This proves the essential self-adjointness claimed in (2).

The semigroup. We define $p_t(x,y) = \sum_j e^{-\lambda_j t} \phi_j(x) \phi_j(y)$. We first show that the series converges in $C^\infty((0,\infty) \times S_N \times S_N)$. Since ρ is bounded above and below, the min–max principle for $-L$ and for $-\Delta$ implies that $\lambda_j \ge e^{-2\beta \operatorname{osc} H} \lambda_j(-\Delta)$; see Lemma 3.6(5), whose proof uses only (3). The eigenvalues of $-\Delta$ are $k(k+N-2)/N$, $k \ge 0$, with multiplicities bounded by a polynomial in k. Hence $\sum_j (1+\lambda_j)^{-m} < \infty$ for some m, and together with the bounds on $\|\phi_j\|_{C^k}$ this gives the convergence. For $f \in L^2(\pi)$ and $t > 0$, we have $P_t f = \sum_j e^{-\lambda_j t} \hat{f}_j \phi_j$. The series $\sum_j e^{-\lambda_j t} \hat{f}_j \phi_j(x) = \int f(y) p_t(x,y)\, \pi(\mathrm{d}y)$ converges uniformly in x, which identifies $P_t f$ with a continuous function. The symmetry of p_t is evident, and $\int p_t(x,y)\, \pi(\mathrm{d}y) = P_t 1(x) = 1$. The Chapman–Kolmogorov equations follow from $e^{-(s+t)A} = e^{-sA} e^{-tA}$ and the continuity of the kernels. Reversibility follows from the symmetry of p_t, and it implies invariance. We prove that $p_t \ge 0$ below.

The diffusion. We extend $\sigma(y) = \sqrt{2} P_y$ and $b(y) = \beta \nabla H(y) - (N-1)y/N$ from a neighborhood of S_N to smooth compactly supported functions on $\mathbb{R}^N$. With these Lipschitz coefficients, (3.6) has a unique strong solution. Let $u(y) = |y|^2 - N$. By Itô's formula, the drift and diffusion coefficients of $u(X_t)$ are $2y \cdot b(y) + \operatorname{Tr}(\sigma\sigma^{\mathsf{T}})(y)$ and $2y^{\mathsf{T}} \sigma(y)$, evaluated at $y = X_t$. Both are smooth and vanish on S_N, because $y^{\mathsf{T}} P_y = 0$, $y \cdot \nabla H(y) = 0$, and $\operatorname{Tr}(2P_y^2) = 2(N-1)$ for $y \in S_N$. As S_N is a regular level set of u, near S_N these coefficients have the form $u(y)\alpha(y)$ and $u(y)\gamma(y)$ with α and γ smooth. Then $u(X_t)$ solves a linear equation with zero initial value up to the exit time from a neighborhood of S_N, which implies that $u(X_t) = 0$ for all t. In particular, the law of X does not depend on the extension.

Next, let F be a smooth function on $\mathbb{R}^N$. For $y \in S_N$, the Laplace–Beltrami operator of the sphere of radius $\sqrt{N}$ satisfies $\Delta(F|_{S_N})(y) = \operatorname{Tr}(P_y \nabla^2 F(y) P_y) - (N-1) y \cdot \nabla F(y)/N$,

where ∇ and ∇^2 on the right are Euclidean. By Itô's formula, we have

$$\mathrm{d}F(X_t) = LF(X_t)\,\mathrm{d}t + \sqrt{2}\,\nabla F(X_t)\cdot P_{X_t}\,\mathrm{d}W_t, \tag{3.7}$$

which proves the martingale property. We now apply (3.7) to $e^{\lambda_j t}\phi_j(X_t)$, with ϕ_j extended to a smooth function on $\mathbb{R}^N$. Since $L\phi_j = -\lambda_j\phi_j$, the process $e^{\lambda_j t}\phi_j(X_t)$ is a martingale; here the stochastic integral is a martingale because its integrand is bounded. Taking expectations, we find $\mathbb{E}_x\phi_j(X_t) = e^{-\lambda_j t}\phi_j(x) = P_t\phi_j(x)$ for every x. The law of X_t under $\mathbb{P}_x$ and the finite signed measure $p_t(x,y)\,\pi(\,\mathrm{d}y)$ therefore agree on finite linear combinations of eigenfunctions. These combinations are dense in the continuous functions on S_N in the uniform norm. Indeed, smooth functions are dense, and for smooth f we have $\hat{f}_j = \lambda_j^{-k}\langle(-L)^k f, \phi_j\rangle$ for $j \geq 1$ and every k, and the bounds above then show that the expansion of f converges uniformly. We conclude that the two measures coincide, which proves that $\mathbb{P}_x(X_t \in \cdot) = P_t(x,\cdot)$ and that $p_t \geq 0$. □

Lemma 3.6 (Spectral calculus)**.** *In the setting of Lemma 3.5, let $f \in L^2(\pi)$ and $t \geq 0$.*

1. *We have $P_t f = \sum_j e^{-t\lambda_j}\hat{f}_j\phi_j$ and $\langle f, P_t f\rangle = \sum_j e^{-t\lambda_j}\hat{f}_j^2$.*
2. *If $f \in \mathcal{D}(\mathcal{E})$, then $\|f - P_t f\|_2^2 \leq t\mathcal{E}(f)$ and $\langle f, (1-P_t)f\rangle \leq t\mathcal{E}(f)$.*
3. *If X_0 has law π, then $\mathbb{E}_\pi[f(X_0)g(X_t)] = \langle f, P_t g\rangle$ for every $g \in L^2(\pi)$, and*
$$\mathbb{E}_\pi\,|f(X_t) - f(X_0)|^2 = 2\langle f, (1-P_t)f\rangle \leq 2t\mathcal{E}(f) \qquad \text{if } f \in \mathcal{D}(\mathcal{E}).$$
4. *For $\Lambda > 0$, let $\Pi_\Lambda = \mathbf{1}_{(0,\Lambda]}(A)$, so that $\Pi_\Lambda f = \sum_{j:0<\lambda_j\leq\Lambda}\hat{f}_j\phi_j$. Then $\operatorname{rank}\Pi_\Lambda = \#\{j \geq 1 : \lambda_j \leq \Lambda\}$, and for every integer $k \geq 0$ we have $\lambda_k \leq \Lambda$ if and only if $\operatorname{rank}\Pi_\Lambda \geq k$. If $\pi(f) = 0$, then $\langle f, P_t f\rangle \geq e^{-t\Lambda}\,\|\Pi_\Lambda f\|_2^2$. If $f \in \mathcal{D}(\mathcal{E})$, then $\|f - \pi(f) - \Pi_\Lambda f\|_2^2 \leq \mathcal{E}(f)/\Lambda$.*
5. Min–max principle. *For every $k \geq 0$,*
$$\lambda_k = \min_V \max_{f\in V\setminus\{0\}} \frac{\mathcal{E}(f)}{\|f\|_2^2},$$
where V ranges over the $(k+1)$-dimensional subspaces of $\mathcal{D}(\mathcal{E})$. The minimum is attained at the span of $\phi_0, \ldots, \phi_k$, which consists of smooth functions, so V may equally range over subspaces of smooth functions.

Items (1)–(3) and the two estimates for functions in item (4) extend to vector-valued functions, by applying the operators componentwise and summing over components, with $\mathcal{E}_{\mathrm{vec}}$ in place of $\mathcal{E}$. The rank identity and the min–max principle concern the scalar operator A. For example, $\mathbb{E}_\pi\,|B(X_t) - B(X_0)|^2 \leq 2t\mathcal{E}_{\mathrm{vec}}(B)$.

Proof. Item (1) is the spectral theorem for $P_t = e^{-tA}$, together with Lemma 3.5(3). For (2), we have

$$\|f - P_t f\|_2^2 = \sum_j \left(1 - e^{-t\lambda_j}\right)^2 \hat{f}_j^2 \leq \sum_j t\lambda_j\hat{f}_j^2 = t\mathcal{E}(f),$$

where we used $(1-e^{-u})^2 \le u$ for $u \ge 0$. For $u \le 1$, this follows from $1-e^{-u} \le u$, and for $u \ge 1$, the left side is at most 1. The second bound in (2) follows in the same way from $1-e^{-u} \le u$.

For (3), recall from Lemma 3.5(4)–(5) that the joint law of (X_0, X_t) under $\mathbb{P}_\pi$ is $\pi(\mathrm{d}x)P_t(x, \mathrm{d}y)$. Then $\mathbb{E}_\pi[f(X_0)g(X_t)] = \int f(x)P_tg(x)\,\pi(\mathrm{d}x) = \langle f, P_tg\rangle$. Both endpoints have law π, and therefore

$$\mathbb{E}_\pi\,|f(X_t)-f(X_0)|^2 = 2\,\|f\|_2^2 - 2\langle f, P_tf\rangle = 2\langle f, (1-P_t)f\rangle,$$

and the bound follows from (2).

For (4), note that the range of Π_Λ is spanned by the ϕ_j with $j \ge 1$ and $\lambda_j \le \Lambda$, which gives the rank. Since $\lambda_0 = 0$ and the eigenvalues are nondecreasing, $\lambda_k \le \Lambda$ holds if and only if at least $k+1$ indices $j \ge 0$ satisfy $\lambda_j \le \Lambda$, that is, if and only if $\operatorname{rank}\Pi_\Lambda \ge k$. If $\pi(f) = 0$, then $\hat f_0 = 0$, and every term of the series for $\langle f, P_tf\rangle$ in (1) is nonnegative. Keeping only the terms with $0 < \lambda_j \le \Lambda$, we find $\langle f, P_tf\rangle \ge e^{-t\Lambda}\,\|\Pi_\Lambda f\|_2^2$. Finally, since $\hat f_0 = \pi(f)$, we have $f - \pi(f) - \Pi_\Lambda f = \sum_{j:\lambda_j>\Lambda} \hat f_j\phi_j$, and the squared norm of this function is at most $\Lambda^{-1}\sum_j \lambda_j\hat f_j^2 = \mathcal{E}(f)/\Lambda$.

For (5), let V be a $(k+1)$-dimensional subspace of $\mathcal{D}(\mathcal{E})$. It contains a nonzero f orthogonal to $\phi_0, \dots, \phi_{k-1}$ (no condition if $k = 0$), and by Lemma 3.5(3) this f satisfies $\mathcal{E}(f) = \sum_{j\ge k}\lambda_j\hat f_j^2 \ge \lambda_k\,\|f\|_2^2$. On the other hand, every f in the span of $\phi_0, \dots, \phi_k$ satisfies $\mathcal{E}(f) = \sum_{j\le k}\lambda_j\hat f_j^2 \le \lambda_k\,\|f\|_2^2$. This proves (5), including the attainment of the minimum by smooth functions. This is the classical min–max principle; compare [66, equation (2)] for the equivalent weighted formulation. The vector-valued estimates follow by summing over components. □

Remark 3.7 (Fields of class C^2). Section 18 applies these facts to fields that are only twice continuously differentiable. The proofs of Lemma 3.5(1)–(3), except for the smoothness of the eigenfunctions and the essential self-adjointness, use only that H is continuously differentiable. The proofs of Lemma 3.6(1), (2), and (4), and of the min–max principle over $\mathcal{D}(\mathcal{E})$, use only Lemma 3.5(3). These statements hold for every H of class C^2, with $P_t = e^{-tA}$ defined by the spectral theorem. The last sentence of Lemma 3.6(5), on subspaces of smooth functions, uses the smoothness of the eigenfunctions; for fields of class C^2, it is replaced by the corresponding statement of Lemma 18.3.

3.4 Markov semigroups and relative entropy

In this section, E and F are standard Borel spaces. A Markov kernel K from E to F acts on probability measures on E by $\mu K(\cdot) = \int K(x,\cdot)\,\mu(\mathrm{d}x)$ and on bounded measurable functions on F by $Kf(x) = \int f(y)K(x, \mathrm{d}y)$. A Markov kernel on E is a Markov kernel from E to E. A *Markov semigroup* is a family $(P_t)_{t\ge0}$ of Markov kernels on E such that

$P_0(x,\cdot) = \delta_x$ and $P_{s+t} = P_s P_t$ for all $s,t \geq 0$. A probability measure π is *invariant* if $\pi P_t = \pi$ for all t. Recall from Lemma 3.5(4) that the Langevin semigroup is a Markov semigroup on S_N with invariant law π. Nothing in this section uses reversibility.

The following facts are classical. For data processing, we refer to [55] and [127, Theorem 2.17 and Corollary 2.18]. We include the short proofs.

Lemma 3.8 (Contraction)**.** *Let K be a Markov kernel from E to F, and let μ, ν be probability measures on E. Then*

$$\|\mu K - \nu K\|_{\mathrm{TV}} \leq \|\mu - \nu\|_{\mathrm{TV}}, \qquad D(\mu K \| \nu K) \leq D(\mu\|\nu).$$

Let (P_t) be a Markov semigroup with invariant law π. Then for every probability measure μ, the functions $t \mapsto \|\mu P_t - \pi\|_{\mathrm{TV}}$ and $t \mapsto D(\mu P_t\|\pi)$ are nonincreasing on $[0,\infty)$.

Proof. Let E_+ be a positive set of a Hahn decomposition of the signed measure $\mu - \nu$. Then $\|\mu-\nu\|_{\mathrm{TV}} = (\mu-\nu)(E_+) = -(\mu-\nu)(E_+^c)$. For every measurable function u with $0 \leq u \leq 1$, we have $(\mu-\nu)(E_+^c) \leq \int u \,\mathrm{d}(\mu-\nu) \leq (\mu-\nu)(E_+)$, which gives $\left|\int u \,\mathrm{d}(\mu-\nu)\right| \leq \|\mu-\nu\|_{\mathrm{TV}}$. Choosing $u = K(\cdot, A)$, we obtain $|\mu K(A) - \nu K(A)| \leq \|\mu-\nu\|_{\mathrm{TV}}$ for every measurable $A \subset F$.

For relative entropy, we use the Donsker–Varadhan formula $D(\mu\|\nu) = \sup_u\{\int u \,\mathrm{d}\mu - \log \int e^u \,\mathrm{d}\nu\}$, where the supremum is over bounded measurable functions [127, Theorem 4.6]. Let u be a bounded measurable function on F. By Jensen's inequality, we have $K(e^u) \geq e^{Ku}$ pointwise. This implies that

$$\int u \,\mathrm{d}(\mu K) - \log \int e^u \,\mathrm{d}(\nu K) \leq \int Ku \,\mathrm{d}\mu - \log \int e^{Ku} \,\mathrm{d}\nu \leq D(\mu\|\nu).$$

In the last inequality, we used that Ku is a bounded measurable function on E. Taking the supremum over u proves that $D(\mu K\|\nu K) \leq D(\mu\|\nu)$.

For the last assertion, let $s,t \geq 0$. Since $\mu P_{t+s} = (\mu P_t)P_s$ and $\pi = \pi P_s$, the two inequalities with $K = P_s$ imply that $\|\mu P_{t+s} - \pi\|_{\mathrm{TV}} \leq \|\mu P_t - \pi\|_{\mathrm{TV}}$ and $D(\mu P_{t+s}\|\pi) \leq D(\mu P_t\|\pi)$. □

Lemma 3.9 (Binary data processing)**.** *Let μ and ν be probability measures on a measurable space, let B be a measurable set, and let $p = \mu(B)$ and $a = \nu(B)$. Then*

$$D(\mu\|\nu) \geq d(p\|a) \geq p \log \frac{1}{a} - \log 2. \tag{3.8}$$

In particular, if $0 < a < 1$, then $p \leq (D(\mu\|\nu) + \log 2)/\log(1/a)$. If $a = 0$ and $D(\mu\|\nu) < \infty$, then $p = 0$.

Proof. Let K be the kernel that sends x to the point mass at $\mathbf{1}_B(x) \in \{0,1\}$. Then μK and νK are the Bernoulli laws with parameters p and a, and Lemma 3.8, applied to this kernel

into $\{0,1\}$, implies that $D(\mu\|\nu) \geq D(\mu K\|\nu K) = d(p\|a)$. For $0 < a < 1$, expanding the definition, we find

$$d(p\|a) = p\log\frac{1}{a} + (1-p)\log\frac{1}{1-a} - h(p) \geq p\log\frac{1}{a} - \log 2,$$

since the middle term is nonnegative and $h(p) \leq \log 2$. If $a = 0 < p$, then $d(p\|a) = \infty = p\log(1/a)$. If $a = p = 0$, the right side of (3.8) is $-\log 2$ by the convention $0 \cdot \infty = 0$. If $a = 1$, the right side is $-\log 2$ as well. In these cases, (3.8) holds because $d(p\|a) \geq 0$. The two final assertions follow by rearranging (3.8). □

The next proposition starts from a bound on the mass that the law at a single time T gives to a set B of almost full Gibbs measure. It converts this bound into separation in total variation on the whole interval $[0,T]$, and it transfers the separation to initial laws of finite relative entropy.

Proposition 3.10 (Separation in total variation and transfer by relative entropy)**.** *Let (P_t) be a Markov semigroup on E with invariant law π. Let ν be a probability measure on E, let $T \geq 0$, let $B \subset E$ be measurable, and let $a = (\nu P_T)(B)$ and $\eta = \pi(B^c)$. Then*

$$\inf_{0\leq t\leq T} \|\nu P_t - \pi\|_{\mathrm{TV}} \geq 1 - \eta - a. \tag{3.9}$$

Let μ be another probability measure with $D(\mu\|\nu) < \infty$. If $0 < a < 1$, then

$$\inf_{0\leq t\leq T} \|\mu P_t - \pi\|_{\mathrm{TV}} \geq 1 - \eta - \min\Big\{1, \frac{D(\mu\|\nu) + \log 2}{\log(1/a)}\Big\}. \tag{3.10}$$

If $a = 0$, then $(\mu P_T)(B) = 0$ and $\inf_{0\leq t\leq T} \|\mu P_t - \pi\|_{\mathrm{TV}} \geq 1 - \eta$.

Proof. Let ρ be a probability measure on E. Lemma 3.8 implies that the function $t \mapsto \|\rho P_t - \pi\|_{\mathrm{TV}}$ is nonincreasing. Testing the set B at time T, we obtain

$$\inf_{0\leq t\leq T} \|\rho P_t - \pi\|_{\mathrm{TV}} = \|\rho P_T - \pi\|_{\mathrm{TV}} \geq \pi(B) - (\rho P_T)(B) = 1 - \eta - (\rho P_T)(B). \tag{3.11}$$

With $\rho = \nu$, this is (3.9).

Now let $\rho = \mu$ and $p = (\mu P_T)(B)$. By Lemma 3.8 with $K = P_T$, we have $D(\mu P_T\|\nu P_T) \leq D(\mu\|\nu)$. If $0 < a < 1$, then Lemma 3.9, applied to μP_T, νP_T, and B, implies that $p \leq (D(\mu\|\nu) + \log 2)/\log(1/a)$. Together with $p \leq 1$ and (3.11), this proves (3.10). If $a = 0$, then $D(\mu P_T\|\nu P_T) < \infty$ and Lemma 3.9 gives $p = 0$. The last assertion then follows from (3.11). □

Remark 3.11 (Polynomial endpoint bounds). In our applications, the semigroup, the invariant law, the set B, and the initial laws depend on the realization of the field,

and Proposition 3.10 is applied for each fixed realization. For example, suppose that $(\nu P_T)(B) \le N^{-\zeta}$ for some $\zeta > 0$. The proof of Proposition 3.10 shows that

$$(\mu P_T)(B) \le \frac{D(\mu\|\nu) + \log 2}{\zeta \log N}.$$

In particular, a field-dependent initial law μ has vanishing mass in B at time T as soon as its relative entropy $D(\mu\|\nu)$ with respect to ν, computed in the realized field, is $o(\log N)$ on the event, or in the probability mode, used in the statement at hand. The proposition concerns only the law at time T, and it gives no information about the path before time T.

The last lemma of this section is the chain identity for mutual information, known as the golden formula [127, Theorem 4.1], with the reference law π in the role of the output law.

Lemma 3.12 (Endpoint mutual information)**.** *Let K be a Markov kernel on E, let π and μ be probability measures on E, and let (X_0, X_1) have joint law $\mu(\mathrm{d}x)K(x, \mathrm{d}y)$. Then, in $[0,\infty]$,*

$$\int D\big(K(x,\cdot) \,\big\|\, \pi\big)\, \mu(\mathrm{d}x) = I(X_0; X_1) + D(\mu K\|\pi). \tag{3.12}$$

In particular, if $D(\mu K\|\pi) < \infty$, then $I(X_0; X_1) = \int D(K(x,\cdot)\|\pi)\, \mu(\mathrm{d}x) - D(\mu K\|\pi)$. If $\pi K = \pi$, then the stationary pair, with $\mu = \pi$, satisfies

$$I(X_0; X_1) = \int D\big(K(x,\cdot) \,\big\|\, \pi\big)\, \pi(\mathrm{d}x). \tag{3.13}$$

For a Markov semigroup, we write $I_\mu(X_0; X_t)$ for the mutual information of (X_0, X_t) when X_0 has law μ and $K = P_t$, and $I(X_0; X_t) = I_\pi(X_0; X_t)$ in the stationary case.

Proof. Let $\Gamma(\mathrm{d}x, \mathrm{d}y) = \mu(\mathrm{d}x)K(x, \mathrm{d}y)$ and $\mu_1 = \mu K$, and recall that $I(X_0; X_1) = D(\Gamma\|\mu \otimes \mu_1)$. The identity (3.12) follows from the two identities

$$D(\Gamma\|\mu \otimes \pi) = \int D\big(K(x,\cdot) \,\big\|\, \pi\big)\, \mu(\mathrm{d}x), \tag{3.14}$$

$$D(\Gamma\|\mu \otimes \pi) = D(\Gamma\|\mu \otimes \mu_1) + D(\mu_1\|\pi). \tag{3.15}$$

We begin with (3.14). Suppose that $\Gamma \ll \mu \otimes \pi$ with density k. By Fubini's theorem, for μ-almost every x we have $K(x, A) = \int_A k(x, y)\, \pi(\mathrm{d}y)$ for all sets A in a countable generating algebra, and then for all measurable A by the monotone class theorem. It follows that $K(x,\cdot) \ll \pi$ with density $k(x,\cdot)$ for μ-almost every x, and Tonelli's theorem, applied to $k \log k + e^{-1} \ge 0$, gives (3.14). If Γ is not absolutely continuous with respect to $\mu \otimes \pi$, the left side of (3.14) is infinite, and so is the right side. Indeed, if the right side were finite, then $K(x,\cdot) \ll \pi$ for μ-almost every x, and a jointly measurable choice of the

densities, which exists on standard Borel spaces, would be a density of Γ with respect to $\mu \otimes \pi$.

For (3.15), suppose first that $\mu_1 \not\ll \pi$. Since μ_1 is the second marginal of Γ, Lemma 3.8, applied to the projection kernel from $E \times E$ to E, implies that $D(\Gamma \| \mu \otimes \pi) \geq D(\mu_1 \| \pi) = \infty$, and both sides are infinite. Suppose next that $\mu_1 \ll \pi$ with density r. Then $\mu \otimes \mu_1$ has density $r(y)$ with respect to $\mu \otimes \pi$, so $\Gamma \ll \mu \otimes \mu_1$ implies $\Gamma \ll \mu \otimes \pi$. Conversely, suppose that $\Gamma \ll \mu \otimes \pi$ and that $(\mu \otimes \mu_1)(E') = 0$ for a measurable set $E' \subset E \times E$. Then $r(y) = 0$ for $(\mu \otimes \pi)$-almost every $(x, y) \in E'$, so the set $E' \cap \{(x, y) : r(y) > 0\}$ is $(\mu \otimes \pi)$-null and hence Γ-null. Together with $\Gamma(\{(x, y) : r(y) = 0\}) = \mu_1(\{r = 0\}) = 0$, this gives $\Gamma(E') = 0$. We conclude that $\Gamma \ll \mu \otimes \pi$ if and only if $\Gamma \ll \mu \otimes \mu_1$, and if neither holds, both sides of (3.15) are infinite. Otherwise, let k be the density of Γ with respect to $\mu \otimes \mu_1$. Then Γ has density $k(x, y)r(y)$ with respect to $\mu \otimes \pi$, and we have

$$D(\Gamma \| \mu \otimes \pi) = \int \log k \, \mathrm{d}\Gamma + \int \log r(y) \, \Gamma(\mathrm{d}x, \mathrm{d}y) = D(\Gamma \| \mu \otimes \mu_1) + D(\mu_1 \| \pi).$$

The splitting of the integral is legitimate because the negative parts of $\log k$ and $\log r$ are Γ-integrable. For instance, $\int (\log k)^- \, \mathrm{d}\Gamma = \int k (\log k)^- \, \mathrm{d}(\mu \otimes \mu_1) \leq e^{-1}$.

The remaining assertions follow from (3.12). If $D(\mu K \| \pi) < \infty$, we may subtract it, and if $\pi K = \pi$ and $\mu = \pi$, then $D(\mu K \| \pi) = 0$. □

For the Langevin semigroup, $P_t(x, \cdot)$ has the continuous density $p_t(x, \cdot)$ for $t > 0$ by Lemma 3.5(4), and the measure-theoretic points in the proof are immediate there. At $t = 0$, the stationary pair lies on the diagonal, which has $\pi \otimes \pi$-measure zero, and $I(X_0; X_0) = \infty$.

3.5 The Crisanti–Sommers functional and the static thresholds

The free energy of a spherical spin glass is given by a variational formula over probability measures on $[0, 1)$. This formula was derived by Crisanti and Sommers [52] with the replica method, in a form equivalent to Parisi's formula. It was proved by Talagrand for even mixtures [146] and by Chen for general mixtures [46], and Huang and Sellke gave a constructive proof of the lower bound [90]. We need the formula for covariances with a term of degree one, which appear in Section 14.

Definition 3.13 (Crisanti–Sommers functional)**.** Let $\mathscr{M}$ be the set of Borel probability measures ζ on $[0, 1)$ such that $\zeta([0, \hat{q}]) = 1$ for some $\hat{q} < 1$. For $\zeta \in \mathscr{M}$ define

$$\hat{x}_\zeta(t) = \int_t^1 \zeta([0, s]) \, \mathrm{d}s = \int \big(1 - \max(t, s)\big) \, \zeta(\mathrm{d}s), \qquad 0 \leq t \leq 1.$$

A *covariance polynomial* is a polynomial $\eta(t) = \sum_{p=1}^{P} a_p t^p$ with $a_p \geq 0$; a term of degree one is allowed. For a covariance polynomial η and $\beta \geq 0$, the *Crisanti–Sommers functional* is

$$\mathcal{P}_\beta(\zeta;\eta) = \frac{1}{2}\biggl[\beta^2\Bigl(\eta(1) - \int \eta \,\mathrm{d}\zeta\Bigr) + \int_0^{\hat{q}} \frac{\mathrm{d}t}{\hat{x}_\zeta(t)} + \log(1-\hat{q})\biggr], \tag{3.16}$$

where $\hat{q} < 1$ is any number with $\zeta([0,\hat{q}]) = 1$, and we set $\mathcal{P}_\beta(\eta) = \inf_{\zeta\in\mathscr{M}} \mathcal{P}_\beta(\zeta;\eta)$.

The second expression for $\hat{x}_\zeta$ follows from Fubini's theorem. The function $\hat{x}_\zeta$ is Lipschitz, with derivative $-\zeta([0,t])$ almost everywhere, and $\hat{x}_\zeta(t) = 1-t$ for $t \geq \hat{q}$. Since ζ is carried by $[0,\hat{q}]$, the second expression gives $\hat{x}_\zeta(t) \geq 1 - \max(t,\hat{q}) > 0$ for $t < 1$. Further, (3.16) does not depend on the choice of $\hat{q}$, because $\int_{\hat{q}}^{\hat{q}'} \mathrm{d}t/(1-t) = \log(1-\hat{q}) - \log(1-\hat{q}')$. Using $\eta(0) = 0$, we also have

$$\eta(1) - \int \eta \,\mathrm{d}\zeta = \int_0^1 \eta'(t)\,\zeta([0,t])\,\mathrm{d}t = \eta'(0)\hat{x}_\zeta(0) + \int_0^1 \eta''(t)\hat{x}_\zeta(t)\,\mathrm{d}t. \tag{3.17}$$

For the first equality, we wrote $\eta(1) - \eta(s) = \int_s^1 \eta'(t)\,\mathrm{d}t$ and used Fubini's theorem. For the second, we integrated by parts, using $\hat{x}_\zeta'(t) = -\zeta([0,t])$ and $\hat{x}_\zeta(1) = 0$. By (3.17), $\mathcal{P}_\beta(\zeta;\eta)$ coincides with the functional $\mathcal{P}(x;\xi)$ of [90, (1.4)] for $\xi = \beta^2\eta$ and $x(t) = \zeta([0,t])$.

Two choices of ζ appear repeatedly. For $\zeta = \delta_0$, we have $\hat{x}_{\delta_0}(t) = 1-t$ and $\mathcal{P}_\beta(\delta_0;\eta) = \beta^2\eta(1)/2$. For $\zeta = x\delta_0 + (1-x)\delta_Q$ with $0 < x < 1$ and $0 < Q < 1$, we have $\hat{x}_\zeta(t) = 1 - Q + x(Q-t)$ for $t \leq Q$ and $\hat{x}_\zeta(t) = 1-t$ for $t \geq Q$. This is the function $\hat{x}$ used in Section 5.

Theorem 3.14 (Parisi formula; [46, 146])**.** *Let η be a covariance polynomial, let $\beta \geq 0$, and let G_N be a centered Gaussian process on S_N with continuous sample paths and covariance $\mathbb{E}G_N(x)G_N(y) = N\eta(R(x,y))$. Then*

$$\lim_{N\to\infty} \frac{1}{N}\mathbb{E}\log\int_{S_N} e^{\beta G_N}\,\mathrm{d}\nu_N = \mathcal{P}_\beta(\eta).$$

In particular, $F(\beta) = \mathcal{P}_\beta(\xi)$ for every finite mixture ξ and every $\beta \geq 0$.

For even η, this is [146, Theorem 1.1]. For general covariance polynomials, including odd degrees and a term of degree one, it is [46, Theorem 1]. The summability hypothesis there holds trivially, because the coefficients $\beta_p = \beta\sqrt{a_p}$ in the notation of [46, (1.2)] are nonnegative and only finitely many are nonzero. Chen states the formula in Parisi's form, and its equality with the Crisanti–Sommers form is [146, Theorem 4.1]. In exactly the form (3.16), the statement is [90, Theorem 1.1], which also gives convergence in probability of $N^{-1}\log\int e^{\beta G_N}\,\mathrm{d}\nu_N$. The value of $\int e^{\beta G_N}\,\mathrm{d}\nu_N$ depends only on the law of G_N, and any realization may be used, for example the analog of (3.2) with p ranging over the degrees of η.

The minimizer of the Crisanti–Sommers functional is characterized by its first variation. For $\zeta \in \mathscr{M}$, let

$$f_\zeta(s) = \beta^2\eta(s) - \int_0^s \frac{s-t}{\hat{x}_\zeta(t)^2}\,\mathrm{d}t, \qquad 0 \le s < 1, \tag{3.18}$$

and let $S_\zeta = \{s \in [0,1) : f_\zeta(s) = \sup_{[0,1)} f_\zeta\}$. With $\eta = \xi$ and $\zeta = x\delta_0 + (1-x)\delta_Q$, the function $-f_\zeta$ is the function V of Section 5.

Lemma 3.15 (First variation and the optimality criterion)**.** *Let η be a covariance polynomial and $\beta \ge 0$.*

1. *For every $\zeta \in \mathscr{M}$, the function f_ζ is continuous on $[0,1)$ and tends to $-\infty$ at 1. In particular, S_ζ is a nonempty compact subset of $[0,1)$.*
2. *For $\zeta, \mu \in \mathscr{M}$, the function $\varepsilon \mapsto \mathcal{P}_\beta((1-\varepsilon)\zeta + \varepsilon\mu; \eta)$ is convex and differentiable on $[0,1]$, and its derivative at $\varepsilon = 0$ is $-\int f_\zeta\,\mathrm{d}(\mu-\zeta)/2$. In particular,*

$$\mathcal{P}_\beta(\mu;\eta) - \mathcal{P}_\beta(\zeta;\eta) \ge \frac{1}{2}\Big(\int f_\zeta\,\mathrm{d}\zeta - \int f_\zeta\,\mathrm{d}\mu\Big). \tag{3.19}$$

3. *The functional $\mathcal{P}_\beta(\cdot\,;\eta)$ has a unique minimizer on $\mathscr{M}$, and $\zeta \in \mathscr{M}$ is this minimizer if and only if $\zeta(S_\zeta) = 1$.*

We call the characterization in (3) the *Crisanti–Sommers optimality criterion*. Together with existence and uniqueness, it is due to Talagrand [146, Proposition 2.1]. In the present normalization it is [90, Lemma 1.2]; see also [143, Proposition 9]. For $\eta = \xi$, we call the minimizer the *Parisi measure* of (ξ, β). We include the short proof of the characterization, since its first-variation formula is used in Sections 5 and 14.

Proof. For (1), continuity is clear. Fix $\hat{q} < 1$ with $\zeta([0,\hat{q}]) = 1$. For $\hat{q} < s < 1$, using $\hat{x}_\zeta(t) = 1 - t$ on $[\hat{q}, s]$, we find

$$f_\zeta(s) \le \beta^2\eta(1) - \int_{\hat{q}}^s \frac{s-t}{(1-t)^2}\,\mathrm{d}t = \beta^2\eta(1) - \log\frac{1-\hat{q}}{1-s} + 1 - \frac{1-s}{1-\hat{q}}.$$

The right side tends to $-\infty$ as $s \to 1$. Hence the supremum of f_ζ is attained, and S_ζ is a nonempty closed subset of an interval $[0, s_0]$ with $s_0 < 1$.

For (2), let $\zeta_\varepsilon = (1-\varepsilon)\zeta + \varepsilon\mu$ and choose $\hat{q} < 1$ such that $\zeta([0,\hat{q}]) = \mu([0,\hat{q}]) = 1$. The map $\varepsilon \mapsto \hat{x}_{\zeta_\varepsilon} = \hat{x}_\zeta + \varepsilon(\hat{x}_\mu - \hat{x}_\zeta)$ is affine, and $\hat{x}_{\zeta_\varepsilon} \ge 1 - \hat{q}$ on $[0,\hat{q}]$ for $\varepsilon \in [0,1]$. The energy term in (3.16) is affine in ε, and the entropy term is convex in ε because $u \mapsto 1/u$ is convex on $(0,\infty)$. Differentiating under the integral, which is justified by the lower bound $1 - \hat{q}$ on the denominators, we find that the derivative at $\varepsilon = 0$ is

$$-\frac{1}{2}\Big[\beta^2\int \eta\,\mathrm{d}(\mu-\zeta) + \int_0^{\hat{q}} \frac{\hat{x}_\mu(t) - \hat{x}_\zeta(t)}{\hat{x}_\zeta(t)^2}\,\mathrm{d}t\Big] = -\frac{1}{2}\int G(s)\,(\mu-\zeta)(\,\mathrm{d}s),$$

where $G(s) = \beta^2\eta(s) + \int_0^{\hat{q}} (1 - \max(t,s))\hat{x}_\zeta(t)^{-2}\,\mathrm{d}t$. Here we used the second expression for $\hat{x}_\zeta$ in Definition 3.13 and Fubini's theorem. For $s \le \hat{q}$, the identity $1 - \max(t,s) = (1-t) - (s-t)_+$ gives $G(s) = f_\zeta(s) + \int_0^{\hat{q}} (1-t)\hat{x}_\zeta(t)^{-2}\,\mathrm{d}t$. The last term does not depend on s, and $\mu - \zeta$ has total mass zero and is carried by $[0,\hat{q}]$. It follows that the derivative equals $-\int f_\zeta\,\mathrm{d}(\mu-\zeta)/2$. A convex differentiable function on $[0,1]$ lies above its tangent line at 0, and evaluating at $\varepsilon = 1$ gives (3.19).

For (3), suppose first that $\zeta(S_\zeta) = 1$. Then $\int f_\zeta\,\mathrm{d}\zeta = \max f_\zeta \ge \int f_\zeta\,\mathrm{d}\mu$ for every $\mu \in \mathscr{M}$, and (3.19) implies that ζ is a minimizer. Conversely, let ζ be a minimizer and $r \in [0,1)$. With $\mu = \delta_r$ in (2), the derivative at $\varepsilon = 0$ is nonnegative, which means that $f_\zeta(r) \le \int f_\zeta\,\mathrm{d}\zeta$. Taking the supremum over r, we find $\max f_\zeta \le \int f_\zeta\,\mathrm{d}\zeta \le \max f_\zeta$. This shows that $f_\zeta = \max f_\zeta$ holds ζ-almost surely, that is, $\zeta(S_\zeta) = 1$. For uniqueness, let $\zeta \ne \mu$ be two minimizers. Since $\hat{x}_\zeta'(t) = -\zeta([0,t])$ almost everywhere and distribution functions are right-continuous, the continuous functions $\hat{x}_\zeta$ and $\hat{x}_\mu$ differ on an interval contained in $[0,\hat{q}]$. By the strict convexity of $u \mapsto 1/u$, the entropy term is then strictly convex along the segment from ζ to μ. The midpoint of the segment then has a strictly smaller value, which is a contradiction. The existence of a minimizer is part of [146, Proposition 2.1] and [90, Lemma 1.2]. □

Definition 3.16 (Static and dynamical thresholds)**.** Let ξ be a finite mixture and $\beta \ge 0$. For $0 < r < 1$, let

$$A(r) = \frac{(1-r)\xi'(r)}{r}, \qquad \Phi_\beta(r) = \beta^2\xi(r) + r + \log(1-r).$$

The *dynamical threshold* $\beta_\mathrm{d} = \beta_\mathrm{d}(\xi)$ and the *static threshold* $\beta_\mathrm{s} = \beta_\mathrm{s}(\xi)$ are defined by

$$\beta_\mathrm{d}^{-2} = \sup_{0<r<1} A(r), \qquad \beta_\mathrm{s}^2 = \inf_{0<r<1} \frac{-r - \log(1-r)}{\xi(r)}.$$

The pair (ξ,β) is *strictly replica symmetric* if $\Phi_\beta(r) < 0$ for all $r \in (0,1)$.

In the proof of Lemma 3.18 below, we show that Φ_β is the first-variation function f_{δ_0} of (3.18) for $\eta = \xi$. By Lemma 3.15(3), (ξ,β) is strictly replica symmetric if and only if δ_0 is its Parisi measure and $S_{\delta_0} = \{0\}$. This is the definition of Huang and Sellke [90, Definition 3], with the temperature absorbed into the covariance. Several results below require in addition that $\beta^2\xi''(0) < 1$, and we then say that (ξ,β) is strictly replica symmetric with $\beta^2\xi''(0) < 1$.

Remark 3.17 (Elementary properties of the thresholds). Both thresholds lie in $(0,\infty)$. Indeed, ξ and ξ' are positive on $(0,1)$. The function A is positive and continuous on $(0,1)$, and it tends to $\xi''(0)$ at 0 and to 0 at 1. The static ratio $(-r - \log(1-r))/\xi(r)$ is

positive and continuous on $(0,1)$, and it tends to ∞ at 1 and to $1/\xi''(0)$ at 0, which is ∞ if $\xi''(0) = 0$. Further, $\Phi_\beta(r) \to 0$ as $r \to 0$, and

$$\Phi'_\beta(r) = \beta^2 \xi'(r) - \frac{r}{1-r} = \frac{r}{1-r}\big(\beta^2 A(r) - 1\big), \qquad 0 < r < 1. \tag{3.20}$$

If $\beta \le \beta_{\mathrm{d}}$, then $\beta^2 A \le 1$, and (3.20) implies that Φ_β is nonincreasing and hence nonpositive on $(0,1)$. This means that $\beta^2 \le (-r - \log(1-r))/\xi(r)$ for every r, that is, $\beta \le \beta_{\mathrm{s}}$. We conclude that $\beta_{\mathrm{d}} \le \beta_{\mathrm{s}}$. Part II concerns $\beta > \beta_{\mathrm{s}}$, and Part III concerns $\beta_{\mathrm{d}} < \beta < \beta_{\mathrm{s}}$.

Lemma 3.18 (The replica-symmetric regime)**.** *Let ξ be a finite mixture and $\beta \ge 0$.*

1. *We have $F(\beta) \le \beta^2\xi(1)/2$. Moreover, $F(\beta) = \beta^2\xi(1)/2$ if and only if $\Phi_\beta(r) \le 0$ for all $r \in (0,1)$, if and only if $\beta \le \beta_{\mathrm{s}}$. In this case, δ_0 is the Parisi measure of (ξ,β).*
2. *If $\beta < \beta_{\mathrm{s}}$, then (ξ,β) is strictly replica symmetric with $\beta^2\xi''(0) < 1$.*
3. *If $0 < \beta \le \beta_{\mathrm{s}}$, then for every $\varepsilon > 0$, for all sufficiently large N,*

$$\mathbb{P}\Big(\Big|\frac{1}{N}\log Z_{N,\beta} - \frac{\beta^2\xi(1)}{2}\Big| > \varepsilon\Big) \le 2\exp\Big(-\frac{\varepsilon^2 N}{8\beta^2\xi(1)}\Big). \tag{3.21}$$

For even mixtures, Talagrand proved that the strict form of the criterion in (1) is sufficient and that the non-strict form is necessary [146, Proposition 2.3]; see also [13, (5.2)] for the formula for β_{s}. The proof of the equivalence below uses only Theorem 3.14 and Lemma 3.15, and so it applies to all finite mixtures, including those with odd degrees.

Proof. For (1), by Jensen's inequality and (3.1), we have $F_N(\beta) \le N^{-1}\log \mathbb{E}Z_{N,\beta} = \beta^2\xi(1)/2$ for all N, and $F(\beta) \le \beta^2\xi(1)/2$ follows. For $\zeta = \delta_0$, we have $\hat{x}_{\delta_0}(t) = 1 - t$ and

$$\int_0^s \frac{s-t}{(1-t)^2}\,\mathrm{d}t = \int_0^s \Big(\frac{1}{1-t} - \frac{1-s}{(1-t)^2}\Big)\,\mathrm{d}t = -\log(1-s) - s.$$

Comparing with (3.18) for $\eta = \xi$, we find that $f_{\delta_0} = \Phi_\beta$ and $f_{\delta_0}(0) = 0$. By Lemma 3.15(3), δ_0 is the Parisi measure if and only if $0 \in S_{\delta_0}$, that is, if and only if $\Phi_\beta \le 0$ on $(0,1)$. Recall that $\mathcal{P}_\beta(\delta_0;\xi) = \beta^2\xi(1)/2$. By Theorem 3.14 and the uniqueness of the minimizer, we find that $F(\beta) = \beta^2\xi(1)/2$ if and only if δ_0 is the Parisi measure. Finally, since $\xi(r) > 0$ for $0 < r < 1$, the inequality $\Phi_\beta(r) \le 0$ is equivalent to $\beta^2 \le (-r - \log(1-r))/\xi(r)$. This holds for all $r \in (0,1)$ if and only if $\beta \le \beta_{\mathrm{s}}$.

For (2), let $\beta < \beta_{\mathrm{s}}$. For each $r \in (0,1)$, we have $\beta^2 < \beta_{\mathrm{s}}^2 \le (-r - \log(1-r))/\xi(r)$, which gives $\Phi_\beta(r) < 0$. If $\xi''(0) = 2c_2 > 0$, then $(-r - \log(1-r))/\xi(r) \to 1/\xi''(0)$ as $r \to 0$ by Remark 3.17. This implies $\beta_{\mathrm{s}}^2 \le 1/\xi''(0)$, and therefore $\beta^2\xi''(0) < 1$. If $\xi''(0) = 0$, the inequality $\beta^2\xi''(0) < 1$ is trivial.

For (3), note that $F_N(\beta) \to \beta^2\xi(1)/2$ by (1) and Theorem 3.14. Then $\big|F_N(\beta) - \beta^2\xi(1)/2\big| \le \varepsilon/2$ for all sufficiently large N, and for such N the event in (3.21) is contained in the event $\big|\log Z_{N,\beta} - \mathbb{E}\log Z_{N,\beta}\big| > \varepsilon N/2$. Using Lemma 3.3 with $B = S_N$, $h = 0$, and $t = \varepsilon/2$, we obtain (3.21). □

3.6 Two inputs from Huang, Montanari, and Pham

We state two results of [93] in our notation. Both hold when (ξ, β) is strictly replica symmetric with $\beta^2\xi''(0) < 1$, which by Lemma 3.18(2) is the case for every $\beta < \beta_{\mathrm{s}}$. In [93], the temperature is absorbed into the covariance. The field has covariance $N\eta(R(x,y))$, where $\eta(t) = \sum_{p\geq 2} \gamma_p^2 t^p$ and $\sum_p 2^p\gamma_p^2 < \infty$, and the Gibbs measure has density proportional to e^G with respect to the uniform measure on the sphere of radius $\sqrt{N}$ [93, (1.1)–(1.2)]. For our model at inverse temperature β, this means that $\eta = \beta^2\xi$ and $G = \beta H_N$, and the summability condition holds for finite mixtures. Their condition [93, (2.21)], reproduced as [93, (7.9)], reads

$$\eta''(0) < 1, \qquad \eta(t) + t + \log(1-t) < 0 \quad \text{for all } t \in (0,1). \tag{3.22}$$

For $\eta = \beta^2\xi$, this says exactly that (ξ, β) is strictly replica symmetric with $\beta^2\xi''(0) < 1$.

Lemma 3.19 (Normalized partition function; [93, Lemma 3.2 and Remark 3.3])**.** *Let $\eta(t) = \sum_{p=2}^{P} a_p t^p$ be a polynomial with nonnegative coefficients satisfying* (3.22)*. For $n \geq 2$, let G_n be a centered Gaussian process on S_n with continuous sample paths and covariance $n\eta(R(x,y))$, and let*

$$W_n = e^{-n\eta(1)/2} \int_{S_n} e^{G_n}\, \mathrm{d}\nu_n,$$

so that $\mathbb{E}W_n = 1$. Then W_n converges in distribution to $e^{\sigma g - \sigma^2/2}$, where g is a standard Gaussian variable and $\sigma^2 = -\log(1-\eta''(0))/2$. Moreover, the following hold.

1. *The lower tail of W_n is tight: $\lim_{t\downarrow 0} \limsup_{n\to\infty} \mathbb{P}(W_n < t) = 0$.*
2. *For each n, let Q_n be a probability measure under which G_n has the law above, and let P_n be the probability measure with $\mathrm{d}\mathsf{P}_n/\,\mathrm{d}\mathsf{Q}_n = W_n$. Then for all events A_n and all $t > 0$,*

$$\mathsf{Q}_n(A_n) \leq t^{-1}\mathsf{P}_n(A_n) + \mathsf{Q}_n(W_n < t).$$

   In particular, $\mathsf{P}_n(A_n) \to 0$ implies $\mathsf{Q}_n(A_n) \to 0$.

If (ξ, β) is strictly replica symmetric with $\beta^2\xi''(0) < 1$, then $Z_{N,\beta}/\mathbb{E}Z_{N,\beta}$ converges in distribution to $e^{\sigma g - \sigma^2/2}$ with $\sigma^2 = -\log(1-\beta^2\xi''(0))/2$.

Proof. The convergence in distribution is [93, Lemma 3.2], applied with their mixture equal to η, that is, with $\gamma_p^2 = a_p$. As noted in [93, Remark 3.3], it is proved there under (3.22) alone. The law of W_n depends only on the covariance of G_n, and the realization in [93, (1.1)] may be replaced by any other. As in (3.1), Fubini's theorem gives $\mathbb{E}W_n = 1$. For (1), note that the set $\{w \leq t\}$ is closed. By the portmanteau theorem, we have $\limsup_n \mathbb{P}(W_n < t) \leq \mathbb{P}(e^{\sigma g - \sigma^2/2} \leq t)$, which tends to zero as $t \downarrow 0$ because $e^{\sigma g - \sigma^2/2} > 0$ almost surely. For (2), we have

$$\mathsf{Q}_n(A_n) \leq \mathsf{Q}_n(A_n \cap \{W_n \geq t\}) + \mathsf{Q}_n(W_n < t) \leq t^{-1}\int_{A_n} W_n\, \mathrm{d}\mathsf{Q}_n + \mathsf{Q}_n(W_n < t),$$

and the integral equals $\mathsf{P}_n(A_n)$. The last assertion of (2) follows by letting $n \to \infty$ and then $t \downarrow 0$, and using (1). The final statement is the case $n = N$, $\eta = \beta^2\xi$, and $G_N = \beta H_N$, with $\mathbb{E}Z_{N,\beta}$ given by (3.1). □

Item (2) is the one-sided contiguity of [93, Corollary 3.5], which deduces it from [93, Lemma 3.2] by Le Cam's first lemma. In Section 12, the lemma is applied in dimension $n = N - 1$.

Lemma 3.20 (Replica concentration; [93, Section 7.1 and Lemma 7.7])**.** *Let ξ be a finite mixture and let $\beta > 0$ be such that (ξ, β) is strictly replica symmetric with $\beta^2\xi''(0) < 1$. For every $s \in (0,1)$, there exists $c_s > 0$ such that for all sufficiently large N, with probability at least $1 - e^{-c_s N}$, there exists a Borel set $\mathcal{T}_s \subset S_N$ with*

$$\pi(\mathcal{T}_s^c) \le e^{-c_s N} \quad \text{and} \quad \pi\big(\{y : |R(x,y)| > s\}\big) \le e^{-c_s N} \quad \text{for every } x \in \mathcal{T}_s. \tag{3.23}$$

On the same event,

$$(\pi \otimes \pi)\big(\{(x,y) : |R(x,y)| > s\}\big) \le 2e^{-c_s N}. \tag{3.24}$$

Proof. We apply the results of [93, Section 7.1] to the model [93, (7.7)–(7.8)] with mixture $\eta = \beta^2\xi$ and external field $u = 0$. Quadratic terms and odd degrees are allowed there, and the standing assumption [93, (7.9)] is (3.22), which holds by hypothesis. For $u = 0$, the smallness condition $\|u\| \le u(\delta)\sqrt{N}$ of [93, Lemma 7.7] holds for every δ. Fix $s \in (0,1)$ and take $\delta = s$. The typical set $T(s)$ of [93, (7.31)] consists of the points $x \in S_N$ such that

$$\int_{\{y : |R(x,y)| > s\}} e^{\beta H_N(y)}\, \nu_N(\,\mathrm{d}y) < e^{-c_1(s)N} \min\big\{Z_{N,\beta},\, e^{N\beta^2\xi(1)/2}\big\},$$

for a constant $c_1(s) > 0$. Dividing by $Z_{N,\beta}$, we find that every $x \in T(s)$ satisfies $\pi(\{y : |R(x,y)| > s\}) < e^{-c_1(s)N}$. By [93, (7.34)], with probability at least $1 - e^{-c_2(s)N}$, we have $\pi(T(s)) \ge 1 - e^{-c_2(s)N}$. The set $T(s)$ is Borel, since the function $x \mapsto \int_{\{|R(x,y)|>s\}} e^{\beta H_N(y)} \nu_N(\,\mathrm{d}y)$ is lower semicontinuous by Fatou's lemma. We conclude that (3.23) holds with $\mathcal{T}_s = T(s)$ and $c_s = \min(c_1(s), c_2(s))$. Integrating the second bound in (3.23) over $x \in \mathcal{T}_s$, and bounding the contribution of $\mathcal{T}_s^c$ by its mass, we obtain (3.24). □

The proof of [93, Lemma 7.7] uses that $N^{-1}\log Z_{N,\beta} \to \beta^2\xi(1)/2$ in probability, together with the concentration of $\log Z_{N,\beta}$. Under the hypothesis of the lemma, we have $\Phi_\beta < 0$ on $(0,1)$, hence $\beta \le \beta_{\mathrm{s}}$ by Lemma 3.18(1), and this input is Lemma 3.18(3); it holds for all finite mixtures, including those with odd degrees. The lemma holds for each fixed s, or for finitely many values of s at once, and it gives no rate that is uniform as $s \downarrow 0$.

PART II

The condensed phase

SECTION 4

Total variation lower bounds from Gibbs concentration and hitting estimates

In this section, we reduce the separation from equilibrium of the dynamics started from the uniform law to estimates on the critical points of the center field H_Q and on the Gibbs measure of the bands around them. The main statement is Proposition 4.5. It shows that separation up to a deterministic time T_N follows from three properties of the critical points in a fixed energy window: the union of their bands carries almost all of the Gibbs mass once the window is large; their expected number is bounded; and the probability that the diffusion started from ν_N comes close to one of them before time T_N, averaged with the weights of the Kac–Rice formula, tends to zero. Proposition 4.9 reduces the first property to three estimates on single bands. In Section 4.5, we use the rotation invariance of the disorder to transfer the conclusion from the uniform initial law to every initial law independent of the field (Theorem 4.11), to field-dependent choices among polynomially many prescribed laws (Corollary 4.14), and to field-dependent initial laws whose relative entropy with respect to ν_N is $o(\log N)$ (Corollary 4.15).

No estimate on the critical points or on the Gibbs measure is proved here. The only property of the disorder that we use is the invariance of its law under rotations. From Section 3, we use the representation of H_N by Gaussian tensors (Lemma 3.2), the construction of the dynamics (Lemma 3.5), the contraction and data processing inequalities (Lemma 3.8), binary data processing (Lemma 3.9), and the endpoint estimate of Proposition 3.10. The results of this section are applied in Section 13, where the three properties above are verified using Sections 5 to 12.

4.1 Setting and notation

Throughout this part, $\xi(r) = \sum_p c_p r^p$ is a nonzero finite mixture with $c_p \geq 0$, and $\mathcal{P} = \{p : c_p > 0\}$. For $N \geq 2$, we let $\mathcal{H}_N$ be the finite-dimensional real vector space spanned by the homogeneous polynomials on $\mathbb{R}^N$ whose degrees lie in $\mathcal{P}$, with the topology given by any norm $|\cdot|$. The field H_N is a centered Gaussian random element of $\mathcal{H}_N$ (Lemma 3.2). We write $\mathbb{P}_H$ and $\mathbb{E}_H$ for its law and the corresponding expectation, and H for a generic element of $\mathcal{H}_N$. For $g \in O(N)$, we set $H^g = H \circ g^{-1}$, and for a probability measure μ on S_N we let $g_*\mu = \mu \circ g^{-1}$ be its image under g. The covariance

of H_N depends only on $R(x,y)$, and Lemma 3.2 shows that

$$H_N^g \overset{d}{=} H_N \qquad \text{for every } g \in O(N). \tag{4.1}$$

Fix $\beta \geq 0$. For $H \in \mathcal{H}_N$, we write $\pi^H = \pi^H_{N,\beta}$ for the Gibbs measure, $P_t^H = e^{tL_{\beta,H}}$ for the semigroup, and $P_t^H(x,A)$ for its transition kernel. For a probability measure μ on S_N, we set $(\mu P_t^H)(A) = \int \mu(\,\mathrm{d}x) P_t^H(x,A)$. We let $\mathbb{P}_\mu^H$ be the law of the diffusion $(X_t)_{t\geq 0}$ with generator $L_{\beta,H}$ and initial law μ; we write $\mathbb{P}_x^H$ when $\mu = \delta_x$. The initial point is always independent of the driving noise. For a Borel set $D \subset S_N$, we set

$$Z_H(D) = \int_D e^{\beta H}\,\mathrm{d}\nu_N,$$

so that $Z_{N,\beta} = Z_H(S_N)$ and $\pi^H(D) = Z_H(D)/Z_H(S_N)$.

Fix $Q \in (0,1)$. Recall the center field $H_Q(u) = H(\sqrt{Q}\,u)$ and the radial derivative $K_Q(u) = \frac{\mathrm{d}}{\mathrm{d}t} H(t\sqrt{Q}\,u)\big|_{t=1}$ for $u \in S_N$. Here H is the polynomial itself, and $K_Q(u) = \sqrt{Q}\,u \cdot \nabla H(\sqrt{Q}\,u)$, where ∇ is the Euclidean gradient. We write ∇_{S_N} for the gradient along S_N.

Definition 4.1 (Critical points in a window and their bands)**.** Let $Q \in (0,1)$, let $U \subset \mathbb{R}$ be a compact interval, let $(m_N)_{N\geq 2}$ be a deterministic real sequence, and let $b \in (0, \sqrt{Q})$. For $H \in \mathcal{H}_N$ and $M > 0$, we define

$$\mathcal{C}_{N,M}(H) = \big\{u \in S_N :\ \nabla_{S_N} H_Q(u) = 0,\ |H_Q(u) - m_N| \leq M,\ K_Q(u)/N \in U\big\}.$$

For $u \in S_N$, the band of u is $B_u = \{x \in S_N : |R(x,u) - \sqrt{Q}| \leq b\}$, and we set

$$B_{N,M}(H) = \bigcup_{u \in \mathcal{C}_{N,M}(H)} B_u.$$

For $a \in (0,1)$ and $u \in S_N$, we let $C_a(u) = \{x \in S_N : |R(x,u)| \geq a\}$.

The elements of $\mathcal{C}_{N,M}(H)$ are the critical points of H_Q in the energy window $[m_N - M, m_N + M]$ whose normalized radial derivative lies in U. In Section 13, U is a small interval around K_0 and $m_N = NE_0 + O(\log N)$, where the ground-state energy E_0 and the radial derivative K_0 are given by Proposition 5.11; in this section, U, (m_N), and b are arbitrary. We count points of S_N. When ξ is even, we have $H_Q(-u) = H_Q(u)$ and $K_Q(-u) = K_Q(u)$. Then $\mathcal{C}_{N,M}(H) = -\mathcal{C}_{N,M}(H)$, and both bands B_u and $B_{-u} = -B_u$ enter the union. If $a < \sqrt{Q} - b$, then

$$B_u \subset C_a(u),$$

since every $x \in B_u$ satisfies $R(x,u) \geq \sqrt{Q} - b > a$.

We say that the bands carry almost all of the Gibbs mass if

$$\lim_{M\to\infty} \limsup_{N\to\infty} \mathbb{E}_H\big[\pi^H\big(B_{N,M}(H)^c\big)\big] = 0. \tag{4.2}$$

Because $0 \le \pi^H(B_{N,M}(H)^c) \le 1$, this is equivalent to

$$\lim_{M\to\infty} \limsup_{N\to\infty} \mathbb{P}_H\big\{\pi^H\big(B_{N,M}(H)^c\big) > \eta\big\} = 0 \qquad \text{for every } \eta > 0,$$

by Markov's inequality in one direction and the bound $\mathbb{E}Y \le \eta + \mathbb{P}\{Y > \eta\}$ for $Y \in [0,1]$ in the other. The order of the limits is part of the meaning of (4.2): M is fixed while N tends to infinity, and only afterward does M tend to infinity. The same convention applies to every iterated limit in this part; no parameter is allowed to grow with N.

4.2 Measurability and rotations

The arguments below integrate quantities such as $P_T^H(x, B_{N,M}(H))$ over the field, the initial point, and the rotation group. The next two lemmas show that these quantities are Borel functions and that they transform correctly under rotations.

Lemma 4.2 (Measurable path laws and rotation covariance)**.** *Let $N \ge 2$, $\beta \ge 0$, and $T \in (0,\infty)$. For $H \in \mathcal{H}_N$ and $x \in S_N$, let $\mathbb{P}_x^H$ also denote the law of $(X_t)_{t\in[0,T]}$ on $C([0,T], S_N)$.*

(a) The map $(H,x) \mapsto \mathbb{P}_x^H$ is a Borel probability kernel from $\mathcal{H}_N \times S_N$ to $C([0,T],S_N)$. In particular, $(H,x) \mapsto P_t^H(x,A)$ is Borel for every $t \in [0,T]$ and every Borel set $A \subset S_N$.

(b) For every $g \in O(N)$, $x \in S_N$, $t \ge 0$, and Borel set $A \subset S_N$,

$$P_t^{Hg}(gx, gA) = P_t^H(x,A), \qquad \pi^{Hg} = g_*\pi^H. \tag{4.3}$$

(c) For every Borel probability kernel $H \mapsto \mu^H$ from $\mathcal{H}_N$ to S_N and every $t \ge 0$, the map $H \mapsto \big\|\mu^H P_t^H - \pi^H\big\|_{\mathrm{TV}}$ is Borel, and

$$\inf_{0\le t\le T} \Big\|\mu^H P_t^H - \pi^H\Big\|_{\mathrm{TV}} = \Big\|\mu^H P_T^H - \pi^H\Big\|_{\mathrm{TV}}. \tag{4.4}$$

Proof. We realize the diffusion as the solution of an Itô equation in $\mathbb{R}^N$ whose coefficients depend continuously on H. Part (a) then follows from the continuity of the solution in (H,x), and part (b) from the transformation of the equation under g.

The stochastic differential equation. For $x \in \mathbb{R}^N$, let $\Pi_x = \mathrm{Id} - xx^{\mathsf{T}}/N$; for $x \in S_N$, this is the orthogonal projection onto the tangent space of S_N at x. Fix a smooth function $\chi\colon \mathbb{R}^N \to [0,1]$ that depends only on $|x|$, equals one on $\{N/2 \le |x|^2 \le 2N\}$, and vanishes outside $\{N/4 \le |x|^2 \le 4N\}$. Let W be a standard Brownian motion in $\mathbb{R}^N$, and consider

$$\mathrm{d}X_t = \sqrt{2}\,\chi(X_t)\Pi_{X_t}\,\mathrm{d}W_t + \chi(X_t)\Big(\beta\Pi_{X_t}\nabla H(X_t) - \frac{N-1}{N}X_t\Big)\,\mathrm{d}t, \qquad X_0 = x \in S_N. \tag{4.5}$$

For $x \in S_N$, we have $\Pi_x \nabla H(x) = \nabla_{S_N} H(x)$. Then (4.5) is the equation (3.6) of Lemma 3.5, with a particular extension of its coefficients off a neighborhood of S_N. That lemma shows that (4.5) has a unique strong solution $X^{H,x}$, which stays on S_N and is the diffusion with generator $L_{\beta,H}$ started at x, so that $\mathbb{P}_x^H$ is the law of $X^{H,x}$.

We now show that $X^{H,x}$ depends continuously on (H, x). The coefficients of (4.5) are smooth with compact support in x, and the drift is affine in H. For every $R > 0$, there exists $C = C(N, \beta, R)$ such that, for $|H|, |H'| \le R$, both coefficients are C-Lipschitz in x and the drift changes by at most $C|H - H'|$ when H is replaced by H'. We set $X_t^{(0)} = x$ and define $X^{(k+1)}$ by inserting $X^{(k)}$ on the right side of (4.5). Using Doob's maximal inequality and the Itô isometry, we find

$$\mathbb{E} \sup_{s \le t} \big| X_s^{(k+1)} - X_s^{(k)} \big|^2 \le C' \int_0^t \mathbb{E} \sup_{v \le s} \big| X_v^{(k)} - X_v^{(k-1)} \big|^2 \, \mathrm{d}s,$$

so the left side is at most $C''(C't)^k/k!$. The iterates converge almost surely, uniformly on $[0, T]$, and by uniqueness the limit is $X^{H,x}$. By induction on k, the joint law of $(X^{(k)}, W)$ is the same for all standard Brownian motions W, because the stochastic integral in (4.5) is a limit in probability of Riemann sums, which are continuous functions of $(X^{(k)}, W)$. Hence the law of $X^{H,x}$ does not depend on the choice of the driving Brownian motion. Using the same inequalities and Gronwall's lemma, we obtain

$$\mathbb{E} \sup_{t \le T} \big| X_t^{H,x} - X_t^{H',x'} \big|^2 \le C_T \big(|x - x'|^2 + |H - H'|^2 \big), \qquad |H|, |H'| \le R, \tag{4.6}$$

where both solutions are driven by the same W and C_T depends on N, β, R, and T.

Measurability. If $(H_j, x_j) \to (H, x)$, then (4.6) implies that $X^{H_j,x_j} \to X^{H,x}$ in probability in $C([0,T], \mathbb{R}^N)$, and in particular $\mathbb{P}_{x_j}^{H_j} \to \mathbb{P}_x^H$ weakly. Then $(H, x) \mapsto \int \Phi \, \mathrm{d}\mathbb{P}_x^H$ is continuous for all bounded continuous functions Φ on $C([0,T], S_N)$, and by the monotone class theorem it is Borel for all bounded Borel Φ. This proves (a).

Rotations. Fix $g \in O(N)$. Note that χ depends only on $|x|$, that $\Pi_{gx} = g \Pi_x g^{-1}$, and that $\nabla H^g(gx) = g \nabla H(x)$. Then the process $gX^{H,x}$ solves (4.5) with H^g in place of H, initial point gx, and the Brownian motion gW in place of W. By pathwise uniqueness, $gX^{H,x}$ is the solution of (4.5) for the field H^g and the initial point gx, driven by gW. Since gW is a standard Brownian motion and the law of the solution does not depend on the driving Brownian motion, $gX^{H,x}$ has the law of $X^{H^g,gx}$. Evaluating at time t, we obtain the first identity in (4.3). For the second identity, we use the invariance of ν_N under g:

$$\int_{gA} e^{\beta H^g} \, \mathrm{d}\nu_N = \int_{gA} e^{\beta H(g^{-1}y)} \, \nu_N(\mathrm{d}y) = \int_A e^{\beta H} \, \mathrm{d}\nu_N.$$

Taking $A = S_N$ and dividing, we obtain $\pi^{H^g}(gA) = \pi^H(A)$.

Total variation. Let $\mathcal{A}$ be the algebra generated by a countable base of the topology of S_N; it is countable. For finite Borel measures λ, ρ on S_N, we claim that the supremum of $|\lambda(A) - \rho(A)|$ over $A \in \mathcal{A}$ equals the supremum over all Borel sets. Indeed, given a Borel set A and $\varepsilon > 0$, by regularity there exist a compact set K and an open set O with $K \subset A \subset O$ and $(\lambda + \rho)(O \setminus K) < \varepsilon$. Finitely many base sets contained in O cover K, and their union $A' \in \mathcal{A}$ satisfies $|\lambda(A) - \lambda(A')| + |\rho(A) - \rho(A')| \le (\lambda + \rho)(O \setminus K) < \varepsilon$. For each $A \in \mathcal{A}$, the map $H \mapsto (\mu^H P_t^H)(A)$ is Borel by (a), and $H \mapsto \pi^H(A)$ is continuous by dominated convergence. We deduce that $H \mapsto \|\mu^H P_t^H - \pi^H\|_{\mathrm{TV}}$ is Borel, as a countable supremum of Borel functions. Finally, π^H is invariant for P_t^H by Lemma 3.5, and Lemma 3.8 implies that $t \mapsto \|\mu^H P_t^H - \pi^H\|_{\mathrm{TV}}$ is nonincreasing, which gives (4.4). □

Lemma 4.3 (Measurable and equivariant critical sets)**.** *Let Q, U, (m_N), and b be as in Definition 4.1, and let $M > 0$, $a \in (0,1)$, and $T \in (0,\infty)$.*

(a) The sets $\{(H,u) : u \in \mathcal{C}_{N,M}(H)\}$ and $\{(H,x) : x \in B_{N,M}(H)\}$ are closed in $\mathcal{H}_N \times S_N$. For every $g \in O(N)$ and $H \in \mathcal{H}_N$,

$$\mathcal{C}_{N,M}(H^g) = g\,\mathcal{C}_{N,M}(H), \qquad B_{N,M}(H^g) = g\,B_{N,M}(H). \tag{4.7}$$

(b) The function $\theta(H,u) = \mathbb{P}^H_{\nu_N}\{X_t \in C_a(u) \text{ for some } t \in [0,T]\}$ is Borel on $\mathcal{H}_N \times S_N$, and the probability $\mathbb{P}^H_{\nu_N}\{X_t \in C_a(u) \text{ for some } t \in [0,T] \text{ and some } u \in \mathcal{C}_{N,M}(H)\}$ is a Borel function of H.

(c) The set of H for which $\mathcal{C}_{N,M}(H)$ is finite is Borel, and $S(H) = \sum_{u \in \mathcal{C}_{N,M}(H)} \theta(H,u) \in [0,\infty]$ is a Borel function of H; it equals $+\infty$ when $\mathcal{C}_{N,M}(H)$ is infinite. Moreover, for all H,

$$\begin{aligned}\sup_{0\le t\le T} (\nu_N P_t^H)\Big(\bigcup_{u\in\mathcal{C}_{N,M}(H)} C_a(u)\Big) &\le \mathbb{P}^H_{\nu_N}\{X_t \in C_a(u) \text{ for some } t \le T \text{ and } u \in \mathcal{C}_{N,M}(H)\} \\ &\le S(H).\end{aligned} \tag{4.8}$$

Proof. For (a), we first note that the maps $(H,u) \mapsto H_Q(u)$, $K_Q(u)$, and $\nabla_{S_N} H_Q(u)$ are continuous, because the coefficients of these polynomials in u depend linearly on H. The first set is closed because U is closed. The second set is the projection, along the compact factor S_N of the variable u, of the closed set $\{(H,u,x) : u \in \mathcal{C}_{N,M}(H),\ |R(x,u) - \sqrt{Q}| \le b\}$. Such a projection is closed: if $(H_j, x_j) \to (H,x)$ with witnesses u_j, we may pass to a subsequence along which u_j converges to some u, and then (H,u,x) belongs to the closed set. For the equivariance, note that $(H^g)_Q(v) = H_Q(g^{-1}v)$ for $v \in S_N$, so that

$$(H^g)_Q(gu) = H_Q(u), \qquad (K^g)_Q(gu) = K_Q(u), \qquad \nabla_{S_N}(H^g)_Q(gu) = g\nabla_{S_N} H_Q(u),$$

where $(K^g)_Q$ is the radial derivative of H^g. From these identities, we see that $u \in \mathcal{C}_{N,M}(H)$ if and only if $gu \in \mathcal{C}_{N,M}(H^g)$. Since $R(gx,gu) = R(x,u)$, we have $B_{gu} = gB_u$, and the second identity in (4.7) follows.

For (b), the map $(\omega, u) \mapsto \max_{t\le T} |R(\omega(t), u)|$ is continuous on $C([0,T], S_N) \times S_N$, and the set $\{(\omega,u) : \max_{t\le T} |R(\omega(t),u)| \ge a\}$ is closed. Integrating its indicator against the Borel kernel $\mathbb{P}_x^H$ of Lemma 4.2(a), and then against $\nu_N(\,\mathrm{d}x)$, we find that θ is Borel. The set of pairs (H,ω) for which some $u \in \mathcal{C}_{N,M}(H)$ satisfies $\max_{t\le T} |R(\omega(t),u)| \ge a$ is closed, by the projection argument of (a), and the same integration proves the second assertion. The first assertion is also the case $\mathcal{A} = \{(\omega,u) : \max_{t\le T} |R(\omega(t),u)| \ge a\}$ of Lemma 7.21.

For (c), let $O \subset S_N$ be open and let $O_k = \{x \in S_N : \operatorname{dist}(x, O^c) \ge 1/k\}$. For $l \in \mathbb{N}$, the event $\{\#(\mathcal{C}_{N,M}(H) \cap O) \ge l\}$ is the union over k of the events that there exist $u_1, \dots, u_l \in \mathcal{C}_{N,M}(H) \cap O_k$ with $|u_i - u_j| \ge 1/k$ for $i \ne j$. Each of these is the projection along the compact factor $(S_N)^l$ of a closed set, and it is closed. Hence $H \mapsto \#(\mathcal{C}_{N,M}(H) \cap O)$ is Borel for all open O, and the set F where $\mathcal{C}_{N,M}(H)$ is finite is Borel. On $F_l = \{H : \#\mathcal{C}_{N,M}(H) \le l\}$, the counting measure $n_H = \sum_{u \in \mathcal{C}_{N,M}(H)} \delta_u$ is finite. The Borel sets A for which $H \mapsto n_H(A)\mathbf{1}_{F_l}(H)$ is Borel form a class that contains the open sets and is closed under proper differences and increasing unions; by Dynkin's theorem, it contains all Borel sets. Applying the monotone class theorem to θ, starting from products $\mathbf{1}_{A_1}(H)\mathbf{1}_{A_2}(u)$, we find that $H \mapsto \int \theta(H,u)\, n_H(\,\mathrm{d}u) = S(H)$ is Borel on each F_l. Off F, every term satisfies $\theta(H,u) \ge \mathbb{P}^H_{\nu_N}\{X_0 \in C_a(u)\} = \nu_N(C_a(u))$, and $\nu_N(C_a(u)) > 0$ does not depend on u, because $a < 1$ and ν_N is rotation invariant. Then $S = +\infty$ off F, and S is Borel. In (4.8), the first inequality holds because $X_t \in \bigcup_u C_a(u)$ for one $t \le T$ is contained in the event on the right. The second is countable subadditivity when $\mathcal{C}_{N,M}(H)$ is finite, and it is trivial otherwise. □

4.3 Separation from Gibbs concentration, counts, and hitting bounds

Each of the separation statements in this part ends with the same order of limits: a window half-width M is fixed for a given tolerance, N tends to infinity, and only then is the tolerance removed. We isolate this step.

Lemma 4.4 (Testing the band union)**.** *Assume that the bands carry almost all of the Gibbs mass,* (4.2)*. Let $\Delta_N(H) \in [0,1]$ be random variables. Suppose that, for every fixed $M > 0$, there exist random variables $\mathcal{E}_{N,M}(H) \ge 0$ with $\mathcal{E}_{N,M} \to 0$ in $\mathbb{P}_H$-probability as $N \to \infty$, such that*

$$\Delta_N(H) \ge 1 - \pi^H\big(B_{N,M}(H)^c\big) - \mathcal{E}_{N,M}(H) \qquad \text{for all } N \text{ and } H.$$

Then $\Delta_N \to 1$ in $\mathbb{P}_H$-probability.

Proof. Fix $\varepsilon, \delta \in (0,1)$. By (4.2), there exists M such that $\limsup_N \mathbb{E}_H \pi^H(B^c_{N,M}) < \varepsilon\delta/4$. Using Markov's inequality, we find that $\limsup_N \mathbb{P}_H\{\pi^H(B^c_{N,M}) > \varepsilon/2\} \le \delta/2$. With this M fixed, $\mathbb{P}_H\{\mathcal{E}_{N,M} > \varepsilon/2\} \to 0$. Outside these two events we have $\Delta_N \ge 1 - \varepsilon$, so $\limsup_N \mathbb{P}_H\{\Delta_N < 1 - \varepsilon\} \le \delta/2$. Since δ is arbitrary, this completes the proof. □

In the applications, $\Delta_N = \inf_{0\le t\le T_N} \|\mu^H P_t^H - \pi^H\|_{\mathrm{TV}}$ for an initial law μ^H, and Proposition 3.10 implies that $\Delta_N \ge 1 - \pi^H(B_{N,M}^c) - (\mu^H P_{T_N}^H)(B_{N,M})$. The next proposition is the reduction for the uniform initial law.

Proposition 4.5 (Separation from Gibbs concentration, counts, and hitting bounds)**.** *Let $\beta > 0$, let Q, U, (m_N), and b be as in Definition 4.1, let $a \in (0, \sqrt{Q} - b)$, and let (T_N) be deterministic times. For $u \in S_N$, set*

$$\theta_N(H,u) = \mathbb{P}_{\nu_N}^H\big\{X_t \in C_a(u) \text{ for some } t \in [0,T_N]\big\}.$$

Assume the following.

(i) The bands carry almost all of the Gibbs mass, (4.2)*.*

(ii) For every fixed $M > 0$, $A_M = \limsup_{N\to\infty} \mathbb{E}_H|\mathcal{C}_{N,M}(H)| < \infty$.

(iii) For every fixed $M > 0$, there exist numbers $q_{N,M} \ge 0$ with $\lim_{N\to\infty} q_{N,M} = 0$ such that, for all sufficiently large N,

$$\mathbb{E}_H \sum_{u\in\mathcal{C}_{N,M}(H)} \theta_N(H,u) \le q_{N,M}\, \mathbb{E}_H|\mathcal{C}_{N,M}(H)|. \tag{4.9}$$

For $M > 0$, let $S_{N,M}(H) = \sum_{u\in\mathcal{C}_{N,M}(H)} \theta_N(H,u)$. Then $\lim_{N\to\infty} \mathbb{E}_H S_{N,M} = 0$ for each fixed M, and, for all H,

$$\begin{aligned} \sup_{0\le t\le T_N} (\nu_N P_t^H)\big(B_{N,M}(H)\big) &\le S_{N,M}(H), \\ \inf_{0\le t\le T_N} \Big\|\nu_N P_t^H - \pi^H\Big\|_{\mathrm{TV}} &\ge 1 - \pi^H\big(B_{N,M}(H)^c\big) - S_{N,M}(H). \end{aligned} \tag{4.10}$$

In particular, $\inf_{0\le t\le T_N} \|\nu_N P_t^H - \pi^H\|_{\mathrm{TV}} \to 1$ in $\mathbb{P}_H$-probability.

Proof. Fix M. By (ii) and (iii), we have $\mathbb{E}_H S_{N,M} \le q_{N,M}(A_M + 1)$ for all sufficiently large N, and the right side tends to zero. Because $a < \sqrt{Q} - b$, every band satisfies $B_u \subset C_a(u)$, so that $B_{N,M}(H) \subset \bigcup_{u\in\mathcal{C}_{N,M}(H)} C_a(u)$. The first bound in (4.10) now follows from (4.8). For every $t \le T_N$, we test the set $B_{N,M}(H)$ in the definition of the total variation distance and obtain

$$\begin{aligned} \Big\|\nu_N P_t^H - \pi^H\Big\|_{\mathrm{TV}} &\ge \pi^H\big(B_{N,M}(H)\big) - (\nu_N P_t^H)\big(B_{N,M}(H)\big) \\ &\ge 1 - \pi^H\big(B_{N,M}(H)^c\big) - S_{N,M}(H), \end{aligned}$$

which is the second bound. Using Markov's inequality, we find that $S_{N,M} \to 0$ in $\mathbb{P}_H$-probability for each fixed M. Lemma 4.4, applied with $\Delta_N = \inf_{t\le T_N} \|\nu_N P_t^H - \pi^H\|_{\mathrm{TV}}$ and $\mathcal{E}_{N,M} = S_{N,M}$, completes the proof. The function Δ_N is Borel by Lemma 4.2(c), and $S_{N,M}$ is Borel by Lemma 4.3(c). □

Remark 4.6 (The weights in hypothesis (iii)). The left side of (4.9) is the expected sum of the hitting probabilities over the critical points in the window. By the marked Kac–Rice formula of Theorem 7.17, we can write it as the integral, against the intensity measure of $\mathcal{C}_{N,M}$, of the expectation of $\theta_N(H,u)$ under the law of the field conditioned on a critical point at u with prescribed marks, reweighted by $|\det \nabla^2_{S_N} H_Q(u)|$ and normalized to a probability measure. Since the total mass of the intensity measure is $\mathbb{E}_H|\mathcal{C}_{N,M}|$, (4.9) holds whenever this determinant-weighted conditional hitting probability is at most $q_{N,M}$ at every point of the window. The hitting estimate for the caps under critical-point conditioning in Section 8 bounds the hitting probability under the ordinary Gaussian conditioning on the values of $H_Q(u)$, $K_Q(u)$, and $\nabla_{S_N} H_Q(u)$, without the determinant. The weight cannot simply be dropped, since the determinant and the trajectory both depend on the whole field. Lemma 7.22 converts the ordinary bound into (4.9), at the cost of a power, using the moments of the normalized determinant from Proposition 7.7. This is carried out in Proposition 13.1.

Remark 4.7 (The order of the parameters). In Proposition 4.5, the times T_N do not depend on M, and hypothesis (iii) must hold for every fixed M with the same T_N. In Section 13, $T_N = e^{\gamma N}$, and γ comes from a hitting estimate that holds uniformly over a fixed compact rectangle of normalized marks $(h,k) = (H_Q(u)/N, K_Q(u)/N)$. When m_N/N converges, this rectangle contains the marks $[(m_N - M)/N, (m_N + M)/N] \times U$ of every fixed window once N is large enough, depending on M. We therefore fix the rectangle, a, and γ first. The half-width M is chosen afterward, depending on the tolerance, and only the constants and the threshold for N in (iii) depend on M. A hitting estimate whose time scale depended on M would not suffice, because (4.2) lets M tend to infinity after N.

Remark 4.8 (An endpoint formulation). For separation alone, the bound on the whole path in (4.10) is more than needed. If $\lim_{N\to\infty} \mathbb{E}_H(\nu_N P^H_{T_N})(B_{N,M}(H)) = 0$ for each fixed M, then the same conclusion follows from Proposition 3.10, Markov's inequality, and Lemma 4.4. The results of Section 4.5 use this endpoint form.

4.4 Concentration on the bands of a fixed window from a coarse cover

The hypothesis (4.2) concerns critical points in a fixed window $[m_N - M, m_N + M]$. In the next proposition, we derive it from three estimates that are easier to verify directly: the bands of a coarser family of critical points carry almost all of the Gibbs mass; the band of a critical point in a fixed window is rarely much lighter than its energy suggests; and the sum of the partition functions $Z_H(B_u)$ of the bands of critical points with $H_Q(u) < m_N - M$ decays exponentially in M. The estimates of Sections 9 and 12 have this form.

Proposition 4.9 (Concentration on the bands of a fixed window from a coarse cover)**.** *Let $\beta > 0$ and let Q, U, (m_N), and b be as in Definition 4.1. Assume that*

$$\lim_{L\to\infty} \liminf_{N\to\infty} \mathbb{P}_H\Big\{\mathcal{C}_{N,L}(H) \neq \varnothing \text{ and } \max_{S_N} H_Q \leq m_N + L\Big\} = 1. \tag{4.11}$$

Let $D_N > 0$ be deterministic, and let $\mathcal{C}_N^{\mathrm{co}}(H)$ be finite sets of critical points u of H_Q with $K_Q(u)/N \in U$, whose graphs $\{(H,u) : u \in \mathcal{C}_N^{\mathrm{co}}(H)\}$ are Borel. Assume the following.

(i) (Coarse cover.) $\lim_{N\to\infty} \mathbb{E}_H\, \pi^H\Big(\Big[\bigcup_{u\in\mathcal{C}_N^{\mathrm{co}}(H)} B_u\Big]^c\Big) = 0.$

(ii) (Light bands are rare.) For every fixed $L > 0$,

$$\lim_{\delta\downarrow 0} \limsup_{N\to\infty} \mathbb{E}_H \sum_{u\in\mathcal{C}_{N,L}(H)} \mathbf{1}\big\{Z_H(B_u) < \delta D_N e^{\beta(H_Q(u)-m_N)}\big\} = 0.$$

(iii) (Critical points below the window.) There exist $C_*, c_* > 0$ such that, for every fixed $M > 0$,

$$\limsup_{N\to\infty} D_N^{-1}\, \mathbb{E}_H \sum_{\substack{u\in\mathcal{C}_N^{\mathrm{co}}(H)\\ H_Q(u)<m_N-M}} Z_H(B_u) \leq C_* e^{-c_* M}.$$

Then the bands carry almost all of the Gibbs mass, (4.2)*.*

We call the band B_u of a point $u \in \mathcal{C}_{N,L}(H)$ *light* if $Z_H(B_u) < \delta D_N e^{\beta(H_Q(u)-m_N)}$, so that (ii) bounds the expected number of light bands. The quantities in (i)–(iii) are Borel functions of H. Indeed, the map $(H,u) \mapsto Z_H(B_u)$ is continuous, by dominated convergence, since the two latitude spheres bounding B_u have ν_N-measure zero. For fixed N, L, and δ, let $f(H,u) = \delta D_N e^{\beta(H_Q(u)-m_N)} - Z_H(B_u)$. The event that at least l points $u \in \mathcal{C}_{N,L}(H)$ satisfy $f(H,u) > 0$ is the union over $k \geq 1$ of the projections, along the compact factor $(S_N)^l$, of the closed sets defined by $u_i \in \mathcal{C}_{N,L}(H)$, $f(H,u_i) \geq 1/k$, and $|u_i - u_j| \geq 1/k$ for $i \neq j$. These projections are closed, so the count in (ii) is Borel, also on fields with infinitely many critical points. The graph of $\mathcal{C}_N^{\mathrm{co}}$ is a Borel set with finite sections. By the Lusin–Novikov uniformization theorem [98, Theorem 18.10], it is a countable disjoint union of graphs of Borel maps u_j defined on Borel subsets of $\mathcal{H}_N$. Summing over this enumeration shows that the sum in (iii) is Borel, and that the union of the coarse bands has a Borel graph in (H,x), so that its Gibbs mass in (i) is Borel.

Proof. We first use (ii) and the existence part of (4.11) to bound $Z_H(S_N)$ from below by one band. Then we split the missing Gibbs mass into the part missed by the coarse family and the part carried by the bands of critical points below the window.

A lower bound for the partition function. Fix $\varepsilon \in (0,1)$. By (4.11), there exists L such that $\limsup_N \mathbb{P}_H\{\mathcal{C}_{N,L}(H) = \varnothing\} < \varepsilon$. With L fixed, by (ii) there exists $\delta > 0$ such that the expected number of $u \in \mathcal{C}_{N,L}(H)$ with $Z_H(B_u) < \delta D_N e^{\beta(H_Q(u)-m_N)}$ has lim sup in N below ε. Using Markov's inequality, we see that the probability that there exists such a u has lim sup below ε. If $\mathcal{C}_{N,L}(H)$ contains a point u that is not of this kind, then $H_Q(u) \geq m_N - L$, and

$$Z_H(S_N) \geq Z_H(B_u) \geq \delta D_N e^{\beta(H_Q(u)-m_N)} \geq \eta D_N, \qquad \eta = \delta e^{-\beta L}.$$

We conclude that

$$\limsup_{N\to\infty} \mathbb{P}_H\big\{Z_H(S_N) < \eta D_N\big\} < 2\varepsilon. \tag{4.12}$$

This uses only the expected number of light bands; no band is selected in a way that depends on its determinant or on the field.

Critical points below the window. Let $G_{N,M} = \{\max_{S_N} H_Q \leq m_N + M\}$. Note that $\lim_{M\to\infty} \limsup_N \mathbb{P}_H(G_{N,M}^c) = 0$ by (4.11). On $G_{N,M}$, every $u \in \mathcal{C}_N^{\mathrm{co}}(H) \setminus \mathcal{C}_{N,M}(H)$ satisfies $H_Q(u) < m_N - M$. Indeed, u is a critical point with $K_Q(u)/N \in U$, and $u \notin \mathcal{C}_{N,M}(H)$ means that $|H_Q(u) - m_N| > M$, while $H_Q(u) \leq m_N + M$ on $G_{N,M}$. It follows that, on $G_{N,M}$,

$$B_{N,M}(H)^c \subset \Big[\bigcup_{u\in\mathcal{C}_N^{\mathrm{co}}(H)} B_u\Big]^c \cup \bigcup_{\substack{u\in\mathcal{C}_N^{\mathrm{co}}(H)\\ H_Q(u)<m_N-M}} B_u,$$

and on $G_{N,M} \cap \{Z_H(S_N) \geq \eta D_N\}$ we obtain

$$\pi^H\big(B_{N,M}(H)^c\big) \leq \pi^H\Big(\Big[\bigcup_{u\in\mathcal{C}_N^{\mathrm{co}}(H)} B_u\Big]^c\Big) + \frac{1}{\eta D_N} \sum_{\substack{u\in\mathcal{C}_N^{\mathrm{co}}(H)\\ H_Q(u)<m_N-M}} Z_H(B_u). \tag{4.13}$$

A point covered by several bands is counted several times in the last sum, which only increases it; we therefore do not need the bands to be disjoint.

Conclusion. We bound $\pi^H(B_{N,M}^c)$ by one outside $G_{N,M} \cap \{Z_H(S_N) \geq \eta D_N\}$, take expectations in (4.13), and let $N \to \infty$ with M fixed. Using (i), (4.12), and (iii), we find

$$\limsup_{N\to\infty} \mathbb{E}_H \pi^H\big(B_{N,M}(H)^c\big) \leq 2\varepsilon + \limsup_{N\to\infty} \mathbb{P}_H(G_{N,M}^c) + \eta^{-1} C_* e^{-c_* M}.$$

Letting $M \to \infty$ with L, δ, and η fixed, the right side tends to 2ε. Since ε is arbitrary, (4.2) follows. The parameters are chosen in the order L, δ, M, and each is fixed before $N \to \infty$. □

Remark 4.10 (On hypothesis (ii)). Hypothesis (ii) bounds the expected number of critical points in a fixed window whose band is light. Suppose that the counts in the window

$\mathcal{C}_{N,L}$ are bounded as in Proposition 4.5(ii), and that the determinant-weighted conditional probability of $\{Z_H(B_u) < \delta D_N e^{\beta(H_Q(u)-m_N)}\}$ is at most $\rho_L(\delta) + o_N(1)$, uniformly over the points and marks of this window, where $\rho_L(\delta) \to 0$ as $\delta \downarrow 0$. Using the marked Kac–Rice formula of Theorem 7.17, we find that the expected number of light bands is at most $(A_L + o_N(1))(\rho_L(\delta) + o_N(1))$, which gives (ii). If the existence part of (4.11) is known only for a subfamily of $\mathcal{C}_{N,L}(H)$, for instance the critical points on which an additional selection event holds, the proof applies with that subfamily in the existence event. The light bands are still counted over all of $\mathcal{C}_{N,L}(H)$, which can only increase their number, and no conditional law normalized by the probability of the selection event is needed.

4.5 Initial laws other than the uniform law

We now transfer separation from the uniform initial law to other initial laws. The first result uses only the rotation invariance (4.1). It is an instance of the classical symmetrization over a compact group: the expected distance to equilibrium is a convex and rotation-invariant function of the initial law, so it is smallest at the rotation average of any initial law, which is ν_N. The same averaging over a group underlies the Hunt–Stein theorem in statistics; see Lehmann and Romano [113, Section 8.5].

Theorem 4.11 (Initial laws independent of the field)**.** *Let $\beta > 0$ and $N \geq 2$. For every Borel probability measure μ on S_N and every $T \geq 0$,*

$$\mathbb{E}_H\big[1 - \big\|\mu P_T^H - \pi^H\big\|_{\mathrm{TV}}\big] \leq \mathbb{E}_H\big[1 - \big\|\nu_N P_T^H - \pi^H\big\|_{\mathrm{TV}}\big]. \tag{4.14}$$

Further, for deterministic times T_N and every $\varepsilon > 0$,

$$\sup_{\mu} \mathbb{P}_H\Big\{\inf_{0\leq t\leq T_N} \big\|\mu P_t^H - \pi^H\big\|_{\mathrm{TV}} < 1 - \varepsilon\Big\} \leq \varepsilon^{-1}\, \mathbb{E}_H\big[1 - \big\|\nu_N P_{T_N}^H - \pi^H\big\|_{\mathrm{TV}}\big], \tag{4.15}$$

where the supremum is over all Borel probability measures μ on S_N. In particular, if $\inf_{0\leq t\leq T_N} \|\nu_N P_t^H - \pi^H\|_{\mathrm{TV}} \to 1$ in $\mathbb{P}_H$-probability, then the left side of (4.15) *tends to zero for every $\varepsilon > 0$. The same convergence holds for every random probability measure μ_N on S_N that is independent of H_N, in the joint probability of H_N and μ_N.*

Proof. We show that the expected distance is invariant under rotations of the initial law and convex in it, and then average over the rotation group.

Averaging over rotations. Let $\mathrm{d}g$ be the normalized Haar measure on $O(N)$, which is invariant under left and right multiplication. For a bounded Borel function f on S_N, let $\varphi(x) = \int f(gx)\,\mathrm{d}g$. For $h \in O(N)$, we have $\varphi(hx) = \int f(ghx)\,\mathrm{d}g = \varphi(x)$ by right invariance. The group $O(N)$ acts transitively on S_N, and φ is constant. Integrating

against ν_N and using Fubini's theorem and the rotation invariance of ν_N, we find that the constant is $\int f \, \mathrm{d}\nu_N$. Since f is arbitrary, we obtain $\iint f(gx) \, \mu(\mathrm{d}x) \, \mathrm{d}g = \int f \, \mathrm{d}\nu_N$, that is,

$$\int_{O(N)} g_* \mu \, \mathrm{d}g = \nu_N. \tag{4.16}$$

Invariance and convexity. For a probability measure λ on S_N, let $F(\lambda) = \mathbb{E}_H \left\| \lambda P_T^H - \pi^H \right\|_{\mathrm{TV}}$; the integrand is Borel by Lemma 4.2(c). Fix $g \in O(N)$. By (4.3), we have $(g_*\mu) P_T^{Hg} = g_*(\mu P_T^H)$ and $\pi^{Hg} = g_* \pi^H$. As g_* preserves the total variation distance, it follows that

$$\left\| (g_*\mu) P_T^{Hg} - \pi^{Hg} \right\|_{\mathrm{TV}} = \left\| \mu P_T^H - \pi^H \right\|_{\mathrm{TV}}.$$

Taking expectations and using (4.1), we obtain $F(g_*\mu) = F(\mu)$. On the other hand, by (4.16) and Fubini's theorem, we have $\nu_N P_T^H = \int (g_*\mu) P_T^H \, \mathrm{d}g$ for all H. For every Borel set A, we therefore have

$$\left| (\nu_N P_T^H)(A) - \pi^H(A) \right| \le \int \left| ((g_*\mu) P_T^H)(A) - \pi^H(A) \right| \mathrm{d}g \le \int \left\| (g_*\mu) P_T^H - \pi^H \right\|_{\mathrm{TV}} \mathrm{d}g.$$

Taking the supremum over A, we find $\left\| \nu_N P_T^H - \pi^H \right\|_{\mathrm{TV}} \le \int \left\| (g_*\mu) P_T^H - \pi^H \right\|_{\mathrm{TV}} \mathrm{d}g$. The integrand is jointly Borel in (H, g): the map $(H, g, x) \mapsto P_T^H(gx, A)$ is Borel by Lemma 4.2(a), and the distance is a supremum over the countable algebra $\mathcal{A}$ in the proof of that lemma. Taking expectations and using Tonelli's theorem, we conclude that $F(\nu_N) \le \int F(g_*\mu) \, \mathrm{d}g = F(\mu)$, which is (4.14).

Uniformity. Recall from (4.4) that the infimum over $t \le T_N$ in (4.15) equals the distance at T_N. Applying Markov's inequality to the nonnegative deficit $1 - \left\| \mu P_{T_N}^H - \pi^H \right\|_{\mathrm{TV}}$ and using (4.14), we obtain (4.15) for each μ; the bound does not depend on μ. If the deficit $Y_N = 1 - \left\| \nu_N P_{T_N}^H - \pi^H \right\|_{\mathrm{TV}} \in [0, 1]$ of the uniform initial law tends to zero in probability, then $\mathbb{E}_H Y_N \to 0$, because $\mathbb{E}_H Y_N \le \eta + \mathbb{P}_H\{Y_N > \eta\}$ for every $\eta > 0$. For a random initial law μ_N independent of H_N, we condition on μ_N, apply (4.15) to its realization, and integrate. The map $(H, \mu) \mapsto \left\| \mu P_{T_N}^H - \pi^H \right\|_{\mathrm{TV}}$ is jointly measurable, by the countable supremum above and the measurability of $\mu \mapsto \int f \, \mathrm{d}\mu$ for bounded Borel f. □

Remark 4.12 (The order of the supremum and the probability). The supremum in (4.15) is outside the probability: for each initial law fixed before the field is drawn, the probability of failure is small, uniformly in the law. The theorem does not give one event of high probability on which the conclusion holds for every initial law, and no such event exists: the field-dependent choice $\mu = \pi^H$ is at distance zero from π^H at all times. The remaining results of this section allow field-dependent initial laws under explicit restrictions.

Lemma 4.13 (Equal expected mass from every starting point). *Let $T \geq 0$, and let $B(H) \subset S_N$ be sets such that $\{(H,x) : x \in B(H)\}$ is Borel and $B(H^g) = gB(H)$ for all $g \in O(N)$ and $H \in \mathcal{H}_N$. Set $q = \mathbb{E}_H(\nu_N P_T^H)(B(H_N))$. Then, for every $x \in S_N$ and every Borel probability measure μ on S_N,*

$$\mathbb{E}_H P_T^H\big(x, B(H)\big) = q, \qquad \mathbb{E}_H(\mu P_T^H)\big(B(H)\big) = q. \tag{4.17}$$

Proof. The function $\varrho(H,x) = P_T^H(x, B(H))$ is Borel on $\mathcal{H}_N \times S_N$, since it is the integral of the indicator of the Borel graph of B against the kernel of Lemma 4.2(a). By (4.3) and the equivariance of B, we have $\varrho(H^g, gx) = P_T^{H^g}(gx, gB(H)) = \varrho(H,x)$. Then (4.1) implies that $\mathbb{E}_H\varrho(H,gx) = \mathbb{E}_H\varrho(H^g,gx) = \mathbb{E}_H\varrho(H,x)$. Because $O(N)$ acts transitively on S_N, the function $x \mapsto \mathbb{E}_H\varrho(H,x)$ is constant. Integrating it against ν_N and using Fubini's theorem identifies the constant as q, and integrating it against μ gives the second identity. □

For the remaining results, we let Q, U, (m_N), and b be as in Definition 4.1. Recall from Lemma 4.3(a) that the band unions $B_{N,M}(H)$ have Borel graphs and satisfy $B_{N,M}(H^g) = gB_{N,M}(H)$; in particular, Lemma 4.13 applies to them. For deterministic times T_N, we write

$$q_{N,M} = \mathbb{E}_H(\nu_N P_{T_N}^H)\big(B_{N,M}(H)\big).$$

In Proposition 13.1, this endpoint mass is shown to be at most a negative power of N for each fixed M. Such a bound leaves room for initial laws that depend on the field in a limited way. We will use the bound

$$q_{N,M} \leq C_M N^{-\zeta} \qquad \text{for all sufficiently large } N, \tag{4.18}$$

where $\zeta > 0$ and, for each fixed $M > 0$, C_M is a constant.

The next result allows a field-dependent choice among K_N prescribed initial laws. It uses a union bound and allows K_N to grow polynomially when (4.18) holds.

Corollary 4.14 (Selection among prescribed initial laws). *Let $\beta > 0$, let (T_N) be deterministic times, and suppose that the bands carry almost all of the Gibbs mass,* (4.2)*. For each N, let $\mu_{N,1}, \dots, \mu_{N,K_N}$ be Borel probability measures on S_N that do not depend on H. If $K_N q_{N,M} \to 0$ as $N \to \infty$ for every fixed $M > 0$, then*

$$\inf\Big\{ \Big\| \Big(\sum_{j=1}^{K_N} w_j \mu_{N,j} \Big) P_t^H - \pi^H \Big\|_{\mathrm{TV}} : \ w_j \geq 0, \ \sum_{j=1}^{K_N} w_j = 1, \ 0 \leq t \leq T_N \Big\} \longrightarrow 1 \tag{4.19}$$

in $\mathbb{P}_H$-probability. In particular, the conclusion holds for weights $w_j = w_j(H)$ that are arbitrary Borel functions of the field, and for a field-dependent choice of one of the laws $\mu_{N,j}$. For collections that are random but independent of H_N, with the same deterministic K_N, the conclusion holds in the joint probability.

Proof. Fix $M > 0$ and set $a_j(H) = (\mu_{N,j} P^H_{T_N})(B_{N,M}(H))$. By Lemma 4.13, $\mathbb{E}_H a_j = q_{N,M}$ for all j, and

$$\mathbb{E}_H \max_{j \le K_N} a_j(H) \le \sum_{j \le K_N} \mathbb{E}_H a_j(H) = K_N q_{N,M} \longrightarrow 0.$$

For every choice of weights, the mass of $B_{N,M}(H)$ under $(\sum_j w_j \mu_{N,j}) P^H_{T_N}$ is $\sum_j w_j a_j(H) \le \max_j a_j(H)$. Applying Proposition 3.10 to each mixture, we find that the infimum $\Delta_N(H)$ in (4.19) satisfies

$$\Delta_N(H) \ge 1 - \pi^H\big(B_{N,M}(H)^c\big) - \max_{j \le K_N} a_j(H).$$

Then Lemma 4.4, with $\mathcal{E}_{N,M} = \max_j a_j$, implies (4.19). The infimum Δ_N is a Borel function of H. Indeed, by (4.4) it is the infimum over the weights of the distance at T_N; this distance changes by at most $\sum_j |w_j - w'_j|$ when w is replaced by w', so the infimum may be taken over the countable set of rational weights. For an independent random collection, we condition on its realization. The bound $\mathbb{E}_H \max_j a_j \le K_N q_{N,M}$ holds for all realizations, and it can be integrated. □

Finally, we measure the dependence of the initial law on the field by its relative entropy with respect to ν_N, which may be large on a set of fields of small probability.

Corollary 4.15 (Field-dependent initial laws with small relative entropy)**.** *Let $\beta > 0$ and let (T_N) be deterministic times. Assume that the bands carry almost all of the Gibbs mass,* (4.2)*, and that* (4.18) *holds for some $\zeta > 0$. Let $H \mapsto \mu^H_N$ be a Borel probability kernel from $\mathcal{H}_N$ to S_N such that*

$$\frac{D(\mu^H_N \| \nu_N)}{\log N} \longrightarrow 0 \qquad \text{in } \mathbb{P}_H\text{-probability.} \tag{4.20}$$

Then $\inf_{0 \le t \le T_N} \big\| \mu^H_N P^H_t - \pi^H \big\|_{\mathrm{TV}} \to 1$ in $\mathbb{P}_H$-probability.

Proof. Fix $M > 0$ and $\tau \in (0, \zeta)$, and let $a_{N,M}(H) = (\nu_N P^H_{T_N})(B_{N,M}(H))$. By Markov's inequality and (4.18), we have

$$\mathbb{P}_H\big\{a_{N,M} > N^{-\tau}\big\} \le C_M N^{\tau - \zeta} \longrightarrow 0.$$

Fix a field H with $a_{N,M}(H) \le N^{-\tau}$ and $D(\mu^H_N \| \nu_N) < \infty$. If $a_{N,M}(H) = 0$, then by Proposition 3.10 we have

$$\inf_{0 \le t \le T_N} \Big\| \mu^H_N P^H_t - \pi^H \Big\|_{\mathrm{TV}} \ge 1 - \pi^H\big(B_{N,M}(H)^c\big).$$

If $0 < a_{N,M}(H) \le N^{-\tau} < 1$, then Proposition 3.10 implies that

$$\inf_{0 \le t \le T_N} \Big\| \mu^H_N P^H_t - \pi^H \Big\|_{\mathrm{TV}} \ge 1 - \pi^H\big(B_{N,M}(H)^c\big) - \frac{D(\mu^H_N \| \nu_N) + \log 2}{\tau \log N}.$$

Let $\mathcal{E}_{N,M}(H)$ be the last fraction when $a_{N,M}(H) \leq N^{-\tau}$ and $D(\mu_N^H \| \nu_N) < \infty$, and let $\mathcal{E}_{N,M}(H) = 1$ otherwise. Then (4.20) implies that the fraction tends to zero in probability and that the probability that the relative entropy is infinite tends to zero. Therefore $\mathcal{E}_{N,M} \to 0$ in probability, and Lemma 4.4 completes the proof. □

SECTION 5

Statics of the condensed phase

In this section, we prove the static results on which Part II rests. The central object is the first-variation function V_β of the Parisi measure, defined in (5.1) below, whose zero set is the contact set. In Theorem 5.7, we show for every finite mixture and every $\beta > 0$ that an overlap $q \in [0, 1)$ is multisamplable in the sense of Subag (Definition 5.4) if and only if $V_\beta(q) = 0$. At every positive zero, the minimizer of the ground-state problem on the sphere of squared radius qN and the Parisi measure of the residual model are obtained from the Parisi measure by restriction and rescaling. At every other overlap, the generalized TAP functional has an explicit gap. Lemma 5.8 restates this gap as an exponential bound for replicas with mutual overlaps close to q.

In Section 5.3, we specialize to pairs (ξ, β) that are strictly 1RSB with a nondegenerate contact at Q, in the sense of Definition 1.1, which we restate as Definition 5.9. Such pairs are condensed (Lemma 5.10). Proposition 5.11 collects the static consequences of these 1RSB assumptions. The Parisi measure has two atoms 0 and Q, and the residual model at Q is strictly replica symmetric with $\beta^2\psi_Q''(0) < 1$. The ground state of H_Q is strictly one-step replica symmetry breaking at zero temperature, with explicit ground-state energy E_0 and radial derivative K_0, and K_0 lies strictly above the edge $2\sqrt{b_Q}$ of the spectrum of the GOE part of the tangent Hessian. The multisamplable overlaps are 0 and Q. In Section 5.4, we record the covariance of the value and the radial derivative of H_Q at a point, the identities satisfied by E_0 and K_0, and the conditional law of the Hessian. Section 5.5 shows that the 1RSB assumptions persist under small changes of ξ and β and that the free energy is differentiable at every inverse temperature at which they hold. Finally, Section 5.6 verifies the 1RSB assumptions for the classes of Corollary 1.3. These are the covariances for which $\xi''^{-1/2}$ is strictly convex (Proposition 5.17), in particular those in a degree window (Proposition 5.18), the pure models (Proposition 5.20), the small even perturbations of pure even models (Proposition 5.23), and the explicit covariance $r^4 + br^{12}$ of Proposition 5.24, for which $\xi''^{-1/2}$ is not convex. The finite rational computations behind the last example are in Section A.

From Section 3, we use the representation of the field (Lemma 3.2), the Crisanti–Sommers functional, the Parisi formula, and the Crisanti–Sommers optimality criterion (Definition 3.13, Theorem 3.14, and Lemma 3.15), and the thresholds and the replica-symmetric regime (Definition 3.16 and Lemma 3.18). No other result of the paper is used here. The results of this section are used as follows. Section 6 uses Theorem 5.7,

Lemma 5.8, Proposition 5.11, and Corollary 5.16. The radius $\sqrt{Q}$, the values E_0 and K_0, and the Hessian margin of Propositions 5.11 and 5.13 and Lemma 5.14 define the critical points studied in Section 4 and in Sections 7 and 9 to 12. The bound of Lemma 5.22 is used in Section 9, and the strict replica symmetry of the residual model in Section 12. The membership statements of Section 5.6 enter only the proof of Corollary 1.3 in Section 13.

Theorem 5.7 belongs to Subag's generalized TAP theory [143]. That every zero of V_β is multisamplable is [143, Theorem 10, (1.39)], and the restricted and rescaled minimizers at a zero are [143, Proposition 11, (1.43)–(1.44)]. The converse, that every multisamplable overlap is a zero of V_β, is proved in [143, Theorem 10, (1.40)] for generic mixtures. We prove it for every finite mixture by an explicit variation of the ground profile. We derive the whole theorem from the ground-state and generalized TAP formulas recalled below (Theorems 5.2 and 5.5), using the identity (5.9) for both directions. The ground-state formula that we use is due to Chen and Sen [48] and, independently, to Jagannath and Tobasco [96]. The two-atom stationarity equations are the classical one-step equations of Crisanti and Sommers [52]. The bound on the number of contacts under strict convexity of $\xi''^{-1/2}$ is their argument as presented by Talagrand [146, Proposition 2.2], and its zero-temperature version is [90, Lemma 3.21]. For pure models, the static parameters were computed from the TAP approach by Subag [142, Theorems 1 and 2].

Notation. Throughout the section, $\xi(r) = \sum_{p=2}^{P} c_p r^p$ is a finite mixture in the sense of Definition 3.1, with $c_P > 0$, and $\beta > 0$. Only Sections 5.3 to 5.6 assume that the minimum degree is at least four. We use the set $\mathscr{M}$, the functions $\hat{x}_\zeta$, and the functional $\mathcal{P}_\beta(\zeta;\eta)$ of Definition 3.13. For $\zeta \in \mathscr{M}$, we define

$$V_\zeta(r) = \int_0^r \frac{r-t}{\hat{x}_\zeta(t)^2}\,\mathrm{d}t - \beta^2\xi(r), \qquad 0 \le r < 1. \tag{5.1}$$

In the notation of (3.18) with $\eta = \xi$, we have $V_\zeta = -f_\zeta$. The function $\hat{x}_\zeta$ is continuous and positive on $[0,1)$. Hence V_ζ is twice continuously differentiable on $[0,1)$, and

$$V_\zeta(0) = V_\zeta'(0) = 0, \qquad V_\zeta'(r) = \int_0^r \frac{\mathrm{d}t}{\hat{x}_\zeta(t)^2} - \beta^2\xi'(r), \qquad V_\zeta''(r) = \frac{1}{\hat{x}_\zeta(r)^2} - \beta^2\xi''(r). \tag{5.2}$$

We write ζ_β for the Parisi measure of (ξ,β), that is, the minimizer of $\mathcal{P}_\beta(\cdot\,;\xi)$ given by Lemma 3.15, and set $\hat{x}_\beta = \hat{x}_{\zeta_\beta}$ and $V_\beta = V_{\zeta_\beta}$. We call V_ζ the *first-variation function* of ζ; it is the function $-f_\zeta$ of (3.18), with the sign chosen so that $V_\beta \ge 0$ (Lemma 5.6). The zero set of V_β in $[0,1)$ is the *contact set* of (ξ,β), and its elements are the *contact points*. This interpretation agrees with the coincidence set in the obstacle formulation of [97, Theorem 1.1], after absorbing β^2 into the covariance. Recall from Theorem 3.14 that $F(\beta) = \mathcal{P}_\beta(\zeta_\beta;\xi)$, and by Lemma 3.15(3), ζ_β is carried by the set of minimizers of V_β on $[0,1)$; since this set is closed, it contains the support of ζ_β.

For $0 \le q < 1$, the *residual mixture* at q is

$$\psi_q(t) = \xi(q + (1-q)t) - \xi(q) - (1-q)\xi'(q)t = \sum_{k=2}^{P} \Big(\sum_{p=k}^{P} c_p \binom{p}{k} q^{p-k} \Big) (1-q)^k t^k. \quad (5.3)$$

Its coefficients are nonnegative, it has no terms of degree zero or one, and its coefficient of t^P is $c_P(1-q)^P > 0$. This means that ψ_q is a finite mixture. Note that $\psi_0 = \xi$. For $0 < Q < 1$ and $u \in S_N$, we write

$$H_Q(u) = H_N(\sqrt{Q}\, u), \qquad K_Q(u) = \frac{\mathrm{d}}{\mathrm{d}t} H_N(t\sqrt{Q}\, u)\Big|_{t=1} = u \cdot \nabla H_Q(u),$$

where H_N is the polynomial of (3.2) and ∇ is the Euclidean gradient. We also set $a_Q = Q\xi'(Q)$ and $b_Q = Q^2\xi''(Q)$. We call $K_Q(u)$ the *radial derivative* of H_Q at u.

5.1 Ground states and the generalized TAP functional

For $0 < q \le 1$, the field $u \mapsto H_N(\sqrt{q}\, u)$ on S_N is centered Gaussian with covariance $N\xi(qR(u,v))$, and its maximum is the maximum of H_N on the sphere of squared radius qN. The limit of this maximum, divided by N, is given by a zero-temperature analog of the Crisanti–Sommers formula.

Definition 5.1 (Zero-temperature Crisanti–Sommers functional)**.** Let $0 < q \le 1$. An *admissible ground profile* on $[0, q]$ is a function

$$\ell(v) = L - \int_0^v \alpha(w)\, \mathrm{d}w, \qquad 0 \le v \le q,$$

where $\alpha \colon [0, q) \to [0, \infty)$ is nondecreasing, right-continuous, and integrable, and $L > \int_0^q \alpha(w)\, \mathrm{d}w$. The *zero-temperature Crisanti–Sommers functional* is

$$\mathcal{Q}_q(\ell) = \frac{1}{2} \Big(\int_0^q \xi''(v)\ell(v)\, \mathrm{d}v + \int_0^q \frac{\mathrm{d}v}{\ell(v)} \Big). \quad (5.4)$$

Note that an admissible profile is nonincreasing. It follows that $\ell \ge \ell(q) > 0$ on $[0, q]$ and that $\mathcal{Q}_q(\ell)$ is finite. The admissible profiles form a convex set, on which $\mathcal{Q}_q$ is strictly convex, because $z \mapsto 1/z$ is strictly convex on $(0, \infty)$.

Theorem 5.2 (Ground-state energy; Chen–Sen)**.** *Let $0 < q \le 1$. The limit*

$$E(q) = \lim_{N \to \infty} \frac{1}{N} \mathbb{E} \max_{u \in S_N} H_N(\sqrt{q}\, u)$$

exists, and $E(q) = \min_\ell \mathcal{Q}_q(\ell)$, where the minimum is over admissible ground profiles on $[0, q]$ and is attained at exactly one profile.

We also set $E(0) = 0$. Theorem 5.2 is [48, Theorem 1] after a change of variables, as we now explain. Chen and Sen consider a field on S_N with covariance $N\eta(R)$, where $\eta(r) = \sum_{p \geq 2} \gamma_p^2 r^p$, and an external field h [48, (1)]. We take $h = 0$ and $\eta(r) = \xi(qr)$, which is the covariance of $u \mapsto H_N(\sqrt{q}\, u)$; their summability hypothesis $\sum_p 2^p \gamma_p^2 < \infty$ holds because only finitely many γ_p are nonzero, and their formula allows odd degrees. Their theorem [48, (8)–(11)] states that $N^{-1} \max_{S_N}$ of the field converges in L^1 and almost surely to

$$\min \frac{1}{2}\biggl(\eta'(1) L - \int_0^1 \eta''(r) \int_0^r \alpha(w)\, \mathrm{d}w\, \mathrm{d}r + \int_0^1 \frac{\mathrm{d}r}{L - \int_0^r \alpha(w)\, \mathrm{d}w} \biggr),$$

where the minimum is over $L > \int_0^1 \alpha$ and nonnegative, nondecreasing, right-continuous α on $[0,1)$, and is attained at a unique pair. Given such a pair, let $b(r) = L - \int_0^r \alpha(w)\, \mathrm{d}w$ and $\ell(v) = q b(v/q)$. Then ℓ is an admissible ground profile on $[0,q]$, with slope $\alpha(\cdot/q)$, and every admissible profile arises in this way from exactly one pair. Using $\eta'(0) = 0$, $\eta''(r) = q^2 \xi''(qr)$, and the substitution $v = qr$, we find

$$\eta'(1) L - \int_0^1 \eta''(r) \int_0^r \alpha(w)\, \mathrm{d}w\, \mathrm{d}r = \int_0^1 \eta''(r) b(r)\, \mathrm{d}r = \int_0^q \xi''(v) \ell(v)\, \mathrm{d}v,$$
$$\int_0^1 \frac{\mathrm{d}r}{b(r)} = \int_0^q \frac{\mathrm{d}v}{\ell(v)}.$$

This shows that the objective of Chen and Sen at (L, α) equals $\mathcal{Q}_q(\ell)$, and Theorem 5.2 follows.

Definition 5.3 (Generalized TAP functional)**.** For $0 \leq q < 1$, let $F_{\mathrm{res}}(\beta, q) = \mathcal{P}_\beta(\psi_q)$. By Theorem 3.14, this is the limiting free energy of a field with covariance $N\psi_q(R)$ at inverse temperature β. The *generalized TAP functional* is

$$T_\beta(q) = \beta E(q) + \frac{1}{2} \log(1 - q) + F_{\mathrm{res}}(\beta, q). \tag{5.5}$$

Definition 5.4 (Multisamplable overlaps)**.** For $q \in [0,1)$, an integer $k \geq 1$, and $\rho > 0$, let

$$p_{N,k,\rho}(q) = \mathbb{E}\, \pi^{\otimes k}\Bigl(\bigl| R(\sigma^i, \sigma^j) - q \bigr| < \rho \text{ for all } 1 \leq i < j \leq k \Bigr), \tag{5.6}$$

where $\sigma^1, \dots, \sigma^k$ are independent samples from π. The overlap q is *multisamplable* if $\lim_{N \to \infty} N^{-1} \log p_{N,k,\rho}(q) = 0$ for every $k \geq 1$ and every $\rho > 0$.

Note that for $k = 1$ the event in (5.6) is the whole space; only $k \geq 2$ matters.

Theorem 5.5 (Generalized TAP representation; Subag)**.** *Let $0 \leq q < 1$.*

1. *We have $T_\beta(q) \leq F(\beta)$, with equality if and only if q is multisamplable.*

2. *If* $\limsup_{N\to\infty} N^{-1}\log p_{N,k,\rho}(q) = 0$ *for every* $k \geq 1$ *and every* $\rho > 0$*, then* q *is multisamplable.*

Part (1) is [143, Theorem 5, (1.24)], and part (2) is [143, footnote 4 to Definition 2, and Remark 25]. Subag's mixtures $\nu(t) = \sum_p \gamma_p^2 t^p$ have exponentially decaying coefficients, which holds here with $\gamma_p^2 = c_p$. Neither statement assumes genericity, and odd degrees are allowed. In his notation, the sphere of radius $\sqrt{N}$ carries the normalized surface measure, and his limiting free energy is our $F(\beta)$. His ground energy $E_\star(q)$ in [143, (1.16)] is the maximum over the sphere of squared radius Nq divided by N, which is our $E(q)$. His residual free energy $F_\beta(q)$ is defined from the mixture ν_q of [143, (1.22)], which is our ψ_q. We conclude that $F_\beta(q) = F_{\mathrm{res}}(\beta, q)$.

5.2 Multisamplable overlaps and the contact set

We first show that 0 always belongs to the contact set. This is [143, Lemma 37] in our notation; we include the short proof.

Lemma 5.6 (Zero is a contact point)**.** *The minimum of* V_β *on* $[0,1)$ *equals* $V_\beta(0) = 0$*. In particular,* $V_\beta \geq 0$ *on* $[0,1)$*, and* ζ_β *is carried by the contact set. If the minimum degree of* ξ *is at least three, then* 0 *belongs to the support of* ζ_β*.*

Proof. Let a_0 be the smallest point of the support of ζ_β, so that $0 \leq a_0 < 1$, and recall that a_0 minimizes V_β. If $a_0 = 0$, the minimum of V_β is $V_\beta(0) = 0$. Suppose that $a_0 > 0$. Then $\zeta_\beta([0,t]) = 0$ for $t < a_0$, and $\hat{x}_\beta$ is constant on $[0,a_0]$, equal to $\hat{x}_\beta(0)$. Using (5.2), we see that the function $G = -V_\beta'$ is

$$G(r) = \beta^2\xi'(r) - \frac{r}{\hat{x}_\beta(0)^2}, \qquad 0 \leq r \leq a_0.$$

We have $G(0) = 0$ because $\xi'(0) = 0$, and $G(a_0) = 0$ because a_0 is an interior minimum of V_β. Further, $G'' = \beta^2\xi''' \geq 0$ on $[0,a_0]$, since the coefficients of ξ are nonnegative. A convex function that vanishes at both endpoints of an interval is nonpositive on it, so $G \leq 0$ on $[0,a_0]$ and

$$V_\beta(a_0) - V_\beta(0) = -\int_0^{a_0} G(r)\,\mathrm{d}r \geq 0.$$

Since a_0 is a global minimizer, we also have $V_\beta(a_0) \leq V_\beta(0)$. We conclude that the minimum of V_β equals $V_\beta(0) = 0$ in this case too.

If the minimum degree is at least three, then $\xi''' > 0$ on $(0,1)$, and G is strictly convex. In the case $a_0 > 0$, this gives $G < 0$ on $(0,a_0)$ and $V_\beta(a_0) > V_\beta(0)$, which contradicts the minimality of a_0. This contradiction shows that $a_0 = 0$. □

For $q \in (0,1)$, we define $\zeta_{\beta,q}$ as the probability measure on $[0,1)$ with distribution function

$$\zeta_{\beta,q}([0,t]) = \zeta_\beta([0, q + (1-q)t]), \qquad 0 \le t \le 1. \tag{5.7}$$

Equivalently, $\zeta_{\beta,q}$ is the image of ζ_β under $\tau(r) = \max\{0, (r-q)/(1-q)\}$. It satisfies $\zeta_{\beta,q}(\{0\}) = \zeta_\beta([0,q])$; this mass may be zero. Note that $\zeta_{\beta,q} \in \mathscr{M}$.

Theorem 5.7 (Multisamplable overlaps and the contact set)**.** *Let ξ be a finite mixture and $\beta > 0$.*

1. *An overlap $q \in [0,1)$ is multisamplable if and only if $V_\beta(q) = 0$.*
2. *Let $q \in (0,1)$ with $V_\beta(q) = 0$. Then the unique minimizer of $\mathcal{Q}_q$ is $\ell_q(v) = \beta \hat{x}_\beta(v)$, $0 \le v \le q$, so that $E(q) = \mathcal{Q}_q(\ell_q)$. Further, $\zeta_{\beta,q}$ is the Parisi measure of (ψ_q, β).*
3. *Let $q \in (0,1)$ with $V_\beta(q) > 0$, and let*

$$M_q = \frac{1}{\beta^3} \int_0^q \frac{(q-v)^2}{\hat{x}_\beta(v)^3}\, \mathrm{d}v.$$

Then $0 < M_q \le q^3/(3\beta^3 \hat{x}_\beta(q)^3)$, and

$$T_\beta(q) \le F(\beta) - \frac{V_\beta(q)^2}{8\beta^3 M_q}. \tag{5.8}$$

In particular, for every compact set $K \subset (0,1)$ that contains no contact point, $\inf_{q \in K} V_\beta(q)^2/M_q > 0$.

Before the proof, we describe its structure. For every $q \in (0,1)$, the identity (5.9) below writes $F(\beta)$ as the sum of $\beta \mathcal{Q}_q(\ell_q)$, the term $\frac{1}{2}\log(1-q)$, and the Crisanti–Sommers functional of ψ_q evaluated at $\zeta_{\beta,q}$. Since ℓ_q and $\zeta_{\beta,q}$ are trial points of two variational problems, this identity yields an upper bound for $T_\beta(q)$. If $V_\beta(q) > 0$, a variation of ℓ_q in an explicit admissible direction lowers the zero-temperature functional by an amount of order $V_\beta(q)^2$, and this leads to (5.8). If $V_\beta(q) = 0$, we show by a convexity inequality that ℓ_q is optimal and by the optimality criterion that $\zeta_{\beta,q}$ is optimal; then the upper bound is an equality. Theorem 5.5 then identifies the multisamplable overlaps.

Proof. Throughout the proof, we fix $q \in (0,1)$ and write $s = 1-q$, $\zeta = \zeta_\beta$, $\hat{x} = \hat{x}_\beta$, and $V = V_\beta$. We also write $\mathcal{P}_\beta(\cdot\,; \psi_q)$ for the Crisanti–Sommers functional of the residual mixture.

A decomposition at overlap q. We use the variational split behind [143, proof of Lemma 34, (6.8)], evaluated here at trial profiles for every q. The function $\ell_q = \beta\hat{x}$ is an admissible ground profile on $[0,q]$: its slope is $\beta\zeta([0,\cdot])$, which is nonnegative,

nondecreasing, right-continuous, and bounded, and $\ell_q(q) = \beta\hat{x}(q) > 0$. For the residual measure, we use the definition (5.7) and the substitution $w = q + su$, and find that

$$\hat{x}_{\zeta_{\beta,q}}(t) = \int_t^1 \zeta([0, q+su])\,\mathrm{d}u = \frac{1}{s}\int_{q+st}^1 \zeta([0,w])\,\mathrm{d}w = \frac{\hat{x}(q+st)}{s}, \qquad 0 \le t \le 1.$$

Fix $\hat{q} \in (q,1)$ with $\zeta([0,\hat{q}]) = 1$, and let $\hat{q}' = (\hat{q}-q)/s$, so that $\zeta_{\beta,q}([0,\hat{q}']) = 1$. Using the last display and the substitution $v = q + st$, we find

$$\int_0^{\hat{q}'} \frac{\mathrm{d}t}{\hat{x}_{\zeta_{\beta,q}}(t)} + \log(1-\hat{q}') = \int_q^{\hat{q}} \frac{\mathrm{d}v}{\hat{x}(v)} + \log\frac{1-\hat{q}}{s}.$$

For the energy terms, we use (3.17) for ψ_q and ξ. We have $\psi_q'(t) = s(\xi'(q+st) - \xi'(q))$, and the same substitution yields

$$\psi_q(1) - \int \psi_q\,\mathrm{d}\zeta_{\beta,q} = \int_0^1 \psi_q'(t)\,\zeta([0,q+st])\,\mathrm{d}t = \int_q^1 \xi'(v)\zeta([0,v])\,\mathrm{d}v - \xi'(q)\hat{x}(q).$$

Inserting the last two displays into (3.16), we find

$$\mathcal{P}_\beta(\zeta_{\beta,q};\psi_q) = \frac{\beta^2}{2}\Big(\int_q^1 \xi'(v)\zeta([0,v])\,\mathrm{d}v - \xi'(q)\hat{x}(q)\Big) + \frac{1}{2}\int_q^{\hat{q}} \frac{\mathrm{d}v}{\hat{x}(v)} + \frac{1}{2}\log(1-\hat{q}) - \frac{1}{2}\log s.$$

For the zero-temperature functional, we integrate by parts, using $\hat{x}'(v) = -\zeta([0,v])$ for almost every v and $\xi'(0) = 0$, and obtain

$$\beta\mathcal{Q}_q(\ell_q) = \frac{\beta^2}{2}\Big(\xi'(q)\hat{x}(q) + \int_0^q \xi'(v)\zeta([0,v])\,\mathrm{d}v\Big) + \frac{1}{2}\int_0^q \frac{\mathrm{d}v}{\hat{x}(v)}.$$

We add the last two identities and $\frac{1}{2}\log s$. The terms $\xi'(q)\hat{x}(q)$ and $\log s$ cancel, and the integrals combine into integrals over $[0,1]$ and $[0,\hat{q}]$. Then (3.17) for ξ implies that

$$F(\beta) = \mathcal{P}_\beta(\zeta;\xi) = \beta\mathcal{Q}_q(\ell_q) + \frac{1}{2}\log s + \mathcal{P}_\beta(\zeta_{\beta,q};\psi_q). \tag{5.9}$$

This holds for every $q \in (0,1)$, whether or not q is a contact point. By Theorem 5.2, $E(q) = \min \mathcal{Q}_q \le \mathcal{Q}_q(\ell_q)$, and by definition, $F_{\mathrm{res}}(\beta,q) = \min \mathcal{P}_\beta(\,\cdot\,;\psi_q) \le \mathcal{P}_\beta(\zeta_{\beta,q};\psi_q)$. Then (5.9) implies that

$$T_\beta(q) \le F(\beta) - \beta\big(\mathcal{Q}_q(\ell_q) - E(q)\big). \tag{5.10}$$

Away from the contacts. For $\varepsilon \ge 0$, let $\ell_\varepsilon(v) = \ell_q(v) + \varepsilon(q-v)$. Its slope is $\beta\zeta([0,\cdot]) + \varepsilon$, its value at 0 is $\beta\hat{x}(0) + \varepsilon q$, and $\ell_\varepsilon(q) = \ell_q(q) > 0$. Thus ℓ_ε is admissible, and $\ell_\varepsilon \ge \ell_q \ge \beta\hat{x}(q)$ on $[0,q]$. Since ℓ_ε is bounded below by $\beta\hat{x}(q) > 0$, uniformly in ε, we may differentiate under the integrals in (5.4), and we find

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\mathcal{Q}_q(\ell_\varepsilon)\Big|_{\varepsilon=0} &= \frac{1}{2}\int_0^q \Big(\xi''(v) - \frac{1}{\beta^2\hat{x}(v)^2}\Big)(q-v)\,\mathrm{d}v \\ &= -\frac{1}{2\beta^2}\int_0^q V''(v)(q-v)\,\mathrm{d}v = -\frac{V(q)}{2\beta^2}.\end{aligned}$$

In the second step, we used (5.2), and in the last step we integrated by parts and used $V(0) = V'(0) = 0$. Further, for every $\varepsilon \geq 0$,

$$\frac{\mathrm{d}^2}{\mathrm{d}\varepsilon^2}\mathcal{Q}_q(\ell_\varepsilon) = \int_0^q \frac{(q-v)^2}{\ell_\varepsilon(v)^3}\,\mathrm{d}v \leq \int_0^q \frac{(q-v)^2}{\ell_q(v)^3}\,\mathrm{d}v = M_q.$$

Taylor's theorem implies that $\mathcal{Q}_q(\ell_\varepsilon) \leq \mathcal{Q}_q(\ell_q) - \varepsilon V(q)/(2\beta^2) + \varepsilon^2 M_q/2$ for every $\varepsilon \geq 0$. If $V(q) > 0$, we choose $\varepsilon = V(q)/(2\beta^2 M_q)$ and obtain

$$E(q) \leq \mathcal{Q}_q(\ell_\varepsilon) \leq \mathcal{Q}_q(\ell_q) - \frac{V(q)^2}{8\beta^4 M_q}.$$

Inserting this bound into (5.10) gives (5.8). The bounds on M_q follow from $\beta\hat{x}(q) \leq \ell_q \leq \beta$ on $[0,q]$ and $\int_0^q (q-v)^2\,\mathrm{d}v = q^3/3$. On a compact set $K \subset (0,1)$ without contact points, the continuous function V has a positive minimum, while $M_q \leq q^3/(3\beta^3\hat{x}(\max K)^3)$ is bounded, which proves the last assertion of (3).

At a contact: the ground profile. Suppose that $V(q) = 0$. Then q is an interior minimum of $V \geq 0$ (Lemma 5.6), and we also have $V'(q) = 0$. Let $\ell(v) = L - \int_0^v \alpha(w)\,\mathrm{d}w$ be an admissible ground profile on $[0,q]$, and let m be the Borel measure on $[0,q)$ with $m([0,v]) = \alpha(v)$; it has an atom of mass $\alpha(0)$ at 0 and may have infinite total mass. Similarly, the slope $\beta\zeta([0,\cdot])$ of ℓ_q is the distribution function of $m_q = \beta\zeta$ restricted to $[0,q)$. By Fubini's theorem, $\int_0^v \alpha(w)\,\mathrm{d}w = \int_{[0,v]}(v-t)\,m(\mathrm{d}t)$, and therefore

$$\ell(v) = L - \int_{[0,v]}(v-t)\,m(\mathrm{d}t), \qquad \ell_q(v) = \beta\hat{x}(0) - \int_{[0,v]}(v-t)\,m_q(\mathrm{d}t).$$

Using the convexity inequality $1/z \geq 1/z_0 - (z-z_0)/z_0^2$ for $z, z_0 > 0$ and (5.2), we find that

$$\begin{aligned}\mathcal{Q}_q(\ell) - \mathcal{Q}_q(\ell_q) &\geq \frac{1}{2}\int_0^q \Bigl(\xi''(v) - \frac{1}{\ell_q(v)^2}\Bigr)\bigl(\ell(v) - \ell_q(v)\bigr)\,\mathrm{d}v \\ &= -\frac{1}{2\beta^2}\int_0^q V''(v)\bigl(\ell(v) - \ell_q(v)\bigr)\,\mathrm{d}v.\end{aligned} \tag{5.11}$$

We compute the contribution of ℓ. The constant L contributes $-L(V'(q) - V'(0)) = 0$. For $0 \leq t < q$, we integrate by parts and obtain

$$\int_t^q V''(v)(v-t)\,\mathrm{d}v = (q-t)V'(q) - V(q) + V(t) = V(t).$$

To exchange the order of integration below, note that $\int_t^q |V''(v)|\,(v-t)\,\mathrm{d}v \leq \|V''\|_{\infty,[0,q]}\,(q-t)^2/2$, and by Fubini's theorem for nonnegative functions, $\int_{[0,q)}(q-t)^2\,m(\mathrm{d}t) \leq q\int_{[0,q)}(q-t)\,m(\mathrm{d}t) = q\int_0^q \alpha(w)\,\mathrm{d}w < \infty$. Hence Fubini's

theorem applies to $V''(v)(v-t)$ on $\{0 \le t \le v \le q\}$ with respect to $m(\mathrm{d}t)\,\mathrm{d}v$, and we obtain

$$-\int_0^q V''(v)\ell(v)\,\mathrm{d}v = \int_{[0,q)} \Big(\int_t^q V''(v)(v-t)\,\mathrm{d}v\Big) m(\mathrm{d}t) = \int_{[0,q)} V(t)\,m(\mathrm{d}t).$$

For ℓ_q, the same computation yields $\int_{[0,q)} V\,\mathrm{d}m_q$, which is zero, because ζ is carried by the zeros of V (Lemma 5.6). Inserting these two identities into (5.11), we find

$$\mathcal{Q}_q(\ell) - \mathcal{Q}_q(\ell_q) \ge \frac{1}{2\beta^2}\int_{[0,q)} V(t)\,m(\mathrm{d}t) \ge 0. \tag{5.12}$$

We see that ℓ_q minimizes $\mathcal{Q}_q$. If $\mathcal{Q}_q(\ell) = \mathcal{Q}_q(\ell_q)$, then equality holds in (5.11), that is, $\int_0^q (\ell-\ell_q)^2/(\ell\ell_q^2)\,\mathrm{d}v = 0$, and the continuous functions ℓ and ℓ_q coincide. This proves the first assertion of (2), and $E(q) = \mathcal{Q}_q(\ell_q)$ by Theorem 5.2.

At a contact: the residual measure. Still assuming $V(q) = 0$, let $W = V_{\zeta_{\beta,q}}$ be the function (5.1) for the mixture ψ_q and the measure $\zeta_{\beta,q}$, that is,

$$W(t) = \int_0^t \frac{t-u}{\hat{x}_{\zeta_{\beta,q}}(u)^2}\,\mathrm{d}u - \beta^2\psi_q(t), \qquad 0 \le t < 1.$$

Using $\hat{x}_{\zeta_{\beta,q}}(u) = \hat{x}(q+su)/s$ and the substitution $v = q+su$, we find $\int_0^t (t-u)\hat{x}_{\zeta_{\beta,q}}(u)^{-2}\,\mathrm{d}u = \int_q^{q+st}(q+st-v)\hat{x}(v)^{-2}\,\mathrm{d}v$. On the other hand, by Taylor's formula with integral remainder and (5.2), we have

$$V(q+st) - V(q) - stV'(q) = \int_q^{q+st} \frac{q+st-v}{\hat{x}(v)^2}\,\mathrm{d}v - \beta^2\big(\xi(q+st) - \xi(q) - st\xi'(q)\big).$$

We have $V(q) = V'(q) = 0$, and comparing the two displays shows that

$$W(t) = V(q+st), \qquad 0 \le t < 1. \tag{5.13}$$

In particular, $W \ge 0 = W(0)$. Recall that $\zeta_{\beta,q}$ is the image of ζ under $\tau(r) = \max\{0, (r-q)/s\}$. For $r \le q$, we have $W(\tau(r)) = W(0) = 0$. For ζ-almost every $r > q$, we have $W(\tau(r)) = V(r) = 0$, because ζ is carried by the zeros of V. This shows that $\zeta_{\beta,q}$ is carried by the zeros of W, which are its global minimizers, that is, the maximizers of $f_{\zeta_{\beta,q}} = -W$ for the mixture ψ_q. Using Lemma 3.15(3), we find that $\zeta_{\beta,q}$ is the Parisi measure of (ψ_q, β). This completes the proof of (2).

Conclusion. If $q \in (0,1)$ and $V(q) > 0$, then $T_\beta(q) < F(\beta)$ by (3), and q is not multisamplable by Theorem 5.5(1). If $q \in (0,1)$ and $V(q) = 0$, then by (2) both variational problems in (5.9) are solved by the trial points, that is, $E(q) = \mathcal{Q}_q(\ell_q)$ and $F_{\mathrm{res}}(\beta, q) = \mathcal{P}_\beta(\zeta_{\beta,q}; \psi_q)$. Then (5.9) gives $T_\beta(q) = F(\beta)$, and q is multisamplable by Theorem 5.5(1). Finally, $V(0) = 0$, and since $E(0) = 0$ and $\psi_0 = \xi$, we have $T_\beta(0) = \mathcal{P}_\beta(\xi) = F(\beta)$; by Theorem 5.5(1), 0 is multisamplable. This proves (1). □

The next lemma restates the bound for overlaps outside the contact set in the form used in Section 6. The point is that the replica number, the tolerance, and the exponential rate are fixed before $N \to \infty$, and that the bound holds along the full sequence.

Lemma 5.8 (Replicas at an overlap outside the contact set)**.** *Let ξ be a finite mixture, $\beta > 0$, and $q \in (0,1)$ with $V_\beta(q) > 0$. There exist an integer $k \geq 2$ and numbers $\rho \in (0,1)$ and $c > 0$, depending only on ξ, β, and q, such that $p_{N,k,\rho}(q) \leq e^{-cN}$ for all sufficiently large N.*

Proof. By Theorem 5.7(1), q is not multisamplable. Every $\limsup_N N^{-1} \log p_{N,k,\rho}(q)$ is at most zero, because $p_{N,k,\rho}(q) \leq 1$. By Theorem 5.5(2), there exist $k \geq 1$ and $\rho > 0$ for which this lim sup is negative, say smaller than $-c$ with $c > 0$. Then $p_{N,k,\rho}(q) \leq e^{-cN}$ for all sufficiently large N. We have $k \geq 2$, because $p_{N,1,\rho}(q) = 1$. Finally, since $p_{N,k,\rho}(q)$ is nondecreasing in ρ, we may replace ρ by $\min\{\rho, 1/2\}$. □

5.3 Strict one-step replica symmetry breaking

We restate Definition 1.1.

Definition 5.9 (Strict one-step replica symmetry breaking)**.** Let ξ be a finite mixture with minimum degree at least four, and let $\beta > 0$. The pair (ξ, β) is *strictly 1RSB with a nondegenerate contact at Q* if there exist $x, Q \in (0,1)$ such that, with $s = 1 - Q$, $C = s + xQ$,

$$\hat{x}(v) = \begin{cases} C - xv, & 0 \leq v \leq Q, \\ 1 - v, & Q \leq v < 1, \end{cases} \tag{5.14}$$

and

$$V(r) = \int_0^r \frac{r - v}{\hat{x}(v)^2} \, \mathrm{d}v - \beta^2 \xi(r), \qquad 0 \leq r < 1,$$

we have

$$V(Q) = V'(Q) = 0, \qquad V(r) > 0 \quad \text{for } r \in (0,1) \setminus \{Q\}, \qquad V''(Q) > 0. \tag{5.15}$$

We call x and Q the *parameters*, and we refer to the three conditions in (5.15) as the *1RSB assumptions*.

The function $\hat{x}$ is $\hat{x}_\zeta$ for $\zeta = x\delta_0 + (1-x)\delta_Q$, as noted after Definition 3.13, and $V = V_\zeta$. Then V is twice continuously differentiable on $[0,1)$, $V(0) = V'(0) = 0$, and $V''(Q) = s^{-2} - \beta^2 \xi''(Q)$. By Lemma 3.15(3), as in the proof of Proposition 5.11(1) below, the first two conditions in (5.15) imply that ζ is the Parisi measure, so that $V = V_\beta$ and the contact set of (ξ, β) is exactly $\{0, Q\}$. Conversely, if the Parisi measure is $x\delta_0 + (1-x)\delta_Q$ and the contact set is $\{0, Q\}$, then $V = V_\beta \geq 0$ by Lemma 5.6, Q is an interior minimum of V, and the first two conditions hold. In this sense, the pair is strictly 1RSB: the Parisi

measure has two atoms and the contact set is exactly its support. In particular, the parameters are determined by (ξ, β). The third condition says that the contact at Q is nondegenerate. Since $V''(Q) = s^{-2} - \beta^2 \xi''(Q)$, it is equivalent to $\beta^2 s^2 \xi''(Q) < 1$, that is, to the positivity at Q of the replicon eigenvalue $\hat{x}(Q)^{-2} - \beta^2 \xi''(Q)$ of [25, Section 1.3], written with the temperature absorbed into the covariance.

This terminology follows Huang and Sellke [90], who call a model strictly replica symmetric if its contact set is $\{0\}$ [90, Definition 3] (compare Definition 3.16) and, at zero temperature, strictly 1RSB if the zero set of the corresponding function is $\{0, 1\}$ [90, Definition 4]; Definition 5.9 is the analogous notion at positive temperature, with the nondegeneracy condition added. Throughout Sections 5.3 and 5.4, when (ξ, β) satisfies the 1RSB assumptions, we write

$$z = \frac{xQ}{s}, \qquad y = \frac{Q}{\beta s}, \qquad c_0 = \beta x. \tag{5.16}$$

In this section, the letter C always denotes the constant $C = s + xQ$ of Definition 5.9, never a generic constant.

Lemma 5.10 (Strictly 1RSB pairs are condensed)**.** *If (ξ, β) is strictly 1RSB with a nondegenerate contact at Q, then $\beta > \beta_{\mathrm{s}}(\xi)$.*

Proof. For $0 \le v < Q$, we have $\hat{x}(v) = 1 - Q + x(Q - v) < 1 - v$, since $x < 1$. The equation $V(Q) = 0$ implies that

$$\beta^2 \xi(Q) = \int_0^Q \frac{Q - v}{\hat{x}(v)^2} \, \mathrm{d}v > \int_0^Q \frac{Q - v}{(1 - v)^2} \, \mathrm{d}v = -Q - \log(1 - Q),$$

where the last integral was computed in the proof of Lemma 3.18. Dividing by $\xi(Q) > 0$ and comparing with the definition of β_{s} in Definition 3.16, we obtain $\beta^2 > \beta_{\mathrm{s}}^2$. □

Proposition 5.11 (Static consequences of strict 1RSB)**.** *Let (ξ, β) be strictly 1RSB with parameters x, Q and a nondegenerate contact at Q, and let s, C, $\hat{x}$, V, z, y, and c_0 be as in Definition 5.9 and* (5.16)*.*

1. (Parisi measure.) *The Parisi measure of (ξ, β) is $\zeta_\beta = x\delta_0 + (1 - x)\delta_Q$, so that $\hat{x}_\beta = \hat{x}$ and $V_\beta = V$. The parameters satisfy*

$$\beta^2 \xi'(Q) = \frac{Q}{sC}, \qquad \beta^2 x^2 \xi(Q) = \log \frac{C}{s} - \frac{xQ}{C}, \tag{5.17}$$

and

$$F(\beta) = \frac{1}{2}\Big(\beta^2\big(\xi(1) - (1 - x)\xi(Q)\big) + \frac{1}{x}\log\frac{C}{s} + \log s\Big). \tag{5.18}$$

2. (Residual model.) *We have $\beta^2 s^2 \xi''(Q) < 1$ and*

$$\beta^2 \psi_Q(t) + t + \log(1 - t) = -V(Q + st) < 0, \qquad 0 < t < 1. \tag{5.19}$$

There exists $\kappa_{\mathrm{res}} = \kappa_{\mathrm{res}}(\xi,\beta) > 0$ *such that* $V(Q+st) \ge \kappa_{\mathrm{res}} t^2$ *for* $0 < t < 1$. *In particular,* (ψ_Q,β) *is strictly replica symmetric with* $\beta^2\psi_Q''(0) < 1$, *its Parisi measure is* δ_0, *and* $F_{\mathrm{res}}(\beta,Q) = \beta^2\psi_Q(1)/2$.

3. (Ground state at radius $\sqrt{Q}$.) *The unique minimizer of* $\mathcal{Q}_Q$ *is* $\ell(v) = \beta\hat{x}(v) = \beta(C - xv)$, $0 \le v \le Q$, *and*

$$E(Q) = E_0, \qquad \text{where} \quad E_0 = \frac{a_Q + z\xi(Q)}{y}. \tag{5.20}$$

The function $g(r) = \beta^{-2}V(Qr)$ *on* $[0,1]$ *satisfies* $g(0) = g(1) = 0$ *and* $g(r) > 0$ *for* $0 < r < 1$, *and*

$$g(r) = \xi(Q) - \xi(Qr) - a_Q\Big(\frac{1+z}{z^2}\log\big(1+(1-r)z\big) - \frac{1-r}{z}\Big), \qquad 0 \le r \le 1. \tag{5.21}$$

4. (Energy and radial derivative.) *Let* $K_0 = y + b_Q/y$. *Then*

$$y^2 = (1+z)a_Q, \qquad \frac{\xi(Q)}{a_Q} = \frac{(1+z)\log(1+z) - z}{z^2}, \qquad c_0 = \frac{z}{y} < \beta, \tag{5.22}$$

and

$$\frac{b_Q}{y^2} = \beta^2 s^2 \xi''(Q) < 1, \qquad K_0 - 2\sqrt{b_Q} = \frac{(y - \sqrt{b_Q})^2}{y} > 0. \tag{5.23}$$

5. (TAP identity.) *We have*

$$F(\beta) = \beta E_0 + \frac{1}{2}\log s + \frac{\beta^2}{2}\psi_Q(1), \tag{5.24}$$

the set of multisamplable overlaps in $[0,1)$ *is* $\{0,Q\}$, *and* (5.8) *holds for every* $q \in (0,1) \setminus \{Q\}$, *with* $\hat{x}_\beta = \hat{x}$ *in the definition of* M_q.

In the language of [90, Definition 4], part (3) says that the field H_Q, whose covariance is $N\xi(QR)$, is strictly 1RSB at zero temperature: the function g is the function of [90, (2.3)] for the covariance $r \mapsto \xi(Qr)$ and the normalized profile $\ell(Qr)/Q$, and its zero set is $\{0,1\}$. We compare the parameters with those of [90] after the proof.

Proof. We prove the five parts in order, except that part (3) uses the first two identities of (5.22), which are proved in part (4) from part (1) alone.

Part (1). Let $\zeta = x\delta_0 + (1-x)\delta_Q$. Then $\hat{x}_\zeta = \hat{x}$, $V_\zeta = V$, and $f_\zeta = -V$ in (3.18). By (5.15) and $V(0) = 0$, the set of maximizers of f_ζ is $\{0,Q\}$, which carries ζ. We conclude from Lemma 3.15(3) that ζ is the Parisi measure. Direct integration gives

$$\int_0^Q \frac{\mathrm{d}v}{(C-xv)^2} = \frac{Q}{sC}, \qquad \int_0^Q \frac{Q-v}{(C-xv)^2}\,\mathrm{d}v = \frac{1}{x^2}\Big(\log\frac{C}{s} - \frac{xQ}{C}\Big),$$

where we used $C - xQ = s$. Using (5.2), we see that the equations $V'(Q) = 0$ and $V(Q) = 0$ are the two equations in (5.17). For (5.18), we evaluate (3.16) at ζ with $\hat{q} = Q$, using $\int \xi\,\mathrm{d}\zeta = (1-x)\xi(Q)$ and $\int_0^Q \mathrm{d}v/(C-xv) = x^{-1}\log(C/s)$, and recall that $F(\beta) = \mathcal{P}_\beta(\zeta;\xi)$.

Part (2). On $[Q,1)$, we have $\hat{x}(v) = 1 - v$ and $V''(v) = (1-v)^{-2} - \beta^2\xi''(v)$. By Taylor's formula and $V(Q) = V'(Q) = 0$, for $0 \le t < 1$,

$$\begin{aligned} V(Q+st) &= \int_Q^{Q+st} \frac{Q+st-v}{(1-v)^2}\,\mathrm{d}v - \beta^2\psi_Q(t) \\ &= \int_0^t \frac{t-w}{(1-w)^2}\,\mathrm{d}w - \beta^2\psi_Q(t) = -t - \log(1-t) - \beta^2\psi_Q(t), \end{aligned}$$

where we substituted $v = Q + sw$ in the second step and used the computation in the proof of Lemma 3.18 in the third. With (5.15), this proves (5.19). Further, since $V''(Q) = s^{-2} - \beta^2\xi''(Q)$, the condition $V''(Q) > 0$ is equivalent to $\beta^2 s^2 \xi''(Q) < 1$. Let $\kappa_{\mathrm{res}} = \inf_{0<t<1} V(Q+st)/t^2$. As $t \to 0$, the ratio tends to $s^2V''(Q)/2 > 0$. As $t \to 1$, the ratio tends to infinity, because $-\log(1-t) \to \infty$ while t and $\psi_Q(t)$ stay bounded. Since the ratio is continuous and positive on $(0,1)$, we conclude that $\kappa_{\mathrm{res}} > 0$. Using $\psi_Q''(0) = s^2\xi''(Q)$, we see that the residual pair (ψ_Q, β) is strictly replica symmetric with $\beta^2\psi_Q''(0) < 1$, in the sense of Definition 3.16. Then Lemma 3.18(1) shows that its Parisi measure is δ_0 and that $F_{\mathrm{res}}(\beta, Q) = \mathcal{P}_\beta(\psi_Q) = \beta^2\psi_Q(1)/2$.

Part (3). By (1), $V_\beta(Q) = V(Q) = 0$, and Theorem 5.7(2) implies that $\ell = \beta\hat{x}_\beta = \beta\hat{x}$ is the unique minimizer of $\mathcal{Q}_Q$ and that $E(Q) = \mathcal{Q}_Q(\ell)$. We compute

$$\begin{aligned} \int_0^Q \xi''(v)\beta(C - xv)\,\mathrm{d}v &= \beta\big((C-xQ)\xi'(Q) + x\xi(Q)\big) = \beta\big(s\xi'(Q) + x\xi(Q)\big), \\ \int_0^Q \frac{\mathrm{d}v}{\beta(C-xv)} &= \frac{1}{\beta x}\log\frac{C}{s}. \end{aligned}$$

Using $C/s = 1+z$, $\beta s\xi'(Q) = a_Q/y$, $\beta x = z/y$, and $1/(\beta x) = y/z$, we find

$$2\mathcal{Q}_Q(\ell) = \frac{a_Q + z\xi(Q)}{y} + \frac{y}{z}\log(1+z).$$

By the first two identities in (5.22), we have $(y/z)\log(1+z) = (1+z)a_Q\log(1+z)/(yz) = (a_Q + z\xi(Q))/y$. This proves (5.20).

For g, we have $g(0) = \beta^{-2}V(0) = 0$ and $g(1) = \beta^{-2}V(Q) = 0$, and $g > 0$ on $(0,1)$ by (5.15). To prove (5.21), let $\tilde{g}$ be its right side. Then $\tilde{g}(1) = 0 = g(1)$, and it suffices to show that $g' = \tilde{g}'$ on $[0,1]$. By (5.2),

$$g'(r) = \frac{Q}{\beta^2}V'(Qr) = \frac{Q}{\beta^2}\int_0^{Qr}\frac{\mathrm{d}v}{(C-xv)^2} - Q\xi'(Qr) = \frac{Q^2 r}{\beta^2 C(C-xQr)} - Q\xi'(Qr).$$

Since $C = s(1+z)$, $C - xQr = s(1+(1-r)z)$, and $Q^2/(\beta^2 s^2) = y^2 = (1+z)a_Q$, the first term equals $a_Q r/(1+(1-r)z)$. On the other hand,

$$\tilde{g}'(r) = -Q\xi'(Qr) - a_Q\Big(-\frac{1+z}{z(1+(1-r)z)} + \frac{1}{z}\Big) = -Q\xi'(Qr) + \frac{a_Q r}{1+(1-r)z},$$

and we conclude that $g' = \tilde{g}'$.

Part (4). Recall that $C = s(1+z)$. The first equation in (5.17) reads $\beta^2\xi'(Q) = Q/(s^2(1+z))$. Multiplying by Q/β^2 gives $a_Q = y^2/(1+z)$. Further, $\log(C/s) = \log(1+z)$, $xQ/C = z/(1+z)$, and $\beta^2 x^2 = z^2/y^2 = z^2/((1+z)a_Q)$, and the second equation in (5.17) reads

$$\frac{z^2\xi(Q)}{(1+z)a_Q} = \log(1+z) - \frac{z}{1+z},$$

which is the second identity in (5.22). Next, $z/y = (xQ/s)(\beta s/Q) = \beta x = c_0$, and $c_0 < \beta$ because $x < 1$. Further, $b_Q/y^2 = Q^2\xi''(Q)\beta^2 s^2/Q^2 = \beta^2 s^2 \xi''(Q)$, which is less than one by (2). Finally, $K_0 - 2\sqrt{b_Q} = (y^2 - 2y\sqrt{b_Q} + b_Q)/y$, which is positive because $y > \sqrt{b_Q}$.

Part (5). By (1), the contact set is the zero set of V, which is $\{0, Q\}$ by (5.15). Using Theorem 5.7(1), we find that the multisamplable overlaps are 0 and Q, and Theorem 5.7(3) implies (5.8) for every other $q \in (0,1)$. We have $T_\beta(Q) = F(\beta)$ by Theorem 5.5(1). Inserting $E(Q) = E_0$ from (3) and $F_{\mathrm{res}}(\beta, Q) = \beta^2\psi_Q(1)/2$ from (2) into (5.5) yields (5.24). □

Remark 5.12 (Comparison with Huang and Sellke). For the covariance $\eta(r) = \xi(Qr)$, we have $\eta(1) = \xi(Q)$, $\eta'(1) = a_Q$, and $\eta''(1) = b_Q$. The normalized ground profile is $\ell(Qr)/Q = (1+z-zr)/y$, which has initial value $(1+z)/y$ and constant slope $c_0 = z/y$. Recall from (5.22) that z solves $((1+z)\log(1+z) - z)/z^2 = \eta(1)/\eta'(1)$ and $y = \sqrt{(1+z)\eta'(1)}$. These are the parameters z, y, $L = (1+z)/y$, and $u = z/y$ of [90, Lemma 3.5], whose energy $(\eta'(1) + z\eta(1))/y$ is our E_0. The zero-temperature contact inequality [90, (3.4)], with equality exactly at 0 and 1, is the positivity of (5.21) on $(0,1)$. The radial value $R_0 = y + \eta''(1)/y$ in [90, Section 3.2, p. 15, display preceding (3.6)] is our K_0, because their radial derivative of H_Q at u is $K_Q(u)/\sqrt{N}$ and they normalize it by $\sqrt{N}$. Finally, [90, Lemma 3.6] gives $y^2 \geq \eta''(1)$ for every strictly 1RSB covariance; the strict inequality $b_Q < y^2$ in (5.23) comes from $V''(Q) > 0$. In [90], strict 1RSB at zero temperature is an assumption; here it is a consequence of the 1RSB assumptions at positive temperature.

5.4 Value, radial derivative, and Hessian at radius $\sqrt{Q}$

In this section, (ξ, β) is strictly 1RSB with parameters x, Q and a nondegenerate contact at Q, and E_0, K_0, c_0, y, z are given by Proposition 5.11 and (5.16). Let

$$\Sigma_Q = \begin{pmatrix} \xi(Q) & a_Q \\ a_Q & a_Q + b_Q \end{pmatrix}. \tag{5.25}$$

For $t > 2$, let

$$\Omega(t) = \int_{-2}^{2} \log(t-\lambda)\,\rho_{\mathrm{sc}}(\,\mathrm{d}\lambda), \qquad \rho_{\mathrm{sc}}(\,\mathrm{d}\lambda) = \frac{\sqrt{4-\lambda^2}}{2\pi}\,\mathrm{d}\lambda,$$

and $\Omega(2) = \lim_{t \downarrow 2} \Omega(t)$. For $n \geq 1$, a GOE matrix G_n is a real symmetric $n \times n$ random matrix whose entries on and above the diagonal are independent centered Gaussian variables, with variance $1/n$ off the diagonal and $2/n$ on the diagonal. Its empirical spectral distribution converges to ρ_{sc}, whose support is $[-2, 2]$.

Proposition 5.13 (Value and radial derivative)**.** *Let (ξ, β) be strictly 1RSB with a nondegenerate contact at Q.*

1. *For every $u \in S_N$, the vector $(H_Q(u), K_Q(u))$ is centered Gaussian with covariance $N\Sigma_Q$. With $w_p = c_p Q^p / \xi(Q)$, we have*

$$\det \Sigma_Q = \xi(Q)^2 \Big(\sum_p p^2 w_p - \Big(\sum_p p w_p \Big)^2 \Big),$$

   which is $\xi(Q)^2$ times the variance of p under the probability weights (w_p). In particular, Σ_Q is invertible if and only if ξ is not pure.
2. *We have*

$$\Sigma_Q \begin{pmatrix} c_0 \\ 1/y \end{pmatrix} = \begin{pmatrix} E_0 \\ K_0 \end{pmatrix}. \tag{5.26}$$

   If ξ is not pure, then $\Sigma_Q^{-1}(E_0, K_0)^{\mathsf{T}} = (c_0, 1/y)^{\mathsf{T}}$. If $\xi(r) = c_p r^p$ is pure, then $K_Q = pH_Q$ identically, $K_0 = pE_0$, and $E_0/\xi(Q) = c_0 + p/y$.
3. *Suppose that ξ is not pure, and define, for $e \in \mathbb{R}$ and $k \geq 2\sqrt{b_Q}$,*

$$\Theta(e, k) = \frac{1}{2} + \frac{1}{2} \log \frac{b_Q}{a_Q} - \frac{1}{2} (e, k) \Sigma_Q^{-1} (e, k)^{\mathsf{T}} + \Omega\big(k/\sqrt{b_Q}\big). \tag{5.27}$$

   Then

$$\Theta(E_0, K_0) = 0, \qquad \partial_e \Theta(E_0, K_0) = -c_0, \qquad \partial_k \Theta(E_0, K_0) = 0,$$

   and K_0 is the unique maximizer of $k \mapsto \Theta(E_0, k)$ on $[2\sqrt{b_Q}, \infty)$.
4. *Suppose that $\xi(r) = c_p r^p$ is pure, and define, for $e \geq 2\sqrt{b_Q}/p$,*

$$\Theta_{\mathrm{pure}}(e) = \frac{1}{2} + \frac{1}{2} \log \frac{b_Q}{a_Q} - \frac{e^2}{2\xi(Q)} + \Omega\big(pe/\sqrt{b_Q}\big).$$

   Then $E_0 > 2\sqrt{b_Q}/p$, $\Theta_{\mathrm{pure}}(E_0) = 0$, and $\Theta'_{\mathrm{pure}}(E_0) = -c_0$.

In Section 10, Θ appears as the exponential growth rate of the Kac–Rice density of critical points of H_Q with normalized value e and normalized radial derivative k; for mixed spherical models this density was computed by Ben Arous, Subag, and Zeitouni [30, Theorem 5, (3.4), and Section 4.1, (4.2)]. Part (3) says that (E_0, K_0) is a point of zero complexity, at which the complexity decreases in the energy with slope c_0 and is stationary in the radial derivative. We only use the explicit formula here.

Proof. For (1), let $u \in S_N$. By Lemma 3.2(1), $\mathbb{E}H_N(t\sqrt{Q}\,u)H_N(t'\sqrt{Q}\,u) = N\xi(tt'Q)$ for $t, t' > 0$, since $|u|^2 = N$. Differentiating once in t', and once in each of t and t', at $t = t' = 1$, we find

$$\mathbb{E}H_Q(u)K_Q(u) = NQ\xi'(Q) = Na_Q, \qquad \mathbb{E}K_Q(u)^2 = N\big(Q\xi'(Q) + Q^2\xi''(Q)\big).$$

Differentiation under the expectation is justified because H_N is a polynomial with Gaussian coefficients. With $w_p = c_pQ^p/\xi(Q)$, we have $\xi(Q) = \xi(Q)\sum_p w_p$, $a_Q = \xi(Q)\sum_p pw_p$, and $a_Q + b_Q = \sum_p (p + p(p-1))c_pQ^p = \xi(Q)\sum_p p^2 w_p$, which implies the formula for $\det \Sigma_Q$. The variance of p under (w_p) vanishes if and only if exactly one w_p is positive.

For (2), we use $c_0 = z/y$ and $y^2 = (1+z)a_Q$ from (5.22), and multiply:

$$\Sigma_Q \begin{pmatrix} z/y \\ 1/y \end{pmatrix} = \frac{1}{y}\begin{pmatrix} z\xi(Q) + a_Q \\ za_Q + a_Q + b_Q \end{pmatrix} = \begin{pmatrix} (a_Q + z\xi(Q))/y \\ (y^2 + b_Q)/y \end{pmatrix} = \begin{pmatrix} E_0 \\ K_0 \end{pmatrix}.$$

If Σ_Q is invertible, this is the inverse statement. If $\xi = c_p r^p$, then by Euler's identity for homogeneous polynomials, we have $K_Q(u) = u \cdot \nabla H_Q(u) = pH_Q(u)$. In this case $a_Q = p\xi(Q)$ and $b_Q = p(p-1)\xi(Q)$, so that $\Sigma_Q = \xi(Q)\begin{pmatrix} 1 & p \\ p & p^2 \end{pmatrix}$, and (5.26) reads $E_0 = \xi(Q)(c_0 + p/y)$ and $K_0 = p\xi(Q)(c_0 + p/y) = pE_0$.

For (3) and (4), we first compute Ω to the right of 2. For $t > 2$, we write $t = \omega + 1/\omega$ with $\omega > 1$. Then

$$\Omega'(t) = \int \frac{\rho_{\mathrm{sc}}(\,\mathrm{d}\lambda)}{t - \lambda} = \frac{t - \sqrt{t^2-4}}{2} = \frac{1}{\omega}, \qquad \Omega(t) = \log\omega + \frac{1}{2\omega^2}. \tag{5.28}$$

The first identity is the classical Stieltjes transform of the semicircle law; the resulting logarithmic potential is also [30, (3.3)]; it follows from the substitution $\lambda = 2\cos\theta$ and the elementary integral $\int_0^\pi (t - 2\cos\theta)^{-1}\,\mathrm{d}\theta = \pi(t^2-4)^{-1/2}$. For the second, note that $\frac{\mathrm{d}}{\mathrm{d}t}(\log\omega + \frac{1}{2\omega^2}) = (\frac{1}{\omega} - \frac{1}{\omega^3})\frac{\mathrm{d}\omega}{\mathrm{d}t} = \frac{1}{\omega}$, since $\frac{\mathrm{d}t}{\mathrm{d}\omega} = 1 - \omega^{-2}$, and that both sides of the second identity in (5.28) equal $\log t + o(1)$ as $t \to \infty$. Further, $\Omega''(t) = -\int (t-\lambda)^{-2}\rho_{\mathrm{sc}}(\,\mathrm{d}\lambda) < 0$ for $t > 2$, and Ω is continuous on $[2, \infty)$.

We have $K_0/\sqrt{b_Q} = \omega + 1/\omega$ with $\omega = y/\sqrt{b_Q} > 1$, by (5.23). Using (5.28), we obtain

$$\frac{1}{2}\log b_Q + \Omega\big(K_0/\sqrt{b_Q}\big) = \log y + \frac{b_Q}{2y^2}, \qquad \frac{1}{\sqrt{b_Q}}\Omega'\big(K_0/\sqrt{b_Q}\big) = \frac{1}{y}. \tag{5.29}$$

By (5.26), the quadratic term at (E_0, K_0) is $(E_0, K_0)\Sigma_Q^{-1}(E_0, K_0)^{\mathsf{T}} = c_0E_0 + K_0/y$, and

$$c_0E_0 + \frac{K_0}{y} = \frac{za_Q + z^2\xi(Q)}{y^2} + 1 + \frac{b_Q}{y^2}.$$

When we insert this and (5.29) into (5.27), the constant terms and the terms $b_Q/(2y^2)$ cancel, which gives

$$\Theta(E_0, K_0) = \frac{1}{2}\log\frac{y^2}{a_Q} - \frac{za_Q + z^2\xi(Q)}{2y^2} = \frac{1}{2}\log(1+z) - \frac{z}{2(1+z)}\Big(1 + \frac{z\xi(Q)}{a_Q}\Big) = 0.$$

In the second step, we used $y^2 = (1+z)a_Q$, and in the last step the second identity in (5.22), which gives $1 + z\xi(Q)/a_Q = (1+z)\log(1+z)/z$. The gradient of the positive quadratic cost $\frac{1}{2}(e,k)\Sigma_Q^{-1}(e,k)^{\mathsf{T}}$ at (E_0, K_0) is $\Sigma_Q^{-1}(E_0,K_0)^{\mathsf{T}} = (c_0, 1/y)^{\mathsf{T}}$, and by (5.29) we find $\partial_e\Theta(E_0,K_0) = -c_0$ and $\partial_k\Theta(E_0,K_0) = -1/y + 1/y = 0$. The matrix Σ_Q^{-1} is positive definite and $\Omega'' < 0$ to the right of 2, so the function $k \mapsto \Theta(E_0,k)$ is strictly concave on $(2\sqrt{b_Q},\infty)$ and continuous on $[2\sqrt{b_Q},\infty)$. It follows that its stationary point K_0 is its unique maximizer. This proves (3).

For (4), we have $pE_0 = K_0 > 2\sqrt{b_Q}$ by (2) and (5.23). By (2), $E_0^2/\xi(Q) = E_0(c_0 + p/y) = c_0E_0 + K_0/y$, and the computation of $\Theta(E_0,K_0)$ above applies without change. It shows that $\Theta_{\mathrm{pure}}(E_0) = 0$. Finally, by (5.29),

$$\Theta'_{\mathrm{pure}}(E_0) = -\frac{E_0}{\xi(Q)} + \frac{p}{\sqrt{b_Q}}\Omega'\big(K_0/\sqrt{b_Q}\big) = -\frac{E_0}{\xi(Q)} + \frac{p}{y} = -c_0. \qquad \square$$

The next lemma gives the law of the Riemannian Hessian of H_Q at a fixed point, given the value, the radial derivative, and a vanishing tangent gradient. It is the classical computation of Auffinger, Ben Arous, and Černý [12] and Auffinger and Ben Arous [8]; in exactly this form it is [90, (2.11) and Lemma 2.14] for the covariance $r \mapsto \xi(Qr)$. The same computation appears in Lemma 7.4, where the conditioning is used in the Kac–Rice formula. We include the short proof, with the exact factor $(N-1)/N$.

Lemma 5.14 (Conditional law of the Hessian)**.** *Let ξ be a finite mixture, $0 < Q < 1$, $N \geq 2$, and $u \in S_N$. Let $e_2, \dots, e_N$ be an orthonormal basis of the tangent space $\{w : w \cdot u = 0\}$, and identify the Riemannian Hessian $\nabla^2_{S_N} H_Q(u)$ with the matrix $(\nabla^2_{S_N} H_Q(u)[e_i, e_j])_{i,j=2}^N$.*

1. *The matrix $\nabla^2_{S_N} H_Q(u) + N^{-1}K_Q(u)I_{N-1}$ is independent of $(H_Q(u), K_Q(u), \nabla_{S_N} H_Q(u))$ and has the law of $\sqrt{(N-1)b_Q/N}\, G_{N-1}$. Consequently, conditionally on $H_Q(u) = Ne$, $K_Q(u) = Nk$, and $\nabla_{S_N} H_Q(u) = 0$, the Hessian has the law*

$$\nabla^2_{S_N} H_Q(u) \overset{\text{law}}{=} \sqrt{\frac{(N-1)b_Q}{N}}\, G_{N-1} - kI_{N-1}. \tag{5.30}$$

   When ξ is pure of degree p, the conditioning on $K_Q(u)$ is redundant and only values with $k = pe$ occur.
2. *If (ξ, β) is strictly 1RSB with a nondegenerate contact at Q, let $g_0 = K_0 - 2\sqrt{b_Q} = (y - \sqrt{b_Q})^2/y > 0$. Then the interval $U_0 = (K_0 - g_0/4, K_0 + g_0/4)$ satisfies $\inf U_0 = 2\sqrt{b_Q} + 3g_0/4$.*

By (1), for each fixed $k \in U_0$, the conditional empirical spectral distribution of the Hessian converges to the semicircle law scaled by $\sqrt{b_Q}$ and translated by $-k$. Its limiting support is $[-k - 2\sqrt{b_Q}, -k + 2\sqrt{b_Q}]$, whose upper endpoint is at most $-3g_0/4$. Localization of every eigenvalue at finite N requires a fixed enlargement of this interval; the

corresponding edge estimates are proved in Section 7. Part (2) itself is a deterministic statement about the location of U_0. The law in (1) is restated, for the version of the conditional law fixed in Definition 7.3, as Corollary 7.6, which is the form used in the later sections.

Proof. For (1), by rotation invariance (Lemma 3.2(2)), we may assume that $u = \sqrt{N}e_1$ and that $e_2, \dots, e_N$ are the remaining standard basis vectors. The field H_Q, extended to $\mathbb{R}^N$, is centered Gaussian with covariance $\mathbb{E}H_Q(v)H_Q(w) = N\eta(v \cdot w/N)$, where $\eta(r) = \xi(Qr)$, so that $\eta'(1) = a_Q$ and $\eta''(1) = b_Q$. Covariances of derivatives of H_Q at u are derivatives of $N\eta(v \cdot w/N)$ in v and w, evaluated at $v = w = u$. For indices $i, j, k, l \geq 2$, we find

$$\mathbb{E}\big[\partial_{ij}H_Q(u)\,\partial_{kl}H_Q(u)\big] = \partial_{v_i}\partial_{v_j}\partial_{w_k}\partial_{w_l}N\eta\Big(\frac{v \cdot w}{N}\Big)\Big|_{v=w=u} = \frac{\eta''(1)}{N}\big(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}\big).$$

Indeed, every other term of this fourth derivative contains a factor w_i, w_j, v_k, or v_l, which vanishes at $v = w = u$. Similarly, the covariances of $\partial_{ij}H_Q(u)$, with $i, j \geq 2$, with $H_Q(u)$, with $\partial_1 H_Q(u)$, and with $\partial_k H_Q(u)$ for $k \geq 2$, are derivatives of the covariance in which every term contains a factor w_i or w_j, and they vanish. Since these variables are jointly Gaussian, the tangent block $(\partial_{ij}H_Q(u))_{i,j\geq 2}$ is independent of $H_Q(u)$, of $K_Q(u) = \sqrt{N}\partial_1 H_Q(u)$, and of $\nabla_{S_N}H_Q(u) = (\partial_k H_Q(u))_{k\geq 2}$. Its entries are independent up to symmetry, with variance b_Q/N off the diagonal and $2b_Q/N$ on it, which is the law of $\sqrt{(N-1)b_Q/N}\,G_{N-1}$. By the formula for the Riemannian Hessian in the proof of Lemma 3.2,

$$\nabla^2_{S_N}H_Q(u)[e_i, e_j] = \partial_{ij}H_Q(u) - N^{-1}\big(u \cdot \nabla H_Q(u)\big)\delta_{ij} = \partial_{ij}H_Q(u) - N^{-1}K_Q(u)\delta_{ij},$$

which proves the first assertion. The conditional law (5.30) follows from this independence. In the pure case, $K_Q = pH_Q$ by Euler's identity.

For (2), the identity for g_0 is (5.23), and $K_0 - g_0/4 = 2\sqrt{b_Q} + 3g_0/4$. □

5.5 Continuation of the 1RSB solution and the derivative of the free energy

In this section, we fix a finite set $\mathcal{D}$ of integers, each at least four, and for $c = (c_p)_{p\in\mathcal{D}} \in \mathbb{R}^{\mathcal{D}}$ we write $\xi_c(r) = \sum_{p\in\mathcal{D}} c_p r^p$. For (c, β) with $c \in [0,\infty)^{\mathcal{D}} \setminus \{0\}$ and $\beta > 0$, we write $V_{c,\beta,x,Q}$ for the function V of Definition 5.9 built from ξ_c, β, and parameters $x, Q \in (0,1)$.

Lemma 5.15 (Continuation of the 1RSB solution)**.** *Let $\bar{c} \in [0,\infty)^{\mathcal{D}} \setminus \{0\}$ and $\bar{\beta} > 0$ be such that $(\xi_{\bar{c}}, \bar{\beta})$ is strictly 1RSB with parameters $\bar{x}, \bar{Q}$ and a nondegenerate contact at $\bar{Q}$. There exist a neighborhood $\mathcal{U}$ of $(\bar{c}, \bar{\beta})$ in $\mathbb{R}^{\mathcal{D}} \times (0,\infty)$ and real-analytic functions $x, Q\colon \mathcal{U} \to (0,1)$ with $x(\bar{c}, \bar{\beta}) = \bar{x}$ and $Q(\bar{c}, \bar{\beta}) = \bar{Q}$, such that for every $(c, \beta) \in \mathcal{U}$ with $c \in [0,\infty)^{\mathcal{D}} \setminus \{0\}$, the pair (ξ_c, β) is strictly 1RSB with parameters $x(c,\beta), Q(c,\beta)$ and a nondegenerate contact at $Q(c,\beta)$.*

Proof. The proof has two steps. We first continue the two contact equations $V(Q) = V'(Q) = 0$ by the implicit function theorem. Because V vanishes at the two contacts 0 and Q, continuity alone does not give positivity of the continued function near them. In the second step, we therefore prove positivity separately near 0, near the moving contact, near 1, and on the remaining compact set.

The contact equations. For $(x, Q) \in (0,1)^2$, $c \in \mathbb{R}^{\mathcal{D}}$, and $\beta > 0$, let $C = 1 - Q + xQ > 0$ and

$$H_1 = V_{c,\beta,x,Q}(Q) = \frac{1}{x^2}\Big(\log\frac{C}{1-Q} - \frac{xQ}{C}\Big) - \beta^2\xi_c(Q),$$
$$H_2 = V'_{c,\beta,x,Q}(Q) = \frac{Q}{(1-Q)C} - \beta^2\xi'_c(Q),$$

where we used the integrals in the proof of Proposition 5.11(1). These are real-analytic functions of (x, Q, c, β). We write $J_j = \int_0^Q (Q-v)^j (C-xv)^{-3}\,\mathrm{d}v$ for $j = 0, 1, 2$. Differentiating the integral expressions $H_1 = \int_0^Q (Q-v)(C-xv)^{-2}\,\mathrm{d}v - \beta^2\xi_c(Q)$ and $H_2 = \int_0^Q (C-xv)^{-2}\,\mathrm{d}v - \beta^2\xi'_c(Q)$, and using $\partial_x(C-xv) = Q-v$ and $\partial_Q(C-xv) = -(1-x)$, we find

$$\begin{pmatrix} \partial_x H_1 & \partial_Q H_1 \\ \partial_x H_2 & \partial_Q H_2 \end{pmatrix} = \begin{pmatrix} -2J_2 & H_2 + 2(1-x)J_1 \\ -2J_1 & V''(Q) + 2(1-x)J_0 \end{pmatrix}, \tag{5.31}$$

where $V''(Q) = (1-Q)^{-2} - \beta^2\xi''_c(Q)$. In $\partial_Q H_1$, the boundary term vanishes because the integrand $(Q-v)(C-xv)^{-2}$ vanishes at $v = Q$; in $\partial_Q H_2$, the boundary term is $(C-xQ)^{-2} = (1-Q)^{-2}$. At $(\bar{x}, \bar{Q}, \bar{c}, \bar{\beta})$, we have $H_2 = 0$, and the determinant of (5.31) is

$$-2J_2 V''(\bar{Q}) - 4(1-\bar{x})\big(J_0 J_2 - J_1^2\big) < 0.$$

Here $J_2 > 0$, $V''(\bar{Q}) > 0$ by (5.15), and $J_0 J_2 > J_1^2$ by the Cauchy–Schwarz inequality, which is strict because $v \mapsto \bar{Q} - v$ is not constant. By the implicit function theorem, there exist a neighborhood $\mathcal{U}_1$ of $(\bar{c}, \bar{\beta})$ and real-analytic functions x, Q on $\mathcal{U}_1$ with values in $(0,1)$, such that $H_1 = H_2 = 0$ at $(x(c,\beta), Q(c,\beta), c, \beta)$.

Positivity. For $(c, \beta) \in \mathcal{U}_1$ with $c \in [0,\infty)^{\mathcal{D}} \setminus \{0\}$, let $\tilde{V} = V_{c,\beta,x(c,\beta),Q(c,\beta)}$, $\tilde{Q} = Q(c,\beta)$, and let V_0 be the function V of $(\xi_{\bar{c}}, \bar{\beta})$. We shrink $\mathcal{U}_1$ to a neighborhood $\mathcal{U}$ with compact closure in $\mathcal{U}_1$, on which $\big|Q(c,\beta) - \bar{Q}\big|$ is small and $\beta^2 \sum_p |c_p| \le M$ for some $M < \infty$. We cover $(0,1)$ by four regions.

Near the origin: since the function $\hat{x}$ of (5.14) satisfies $0 < \hat{x} \le 1$, and $\xi_c(r) \le r^4 \sum_p c_p$ for $0 \le r \le 1$, we have $\tilde{V}(r) \ge r^2/2 - Mr^4 \ge r^2/4$ for $0 < r \le r_0 = (4M)^{-1/2}$.

Near the contact: let $\lambda = V_0''(\bar{Q}) > 0$. The function $V'' = \hat{x}^{-2} - \beta^2\xi''_c$ is continuous in r and in the parameters, on compact subsets of $[0,1)$, because $\hat{x}$ is continuous in (v, x, Q) and positive for $v < 1$. After shrinking $\mathcal{U}$, there exists $\delta > 0$ such that $\tilde{V}'' \ge \lambda/2$ on $I = [\bar{Q} - 2\delta, \bar{Q} + 2\delta]$ and $\big|\tilde{Q} - \bar{Q}\big| \le \delta$. Using $\tilde{V}(\tilde{Q}) = \tilde{V}'(\tilde{Q}) = 0$ and Taylor's formula, we find $\tilde{V}(r) \ge \lambda(r - \tilde{Q})^2/4$ for $r \in I$, which is positive for $r \ne \tilde{Q}$.

Near one: for $r \geq \tilde{Q}$, we have, by the computation in the proof of Proposition 5.11(2),

$$\tilde{V}(r) = -\log\frac{1-r}{1-\tilde{Q}} - \frac{r-\tilde{Q}}{1-\tilde{Q}} - \beta^2\big(\xi_c(r) - \xi_c(\tilde{Q}) - (r-\tilde{Q})\xi_c'(\tilde{Q})\big).$$

All terms except $-\log(1-r)$ are bounded uniformly over $\mathcal{U}$ and $r \in [\tilde{Q}, 1)$, by a constant M_1. Then $\tilde{V}(r) \geq -\log(1-r) - M_1 > 0$ for $1 - e^{-M_1} < r < 1$; we may assume that $1 - e^{-M_1} > \bar{Q} + 2\delta$.

The remaining set $K = [r_0, 1 - e^{-M_1}] \setminus (\bar{Q} - \delta, \bar{Q} + \delta)$ is compact and does not contain 0, $\bar{Q}$, or 1, so V_0 has a positive minimum m_0 on K. The function $(c, \beta, r) \mapsto \tilde{V}(r)$ is continuous on the closure of $\mathcal{U}$ times K. After shrinking $\mathcal{U}$, we have $\tilde{V} \geq m_0/2$ on K.

Since these four regions cover $(0,1)$, we have $\tilde{V} > 0$ on $(0,1) \setminus \{\tilde{Q}\}$. Together with $\tilde{V}(\tilde{Q}) = \tilde{V}'(\tilde{Q}) = 0$ and $\tilde{V}''(\tilde{Q}) \geq \lambda/2$, this proves that (ξ_c, β) satisfies the 1RSB assumptions with parameters $x(c,\beta)$ and $Q(c,\beta)$. □

The proof also shows that the margins $r^2/4$ near the origin, $\lambda(r-\tilde{Q})^2/4$ near the contact, and $m_0/2$ on the compact set hold uniformly on $\mathcal{U}$.

Corollary 5.16 (Derivative of the free energy)**.** *Let (ξ, β_0) be strictly 1RSB with parameters x, Q and a nondegenerate contact at Q. Then F is real-analytic on a neighborhood of β_0, and*

$$F'(\beta_0) = \beta_0\big(\xi(1) - (1-x)\xi(Q)\big). \tag{5.32}$$

Moreover, $F_N'(\beta_0) = N^{-1}\mathbb{E}\,\pi(H_N)$ for every N, and $F_N'(\beta_0) \to F'(\beta_0)$ as $N \to \infty$.

Proof. Let $\mathcal{D}$ be the set of supported degrees of ξ, and apply Lemma 5.15 with c fixed. For β near β_0, the pair (ξ, β) satisfies the 1RSB assumptions with real-analytic parameters $x(\beta), Q(\beta)$, and by (5.18), $F(\beta) = \mathcal{F}(x(\beta), Q(\beta), \beta)$, where

$$\mathcal{F}(x, Q, \beta) = \frac{1}{2}\Big(\beta^2\big(\xi(1) - (1-x)\xi(Q)\big) + \frac{1}{x}\log\frac{1-Q+xQ}{1-Q} + \log(1-Q)\Big).$$

This proves analyticity. By direct differentiation, we have

$$\partial_x\mathcal{F} = -\frac{1}{2}V_{x,Q}(Q), \qquad \partial_Q\mathcal{F} = \frac{1}{2}(1-x)V_{x,Q}'(Q),$$

where $V_{x,Q}$ is the function of Definition 5.9 with parameters x, Q at inverse temperature β; for the first identity, we used the expression for H_1 in the proof of Lemma 5.15, and for the second, the expression for H_2. Both vanish along the branch. Using the chain rule, we find that $F'(\beta_0) = \partial_\beta\mathcal{F}(x, Q, \beta_0)$, which is (5.32).

For each N, the function F_N is convex and differentiable, with $F_N'(\beta) = N^{-1}\mathbb{E}\,\pi(H_N)$, by differentiation under the expectation (the field is a polynomial with Gaussian coefficients, and $\sup_{S_N}|H_N|$ has Gaussian tails). For $h > 0$, convexity implies that

$$\frac{F_N(\beta_0) - F_N(\beta_0 - h)}{h} \leq F_N'(\beta_0) \leq \frac{F_N(\beta_0 + h) - F_N(\beta_0)}{h}.$$

By Theorem 3.14, $F_N \to F$ at every inverse temperature. Letting $N \to \infty$ and then $h \downarrow 0$, and using the differentiability of F at β_0, we obtain $F_N'(\beta_0) \to F'(\beta_0)$. □

5.6 Classes of strictly 1RSB covariances

Proposition 5.17 (Strict convexity of $\xi''^{-1/2}$)**.** *Let ξ be a finite mixture with minimum degree at least four, such that $r \mapsto \xi''(r)^{-1/2}$ is strictly convex on $(0,1]$. Then for every $\beta > \beta_{\mathrm{s}}(\xi)$, the pair (ξ,β) is strictly 1RSB with a nondegenerate contact at Q, where its parameters x, Q are given by $\zeta_\beta = x\delta_0 + (1-x)\delta_Q$.*

The argument that bounds the number of contacts is that of Crisanti and Sommers [52], as presented by Talagrand [146, Proposition 2.2], who treats even mixtures. The optimality criterion of Lemma 3.15 holds for all finite mixtures, and the argument applies to odd degrees as well; we include it with the strictness made explicit. The contact-count argument also appears in the rule of signs of [97, Theorem 1.3(1) and Remark 1.4] when $(\xi''^{-1/2})'' > 0$. For the ground state, the analogous criterion is [90, Lemma 3.21]; we do not use it.

Proof. Fix $\beta > \beta_{\mathrm{s}}(\xi)$, and write $\zeta = \zeta_\beta$, $\hat{x} = \hat{x}_\beta$, and $V = V_\beta$. By Lemma 5.6, $V \geq 0$ on $[0,1)$, the support of ζ consists of contact points, and 0 belongs to the support, since the minimum degree is at least three. Let

$$D(r) = \bigl(\beta^2 \xi''(r)\bigr)^{-1/2} - \hat{x}(r), \qquad 0 < r < 1.$$

The function $\hat{x}$ is concave, because its almost everywhere derivative $-\zeta([0,r])$ is nonincreasing. Then D is strictly convex on $(0,1)$, and it has at most two zeros there. By (5.2), for $0 < r < 1$ we have $V''(r) = 0$ if and only if $D(r) = 0$.

At most two contacts. Let $a < b$ be two contact points. Then $V(a) = V(b) = 0$ and $V'(a) = V'(b) = 0$; for $a = 0$ this is (5.2), and otherwise it holds because $V \geq 0$ has interior minima at a and b. The function V does not vanish identically on $[a,b]$: otherwise $V'' = 0$ on (a,b), and D would vanish on an interval. The maximum of V on $[a,b]$ is therefore positive and is attained at some $c \in (a,b)$, where $V'(c) = 0$. Applying Rolle's theorem to V' on $[a,c]$ and on $[c,b]$, we obtain zeros of V'' in (a,c) and in (c,b), that is, two distinct zeros of D in (a,b). If there were three contacts $a < b < b'$, we would obtain two zeros of D in (a,b) and two in (b,b'), which is impossible. We conclude that there are at most two contact points.

Two atoms. As $\beta > \beta_{\mathrm{s}}(\xi)$, Lemma 3.18(1) implies that $\zeta \neq \delta_0$. The support of ζ contains 0, is not $\{0\}$, and consists of at most two contacts. This implies that $\zeta = x\delta_0 + (1-x)\delta_Q$ with $x, Q \in (0,1)$, and the contact set is $\{0, Q\}$. Then $\hat{x}$ is the function (5.14) with these

parameters, V is the function of Definition 5.9, $V(Q) = V'(Q) = 0$, and $V > 0$ on $(0,1) \setminus \{Q\}$.

Strict curvature at Q. Since Q is an interior minimum of V, we have $V''(Q) \geq 0$. If $V''(Q) = 0$, then $D(Q) = 0$. The argument above, applied to the contacts $0 < Q$, shows that the function D also has two zeros in $(0, Q)$. It would then have three zeros, which is impossible. Therefore $V''(Q) > 0$, and (ξ, β) satisfies (5.15). □

The next proposition gives a criterion for the convexity hypothesis of Proposition 5.17 that involves only the smallest and the largest supported degrees.

Proposition 5.18 (A condition on the smallest and largest degrees)**.** *Let ξ be a finite mixture whose smallest supported degree is $d \geq 4$ and whose largest supported degree is P. If*

$$6(d-2)(P-2) > (d+P-5)^2, \tag{5.33}$$

which is equivalent to $P - d < d - 1 + \sqrt{3d(d-2)}$, then $(\xi''^{-1/2})'' > 0$ on $(0,1]$. In particular, $\xi''^{-1/2}$ is strictly convex on $(0,1]$, and (ξ, β) is strictly 1RSB with a nondegenerate contact for every $\beta > \beta_s(\xi)$. The condition (5.33) *holds whenever all supported degrees of ξ lie in one of the sets $\{4,5,\dots,11\}$, $\{6,7,\dots,18\}$, and $\{8,9,\dots,26\}$.*

The proof applies the Bhatia–Davis bound $\mathrm{Var}(I) \leq (m-n)(M-m)$ for a random variable I with values in $[n, M]$ and mean m [31] to a tilted distribution of the degrees. No restriction on the number of supported degrees, their parity, or the sizes of the coefficients is needed.

Proof. We write $v(r) = \xi''(r) = \sum_i a_i r^i$, where $i = p - 2$ ranges over the shifted supported degrees and $a_i = p(p-1)c_p > 0$. Fix $r \in (0,1]$, and let I be a random variable with $\mathbb{P}(I = i) = a_i r^i / v(r)$, with mean m and variance σ^2. Differentiating term by term, we have

$$\frac{r v'(r)}{v(r)} = m, \qquad \frac{r^2 v''(r)}{v(r)} = \mathbb{E}[I(I-1)] = m^2 + \sigma^2 - m,$$

and therefore

$$\frac{r^2}{v(r)^2}\big(3v'(r)^2 - 2v(r)v''(r)\big) = 3m^2 - 2(m^2 + \sigma^2 - m) = m^2 + 2m - 2\sigma^2. \tag{5.34}$$

Let $n = d - 2 \geq 2$ and $M = P - 2$. Since $n \leq I \leq M$, we have $0 \leq \mathbb{E}[(I-n)(M-I)] = (m-n)(M-m) - \sigma^2$, which is the Bhatia–Davis bound. It follows that the right side of (5.34) is at least

$$f(m) = m^2 + 2m - 2(m-n)(M-m) = 3m^2 - 2(n+M-1)m + 2nM, \qquad n \leq m \leq M.$$

The vertex of this quadratic is at $m_* = (n + M - 1)/3$. If $M \leq 2n + 1$, then $m_* \leq n$, and $f(m) \geq f(n) = n(n+2) > 0$ on $[n, M]$. Otherwise, $m_* \in (n, M)$, and the minimum of f

on $[n, M]$ is $f(m_*) = 2nM - (n + M - 1)^2/3$, which is positive by (5.33). In both cases, $3v'^2 - 2vv'' > 0$ on $(0, 1]$, and

$$\bigl(v^{-1/2}\bigr)'' = \frac{3v'^2 - 2vv''}{4v^{5/2}} > 0 \qquad \text{on } (0, 1].$$

A function with positive second derivative is strictly convex, and Proposition 5.17 implies that the 1RSB assumptions hold for every $\beta > \beta_s(\xi)$.

For the equivalent form, let $k = P - d$. Expanding both sides, we find

$$6(d-2)(P-2) - (d+P-5)^2 = 3d(d-2) - \bigl(k - (d-1)\bigr)^2.$$

Since $d - 1 - \sqrt{3d(d-2)} < 0 \le k$ for $d \ge 4$, the left side is positive if and only if $k < d - 1 + \sqrt{3d(d-2)}$. For the three sets, let $h(d) = d - 1 + \sqrt{3d(d-2)}$, which is increasing in d, and let $[d_1, P_1]$ be the set. If $d_1 \le d \le P \le P_1$, then $P - d \le P_1 - d_1$, and it suffices to check $P_1 - d_1 < h(d_1)$. For $(d_1, P_1) = (4, 11), (6, 18), (8, 26)$, the left side of (5.33) is 108, 384, 864 and the right side is 100, 361, 841. By the equivalence just proved, this means that $P_1 - d_1 < h(d_1)$. □

Remark 5.19 (Optimality among criteria in the extreme degrees). The inequality (5.33) cannot be relaxed to a criterion that guarantees $(\xi''^{-1/2})'' > 0$ and depends only on d and P. Indeed, suppose that $6(d-2)(P-2) < (d+P-5)^2$. Then $f(m_*) < 0$, and $M > 2n + 1$ (otherwise $f \ge n(n+2) > 0$ on $[n, M]$), which means that $n < m_* < M$. The two-point law on $\{n, M\}$ with mean m_* attains the Bhatia–Davis bound, and it is the law of I at $r = 1$ for suitable positive coefficients on the degrees d and P. For these coefficients, the right side of (5.34) at $r = 1$ is $f(m_*) < 0$. Then $(\xi''^{-1/2})'' < 0$ near $r = 1$, and this persists after adding sufficiently small positive coefficients on intermediate degrees. For example, $\xi(r) = \frac{19}{12}r^4 + \frac{5}{132}r^{12}$ has $v(r) = 19r^2 + 5r^{10}$; at $r = 1$ we find $m = 11/3$, $\sigma^2 = 95/9$, and $m^2 + 2m - 2\sigma^2 = -1/3$. When (5.33) holds with equality, which can happen only if $d + P$ is odd, the same computation shows that $(\xi''^{-1/2})'' \ge 0$, with isolated zeros, and $\xi''^{-1/2}$ may still be strictly convex; the strict inequality is only a sufficient condition. Proposition 5.24 below gives a covariance with degrees 4 and 12 that lies outside the window, for which $\xi''^{-1/2}$ is not convex, and which is still strictly 1RSB with a nondegenerate contact at one temperature.

We now turn to pure models, whose static parameters Subag computed rigorously from the generalized TAP representation, without the Parisi formula (see Remark 5.21). For $z > 0$, let

$$v(z) = \frac{(1+z)\log(1+z) - z}{z^2} = \int_0^1 \frac{1-t}{1+zt}\,\mathrm{d}t. \tag{5.35}$$

The second expression follows by computing the integral. Differentiating under the integral, we see that v is strictly decreasing, with $v(0+) = 1/2$ and $v(\infty) = 0$. Under the 1RSB assumptions, (5.22) implies that the parameter z of (5.16) satisfies $v(z) = \xi(Q)/a_Q$.

Proposition 5.20 (Pure models). *Let $p \geq 4$ and $\xi(r) = r^p$. Let $z_p > 0$ be the unique solution of $v(z_p) = 1/p$, and let*

$$q_s = \frac{z_p}{1+z_p}, \qquad e_p = \frac{p+z_p}{\sqrt{p(1+z_p)}}.$$

1. *We have $z_p > 2(p-2)$, q_s is the unique minimizer of $r \mapsto (-r - \log(1-r))/r^p$ on $(0,1)$, and*
$$\beta_{\mathrm{s}}(p)^2 = \frac{1}{p q_s^{p-2}(1-q_s)}.$$
2. *Let $\beta > \beta_{\mathrm{s}}(p)$. Then (ξ, β) is strictly 1RSB with a nondegenerate contact. Its parameter Q is the unique solution in $(q_s, 1)$ of*
$$\beta^2 p(1+z_p)Q^{p-2}(1-Q)^2 = 1, \tag{5.36}$$
and $x = z_p(1-Q)/Q$. Moreover, $V''(Q) = (1-Q)^{-2}\big(1-(p-1)/(1+z_p)\big)$.
3. *We have $E(q) = e_p q^{p/2}$ for $0 \leq q \leq 1$, $E_0 = e_p Q^{p/2}$, $K_0 = pE_0$, and*
$$F(\beta) = \beta e_p Q^{p/2} + \frac{1}{2}\log(1-Q) + \frac{\beta^2}{2}\big(1 - Q^p - pQ^{p-1}(1-Q)\big).$$

Proof. Recall that v is strictly decreasing from $1/2$ to 0. As $1/p \in (0, 1/2)$, the solution z_p exists and is unique.

The bound on z_p. Let $m = p - 2 \geq 2$. Since v is decreasing, $z_p > 2m$ is equivalent to $v(2m) > 1/(m+2)$. Multiplying by $4m^2/(2m+1)$ and rearranging, this is equivalent to $h_0(m) > 0$, where

$$h_0(m) = \log(2m+1) - \frac{2m(3m+2)}{(m+2)(2m+1)}.$$

A direct computation gives $h_0'(m) = 4m^2(m-1)/((m+2)^2(2m+1)^2) > 0$ for $m > 1$. At $m = 2$, since $\log 5 = 2\,\mathrm{atanh}(2/3)$ and the series of atanh has positive terms,

$$\log 5 > 2\Big(\frac{2}{3} + \frac{(2/3)^3}{3} + \frac{(2/3)^5}{5} + \frac{(2/3)^7}{7}\Big) = \frac{8}{5} + \frac{4}{15309},$$

that is, $h_0(2) = \log 5 - 8/5 > 0$. Since h_0 is increasing, we have $h_0(m) > 0$ for $m \geq 2$, and therefore $z_p > 2(p-2)$.

The static threshold. The derivative of $(-r - \log(1-r))/r^p$ has the sign of $D_p(r) = r^2/(1-r) - p(-r - \log(1-r))$. Substituting $r = z/(1+z)$, which is an increasing bijection of $(0,\infty)$ onto $(0,1)$, we find $r^2/(1-r) = z^2/(1+z)$ and $-r - \log(1-r) = z^2 v(z)/(1+z)$, so that

$$D_p\Big(\frac{z}{1+z}\Big) = \frac{z^2}{1+z}\big(1 - pv(z)\big).$$

This is negative for $z < z_p$ and positive for $z > z_p$, which shows that q_s is the unique minimizer. Since $D_p(q_s) = 0$, the minimum equals $q_s^2/(p(1-q_s)q_s^p)$, which is the formula for $\beta_{\mathrm{s}}(p)^2$.

The 1RSB assumptions. The function $\xi''(r)^{-1/2} = (p(p-1))^{-1/2} r^{-(p-2)/2}$ has positive second derivative on $(0,1]$. Proposition 5.17 implies that (ξ, β) satisfies the 1RSB assumptions for every $\beta > \beta_{\mathrm{s}}(p)$. Let x, Q be its parameters. By (5.22) and $\xi(Q)/a_Q = 1/p$, we have $v(z) = 1/p$. This means that $z = xQ/(1-Q) = z_p$, that is, $x = z_p(1-Q)/Q$. As $C = (1-Q)(1+z_p)$, the first equation in (5.17) becomes (5.36). We have $Q > z_p/(1+z_p) = q_s$, because $x < 1$. To see that (5.36) has only one solution in $(q_s, 1)$, note that $r \mapsto r^{p-2}(1-r)^2$ is strictly decreasing on $((p-2)/p, 1)$, and that $q_s > (p-2)/p$, because $z_p > 2(p-2)$ gives $q_s > 2(p-2)/(2p-3) > (p-2)/p$. Further, by (5.36), we have $\beta^2 p(p-1)Q^{p-2} = (p-1)/((1+z_p)(1-Q)^2)$, and $V''(Q) = (1-Q)^{-2} - \beta^2 \xi''(Q)$ has the stated form; it is positive because $z_p > p-2$.

Energies and free energy. For $\xi = r^p$, we have $\xi(Q) = Q^p$, $a_Q = pQ^p$, $b_Q = p(p-1)Q^p$, and $y = \sqrt{(1+z_p)a_Q} = \sqrt{p(1+z_p)}\, Q^{p/2}$. Using (5.20), we find $E_0 = (p+z_p)Q^p/y = e_p Q^{p/2}$, and $K_0 = pE_0$ by Proposition 5.13(2). By homogeneity, $\max_{u \in S_N} H_N(\sqrt{q}\, u) = q^{p/2} \max_{u \in S_N} H_N(u)$ for every $q \in (0,1]$ and every N, which implies that $E(q) = q^{p/2} E(1)$. Taking $q = Q$, we find $E(1) = e_p$, and then $E(q) = e_p q^{p/2}$ for $0 \le q \le 1$. The formula for $F(\beta)$ is (5.24), since $\psi_Q(1) = 1 - Q^p - pQ^{p-1}(1-Q)$. □

Remark 5.21 (Subag's formulas for pure models). Subag obtained the formulas of Proposition 5.20 for the static overlap, the static threshold, the ground-state energy, the largest multisamplable overlap, and the free energy at all temperatures, from the generalized TAP representation and without the Parisi formula [142, Theorems 1 and 2, (1.9)–(1.13)]. The free energy, the ground-state energy, and the 1RSB parameters had been derived earlier in physics, by the replica method and the TAP approach [52, 53, 108]; see [142, Section 1.1]. His critical overlap q_c, defined by [142, (1.9)], is our q_s: multiplying $D_p(q_s) = 0$ by $1 - q_s$ gives $p(1-q_s)\log(1-q_s) + pq_s - (p-1)q_s^2 = 0$. His critical inverse temperature [142, (1.10)] is $q_s^{1-p/2}(p(1-q_s))^{-1/2} = \beta_{\mathrm{s}}(p)$. His ground-state energy [142, (1.11)] equals e_p, and his equation [142, (1.12)] for the largest multisamplable overlap is equivalent to (5.36), where $z_p > p-2$ selects the root; [142, (1.13)] is the formula for $F(\beta)$ in Proposition 5.20(3). Our derivation of these parameters from the 1RSB assumptions, through Propositions 5.11 and 5.17, is independent of his argument. What the later sections need beyond his formulas is the strict 1RSB structure with a nondegenerate contact, in particular the strict inequalities $V > 0$ on $(0,1) \setminus \{Q\}$ and $V''(Q) > 0$.

The next bound is used in Section 9 to control the variance of the residual field at latitudes $a \ge \sqrt{Q}$.

Lemma 5.22 (A bound on the pure residual coefficient)**.** *Let $p \ge 4$, $\xi(r) = r^p$, $\beta > \beta_{\mathrm{s}}(p)$, and let Q be as in Proposition 5.20. For $0 \le a \le 1$, let*

$$C_\beta(a) = \beta^2\big(1 - a^{2p} - pa^{2p-2}(1-a^2)\big).$$

Then $C_\beta(\sqrt{Q}) = \beta^2 \psi_Q(1)$,

$$C_\beta(\sqrt{Q}) < \frac{159}{320} < \frac{1}{2}, \qquad C'_\beta(a) = -2\beta^2 p(p-1) a^{2p-3}(1-a^2) \le 0. \tag{5.37}$$

Proof. The identity $C_\beta(\sqrt{Q}) = \beta^2 \psi_Q(1)$ follows from (5.3) with $\xi = r^p$, and the derivative is a direct computation. Let $m = p-2$, $z = z_p$, and $t = (1-Q)/Q$. By (5.3), $\psi_Q(1) = Q^p \sum_{k=2}^p \binom{p}{k} t^k$, and by (5.36), $\beta^2 = 1/(p(1+z)Q^{p-2}(1-Q)^2)$. We obtain

$$C_\beta(\sqrt{Q}) = \frac{Q^2}{p(1+z)(1-Q)^2} \sum_{k=2}^{p} \binom{p}{k} t^k = \frac{1}{p(1+z)} \sum_{k=2}^{p} \binom{p}{k} t^{k-2}.$$

For $2 \le k \le p$, we have $\binom{p}{k} = \frac{p(p-1)}{k(k-1)} \binom{p-2}{k-2} \le \frac{p(p-1)}{2} \binom{p-2}{k-2}$. By the binomial theorem,

$$C_\beta(\sqrt{Q}) \le \frac{p-1}{2(1+z)} \sum_{k=2}^{p} \binom{p-2}{k-2} t^{k-2} = \frac{m+1}{2(1+z)} (1+t)^m.$$

By Proposition 5.20(1), we have $z > 2m$, and therefore $(m+1)/(2(1+z)) < (m+1)/(2(1+2m)) \le 3/10$, where the last inequality holds because $m \ge 2$ and $(m+1)/(1+2m)$ is decreasing in m. Further, $zt = x < 1$ by Proposition 5.20(2). Then $t < 1/z < 1/(2m)$ and $(1+t)^m \le e^{mt} < e^{1/2}$. It remains to bound $e^{1/2}$. For $j \ge 8$, the ratio of consecutive terms of $\sum_j 2^{-j}/j!$ is $1/(2(j+1)) \le 1/18$, and a geometric bound for the tail gives

$$e^{1/2} < \sum_{j=0}^{7} \frac{2^{-j}}{j!} + \frac{2^{-8}}{8!} \cdot \frac{1}{1-1/18} = \frac{48217579}{29245440} < \frac{53}{32}.$$

Combining these bounds, we obtain $C_\beta(\sqrt{Q}) < \frac{3}{10} \cdot \frac{53}{32} = \frac{159}{320}$. □

Proposition 5.23 (Small even perturbations of pure even models)**.** *Let $d \ge 4$ be even, let $P > d$, and let $\beta > \beta_s(d)$. There exists $\varepsilon_0 = \varepsilon_0(d, P, \beta) > 0$ with the following property. For every $\varepsilon \in [0, \varepsilon_0)$ and every family (b_p), indexed by the even integers p with $d < p \le P$, such that $b_p \ge 0$ and $\sum_p b_p \le 1$, the covariance*

$$\xi(r) = r^d + \varepsilon \sum_p b_p r^p$$

is such that (ξ, β) is strictly 1RSB with a nondegenerate contact. Moreover, its parameters converge to those of (r^d, β) as $\varepsilon \to 0$, uniformly over the families (b_p).

Proof. Recall from Proposition 5.20 that (r^d, β) satisfies the 1RSB assumptions. Let $\mathcal{D} = \{d\} \cup \{p \text{ even} : d < p \le P\}$, and apply Lemma 5.15 at the coefficient vector e_d of r^d and at β, with β held fixed. Fix $\varepsilon_0 > 0$ such that the compact convex set $B = \{(c, \beta) : \sum_{p \in \mathcal{D}} |c_p - (e_d)_p| \le \varepsilon_0\}$, with β fixed, is contained in the neighborhood $\mathcal{U}$ given by the

lemma. The coefficient vector of $r^d + \varepsilon \sum_p b_p r^p$ is $e_d + \varepsilon b$, which is nonnegative, nonzero, and at ℓ^1 distance at most $\varepsilon < \varepsilon_0$ from e_d. Using Lemma 5.15, we find that it satisfies the 1RSB assumptions with parameters $x(e_d + \varepsilon b, \beta)$ and $Q(e_d + \varepsilon b, \beta)$. The functions x and Q are smooth on $\mathcal{U}$, and their derivatives are bounded on B, and the mean value theorem shows that the parameters differ from those of (r^d, β) by $O(\varepsilon)$, uniformly in b. □

The proof uses neither the parity of d nor that of the perturbing degrees, which may be any finite set of degrees larger than d. The threshold ε_0 depends on β and P; we make no claim of uniformity as $\beta \downarrow \beta_s(d)$ or as $P \to \infty$.

Finally, we give a covariance that satisfies the 1RSB assumptions although $\xi''^{-1/2}$ is not convex. For related two-degree phase diagrams, Auffinger and Zhou classify the Parisi measure of spherical $p + s$ models at zero temperature [11, Theorems 2–4]; the proposition below verifies the positive-temperature contact conditions directly.

Proposition 5.24 (A strictly 1RSB example without convexity)**.** *Let* $Q = 9/10$, $x = 5/9$, *and*

$$t = \frac{45 - 24\log 6}{72\log 6 - 85}, \qquad b = t\Big(\frac{10}{9}\Big)^8, \qquad \xi(r) = r^4 + br^{12}, \qquad \beta^2 = \frac{15}{Q^3(4 + 12t)}. \tag{5.38}$$

Then $4539/100000 < t < 4541/100000$*; in particular* $b > 0$ *and* $\beta > 0$*. The pair* (ξ, β) *is strictly 1RSB with parameters* x, Q *and a nondegenerate contact at* Q*, and* $\beta^2(1 - Q)^2\xi''(Q) < 2/3$*. Further,* $(\xi''^{-1/2})''(5/6) < 0$*, so* $\xi''^{-1/2}$ *is not convex on* $(0, 1]$*.*

Numerically, $t \approx 0.0453970$, $b \approx 0.1054599$, and $\beta \approx 2.1277773$; these decimals are not used in the proof. The degrees 4 and 12 violate (5.33), since $6 \cdot 2 \cdot 10 = 120 < 121 = 11^2$, and Proposition 5.17 does not apply.

Proof. The proof uses two finite computations, which are carried out in Section A: the enclosure of t (Lemma A.1) and the sign pattern of a polynomial of degree twelve (Lemma A.3). We have $s = 1 - Q = 1/10$, $C = s + xQ = 3/5$, $C/s = 6$, and $z = xQ/s = 5$.

The contact equations. By Lemma A.1, $t_- < t < t_+$, where $t_- = 4539/100000$ and $t_+ = 4541/100000$; in particular t, b, and β^2 are positive. Solving the definition of t for $\log 6$, we find

$$\frac{1 + t}{4 + 12t} = \frac{6\log 6 - 5}{25}. \tag{5.39}$$

We have $bQ^8 = t$, $\xi(Q) = Q^4(1 + t)$, and $\xi'(Q) = Q^3(4 + 12t)$. It follows that

$$\beta^2\xi'(Q) = 15 = \frac{Q}{sC}, \qquad \beta^2 x^2 \xi(Q) = \frac{25(1 + t)}{6(4 + 12t)} = \log 6 - \frac{5}{6} = \log\frac{C}{s} - \frac{xQ}{C},$$

where we used (5.39). By the integrals in the proof of Proposition 5.11(1), these are the equations $V'(Q) = 0$ and $V(Q) = 0$.

Below the contact. For $0 \le r \le 1$, let $U(r) = x^2 V(Qr)$. For $0 \le q \le Q$, direct integration gives $\int_0^q (q-v)(C-xv)^{-2}\,\mathrm{d}v = x^{-2}(\log(C/(C-xq)) - xq/C)$. Since $C/(C-xQr) = 6/(6-5r)$, $xQr/C = 5r/6$, and $x^2\beta^2 Q^4 = 25/(6(4+12t))$, we obtain

$$U(r) = -\log\Big(1-\frac{5r}{6}\Big) - \frac{5r}{6} - \frac{25}{6(4+12t)}\big(r^4 + tr^{12}\big). \tag{5.40}$$

We have $U(0) = U'(0) = 0$, and $U(1) = x^2 V(Q) = 0$ and $U'(1) = x^2 Q V'(Q) = 0$ by the contact equations. Differentiating twice, we find

$$U''(r) = \frac{25\,P_t(r)}{6(4+12t)(6-5r)^2}, \qquad P_t(r) = 24 + 72t - \big(12r^2 + 132tr^{10}\big)(6-5r)^2. \tag{5.41}$$

By Lemma A.3, there exist $a_0 \in (357/1000, 359/1000)$ and $b_0 \in (890/1000, 892/1000)$ such that $U'' > 0$ on $(0, a_0)$, $U'' < 0$ on (a_0, b_0), and $U'' > 0$ on $(b_0, 1)$. Using $U'(0) = 0$, we find $U' > 0$ on $(0, a_0]$, and using $U'(1) = 0$, we find $U' < 0$ on $[b_0, 1)$. As U' is strictly decreasing on $[a_0, b_0]$, it has exactly one zero $c \in (a_0, b_0)$. We conclude that $U' > 0$ on $(0, c)$ and $U' < 0$ on $(c, 1)$. Together with $U(0) = U(1) = 0$, this shows that $U > 0$ on $(0, 1)$, that is, $V > 0$ on $(0, Q)$.

Above the contact. By (5.38) and $bQ^{10} = Q^2 t$,

$$\mu_Q = \beta^2 s^2 \xi''(Q) = \frac{15\,(12Q^2 + 132tQ^2)}{100\,Q^3(4+12t)} = \frac{15\,(12+132t)}{90\,(4+12t)} = \frac{12+132t}{24+72t}.$$

The right side is increasing in t, with derivative $2304/(24+72t)^2$, and therefore $\mu_Q < (12+132t_+)/(24+72t_+) = 149951/227246 < 2/3$. For $p \in \{4, 12\}$, the derivative of $(1-q)^2 q^{p-2}$ is $(1-q)q^{p-3}((p-2) - pq)$, which is negative for $q \ge Q$, since $Q = 9/10 > 5/6 \ge (p-2)/p$. This implies that $\beta^2(1-q)^2\xi''(q) \le \mu_Q < 1$ for $Q \le q < 1$. As $\hat{x}(q) = 1-q$ there, we have $V''(q) = (1-q)^{-2}(1 - \beta^2(1-q)^2\xi''(q)) > 0$ on $[Q, 1)$. In particular, $V''(Q) > 0$, and since $V(Q) = V'(Q) = 0$, we have $V > 0$ on $(Q, 1)$. We have now verified (5.15).

Nonconvexity. Let $v = \xi''$. Then $v(r) = 12r^2(1 + w(r))$ with $w(r) = 11br^8$. We have $rv'(r) = 12r^2(2 + 10w(r))$ and $r^2 v''(r) = 12r^2(2 + 90w(r))$, and

$$\frac{r^2\big(3v'(r)^2 - 2v(r)v''(r)\big)}{(12r^2)^2} = 3(2+10w)^2 - 2(1+w)(2+90w) = 8(1-3w)(1-5w),$$

where $w = w(r)$. At $r_0 = 5/6$, we have $w(r_0) = 11t(10/9)^8(5/6)^8 = 11t(25/27)^8$, and Lemma A.1 gives $1/4 < 11t_-(25/27)^8 < w(r_0) < 11t_+(25/27)^8 < 3/10$; the two outer inequalities are rational comparisons checked in Section A. Then $1 - 3w(r_0) > 0 > 1 - 5w(r_0)$, which implies that $3v'^2 - 2vv'' < 0$ and $(v^{-1/2})'' = (3v'^2 - 2vv'')/(4v^{5/2}) < 0$ at r_0. □

SECTION 6

The directing measure

In this section, we prove Theorem 6.3. It concerns the overlaps of independent samples $\sigma^1, \sigma^2, \dots$ from the Gibbs measure in the condensed phase. By the theorem of Dovbysh and Sudakov, every limit in distribution of the overlap array $(R(\sigma^i, \sigma^j))_{i,j \geq 1}$, averaged over the disorder, can be generated in two stages. One first draws a random probability measure G on the unit ball of a Hilbert space and then independent vectors $v_1, v_2, \dots$ with law G, and one sets $R_{ij} = v_i \cdot v_j$ for $i \neq j$. We call G a directing measure of the limit. Two samples whose directing vectors are close have overlap close to the squared norm of either vector. In this way, the squared norms of the vectors in the support of G are overlap values that many independent samples realize simultaneously. Theorem 6.3 states that if (ξ, β) is strictly 1RSB with parameters x and Q and a nondegenerate contact at Q, then every directing measure of every limit is carried by the vectors of squared norm Q.

This is the property of the Gibbs measure that the coarse cover by bands around critical points needs in Sections 9 and 12 (see Proposition 12.10). There, a Gibbs sample is placed in a band around a critical point of the center field H_Q as follows. Among a fixed finite number of additional independent replicas, we find n replicas whose overlaps with each other and with the given sample are all close to Q. The replicated estimates of Subag [143] show that the normalized average of these replicas has nearly maximal energy, and a localization argument then places it near a critical point of H_Q (Proposition 9.12 and Lemma 12.8). In the limit, such replicas exist with probability close to one if the directing vector v_1 of the given sample lies in the support of G and $\|v_1\|^2$ is close to Q. Indeed, the vectors in a small ball around v_1 then have inner products close to Q with each other and with v_1. An atom of G at the origin would describe a positive fraction of samples that are asymptotically orthogonal to all other samples, and such a sample has overlap close to Q with no additional replica. For a sample whose directing vector has another squared norm q, the additional replicas found in this way would have overlaps close to q instead of Q. Theorem 6.3 excludes both possibilities.

We use the following results of Section 5. Under the 1RSB assumptions of Definition 5.9, the first-variation function V is positive on $(0,1) \setminus \{Q\}$, and Theorem 5.7 and Proposition 5.11 show that 0 and Q are the only multisamplable overlaps. Lemma 5.8 expresses the failure of multisamplability at every other $q \in (0,1)$ as an exponential bound, valid for all large N, on the probability that a fixed number of samples have pairwise overlaps near q. Finally, F is differentiable at β (Corollary 5.16). From Section 3,

we use the tensor representation of H_N (Lemma 3.2), the concentration of free energies (Lemma 3.3), and the Parisi formula (Theorem 3.14).

The proof has three steps. In Section 6.3, we show that replica exclusion at an overlap q, that is, the convergence to zero of the probability that a fixed number of samples have pairwise overlaps near q, forbids directing vectors of squared norm near q (Lemma 6.5). We exclude vectors of norm one separately, by the entropy estimate of Lemma 6.4, since multisamplability concerns overlaps below one. After this step, the squared norms lie in $\{0, Q\}$. In Section 6.4, we derive one identity of Ghirlanda–Guerra type, with test function $\xi(R)$, from Gaussian integration by parts and the self-averaging of the energy at the point β where F is differentiable (Lemma 6.7). In Section 6.5, we show that this identity leaves no mass at the origin. The argument uses a single identity of this type, for the original Hamiltonian without a perturbation. It uses no ultrametricity, no atomicity of the directing measure, and no Poisson–Dirichlet structure. Odd degrees are allowed throughout: $\xi(R)$ may be negative at negative overlaps, and the last step uses only the upper bound $\xi(R) \le \xi(Q)$ for $|R| \le Q$.

Notation. Throughout this section, ξ is a finite mixture and $\beta > 0$. Conditionally on H_N, let $\sigma^1, \sigma^2, \dots$ be independent samples from π. We write $R_{ij} = R(\sigma^i, \sigma^j)$, so that $R_{ii} = 1$, and $\langle \cdot \rangle$ for the average over the samples; thus $\langle f \rangle = \int f \, \mathrm{d}\pi^{\otimes n}$ for a bounded Borel function f of $\sigma^1, \dots, \sigma^n$. Recall that $F_N(\beta) = N^{-1}\mathbb{E} \log Z_{N,\beta}$ and $F(\beta) = \lim_N F_N(\beta)$. We set

$$e_N = \mathbb{E}\Big\langle \frac{H_N(\sigma^1)}{N} \Big\rangle, \qquad \varepsilon_N = \mathbb{E}\Big\langle \Big| \frac{H_N(\sigma^1)}{N} - e_N \Big| \Big\rangle. \tag{6.1}$$

Recall from Definition 5.9 that (ξ, β) is strictly 1RSB with parameters $x, Q \in (0,1)$ and a nondegenerate contact at Q if ξ has minimum degree at least four and the function

$$V(r) = \int_0^r \frac{r - v}{\hat{x}(v)^2} \, \mathrm{d}v - \beta^2 \xi(r), \qquad \hat{x}(v) = \begin{cases} 1 - Q + x(Q - v), & 0 \le v \le Q, \\ 1 - v, & Q \le v < 1, \end{cases}$$

satisfies $V(Q) = V'(Q) = 0$, $V''(Q) > 0$, and $V(r) > 0$ for every $r \in (0,1) \setminus \{Q\}$.

6.1 Overlap arrays and directing measures

Let $\mathcal{M} = [-1,1]^{\mathbb{N}\times\mathbb{N}}$ be the space of arrays $(r_{ij})_{i,j\ge1}$ with entries in $[-1,1]$, with the product topology. It is compact and metrizable. We also write R_{ij} for the coordinate maps on $\mathcal{M}$. For each N, let $\mathcal{L}_N$ be the annealed law of the overlap array, that is,

$$\mathcal{L}_N(A) = \mathbb{E}\, \pi^{\otimes\infty}\big(\{(R(\sigma^i, \sigma^j))_{i,j\ge1} \in A\}\big)$$

for Borel sets $A \subset \mathcal{M}$. By the compactness of $\mathcal{M}$, every subsequence of $(\mathcal{L}_N)$ has a further subsequence that converges weakly. We call the limits *limiting overlap laws*. Weak

convergence on $\mathcal{M}$ is the convergence of the joint laws of all finite collections of entries. For a limiting overlap law $\mathcal{L}$, we write $\mathbb{E}_{\mathcal{L}}$ for the expectation with respect to $\mathcal{L}$.

Each $\mathcal{L}_N$ is carried by the set of symmetric arrays with diagonal entries one whose finite principal submatrices are positive semidefinite, since $R(\sigma^i, \sigma^j) = U(\sigma^i) \cdot U(\sigma^j)$ is a Gram matrix. This set is closed. Further, $\mathcal{L}_N$ is invariant under the maps of $\mathcal{M}$ that permute finitely many indices, because the samples are exchangeable. Both properties pass to weak limits: the first by the portmanteau theorem, since the set is closed, and the second because the permutation maps are continuous. Every limiting overlap law $\mathcal{L}$ is the law of a symmetric, nonnegative definite, weakly exchangeable array with diagonal entries one. We use two further consequences of weak convergence along a subsequence (N_m) with limit $\mathcal{L}$. If $E \subset \mathcal{M}$ is open, then $\mathcal{L}(E) \le \liminf_m \mathcal{L}_{N_m}(E)$. If f is a bounded continuous function on $\mathcal{M}$ that depends on finitely many coordinates, then $\mathbb{E}_{\mathcal{L}} f = \lim_m \mathbb{E}\langle f \rangle$, where on the right f is evaluated at the overlaps of the samples.

Let B be the closed unit ball of the Hilbert space ℓ^2, and let $B(v, r)$ be the open ball of radius r around $v \in \ell^2$. For a Borel probability measure G on B, let $\mathcal{L}^G$ be the law of the array with entries $v_i \cdot v_j$ for $i \neq j$ and 1 on the diagonal, where $v_1, v_2, \dots$ are independent with law G.

Definition 6.1 (Directing measure)**.** Let $\mathcal{L}$ be a Borel probability measure on $\mathcal{M}$. A *directing measure* of $\mathcal{L}$ is a family $(G_\omega)_{\omega \in \Omega}$ of Borel probability measures on B, indexed by a probability space $(\Omega, \mathcal{F}, \Pr)$, such that for every Borel set $A \subset \mathcal{M}$, the map $\omega \mapsto \mathcal{L}^{G_\omega}(A)$ is measurable and

$$\mathcal{L}(A) = \int_\Omega \mathcal{L}^{G_\omega}(A) \, \Pr(\, \mathrm{d}\omega).$$

For each fixed ω, the array law $\mathcal{L}^{G_\omega}$ is generated by independent vectors with law G_ω, and $\mathcal{L}$ is the mixture of these array laws. If $\omega \mapsto G_\omega(A)$ is measurable for every Borel set $A \subset \mathsf{B}$, this also gives a joint construction, in which one draws ω with law Pr and then independent vectors $v_1, v_2, \dots$ with law G_ω, and sets $R_{ij} = v_i \cdot v_j$ for $i \neq j$. The definition does not require this measurability, and the proofs below use only the measurable functions $\omega \mapsto \mathcal{L}^{G_\omega}(A)$.

Theorem 6.2 (Dovbysh–Sudakov; [123, Proposition 1])**.** *Every limiting overlap law has a directing measure.*

This is the representation theorem of Dovbysh and Sudakov [63], in the form proved by Panchenko [123, Proposition 1]. That proposition applies to the law of every symmetric, nonnegative definite, weakly exchangeable array, and we checked above that limiting overlap laws are of this kind. It states that such a law is a mixture, over a probability space $(\Omega, \mathcal{F}, \Pr)$, of laws $\mathcal{Q}_\omega$ on $\mathcal{M}$ such that $\omega \mapsto \mathcal{Q}_\omega(A)$ is measurable for every Borel set A, and such that each $\mathcal{Q}_\omega$ is the law of the array $(h_i \cdot h_j + a_i \mathbf{1}\{i = j\})_{i,j \ge 1}$ for an independent sequence of pairs (h_i, a_i) with a common law η_ω on $\ell^2 \times [0, \infty)$ [123, (1.1)–(1.2)]. Since

the diagonal entries equal one $\mathcal{L}$-almost surely, we have $\int \mathcal{Q}_\omega(\{R_{11} = 1\}) \Pr(\mathrm{d}\omega) = 1$, and the measurable set Ω_0 of those ω with $\mathcal{Q}_\omega(\{R_{11} = 1\}) = 1$ has $\Pr(\Omega_0) = 1$. For $\omega \in \Omega_0$, the measure η_ω is carried by $\{\|h\|^2 + a = 1\}$, hence by $\{\|h\| \le 1\}$. We let G_ω be the image of η_ω under $(h, a) \mapsto h$. Then $\mathcal{L}^{G_\omega} = \mathcal{Q}_\omega$. For $\omega \notin \Omega_0$, we let $G_\omega = \delta_0$. This defines a directing measure of $\mathcal{L}$. A directing measure is not unique; for example, it may be composed with any linear isometry of ℓ^2. The following theorem applies to every directing measure.

Theorem 6.3 (Directing vectors have squared norm Q)**.** *Let (ξ, β) be strictly 1RSB with parameters x and Q and a nondegenerate contact at Q (Definition 5.9). Let $\mathcal{L}$ be a limiting overlap law, and let $(G_\omega)_{\omega \in \Omega}$ be a directing measure of $\mathcal{L}$. Then, for* Pr*-almost every ω,*

$$G_\omega\big(\{v \in \mathsf{B} : \|v\|^2 = Q\}\big) = 1.$$

Equivalently, for Pr-almost every ω, if $v_1, v_2, \dots$ are independent with law G_ω, then almost surely $\|v_i\|^2 = Q$ for every i. Then $\|v_i - v_j\|^2 = 2(Q - v_i \cdot v_j)$ for $i \neq j$, so that distances between directing vectors are functions of the array. In particular, $|R_{ij}| \le Q$ for all $i \neq j$, $\mathcal{L}$-almost surely. When the directing measure is a random measure G, that is, when $\omega \mapsto G_\omega(A)$ is measurable for every Borel set $A \subset \mathsf{B}$, the conclusion reads $G(\{\|v\|^2 = Q\}) = 1$ almost surely.

6.2 Pairs of samples near overlap one

The results of Section 5 concern overlaps in $[0, 1)$, and they do not exclude directing vectors of norm one. We exclude them with an entropy estimate. On an event of probability close to one, the density of π with respect to ν_N is at most $e^{2\beta K N}$, while a spherical cap $\{R(\cdot, u) > 1 - \delta\}$ has ν_N-mass at most $\sqrt{N}(2\delta)^{(N-3)/2}$; for small δ, the second factor wins.

Lemma 6.4 (A field bound and a cap estimate)**.** *Let $\mathcal{P} = \{p : c_p > 0\}$, and set*

$$a_p = \big(2p \log(1 + 4p) + 2\big)^{1/2}, \qquad K = 2 \sum_{p \in \mathcal{P}} \sqrt{c_p}\, a_p.$$

1. *For every $N \ge 2$,*
$$\mathbb{P}\Big(\sup_{x \in S_N} |H_N(x)| > KN \Big) \le 2\,|\mathcal{P}|\, e^{-N}. \tag{6.2}$$
2. *For every $N \ge 3$, every $u \in S_N$, and every $\delta \in (0, 1/2)$,*
$$\nu_N\big(\{x \in S_N : R(x, u) > 1 - \delta\}\big) \le \sqrt{N}\,(2\delta)^{(N-3)/2}. \tag{6.3}$$
3. *Let $\delta = e^{-4\beta K - 4}/2$. For every $N \ge 3$,*
$$\mathbb{E}\,\pi^{\otimes 2}\big(\{R_{12} > 1 - \delta\}\big) \le 2\,|\mathcal{P}|\, e^{-N} + e^{6\beta K + 6}\sqrt{N}\, e^{-2N}. \tag{6.4}$$

Part (1) is an explicit version of the bound on $\sup|H_N|$ in Lemma 3.2(3). We include its short proof, since it gives a value of K and holds for every N.

Proof. For (1), let $v \in \mathbb{R}^N$ be a unit vector. By (3.2), we have

$$H_N(\sqrt{N}v) = \sqrt{N}\sum_{p\in\mathcal{P}} \sqrt{c_p}\, T_p(v,\dots,v), \qquad T_p(v_1,\dots,v_p) = \sum_{i_1,\dots,i_p=1}^{N} g^{(p)}_{i_1\dots i_p} v_{1,i_1}\cdots v_{p,i_p},$$

where $v_{k,i}$ is the i-th coordinate of v_k. For unit vectors $v_1,\dots,v_p$, the variable $T_p(v_1,\dots,v_p)$ is centered Gaussian with variance $\prod_k |v_k|^2 = 1$. Let $\|T_p\|$ be the supremum of $|T_p(v_1,\dots,v_p)|$ over unit vectors $v_1,\dots,v_p$. Fix $p \in \mathcal{P}$ and let $\varepsilon = 1/(2p)$. A maximal ε-separated subset $\mathcal{N}$ of the unit sphere of $\mathbb{R}^N$ is an ε-net. The balls of radius $\varepsilon/2$ around its points are disjoint and contained in the ball of radius $1+\varepsilon/2$, and comparing volumes, as in [147, Appendix A.8, Proposition A.8.1], we find $|\mathcal{N}| \le (1+2/\varepsilon)^N = (1+4p)^N$. Given unit vectors $v_1,\dots,v_p$, choose $w_k \in \mathcal{N}$ with $|v_k - w_k| \le \varepsilon$. By multilinearity,

$$T_p(v_1,\dots,v_p) - T_p(w_1,\dots,w_p) = \sum_{k=1}^{p} T_p(w_1,\dots,w_{k-1}, v_k - w_k, v_{k+1},\dots,v_p),$$

and each term on the right has absolute value at most $\varepsilon\|T_p\|$. Let M_p be the maximum of $|T_p|$ over $\mathcal{N}^p$. Taking the supremum over $v_1,\dots,v_p$, we obtain $\|T_p\| \le M_p + p\varepsilon\|T_p\| = M_p + \|T_p\|/2$, that is, $\|T_p\| \le 2M_p$. The set $\mathcal{N}^p$ has at most $(1+4p)^{pN}$ elements. Using a union bound and the Gaussian tail bound $\mathbb{P}(|\gamma| > t) \le 2e^{-t^2/2}$ for a standard Gaussian variable γ, we obtain

$$\mathbb{P}\big(\|T_p\| > 2a_p\sqrt{N}\big) \le \mathbb{P}\big(M_p > a_p\sqrt{N}\big) \le 2(1+4p)^{pN} e^{-a_p^2 N/2} = 2e^{-N},$$

where we used $a_p^2/2 = p\log(1+4p) + 1$. On the intersection of the events $\{\|T_p\| \le 2a_p\sqrt{N}\}$, $p \in \mathcal{P}$, we have $|H_N(x)| \le \sqrt{N}\sum_p \sqrt{c_p}\,\|T_p\| \le KN$ for every $x \in S_N$. This proves (1).

For (2), recall that for $N \ge 3$ and $u \in S_N$, the law of $R(x,u)$ under $\nu_N(\mathrm{d}x)$ has the density $d_N(1-r^2)^{(N-3)/2}$ on $[-1,1]$, where $d_N^{-1} = \int_{-1}^{1}(1-r^2)^{(N-3)/2}\,\mathrm{d}r$. Restricting this integral to $|r| \le N^{-1/2}$, we find

$$d_N^{-1} \ge 2N^{-1/2}\Big(1-\frac{1}{N}\Big)^{(N-3)/2} \ge 2N^{-1/2}e^{-1/2},$$

where we used $\log(1-1/N) \ge -1/(N-1)$, which follows from $\log(1-t) \ge -t/(1-t)$, and $(N-3)/(N-1) \le 1$. We conclude that $d_N \le e^{1/2}\sqrt{N}/2 < \sqrt{N}$. For $r \in (1-\delta, 1]$, we have $1-r^2 = (1-r)(1+r) \le 2\delta$. Integrating over this interval, whose length is $\delta < 1$, we obtain $\nu_N(\{R(\cdot,u) > 1-\delta\}) \le d_N\delta(2\delta)^{(N-3)/2} \le \sqrt{N}(2\delta)^{(N-3)/2}$.

For (3), let $\mathcal{K}$ be the event $\{\sup_{S_N} |H_N| \le KN\}$. On $\mathcal{K}$, we have $Z_{N,\beta} \ge e^{-\beta KN}$, and the density of π with respect to ν_N is at most $e^{2\beta KN}$. By (2), on $\mathcal{K}$,

$$\pi^{\otimes 2}\big(\{R_{12} > 1-\delta\}\big) = \int \pi\big(\{x : R(x,y) > 1-\delta\}\big)\, \pi(\,\mathrm{d}y) \le e^{2\beta KN}\sqrt{N}\,(2\delta)^{(N-3)/2}.$$

The integration in y does not produce a second density factor, since π is a probability measure. Inserting $\log(2\delta) = -4\beta K - 4$, we find that the right side equals $\sqrt{N}\exp(2\beta KN - (N-3)(2\beta K+2)) = e^{6\beta K+6}\sqrt{N}\,e^{-2N}$. Off $\mathcal{K}$, we bound $\pi^{\otimes 2}(\{R_{12} > 1-\delta\})$ by one. Taking expectations and using (1), we obtain (6.4). □

6.3 Replica exclusion and the norms of directing vectors

Recall from (5.6) the annealed probability $p_{N,k,\rho}(q)$ that k samples have all their pairwise overlaps within ρ of q, and recall that q is multisamplable if $N^{-1}\log p_{N,k,\rho}(q) \to 0$ for every k and ρ (Definition 5.4, following Subag [143, Definition 2]). Theorem 5.5(2), which is [143, footnote 4 and Remark 25], shows that the definition does not change if the limit is replaced by the limit superior. This form matters here. If q is not multisamplable, there exist k and ρ with $\limsup_N N^{-1}\log p_{N,k,\rho}(q) < 0$. Then $p_{N,k,\rho}(q)$ tends to zero along the full sequence of dimensions, and hence along the subsequence that defines any given limiting overlap law. With the limit alone, the failure of multisamplability would give decay only along some subsequence, which need not be related to the limiting overlap law at hand. Under the 1RSB assumptions, we have $V_\beta = V$ by Proposition 5.11(1), and every $q \in (0,1)\setminus\{Q\}$ has $V(q) > 0$. For such q, Lemma 5.8 implies that there exist $k \ge 2$, $\rho \in (0,1)$, and $c > 0$ such that $p_{N,k,\rho}(q) \le e^{-cN}$ for all sufficiently large N.

The next lemma converts replica exclusion into a restriction on the directing measure. Its mechanism is the standard support argument for Dovbysh–Sudakov representations: vectors in a small ball around a point v of the support have mutual inner products close to $\|v\|^2$.

Lemma 6.5 (Replica exclusion restricts the directing norms)**.** *Let $\mathcal{A} \subset [0,1)$ be closed relative to $[0,1)$, with $0 \in \mathcal{A}$. Suppose that for every $q \in (0,1)\setminus\mathcal{A}$, there exist an integer $k_q \ge 2$ and $\rho_q > 0$ such that $p_{N,k_q,\rho_q}(q) \to 0$ as $N \to \infty$. Let $\mathcal{L}$ be a limiting overlap law and $(G_\omega)_{\omega\in\Omega}$ a directing measure of $\mathcal{L}$. Then, for* Pr*-almost every ω,*

$$G_\omega\big(\{v \in \mathsf{B} : \|v\|^2 \in \mathcal{A}\}\big) = 1.$$

Proof. Since $\mathcal{A}$ is closed relative to $[0,1)$, it is a Borel subset of $[0,1]$, and the set $\{v \in \mathsf{B} : \|v\|^2 \in \mathcal{A}\}$ is a Borel subset of B, as the preimage of $\mathcal{A}$ under a continuous map. The same holds for $(0,1)\setminus\mathcal{A}$. We next recall a fact about supports. For a Borel probability measure G on B, its support $\operatorname{supp} G$ is the set of $v \in \mathsf{B}$ with $G(B(v,r)) > 0$ for every $r > 0$.

The complement of $\operatorname{supp} G$ in B is a union of relatively open balls of G-measure zero. As ℓ^2 is separable, countably many of these balls cover it. This implies that $G(\operatorname{supp} G) = 1$.

Fix a subsequence (N_m) along which $\mathcal{L}_{N_m}$ converges to $\mathcal{L}$. Let $q \in (0,1) \setminus \mathcal{A}$, and write $k = k_q$ and $\rho = \rho_q$. The set

$$E_q = \Big\{ \max_{1 \le i < j \le k} \big| R_{ij} - q \big| < \rho \Big\} \subset \mathcal{M}$$

is open, since it is defined by strict inequalities between finitely many coordinates. By the portmanteau theorem and the hypothesis, we have $\mathcal{L}(E_q) \le \liminf_m p_{N_m,k,\rho}(q) = 0$. By Definition 6.1, there exists a Pr-null set Ω_q such that for $\omega \notin \Omega_q$,

$$G_\omega^{\otimes k}\Big(\Big\{ (v_1, \dots, v_k) : \max_{1 \le i < j \le k} \big| v_i \cdot v_j - q \big| < \rho \Big\} \Big) = \mathcal{L}^{G_\omega}(E_q) = 0. \tag{6.5}$$

Let $\omega \notin \Omega_q$, and suppose that some $v \in \operatorname{supp} G_\omega$ satisfies $\Big| \|v\|^2 - q \Big| < \rho/2$. Let $r = \rho/4$. For $v_i, v_j \in B(v,r)$, since all vectors in B have norm at most one, we have

$$\Big| v_i \cdot v_j - \|v\|^2 \Big| \le \big| (v_i - v) \cdot v_j \big| + \big| v \cdot (v_j - v) \big| < 2r = \rho/2,$$

and therefore $\big| v_i \cdot v_j - q \big| < \rho$. The left side of (6.5) is then at least $G_\omega(B(v,r))^k > 0$, which is a contradiction. We conclude that for $\omega \notin \Omega_q$, no point of $\operatorname{supp} G_\omega$ has squared norm in the interval $I_q = (q - \rho_q/2, q + \rho_q/2)$.

The intervals I_q, $q \in (0,1) \setminus \mathcal{A}$, cover $(0,1) \setminus \mathcal{A}$. Every subset of $\mathbb{R}$ has the Lindelöf property, which means here that countably many of these intervals, say $I_{q_1}, I_{q_2}, \dots$, already cover $(0,1) \setminus \mathcal{A}$. For ω outside the null set $\bigcup_l \Omega_{q_l}$, no point of $\operatorname{supp} G_\omega$ has squared norm in $(0,1) \setminus \mathcal{A}$. Together with $G_\omega(\operatorname{supp} G_\omega) = 1$, this implies that $G_\omega(\{\|v\|^2 \in (0,1) \setminus \mathcal{A}\}) = 0$.

It remains to exclude norm one. Let δ be as in Lemma 6.4(3). The set $\{R_{12} > 1 - \delta\} \subset \mathcal{M}$ is open, and by (6.4) and the portmanteau theorem, we have $\mathcal{L}(\{R_{12} > 1 - \delta\}) = 0$. As above, there exists a Pr-null set Ω_1 such that $G_\omega^{\otimes 2}(\{v_1 \cdot v_2 > 1 - \delta\}) = 0$ for $\omega \notin \Omega_1$. Let $\omega \notin \Omega_1$, and suppose that $v \in \operatorname{supp} G_\omega$ has $\|v\| = 1$. For $v_1, v_2 \in B(v, \delta/2)$, the estimate displayed above, with $r = \delta/2$, shows that $v_1 \cdot v_2 > \|v\|^2 - \delta = 1 - \delta$. Then $G_\omega^{\otimes 2}(\{v_1 \cdot v_2 > 1 - \delta\}) \ge G_\omega(B(v, \delta/2))^2 > 0$, which is a contradiction. Hence $\operatorname{supp} G_\omega$ contains no vector of norm one, and $G_\omega(\{\|v\| = 1\}) = 0$.

The interval $[0,1]$ is the union of $\mathcal{A}$, $(0,1) \setminus \mathcal{A}$, and $\{1\}$, because $0 \in \mathcal{A} \subset [0,1)$. For ω outside the null set $\Omega_1 \cup \bigcup_l \Omega_{q_l}$, the last two sets have G_ω-measure zero, which proves the lemma. □

6.4 A Ghirlanda–Guerra identity for the covariance

The Ghirlanda–Guerra identities were discovered by Ghirlanda and Guerra [83], and they are closely related to the stochastic stability of Aizenman and Contucci [4]. In their full

form, they hold for Hamiltonians with a small generic perturbation. We need only the single identity whose test function is the covariance $\xi(R)$, for the original Hamiltonian at a fixed β where F is differentiable. It follows from Gaussian integration by parts once the energy H_N/N is known to concentrate under the annealed Gibbs measure. This self-averaging at a point of differentiability is the mechanism of Panchenko [124]. The main theorem of [124], Theorem 1.1 there, is stated for Ising mixed p-spin models, but the remark that follows it proves the self-averaging for general Gibbs measures. We use it in that form.

Proposition 6.6 (Self-averaging of the energy; [124, Remark following Theorem 1.1])**.** *Let ξ be a finite mixture and $\beta > 0$, and suppose that F is differentiable at β. Then $\varepsilon_N \to 0$ as $N \to \infty$.*

Proof. The remark following [124, Theorem 1.1] concerns probability measures ν_N on a common measurable space, random measurable functions A_N on that space, and the Gibbs averages $\langle\cdot\rangle_t$ with density proportional to e^{tA_N} with respect to ν_N. Let $\psi_N(t) = N^{-1}\log\int e^{tA_N}\,\mathrm{d}\nu_N$. The remark states that if $\mathbb{E}\,|\psi_N(t) - \mathbb{E}\psi_N(t)| \to 0$ and $\mathbb{E}\psi_N(t)$ converges to a limit $P(t)$ for every t in a neighborhood of t_0, and if P is differentiable at t_0, then

$$\lim_{N\to\infty}\frac{1}{N}\mathbb{E}\big\langle\, |A_N - \mathbb{E}\langle A_N\rangle_{t_0}|\,\big\rangle_{t_0} = 0,$$

under measurability and integrability conditions that are implicit in its proof [124, (7), p. 190]. We take the disjoint union of the spheres S_N as the common space, with ν_N the uniform law on S_N, $A_N = H_N$ on S_N, $t_0 = \beta$, and $P = F$; the conclusion is then $\varepsilon_N \to 0$.

We verify the hypotheses. For every $t > 0$, we have $\psi_N(t) = N^{-1}\log Z_{N,t}$, and using Lemma 3.3 with $B = S_N$ and $h = 0$, we find

$$\begin{aligned}\mathbb{E}\,|\psi_N(t) - \mathbb{E}\psi_N(t)| &= \int_0^\infty \mathbb{P}\big(\,|\psi_N(t) - \mathbb{E}\psi_N(t)| \ge s\big)\,\mathrm{d}s\\ &\le \int_0^\infty 2e^{-s^2N/(2t^2\xi(1))}\,\mathrm{d}s = t\Big(\frac{2\pi\xi(1)}{N}\Big)^{1/2}.\end{aligned}$$

By Theorem 3.14, we have $\mathbb{E}\psi_N(t) = F_N(t) \to F(t)$ for every $t \ge 0$, and F is differentiable at β by assumption. For the integrability conditions, we write $H_N(x) = g\cdot\Psi(x)$ as in the proof of Lemma 3.3, where g is a standard Gaussian vector of finite dimension and $|\Psi(x)|^2 = N\xi(1)$ for $x \in S_N$. Then H_N is continuous on S_N and $\sup_{S_N}|H_N| \le |g|\,(N\xi(1))^{1/2}$, which has finite moments of all orders. This justifies the differentiations in t under the Gibbs average and the expectation that occur in the proof of [124, (7)]. □

Recall from Corollary 5.16 that, under the 1RSB assumptions with parameters x and Q, the free energy F is differentiable at β, with $F'(\beta) = \beta(\xi(1) - (1-x)\xi(Q))$ by (5.32), and that $e_N = F_N'(\beta) \to F'(\beta)$ as $N \to \infty$.

Lemma 6.7 (The Ghirlanda–Guerra identity for $\xi(R)$)**.** *Let (ξ, β) be strictly 1RSB with parameters x and Q and a nondegenerate contact at Q.*

1. *For every $N \geq 2$, every $n \geq 1$, and every bounded Borel function f on S_N^n, evaluated at $(\sigma^1, \dots, \sigma^n)$,*

$$\left| \mathbb{E}\langle f\xi(R_{1,n+1})\rangle - \frac{(1-x)\xi(Q)}{n}\mathbb{E}\langle f\rangle - \frac{1}{n}\sum_{j=2}^{n}\mathbb{E}\langle f\xi(R_{1j})\rangle \right| \leq \frac{\|f\|_\infty}{\beta n}\big(\varepsilon_N + \big|e_N - F'(\beta)\big|\big). \quad (6.6)$$

   Moreover, $\varepsilon_N + |e_N - F'(\beta)| \to 0$ as $N \to \infty$.

2. *Let $\mathcal{L}$ be a limiting overlap law. For every $n \geq 1$ and every bounded Borel function f of the coordinates $(R_{ij})_{1 \leq i < j \leq n}$,*

$$\mathbb{E}_{\mathcal{L}}\big[f\xi(R_{1,n+1})\big] = \frac{(1-x)\xi(Q)}{n}\mathbb{E}_{\mathcal{L}} f + \frac{1}{n}\sum_{j=2}^{n}\mathbb{E}_{\mathcal{L}}\big[f\xi(R_{1j})\big]. \quad (6.7)$$

   In particular, $\mathbb{E}_{\mathcal{L}}\xi(R_{12}) = (1-x)\xi(Q)$.

For $n = 1$, the sums over j are empty, and in (2) the function f is a constant.

Proof. We begin with the last assertion of (1). By Corollary 5.16, we have $e_N = F_N'(\beta) \to F'(\beta)$, and F is differentiable at β. Then Proposition 6.6 implies that $\varepsilon_N \to 0$.

Next, we derive an exact identity at finite N by Gaussian integration by parts. We write $H_N(x) = g \cdot \Psi(x) = \sum_\alpha g_\alpha \Psi_\alpha(x)$ as in the proof of Proposition 6.6. By the computation in the proof of Lemma 3.2(1), we have $\sum_\alpha \Psi_\alpha(x)\Psi_\alpha(y) = N\xi(R(x,y))$. Fix $n \geq 1$ and a bounded Borel function f on S_N^n, and for each index α, let $\Phi_\alpha(g) = \langle f\Psi_\alpha(\sigma^1)\rangle$; this Gibbs average depends on g through π. Differentiating the n Gibbs densities and the normalization $Z_{N,\beta}^{-n}$, we find

$$\partial_{g_\alpha}\Phi_\alpha(g) = \beta\Big\langle f\Psi_\alpha(\sigma^1)\Big(\sum_{l=1}^{n}\Psi_\alpha(\sigma^l) - n\Psi_\alpha(\sigma^{n+1})\Big)\Big\rangle.$$

The function f is not differentiated, since it depends on the samples and not on g. The bound $|\Psi_\alpha| \leq (N\xi(1))^{1/2}$ on S_N implies that Φ_α is bounded and smooth with bounded first derivatives. Then Gaussian integration by parts gives $\mathbb{E}[g_\alpha\Phi_\alpha(g)] = \mathbb{E}[\partial_{g_\alpha}\Phi_\alpha(g)]$. Summing over α and dividing by N, we obtain

$$\mathbb{E}\Big\langle f\frac{H_N(\sigma^1)}{N}\Big\rangle = \beta\,\mathbb{E}\Big\langle f\Big(\xi(1) + \sum_{j=2}^{n}\xi(R_{1j}) - n\xi(R_{1,n+1})\Big)\Big\rangle, \quad (6.8)$$

where the term $l = 1$ contributed $\xi(R_{11}) = \xi(1)$.

Taking $n = 1$ and $f = 1$ in (6.8), we have $e_N = \beta(\xi(1) - \mathbb{E}\langle \xi(R_{12})\rangle)$. Inserting this into (6.8) and rearranging, we find

$$n\mathbb{E}\langle f\xi(R_{1,n+1})\rangle - \sum_{j=2}^{n} \mathbb{E}\langle f\xi(R_{1j})\rangle - \mathbb{E}\langle \xi(R_{12})\rangle \mathbb{E}\langle f\rangle = -\frac{1}{\beta}\mathbb{E}\Big\langle f\Big(\frac{H_N(\sigma^1)}{N} - e_N\Big)\Big\rangle.$$

The right side has absolute value at most $\|f\|_\infty \varepsilon_N/\beta$. By (5.32) and the formula for e_N, we have $|\mathbb{E}\langle \xi(R_{12})\rangle - (1-x)\xi(Q)| = |e_N - F'(\beta)|/\beta$, and $|\mathbb{E}\langle f\rangle| \le \|f\|_\infty$. Combining these bounds and dividing by n, we obtain (6.6).

For (2), let (N_m) be a subsequence along which $\mathcal{L}_{N_m}$ converges to $\mathcal{L}$, and write $R^{(n)} = (R_{ij})_{1\le i<j\le n}$. Suppose first that f is continuous. Then $f(R^{(n)})$ and $f(R^{(n)})\xi(R_{1j})$, $2 \le j \le n+1$, are bounded continuous functions on $\mathcal{M}$ that depend on finitely many coordinates. We apply (6.6) to the bounded Borel function $(\sigma^1, \dots, \sigma^n) \mapsto f((R(\sigma^i, \sigma^j))_{1\le i<j\le n})$ and let $m \to \infty$. Using the last assertion of (1), we find that the right side of (6.6) tends to zero, which proves (6.7) for continuous f. For general f, both sides of (6.7) have the form $\int f \,\mathrm{d}\mu$ for a finite signed Borel measure μ on $[-1,1]^{n(n-1)/2}$, since ξ is bounded on $[-1,1]$; for the left side, $\mu(A) = \mathbb{E}_{\mathcal{L}}[\mathbf{1}_A(R^{(n)})\xi(R_{1,n+1})]$. Two finite signed Borel measures on a compact metric space that agree on continuous functions are equal, and (6.7) extends to every bounded Borel f. For $n = 1$ and $f = 1$, it reads $\mathbb{E}_{\mathcal{L}}\xi(R_{12}) = (1-x)\xi(Q)$. □

6.5 Proof of Theorem 6.3

Let $\mathcal{L}$ be a limiting overlap law and $(G_\omega)_{\omega\in\Omega}$ a directing measure of $\mathcal{L}$. The proof has three steps: we restrict the squared norms to $\{0, Q\}$, derive a recursion from Lemma 6.7, and use it to remove the mass at the origin.

Squared norms in $\{0, Q\}$. Under the 1RSB assumptions, $V_\beta = V$ by Proposition 5.11(1), and $V(q) > 0$ for every $q \in (0,1) \setminus \{Q\}$. For each such q, by Lemma 5.8, there exist $k_q \ge 2$ and $\rho_q > 0$ such that $p_{N,k_q,\rho_q}(q) \to 0$. Applying Lemma 6.5 with $\mathcal{A} = \{0, Q\}$, we find that for Pr-almost every ω, the measure G_ω is carried by the vectors of squared norm 0 or Q. For such ω and independent v_1, v_2 with law G_ω, we have $|v_1 \cdot v_2| \le \|v_1\|\,\|v_2\| \le Q$ almost surely. Integrating over ω and using the weak exchangeability of $\mathcal{L}$, we conclude that $|R_{ij}| \le Q$ for all $i \ne j$, $\mathcal{L}$-almost surely. Since the coefficients of ξ are nonnegative, we have, $\mathcal{L}$-almost surely,

$$\xi(R_{ij}) \le \sum_p c_p |R_{ij}|^p \le \xi(Q) \qquad (i \ne j). \tag{6.9}$$

Note that (6.9) holds also when ξ has odd degrees, in which case $\xi(R_{ij})$ may be negative.

A recursion. For $n \ge 1$, let

$$A_n = \{\xi(R_{1j}) = 0 \text{ for } 2 \le j \le n\} \subset \mathcal{M}, \qquad a_n = \mathcal{L}(A_n),$$

so that $A_1 = \mathcal{M}$ and $a_1 = 1$. The indicator $\mathbf{1}_{A_n}$ is a bounded Borel function of $(R_{ij})_{1\le i<j\le n}$. On A_n, we have $\xi(R_{1j}) = 0$ for $2 \le j \le n$, so the sum in (6.7) vanishes for $f = \mathbf{1}_{A_n}$, and Lemma 6.7(2) gives

$$\mathbb{E}_{\mathcal{L}}\big[\mathbf{1}_{A_n}\xi(R_{1,n+1})\big] = \frac{(1-x)\xi(Q)}{n}\, a_n.$$

On the other hand, $A_{n+1} = A_n \cap \{\xi(R_{1,n+1}) = 0\}$. The integrand on the left vanishes on A_{n+1}, and on $A_n \setminus A_{n+1}$ it is at most $\xi(Q)$ by (6.9). The left side is therefore at most $\xi(Q)(a_n - a_{n+1})$. Comparing the two bounds and dividing by $\xi(Q) > 0$ yields

$$a_{n+1} \le \Big(1 - \frac{1-x}{n}\Big) a_n.$$

Every factor lies in $(0,1)$, because $0 < 1 - x < 1$. Iterating from $a_1 = 1$, we obtain

$$a_n \le \prod_{j=1}^{n-1} \Big(1 - \frac{1-x}{j}\Big) \le \exp\Big(-(1-x)\sum_{j=1}^{n-1}\frac{1}{j}\Big) \le n^{-(1-x)}, \tag{6.10}$$

where we used $\log(1-t) \le -t$ and $\sum_{j=1}^{n-1} j^{-1} \ge \int_1^n \mathrm{d}t/t = \log n$.

No mass at the origin. Fix $\omega \in \Omega$, and let $v_1, v_2, \dots$ be independent with law G_ω. If $v_1 = 0$, then $R_{1j} = v_1 \cdot v_j = 0$ and $\xi(R_{1j}) = 0$ for every $j \ge 2$, and the array belongs to A_n for every n. This implies that $G_\omega(\{0\}) \le \mathcal{L}^{G_\omega}(A_n)$ for every n. The numbers $\mathcal{L}^{G_\omega}(A_n)$ are nonincreasing in n, and by Definition 6.1 and (6.10),

$$\int_\Omega \mathcal{L}^{G_\omega}(A_n)\ \Pr(\,\mathrm{d}\omega) = a_n \le n^{-(1-x)}.$$

By dominated convergence, $\lim_n \mathcal{L}^{G_\omega}(A_n) = 0$ for Pr-almost every ω, and therefore $G_\omega(\{0\}) = 0$ for Pr-almost every ω. Together with the first step, this shows that G_ω is carried by $\{\|v\|^2 = Q\}$ for Pr-almost every ω, which completes the proof of Theorem 6.3.

SECTION 7

Determinant moments and the Kac–Rice formula

This section relates the critical points of the center field H_Q to estimates under ordinary Gaussian conditioning at a single deterministic point. Lemma 7.1 records the joint law of the value, the radial derivative, the tangent gradient, and the Hessian of H_Q at a point, and Definition 7.3 fixes the version of the conditional law given these quantities that is used throughout Part II; Corollary 7.6 states the resulting law of the Hessian at a conditioned critical point. Proposition 7.7 bounds every fixed moment of the normalized absolute Hessian determinant when the normalized radial derivative lies a fixed distance above the edge of the semicircle law. Section 7.4 adds a normalization of the expected determinant by the semicircle law, bounds for bordered determinants, and the same moment bound under conditioning at two points with small overlap. Theorem 7.17 is the Kac–Rice formula with bounded Borel weights that may depend on the whole field, and Corollary 7.18 records the resulting exact integrals for expected numbers of critical points. Lemma 7.21 shows that hitting probabilities of the Langevin diffusion are such weights. Finally, Theorem 7.23 combines these results with Hölder's inequality (Lemma 7.22): a bound on a hitting probability under ordinary conditioning becomes a bound on the expected sum of the hitting probabilities over critical points. Corollary 7.24 specializes it to the critical points in a fixed window.

The section uses only the definition of the model (Definition 3.1), the construction of the Langevin dynamics (Definition 3.4 and Lemma 3.5), and the measurability of its path laws (Lemma 4.2). Corollary 7.24 takes as hypotheses a hitting estimate for the caps under critical-point conditioning, which is proved in Section 8, and a bound on expected counts, which is proved in Sections 9 and 11. The results are used as follows: Definition 7.3 and Corollary 7.6 in Section 8; Theorem 7.17 in the criteria of Section 4 for separation and for the concentration of the Gibbs measure on bands; Theorem 7.23 and the argument of Corollary 7.24 in Proposition 13.1; Proposition 7.7, Theorem 7.17, and Corollary 7.18 in the estimates for counts and for the concentration on bands of Sections 9 to 12; and Proposition 7.16 and Lemmas 7.11 and 7.13 in the two-point estimates of Section 11. No asymptotics of expected counts are proved here.

The Kac–Rice formula is classical [3, Theorems 11.2.1 and 12.1.1], [14, Theorems 6.2 and 6.4]. Its use for spherical spin glasses goes back to Auffinger, Ben Arous, and Černý [12] and Auffinger and Ben Arous [8]. The one-point law of a mixed model at a point of a sphere of smaller radius is computed in [30, Section 4.1]. The determinant-

weighted Hölder argument of Lemma 7.22 is a form of the conditional Cauchy–Schwarz argument of Huang and Sellke [90, Corollary 3.16]. Two points require more care than in these references. First, the weights to which we apply the formula are probabilities of trajectory events of the diffusion in the given field. They are bounded Borel functions of the whole field, but they need not be continuous and they are not functions of finitely many derivatives at the point. The particular textbook and spin-glass statements cited above assume that these weights, which are called marks there, are continuous [14, Theorem 6.4], or that they are given by a tame auxiliary field [141, Lemma 36]. General Kac–Rice formulas allowing Borel weights that depend on the whole random field were subsequently established by Stecconi [139, Theorem 29]; see also [116, Proposition 6.1]. We retain a direct proof in coefficient space to specify the conditioning version and the observables used here. Second, the event that the gradient vanishes at a given point has probability zero, and a conditional law given this event is defined only up to a choice of version. We therefore fix an explicit version in Definition 7.3 and prove the formula in the coefficient space of the polynomial field (Proposition 7.16). The determinant bounds are proved directly from Gaussian concentration, and this yields all fixed moments uniformly in the dimension.

7.1 Standing notation

Throughout this section, $\xi(r) = \sum_{p\in\mathcal{P}} c_p r^p$ is a finite mixture as in Definition 3.1, where $\mathcal{P} = \{p \geq 2 : c_p > 0\}$ is finite and nonempty, and $d = \min \mathcal{P}$. Recall from Definition 3.1 that ξ is *pure* if $\mathcal{P} = \{p\}$ for a single degree p, and *non-pure* otherwise. Part II assumes $d \geq 4$; in this section the assumption is used only in Lemma 7.13 and Corollary 7.24. We fix $Q \in (0, 1]$. The mixture and Q do not depend on N. We assume $N \geq 2$ unless stated otherwise and write $n = N - 1$. Constants $C, c > 0$ depend only on the quantities indicated and may change from line to line.

The field and its coefficients. Let $Z = (Z_{p,i_1,\dots,i_p})$, indexed by $p \in \mathcal{P}$ and $i_1, \dots, i_p \in \{1, \dots, N\}$, be a standard Gaussian vector in $\mathbb{R}^{\ell_N}$, where $\ell_N = \sum_{p\in\mathcal{P}} N^p$. For $z \in \mathbb{R}^{\ell_N}$ and $x \in \mathbb{R}^N$, we set

$$H_z(x) = \sum_{p\in\mathcal{P}} \sqrt{c_p}\, N^{-(p-1)/2} \sum_{i_1,\dots,i_p=1}^{N} z_{p,i_1,\dots,i_p}\, x_{i_1} \cdots x_{i_p}. \tag{7.1}$$

Then $\mathbb{E}H_Z(x)H_Z(y) = \sum_p c_p N^{1-p}(x \cdot y)^p = N\xi(x \cdot y/N)$ for all $x, y \in \mathbb{R}^N$. In particular, the restriction of H_Z to S_N has the law of H_N. We take $H_N = H_Z$ and always use the polynomial extension (7.1) of H_N to $\mathbb{R}^N$. We write $H_{N,p}$ for the homogeneous part of degree p of H_N. As in Section 4, let $\mathcal{H}_N$ be the finite-dimensional space of real polynomials on $\mathbb{R}^N$ that are sums of homogeneous polynomials with degrees in $\mathcal{P}$, with the topology of any norm. The map $z \mapsto H_z$ is linear from $\mathbb{R}^{\ell_N}$ into $\mathcal{H}_N$.

A *field-dependent weight* is a Borel function $\varphi\colon \mathcal{H}_N \times S_N \to \mathbb{R}$. Its value $\varphi(H_N, u)$ may depend on the whole field, not only on finitely many derivatives at u. In the literature on marked point processes, such a function is called a mark; we reserve this word for the value and the radial derivative of a critical point, defined below. The function $(z,u) \mapsto \varphi(H_z, u)$ is Borel, because $(z,u) \mapsto (H_z, u)$ is continuous.

The center field. For $u \in \mathbb{R}^N$ we write

$$H_Q(u) = H_N(\sqrt{Q}\,u), \qquad K_Q(u) = \frac{\mathrm{d}}{\mathrm{d}t} H_N(t\sqrt{Q}\,u)\Big|_{t=1} = u \cdot \nabla H_Q(u),$$

and $H_{z,Q}$, $K_{z,Q}$ for the same objects built from H_z. By homogeneity, $H_Q = \sum_p Q^{p/2} H_{N,p}$ and $K_Q = \sum_p p Q^{p/2} H_{N,p}$. For $u \in S_N$ we have $K_Q(u) = \sqrt{QN}\,\partial_R H_N(\sqrt{Q}\,u)$, where ∂_R is the derivative in the unit outward normal direction. In other words, K_Q is the Euler radial derivative. We call $K_Q(u)$ the *radial derivative* of H_Q at u. We normalize the radial derivatives of critical points by N; at a deterministic point, $\operatorname{Var} K_Q(u) = N(a_Q + b_Q)$ by Lemma 7.1. The covariance of H_Q is $\mathbb{E} H_Q(x) H_Q(y) = N\xi_Q(x \cdot y/N)$, where $\xi_Q(t) = \xi(Qt)$. Recall the center covariance parameters

$$\xi_Q(1) = \xi(Q), \qquad a_Q = \xi_Q'(1) = Q\xi'(Q), \qquad b_Q = \xi_Q''(1) = Q^2\xi''(Q),$$

and Σ_Q is the 2×2 matrix with rows $(\xi(Q), a_Q)$ and $(a_Q, a_Q + b_Q)$.

For $u \in S_N$, the tangent space is $u^\perp = \{v \in \mathbb{R}^N : v \cdot u = 0\}$. We write $\nabla_{S_N} H_Q(u) \in u^\perp$ for the spherical gradient, which is the orthogonal projection of $\nabla H_Q(u)$ onto $u^\perp$, and $\nabla^2_{S_N} H_Q(u)$ for the Riemannian Hessian, a symmetric operator on $u^\perp$. We also write $\nabla^2_\perp H_Q(u)$ for the compression of the Euclidean Hessian to $u^\perp$, that is, the operator $v \mapsto P_u \nabla^2 H_Q(u) v$ on $u^\perp$, where $P_u = \mathrm{Id} - uu^{\mathsf{T}}/N$. A critical point of H_Q is a point $u \in S_N$ with $\nabla_{S_N} H_Q(u) = 0$. Its *marks* are the normalized value $h = H_Q(u)/N$ and the normalized radial derivative $k = K_Q(u)/N$. The surface measure of S_N is

$$|S_N| = \frac{2\pi^{N/2}}{\Gamma(N/2)} N^{(N-1)/2},$$

and ν_N is the normalized surface measure.

Random matrices. For $m \geq 1$, a standard GOE matrix G_m is a random real symmetric $m \times m$ matrix whose entries on and above the diagonal are independent centered Gaussians, with variance $1/m$ off the diagonal and $2/m$ on it. We set

$$W_{n,N} = \sqrt{b_Q n/N}\, G_n, \qquad \mu_{n,N}(k) = \mathbb{E}\big|\det(W_{n,N} - k\,\mathrm{Id}_n)\big| \qquad (k \in \mathbb{R}).$$

The entries of $W_{n,N}$ have variance b_Q/N off the diagonal and $2b_Q/N$ on it. Note the normalization by N, not by n. For every symmetric M we have $\operatorname{Tr}(W_{n,N}M) = \sum_i W_{ii}M_{ii} + 2\sum_{i<j} W_{ij}M_{ij}$, so that

$$\operatorname{Var}\operatorname{Tr}(W_{n,N}M) = \frac{2b_Q}{N}\sum_i M_{ii}^2 + \frac{4b_Q}{N}\sum_{i<j} M_{ij}^2 = \frac{2b_Q}{N}\operatorname{Tr}(M^2). \tag{7.2}$$

A symmetric operator on an n-dimensional Euclidean space *has the law of* $W_{n,N}$ if its matrix in one orthonormal basis does. By the orthogonal invariance of the GOE, its matrix in every orthonormal basis then has this law. The matrix $W_{n,N}$ has a density on the space of symmetric matrices, so $\det(W_{n,N} - k\,\mathrm{Id}_n) \neq 0$ almost surely and $\mu_{n,N}(k) > 0$ for every k.

7.2 The one-point law and the conditioned laws

Lemma 7.1 (One-point law)**.** *Let $u \in S_N$. The three Gaussian objects*

$$\big(H_Q(u), K_Q(u)\big), \qquad \nabla_{S_N} H_Q(u), \qquad \nabla^2_\perp H_Q(u)$$

are centered and independent. The pair $(H_Q(u), K_Q(u))$ has covariance matrix $N\Sigma_Q$, the gradient $\nabla_{S_N} H_Q(u)$ is $\sqrt{a_Q}$ times a standard Gaussian vector in $u^\perp$, and $\nabla^2_\perp H_Q(u)$ has the law of $W_{n,N}$. Moreover,

$$\nabla^2_{S_N} H_Q(u) = \nabla^2_\perp H_Q(u) - \frac{K_Q(u)}{N}\,\mathrm{Id}_{u^\perp}. \tag{7.3}$$

If ξ is a non-pure mixture, then $\det \Sigma_Q > 0$. If ξ is pure of degree p, then $K_Q = pH_Q$ on $\mathbb{R}^N$, and $H_Q(u)$, $\nabla_{S_N} H_Q(u)$, and $\nabla^2_\perp H_Q(u)$ are independent, with $\operatorname{Var} H_Q(u) = N\xi(Q)$.

Proof. The coefficients of the polynomial H_Q are linear in Z. Hence its derivatives at fixed points are jointly Gaussian and centered, and their covariances are derivatives of

$$\kappa(x,y) = \mathbb{E} H_Q(x) H_Q(y) = N\xi_Q(t), \qquad t = \frac{x\cdot y}{N}.$$

The function κ is invariant under orthogonal transformations of $\mathbb{R}^N$, hence so is the law of H_Q, and we may assume that $u = \sqrt{N} e_N$. Then $u^\perp$ is spanned by $e_1, \dots, e_n$, and $u_a = 0$ for $a \le n$. In this proof, the indices a, b, i, j range over $\{1, \dots, n\}$, and all derivatives are evaluated at $x = y = u$, where $t = 1$.

Using $\partial_{x_a} t = y_a/N$ and $\partial_{y_a} t = x_a/N$, we have

$$\partial_{x_a}\partial_{y_b}\kappa = \xi_Q''(t)\frac{y_a x_b}{N} + \xi_Q'(t)\delta_{ab} = a_Q\delta_{ab}, \qquad \partial_{x_i}\partial_{x_j}\partial_{y_a}\partial_{y_b}\kappa = \frac{b_Q}{N}\big(\delta_{ia}\delta_{jb} + \delta_{ib}\delta_{ja}\big).$$

In the second identity, every other term of the derivative contains a factor x_c or y_c with $c \le n$, and vanishes. From the first identity, we see that $\nabla_{S_N} H_Q(u)$ has covariance $a_Q\,\mathrm{Id}_{u^\perp}$. From the second, we see that the entries of $\nabla^2_\perp H_Q(u)$ on and above the diagonal are uncorrelated, with variance b_Q/N off the diagonal and $2b_Q/N$ on it, which is the law of $W_{n,N}$.

Since $K_Q(y) = y\cdot\nabla H_Q(y)$ and $y\cdot\nabla_y t = t$, we have $\mathbb{E} H_Q(x) K_Q(y) = y\cdot\nabla_y\kappa = Nf(t)$ with $f(t) = t\xi_Q'(t)$, and $\mathbb{E} K_Q(x) K_Q(y) = x\cdot\nabla_x(Nf(t)) = Ntf'(t)$. At $t = 1$ we obtain

$$\mathbb{E} H_Q(u)^2 = N\xi(Q), \qquad \mathbb{E} H_Q(u) K_Q(u) = Na_Q, \qquad \mathbb{E} K_Q(u)^2 = N(a_Q + b_Q),$$

so $(H_Q(u), K_Q(u))$ has covariance $N\Sigma_Q$. For the cross covariances, we compute

$$\begin{aligned}
&\mathbb{E} H_Q(x)\, \partial_{y_a} H_Q(y) = \xi_Q'(t)\, x_a, \qquad \mathbb{E} K_Q(x)\, \partial_{y_a} H_Q(y) = f'(t)\, x_a,\\
&\mathbb{E} \partial_{x_i}\partial_{x_j} H_Q(x)\, H_Q(y) = \xi_Q''(t)\frac{y_i y_j}{N}, \qquad \mathbb{E} \partial_{x_i}\partial_{x_j} H_Q(x)\, K_Q(y) = f''(t)\frac{y_i y_j}{N},\\
&\mathbb{E} \partial_{x_i}\partial_{x_j} H_Q(x)\, \partial_{y_a} H_Q(y) = \xi_Q'''(t)\frac{y_i y_j x_a}{N^2} + \xi_Q''(t)\frac{\delta_{ia} y_j + \delta_{ja} y_i}{N}.
\end{aligned}$$

All of these vanish at $x = y = u$. Since the three objects are jointly Gaussian with vanishing cross covariances, they are independent.

For (7.3), let $e \in u^\perp$ be a unit vector and let $\gamma(s) = \cos(s/\sqrt{N})\, u + \sqrt{N}\sin(s/\sqrt{N})\, e$ be the unit-speed geodesic of S_N with $\gamma(0) = u$ and $\gamma'(0) = e$. Then $\gamma''(0) = -u/N$, and

$$\begin{aligned}
\nabla^2_{S_N} H_Q(u)(e,e) &= \frac{\mathrm{d}^2}{\mathrm{d}s^2} H_Q(\gamma(s))\Big|_{s=0} = e \cdot \nabla^2 H_Q(u)\, e - \frac{u \cdot \nabla H_Q(u)}{N}\\
&= e \cdot \nabla^2_\perp H_Q(u)\, e - \frac{K_Q(u)}{N}.
\end{aligned}$$

Polarization gives (7.3).

The weights $w_p = c_p Q^p/\xi(Q)$, $p \in \mathcal{P}$, form a probability vector. We have $\xi(Q) = \sum_p c_p Q^p$, $a_Q = \sum_p p c_p Q^p$, and $a_Q + b_Q = \sum_p p^2 c_p Q^p$, and therefore

$$\det \Sigma_Q = \xi(Q)(a_Q + b_Q) - a_Q^2 = \xi(Q)^2 \Big(\sum_p w_p p^2 - \Big(\sum_p w_p p \Big)^2 \Big),$$

which is $\xi(Q)^2$ times the variance of p under the weights w. It is positive exactly when $\mathcal{P}$ has at least two elements. If $\mathcal{P} = \{p\}$, then H_Q is homogeneous of degree p, and Euler's identity implies that $K_Q = pH_Q$. The independence assertion in the pure case is contained in the computation above. □

Remark 7.2 (Comparison with the literature). Lemma 7.1 is the one-point structure of [30, Section 4.1, (4.1)]. There the field lives on a sphere of radius $q\sqrt{N}$ and the radial derivative is the unit normal derivative. With $q = \sqrt{Q}$, their radial variable is $K_Q(u)/\sqrt{QN}$, and their derivatives are taken in the variable $\sqrt{Q}\, u$. Moreover, [30] normalizes the mixture so that its value at 1 is 1. To compare, we set $\alpha = \xi(1)$, $\nu = \xi/\alpha$, and $\hat{H}_N = H_N/\sqrt{\alpha}$. Then values, radial derivatives, gradients, and Hessians of H_N are $\sqrt{\alpha}$ times those of $\hat{H}_N$, and gradients and Hessians in the variable u carry the additional factors $\sqrt{Q}$ and Q from the homothety $u \mapsto \sqrt{Q}\, u$. We have included the computation to fix these normalizations, and it holds for every $\xi(1) > 0$.

We now fix the conditional laws given the value, the radial derivative, and the tangent gradient at a point. For a non-pure mixture and $u \in S_N$, let $\Lambda_u \colon \mathbb{R}^{\ell_N} \to \mathbb{R}^2 \times u^\perp$ be the linear map

$$\Lambda_u z = \big(H_{z,Q}(u),\, K_{z,Q}(u),\, \nabla_{S_N} H_{z,Q}(u) \big),$$

and in the pure case let $\Lambda_u z = (H_{z,Q}(u), \nabla_{S_N} H_{z,Q}(u)) \in \mathbb{R} \times u^\perp$. We write $\mathcal{T}_u$ for the target space of Λ_u and Λ_u^* for the adjoint of Λ_u with respect to the Euclidean inner products. By Lemma 7.1, the covariance operator $\mathcal{C}_u = \Lambda_u \Lambda_u^*$ of $\Lambda_u Z$ is $N\Sigma_Q \oplus a_Q \operatorname{Id}_{u^\perp}$ for a non-pure mixture and $N\xi(Q) \oplus a_Q \operatorname{Id}_{u^\perp}$ in the pure case, and it is invertible on $\mathcal{T}_u$.

Definition 7.3 (Conditioned laws)**.** For $u \in S_N$ and $a \in \mathcal{T}_u$, let

$$Z^{u,a} = \Lambda_u^* \mathcal{C}_u^{-1} a + \big(\operatorname{Id} - \Lambda_u^* \mathcal{C}_u^{-1} \Lambda_u\big) Z. \tag{7.4}$$

For a non-pure mixture and $(h,k) \in \mathbb{R}^2$, we write $\mathbb{P}_{h,k,u}$ and $\mathbb{E}_{h,k,u}$ for probability and expectation under which the field is $H_{Z^{u,a}}$ with $a = (Nh, Nk, 0)$; thus $\mathbb{E}_{h,k,u}[\Phi(H_N)] = \mathbb{E}[\Phi(H_{Z^{u,(Nh,Nk,0)}})]$ for every Borel function $\Phi \geq 0$ on $\mathcal{H}_N$. In the pure case, $\mathbb{P}_{h,u}$ and $\mathbb{E}_{h,u}$ are defined in the same way with $a = (Nh, 0)$, and for $k = ph$ we also write $\mathbb{E}_{h,k,u} = \mathbb{E}_{h,u}$. We call these the *conditioned laws* at u with marks (h,k). Similarly, with $\Gamma_u z = \nabla_{S_N} H_{z,Q}(u)$, we write $\mathbb{E}_u^\nabla$ for expectation under which the field is $H_{Z_0^u}$, where $Z_0^u = (\operatorname{Id} - a_Q^{-1} \Gamma_u^* \Gamma_u) Z$.

Lemma 7.4 (Properties of the conditioned laws)**.** *Let $u \in S_N$.*

1. *For every $a \in \mathcal{T}_u$, we have $\Lambda_u Z^{u,a} = a$. The vector $(\operatorname{Id} - \Lambda_u^* \mathcal{C}_u^{-1} \Lambda_u) Z$ is independent of $\Lambda_u Z$, and $Z = Z^{u, \Lambda_u Z}$. In particular, $a \mapsto \operatorname{Law}(Z^{u,a})$ is a regular version of the conditional law of Z given $\Lambda_u Z = a$, and it is continuous in a. In the same way, $\operatorname{Law}(Z_0^u)$ is a version of the conditional law of Z given $\nabla_{S_N} H_Q(u) = 0$.*
2. *Under $\mathbb{P}_{h,k,u}$, we have $H_Q(u) = Nh$, $K_Q(u) = Nk$, and $\nabla_{S_N} H_Q(u) = 0$ surely, and $\nabla^2_{S_N} H_Q(u)$ has the law of $W_{n,N} - k \operatorname{Id}_n$; in the pure case this holds with $k = ph$. In particular, $\mathbb{E}_{h,k,u} |\det \nabla^2_{S_N} H_Q(u)| = \mu_{n,N}(k) > 0$.*
3. *For a non-pure mixture and every field-dependent weight $\varphi \geq 0$, the function $(u,h,k) \mapsto \mathbb{E}_{h,k,u} \varphi(H_N, u)$ is Borel on $S_N \times \mathbb{R}^2$. In the pure case the function $(u,h) \mapsto \mathbb{E}_{h,u} \varphi(H_N, u)$ is Borel on $S_N \times \mathbb{R}$; equivalently, $(u,h,k) \mapsto \mathbb{E}_{h,k,u} \varphi(H_N, u)$ is Borel on $S_N \times \{(h,k) : k = ph\}$. In both cases, $u \mapsto \mathbb{E}_u^\nabla \varphi(H_N, u)$ is Borel on S_N.*
4. *For a non-pure mixture and every field-dependent weight $\varphi \geq 0$,*

$$\mathbb{E}_u^\nabla \varphi(H_N, u) = \int_{\mathbb{R}^2} p_N(h,k)\, \mathbb{E}_{h,k,u} \varphi(H_N, u)\, \mathrm{d}h\, \mathrm{d}k, \tag{7.5}$$

where $p_N(h,k) = N(2\pi)^{-1} (\det \Sigma_Q)^{-1/2} \exp(-N(h,k)\Sigma_Q^{-1}(h,k)^\mathsf{T}/2)$ is the density of the centered Gaussian law with covariance Σ_Q/N, and in the pure case the same identity holds with $\int_{\mathbb{R}} p_N^{\mathrm{pure}}(h)\, \mathbb{E}_{h,u} \varphi(H_N, u)\, \mathrm{d}h$ on the right, where $p_N^{\mathrm{pure}}(h) = (N/(2\pi\xi(Q)))^{1/2} e^{-Nh^2/(2\xi(Q))}$.

Proof. For (1), write $\Pi_u = \Lambda_u^* \mathcal{C}_u^{-1} \Lambda_u$. We have $\Lambda_u \Pi_u = \Lambda_u$ because $\Lambda_u \Lambda_u^* = \mathcal{C}_u$, which implies that $\Lambda_u Z^{u,a} = a$. The covariance between $(\operatorname{Id} - \Pi_u) Z$ and $\Lambda_u Z$ is $(\operatorname{Id} - \Pi_u)\Lambda_u^* = \Lambda_u^* - \Lambda_u^* \mathcal{C}_u^{-1} \mathcal{C}_u = 0$, and jointly Gaussian vectors with zero covariance are independent. The identity $Z = \Lambda_u^* \mathcal{C}_u^{-1} \Lambda_u Z + (\operatorname{Id} - \Pi_u) Z$ is $Z = Z^{u, \Lambda_u Z}$. These facts

imply that $\mathrm{Law}(Z^{u,a})$ is a regular version of the conditional law, and it is continuous in a because a enters (7.4) linearly. The same argument applies to Γ_u, whose covariance is $a_Q \,\mathrm{Id}_{u^\perp}$.

For (2), the first assertion follows from $\Lambda_u Z^{u,a} = a$. The operator $\nabla^2_\perp H_{z,Q}(u)$ is linear in z, and by Lemma 7.1 it is uncorrelated with $\Lambda_u Z$. In other words, $\nabla^2_\perp H_{\Lambda_u^* b,Q}(u) = 0$ for every $b \in \mathcal{T}_u$. Applying this with $b = \mathcal{C}_u^{-1} a$ and $b = \mathcal{C}_u^{-1}\Lambda_u Z$, we find that $\nabla^2_\perp H_{Z^{u,a},Q}(u) = \nabla^2_\perp H_{Z,Q}(u)$, which has the law of $W_{n,N}$. Combining this with (7.3) and $K_Q(u) = Nk$, we obtain the law of the intrinsic Hessian, and the last assertion follows from the definition of $\mu_{n,N}$.

For (3), regard Λ_u as a map into $\mathbb{R}^2 \times \mathbb{R}^N$, which depends smoothly on $u \in \mathbb{R}^N$. On $\mathbb{R}^2 \times \mathbb{R}^N$, the operator $\mathcal{C}_u^+ = (N\Sigma_Q)^{-1} \oplus a_Q^{-1} P_u$ is smooth in u and agrees with $\mathcal{C}_u^{-1}$ on $\mathcal{T}_u$. Then the map $(z,u,h,k) \mapsto Z^{u,(Nh,Nk,0)}$, computed from z in place of Z, is continuous, and $(z,u,h,k) \mapsto \varphi(H_{Z^{u,(Nh,Nk,0)}}, u)$ is Borel. Integrating out z against the standard Gaussian measure, we find by Tonelli's theorem that $(u,h,k) \mapsto \mathbb{E}_{h,k,u}\varphi(H_N, u)$ is Borel. The pure case and $\mathbb{E}_u^\nabla$ are identical.

For (4), write $\Lambda_u z = (\Lambda'_u z, \Gamma_u z)$ with $\Lambda'_u z = (H_{z,Q}(u), K_{z,Q}(u))$. We have $\Lambda'_u \Gamma_u^* = 0$ by Lemma 7.1. Then $\mathcal{C}_u^{-1}$ is block diagonal, $\Pi_u = \Lambda_u'^* (N\Sigma_Q)^{-1}\Lambda'_u + a_Q^{-1}\Gamma_u^*\Gamma_u$, and we have

$$Z_0^u = \Lambda_u'^* (N\Sigma_Q)^{-1}\Lambda'_u Z + (\mathrm{Id} - \Pi_u)Z = Z^{u,(\Lambda'_u Z,0)}.$$

By (1), $\Lambda'_u Z$ is independent of $(\mathrm{Id} - \Pi_u)Z$, and it has the law $N(0, N\Sigma_Q)$. Conditioning on $\Lambda'_u Z = (Nh, Nk)$ and changing variables, the density of (h,k) is p_N, since Σ_Q/N has determinant $\det \Sigma_Q / N^2$. This proves (7.5). The pure case is the same with $\Lambda'_u z = H_{z,Q}(u)$. □

The conditioned laws depend on no choice of frame on S_N, and they are defined for every value of the marks. Any regular version of the conditional law of Z given $\Lambda_u Z$ agrees with $\mathrm{Law}(Z^{u,a})$ only for almost every a. This says nothing at the single target value 0 of the gradient, which is the value that enters the Kac–Rice formula. All conditioned expectations in Part II refer to the laws of Definition 7.3. Since $z \mapsto H_z(x)$ is linear, (7.4) says that the field under $\mathbb{P}_{h,k,u}$ is

$$H_N(x) + \mathrm{Cov}\big(H_N(x), \Lambda_u Z\big)\, \mathcal{C}_u^{-1}\big(a - \Lambda_u Z\big), \qquad a = (Nh, Nk, 0),$$

which is the explicit Gaussian regression used in Definition 8.1. The conditioned hitting estimate of Section 8 is therefore stated under the same laws.

7.3 Gaussian concentration and determinant moments

We first record two Gaussian estimates. Part (1) of the next lemma is the Gaussian concentration inequality in the form [111, Proposition 4.1 and (4.1)], which also underlies

Lemma 3.3; we include its proof through Gross's logarithmic Sobolev inequality [84] and Herbst's argument (see [17, 112]); part (2) bounds the extreme eigenvalues of a Gaussian matrix whose covariance is dominated by that of the GOE. We give the short proofs because we need dimension-free constants and degenerate covariances.

Lemma 7.5 (Gaussian concentration and an edge bound)**.** *1. Let Y be a standard Gaussian vector in $\mathbb{R}^m$ and let $f\colon \mathbb{R}^m \to \mathbb{R}$ be L-Lipschitz. Then for every $t \in \mathbb{R}$,*

$$\mathbb{E}\exp\big(t(f(Y)-\mathbb{E}f(Y))\big) \le \exp(t^2L^2/2), \tag{7.6}$$

and, if $L>0$, $\mathbb{P}(f(Y)\ge \mathbb{E}f(Y)+s)\le \exp(-s^2/(2L^2))$ for every $s>0$. If $L=0$, then f is constant and this probability is zero.

2. Let $b>0$ and $1\le n\le N$, and let W be a centered Gaussian random symmetric $n\times n$ matrix such that $\operatorname{Var}\operatorname{Tr}(WM)\le (2b/N)\operatorname{Tr}(M^2)$ for every symmetric M. Then

$$\mathbb{E}\lambda_{\max}(W)\le 2\sqrt{bn/N}, \qquad \mathbb{P}\big(\,\|W\|_{\mathrm{op}} > 2\sqrt{b}+s\big)\le 2\exp\big(-Ns^2/(4b)\big) \quad (s>0). \tag{7.7}$$

The covariance of W may be degenerate.

Proof. We begin with (1). Let γ be the standard Gaussian measure on $\mathbb{R}^m$ and let $Q_sg(y)=\mathbb{E}g(e^{-s}y+\sqrt{1-e^{-2s}}\,Y)$ be the Mehler semigroup, with generator $\mathcal{L}=\Delta-y\cdot\nabla$. We use three standard facts: the invariance of γ under Q_s, the Gaussian integration by parts formula $\int(\mathcal{L}\psi)\chi\,\mathrm{d}\gamma=-\int\nabla\psi\cdot\nabla\chi\,\mathrm{d}\gamma$, and the relation $\nabla Q_sg=e^{-s}Q_s\nabla g$. Let g be smooth and bounded, with bounded derivatives and $\inf g>0$. Then

$$\frac{\mathrm{d}}{\mathrm{d}s}\int Q_sg\log Q_sg\,\mathrm{d}\gamma=\int(\mathcal{L}Q_sg)(1+\log Q_sg)\,\mathrm{d}\gamma=-\int\frac{|\nabla Q_sg|^2}{Q_sg}\,\mathrm{d}\gamma.$$

Since $Q_sg\to\int g\,\mathrm{d}\gamma$ as $s\to\infty$, integrating in s gives

$$\int g\log g\,\mathrm{d}\gamma-\int g\,\mathrm{d}\gamma\,\log\int g\,\mathrm{d}\gamma=\int_0^\infty\int\frac{|\nabla Q_sg|^2}{Q_sg}\,\mathrm{d}\gamma\,\mathrm{d}s.$$

By the Cauchy–Schwarz inequality, $|Q_s\nabla g|^2\le Q_s(|\nabla g|^2/g)\,Q_sg$. Together with $\nabla Q_sg=e^{-s}Q_s\nabla g$, this implies that $|\nabla Q_sg|^2/Q_sg\le e^{-2s}Q_s(|\nabla g|^2/g)$. Inserting this and using the invariance of γ, we obtain the logarithmic Sobolev inequality

$$\int g\log g\,\mathrm{d}\gamma-\int g\,\mathrm{d}\gamma\,\log\int g\,\mathrm{d}\gamma\le\frac12\int\frac{|\nabla g|^2}{g}\,\mathrm{d}\gamma.$$

Now let f be smooth and bounded with $|\nabla f|\le L$, and set $\psi(t)=\log\mathbb{E}e^{tf(Y)}$. Applying the last inequality to $g=e^{tf}$ with $t>0$, we obtain $t\psi'(t)-\psi(t)\le t^2L^2/2$, which means that $(\psi(t)/t)'\le L^2/2$. As $\psi(t)/t\to\mathbb{E}f(Y)$ when $t\downarrow 0$, we have $\psi(t)\le t\mathbb{E}f(Y)+$

$t^2L^2/2$, which is (7.6). Applying this to $-f$ covers $t<0$. A general L-Lipschitz f is the pointwise limit of smooth bounded functions with $|\nabla f|\le L$, obtained by truncation and mollification. We extend (7.6) to f by dominated convergence, using $|f(y)|\le |f(0)|+L|y|$ and the integrability of $e^{c|Y|}$. The tail bound follows from Markov's inequality with $t=s/L^2$.

For (2), let Γ be the covariance operator of W on the space of symmetric matrices with the inner product $\operatorname{Tr}(AB)$. Then W has the law of $\Gamma^{1/2}\zeta$, where ζ is a standard Gaussian vector in that space, and the hypothesis says that $\|\Gamma^{1/2}\|_{\mathrm{op}}\le\sqrt{2b/N}$. For unit vectors $v,w\in\mathbb{R}^n$, let $X_v=v\cdot Wv=\operatorname{Tr}(Wvv^{\mathsf T})$ and $Y_v=2\sqrt{b/N}\,g\cdot v$, where g is a standard Gaussian vector in $\mathbb{R}^n$. Then

$$\mathbb{E}(X_v-X_w)^2\le\frac{2b}{N}\left\|vv^{\mathsf T}-ww^{\mathsf T}\right\|_F^2=\frac{4b}{N}\bigl(1-(v\cdot w)^2\bigr)\le\frac{8b}{N}(1-v\cdot w)=\mathbb{E}(Y_v-Y_w)^2,$$

where we used $1-(v\cdot w)^2=(1-v\cdot w)(1+v\cdot w)$ and $|v\cdot w|\le 1$. Both processes are continuous on the unit sphere. By the Sudakov–Fernique inequality [3, Theorem 2.2.3], we have $\mathbb{E}\lambda_{\max}(W)=\mathbb{E}\sup_v X_v\le\mathbb{E}\sup_v Y_v=2\sqrt{b/N}\,\mathbb{E}|g|\le 2\sqrt{bn/N}$. The function $\zeta\mapsto\lambda_{\max}(\Gamma^{1/2}\zeta)$ is $\sqrt{2b/N}$-Lipschitz, since $\lambda_{\max}$ is 1-Lipschitz for the Frobenius norm. By part (1) and $\mathbb{E}\lambda_{\max}(W)\le 2\sqrt{b}$, we have $\mathbb{P}(\lambda_{\max}(W)>2\sqrt{b}+s)\le\exp(-Ns^2/(4b))$. The matrix $-W$ satisfies the same hypothesis, and a union bound gives (7.7). □

Combining Lemma 7.4(2) with Lemma 7.5, we record the law of the Hessian at a conditioned critical point and the meaning of a strict gap.

Corollary 7.6 (The Hessian at a conditioned critical point)**.** *Let $u\in S_N$ and $(h,k)\in\mathbb{R}^2$, with $k=ph$ in the pure case. Under $\mathbb{P}_{h,k,u}$, the matrix of $\nabla^2_{S_N}H_Q(u)$ in any orthonormal basis of $u^\perp$ has the law of*

$$\sqrt{b_Q(N-1)/N}\;G_{N-1}-k\,\mathrm{Id}_{N-1}=W_{n,N}-k\,\mathrm{Id}_n.$$

In other words, its entries on and above the diagonal are independent Gaussians with mean $-k\delta_{ij}$, variance b_Q/N off the diagonal, and variance $2b_Q/N$ on it. If $k\ge 2\sqrt{b_Q}+g$ for some $g>0$, then

$$\mathbb{P}_{h,k,u}\Bigl(\nabla^2_{S_N}H_Q(u)\le-\frac{g}{2}\,\mathrm{Id}_{u^\perp}\Bigr)\ge 1-2e^{-Ng^2/(16b_Q)}.$$

Proof. The first assertion is Lemma 7.4(2); it is the law of Lemma 5.14(1), stated for the conditioned laws of Definition 7.3. The shift $-k\,\mathrm{Id}$ comes from (7.3) and $K_Q(u)=Nk$, and the variances are those of Lemma 7.1. By (7.2), $W_{n,N}$ satisfies the hypothesis of Lemma 7.5(2) with $b=b_Q$, and with $s=g/2$ we obtain $\mathbb{P}(\|W_{n,N}\|_{\mathrm{op}}>2\sqrt{b_Q}+g/2)\le 2e^{-Ng^2/(16b_Q)}$. On the complementary event, $W_{n,N}-k\,\mathrm{Id}\le(2\sqrt{b_Q}+g/2-k)\mathrm{Id}\le-(g/2)\mathrm{Id}$. □

We say that a set of marks has a *strict gap* if $k \geq 2\sqrt{b_Q} + g$ on it for a fixed $g > 0$. Corollary 7.6 implies that the conditioned critical points with such marks are nondegenerate local maxima with probability $1 - O(e^{-cN})$. The conditioned hitting estimate of Section 8 does not use this gap, while the determinant bounds below do.

Proposition 7.7 (Normalized determinant moments)**.** *Let $b, g > 0$ and $k_+ < \infty$. Let $1 \leq n \leq N$, let W be a centered Gaussian random symmetric $n \times n$ matrix such that $\operatorname{Var}\operatorname{Tr}(WM) \leq (2b/N)\operatorname{Tr}(M^2)$ for every symmetric M, let C be a deterministic symmetric $n \times n$ matrix with $\|C\|_{\mathrm{op}} \leq g/4$, and let $2\sqrt{b} + g \leq k \leq k_+$. Set*

$$D = \bigl|\det(k\,\mathrm{Id}_n - C - W)\bigr|, \qquad \mu = \mathbb{E}D, \qquad Y = \sum_{j=1}^{n} \log\max\bigl\{|k - \lambda_j(C+W)|,\, g/2\bigr\}.$$

Then $\mu > 0$, and for every real $r \geq 1$ there exists $\mathsf{K}_r = \mathsf{K}_r(b, g, k_+) < \infty$ such that

$$\mathbb{E}(D/\mu)^r \leq \mathbb{E}(e^Y/\mu)^r \leq \mathsf{K}_r. \tag{7.8}$$

The constant K_r does not depend on n, N, k, C, or the covariance of W. There exists $N_0(b, g)$ such that for $N \geq N_0$ one may take $\mathsf{K}_r = 2^r \exp(4r^2 b/g^2)$. In particular, if $u \in S_N$ and $2\sqrt{b_Q} + g \leq k \leq k_+$, then under $\mathbb{P}_{h,k,u}$ the determinant $D = |\det \nabla^2_{S_N} H_Q(u)|$ satisfies $\mathbb{E}_{h,k,u}(D/\mu_{n,N}(k))^r \leq \mathsf{K}_r(b_Q, g, k_+)$; in the pure case this holds for $k = ph$.

For $0 < r < 1$, we have $\mathbb{E}(D/\mu)^r \leq 1$ by Jensen's inequality. The proposition asserts nothing about the dependence between D and other functions of the field, and none is needed below. Both hypotheses on C and W matter: if C were allowed to be random, or W to have a nonzero mean, the determinant could be close to zero with probability close to one.

Proof. The scalar function $x \mapsto \log\max\{|k - x|, g/2\}$ is $(2/g)$-Lipschitz on $\mathbb{R}$. By the Hoffman–Wielandt inequality [89], the ordered eigenvalues of symmetric matrices U, V satisfy $\sum_j (\lambda_j(U) - \lambda_j(V))^2 \leq \|U - V\|_F^2$. (To verify it, expand $\|U - V\|_F^2 = \operatorname{Tr}U^2 + \operatorname{Tr}V^2 - 2\operatorname{Tr}(UV)$ and diagonalize: $\operatorname{Tr}(UV) = \sum_{i,j} \lambda_i(U)\lambda_j(V)|e_i \cdot f_j|^2$ for orthonormal eigenbases (e_i), (f_j); the matrix $(|e_i \cdot f_j|^2)$ is doubly stochastic, and by Birkhoff's theorem and the rearrangement inequality this sum is at most $\sum_j \lambda_j(U)\lambda_j(V)$.) By the Cauchy–Schwarz inequality, Y is a $(2\sqrt{n}/g)$-Lipschitz function of W in the Frobenius norm. Writing $W = \Gamma^{1/2}\zeta$ as in the proof of Lemma 7.5, we find that Y is a Lipschitz function of ζ with constant

$$L = \frac{2}{g}\sqrt{\frac{2bn}{N}} \leq \frac{2\sqrt{2b}}{g}. \tag{7.9}$$

We note that this concerns the clipped function Y, which is defined and Lipschitz on all matrices.

Let $\mathcal{E} = \{\|W\|_{\mathrm{op}} \le 2\sqrt{b} + g/4\}$. We have $\mathbb{P}(\mathcal{E}^c) \le 2e^{-Ng^2/(64b)}$ by (7.7). On $\mathcal{E}$, every eigenvalue of $C + W$ is at most $2\sqrt{b} + g/2 \le k - g/2$. This implies that $D = e^Y$ on $\mathcal{E}$. Further, $D \le e^Y$ everywhere, since $\max\{|k - x|, g/2\} \ge |k - x|$. By (7.6) and the Cauchy–Schwarz inequality,

$$\mathbb{E}e^{tY} \le e^{t\mathbb{E}Y + t^2L^2/2} \quad (t \ge 0), \qquad \mathbb{E}\big[e^Y \mathbf{1}_{\mathcal{E}^c}\big] \le \big(\mathbb{E}e^{2Y}\big)^{1/2} \mathbb{P}(\mathcal{E}^c)^{1/2} \le e^{\mathbb{E}Y + L^2} \mathbb{P}(\mathcal{E}^c)^{1/2}.$$

Since $\mathbb{E}e^Y \ge e^{\mathbb{E}Y}$ by Jensen's inequality, we obtain

$$\mu \ge \mathbb{E}\big[e^Y \mathbf{1}_{\mathcal{E}}\big] \ge e^{\mathbb{E}Y}\Big(1 - e^{L^2}\mathbb{P}(\mathcal{E}^c)^{1/2}\Big). \tag{7.10}$$

By (7.9), there exists $N_0(b, g)$ such that the parenthesis is at least $1/2$ for $N \ge N_0$. For such N, we have $\mu^r \ge 2^{-r} e^{r\mathbb{E}Y}$, and dividing the bound on $\mathbb{E}e^{rY}$ by μ^r yields $\mathbb{E}(e^Y/\mu)^r \le 2^r e^{r^2L^2/2} \le 2^r \exp(4r^2 b/g^2)$, where we used $L^2 \le 8b/g^2$. As $D \le e^Y$, this proves (7.8) for $N \ge N_0$.

It remains to treat the finitely many pairs (n, N) with $n \le N < N_0$. For each such pair, the admissible covariance operators Γ (those with $0 \le \Gamma \le (2b/N)\mathrm{Id}$), the admissible C, and the admissible k form a compact set. The quantities $\mathbb{E}D^r$, $\mathbb{E}e^{rY}$, and μ are continuous on this set: $\Gamma \mapsto \Gamma^{1/2}$ is continuous, and D^r and e^{rY} are bounded by a polynomial in $|\zeta|$ uniformly on the set, and dominated convergence applies. Moreover $\mu > 0$ on the set: at $\zeta = 0$ the matrix $k\,\mathrm{Id} - C \ge (k - g/4)\mathrm{Id}$ is positive definite, and $D > 0$ on a neighborhood of 0 in the support of $\zeta \mapsto \Gamma^{1/2}\zeta$, which has positive probability. By continuity and compactness, $\mathbb{E}e^{rY}/\mu^r$ is bounded on each of these finitely many sets. This proves (7.8) for all N, and also $\mu > 0$.

For the last assertion, recall from Lemma 7.4(2) that under $\mathbb{P}_{h,k,u}$ the matrix of $-\nabla^2_{S_N} H_Q(u)$ in an orthonormal basis of $u^\perp$ has the law of $k\,\mathrm{Id}_n - W_{n,N}$, and $-W_{n,N}$ has the law of $W_{n,N}$. Since $W_{n,N}$ satisfies the hypothesis with $b = b_Q$ by (7.2), we take $C = 0$. □

Corollary 7.8 (Removal of exceptional events)**.** *Under the hypotheses of Proposition 7.7, let $\mathcal{A}$ be an event on the same probability space; it may depend on W and on any other randomness. Then for every $r > 1$ and $s \ge 1$,*

$$\frac{\mathbb{E}[D\mathbf{1}_{\mathcal{A}}]}{\mu} \le \mathsf{K}_r^{1/r}\mathbb{P}(\mathcal{A})^{1-1/r}, \qquad \frac{\mathbb{E}[D^s\mathbf{1}_{\mathcal{A}}]}{\mu^s} \le \mathsf{K}_{2s}^{1/2}\mathbb{P}(\mathcal{A})^{1/2}. \tag{7.11}$$

If moreover $\mathsf{K}_r^{1/r}\mathbb{P}(\mathcal{A})^{1-1/r} \le 1/2$ for some $r > 1$, then $\mu_{\mathrm{good}} = \mathbb{E}[D\mathbf{1}_{\mathcal{A}^c}] \ge \mu/2$ and $\mathbb{E}(D\mathbf{1}_{\mathcal{A}^c}/\mu_{\mathrm{good}})^s \le 2^s\mathsf{K}_s$ for every $s \ge 1$.

Proof. The first bound in (7.11) is Hölder's inequality with exponents r and $r/(r-1)$ together with (7.8), and the second is the Cauchy–Schwarz inequality applied to $D^s\mathbf{1}_{\mathcal{A}}$. Under the additional hypothesis, the first bound implies that $\mu_{\mathrm{good}} = \mu - \mathbb{E}[D\mathbf{1}_{\mathcal{A}}] \ge \mu/2$, and then $\mathbb{E}(D\mathbf{1}_{\mathcal{A}^c}/\mu_{\mathrm{good}})^s \le 2^s\mathbb{E}(D/\mu)^s \le 2^s\mathsf{K}_s$. □

We note that if $\mathbb{P}(\mathcal{A}) \le C_0 e^{-c_0 N}$, then discarding $\mathcal{A}$ changes the expected determinant by a relative amount that is exponentially small, with constants depending only on b, g, k_+, C_0, c_0. In applications $\mathcal{A}$ is a conditional event, such as the failure of a Hessian or derivative bound, and $\mathbb{P}(\mathcal{A}) \le C_0 e^{-c_0 N}$ must hold under the same conditioned law. An unconditional bound on $\mathbb{P}(\mathcal{A})$ is not enough, and a bound $\mathbb{P}(\mathcal{A}) \to 0$ without a rate bounds the relative loss only by $\mathbb{P}(\mathcal{A})^{1-1/r}$. The corollary transfers such a conditional bound once it is established, but it does not prove it.

7.4 The semicircle law and bordered determinants

Proposition 7.7 needs no information about the size of μ. The next proposition identifies μ up to bounded factors in terms of the semicircle law. It also makes explicit the comparison between expected determinants in dimensions that differ by a fixed integer. Let

$$\sigma_{\mathrm{sc}}(\,\mathrm{d}x) = \frac{1}{2\pi}\sqrt{4-x^2}\,\mathbf{1}_{[-2,2]}(x)\,\mathrm{d}x, \qquad \Omega(z) = \int \log(z-x)\,\sigma_{\mathrm{sc}}(\,\mathrm{d}x),$$

and

$$g_{\mathrm{sc}}(z) = \int \frac{\sigma_{\mathrm{sc}}(\,\mathrm{d}x)}{z-x} = \frac{z-\sqrt{z^2-4}}{2}$$

for $z > 2$, so that $\Omega' = g_{\mathrm{sc}}$. (The formula for g_{sc} follows by setting $x = 2\cos\theta$ and using $\int_0^\pi (z - 2\cos\theta)^{-1}\,\mathrm{d}\theta = \pi/\sqrt{z^2-4}$.)

Proposition 7.9 (Determinants and the semicircle law)**.** *Let $d_0 > 0$ and $Z_+ < \infty$. For every $m \ge 1$ and $2 + d_0 \le z \le Z_+$,*

$$e^{m\Omega(z)} \le \mathbb{E}|\det(z\,\mathrm{Id}_m - G_m)| \le C\,e^{m\Omega(z)}, \qquad \mathbb{E}|\det(z\,\mathrm{Id}_m - G_m)|^r \le C_r\,e^{rm\Omega(z)} \quad (r \ge 1). \tag{7.12}$$

The same bound as the second holds for $\mathbb{E}e^{rY_z}$, where $Y_z = \sum_j \log\max\{|z-\lambda_j(G_m)|, d_0/2\}$. The constants depend only on d_0, Z_+, and r.

Proof. The proof has four steps: we compute the signed expectation $P_m(z) = \mathbb{E}\det(z\,\mathrm{Id}_m - G_m)$, locate its roots, compare it with $e^{m\Omega(z)}$ through a Riccati equation, and finally pass to absolute moments by concentration.

The expected characteristic polynomial. We expand the determinant as a sum over permutations σ of $\{1,\dots,m\}$. The term of σ is $\operatorname{sgn}(\sigma)\prod_i (z\delta_{i\sigma(i)} - (G_m)_{i\sigma(i)})$. A diagonal entry occurs at most once in each term, is centered, and is independent of the other factors, and it contributes only through z. An off-diagonal entry $(G_m)_{ab}$ occurs at most twice, and exactly twice when σ exchanges a and b. If σ has a cycle of length at least three, then some off-diagonal entry occurs exactly once and the term has mean zero. Only

involutions contribute. An involution with j transpositions has sign $(-1)^j$, each transposition contributes $\mathbb{E}(G_m)_{ab}^2 = 1/m$, and there are $m!/(2^j j!(m-2j)!)$ such involutions. We conclude that

$$P_m(z) = \sum_{j=0}^{\lfloor m/2 \rfloor} \frac{(-1)^j m!}{2^j j!(m-2j)!\, m^j} z^{m-2j} = m^{-m/2} \operatorname{He}_m(\sqrt{m}\, z), \tag{7.13}$$

where He_m is the monic probabilists' Hermite polynomial.

Location of the roots. The recurrence $\operatorname{He}_{j+1}(x) = x \operatorname{He}_j(x) - j \operatorname{He}_{j-1}(x)$ coincides with the recurrence obtained by expanding the characteristic polynomial of the tridiagonal matrix with zero diagonal and off-diagonal entries $\sqrt{1}, \dots, \sqrt{m-1}$ along its last row. It follows that the roots $\lambda_1, \dots, \lambda_m$ of P_m are the eigenvalues of the tridiagonal matrix J_m with zero diagonal and off-diagonal entries $\sqrt{j/m}$, $1 \le j \le m-1$. Each absolute row sum of the symmetric matrix J_m is at most $2\sqrt{(m-1)/m} < 2$, and by Gershgorin's theorem all roots are real and lie in $(-2, 2)$. In particular $P_m(z) > 0$ for $z \ge 2$.

Comparison with the semicircle law. The Hermite equation $\operatorname{He}_m'' - x \operatorname{He}_m' + m \operatorname{He}_m = 0$ and (7.13) imply that $P_m'' - mzP_m' + m^2 P_m = 0$. For $z > 2$, set $g_m = P_m'/(mP_m) = m^{-1} \sum_j (z - \lambda_j)^{-1}$. Dividing the differential equation by $m^2 P_m$, we find

$$\frac{g_m'}{m} + g_m^2 - z g_m + 1 = 0, \qquad g_m' = -\frac{1}{m} \sum_j (z - \lambda_j)^{-2} < 0. \tag{7.14}$$

The roots of the quadratic $q(y) = y^2 - zy + 1$ are $g_{\mathrm{sc}}(z)$ and $g_+(z) = (z + \sqrt{z^2 - 4})/2$, and (7.14) implies that $q(g_m) = -g_m'/m > 0$, that is, $g_m(z)$ lies outside $[g_{\mathrm{sc}}(z), g_+(z)]$ for every $z > 2$. The sets where $g_m < g_{\mathrm{sc}}$ and where $g_m > g_+$ are open, disjoint, and cover the connected set $(2, \infty)$. Since $g_m(z) \to 0$ and $g_+(z) \to \infty$ as $z \to \infty$, the first set is all of $(2, \infty)$. Using $q(g_m) = (g_{\mathrm{sc}} - g_m)(g_+ - g_m)$, $g_+ - g_m \ge g_+ - g_{\mathrm{sc}} = \sqrt{z^2 - 4}$, and $z - \lambda_j > z - 2$, we obtain

$$0 \le g_{\mathrm{sc}}(z) - g_m(z) = \frac{q(g_m)}{g_+ - g_m} \le \frac{1}{m(z-2)^2 \sqrt{z^2 - 4}}.$$

Since $(\log P_m)' = m g_m$ and $(m\Omega)' = m g_{\mathrm{sc}}$, and since $\log P_m(z)$ and $m\Omega(z)$ both equal $m \log z + O(z^{-2})$ as $z \to \infty$ for fixed m, integrating from z to ∞ yields

$$0 \le \log P_m(z) - m\Omega(z) \le \int_z^\infty \frac{\mathrm{d}t}{(t-2)^2 \sqrt{t^2 - 4}} \le C_{d_0} \qquad (z \ge 2 + d_0). \tag{7.15}$$

Absolute moments. Let $\mathcal{E}_m = \{\|G_m\|_{\mathrm{op}} \le 2 + d_0/2\}$. By Lemma 7.5(2) with $N = n = m$ and $b = 1$, we have $\mathbb{P}(\mathcal{E}_m^c) \le 2e^{-m d_0^2/16}$. On $\mathcal{E}_m$, the signed determinant is positive

and equals e^{Y_z}, and its absolute value is at most e^{Y_z} everywhere. As in the proof of Proposition 7.7, Y_z is a Lipschitz function of the Gaussian coordinates with constant at most $2\sqrt{2}/d_0$. Using the Cauchy–Schwarz inequality in the numerator and Jensen's inequality in the denominator, we find

$$\frac{\mathbb{E}[e^{Y_z}\mathbf{1}_{\mathcal{E}_m^c}]}{\mathbb{E}e^{Y_z}} \le \exp(8/d_0^2)\mathbb{P}(\mathcal{E}_m^c)^{1/2}.$$

We have $P_m \ge \mathbb{E}e^{Y_z} - 2\mathbb{E}[e^{Y_z}\mathbf{1}_{\mathcal{E}_m^c}]$, and therefore $\mathbb{E}e^{Y_z} \le 2P_m(z)$ once $m \ge m_0(d_0)$. Then (7.6) gives $\mathbb{E}e^{rY_z} \le e^{4r^2/d_0^2}(\mathbb{E}e^{Y_z})^r$. Combining these bounds with (7.15) and $|\det| \le e^{Y_z}$ proves the upper bounds in (7.12) for $m \ge m_0$, and the lower bound follows from $\mathbb{E}|\det(z\,\mathrm{Id}_m - G_m)| \ge P_m(z) \ge e^{m\Omega(z)}$. For the finitely many $m < m_0$, the moments are continuous in z on the compact interval $[2+d_0, Z_+]$ and $e^{m\Omega(z)} \ge d_0^m$, and they are absorbed in the constants. □

Corollary 7.10 (Fixed changes of dimension)**.** *Let $b, g > 0$, $k_+ < \infty$, and $2\sqrt{b} + g \le k \le k_+$. Let $m = N - j$, where j ranges over a fixed finite set of nonnegative integers and $m \ge 1$. Set $b_m = bm/N$ and*

$$\Lambda_{m,N}(k) = m\Big(\frac{1}{2}\log b_m + \Omega\big(k/\sqrt{b_m}\big)\Big).$$

Then the first and every fixed higher moment of $|\det(k\,\mathrm{Id}_m - \sqrt{b_m}\,G_m)|$ satisfy (7.12) *with $e^{m\Omega}$ replaced by $e^{\Lambda_{m,N}(k)}$. For two such dimensions, the difference of the two values of $\Lambda_{m,N}(k)$ is bounded uniformly in N and k. In particular, the corresponding expected determinants are comparable up to constants independent of N and k.*

Proof. We have $|\det(k\,\mathrm{Id}_m - \sqrt{b_m}G_m)| = b_m^{m/2}|\det(z\,\mathrm{Id}_m - G_m)|$ with $z = k/\sqrt{b_m} \ge 2 + g/\sqrt{b}$. For large N the shifts z lie in a fixed compact interval, and Proposition 7.9 applies. The finitely many remaining dimensions have $b_m > 0$ and are absorbed in the constants. For the comparison, consider $F(t) = t\big(\log(bt/N)/2 + \Omega(z_t)\big)$ with $z_t = k\sqrt{N/(bt)}$, for real t between the two dimensions. Since $\Omega' = g_{\mathrm{sc}}$ and $\mathrm{d}z_t/\,\mathrm{d}t = -z_t/(2t)$,

$$F'(t) = \frac{1}{2}\log\frac{bt}{N} + \Omega(z_t) + \frac{1}{2}\big(1 - z_t g_{\mathrm{sc}}(z_t)\big),$$

which is bounded uniformly for $t = N - O(1)$ and z_t in the compact interval. The mean value theorem completes the proof. □

The next lemma bounds determinants of matrices obtained by adding a border of fixed size to a matrix of the form $k\,\mathrm{Id} - W$. It is used in Section 11, where the Hessian at one of two nearby points is compared with a bulk minor.

Lemma 7.11 (Bordered determinants)**.** *Let $b, g > 0$, $k_+ < \infty$, and $2\sqrt{b} + g \le k \le k_+$. Let $1 \le m \le N$, let W have the law of $\sqrt{bm/N}\,G_m$, and set $T = k\,\mathrm{Id}_m - W$. Let $j \ge 1$ be a fixed*

integer, and let U be a random $m \times j$ matrix and C_∂ a random symmetric $j \times j$ matrix on the same probability space; no independence from W is assumed. Set

$$M = \begin{pmatrix} T & U \\ U^{\mathsf{T}} & C_\partial \end{pmatrix}, \qquad \mu_T = \mathbb{E}|\det T|, \qquad Y_T = \sum_{i=1}^{m} \log \max\{|k - \lambda_i(W)|,\, g/2\}.$$

1. *There exists C_j depending only on j, with $C_1 = 1$, such that surely*

$$|\det M| \le C_j\, e^{Y_T} \Big(\|C_\partial\|_{\mathrm{op}} + \frac{2}{g}\, \|U\|_F^2 \Big)^j. \tag{7.16}$$

2. *Suppose that $\mathbb{E}(\|C_\partial\|_{\mathrm{op}} + \|U\|_F^2)^q \le L_q$ for every $q \ge 1$. Then for every $r \ge 1$,*

$$\mathbb{E}|\det M|^r \le C_r\, \mu_T^r, \tag{7.17}$$

   where C_r depends only on b, g, k_+, j, r and L_{2jr}.

3. *Suppose in addition that there exist $s_0 > 0$ and an event $\mathcal{G}$ such that on $\mathcal{G}$ the matrix T is invertible and $|\det(C_\partial - U^{\mathsf{T}} T^{-1} U)| \ge s_0$, and that $\mathsf{K}_2^{1/2} \mathbb{P}(\mathcal{G}^c)^{1/2} \le 1/2$, where $\mathsf{K}_2 = \mathsf{K}_2(b, g, k_+)$. Then $\mathbb{E}|\det M| \ge s_0 \mu_T / 2$, and for every $r \ge 1$,*

$$\mathbb{E}\Big(\frac{|\det M|}{\mathbb{E}|\det M|} \Big)^r \le \frac{2^r C_r}{s_0^r}. \tag{7.18}$$

4. *If instead $M = k\,\mathrm{Id}_{m+j} - C - W'$, where $m + j \le N$, C is deterministic with $\|C\|_{\mathrm{op}} \le g/4$, and W' is a centered Gaussian matrix with $\operatorname{Var} \operatorname{Tr}(W'A) \le (2b/N) \operatorname{Tr}(A^2)$ for every symmetric A, then the normalized moments of $|\det M|$ are bounded by Proposition 7.7 in dimension $m + j$, with no Schur complement hypothesis.*

Proof. For (1), conjugate M by $\mathrm{diag}(O, \mathrm{Id}_j)$, where O is orthogonal and $O^{\mathsf{T}} T O = \mathrm{diag}(t_1, \dots, t_m)$. This replaces U by $O^{\mathsf{T}} U$ and does not change $\det M$, $\|U\|_F$, or Y_T. The determinant is affine in each t_i, and expanding in the t_i gives

$$\det M = \sum_{I \subset \{1,\dots,m\}} \Big(\prod_{i \notin I} t_i \Big) \det \begin{pmatrix} 0_{|I|} & U_I \\ U_I^{\mathsf{T}} & C_\partial \end{pmatrix},$$

where U_I consists of the rows of U with indices in I. If $|I| > j$, the last matrix has rank at most $2j < |I| + j$ and its determinant vanishes. For $|I| \le j$, Laplace expansion along the first $|I|$ rows and then along the first $|I|$ columns writes this determinant as a signed sum, over pairs of sets $J, J' \subset \{1, \dots, j\}$ of size $|I|$, of $\det U_{I,J} \det U_{I,J'}$ times a minor of C_∂ of order $j - |I|$. Each such minor has absolute value at most $\|C_\partial\|_{\mathrm{op}}^{j-|I|}$. By the Cauchy–Schwarz inequality and the Cauchy–Binet formula,

$$\sum_{|I|=\ell} |\det U_{I,J} \det U_{I,J'}| \le \Big(\sum_{|I|=\ell} \det(U_{I,J})^2 \sum_{|I|=\ell} \det(U_{I,J'})^2 \Big)^{1/2} \le \|U\|_F^{2\ell},$$

since each sum is the Gram determinant of ℓ columns of U, which is at most the product of their squared norms. Finally, $\prod_{i\notin I}|t_i| \le e^{Y_T}(g/2)^{-|I|}$. Summing over $\ell = |I| \le j$ and over J, J', we find

$$|\det M| \le e^{Y_T} \sum_{\ell=0}^{j} \binom{j}{\ell}^2 \|C_\partial\|_{\mathrm{op}}^{j-\ell} \Big(\frac{2}{g}\,\|U\|_F^2\Big)^{\ell},$$

which implies (7.16) with $C_j = \max_\ell \binom{j}{\ell}$. For $j = 1$, the sum is $e^{Y_T}(|C_\partial| + (2/g)|U|^2)$. The bound holds also when T is singular.

For (2), we apply the Cauchy–Schwarz inequality to (7.16). Using (7.8) with $C = 0$, we find that $\mathbb{E}e^{2rY_T} \le \mathsf{K}_{2r}\mu_T^{2r}$. Here we used that W satisfies the covariance hypothesis, by (7.2). The border factor has $2r$-th moment at most $\max\{1, 2/g\}^{2jr} L_{2jr}$. This proves (7.17).

For (3), the Schur complement formula implies that $|\det M| = |\det T|\,|\det(C_\partial - U^{\mathsf{T}}T^{-1}U)| \ge s_0|\det T|$ on $\mathcal{G}$. By Corollary 7.8 with $r = 2$,

$$\mathbb{E}|\det M| \ge s_0\mathbb{E}\big[|\det T|\mathbf{1}_{\mathcal{G}}\big] \ge s_0\mu_T\big(1 - \mathsf{K}_2^{1/2}\mathbb{P}(\mathcal{G}^c)^{1/2}\big) \ge \frac{s_0\mu_T}{2}.$$

Combining this with (7.17), we obtain (7.18). Part (4) is Proposition 7.7 in dimension $m + j \le N$. □

Remark 7.12 (The hypothesis in part (3) is needed). Bounded moments of the border give upper bounds relative to μ_T, but they do not bound $\mathbb{E}|\det M|$ from below. For example, let $j = 1$, $U = 0$, and $C_\partial = \mathbf{1}_{\mathcal{A}_N}$, where $\mathcal{A}_N$ is independent of W and $\mathbb{P}(\mathcal{A}_N) = e^{-N}$. All border moments are at most one, but $\mathbb{E}|\det M| = e^{-N}\mu_T$ and $\mathbb{E}|\det M|^r \ge e^{-N}\mu_T^r$, while the normalized r-th moment is at least $e^{(r-1)N}$. The Hessian at a single critical point has no border of this kind, and the Hessians under two-point conditioning are covered by part (4), as the next lemma shows.

We turn to two points. Let $u_1, u_2 \in S_N$ and $\rho = R(u_1, u_2)$ with $|\rho| < 1$. For a non-pure mixture, the *two-point conditioned law* with marks $(h_i, k_i)_{i=1,2}$ is defined as in Definition 7.3, with Λ_u replaced by the joint map $z \mapsto (\Lambda_{u_1}z, \Lambda_{u_2}z)$ and the target $(Nh_1, Nk_1, 0, Nh_2, Nk_2, 0)$, whenever the covariance of this joint vector is invertible on its target space. In the pure case the redundant radial derivatives are omitted, both zero gradients are retained, and the law is the conditional law given

$$\big(H_Q(u_1), H_Q(u_2), \nabla_{S_N}H_Q(u_1), \nabla_{S_N}H_Q(u_2)\big) = (Nh_1, Nh_2, 0, 0),$$

with $k_i = ph_i$. Only the spherical gradients enter: the radial component of the Euclidean gradient is $K_Q/\sqrt{N}$, which is determined by the value in the pure case.

Lemma 7.13 (Hessian determinants under two-point conditioning)**.** *Assume that $d \ge 4$. Let $g > 0$ and let $\mathcal{K}_2 \subset \mathbb{R}^4$ be compact, with $2\sqrt{b_Q} + g \le k_i \le k_+$ for $(h_1, k_1, h_2, k_2) \in \mathcal{K}_2$; in the pure case assume $k_i = ph_i$ on $\mathcal{K}_2$. There exists $\rho_0 \in (0, 1)$, depending only on ξ, Q, g, and $\mathcal{K}_2$, with the following properties for all $N \ge 2$ and all $u_1, u_2 \in S_N$ with $|R(u_1, u_2)| \le \rho_0$.*

1. *The covariance of* $(\Lambda_{u_1} Z, \Lambda_{u_2} Z)$ *is invertible on its target space.*
2. *Under the two-point conditioned law with marks in* $\mathcal{K}_2$*, for* $i = 1, 2$ *and every* $r \geq 1$*, the determinant* $D_i = |\det \nabla^2_{S_N} H_Q(u_i)|$ *satisfies* $\mathbb{E}(D_i / \mathbb{E} D_i)^r \leq \mathsf{K}_r(b_Q, g, k_+)$.

The determinants D_1 *and* D_2 *may be dependent.*

Proof. We first set up coordinates and record the covariances, then prove (1), and finally show that the conditional Hessian at u_1 has the form required by Proposition 7.7.

Coordinates and covariances. We write $\rho = R(u_1, u_2)$ and $\bar\rho = \sqrt{1 - \rho^2}$. By rotation invariance we may assume $u_1 = \sqrt{N} e_N$ and $u_2 = \sqrt{N}(\bar\rho e_{N-1} + \rho e_N)$. The unit vectors $\tau_1 = e_{N-1} \in u_1^\perp$ and $\tau_2 = \rho e_{N-1} - \bar\rho e_N \in u_2^\perp$ span the tangent directions in the plane of u_1 and u_2, and $e_1, \dots, e_{N-2}$ are tangent at both points. Let $Y_i = (H_Q(u_i), K_Q(u_i))/\sqrt{N}$, let $\eta_i = \tau_i \cdot \nabla H_Q(u_i)$, and call $\partial_c H_Q(u_i)$, $c \leq N - 2$, the common gradient components. Differentiating $\kappa(x, y) = N \xi_Q(x \cdot y / N)$ as in the proof of Lemma 7.1, at $x = u_1$, $y = u_2$, and using $u_1 \cdot u_2 = N\rho$, $u_2 \cdot \tau_1 = \sqrt{N}\bar\rho$, $u_1 \cdot \tau_2 = -\sqrt{N}\bar\rho$, and $\tau_1 \cdot \tau_2 = \rho$, we find

$$\begin{gathered}
\operatorname{Cov}(Y_1, Y_2) = \begin{pmatrix} \xi_Q(\rho) & \rho \xi_Q'(\rho) \\ \rho \xi_Q'(\rho) & \rho \xi_Q'(\rho) + \rho^2 \xi_Q''(\rho) \end{pmatrix}, \qquad \operatorname{Cov}(\eta_1, \eta_2) = \rho \xi_Q'(\rho) - \bar\rho^2 \xi_Q''(\rho), \\
\operatorname{Cov}(\eta_1, Y_2) = -\operatorname{Cov}(\eta_2, Y_1) = \bar\rho \big(\xi_Q'(\rho),\, \xi_Q'(\rho) + \rho \xi_Q''(\rho) \big), \\
\operatorname{Cov}\big(\partial_c H_Q(u_1), \partial_{c'} H_Q(u_2) \big) = \xi_Q'(\rho) \delta_{cc'}.
\end{gathered} \tag{7.19}$$

For instance, $\operatorname{Cov}(\eta_1, K_Q(u_2)) = \tau_1 \cdot \nabla_x \big(N f(t) \big) = f'(\rho)\, u_2 \cdot \tau_1$ with $f(t) = t \xi_Q'(t)$. The common gradient components are uncorrelated with Y_1, Y_2, η_1, η_2, and the one-point covariances are those of Lemma 7.1.

Invertibility. By the preceding paragraph, the covariance of all the observations is block diagonal. One block is the covariance $S(\rho)$ of $(Y_1, Y_2, \eta_1, \eta_2)$, a 6×6 matrix (a 4×4 matrix in the pure case) whose entries are polynomials in ρ and $\bar\rho$ that do not depend on N. The other blocks are $N - 2$ copies of the 2×2 matrix with diagonal a_Q and off-diagonal $\xi_Q'(\rho)$. All degrees of ξ are at least d, so $|\xi_Q^{(j)}(\rho)| \leq C|\rho|^{d-j}$ for $|\rho| \leq 1$ and $j \leq 3$. In particular $S(0) = \operatorname{diag}(\Sigma_Q, \Sigma_Q, a_Q, a_Q)$ (with $\tilde\xi(Q)$ in place of Σ_Q in the pure case), which is positive definite by Lemma 7.1, and $|\xi_Q'(\rho)| \leq C|\rho|^{d-1}$. Fix $\rho_1 > 0$ such that for $|\rho| \leq \rho_1$ the smallest eigenvalue of $S(\rho)$ is at least half that of $S(0)$ and $|\xi_Q'(\rho)| \leq a_Q/2$. Then all blocks are invertible, with inverses bounded independently of N. This proves (1) for $\rho_0 \leq \rho_1$.

The conditional Hessian. Consider $\nabla^2_\perp H_Q(u_1)$ in the orthonormal basis $e_1, \dots, e_{N-2}, \tau_1$ of $u_1^\perp$. We describe the covariances of its entries with the observations. Those with the observations at u_1 vanish by Lemma 7.1.

- The entries $\partial_a \partial_b H_Q(u_1)$ with $a, b \le N-2$ are uncorrelated with all the observations at u_2, since every term of the corresponding derivative of κ contains $(u_2)_a$, $(u_2)_b$, $(u_1)_c$, or $(\tau_2)_a$, all of which vanish.
- The entry $\partial_a \partial_{\tau_1} H_Q(u_1)$ with $a \le N-2$ is uncorrelated with Y_2 and η_2 and with all common gradient components except $\partial_a H_Q(u_2)$.
- The corner $v_1 = \tau_1 \cdot \nabla^2 H_Q(u_1)\tau_1$ is uncorrelated with the common gradient components, and

$$\operatorname{Cov}(v_1, Y_2) = \frac{\bar\rho^2}{\sqrt{N}}\big(\xi_Q''(\rho),\, 2\xi_Q''(\rho) + \rho\xi_Q'''(\rho)\big),$$
$$\operatorname{Cov}(v_1, \eta_2) = \frac{1}{\sqrt{N}}\big(2\rho\bar\rho\,\xi_Q''(\rho) - \bar\rho^3\xi_Q'''(\rho)\big).$$

The conditional mean of an entry is its covariance row with the observations, multiplied by the inverse covariance of the observations and by the target vector. The target vanishes on all gradient components. Since the covariance of the observations is block diagonal and the common gradient components form separate blocks with zero target, the first two items show that the conditional mean vanishes on the common block and on the cross row. Only the corner can have a nonzero conditional mean m_1.

To bound m_1, write the 6×6 block in terms of $Y = (Y_1, Y_2)$ and $\eta = (\eta_1, \eta_2)$ and let $c_Y = \operatorname{Cov}(v_1, Y)$ and $c_\eta = \operatorname{Cov}(v_1, \eta)$. The target is $y^* = \sqrt{N}(h_1, k_1, h_2, k_2)$ on Y and 0 on η, and block inversion gives

$$m_1 = \big(c_Y - c_\eta S_{\eta\eta}^{-1} S_{\eta Y}\big)\big(S_{YY} - S_{Y\eta} S_{\eta\eta}^{-1} S_{\eta Y}\big)^{-1} y^*.$$

By the bounds on $\xi_Q^{(j)}$, we have $|c_Y| \le C|\rho|^{d-2}/\sqrt{N}$, $|c_\eta| \le C|\rho|^{d-3}/\sqrt{N}$, and $|S_{\eta Y}| \le C|\rho|^{d-1}$, and $S_{\eta\eta}^{-1}$ is bounded. Then the first factor is at most $C(|\rho|^{d-2} + |\rho|^{2d-4})/\sqrt{N} \le C|\rho|^{d-2}/\sqrt{N}$. The middle factor is the inverse of the conditional covariance of Y given η, which equals the YY block of $S(\rho)^{-1}$ and is bounded. Finally, we have $|y^*| \le C\sqrt{N}$ on $\mathcal{K}_2$. We conclude that $|m_1| \le C_1|\rho|^{d-2}$, with C_1 depending only on ξ, Q, and $\mathcal{K}_2$. This bound includes the extensive size of the targets, and the conditional mean need not be zero.

Gaussian conditioning subtracts a positive semidefinite operator from the covariance. Then the centered part W of the conditional law of $\nabla_\perp^2 H_Q(u_1)$ satisfies $\operatorname{Var}\operatorname{Tr}(WA) \le (2b_Q/N)\operatorname{Tr}(A^2)$, by (7.2) applied to the unconditional law. Its mean is $C = m_1\tau_1\tau_1^{\mathsf T}$, with $\|C\|_{\mathrm{op}} \le C_1|\rho|^{d-2}$. We now choose $\rho_0 \le \rho_1$ such that $C_1\rho_0^{d-2} \le g/4$. Since $K_Q(u_1) = Nk_1$ under the conditioned law, (7.3) implies that $-\nabla_{S_N}^2 H_Q(u_1) = k_1\mathrm{Id} - C - W$ in law, and (2) for $i = 1$ follows from Proposition 7.7. The case $i = 2$ follows by exchanging the roles of u_1 and u_2, and the pure case is the same with $Y_i = H_Q(u_i)/\sqrt{N}$ and $k_i = ph_i$. When $N = 2$, there are no common directions: the common gradient components, the blocks of the observation covariance that they form, and the common Hessian block are empty, $u_1^\perp$ is spanned by τ_1, and the argument applies verbatim. □

Remark 7.14 (Related determinant estimates). Subag [140, Lemma 21] proves exponentially small determinant-weighted edge tails for a fixed number of possibly dependent matrices with GOE marginals and shifts in a compact interval outside $[-2, 2]$. After rescaling and replacing G by $-G$, this is the kind of estimate that Corollary 7.8 provides. In [141, Lemma 38, (12.25)], he records the normalized second moment of the determinant at a critical point in windows of sublinear width in the pure model. The proofs above use the clipped logarithmic determinant and concentration approach of [140, proof of Lemma 21, (8.13)–(8.18)]. Propositions 7.7 and 7.9 give all fixed moments directly, including the covariance-dominated Gaussian matrices and deterministic perturbations in Proposition 7.7. The two-point block structure in the proof of Lemma 7.13 agrees with [30, Lemma 15].

7.5 The Kac–Rice formula with field-dependent weights

We prove the Kac–Rice formula in the coefficient space of the field. The first ingredient is the equal-dimensional area formula [69], together with the corresponding elementary case of Sard's theorem [134]; we include the short proof.

Lemma 7.15 (Equal-dimensional area formula)**.** *Let $V \subset \mathbb{R}^n$ be open and let $\alpha\colon V \to \mathbb{R}^n$ be C^1. Then the image under α of $\{t \in V : \det D\alpha(t) = 0\}$ has Lebesgue measure zero, and for every Borel function $w \geq 0$ on V,*

$$\int_{\mathbb{R}^n} \sum_{t \in V\colon \alpha(t) = y} w(t)\, \mathrm{d}y = \int_V w(t)\, |\det D\alpha(t)|\, \mathrm{d}t. \tag{7.20}$$

The sum may be defined arbitrarily for y in the image of the critical set.

Proof. Let $K \subset V$ be compact, let $\epsilon > 0$, and let L bound $\|D\alpha\|$ on a compact neighborhood of K. By uniform continuity of $D\alpha$, there exists $\delta_0 > 0$ such that for every cube Q of side $\delta \leq \delta_0$ that meets K and contains a critical point t_0, we have $|\alpha(t) - \alpha(t_0) - D\alpha(t_0)(t - t_0)| \leq \epsilon\sqrt{n}\,\delta$ for $t \in \mathsf{Q}$. The image of Q under the affine map $t \mapsto \alpha(t_0) + D\alpha(t_0)(t - t_0)$ lies in an affine subspace of dimension at most $n - 1$, within a ball of radius $L\sqrt{n}\,\delta$. Then $\alpha(\mathsf{Q})$ lies in a box with $n - 1$ sides of length at most $C_n(L + 1)\delta$ and one side of length at most $C_n\epsilon\delta$, whose volume is at most $C_n(L + 1)^{n-1}\epsilon\delta^n$. A grid of cubes of side δ covering K has at most $C_K\delta^{-n}$ cubes meeting K. We conclude that the image of the critical points in K has outer measure at most $C_{K,n}(L + 1)^{n-1}\epsilon$. Letting $\epsilon \to 0$ and exhausting V by compact sets proves the first assertion.

Every regular point has an open neighborhood on which α is a C^1 diffeomorphism onto an open set. The regular set is covered by countably many such neighborhoods V_i, and the sets $P_i = V_i \setminus \bigcup_{i' < i} V_{i'}$ form a Borel partition of it. For each i, by the change of variables formula for nonnegative Borel functions, we have $\int \mathbf{1}_{\alpha(P_i)}(y)\, w((\alpha|_{V_i})^{-1}(y))\, \mathrm{d}y =$

$\int_{P_i} w(t)|\det D\alpha(t)|\,\mathrm{d}t$. Summing over i with Tonelli's theorem proves (7.20): the critical set contributes zero to the right side and only a null set of values y to the left side. □

Next we state the Kac–Rice formula for a general Gaussian field that depends linearly on finitely many coefficients. Let M be a compact smooth Riemannian manifold of dimension n without boundary, let $\phi_1, \dots, \phi_\ell$ be smooth functions on M, and set $f_z = \sum_i z_i \phi_i$ for $z \in \mathbb{R}^\ell$. Let Z be a standard Gaussian vector in $\mathbb{R}^\ell$. For $u \in \mathsf{M}$, let $\Gamma_u \colon \mathbb{R}^\ell \to T_u\mathsf{M}$ be the linear map $z \mapsto \nabla f_z(u)$. If $\Gamma_u \Gamma_u^*$ is positive definite, we write $Z_0^u = (\mathrm{Id} - \Gamma_u^*(\Gamma_u\Gamma_u^*)^{-1}\Gamma_u)Z$, which is a standard Gaussian vector in the subspace $\ker \Gamma_u = \{z : \nabla f_z(u) = 0\}$, and $p_u(0) = (2\pi)^{-n/2}\det(\Gamma_u\Gamma_u^*)^{-1/2}$, which is the density of $\nabla f_Z(u)$ at 0 in any orthonormal basis of $T_u\mathsf{M}$. As in Lemma 7.4(1), the law of Z_0^u is a version of the conditional law of Z given $\nabla f_Z(u) = 0$.

Proposition 7.16 (Kac–Rice formula in coefficient space)**.** *Assume that $\Gamma_u\Gamma_u^*$ is positive definite for every $u \in \mathsf{M}$. Then the set of z for which f_z has a degenerate critical point is closed and Lebesgue null. In particular, almost surely f_Z is a Morse function and has finitely many critical points. For every Borel function $\varphi \ge 0$ on $\mathbb{R}^\ell \times \mathsf{M}$, the function $z \mapsto \sum_{u:\nabla f_z(u)=0} \varphi(z,u)$ is Borel on the set of Morse coefficients, and*

$$\mathbb{E} \sum_{u:\,\nabla f_Z(u)=0} \varphi(Z,u) = \int_{\mathsf{M}} p_u(0)\,\mathbb{E}\Big[\big|\det \nabla^2 f_{Z_0^u}(u)\big|\,\varphi(Z_0^u,u)\Big]\mathrm{vol}(\,\mathrm{d}u). \tag{7.21}$$

The same identity holds for a Borel function φ of either sign whenever $\mathbb{E}\sum_{u:\nabla f_Z(u)=0}|\varphi(Z,u)| < \infty$; by (7.21) *applied to $|\varphi|$, this is equivalent to the finiteness of the right side of* (7.21) *with $|\varphi|$ in place of φ.*

Proof. The proof works in local coordinates on M. We split the coefficients so that the zeros of the gradient become the graph of a function of the remaining coefficients. The area formula then gives the identity in each chart, and a change of variables identifies the result with (7.21).

Local coordinates. Let $t \in V \subset \mathbb{R}^n$ be local coordinates on an open subset of M, and let $\gamma(z,t) = (\partial_{t_1} f_z, \dots, \partial_{t_n} f_z)(t) = \Gamma(t)z$ be the coordinate gradient, where $\Gamma(t)$ is an $n \times \ell$ matrix of rank n depending smoothly on t. Near each point of V we may choose n coefficient indices such that the corresponding $n \times n$ submatrix $\Gamma_1(t)$ of $\Gamma(t)$ is invertible, and by shrinking V we may assume that it is invertible on V. After permuting coordinates, we write $z = (z', z'') \in \mathbb{R}^n \times \mathbb{R}^{\ell-n}$ and $\Gamma(t) = (\Gamma_1(t), \Gamma_2(t))$, so that

$$\gamma(z,t) = 0 \quad\Longleftrightarrow\quad z' = \alpha_{z''}(t), \qquad \alpha_{z''}(t) = -\Gamma_1(t)^{-1}\Gamma_2(t)z''.$$

Differentiating $\Gamma_1(t)\alpha_{z''}(t) + \Gamma_2(t)z'' = 0$ in t gives

$$D\alpha_{z''}(t) = -\Gamma_1(t)^{-1} D_t\gamma\big((\alpha_{z''}(t), z''), t\big). \tag{7.22}$$

At a zero of the gradient, $D_t\gamma$ is the coordinate Hessian. By (7.22), $\det D\alpha_{z''}(t) = 0$ exactly when the critical point t of $f_{(\alpha_{z''}(t),z'')}$ is degenerate. By compactness, M is covered by finitely many such charts $V_1, \dots, V_J$.

Nondegeneracy. The set of pairs (z, u) with $\nabla f_z(u) = 0$ and $\det \nabla^2 f_z(u) = 0$ is closed in $\mathbb{R}^\ell \times \mathsf{M}$, because the Hessian operator depends continuously on (z, u). Its projection to $\mathbb{R}^\ell$ is closed, since M is compact. This is the set $\mathcal{D}$ of coefficients with a degenerate critical point. For a chart V_i and fixed z'', a coefficient (z', z'') has a degenerate critical point in V_i exactly when z' is a critical value of $\alpha_{z''}$, which happens only for a null set of z' by Lemma 7.15. Fubini's theorem then implies that $\mathcal{D}$ is Lebesgue null. The law of Z is absolutely continuous, so f_Z is almost surely a Morse function. Its critical points are then isolated, and since they form a closed subset of the compact manifold M, there are finitely many of them.

Measurability of the sum. At a pair (z_0, u_0) with $\nabla f_{z_0}(u_0) = 0$ and $\det \nabla^2 f_{z_0}(u_0) \neq 0$, the implicit function theorem provides an open set $O \ni z_0$ and a continuous map $u(\cdot)$ on O such that near (z_0, u_0) the critical pairs are exactly the points of the graph $\{(z, u(z)) : z \in O\}$. Countably many such graphs $\mathsf{G}_i = \{(z, u_i(z)) : z \in O_i\}$ cover the set of nondegenerate critical pairs. Let $\mathsf{P}_i = \mathsf{G}_i \setminus \bigcup_{i'<i} \mathsf{G}_{i'}$. Since the projection to z is a homeomorphism from G_i onto O_i, the set $\pi(\mathsf{P}_i)$ is Borel, and

$$\sum_{u:\nabla f_z(u)=0} \varphi(z,u) = \sum_i \mathbf{1}_{\pi(\mathsf{P}_i)}(z)\, \varphi\big(z, u_i(z)\big) \qquad (z \notin \mathcal{D})$$

is Borel. The same enumeration shows that events defined in terms of the critical points, such as unions over critical points of Borel events, are Borel.

The identity in a chart. Let (χ_i) be a smooth partition of unity subordinate to (V_i), fix i, and write $V = V_i$, $\chi = \chi_i$. Let ϕ_ℓ denote the standard Gaussian density on $\mathbb{R}^\ell$. For fixed z'', apply Lemma 7.15 to $\alpha_{z''}$ with the weight $w(t) = \phi_\ell(\alpha_{z''}(t), z'')\chi(t)\varphi((\alpha_{z''}(t), z''), t)$. Using (7.22), integrating over z'', and using Tonelli's theorem, we obtain

$$\begin{aligned}
&\mathbb{E} \sum_{t\in V:\gamma(Z,t)=0} \chi(t)\varphi(Z,t) \\
&\quad = \int_V \int_{\mathbb{R}^{\ell-n}} \frac{\phi_\ell(\alpha_{z''}(t), z'')}{|\det \Gamma_1(t)|} \big|\det D_t\gamma(\bar z, t)\big|\, \chi(t)\varphi(\bar z, t)\, \mathrm{d}z''\, \mathrm{d}t, \qquad \bar z = (\alpha_{z''}(t), z''). \quad (7.23)
\end{aligned}$$

When $\ell = n$, the inner integral is absent.

Identification of the conditional law. Fix $t \in V$. Since the map $(z', z'') \mapsto (\Gamma(t)z, z'')$ is linear with Jacobian $|\det \Gamma_1(t)|$, the density of $\gamma(Z, t)$ at 0 is $p_t^{\mathrm{co}}(0) = \int \phi_\ell(\alpha_{z''}(t), z'')\, \mathrm{d}z'' / |\det \Gamma_1(t)|$. The map $z'' \mapsto (\alpha_{z''}(t), z'')$ is a linear bijection onto $\ker \Gamma(t)$, and it carries Lebesgue measure to a constant multiple of Lebesgue measure

on $\ker \Gamma(t)$. Further, ϕ_ℓ restricted to $\ker \Gamma(t)$ is proportional to the standard Gaussian density of that subspace. We conclude that, for every Borel $\psi \geq 0$,

$$\int \frac{\phi_\ell(\alpha_{z''}(t), z'')}{|\det \Gamma_1(t)|}\, \psi(\alpha_{z''}(t), z'')\, \mathrm{d}z'' = p_t^{\mathrm{co}}(0)\, \mathbb{E}\psi(Z_0^u),$$

where u is the point with coordinates t. The constant is identified by taking $\psi = 1$. Here we used that $\ker \Gamma(t) = \ker \Gamma_u$, since the coordinate gradient and the gradient determine each other. This identification holds at the single value 0 of the gradient, for every Borel integrand.

From coordinates to orthonormal frames. Let $E(t)$ be the matrix expressing the coordinate vectors ∂_{t_i} in an orthonormal frame of $T_u\mathsf{M}$, so that the metric is $G(t) = E(t)^{\mathsf{T}} E(t)$ and $\gamma(z,t)$ is $E(t)^{\mathsf{T}}$ applied to the frame components of $\nabla f_z(u)$. Then $p_t^{\mathrm{co}}(0) = p_u(0)/\sqrt{\det G(t)}$. At a critical point the Christoffel terms vanish, and $D_t\gamma = E^{\mathsf{T}} \nabla^2 f_z(u) E$ and $|\det D_t\gamma| = \det G(t)\, |\det \nabla^2 f_z(u)|$. Inserting these facts into (7.23), the integrand becomes $p_u(0)\, \mathbb{E}[|\det \nabla^2 f_{Z_0^u}(u)| \chi(u) \varphi(Z_0^u, u)] \sqrt{\det G(t)}$, and $\sqrt{\det G(t)}\, \mathrm{d}t$ is the Riemannian volume. Summing over the charts proves (7.21). For φ of either sign with $\mathbb{E} \sum |\varphi(Z,u)| < \infty$, we apply (7.21) to the positive and negative parts of φ. Both resulting identities are between finite quantities, and we subtract them. Integrability of the signed sum alone would not suffice, since cancellations between the critical points can make the signed sum integrable while the right side of (7.21) is undefined. □

We now specialize to the center field on S_N.

Theorem 7.17 (Kac–Rice formula with field-dependent weights)**.** *Let $\mathcal{K} \subset \mathbb{R}^2$ be a Borel set and let φ be a field-dependent weight that is either nonnegative or bounded. Almost surely, H_Q restricted to S_N is a Morse function and has finitely many critical points. If ξ is a non-pure mixture, then*

$$\begin{aligned} &\mathbb{E} \sum_{\substack{u \in S_N:\ \nabla_{S_N} H_Q(u) = 0 \\ (H_Q(u), K_Q(u))/N \in \mathcal{K}}} \varphi(H_N, u) \\ &\qquad = \frac{|S_N|}{(2\pi a_Q)^{n/2}} \int_{S_N} \int_{\mathcal{K}} p_N(h,k)\, \mathbb{E}_{h,k,u}\Big[\big|\det \nabla^2_{S_N} H_Q(u)\big|\, \varphi(H_N, u) \Big]\, \mathrm{d}h\, \mathrm{d}k\, \nu_N(\mathrm{d}u), \end{aligned} \tag{7.24}$$

with p_N as in (7.5)*. If ξ is pure of degree p, let $\mathcal{K}_p = \{h \in \mathbb{R} : (h, ph) \in \mathcal{K}\}$; then*

$$\begin{aligned} &\mathbb{E} \sum_{\substack{u \in S_N:\ \nabla_{S_N} H_Q(u) = 0 \\ H_Q(u)/N \in \mathcal{K}_p}} \varphi(H_N, u) \\ &\qquad = \frac{|S_N|}{(2\pi a_Q)^{n/2}} \int_{S_N} \int_{\mathcal{K}_p} p_N^{\mathrm{pure}}(h)\, \mathbb{E}_{h,u}\Big[\big|\det \nabla^2_{S_N} H_Q(u)\big|\, \varphi(H_N, u) \Big]\, \mathrm{d}h\, \nu_N(\mathrm{d}u). \end{aligned} \tag{7.25}$$

For nonnegative φ both sides may be infinite; for bounded φ both sides are finite. In the pure case there is no two-dimensional density, since Σ_Q is singular.

Proof. We apply Proposition 7.16 with $\mathsf{M} = S_N$, $\ell = \ell_N$, and f_z the restriction of $H_{z,Q}$ to S_N, which is linear in z. By Lemma 7.1, the covariance of $\nabla_{S_N} H_Q(u)$ is $a_Q \,\mathrm{Id}_{u^\perp}$, which is positive definite, and $p_u(0) = (2\pi a_Q)^{-n/2}$. The vector Z_0^u of Proposition 7.16 is the vector of Definition 7.3, and the conditional expectation in (7.21) is $\mathbb{E}_u^{\nabla}$. The Riemannian volume of S_N is $|S_N|\,\nu_N$.

Let $\varphi \geq 0$, and apply (7.21) to the Borel function

$$(z,u) \mapsto \varphi(H_z,u)\,\mathbf{1}\big\{(H_{z,Q}(u),K_{z,Q}(u))/N \in \mathcal{K}\big\}.$$

By Lemma 7.4(4) and (2), the inner expectation equals

$$\int_{\mathbb{R}^2} p_N(h,k)\,\mathbb{E}_{h,k,u}\Big[\big|\det \nabla^2_{S_N} H_Q(u)\big|\,\varphi(H_N,u)\Big]\mathbf{1}_{\mathcal{K}}(h,k)\,\mathrm{d}h\,\mathrm{d}k,$$

since $(H_Q(u),K_Q(u)) = (Nh,Nk)$ surely under $\mathbb{P}_{h,k,u}$. This proves (7.24). The joint measurability needed for the integral is Lemma 7.4(3). No continuity of φ is used, because the incidence argument of Proposition 7.16 applies to every Borel integrand. In the pure case, $K_Q = pH_Q$ and $(H_Q(u),K_Q(u))/N \in \mathcal{K}$ exactly when $H_Q(u)/N \in \mathcal{K}_p$. Then (7.25) follows from the pure part of Lemma 7.4(4).

For bounded φ, take $\varphi = 1$ and $\mathcal{K} = \mathbb{R}^2$. By Lemma 7.4(2), the right side of (7.24) becomes $|S_N|(2\pi a_Q)^{-n/2}\int p_N(h,k)\mu_{n,N}(k)\,\mathrm{d}h\,\mathrm{d}k$. This is finite, as $\mu_{n,N}(k) \leq \mathbb{E}(\|W_{n,N}\|_{\mathrm{op}} + |k|)^n \leq C_N(1+|k|)^n$. This shows that the expected number of critical points is finite. For bounded φ, the hypothesis of the last assertion of Proposition 7.16 then holds, and the formula follows from the positive and negative parts of φ. The pure case is identical. □

Corollary 7.18 (Expected numbers of critical points)**.** *Let $I, U \subset \mathbb{R}$ be Borel sets and set*

$$\mathcal{C}_N(I,U) = \big\{u \in S_N : \nabla_{S_N} H_Q(u) = 0,\ H_Q(u)/N \in I,\ K_Q(u)/N \in U\big\}.$$

If ξ is a non-pure mixture, then

$$\mathbb{E}|\mathcal{C}_N(I,U)| = \frac{|S_N|}{(2\pi a_Q)^{n/2}}\,\frac{N}{2\pi\sqrt{\det\Sigma_Q}}\int_I\int_U e^{-\frac{N}{2}(h,k)\Sigma_Q^{-1}(h,k)^{\mathsf{T}}}\mu_{n,N}(k)\,\mathrm{d}k\,\mathrm{d}h. \tag{7.26}$$

If ξ is pure of degree p, then

$$\mathbb{E}|\mathcal{C}_N(I,U)| = \frac{|S_N|}{(2\pi a_Q)^{n/2}}\Big(\frac{N}{2\pi\xi(Q)}\Big)^{1/2}\int_I \mathbf{1}\{ph \in U\}\,e^{-Nh^2/(2\xi(Q))}\mu_{n,N}(ph)\,\mathrm{d}h. \tag{7.27}$$

Both quantities are finite. With $I = [(m_N - M)/N, (m_N + M)/N]$, these formulas give the expected size of the set $\mathcal{C}_{N,M}$ of critical points in a fixed window.

Proof. We take $\varphi = 1$ and $\mathcal{K} = I \times U$ in Theorem 7.17. By Lemma 7.4(2), the conditioned expectation of the determinant is $\mu_{n,N}(k)$, which depends neither on u nor on h, and the integral over ν_N equals one. Finiteness was shown in the proof of Theorem 7.17. □

No asymptotic estimate of these integrals is part of the statement. The counts in fixed windows are estimated in Sections 9 to 11. We record two reformulations that are used there.

Remark 7.19 (Palm form). For a non-pure mixture, let

$$\Lambda_N(\,\mathrm{d}u\;\mathrm{d}h\;\mathrm{d}k) = \frac{|S_N|}{(2\pi a_Q)^{n/2}}\, p_N(h,k)\,\mu_{n,N}(k)\,\nu_N(\,\mathrm{d}u)\,\mathrm{d}h\,\mathrm{d}k$$

be the intensity measure of the marked critical points, so that $\Lambda_N(S_N \times I \times U) = \mathbb{E}|\mathcal{C}_N(I,U)|$. For $u \in S_N$ and $(h,k) \in \mathbb{R}^2$, the *Palm law* at u with marks (h,k) is the determinant-weighted law

$$\mathbb{E}^{\mathrm{Palm}}_{h,k,u}[\Phi] = \frac{\mathbb{E}_{h,k,u}\big[|\det \nabla^2_{S_N} H_Q(u)|\,\Phi\big]}{\mu_{n,N}(k)}.$$

With this notation, (7.24) reads

$$\mathbb{E} \sum_{\substack{u:\nabla_{S_N} H_Q(u)=0 \\ (H_Q(u),K_Q(u))/N\in\mathcal{K}}} \varphi(H_N,u) = \int_{S_N\times\mathcal{K}} \mathbb{E}^{\mathrm{Palm}}_{h,k,u}\big[\varphi(H_N,u)\big]\,\Lambda_N(\,\mathrm{d}u\;\mathrm{d}h\;\mathrm{d}k).$$

In the pure case, Λ_N has density $|S_N|(2\pi a_Q)^{-n/2} p_N^{\mathrm{pure}}(h)\mu_{n,N}(ph)$ with respect to $\nu_N(\,\mathrm{d}u)\,\mathrm{d}h$, and the Palm law is $\mathbb{E}_{h,u}[|\det \nabla^2_{S_N} H_Q(u)|\,\Phi]/\mu_{n,N}(ph)$. The Palm law and the conditioned law are different laws, and an estimate under one does not transfer to the other without an argument such as Lemma 7.22.

Remark 7.20 (Antipodal points and Subag's tameness condition). The formulas above count points of S_N. In an even model, the critical points come in antipodal pairs, and for a selection invariant under $u \mapsto -u$ the number of pairs is half of (7.26) or (7.27). For a general weight one must instead specify a representative of each pair, or average the two values. A weighted sum cannot be divided by two without such a convention. For pure models of degree $p \geq 3$, Subag [141, Lemma 36] proves a marked Kac–Rice formula under a tameness condition on an auxiliary field. Theorem 7.17 requires only Borel measurability of the weight. This matters for weights given by probabilities of trajectory events, for which the tameness hypotheses of [141, Lemma 36] have not been verified.

7.6 Hitting probabilities as weights

We now show that hitting probabilities of the Langevin diffusion are field-dependent weights. For $H \in \mathcal{H}_N$, $\beta \geq 0$, and $x \in S_N$, let $\mathbb{P}_x^H$ be the law on $C([0,T], S_N)$ of the diffusion of Definition 3.4 started at x. Recall from Lemma 3.5 that it is the law of the solution of the Itô equation (3.6), which has noise coefficient $\sqrt{2}$ and stays on S_N. We write $\mathbb{P}_{\nu_N}^H = \int \mathbb{P}_x^H \, \nu_N(\mathrm{d}x)$ for the law of the diffusion started from ν_N, independently of the field and of the noise.

Lemma 7.21 (Hitting probabilities are field-dependent weights)**.** *Let $N \geq 2$, $\beta \geq 0$, and $T \in (0, \infty)$. For every Borel set $A \subset C([0,T], S_N)$, the map $H \mapsto \mathbb{P}_{\nu_N}^H(A)$ is Borel on $\mathcal{H}_N$. For every Borel set $\mathcal{A} \subset C([0,T], S_N) \times S_N$, the function*

$$\varphi_{\mathcal{A}}(H,u) = \mathbb{P}_{\nu_N}^H\big(\{\omega : (\omega,u) \in \mathcal{A}\}\big)$$

is a field-dependent weight with values in $[0,1]$. In particular, for $a \in [0,1]$,

$$\varphi_{T,a}(H,u) = \mathbb{P}_{\nu_N}^H\Big(\sup_{0 \leq t \leq T} |R(X_t,u)| \geq a\Big) \tag{7.28}$$

is a field-dependent weight. This applies to $T = e^{\gamma N}$ for each fixed N.

Proof. By Lemma 4.2(a), $(H,x) \mapsto \mathbb{P}_x^H$ is a Borel probability kernel from $\mathcal{H}_N \times S_N$ to $C([0,T], S_N)$. Its proof realizes all these laws through the same Brownian motion and shows that they depend continuously on (H,x). For a bounded Borel function F on $C([0,T], S_N)$, the function $(H,x) \mapsto \int F \, \mathrm{d}\mathbb{P}_x^H$ is Borel, and Tonelli's theorem shows that $H \mapsto \int F \, \mathrm{d}\mathbb{P}_{\nu_N}^H$ is Borel. This proves the first assertion. For a Borel set $\mathcal{A}$, the function $(H,u) \mapsto \int \mathbf{1}_{\mathcal{A}}(\omega,u) \, \mathbb{P}_{\nu_N}^H(\mathrm{d}\omega)$ is Borel when $\mathcal{A} = A \times E$ is a product of Borel sets, by the first assertion. The sets $\mathcal{A}$ for which it is Borel form a Dynkin system, and the π–λ theorem extends this to all Borel sets $\mathcal{A}$. Finally, $(\omega,u) \mapsto \sup_{t \leq T} |R(\omega(t),u)|$ is continuous for the uniform topology, since $|R(x,u) - R(x',u')| \leq (|x - x'| + |u - u'|)/\sqrt{N}$ on S_N. It follows that the event in (7.28) is a Borel subset of $C([0,T], S_N) \times S_N$. No bound uniform in T is needed, since N and T are fixed. □

Lemma 4.3(b) states the special case of Lemma 7.21 that is used in Sections 4 and 13.

7.7 Determinant-weighted transfer

The estimate of Section 8 bounds a hitting probability under the conditioned law $\mathbb{P}_{h,k,u}$. The Kac–Rice formula weights this law by the absolute determinant, that is, it averages under the Palm law of Remark 7.19. The following elementary lemma bounds the cost of this change of law.

Lemma 7.22 (Hölder bound for determinant-weighted expectations)**.** *Let $D \geq 0$ be a random variable with $0 < \mu = \mathbb{E}D < \infty$, and suppose that $\mathbb{E}(D/\mu)^r \leq \mathsf{K}$ for some $r > 1$. Let φ be a random variable with $0 \leq \varphi \leq 1$ and $\mathbb{E}\varphi \leq q$. Then*

$$\frac{\mathbb{E}[D\varphi]}{\mu} \leq \mathsf{K}^{1/r} q^{1-1/r}. \tag{7.29}$$

No assumption on the dependence between φ and D is made.

Proof. Let $r' = r/(r-1)$. By Hölder's inequality, $\mathbb{E}[(D/\mu)\varphi] \leq (\mathbb{E}(D/\mu)^r)^{1/r} (\mathbb{E}\varphi^{r'})^{1/r'}$. Since $0 \leq \varphi \leq 1$, we have $\varphi^{r'} \leq \varphi$, and the last factor is at most $q^{1/r'} = q^{1-1/r}$. □

For $r = 2$ the loss is $\mathsf{K}_2^{1/2} q^{1/2}$. If an estimate is already available under the Palm law, there is no loss, and Lemma 7.22 is not needed. The lemma is not a perturbation statement: φ may be strongly correlated with D, and the bound uses only the moment of D/μ and the mean of φ.

Theorem 7.23 (Determinant-weighted transfer)**.** *Let $g > 0$ and let $\mathcal{K} \subset \mathbb{R}^2$ be compact, with $k \geq 2\sqrt{b_Q} + g$ for every $(h,k) \in \mathcal{K}$; in the pure case assume that $k = ph$ for every $(h,k) \in \mathcal{K}$. If $\mathcal{K} \neq \varnothing$, let $k_+ = \max\{k : (h,k) \in \mathcal{K}\}$; if $\mathcal{K} = \varnothing$, set $k_+ = 2\sqrt{b_Q} + g$. For each N, let $\mathcal{W}_N \subset \mathcal{K}$ be a Borel set, let*

$$\mathcal{C}_N = \big\{u \in S_N : \nabla_{S_N} H_Q(u) = 0,\ (H_Q(u), K_Q(u))/N \in \mathcal{W}_N\big\},$$

and let φ_N be a field-dependent weight with values in $[0,1]$. Suppose that $q_N \geq 0$ satisfies

$$\mathbb{E}_{h,k,u} \varphi_N(H_N, u) \leq q_N \qquad \text{for all } u \in S_N \text{ and } (h,k) \in \mathcal{W}_N. \tag{7.30}$$

Then for every $r > 1$ and $N \geq 2$,

$$\mathbb{E} \sum_{u \in \mathcal{C}_N} \varphi_N(H_N, u) \leq \mathsf{K}_r^{1/r} q_N^{1-1/r}\, \mathbb{E}|\mathcal{C}_N|, \tag{7.31}$$

where $\mathsf{K}_r = \mathsf{K}_r(b_Q, g, k_+)$ is the constant of Proposition 7.7. The same bound holds with φ_N replaced by $\theta\varphi_N$, for every field-dependent weight θ with values in $[0,1]$.

Proof. If $\mathcal{K} = \varnothing$, then $\mathcal{W}_N = \mathcal{C}_N = \varnothing$, so both claims are immediate. Otherwise, fix $u \in S_N$ and $(h,k) \in \mathcal{W}_N$, and let $D = |\det \nabla^2_{S_N} H_Q(u)|$ under $\mathbb{P}_{h,k,u}$. Recall from Lemma 7.4(2) that $\mathbb{E}_{h,k,u} D = \mu_{n,N}(k)$. By Proposition 7.7, $\mathbb{E}_{h,k,u}(D/\mu_{n,N}(k))^r \leq \mathsf{K}_r$. Combining Lemma 7.22 with (7.30), we obtain

$$\mathbb{E}_{h,k,u}\big[D\, \varphi_N(H_N, u)\big] \leq \mathsf{K}_r^{1/r} q_N^{1-1/r}\, \mu_{n,N}(k).$$

We insert this bound into (7.24) with $\mathcal{K}$ replaced by $\mathcal{W}_N$ (or into (7.25) in the pure case). The remaining integral is the right side of (7.24) with $\varphi = 1$, which equals $\mathbb{E}|\mathcal{C}_N|$ by

Theorem 7.17 with $\varphi = 1$. This proves (7.31). For the last assertion, $D\theta\varphi_N \le D\varphi_N$ pointwise, and the same integral bounds the left side. Note that the normalizer on the right side of (7.31) remains the expected number of all critical points in $\mathcal{C}_N$, and no lower bound on the Palm law weighted by θ is needed. □

We now apply Theorem 7.23 to hitting probabilities and to the critical points in a fixed window. For a deterministic sequence (m_N), a Borel set $U \subset \mathbb{R}$, and $M > 0$, let

$$\mathcal{C}_{N,M} = \big\{u \in S_N : \nabla_{S_N} H_Q(u) = 0,\ |H_Q(u) - m_N| \le M,\ K_Q(u)/N \in U\big\}. \tag{7.32}$$

We call this set the critical points in a *fixed window*, since its half-width M does not depend on N. The marks of its points lie in $\mathcal{W}_{N,M} = [(m_N - M)/N, (m_N + M)/N] \times U$, intersected with the line $\{k = ph\}$ in the pure case. In an even model, both points of each antipodal pair are counted.

Corollary 7.24 (Critical points in a fixed window)**.** *Assume that $d \ge 4$ and $\beta \ge 0$, and let g, $\mathcal{K}$, and k_+ be as in Theorem 7.23. Suppose the following.*

1. *There exist $a \in (0,1]$, $\gamma > 0$, and numbers q_N with $q_N \to 0$ such that, for all N, all $u \in S_N$, and all $(h,k) \in \mathcal{K}$,*

$$\mathbb{E}_{h,k,u}\mathbb{P}^{H_N}_{\nu_N}\Big(\sup_{0\le t\le e^{\gamma N}} |R(X_t,u)| \ge a\Big) \le q_N.$$

2. *For every $M > 0$ there exists N_M such that $\mathcal{W}_{N,M} \subset \mathcal{K}$ for all $N \ge N_M$.*
3. *For every $M > 0$, $C_M = \sup_N \mathbb{E}|\mathcal{C}_{N,M}| < \infty$.*

Then for every $r > 1$, $M > 0$, and $N \ge N_M$,

$$\mathbb{E}\,\mathbb{P}^{H_N}_{\nu_N}\Big(\exists u \in \mathcal{C}_{N,M},\ \exists t \le e^{\gamma N} :\ |R(X_t,u)| \ge a\Big) \le C_M \mathsf{K}_r^{1/r} q_N^{1-1/r}. \tag{7.33}$$

The same bound holds for every subfamily of $\mathcal{C}_{N,M}$ selected by a field-dependent weight with values in $\{0,1\}$. The exponent γ does not depend on M.

Proof. Let $T_N = e^{\gamma N}$ and $\varphi_N = \varphi_{T_N,a}$. By Lemma 7.21, φ_N is a field-dependent weight with values in $[0,1]$. Fix M and $N \ge N_M$. By hypothesis (1), (7.30) holds on $\mathcal{W}_N = \mathcal{W}_{N,M} \subset \mathcal{K}$, and Theorem 7.23 implies that

$$\mathbb{E}\sum_{u\in\mathcal{C}_{N,M}} \varphi_N(H_N,u) \le \mathsf{K}_r^{1/r} q_N^{1-1/r}\,\mathbb{E}|\mathcal{C}_{N,M}| \le C_M \mathsf{K}_r^{1/r} q_N^{1-1/r}.$$

By Theorem 7.17, $\mathcal{C}_{N,M}$ is almost surely finite, and for each such field the union bound under $\mathbb{P}^{H_N}_{\nu_N}$ bounds the probability on the left of (7.33) by $\sum_{u\in\mathcal{C}_{N,M}} \varphi_N(H_N,u)$. On the Borel set of Morse fields this probability is Borel, by the enumeration of critical points in the proof of Proposition 7.16 together with Lemma 7.21. Defining it to be zero on

the null set of non-Morse fields gives a Borel representative, which is sufficient for its expectation. Taking expectations proves (7.33). For a subfamily selected by a weight θ, use the last assertion of Theorem 7.23. The constant K_r depends only on b_Q, g, k_+, and r, the numbers q_N depend on $\mathcal{K}$, a, and γ, and neither depends on M; the half-width M enters only through C_M and N_M. □

Note that $\mathbb{E}|\mathcal{C}_{N,M}| < \infty$ for each N by Corollary 7.18, and hypothesis (3) is equivalent to $\limsup_N \mathbb{E}|\mathcal{C}_{N,M}| < \infty$. For a non-pure mixture, hypothesis (2) holds, for instance, when $\mathcal{K} = [h_-, h_+] \times [k_-, k_+]$, $U \subset [k_-, k_+]$, and m_N/N converges to a point of (h_-, h_+). In the pure case, replace this rectangle by its intersection with $\{(h,k) : k = ph\}$, as in the definition of $\mathcal{W}_{N,M}$. By Markov's inequality, (7.33) implies that for each fixed M the probability, for the fixed field, that the diffusion started from ν_N enters the cap $C_a(u) = \{x \in S_N : |R(x,u)| \geq a\}$ of Definition 4.1 for some $u \in \mathcal{C}_{N,M}$ before time $e^{\gamma N}$ tends to zero in probability.

Remark 7.25 (Polynomial rates). Let $a_* > 0$ be given by Corollary 8.3 for the fixed ξ, Q, β, and $\mathcal{K}$. Fix $a \in (0, \min\{1, a_*\}]$, and let γ, c, C, and N_0 be given by that corollary. After increasing C to cover the finitely many dimensions $2 \leq N < N_0$, for which the left side of hypothesis (1) is at most one, hypothesis (1) of Corollary 7.24 holds with $q_N = C(N^{1-d/2} + N^{3-d}) + e^{-cN}$. Since $d \geq 4$, we have $N^{3-d} \leq N^{1-d/2}$ and $q_N \leq C' N^{-(d-2)/2}$. Then, for every fixed $M > 0$ and $r > 1$, the left side of (7.33) is at most $C_{M,r} N^{-(d-2)(1-1/r)/2}$ for $N \geq N_M$. For $r = 2$ the exponent is $(d-2)/4$. As Proposition 7.7 holds for every fixed r, every exponent below $(d-2)/2$ is available. The time exponent γ, chosen from Corollary 8.3, does not depend on M or r. These rates are used in Proposition 13.1 and in the selection and entropy statements of Section 4.

Remark 7.26 (Order of the parameters). Table 7.1 lists the parameters of Theorem 7.23 and Corollary 7.24 in the order in which they are fixed. All of them are fixed before $N \to \infty$. Only the count bound C_M and the threshold N_M depend on the window half-width M; in particular, the time exponent γ is chosen before M.

Table 7.1. Parameters in the determinant-weighted transfer, in the order in which they are fixed.

Step	Parameter	Depends on
1	mixture ξ, radius Q, temperature β	—
2	rectangle of marks $\mathcal{K}$, gap g, bound k_+	ξ, Q
3	moment order $r > 1$; constant K_r	b_Q, g, k_+, r
4	cap parameter a, time exponent γ, rate q_N	ξ, Q, β, $\mathcal{K}$, a
5	window half-width M; count bound C_M, threshold N_M	all of the above, (m_N), U, M
6	dimension $N \to \infty$	

SECTION 8

Cap hitting estimates under critical-point conditioning

In this section, we prove Theorem 8.2, the hitting estimate for the caps under critical-point conditioning. It concerns the original Langevin diffusion started from the uniform law ν_N, in a field conditioned to have a critical point of the center field H_Q at a prescribed point u, with prescribed value and radial derivative. The theorem states that this diffusion reaches the set $\{x : |R(x,u)| \geq a\}$ before time $e^{\gamma N}$ with averaged probability at most $C(N^{1-d/2} + N^{3-d}) + e^{-cN}$, for every finite mixture of minimum degree $d \geq 4$. Corollary 8.3 makes the estimate uniform over a compact rectangle of marks, with one time exponent. This corollary is the form of the conditioned hitting estimate used in Proposition 13.1, where it is combined with the determinant-weighted transfer of Theorem 7.23 and the separation criterion of Proposition 4.5.

The section is self-contained. From earlier sections, it uses only the model (Definition 3.1), the construction of the dynamics (Lemma 3.5), and the rotation covariance of the path kernels (Lemma 4.2). It does not use the static results of Section 5, the counts of critical points and the concentration of the Gibbs measure on bands in Sections 9 to 12, or the Kac–Rice formula of Section 7, and it needs no lower bound on the Hessian at the critical point. The spherical harmonic computations, the concentration estimate on the sphere, and the comparison of resolvents that we need are proved here.

Standing notation. Throughout this section, $\xi(r) = \sum_{p\in\mathcal{P}} c_p r^p$ is a finite mixture with $c_p > 0$ for $p \in \mathcal{P}$, minimum degree $d = \min \mathcal{P} \geq 4$, and maximum degree $\bar{p} = \max \mathcal{P}$. We realize $H_N = \sum_{p\in\mathcal{P}} H_{N,p}$, where the $H_{N,p}$ are independent centered Gaussian homogeneous polynomials of degree p on $\mathbb{R}^N$ with $\mathbb{E}H_{N,p}(x)H_{N,p}(y) = Nc_pR(x,y)^p$. For fixed N, the field is a function of the finite-dimensional Gaussian vector J of the coefficients of these polynomials. We write H_J when this dependence matters. The covariance identity $\mathbb{E}H_N(x)H_N(y) = N\xi(R(x,y))$, with $R(x,y) = x\cdot y/N$, holds for all $x, y \in \mathbb{R}^N$. We fix $Q \in (0,1]$ and $\beta \geq 0$, and recall the center field and the radial derivative,

$$\begin{aligned} H_Q(u) &= H_N(\sqrt{Q}\,u) = \sum_{p\in\mathcal{P}} Q^{p/2} H_{N,p}(u), \\ K_Q(u) &= \frac{\mathrm{d}}{\mathrm{d}t} H_N(t\sqrt{Q}\,u)\Big|_{t=1} = \sum_{p\in\mathcal{P}} pQ^{p/2} H_{N,p}(u), \end{aligned} \tag{8.1}$$

and the parameters $\xi(Q)$, $a_Q = Q\xi'(Q)$, $b_Q = Q^2\xi''(Q)$, and

$$\Sigma_Q = \begin{pmatrix} \xi(Q) & a_Q \\ a_Q & a_Q + b_Q \end{pmatrix}.$$

For $u \in S_N$ we write $G_u = \nabla_{S_N} H_Q(u)$, a vector in the tangent space $u^\perp$. We write ∇_E for the Euclidean gradient, $\nabla = \nabla_{S_N}$ for the gradient in the induced metric of S_N, and $\Pi_x = I - xx^{\mathsf{T}}/N$ for the orthogonal projection onto $x^\perp$, so that $\nabla H = \Pi_x \nabla_E H$ at $x \in S_N$. The diffusion with Hamiltonian H has generator $L_{\beta,H} = \Delta_{S_N} + \beta \nabla H \cdot \nabla$ and invariant law π^H, and $\mathbb{P}^H_{\nu_N}$ denotes its path law started from ν_N. For $0 < a < 1$ and $u \in S_N$, we write

$$C_a(u) = \{x \in S_N : |R(x,u)| \geq a\}$$

for the union of the two antipodal caps of latitude a around $\pm u$, and $I(a) = -\frac{1}{2}\log(1 - a^2)$.

We prescribe the values $H_Q(u) = Nh$ and $K_Q(u) = Nk$ and call the pair (h,k) of the energy and the radial derivative per spin the *marks*. We fix bounds $\bar{h}, \bar{k} > 0$ and call a pair (h,k) *admissible* if $|h| \leq \bar{h}$ and $|k| \leq \bar{k}$, and, when ξ is pure of degree p, also $k = ph$. Unless stated otherwise, C and c denote positive constants that depend only on ξ, Q, β, $\bar{h}$, and $\bar{k}$, and that may change from line to line. They never depend on N, u, (h,k), a point of S_N, or the realization of the field on the regularity event of Lemma 8.6.

8.1 The main estimate

We first define the conditioned field. When ξ is a non-pure mixture, Σ_Q is nonsingular (see Proposition 8.7), and we define polynomials $\theta_1, \theta_2, \theta_3$ by

$$\big(\theta_1(r), \theta_2(r)\big) = \big(\xi(\sqrt{Q}\,r),\ \sqrt{Q}\,r\,\xi'(\sqrt{Q}\,r)\big)\Sigma_Q^{-1}, \qquad \theta_3(r) = \frac{\sqrt{Q}\,\xi'(\sqrt{Q}\,r)}{a_Q}. \tag{8.2}$$

When $\xi = c_p r^p$ is pure, we set instead $\theta_1(r) = Q^{-p/2} r^p$, $\theta_2 = 0$, and $\theta_3(r) = Q^{-p/2} r^{p-1}$; the last formula agrees with (8.2).

Definition 8.1 (The conditioned field)**.** Let $u \in S_N$, and for $x \in \mathbb{R}^N$ let $r = x \cdot u/N$. The *Gaussian regression projection* of H_N at u is the random polynomial

$$D_u(x) = \theta_1(r) H_Q(u) + \theta_2(r) K_Q(u) + \theta_3(r)\, G_u \cdot x. \tag{8.3}$$

For an admissible pair (h,k), the *conditioned field* is

$$H^{h,k,u} = H_N + \delta_{h,k,u}, \qquad \delta_{h,k,u}(x) = N\big(\theta_1(r)h + \theta_2(r)k\big) - D_u(x). \tag{8.4}$$

We write $\mathbb{E}_{h,k,u}$ for the expectation under the law of $H^{h,k,u}$.

We show in Proposition 8.7 below that the law of $H^{h,k,u}$ is a version of the conditional law of H_N given $H_Q(u) = Nh$, $K_Q(u) = Nk$, and $G_u = 0$. In the pure case, $K_Q(u) = pH_Q(u)$ identically, and the conditioning is on $H_Q(u) = Nh$ and $G_u = 0$. This is the version given by Gaussian regression, and it depends continuously on (h, k, u).

Theorem 8.2 (Cap hitting bound from the uniform initial law)**.** *Let ξ be a finite mixture of minimum degree $d \geq 4$, and fix $Q \in (0, 1]$, $\beta \geq 0$, and $\bar{h}, \bar{k} > 0$. Then there exists $a_* > 0$ such that, for every $a \in (0, a_*]$, there exist $\gamma, c, C > 0$ and N_0 with the following property. For every $N \geq N_0$, every $u \in S_N$, and every admissible pair (h, k),*

$$\mathbb{E}_{h,k,u}\Big[\mathbb{P}^H_{\nu_N}\Big(\sup_{0\leq t\leq e^{\gamma N}} |R(X_t, u)| \geq a\Big)\Big] \leq C\big(N^{1-d/2} + N^{3-d}\big) + e^{-cN}. \tag{8.5}$$

The constant a_* depends only on $\xi, Q, \beta, \bar{h}, \bar{k}$, and the constants γ, c, C, N_0 depend only on these and on a.

In (8.5), the field H has the law of $H^{h,k,u}$, and X is the diffusion with generator $L_{\beta,H}$; we assume that its initial point and its driving noise are independent of H. The event in (8.5) is the event that the path enters $C_a(u)$ by time $e^{\gamma N}$. Note that both polynomial terms tend to zero, since $d \geq 4$.

Corollary 8.3 (Uniformity over a rectangle of marks)**.** *Let ξ, Q, and β be as in Theorem 8.2, and let $\mathcal{K} \subset \mathbb{R}^2$ be compact. Then there exists $a_* > 0$ such that, for every $a \in (0, a_*]$, there exist $\gamma, c, C > 0$ and N_0, depending only on $\xi, Q, \beta, \mathcal{K}, a$, such that for every $N \geq N_0$,*

$$\sup_{u\in S_N}\sup_{(h,k)\in\mathcal{K}} \mathbb{E}_{h,k,u}\Big[\mathbb{P}^H_{\nu_N}\Big(\sup_{0\leq t\leq e^{\gamma N}} |R(X_t, u)| \geq a\Big)\Big] \leq C\big(N^{1-d/2} + N^{3-d}\big) + e^{-cN}. \tag{8.6}$$

When ξ is pure of degree p, the supremum is over $(h, k) \in \mathcal{K}$ with $k = ph$. If the admissible set of marks is empty, the left side is taken to be zero.

Proof. Fix $\bar{h}, \bar{k}$ with $\mathcal{K} \subset [-\bar{h}, \bar{h}] \times [-\bar{k}, \bar{k}]$ and apply Theorem 8.2. Its constants depend only on the bounds $\bar{h}, \bar{k}$ on the marks, equivalently on the bounds $|H_Q(u)| \leq N\bar{h}$ and $|K_Q(u)| \leq N\bar{k}$ on the conditioned values, and on ξ, Q, β, a. □

Remark 8.4 (Scope). In Proposition 13.1, Corollary 8.3 is applied to a compact rectangle of marks that contains the marks of every fixed energy window $[m_N - M, m_N + M]$. Its constants, and in particular the time exponent γ, do not depend on the half-width M of the window chosen later, as long as the marks of that window lie in $\mathcal{K}$. On the rectangle used there, k also stays at a fixed distance above $2\sqrt{b_Q}$, which the determinant moments of Section 7 require; the conditional law of $\nabla^2 H_Q(u)$ under $\mathbb{E}_{h,k,u}$ is computed in that section. The proof below does not use this condition. We do not obtain uniformity as $Q \downarrow 0$, as the coefficients of ξ degenerate, or as a shrinks with N. For minimum degree $d = 3$, the term N^{3-d} does not vanish, and our argument yields no hitting estimate.

8.2 The mechanism and the origin of the errors

We now explain why the dynamics started from ν_N cannot reach the caps around a critical point of H_Q with a large value within time $e^{\gamma N}$, and where the two polynomial errors in (8.5) come from. The discussion in this section is informal; the precise statements and proofs follow in the next sections.

The unconditioned field. We begin with the case where the field is not conditioned and the axis u is uniform on S_N and independent of H_N. A uniform point has overlap of order $N^{-1/2}$ with u, and $\nu_N(C_a(u)) \le e^{-NI(a)/2}$ for large N. By rotation invariance, the Gibbs mass of $C_a(u)$, averaged over the field and the axis, is also exponentially small. To turn this into a bound that holds for the whole time interval $[0, e^{\gamma N}]$, we introduce a resolvent. Let $\kappa \ge 0$ be a killing rate supported near $C_a(u)$, and let τ be the killing time, that is, the first time at which $\int_0^t \kappa(X_v)\,\mathrm{d}v$ exceeds an independent unit exponential variable. We will see that $\mathbb{E}e^{-\lambda\tau}$ is controlled by $\lambda^{-1}\int \kappa\,\mathrm{d}\pi^H$, together with a cost for the initial law. The relative entropy of ν_N with respect to π^H is of order N, and it enters only through a bound on the Fisher information of the initial density after it is smoothed by the resolvent of the unperturbed generator. With $\lambda = e^{-\gamma N}$, the term $\lambda^{-1}\int \kappa\,\mathrm{d}\pi^H$ is of order $e^{\gamma N - NI(a)/2}$, which is exponentially small when γ is small.

The conditioning and the axis average. Conditioning u to be a critical point of H_Q with a large value changes the whole field, and $H^{h,k,u} = H_N + \delta_{h,k,u}$. The correction is of order N near u, so we cannot treat it as a small perturbation of H_N. The observation that makes a comparison possible is that the correction vanishes to high order at the equator of u. The functions θ_1 and θ_2 are of order $|r|^d$, and θ_3 is of order $|r|^{d-1}$, so that

$$|\delta_{h,k,u}(x)| \le CN|r|^d + C\sqrt{N}|r|^{d-1}|V_x(u)|, \qquad V_x(u) = N^{-1/2}\,G_u \cdot x,$$

where $r = R(x,u)$. A fixed point x has overlap of order $N^{-1/2}$ with a uniform axis u, and $V_x(u)$ is of order one. This means that the correction is of order $N^{1-d/2}$ for typical pairs (x,u), and this is the source of both polynomial errors. We therefore compare the conditioned dynamics at a uniform axis with the unconditioned dynamics, using a smooth truncation δ_α of $\delta_{h,k,u}$ at a small fixed scale α (see (8.27) and the last paragraph of this section), through the density e^w with $w = \beta\delta_\alpha$. The truncation is necessary: near u, the correction itself is of order N, and the axis average of $e^{\beta\delta_{h,k,u}}$ can be exponentially large. For a sufficiently small fixed α, we will see that only two moments of e^w are needed, averaged over the axis for each fixed point x and each fixed field satisfying the regularity bounds of Lemma 8.6:

$$\eta = \mathbb{E}_u|e^w - 1| = O(N^{1-d/2}), \qquad \epsilon = \mathbb{E}_u\big[e^w(1-e^{-w})^2\big] = O(N^{2-d}).$$

The correlated gradient. This raises a difficulty, since the point x in these averages may depend on the field, and G_u is a function of the same field. We cannot treat G_u as a

Gaussian vector independent of x. To overcome this difficulty, we fix the field and x and average u over a latitude $\{u : R(x,u) = r\}$. We show in Lemma 8.9 that V_x concentrates on each latitude, by the logarithmic Sobolev inequality on the sphere (Lemma 8.8), and that its latitude mean can be computed exactly by the Funk–Hecke formula. The top spherical harmonic component of each $H_{N,p}$ contributes $\sqrt{N}|r|^{p-1}$ to this mean, and the lower components contribute a bounded amount. Note that these lower components have covariance weights $\omega_{p,j,N} = O(N^{-j})$, and hence pointwise variances of order N^{1-j} (Lemma 8.5); they must be kept.

The time horizon. The main obstacle is the length $e^{\gamma N}$ of the time interval. A comparison of path laws, for example by Girsanov's formula, costs an amount proportional to the time. To overcome this obstacle, we prove the entropy–resolvent comparison of Lemma 8.11. It compares the Laplace transform $\mathbb{E}e^{-\lambda\tau}$ of the killing time for the perturbed process with an integrated killing rate, in which the initial density is smoothed by the resolvent of the unperturbed generator, at the cost

$$\underbrace{\eta + \sqrt{\epsilon}}_{O(N^{1-d/2})} \quad + \quad \underbrace{\epsilon\,(1 + \mathcal{H}_0)}_{O(N^{3-d})}, \tag{8.7}$$

where $\mathcal{H}_0 = \int (1+f_0)\log(1+f_0)\,\mathrm{d}\mu$, and f_0 is the density of the initial law with respect to the reference measure μ of Lemma 8.11. In our application, μ is the joint law of the field and of a point drawn from the Gibbs measure, $f_0\mu$ is the joint law of the field and of the initial point, and $\mathcal{H}_0 = O(N)$ by (8.51). Since $\int f_0\,\mathrm{d}\mu = 1$ and the function $a \mapsto (1+a)\log(1+a) - a\log a$ is nonnegative and concave, Jensen's inequality shows that $\mathcal{H}_0$ exceeds $\int f_0 \log f_0\,\mathrm{d}\mu$, the entropy of ν_N relative to the Gibbs measure averaged over the field, by at most $2\log 2$. The point is that this cost is not multiplied by the time: the factor $e^{\gamma N}$ appears only in the integrated killing rate, which is exponentially small. At $d = 4$, both terms in (8.7) are $O(N^{-1})$; at $d = 3$, the second is of order one. In the proof of the comparison, we test the resolvent equation of the killed diffusion against a weighted function, and a commutator identity cancels a cross term that would otherwise be of order one.

Truncation and the return to the true process. The comparison lemma requires a bounded perturbation whose size is small compared with the entropy cost of the caps. For this reason, we truncate $\delta_{h,k,u}$ by smooth cutoffs at scale α in both $r = R(x,u)$ and $s = G_u \cdot x/N$. On the support of the cutoffs, the truncated correction is at most $CN\alpha^d$, while a cap of latitude $\alpha/4$ costs $NI(\alpha/4) \geq N\alpha^2/32$. Since $d > 2$, we can choose α small and then γ small, so that the integrated killing rate remains exponentially small. We then return to the true process using two facts. First, after a hit of the cap $C_{\alpha/2}(u)$, the killing rate equals one during a time interval of fixed length, except with exponentially small probability. Second, the true and truncated processes coincide until $|r|$ or $|s|$ reaches

α. Reaching $|r| = \alpha$ requires crossing $C_a(u)$ with $a = \alpha/2$ first. Reaching $|s| = \alpha$ is exponentially unlikely. Indeed, G_u is exactly the Gaussian vector that the conditioning removes. It is independent of the conditioned field and of the whole true trajectory, and s_t is Gaussian with variance at most a_Q/N at each fixed time.

Organization. In Section 8.3, we compute the harmonic decomposition of the field and prove regularity bounds with exponential tails. Section 8.4 constructs the conditioned field by regression and bounds the correction. Section 8.5 proves the latitude concentration and the two moment bounds for e^w. Section 8.6 states and proves the entropy–resolvent comparison for abstract diffusions. Section 8.7 applies it to the truncated process, and Section 8.8 shows that the true process does not reach the cutoff boundary $\{|G_u \cdot x|/N = \alpha\}$. We assemble the proof of Theorem 8.2 in Section 8.9. Table 8.1 lists the parameters in the order in which they are fixed; all of them are fixed before $N \to \infty$.

Table 8.1. Parameters in the proof of Theorem 8.2, in the order in which they are chosen.

Parameter	Role	Depends on
$\xi, Q, \beta, \bar{h}, \bar{k}$	model, radius, temperature, mark bounds	given
C_0	regularity constant (Lemma 8.6)	ξ, Q
α_1	largest cutoff scale (Proposition 8.12)	$\xi, Q, \beta, \bar{h}, \bar{k}$
$a \in (0, \alpha_1/2]$, $\alpha = 2a$	latitude of the caps, cutoff scale	chosen
γ	time exponent	the above and a
C, c, N_0	constants in (8.5)	the above and a

8.3 Harmonic coefficients and regularity of the field

The latitude mean of the tangent gradient term $G_u \cdot x$ involves every spherical harmonic component of $H_{N,p}$, not only the top one. In this section, we compute the variances of these components exactly and prove regularity bounds at their correct scales.

Let $N \geq 3$ and $\alpha_N = (N-2)/2$. The Gegenbauer polynomials $C_\ell^{(\alpha_N)}$ are defined by $(1 - 2rt + t^2)^{-\alpha_N} = \sum_{\ell \geq 0} C_\ell^{(\alpha_N)}(r) t^\ell$. Expanding the left side by the binomial series gives

$$C_\ell^{(\alpha_N)}(r) = \sum_{j=0}^{\lfloor \ell/2 \rfloor} \frac{(-1)^j (\alpha_N)_{\ell-j}}{j!\,(\ell-2j)!} (2r)^{\ell-2j}, \qquad C_\ell^{(\alpha_N)}(1) = \frac{(2\alpha_N)_\ell}{\ell!},$$

where $(t)_j = t(t+1)\cdots(t+j-1)$ and $(t)_0 = 1$; see [62, Section 18.5(iii), (18.5.10)]. We define the normalized zonal polynomials

$$Z_{\ell,N}(r) = \frac{C_\ell^{(\alpha_N)}(r)}{C_\ell^{(\alpha_N)}(1)} = \frac{\ell!}{(2\alpha_N)_\ell} \sum_{j=0}^{\lfloor \ell/2 \rfloor} \frac{(-1)^j (\alpha_N)_{\ell-j}}{j!\,(\ell-2j)!} (2r)^{\ell-2j}, \tag{8.8}$$

so that $\mathsf{Z}_{0,N} = 1$ and $\mathsf{Z}_{\ell,N}(1) = 1$. From the generating function, we also obtain the differential equation

$$(1-r^2)\mathsf{Z}'' - (N-1)r\mathsf{Z}' + \ell(\ell+N-2)\mathsf{Z} = 0, \qquad \mathsf{Z} = \mathsf{Z}_{\ell,N}. \tag{8.9}$$

Lemma 8.5 (Harmonic coefficients)**.** *Let $N \geq 3$.*

1. *For every integer $p \geq 0$,*

$$r^p = \sum_{j=0}^{\lfloor p/2 \rfloor} \omega_{p,j,N}\, \mathsf{Z}_{p-2j,N}(r), \qquad \omega_{p,j,N} = \frac{p!\,\Gamma(\alpha_N)\,(p-2j+\alpha_N)}{2^p j!\,\Gamma(p-j+\alpha_N+1)} \cdot \frac{(2\alpha_N)_{p-2j}}{(p-2j)!}. \tag{8.10}$$

   The coefficients $\omega_{p,j,N}$ are positive and sum to one.

2. *For fixed p and $0 \leq j \leq p/2$, as $N \to \infty$,*

$$\omega_{p,j,N} = \frac{p!}{2^j j!\,(p-2j)!} N^{-j}\big(1 + O_p(N^{-1})\big); \quad \textit{in particular } \omega_{p,j,N} \leq C_p N^{-j}. \tag{8.11}$$

3. *For fixed ℓ and m, the function $\mathsf{Z}_{\ell,N}(r) - r^\ell$ and its first m derivatives are bounded by $C_{\ell,m}/N$ on $[-1,1]$. In particular, for $\ell \geq 1$ and $|r| \leq 1$,*

$$|\mathsf{Z}'_{\ell,N}(r)| \leq C_\ell\big(|r|^{\ell-1} + N^{-1}\big). \tag{8.12}$$

4. *Let $H_{N,p}|_{S_N} = \sum_j H_{p,p-2j}$ be the decomposition into spherical harmonics, where $H_{p,\ell}$ has degree ℓ. The fields $H_{p,p-2j}$, for $p \in \mathcal{P}$ and $0 \leq j \leq p/2$, are independent centered Gaussian fields on S_N, and*

$$\mathbb{E} H_{p,p-2j}(x) H_{p,p-2j}(y) = N c_p\, \omega_{p,j,N}\, \mathsf{Z}_{p-2j,N}(R(x,y)), \qquad x, y \in S_N. \tag{8.13}$$

   For example, at $p = 4$ we have

$$r^4 = \frac{(N-1)(N+1)}{(N+2)(N+4)}\mathsf{Z}_{4,N}(r) + \frac{6(N-1)}{N(N+4)}\mathsf{Z}_{2,N}(r) + \frac{3}{N(N+2)}.$$

In particular, the component of degree two of a quartic field has a variance of order N^{-1} relative to the top component.

Proof. We first recall the uniqueness of zonal harmonics. Fix $z \in S_N$. The Laplacian Δ_{S_N} acts on functions of $r = R(x,z)$ by $g \mapsto N^{-1}\big[(1-r^2)g'' - (N-1)rg'\big]$. Then (8.9) implies that the function $x \mapsto \mathsf{Z}_{\ell,N}(R(x,z))$ is an eigenfunction of Δ_{S_N} with eigenvalue $-\ell(\ell+N-2)/N$; since it is the restriction of a polynomial of degree ℓ, it is a spherical harmonic of degree ℓ, and it is invariant under the rotations fixing z. Conversely, a spherical harmonic of degree ℓ invariant under these rotations is a polynomial in $R(x,z)$ of degree at most ℓ that satisfies (8.9). The polynomial solutions of (8.9) of degree at most ℓ form a one-dimensional space, since the coefficients are determined recursively by the leading one. Such a harmonic is a scalar multiple of $\mathsf{Z}_{\ell,N}(R(\cdot,z))$.

The expansion (8.10). By (8.8), $\omega_{p,j,N} Z_{p-2j,N} = p!\,\Gamma(\alpha_N)(p-2j+\alpha_N)\big(2^p j!\,\Gamma(p-j+\alpha_N+1)\big)^{-1} C_{p-2j}^{(\alpha_N)}$. Inserting the explicit sum for $C_{p-2j}^{(\alpha_N)}$, we find that the coefficient of r^{p-2s} in the right side of (8.10) is

$$\frac{p!}{2^{2s}(p-2s)!} \sum_{j=0}^{s} \frac{(-1)^{s-j}(p-2j+\alpha_N)\,\Gamma(p-s-j+\alpha_N)}{j!\,(s-j)!\,\Gamma(p-j+\alpha_N+1)}.$$

For $s = 0$ this equals one. For $s \geq 1$, recall the beta integral

$$\frac{\Gamma(p-s-j+\alpha_N)}{\Gamma(p-j+\alpha_N+1)} = \frac{1}{s!} \int_0^1 t^{p-s-j+\alpha_N-1}(1-t)^s \,\mathrm{d}t.$$

Writing $y = 1/t$, the binomial theorem and its derivative in y give

$$\sum_{j=0}^{s} \frac{(-1)^{s-j}(p+\alpha_N-2j)\,t^{-j}}{j!\,(s-j)!} = \frac{t^{-s}(1-t)^{s-1}}{s!}\big[(p+\alpha_N)(1-t)-2s\big].$$

Inserting these two identities, we find that the sum over j equals $(s!)^{-2}$ times the integral over $[0,1]$ of

$$t^{p-2s+\alpha_N-1}(1-t)^{2s-1}\big[(p+\alpha_N)(1-t)-2s\big] = \frac{\mathrm{d}}{\mathrm{d}t}\Big[t^{p-2s+\alpha_N}(1-t)^{2s}\Big].$$

Both boundary values vanish, because $p-2s+\alpha_N > 0$ and $2s > 0$. This proves (8.10). The coefficients are positive by their formula, and evaluating (8.10) at $r = 1$ shows that they sum to one.

Asymptotics. For fixed p and j, each gamma ratio and Pochhammer symbol in $\omega_{p,j,N}$ is a finite product of factors $\alpha_N + c$, or of factors $2\alpha_N + c$, with fixed constants c, or the reciprocal of such a product. The substitution $\alpha_N + c = \alpha_N(1 + c/\alpha_N)$ gives

$$\omega_{p,j,N} = \frac{p!}{2^{2j} j!\,(p-2j)!}\,\alpha_N^{-j}\big(1 + O_p(\alpha_N^{-1})\big),$$

which is (8.11), since $\alpha_N^{-j} = 2^j N^{-j}(1 + O_j(N^{-1}))$. In the same way, the leading coefficient of $Z_{\ell,N}$ is $2^\ell(\alpha_N)_\ell/(2\alpha_N)_\ell = 1 + O_\ell(N^{-1})$, and its coefficient of $r^{\ell-2j}$ is $O_\ell(N^{-j})$ for $j \geq 1$. Differentiating this polynomial of fixed degree yields (3).

The harmonic components. The decomposition of $H_{N,p}|_{S_N}$ into spherical harmonics is given by linear projections, and the components are jointly Gaussian with the original fields. Their covariances are obtained by applying the projections in both variables of the kernel $Nc_pR(x,y)^p = Nc_p \sum_j \omega_{p,j,N} Z_{p-2j,N}(R(x,y))$. With one variable fixed, $Z_{\ell,N}(R(x,y))$ is a spherical harmonic of degree ℓ in the other variable. Harmonics of different degrees are orthogonal in $L^2(\nu_N)$, as they are eigenfunctions of Δ_{S_N} with the

distinct eigenvalues $-\ell(\ell+N-2)/N$. The projection onto degree ℓ' in both variables keeps the term with $\ell=\ell'$ and removes the others. This proves (8.13), and it shows that components of different degrees, or of different p, are uncorrelated. Being jointly Gaussian, they are independent. □

The next lemma is a refinement of Lemma 3.2 for the finite mixtures of this section: it also controls the lower harmonic components, at their own scale. For a component $H_{p,\ell}$, we work with its harmonic homogeneous extension of degree ℓ to $\mathbb{R}^N$.

Lemma 8.6 (Regularity with exponential tails)**.** *There exist C_0 and N_0, depending only on ξ and Q, and events Ω_N determined by the coefficient vector J, such that $\mathbb{P}(\Omega_N^c)\le e^{-N}$ for $N\ge N_0$, each Ω_N is invariant under the rotations $H\mapsto H\circ g^{-1}$, $g\in O(N)$, and the following holds on Ω_N.*

1. *For every $p\in\mathcal{P}$ and $m\in\{0,1,2\}$, $\sup_{x\in S_N}\|\nabla_E^m H_{N,p}(x)\|_{\mathrm{op}}\le C_0N^{1-m/2}$.*
2. *For every $p\in\mathcal{P}$ and $0\le j\le p/2$, $\sup_{x\in S_N}|H_{p,p-2j}(x)|\le C_0N^{1-j/2}$.*
3. *With $\|\cdot\|_\infty$ denoting the supremum over S_N,*

$$\begin{aligned}\|H_N\|_\infty+\|H_Q\|_\infty+\|K_Q\|_\infty&\le C_0N,\\ \|\nabla_E H_N\|_\infty+\|\nabla_E H_Q\|_\infty+\|\nabla H_N\|_\infty+\|\nabla H_Q\|_\infty&\le C_0\sqrt{N},\\ \left\|\nabla_E^2 H_N\right\|_\infty+\left\|\nabla_E^2 H_Q\right\|_\infty+\left\|\nabla^2 H_N\right\|_\infty+\left\|\nabla^2 H_Q\right\|_\infty&\le C_0,\end{aligned}\tag{8.14}$$

where the norms of second derivatives are operator norms.

Proof. We first treat a component of degree zero, which occurs as $H_{p,0}$ for even p. It is a centered Gaussian constant Z with $\operatorname{Var}Z=N\sigma^2$, where $\sigma^2=c_p\omega_{p,p/2,N}$. The Gaussian tail bound gives $\mathbb{P}(|Z|>2L\sigma N)\le 2e^{-2L^2N}$ for $L>0$, and on the complementary event $|Z|\le 2L\sigma N\le CLN^{1-p/4}$, which is the value bound required in (2) with $j=p/2$; all derivatives of positive order vanish.

Now let T be an isotropic centered Gaussian homogeneous polynomial of degree $1\le\ell\le\bar p$ on $\mathbb{R}^N$, with $\mathbb{E}T(x)^2=N\sigma^2$ for $|x|^2=N$. By homogeneity and isotropy, $\mathbb{E}T(y)^2=N\sigma^2(|y|^2/N)^\ell$ for all y. Let $\mathcal{T}$ be the symmetric ℓ-linear form with $T(y)=\mathcal{T}(y,\dots,y)$. For unit vectors $v_1,\dots,v_\ell$, the polarization identity

$$\mathcal{T}(v_1,\dots,v_\ell)=\frac{1}{2^\ell\ell!}\sum_{\varepsilon\in\{\pm1\}^\ell}\Big(\prod_i\varepsilon_i\Big)T\Big(\sum_i\varepsilon_iv_i\Big)$$

and $|\sum_i\varepsilon_iv_i|\le\ell$ give $\mathbb{E}\mathcal{T}(v_1,\dots,v_\ell)^2\le D_\ell^2\sigma^2N^{1-\ell}$, where $D_\ell=\ell^\ell/\ell!$. Let $\mathcal{N}$ be a $(4\bar p)^{-1}$-net of the unit sphere; comparing disjoint balls shows that we may take $|\mathcal{N}|\le(1+8\bar p)^N$. Replacing the arguments one at a time by net points, and using that each replacement

costs at most $(4\bar{p})^{-1}\,\|\mathcal{T}\|_{\mathrm{op}}$, we find $\|\mathcal{T}\|_{\mathrm{op}} \le \frac{4}{3}\max_{v_i\in\mathcal{N}}|\mathcal{T}(v_1,\dots,v_\ell)|$. The Gaussian tail bound and a union bound over the $(1+8\bar{p})^{\ell N}$ tuples then give, for $L>0$,

$$\mathbb{P}\big(\|\mathcal{T}\|_{\mathrm{op}} > 2LD_\ell\sigma N^{1-\ell/2}\big) \le 2\exp\big(-N\big[L^2/2-\bar{p}\log(1+8\bar{p})\big]\big). \tag{8.15}$$

For $|x|=\sqrt{N}$ and $0\le m\le \ell$, the mth derivative of T at x is $(\ell)^{\downarrow}_m\,\mathcal{T}(x,\dots,x,\cdot,\dots,\cdot)$, where $(\ell)^{\downarrow}_m = \ell(\ell-1)\cdots(\ell-m+1)$, and has operator norm at most $(\ell)^{\downarrow}_m\,\|\mathcal{T}\|_{\mathrm{op}}\,N^{(\ell-m)/2}$. On the complement of the event in (8.15), it is at most $2LD_\ell(\ell)^{\downarrow}_m\sigma N^{1-m/2}$. Derivatives of order $m>\ell$ vanish.

We use these bounds for each $H_{N,p}$, with $\sigma^2=c_p$, and for the harmonic extension of each $H_{p,p-2j}$, with $\sigma^2 = c_p\omega_{p,j,N} \le CN^{-j}$ by (8.13) and (8.11). This is a fixed finite number M_0 of polynomials, those of degree zero included. We choose L with $L^2/2 \ge \bar{p}\log(1+8\bar{p})+\log(2M_0)+2$. The union of the M_0 exceptional events then has probability at most e^{-N} for large N; any fixed exponential rate is possible by enlarging L. We define Ω_N by the resulting bounds on the suprema in (1) and (2), with a suitable C_0. These suprema are continuous functions of J, which makes Ω_N Borel, and they are invariant under rotations, since the decomposition into components commutes with rotations.

Part (3) follows by summing the components with the weights 1, $Q^{p/2}$, and $pQ^{p/2}$ of (8.1). For the spherical derivatives, recall that $\nabla T = \Pi_x\nabla_E T$ and

$$\nabla^2 T(x) = \Pi_x\nabla_E^2 T(x)\Pi_x - \frac{x\cdot\nabla_E T(x)}{N}\,I_{x^\perp},$$

and that $x\cdot\nabla_E T = \ell T$ for a homogeneous polynomial of degree ℓ. Increasing C_0 once completes the proof. □

8.4 Exact regression of the whole field

In this section, we show that the field of Definition 8.1 has the conditional law, and we bound the correction $\delta_{h,k,u}$. The regression is performed on the whole field H_N, although the conditioned quantities concern the center field at radius $\sqrt{Q}$; the dynamics are always driven by the field on S_N itself.

Proposition 8.7 (Whole-field regression)**.** *Let $u\in S_N$, and let $r = x\cdot u/N$ for $x\in\mathbb{R}^N$.*

1. *If ξ is a non-pure mixture, then $\det\Sigma_Q > 0$, and the coefficient vector of the random polynomial $H_N - D_u$ is independent of the vector $(H_Q(u), K_Q(u), G_u)$. If ξ is pure of degree p, then $K_Q(u) = pH_Q(u)$, and the coefficient vector of $H_N - D_u$ is independent of $(H_Q(u), G_u)$. In both cases, the law of $H^{h,k,u}$ is a version of the conditional law of H_N given $H_Q(u)=Nh$, $K_Q(u)=Nk$, and $G_u=0$.*
2. *For $|r|\le 1$,*

$$\begin{gathered}|\theta_1(r)|+|\theta_2(r)|\le C|r|^d, \qquad |\theta_3'(r)|\le C|r|^{d-2},\\ |\theta_1'(r)|+|\theta_2'(r)|+|\theta_3(r)|\le C|r|^{d-1}.\end{gathered} \tag{8.16}$$

3. The correction has spherical gradient

$$\nabla\delta_{h,k,u}(x) = \Big[\theta_1'(r)\big(Nh - H_Q(u)\big) + \theta_2'(r)\big(Nk - K_Q(u)\big) - \theta_3'(r)\, G_u\cdot x\Big]\frac{\Pi_x u}{N} - \theta_3(r)\Pi_x G_u \tag{8.17}$$

at $x \in S_N$. On Ω_N, for every admissible (h,k) and $x \in S_N$,

$$|\delta_{h,k,u}(x)| \le CN|r|^d + C|r|^{d-1}|G_u\cdot x|, \qquad \|\nabla\delta_{h,k,u}\|_\infty \le C\sqrt{N}. \tag{8.18}$$

Proof. We begin with the covariances. Recall that $\mathbb{E}H_N(x)H_N(y) = N\xi(x\cdot y/N)$ for all $x, y \in \mathbb{R}^N$. Differentiating in y along $t\sqrt{Q}\,u$ and along unit vectors $e \perp u$, we find, for every $x \in \mathbb{R}^N$,

$$\begin{gathered}\operatorname{Cov}\big((H_Q(u), K_Q(u))\big) = N\Sigma_Q, \qquad \operatorname{Cov}(G_u) = a_Q I_{u^\perp},\\ \operatorname{Cov}\big((H_Q(u), K_Q(u)), G_u\big) = 0, \qquad \operatorname{Cov}\big(H_N(x), G_u\big) = \sqrt{Q}\,\xi'(\sqrt{Q}\,r)\Pi_u x,\\ \operatorname{Cov}\big(H_N(x), (H_Q(u), K_Q(u))\big) = N\big(\xi(\sqrt{Q}\,r), \sqrt{Q}\,r\,\xi'(\sqrt{Q}\,r)\big).\end{gathered} \tag{8.19}$$

For example, the identity for $\operatorname{Cov}(H_N(x), G_u)$ follows by differentiating $N\xi(\sqrt{Q}\,x\cdot v/N)$ in v at $v = u$ in a direction $e \perp u$, which yields $\sqrt{Q}\,\xi'(\sqrt{Q}\,r)\,x\cdot e$. The covariance of G_u follows from $\partial_e\partial_{e'}N\xi(Qv\cdot v'/N) = Q\xi'(Q)\,e\cdot e'$ at $v = v' = u$, since $u\cdot e = u\cdot e' = 0$. The same orthogonality shows that $(H_Q(u), K_Q(u))$ and G_u are uncorrelated. With $w_p = c_pQ^p$, we have $\xi(Q) = \sum_p w_p$, $a_Q = \sum_p pw_p$, and $a_Q + b_Q = \sum_p p^2 w_p$, so that

$$\det\Sigma_Q = \Big(\sum_p w_p\Big)\Big(\sum_p p^2 w_p\Big) - \Big(\sum_p pw_p\Big)^2 > 0$$

by the Cauchy–Schwarz inequality, with equality exactly when only one degree occurs.

Consider a non-pure mixture. By (8.19), $D_u(x)$ is the orthogonal projection of $H_N(x)$ onto the span of the coordinates of $(H_Q(u), K_Q(u), G_u)$. Indeed, the coefficients of $H_Q(u)$ and $K_Q(u)$ are $N\big(\xi(\sqrt{Q}\,r), \sqrt{Q}\,r\,\xi'(\sqrt{Q}\,r)\big)(N\Sigma_Q)^{-1} = (\theta_1(r), \theta_2(r))$, the coefficient vector of G_u is $\sqrt{Q}\,\xi'(\sqrt{Q}\,r)\Pi_u x/a_Q$, and $\Pi_u x\cdot G_u = x\cdot G_u$. The family $\{(H_N - D_u)(x) : x \in \mathbb{R}^N\}$ and the vector $(H_Q(u), K_Q(u), G_u)$ are jointly Gaussian with zero cross-covariances, and they are therefore independent. A polynomial of degree at most $\bar{p}$ is determined by its values at finitely many points, and its coefficient vector is a linear function of these values. This proves the independence in (1). Since $H_N = (H_N - D_u) + D_u$ and D_u is linear in the conditioning vector, replacing $(H_Q(u), K_Q(u), G_u)$ by $(Nh, Nk, 0)$ in D_u produces a version of the conditional law; this is the field (8.4). In the pure case, $H_N = H_{N,p}$ is homogeneous of degree p, which implies that $K_Q(u) = pH_Q(u)$. We have $\operatorname{Var} H_Q(u) = Nc_pQ^p$ and $\operatorname{Cov}(H_N(x), H_Q(u)) = Nc_pQ^{p/2}r^p$. By (8.19), the regression coefficients on $H_Q(u)$ and

G_u are $Q^{-p/2}r^p$ and $Q^{-p/2}r^{p-1}$. The same argument applies, and no singular matrix is inverted.

For (2), note that θ_1, θ_2 are linear combinations of $\xi(\sqrt{Q}\,r)$ and $\sqrt{Q}\,r\,\xi'(\sqrt{Q}\,r)$, whose lowest power of r is r^d, and that θ_3 is a multiple of $\xi'(\sqrt{Q}\,r)$, whose lowest power is r^{d-1}. These are fixed polynomials of degree at most $\bar{p}$, which implies (8.16).

For (3), the identity (8.17) is the product rule, with $\nabla r = \Pi_x u/N$ and $\nabla(G_u \cdot x) = \Pi_x G_u$. On Ω_N, we have $|Nh - H_Q(u)| \le (\bar{h} + C_0)N$, $|Nk - K_Q(u)| \le (\bar{k} + C_0)N$, $|G_u| \le C_0\sqrt{N}$, and $|\Pi_x u|/N \le N^{-1/2}$. Inserting these bounds and (8.16) into (8.4) and (8.17), and using $|G_u \cdot x| \le C_0 N$, we obtain (8.18). These bounds control the conditional mean $N(\theta_1 h + \theta_2 k)$, which is of order N near u, as well as the random part of the correction. □

8.5 Latitude concentration and moments of the density ratio

In this section, we bound two moments of the density e^w of the truncated perturbation, averaged over a uniform axis, for every fixed field on Ω_N and every fixed point x. We begin with the concentration inequality on the sphere. This is the curvature criterion of Bakry and Émery [16, Corollary 2, p. 199; Example (b), p. 200] for the logarithmic Sobolev inequality, followed by the Herbst argument; see also [17, Sections 5.4 and 5.7]. The optimal constant on the sphere is due to Mueller and Weissler [121]; see also [16, p. 200]. The curvature constant used below suffices for our purposes.

Lemma 8.8 (Concentration on the sphere)**.** *Let $m \ge 2$, let σ_m be the normalized surface measure on the unit sphere $S^m \subset \mathbb{R}^{m+1}$, and let $g : S^m \to \mathbb{R}$ be L-Lipschitz for the geodesic distance. Then*

$$\int e^{t(g - \int g \,\mathrm{d}\sigma_m)} \,\mathrm{d}\sigma_m \le \exp\Big(\frac{L^2 t^2}{2(m-1)}\Big), \qquad t \in \mathbb{R}. \tag{8.20}$$

Proof. For a function $q > 0$, let $\mathrm{Ent}(q) = \int q \log q \,\mathrm{d}\sigma_m - \int q \,\mathrm{d}\sigma_m \log \int q \,\mathrm{d}\sigma_m$ and $\mathcal{I}(q) = \int |\nabla q|^2/q \,\mathrm{d}\sigma_m$. We first show that

$$\mathrm{Ent}(q) \le \frac{\mathcal{I}(q)}{2(m-1)} \tag{8.21}$$

for smooth $q > 0$. Let $q_t = e^{t\Delta} q$ be the heat flow on S^m, which stays smooth and positive. By integration by parts, we have $\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{Ent}(q_t) = -\mathcal{I}(q_t)$. Writing $v = \log q_t$, we have $\partial_t v = \Delta v + |\nabla v|^2$ and $\mathcal{I}(q_t) = \int q_t |\nabla v|^2$. Differentiating, we find

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{I}(q_t) = \int q_t \Delta |\nabla v|^2 + 2\int q_t \nabla v \cdot \nabla \Delta v + 2\int q_t \nabla v \cdot \nabla |\nabla v|^2,$$

where we moved the Laplacian in the first term. The divergence theorem for $q_t \nabla |\nabla v|^2$, with $\nabla q_t = q_t \nabla v$, shows that the third term equals $-2\int q_t \Delta |\nabla v|^2$. By the Bochner

identity $\frac{1}{2}\Delta|\nabla v|^2 = \nabla v \cdot \nabla \Delta v + |\nabla^2 v|^2 + \mathrm{Ric}(\nabla v, \nabla v)$ and $\mathrm{Ric} = (m-1)g_{S^m}$, we conclude that

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{I}(q_t) = -2\int q_t\big(|\nabla^2 v|^2 + (m-1)|\nabla v|^2\big) \le -2(m-1)\mathcal{I}(q_t).$$

Hence $\mathcal{I}(q_t) \le e^{-2(m-1)t}\mathcal{I}(q)$. By the spherical harmonic expansion, q_t converges uniformly to $\int q \,\mathrm{d}\sigma_m$, and $\mathrm{Ent}(q_t) \to 0$. Integrating in t gives (8.21).

Now let g be smooth with $|\nabla g| \le L$, and let $\psi(t) = \log \int e^{tg}\,\mathrm{d}\sigma_m$. Applying (8.21) to $q = e^{tg}$, for which $\mathcal{I}(q) \le L^2 t^2 \int e^{tg}$, we find $t\psi'(t) - \psi(t) \le L^2t^2/(2(m-1))$. The derivative of $\psi(t)/t$ is at most $L^2/(2(m-1))$, and $\psi(t)/t \to \int g\,\mathrm{d}\sigma_m$ as $t \downarrow 0$. Integrating, we obtain (8.20) for $t > 0$; the case $t < 0$ follows by applying this to $-g$. A Lipschitz function is a uniform limit of smooth functions with the same Lipschitz constant, and dominated convergence completes the proof. □

We now fix $x \in S_N$ and let u be uniform on S_N. Then $r = R(x,u)$ has density

$$p_N(r) = c_N(1-r^2)^{(N-3)/2}, \quad -1 < r < 1, \qquad c_N = \frac{\Gamma(N/2)}{\sqrt{\pi}\,\Gamma((N-1)/2)} \le \sqrt{N}, \quad (8.22)$$

by spherical coordinates and the beta integral; the bound on c_N follows from $\Gamma(z+\frac{1}{2}) \le \sqrt{z}\,\Gamma(z)$. Conditionally on r, we can write $u = rx + \sqrt{1-r^2}\,y$, where y is uniform on the latitude sphere $S_N \cap x^\perp$, which has dimension $N-2$ and radius $\sqrt{N}$. We write $\mathbb{E}_{\mathrm{lat}}$ for the expectation over y at fixed r, and define

$$V_x(u) = N^{-1/2}\, G_u \cdot x = N^{-1/2}\,\nabla H_Q(u) \cdot x. \quad (8.23)$$

Lemma 8.9 (The tangent gradient on a latitude)**.** *Let $N \ge 6$, let the field lie in Ω_N, and fix $x \in S_N$. For $-1 < r < 1$, let $m_x(r) = \mathbb{E}_{\mathrm{lat}} V_x(u)$ and $Z = V_x(u) - m_x(r)$. Then*

$$|m_x(r)| \le C\big(\sqrt{N}|r|^{d-1} + 1\big), \qquad \mathbb{E}_{\mathrm{lat}} e^{tZ} \le e^{Ct^2}, \quad t \in \mathbb{R}. \quad (8.24)$$

The point x in this lemma may be any function of the field: the lemma fixes the field and x before averaging over the axis.

Proof. We first prove the concentration. Along a unit-speed curve in S_N, the derivative of $\Pi_u \nabla_E H_Q(u)$ has norm at most $2|\nabla_E H_Q(u)|/\sqrt{N} + \big\|\nabla_E^2 H_Q(u)\big\|_{\mathrm{op}} \le 3C_0$ by (8.14). Because $|x| = \sqrt{N}$, the function V_x is $3C_0$-Lipschitz on S_N. The map $y \mapsto rx + \sqrt{1-r^2}\,y$ is 1-Lipschitz. After rescaling the latitude sphere to the unit sphere S^{N-2}, the function $y \mapsto V_x(u)$ is $3C_0\sqrt{N}$-Lipschitz. Then Lemma 8.8 with $m = N-2$ implies that $\mathbb{E}_{\mathrm{lat}} e^{tZ} \le \exp(9C_0^2 N t^2/(2(N-3))) \le e^{Ct^2}$.

We next compute the mean, keeping all harmonic components. Let Y be a spherical harmonic of degree ℓ on S_N. Its average over the rotations fixing x is a spherical harmonic of degree ℓ invariant under these rotations, with the value $Y(x)$ at x. By the uniqueness of

zonal harmonics recalled in the proof of Lemma 8.5, it equals $Y(x)\mathsf{Z}_{\ell,N}(R(x,\cdot))$. Evaluated on the latitude of overlap r, this gives $\mathbb{E}_{\mathrm{lat}}Y(u) = \mathsf{Z}_{\ell,N}(r)\,Y(x)$, the latitude-averaging form of the Funk–Hecke identity. For the uniqueness of zonal harmonics used in this argument, see also [7, Theorem 2.8, pp. 19–20]. Using (8.1) and Lemma 8.5(4), we obtain

$$\mathbb{E}_{\mathrm{lat}}H_Q(u) = \sum_{p\in\mathcal{P}} Q^{p/2} \sum_{j=0}^{\lfloor p/2\rfloor} \mathsf{Z}_{p-2j,N}(r)\, H_{p,p-2j}(x). \tag{8.25}$$

For fixed y, we have $\partial_r u = x - ry/\sqrt{1-r^2} = (x-ru)/(1-r^2)$, and $(x-ru)\cdot u = 0$. Since G_u is tangent at u, we have

$$\frac{\mathrm{d}}{\mathrm{d}r}H_Q(u) = \nabla_E H_Q(u)\cdot\frac{x-ru}{1-r^2} = \frac{G_u\cdot x}{1-r^2} = \frac{\sqrt{N}\,V_x(u)}{1-r^2}.$$

Since the law of y does not depend on r, we may differentiate (8.25) under the expectation. We find

$$m_x(r) = \frac{1-r^2}{\sqrt{N}} \sum_{p\in\mathcal{P}} Q^{p/2} \sum_{j=0}^{\lfloor p/2\rfloor} \mathsf{Z}'_{p-2j,N}(r)\, H_{p,p-2j}(x). \tag{8.26}$$

For $j=0$, Lemma 8.6(2) and (8.12) bound the term by $C\sqrt{N}|r|^{p-1} + CN^{-1/2}$. For $j\ge 1$, the component satisfies $|H_{p,p-2j}(x)| \le C_0\sqrt{N}$, and $|\mathsf{Z}'_{p-2j,N}| \le C$ on $[-1,1]$ by Lemma 8.5(3), and the term is bounded. There are finitely many terms, and $|r|^{p-1} \le |r|^{d-1}$ since $p \ge d$. This proves the bound on $m_x(r)$. □

We now define the truncation. We fix once and for all an even C^∞ function $\chi : \mathbb{R} \to [0,1]$ with $\chi = 1$ on $[-1,1]$ and $\chi = 0$ outside $[-2,2]$. For a cutoff scale $\alpha > 0$, $u, x \in S_N$, and an admissible pair (h,k), we set

$$\begin{aligned} r = R(x,u), \qquad s = \frac{G_u\cdot x}{N}, \\ \delta_\alpha(u,x) = \chi(r/\alpha)\chi(s/\alpha)\,\delta_{h,k,u}(x), \qquad w(u,x) = \beta\delta_\alpha(u,x). \end{aligned} \tag{8.27}$$

The function $\delta_\alpha(u,\cdot)$ is smooth on $\mathbb{R}^N$, and it depends on the field through $H_Q(u)$, $K_Q(u)$, and G_u.

Proposition 8.10 (Axis-averaged moments of the density ratio)**.** *There exist $\alpha_0 \in (0, 1/8]$, C_1, and N_0 such that, for every $\alpha \in (0,\alpha_0]$, every $N \ge N_0$, every field in Ω_N, every admissible (h,k), and every $x \in S_N$, with u uniform on S_N,*

$$\mathbb{E}_u\big|e^{w(u,x)} - 1\big| \le C_1 N^{1-d/2}, \tag{8.28}$$

$$\mathbb{E}_u\big[e^{w(u,x)}\big(1-e^{-w(u,x)}\big)^2\big] \le C_1 N^{2-d}. \tag{8.29}$$

Moreover, for all $u \in S_N$,

$$\sup_{x \in S_N} |\delta_\alpha(u,x)| \le C_1 N \alpha^d, \qquad \|\nabla \delta_\alpha(u,\cdot)\|_\infty \le C_1 \sqrt{N} \alpha^{d-1}. \tag{8.30}$$

The constants α_0, C_1, N_0 do not depend on α.

Proof. The proof has three steps. We first bound exponential moments on each latitude, then integrate over the overlap r, and finally verify the deterministic bounds.

Moments on a latitude. Fix $\zeta \ge 0$; we will take $\zeta = 2\beta$. Recall that $|G_u \cdot x| = \sqrt{N}|V_x(u)|$. Using (8.18), we find

$$|\delta_{h,k,u}(x)| \le C_2 N |r|^d + C_2 \sqrt{N} |r|^{d-1} |V_x(u)|. \tag{8.31}$$

Here V_x is the gradient of the sampled field itself. A centered variable with the bound (8.24) satisfies $\mathbb{P}(|Z| > v) \le 2e^{-v^2/(4C)}$ by Chernoff's inequality. It follows that $\mathbb{E}_{\mathrm{lat}} Z^4 \le C$ and $\mathbb{E}_{\mathrm{lat}} e^{t|Z|} \le 2e^{Ct^2}$ for $t \ge 0$. The Cauchy–Schwarz inequality then gives $\mathbb{E}_{\mathrm{lat}}[Z^2 e^{t|Z|}] \le Ce^{2Ct^2}$. We write $V_x = m_x + Z$, insert the bound on m_x from (8.24) into (8.31), and apply these estimates with $t = \zeta C_2 \sqrt{N} |r|^{d-1}$. We obtain

$$\mathbb{E}_{\mathrm{lat}} e^{\zeta|\delta|} \le C_\zeta e^{\Psi_\zeta(r)}, \qquad \mathbb{E}_{\mathrm{lat}}\big[V_x^2 e^{\zeta|\delta|}\big] \le C_\zeta \big(1 + N|r|^{2d-2}\big) e^{\Psi_\zeta(r)}, \tag{8.32}$$

where $\delta = \delta_{h,k,u}(x)$ and

$$\Psi_\zeta(r) = C_\zeta \big(N|r|^d + N|r|^{2d-2} + \sqrt{N}|r|^{d-1}\big).$$

In Ψ_ζ, the first term comes from the value correction, the second from the square of the random coefficient t and from the mean m_x, and the third from the bounded part of m_x.

Integration over the overlap. For $N \ge 6$, the density (8.22) satisfies $p_N(r) \le \sqrt{N} e^{-Nr^2/4}$. We choose $\alpha_0 \le 1/8$ so small that $C_\zeta\big((2\alpha_0)^{d-2} + (2\alpha_0)^{2d-4} + (2\alpha_0)^{d-3}\big) \le 1/8$. Since $\sqrt{N} \le N$ and $d \ge 4$, we then have $\Psi_\zeta(r) \le Nr^2/8$ for $|r| \le 2\alpha_0$. For every fixed $j \ge 0$ and $\alpha \le \alpha_0$, it follows that

$$\int_{|r| \le 2\alpha} p_N(r) |r|^j e^{\Psi_\zeta(r)} \, \mathrm{d}r \le \sqrt{N} \int_{\mathbb{R}} |r|^j e^{-Nr^2/8} \, \mathrm{d}r \le C_j N^{-j/2}. \tag{8.33}$$

Note that $\delta_\alpha^2 e^{\zeta|\delta_\alpha|} \le \mathbf{1}_{\{|r| \le 2\alpha\}} \delta^2 e^{\zeta|\delta|}$, since the cutoffs take values in $[0,1]$ and vanish for $|r| > 2\alpha$. Squaring (8.31), taking the latitude average with (8.32), and integrating with (8.33), we find

$$\mathbb{E}_u\big[\delta_\alpha^2 e^{\zeta|\delta_\alpha|}\big] \le C\big(N^2 N^{-d} + N\, N^{-(d-1)} + N^2 N^{-(2d-2)}\big) \le C N^{2-d}. \tag{8.34}$$

The three terms come from $N^2 r^{2d}$, from Nr^{2d-2} times the bounded part of the second moment of V_x, and from Nr^{2d-2} times Nr^{2d-2}.

We now take $\zeta = 2\beta$. The inequalities $|e^t - 1| \le |t|e^{|t|}$ and $e^t(1-e^{-t})^2 = e^t + e^{-t} - 2 \le t^2 e^{|t|}$, together with the Cauchy–Schwarz inequality, give

$$\mathbb{E}_u|e^w - 1| \le \beta\big(\mathbb{E}_u[\delta_\alpha^2 e^{2\beta|\delta_\alpha|}]\big)^{1/2}, \qquad \mathbb{E}_u[e^w(1-e^{-w})^2] \le \beta^2 \mathbb{E}_u[\delta_\alpha^2 e^{2\beta|\delta_\alpha|}].$$

Together with (8.34), this proves (8.28) and (8.29).

Deterministic bounds. Wherever the cutoff product or one of its derivatives is nonzero, we have $|r| \le 2\alpha$ and $|s| \le 2\alpha$. Since $G_u \cdot x = Ns$, (8.16) and the bounds on Ω_N imply that $|\delta_{h,k,u}| \le CN\alpha^d$ there. In (8.17), the terms with θ_1', θ_2' cost $C\sqrt{N}\alpha^{d-1}$, the term with θ_3' costs $C\alpha^{d-2}N|s|N^{-1/2} \le C\sqrt{N}\alpha^{d-1}$, and the last term costs $C\sqrt{N}\alpha^{d-1}$. Further, $|\nabla r| \le N^{-1/2}$ and $|\nabla s| = |\Pi_x G_u|/N \le C_0 N^{-1/2}$, and each differentiated cutoff contributes at most $C\alpha^{-1}N^{-1/2} \cdot N\alpha^d$. These estimates prove (8.30). □

8.6 An entropy–resolvent comparison

The next lemma does not involve the Gaussian model. It compares the killing time of a perturbed reversible diffusion, killed at a rate κ, with the integrated killing rate against a resolvent of the unperturbed diffusion. The perturbation is averaged over a frozen label, and only the two averaged moments of (8.28) and (8.29) enter. We state the lemma for diffusion Dirichlet forms, so that integration by parts with frozen variables and bounded measurable perturbations has a precise meaning. Its ingredients are standard Dirichlet form calculus [76]; the quantitative comparison below is proved here.

Setting. Let $(\mathcal{U}, \vartheta)$ and (E, μ) be probability spaces and $\bar\mu = \vartheta \otimes \mu$. We think of $u \in \mathcal{U}$ as a frozen label, that is, a parameter that does not move under the dynamics. Let $\mathcal{E}_0$ be a symmetric Dirichlet form on $L^2(\bar\mu)$ with domain $\mathcal{V}$, and let $A_0 \ge 0$ be the associated self-adjoint operator, so that $\mathcal{E}_0(g,v) = \langle A_0 g, v\rangle_{\bar\mu}$ for $g \in \mathrm{Dom}(A_0)$ and $v \in \mathcal{V}$. We assume the following.

(D1) (Conservativeness.) $1 \in \mathcal{V}$ and $\mathcal{E}_0(1,1) = 0$.

(D2) (Carré du champ.) There exists a symmetric bilinear map $(g,v) \mapsto Dg \cdot Dv$ from $\mathcal{V} \times \mathcal{V}$ to $L^1(\bar\mu)$, with $Dg \cdot Dg \ge 0$, such that $\mathcal{E}_0(g,v) = \int Dg \cdot Dv \, \mathrm{d}\bar\mu$.

(D3) (Diffusion property.) If $g, v \in \mathcal{V} \cap L^\infty$, then $gv \in \mathcal{V}$ and $D(gv) \cdot D\psi = g\, Dv \cdot D\psi + v\, Dg \cdot D\psi$ for $\psi \in \mathcal{V}$. If $g \in \mathcal{V} \cap L^\infty$ and $\varphi \in C^1(\mathbb{R})$, then $\varphi(g) \in \mathcal{V}$ and $D\varphi(g) \cdot D\psi = \varphi'(g)\, Dg \cdot D\psi$.

(D4) (Frozen label.) The resolvents $(\lambda + A_0)^{-1}$, $\lambda > 0$, and the map $(g,v) \mapsto Dg \cdot Dv$ send functions that do not depend on u to functions that do not depend on u.

By (D2), $|Dg \cdot Dv| \le (Dg \cdot Dg)^{1/2}(Dv \cdot Dv)^{1/2}$ pointwise. We write $|Dg|^2 = Dg \cdot Dg$, and identify a function on E with a function on $\mathcal{U} \times E$ that does not depend on u. In

our application, E is a product of a parameter space and S_N, the label is the axis, and $Dg \cdot Dv = \nabla g \cdot \nabla v$ with the gradient acting in the variable of S_N only.

Let w be a bounded measurable function on $\mathcal{U} \times E$. We define

$$M = \int e^w \, \mathrm{d}\bar{\mu}, \qquad \mu_w = M^{-1} e^w \bar{\mu}, \qquad \mathcal{E}_w(g, v) = \int Dg \cdot Dv \, \mathrm{d}\mu_w, \quad g, v \in \mathcal{V}.$$

Since w is bounded, the norms of $L^2(\mu_w)$ and $L^2(\bar{\mu})$ are equivalent, and $\mathcal{E}_w$ is a closed form on $L^2(\mu_w)$ with domain $\mathcal{V}$; let $A_w \geq 0$ be its operator. For a smooth w on a manifold, $-A_w = -A_0 + Dw \cdot D$. We assume that there exists a Markov process $(u, X_t)_{t \geq 0}$ on $\mathcal{U} \times E$, whose label coordinate u does not move, and whose transition semigroup acts on bounded measurable functions by e^{-tA_w}, μ_w-almost everywhere.

Lemma 8.11 (Entropy–resolvent comparison)**.** *In the setting above, suppose that for μ-almost every $x \in E$,*

$$\int \left| e^{w(u,x)} - 1 \right| \mathrm{d}\vartheta(u) \leq \eta \leq \frac{1}{2}, \qquad \int e^{w(u,x)} \left(1 - e^{-w(u,x)}\right)^2 \mathrm{d}\vartheta(u) \leq \epsilon. \tag{8.35}$$

Let $f_0 \geq 0$ be a function on E with $\int f_0 \, \mathrm{d}\mu = 1$ and $\mathcal{H}_0 = \int (1 + f_0) \log(1 + f_0) \, \mathrm{d}\mu < \infty$. Let $\kappa \geq 0$ be a bounded measurable function on $\mathcal{U} \times E$. Start the process from the law $\vartheta \otimes (f_0 \mu)$, let $\mathbf{e}$ be a unit exponential variable independent of it, and let $\tau_\kappa = \inf\{t \geq 0 : \int_0^t \kappa(u, X_v) \, \mathrm{d}v \geq \mathbf{e}\}$, with $\inf \emptyset = \infty$, be the killing time at rate κ. For $\lambda > 0$ define

$$f = \lambda(\lambda + A_0)^{-1} f_0, \qquad F = 1 + f, \qquad \mathcal{R}_\lambda = \lambda^{-1} \int \kappa F \, \mathrm{d}\mu_w. \tag{8.36}$$

Then

$$\mathbb{E} e^{-\lambda \tau_\kappa} \leq 4\eta + 5\sqrt{\epsilon} + \mathcal{R}_\lambda + 2\sqrt{\epsilon \mathcal{H}_0 \left[\mathcal{R}_\lambda + \epsilon(1 + \mathcal{H}_0)\right]}, \tag{8.37}$$

with $e^{-\lambda \cdot \infty} = 0$. In particular, for every $T > 0$,

$$\mathbb{P}(\tau_\kappa \leq T) \leq 14\big(\eta + \sqrt{\epsilon} + \epsilon(1 + \mathcal{H}_0) + \mathcal{R}_{1/T}\big). \tag{8.38}$$

The first resolvent in (8.36) acts on the unperturbed form; for unbounded f_0 it is extended to $L^1(\mu)$, where it is a contraction. No constant in (8.37) depends on $\|w\|_\infty$, on λ, or on a spectral gap, and the form need not be irreducible. The boundedness of w is used only to define $\mathcal{E}_w$ and A_w.

Proof. We first treat bounded f_0. An approximation removes this restriction at the end. The proof has four steps: an entropy bound for the resolvent of the unperturbed generator, the probabilistic meaning of the resolvent of the killed diffusion, a weighted energy estimate, and a comparison of initial laws.

Two consequences of the moment bounds. By the first bound in (8.35), we have $1-\eta \le M \le 1+\eta$. Let $k = 1 - e^{-w}$. For every nonnegative measurable Γ on E, Tonelli's theorem and (8.35) give

$$\int k^2 \Gamma \, \mathrm{d}\mu_w \le 2\epsilon \int \Gamma \, \mathrm{d}\mu, \qquad \int \Gamma \, \mathrm{d}\mu_w \le 3 \int \Gamma \, \mathrm{d}\mu. \tag{8.39}$$

Indeed, $e^w k^2 = e^w(1-e^{-w})^2$, $M \ge 1/2$, and $(1+\eta)/(1-\eta) \le 3$. All the small constants below come from (8.39), not from a pointwise bound on e^w.

The resolvent of the unperturbed generator. For bounded f_0, the representation $f = \lambda \int_0^\infty e^{-\lambda t} e^{-tA_0} f_0 \, \mathrm{d}t$ and (D1) show that $f \ge 0$, $\int f \, \mathrm{d}\mu = 1$, and $f \in \mathrm{Dom}(A_0) \cap L^\infty$. By (D4), f does not depend on u. With $F_0 = 1 + f_0$, we have $A_0 F = A_0 f = \lambda(F_0 - F)$. The function $F \ge 1$ is bounded, and (D3) shows that $\log F \in \mathcal{V}$ with $D\log F \cdot D\psi = F^{-1} DF \cdot D\psi$. Let $\mathcal{J} = \int |DF|^2 F^{-1} \, \mathrm{d}\mu$. Testing with $\log F$, we find

$$\begin{aligned} \mathcal{J} &= \langle A_0 F, \log F \rangle_\mu = \lambda \langle F_0 - F, \log F \rangle_\mu \\ &\le \lambda \Big[\int F_0 \log F_0 \, \mathrm{d}\mu - \int F \log F \, \mathrm{d}\mu \Big] \le \lambda \mathcal{H}_0. \end{aligned} \tag{8.40}$$

In the first inequality, we used the tangent line inequality $(a-b)\log b \le a \log a - b \log b - (a-b)$ for $a, b > 0$, and that F_0 and F both have integral two. In the second, we used $F \log F \ge 0$.

The resolvent of the killed diffusion. The form $\mathcal{E}_w(g,v) + \int \kappa g v \, \mathrm{d}\mu_w$ is closed on $\mathcal{V}$; let

$$h = (\lambda + A_w + \kappa)^{-1} \kappa \in \mathrm{Dom}(A_w) \subset \mathcal{V}.$$

The Duhamel formula expresses $e^{-t(A_w+\kappa)}$ as the series

$$e^{-t(A_w+\kappa)} = \sum_{n \ge 0} (-1)^n \int_{0<v_1<\cdots<v_n<t} e^{-v_1 A_w} \kappa e^{-(v_2-v_1)A_w} \kappa \cdots \kappa e^{-(t-v_n)A_w} \, \mathrm{d}v,$$

whose nth term has norm at most $\|\kappa\|_\infty^n t^n / n!$ on bounded functions. Expanding the exponential in $\mathbb{E}_{u,x}\big[e^{-\int_0^t \kappa(u,X_v)\,\mathrm{d}v} g(X_t)\big]$ and using the Markov property, we identify the two series term by term. The result is the Feynman–Kac formula $e^{-t(A_w+\kappa)} g(u,x) = \mathbb{E}_{u,x}\big[e^{-\int_0^t \kappa(u,X_v)\,\mathrm{d}v} g(X_t)\big]$ for bounded g. Conditionally on the path, τ_κ has density $\kappa(u,X_t) e^{-\int_0^t \kappa(u,X_v)\,\mathrm{d}v}$ on $[0,\infty)$. Integrating the Feynman–Kac formula for $g = \kappa$ against $e^{-\lambda t}\,\mathrm{d}t$, we obtain

$$h(u,x) = \mathbb{E}_{u,x} e^{-\lambda \tau_\kappa}. \tag{8.41}$$

Note that $0 \le h \le 1$. This is why the lemma controls the killing time itself.

The commutator. Let $g \in \mathrm{Dom}(A_0)$ not depend on u, and let $v \in \mathcal{V}$ be bounded. Then

$$\mathcal{E}_w(g, v) = \langle (1-k)A_0 g, v\rangle_{\mu_w} + \int k\, Dg \cdot Dv \,\mathrm{d}\mu_w. \tag{8.42}$$

Indeed, since $(1-k)e^w = 1$, the first term on the right equals $M^{-1}\int A_0 g\, v \,\mathrm{d}\bar{\mu} = M^{-1}\int Dg \cdot Dv \,\mathrm{d}\bar{\mu}$, and adding the second term multiplies the integrand by $1 + (e^w - 1) = e^w$. The identity (8.42) is a weak form of $A_w g = (1-k)A_0 g + D_w^*(kDg)$, where D_w^* is the adjoint in $L^2(\mu_w)$; no derivative of w or k appears in it.

The weighted energy. Let $B_h = \int F|Dh|^2 \,\mathrm{d}\mu_w$ and $A_h = \int Fh^2 \,\mathrm{d}\mu_w$, which are finite. By (D3), $Fh \in \mathcal{V}$ and $D(Fh)\cdot Dh = F|Dh|^2 + h\, DF \cdot Dh$. Testing the equation $(\lambda + A_w + \kappa)h = \kappa$ against Fh, and using $h \le 1$, we find

$$B_h + \int h\, DF \cdot Dh \,\mathrm{d}\mu_w + \int (\lambda+\kappa) Fh^2 \,\mathrm{d}\mu_w = \int \kappa Fh \,\mathrm{d}\mu_w \le \int \kappa F \,\mathrm{d}\mu_w = \lambda \mathcal{R}_\lambda. \tag{8.43}$$

The cross term is the delicate one, since $|DF|$ is not small. Using (D3), we write $h\, DF \cdot Dh = \frac{1}{2} DF \cdot D(h^2)$. Applying (8.42) with $g = F$ and $v = h^2$, and using $A_0 F = \lambda(f_0 - f)$ and $(1-k)e^w = 1$, we obtain

$$\int h\, DF \cdot Dh \,\mathrm{d}\mu_w = \frac{\lambda}{2M}\int f_0 h^2 \,\mathrm{d}\bar{\mu} - \frac{\lambda}{2}\int f h^2 \,\mathrm{d}\mu_w + \frac{\lambda}{2}\int k f h^2 \,\mathrm{d}\mu_w + \int k h\, DF \cdot Dh \,\mathrm{d}\mu_w. \tag{8.44}$$

The first term on the right is nonnegative. Since $F - f/2 \ge F/2$, the second term combined with λA_h in (8.43) leaves at least $\lambda A_h/2$. We apply (8.39) to the third and fourth terms, with $\Gamma = f$ and $\Gamma = |DF|^2/F$, which do not depend on u by (D4). Using also $h \le 1$, $\int f \,\mathrm{d}\mu = 1$, and the Cauchy–Schwarz inequality, we obtain

$$\left|\int k f h^2 \,\mathrm{d}\mu_w\right| \le \sqrt{2\epsilon A_h}, \qquad \left|\int k h\, DF \cdot Dh \,\mathrm{d}\mu_w\right| \le \sqrt{2\epsilon \mathcal{J} B_h}. \tag{8.45}$$

For the first bound, we used $\int f h^4 \,\mathrm{d}\mu_w \le A_h$. Inserting (8.44) and (8.45) into (8.43), we have

$$B_h + \frac{\lambda}{2} A_h \le \lambda \mathcal{R}_\lambda + \frac{\lambda}{2}\sqrt{2\epsilon A_h} + \sqrt{2\epsilon \mathcal{J} B_h}.$$

Using Young's inequality in the forms $\frac{\lambda}{2}\sqrt{2\epsilon A_h} \le \frac{\lambda}{4} A_h + \frac{\lambda}{2}\epsilon$ and $\sqrt{2\epsilon \mathcal{J} B_h} \le \frac{1}{2} B_h + \epsilon \mathcal{J}$, and then (8.40), we find

$$\frac{B_h}{\lambda} \le 2\mathcal{R}_\lambda + \epsilon + 2\epsilon \mathcal{H}_0. \tag{8.46}$$

The term λA_h must be kept here, since it absorbs the negative second term of (8.44). Without it, B_h/λ would only be bounded by a constant, and the last term of (8.37) below would be of order $\sqrt{\epsilon \mathcal{H}_0}$, which is $N^{(3-d)/2}$ in our application.

Comparison of initial laws. The actual initial law $\vartheta \otimes (f_0\mu)$ has density $q_0 = Me^{-w}f_0$ with respect to μ_w. Because the first moment bound in (8.35) holds for each x, we have

$$\|q_0 - f_0\|_{L^1(\mu_w)} = M^{-1}\int f_0|M - e^w|\,\mathrm{d}\bar{\mu} \le 2\int f_0\Big(|M-1| + \int |1-e^w|\,\mathrm{d}\vartheta\Big)\,\mathrm{d}\mu \le 4\eta. \tag{8.47}$$

By (8.41), $\mathbb{E}e^{-\lambda\tau_\kappa} = \langle q_0, h\rangle_{\mu_w} \le \langle f_0, h\rangle_{\mu_w} + 4\eta$. To bound the last pairing, we test the equation for h against f. Using (8.42) with $g = f$ and $v = h$, we have $\mathcal{E}_w(f,h) = \lambda\langle f_0 - f, h\rangle_{\mu_w} - \langle kA_0 f, h\rangle_{\mu_w} + \int k\,Df\cdot Dh\,\mathrm{d}\mu_w$. Inserting this into the equation $\lambda\langle h, f\rangle_{\mu_w} + \mathcal{E}_w(h,f) + \langle \kappa h, f\rangle_{\mu_w} = \langle \kappa, f\rangle_{\mu_w}$, we obtain

$$\langle f_0, h\rangle_{\mu_w} = \lambda^{-1}\langle f, \kappa(1-h)\rangle_{\mu_w} + \langle k(f_0 - f), h\rangle_{\mu_w} - \lambda^{-1}\int k\,Df\cdot Dh\,\mathrm{d}\mu_w. \tag{8.48}$$

The first term lies in $[0, \mathcal{R}_\lambda]$. For the second, by the Cauchy–Schwarz inequality and (8.39) with $\Gamma = f_0$ and $\Gamma = f$, we have $\int |k|(f_0 + f)\,\mathrm{d}\mu_w \le 2\sqrt{6\epsilon} \le 5\sqrt{\epsilon}$. For the third, note that $Df = DF$. The argument for (8.45), followed by (8.40) and (8.46), yields

$$\lambda^{-1}\Big|\int k\,Df\cdot Dh\,\mathrm{d}\mu_w\Big| \le \lambda^{-1}\sqrt{2\epsilon\mathcal{J}B_h} \le \sqrt{2\epsilon\mathcal{H}_0\,B_h/\lambda} \le 2\sqrt{\epsilon\mathcal{H}_0\big[\mathcal{R}_\lambda + \epsilon(1+\mathcal{H}_0)\big]}.$$

Combining these bounds proves (8.37) for bounded f_0. All the integrations by parts above are identities between forms, for the functions F, h, Fh, h^2, and $\log F$, which lie in $\mathcal{V}$ by (D3). We use no regularity of these functions or of w beyond membership in $\mathcal{V}$.

Unbounded initial densities. For general f_0, let $f_0^{(n)} = (f_0 \wedge n)/\int (f_0 \wedge n)\,\mathrm{d}\mu$. Then $f_0^{(n)} \to f_0$ in $L^1(\mu)$. For large n, we have $f_0^{(n)} \le 2f_0$. Since $(1+2a)\log(1+2a) \le 2(1+a)\log(1+a) + 2a$ for $a \ge 0$, dominated convergence gives $\mathcal{H}_0^{(n)} \to \mathcal{H}_0$. The resolvent $\lambda(\lambda + A_0)^{-1}$ is a contraction on $L^1(\mu)$, which implies that $f^{(n)} \to f$ in $L^1(\mu)$. Since κ and e^w are bounded, it follows that $\mathcal{R}_\lambda^{(n)} \to \mathcal{R}_\lambda$. The function h does not depend on f_0 and is bounded, so the left side of (8.37) converges as well. This proves (8.37) in general.

The time horizon. Finally, $\mathbf{1}_{\{\tau_\kappa \le T\}} \le e\cdot e^{-\tau_\kappa/T}$. Using (8.37) with $\lambda = 1/T$, together with $2\sqrt{\epsilon\mathcal{H}_0\mathcal{R}} \le \mathcal{R} + \epsilon\mathcal{H}_0$ and $2\epsilon\sqrt{\mathcal{H}_0(1+\mathcal{H}_0)} \le 2\epsilon(1+\mathcal{H}_0)$, we find that $\mathbb{P}(\tau_\kappa \le T) \le e\big(4\eta + 5\sqrt{\epsilon} + 2\mathcal{R}_{1/T} + 3\epsilon(1+\mathcal{H}_0)\big)$. This implies (8.38). □

8.7 Killing-time bound for the truncated process

We now apply Lemma 8.11 to the diffusion whose Hamiltonian is $H_N + \delta_\alpha(u,\cdot)$, with u uniform and independent of the field. Recall the constants α_0 and C_1 of Proposition 8.10.

For a smooth function $\widetilde{H}$ on $\mathbb{R}^N$, the diffusion with generator $L_{\beta,\widetilde{H}}$ is the solution of the stochastic differential equation

$$\mathrm{d}X_t = \sqrt{2}\,\Pi_{X_t}\,\mathrm{d}B_t + \Big[\beta\Pi_{X_t}\nabla_E\widetilde{H}(X_t) - \frac{N-1}{N}X_t\Big]\,\mathrm{d}t \tag{8.49}$$

in $\mathbb{R}^N$, where B is a standard Brownian motion in $\mathbb{R}^N$; compare Lemma 3.5 and (4.5). The coefficients are smooth and in particular locally Lipschitz. Let $\rho_t = |X_t|^2 - N$. Note that $x \cdot \Pi_x v = -\rho\, x \cdot v/N$ and $\operatorname{tr}(\Pi_x \Pi_x) = N - 1 + \rho^2/N^2$ for $\rho = |x|^2 - N$. Then Itô's formula gives $\mathrm{d}\rho_t = \rho_t\,(a_t\,\mathrm{d}t + b_t \cdot \mathrm{d}B_t)$ with locally bounded processes a_t and b_t. This implies that $\rho_t = 0$ for all t when $\rho_0 = 0$, so that the solution started on S_N is global and pathwise unique. When the coefficients of $\widetilde{H}$ range over a finite-dimensional parameter space, localization on bounded sets of parameters and successive approximations construct the solution as a jointly measurable function of the parameters, the initial point, and the path of B.

Proposition 8.12 (The truncated process avoids the caps)**.** *There exists $\alpha_1 \in (0, \alpha_0]$ such that, for every $\alpha \in (0, \alpha_1]$, there exist $\gamma_0, c, C > 0$ and N_0 with the following property. Let Q_α be the joint law of: the coefficient vector J of H_N, conditioned on Ω_N; an independent uniform axis $u \in S_N$; an independent initial point X_0 with law ν_N; and the solution X of* (8.49) *with $\widetilde{H} = H_J + \delta_\alpha(u, \cdot)$, driven by an independent Brownian motion. Then, for every $N \geq N_0$ and every admissible (h, k),*

$$\mathrm{Q}_\alpha\Big(\sup_{0 \leq t \leq e^{\gamma_0 N}} |R(X_t, u)| \geq \frac{\alpha}{2} \Big) \leq C\big(N^{1-d/2} + N^{3-d}\big) + e^{-cN}. \tag{8.50}$$

The same bound holds with γ_0 replaced by any $\gamma \in (0, \gamma_0]$.

Proof. The proof has three steps: we verify the hypotheses of Lemma 8.11, bound the integrated killing rate, and convert the killing at a smooth rate into a hit of the cap $C_{\alpha/2}(u)$.

The setting of the comparison lemma. Let Θ be the space of coefficient vectors, $\mathbb{P}_\Omega(\mathrm{d}J) = \mathbb{P}(\mathrm{d}J \mid \Omega_N)$, $E = \Theta \times S_N$, and

$$\mu(\mathrm{d}J, \mathrm{d}x) = \mathbb{P}_\Omega(\mathrm{d}J)\, \pi^{H_J}(\mathrm{d}x), \qquad f_0(J, x) = Z_{H_J} e^{-\beta H_J(x)},$$

where $Z_{H_J} = \int e^{\beta H_J}\, \mathrm{d}\nu_N$. The label space is $(\mathcal{U}, \vartheta) = (S_N, \nu_N)$, the space of axes. Then $f_0 \mu$ is exactly $\mathbb{P}_\Omega \otimes \nu_N$, the law of the field and the initial point. On Ω_N, we have $|H_J| \leq C_0 N$, which implies $f_0 \leq e^{2\beta C_0 N}$ and

$$\mathcal{H}_0 \leq \log\big(1 + e^{2\beta C_0 N}\big) \int (1 + f_0)\, \mathrm{d}\mu \leq C(1 + N). \tag{8.51}$$

The base form is $\mathcal{E}_0(g, v) = \int \nabla_x g \cdot \nabla_x v\, \mathrm{d}\bar{\mu}$, where $\bar{\mu} = \nu_N \otimes \mu$ and ∇_x is the gradient in the variable of S_N, with domain $\mathcal{V} = L^2\big(\nu_N \otimes \mathbb{P}_\Omega; H^1(S_N, \nu_N)\big)$. Since $e^{-2\beta C_0 N} \leq \mathrm{d}\mu / \mathrm{d}(\mathbb{P}_\Omega \otimes \nu_N) \leq e^{2\beta C_0 N}$, the form norm $\big(\|g\|^2_{L^2(\bar{\mu})} + \mathcal{E}_0(g, g)\big)^{1/2}$ on $\mathcal{V}$ is equivalent to the norm of this Hilbert space. Then $\mathcal{E}_0$ is closed. Constants lie in $\mathcal{V}$ with zero energy, the product and chain rules hold fiberwise, and the operations in (D4) act fiberwise in (u, J).

The equivalence constants grow with N, but they are never used as small constants. For fixed (u, J), the operator A_0 acts as $-L_{\beta,H_J}$, which is symmetric for π^{H_J}.

We take $w = \beta\delta_\alpha(u, x)$, a bounded function of (u, J, x), smooth in x. The conditional law of μ_w given (u, J) is the Gibbs measure of $H_J + \delta_\alpha(u, \cdot)$. For fixed (u, J), the form $\mathcal{E}_w$ is therefore the Dirichlet form of $L_{\beta,H_J+\delta_\alpha(u,\cdot)}$, and A_w acts as $-L_{\beta,H_J+\delta_\alpha(u,\cdot)}$. The solutions of (8.49), constructed for each fixed (u, J) and measurable in these parameters, provide the Markov process required by the lemma, with frozen label u and frozen parameter J. Its initial law $\vartheta \otimes (f_0\mu)$ is the initial law of Q_α. No differentiation in J is needed, since the gradient acts only in x.

For $\alpha \le \alpha_0$, Proposition 8.10 verifies (8.35) for every (J, x) with $J \in \Omega_N$, with

$$\eta = C_1 N^{1-d/2}, \qquad \epsilon = C_1 N^{2-d}, \tag{8.52}$$

and $\eta \le 1/2$ for large N.

The integrated killing rate. Let $\varphi : \mathbb{R} \to [0, 1]$ be smooth and even, with $\varphi(r) = 0$ for $|r| \le \alpha/4$ and $\varphi(r) = 1$ for $|r| \ge \alpha/3$, and let $\kappa(u, J, x) = \varphi(R(x, u))$. Since f does not depend on u and $\int F \,\mathrm{d}\mu = 2$, we integrate over the label first, bound e^w by (8.30), and use $M \ge 1/2$. For fixed x, the set of axes u with $\kappa(u, J, x) \ne 0$ is contained in $C_{\alpha/4}(x)$. By symmetry, $\nu_N(C_b(x))$ does not depend on $x \in S_N$; we denote this common value by $\nu_N(C_b)$. We obtain

$$\mathcal{R}_\lambda \le 4\lambda^{-1} e^{C_1\beta N\alpha^d}\, \nu_N\big(C_{\alpha/4}\big). \tag{8.53}$$

This step uses that (8.30) holds for every u and x, not only on average. By (8.22), for every fixed $b \in (0, 1)$ and large N,

$$\nu_N(C_b) = 2\int_b^1 p_N(r)\,\mathrm{d}r \le 2\sqrt{N}\, e^{-(N-3)I(b)} \le e^{-NI(b)/2}. \tag{8.54}$$

Since $I(b) \ge b^2/2$ and $d > 2$, there exists $\alpha_1 \in (0, \alpha_0]$ such that $C_1\beta\alpha^d \le I(\alpha/4)/8$ for all $\alpha \le \alpha_1$. We fix $\alpha \le \alpha_1$ and set $\gamma_0 = I(\alpha/4)/8$. For $\lambda = (e^{\gamma_0 N} + 1)^{-1}$, (8.53) and (8.54) imply that $\mathcal{R}_\lambda \le 8e^{-NI(\alpha/4)/4}$ for large N. Combining (8.38), (8.51), and (8.52), we obtain

$$\mathsf{Q}_\alpha\big(\tau_\kappa \le e^{\gamma_0 N} + 1\big) \le C\big(N^{1-d/2} + N^{3-d}\big) + e^{-cN}, \tag{8.55}$$

where τ_κ is the killing time at rate κ, defined with an independent unit exponential variable as in Lemma 8.11. In (8.55) we used $N^{2-d} \le N^{1-d/2}$.

From the killing time to the cap. Let $r_t = R(X_t, u)$. By (8.49), r_t has drift $-\frac{N-1}{N} r_t + \beta N^{-1} u \cdot \nabla\widetilde{H}(X_t)$ and martingale part with quadratic variation rate $2|\Pi_{X_t} u|^2/N^2 = 2(1 - r_t^2)/N \le 2/N$. On Ω_N, we have $\big\|\nabla\widetilde{H}\big\|_\infty \le C\sqrt{N}$ by (8.14) and (8.30), and the drift is

bounded by a constant $\bar{d}_r$. Let $\tau_{\alpha/2} = \inf\{t \geq 0 : |r_t| \geq \alpha/2\}$, let $\Delta = \alpha/2 - \alpha/3 = \alpha/6$, and let $t_0 = \Delta/(4(\bar{d}_r + 1)) < 1$. We claim that, for every $T > 0$,

$$\mathsf{Q}_\alpha\Big(\{\tau_{\alpha/2} \leq T\} \cap \Big\{\inf_{\tau_{\alpha/2} \leq t \leq \tau_{\alpha/2}+t_0} |r_t| < \frac{\alpha}{3}\Big\}\Big) \leq 2\exp\Big(-\frac{N\Delta^2}{16t_0}\Big). \tag{8.56}$$

Indeed, after $\tau_{\alpha/2}$, the drift moves r_t by at most $\bar{d}_r t_0 \leq \Delta/4$ during a time t_0. A decrease of $|r_t|$ from at least $\alpha/2$ to below $\alpha/3$ therefore requires an increment of the martingale part of absolute value at least $\Delta/2$, and its quadratic variation during this interval is at most $2t_0/N$. A continuous martingale whose quadratic variation is at most v has an absolute increment at least b with probability at most $2e^{-b^2/(2v)}$, by the exponential supermartingale. Applying this after the stopping time $\tau_{\alpha/2}$, conditionally on the frozen variables, and integrating, we conclude that (8.56) holds. This also covers the case $\tau_{\alpha/2} = 0$.

On the complement of the event in (8.56), and on $\{\tau_{\alpha/2} \leq T\}$, the killing rate equals one on the whole interval $[\tau_{\alpha/2}, \tau_{\alpha/2} + t_0]$. Conditionally on the path, the integrated killing rate is at least t_0 by time $T + t_0$, and the killing occurs by that time with probability at least $1 - e^{-t_0}$. Integrating, and using (8.56), we find

$$\mathsf{Q}_\alpha(\tau_{\alpha/2} \leq T) \leq \big(1 - e^{-t_0}\big)^{-1} \mathsf{Q}_\alpha(\tau_\kappa \leq T + t_0) + 2e^{-N\Delta^2/(16t_0)}.$$

We take $T = e^{\gamma_0 N}$ and use (8.55) with $t_0 < 1$. After adjusting C, c, and N_0, this proves (8.50). The left side of (8.50) decreases when γ_0 is decreased. □

8.8 The cutoff in the gradient variable

The truncated and true processes coincide until $|r|$ or $|s|$ reaches the cutoff scale. In this section, we show that the true process is unlikely to reach the cutoff boundary $|s| = \alpha$. The argument uses the independence of G_u from the conditioned field, under the unrestricted Gaussian law.

Lemma 8.13 (The gradient cutoff is not reached)**.** *Let (h, k) be admissible. Let u be uniform on S_N and independent of H_N, let X_0 have law ν_N, and let X solve* (8.49) *with $\widetilde{H} = H^{h,k,u}$, where X_0 and the Brownian motion are independent of (H_N, u). Let $s_t = G_u \cdot X_t/N$. For every $b > 0$, there exist $\gamma_b, c_b > 0$ and N_0 such that, for $N \geq N_0$,*

$$\mathbb{P}\Big(\sup_{0 \leq t \leq e^{\gamma_b N}} |s_t| \geq b\Big) \leq e^{-c_b N}. \tag{8.57}$$

The same bound, with a smaller c_b, holds when the coefficient vector of H_N is conditioned on Ω_N. The constants do not depend on (h, k).

Proof. We first work under the unrestricted law. Conditionally on u, the coefficient vector of $H_N - D_u$ is independent of G_u, by Proposition 8.7, and G_u is a centered Gaussian vector in $u^\perp$ with covariance $a_Q I_{u^\perp}$, by (8.19). The coefficients of $H^{h,k,u} = (H_N - D_u) + N(\theta_1 h + \theta_2 k)$ are a continuous function of those of $H_N - D_u$ and of u. By the measurable construction of the solutions of (8.49), the whole trajectory $(X_t)_{t \geq 0}$ is a measurable function of these coefficients, of u, of X_0, and of the Brownian motion. Conditionally on u, it is therefore independent of G_u. At each fixed time t, conditionally on u and on the trajectory, s_t is centered Gaussian with variance $a_Q |\Pi_u X_t|^2 / N^2 \leq a_Q / N$, and

$$\mathbb{P}\big(|s_t| \geq b/2\big) \leq 2e^{-Nb^2/(8a_Q)}. \tag{8.58}$$

We choose $0 < \gamma_b < b^2/(32a_Q)$ and the grid of times j/N, $0 \leq j \leq \lceil Ne^{\gamma_b N} \rceil$. A union bound in (8.58) shows that some grid value satisfies $|s_{j/N}| \geq b/2$ with probability at most e^{-cN}. We next control the intervals between grid points; here we use the regularity event once. On Ω_N, we have $|G_u| \leq C_0\sqrt{N}$ and $\big\|\nabla H^{h,k,u}\big\|_\infty \leq C\sqrt{N}$, by (8.14) and (8.18). By (8.49), s_t has drift $-\frac{N-1}{N} s_t + \beta N^{-1} G_u \cdot \nabla H^{h,k,u}(X_t)$, which is bounded by a constant $\bar{d}_s$ since $|s_t| \leq C_0$, and its martingale part has quadratic variation rate $2|\Pi_{X_t} G_u|^2 / N^2 \leq \bar{q}_s / N$. On an interval of length $1/N$, the drift moves s_t by at most $b/4$ for large N. If $|s| < b/2$ at the left endpoint and $|s|$ reaches b in the interval, the martingale increment has absolute value at least $b/4$, while its quadratic variation is at most $\bar{q}_s / N^2$. The martingale inequality used in the proof of Proposition 8.12 shows that this has probability at most $2e^{-b^2 N^2/(32\bar{q}_s)}$. This bound holds conditionally on every field in Ω_N and every axis, because the Brownian motion is independent of them. A union over $O(Ne^{\gamma_b N})$ intervals costs at most e^{-cN^2}. Adding $\mathbb{P}(\Omega_N^c) \leq e^{-N}$ once, we conclude that (8.57) holds.

Under the law conditioned on Ω_N, the probability of the event in (8.57) is at most its unrestricted probability divided by $\mathbb{P}(\Omega_N) \geq 1/2$. This proves the second assertion. The independence of G_u from the trajectory was used only under the unrestricted law. □

8.9 Proof of the main estimate

Proof of Theorem 8.2. We fix the parameters, couple the true and truncated processes, remove the regularity restriction, and finally pass from a uniform axis to a deterministic one.

Choice of parameters. We fix $\xi, Q, \beta, \bar{h}, \bar{k}$, and take C_0 and Ω_N from Lemma 8.6, and α_1 from Proposition 8.12. We set $a_* = \alpha_1/2$. Given $a \in (0, a_*]$, we set $\alpha = 2a$, so that the cap $C_{\alpha/2}(u)$ in (8.50) is $C_a(u)$. We then choose

$$\gamma = \min\{\gamma_0, \gamma_\alpha\},$$

where γ_0 is given by Proposition 8.12 and γ_α by Lemma 8.13 with $b = \alpha$. All these choices precede $N \to \infty$, and none depends on u or (h, k); see Table 8.1.

The coupling. Let the coefficient vector J of H_N have its law conditioned on Ω_N, and sample an independent uniform axis u, an independent initial point X_0 with law ν_N, and an independent Brownian motion B. On this probability space, let X^{tr} solve (8.49) with $\widetilde{H} = H^{h,k,u} = H_N + \delta_{h,k,u}$, and let X^{cut} solve it with $\widetilde{H} = H_N + \delta_\alpha(u,\cdot)$, both from X_0 and driven by B. Let r_t and s_t be defined by (8.27) along each path. On the open set $\{x : |R(x,u)| < \alpha,\ |G_u \cdot x|/N < \alpha\}$, the two cutoffs equal one, and the two Hamiltonians and their gradients agree there. By pathwise uniqueness, the two solutions coincide up to the first time that either of them leaves this set.

Let $T = e^{\gamma N}$, let τ_a^{tr} and τ_a^{cut} be the hitting times of $C_a(u)$ by the two processes, and let $\varsigma^{\mathrm{tr}} = \inf\{t \ge 0 : |s_t^{\mathrm{tr}}| \ge \alpha\}$. We claim that

$$\{\tau_a^{\mathrm{tr}} \le T\} \subset \{\tau_a^{\mathrm{cut}} \le T\} \cup \{\varsigma^{\mathrm{tr}} \le T\}. \tag{8.59}$$

If $|R(X_0,u)| \ge a$, both hitting times are zero. If $|s_0| \ge \alpha$, then $\varsigma^{\mathrm{tr}} = 0$. Otherwise, suppose that $\tau_a^{\mathrm{tr}} \le T$ and $\varsigma^{\mathrm{tr}} > T$. Since $a < \alpha$ and the path is continuous, the true process reaches $|r| = a$ before $|r| = \alpha$, and it does not reach $|s| = \alpha$ before time T. The two processes therefore coincide on $[0, \tau_a^{\mathrm{tr}}]$, and $\tau_a^{\mathrm{cut}} = \tau_a^{\mathrm{tr}} \le T$. This proves (8.59).

The law of (J, u, X^{cut}) is Q_α. Since $\gamma \le \gamma_0$, Proposition 8.12 bounds the probability of the first event in (8.59) by $C(N^{1-d/2} + N^{3-d}) + e^{-cN}$. In the same way, $\gamma \le \gamma_\alpha$ and the restricted bound of Lemma 8.13 imply that the second event has probability at most $e^{-c_\alpha N}$.

Removing the restriction. Let now J have its unrestricted Gaussian law, with the other variables as before. For any event $\mathcal{A}$, we have $\mathbb{P}(\mathcal{A}) \le \mathbb{P}(\mathcal{A} \mid \Omega_N) + \mathbb{P}(\Omega_N^c)$, and $\mathbb{P}(\Omega_N^c) \le e^{-N}$. Under the unrestricted law and conditionally on u, the field $H^{h,k,u}$ has the law $\mathbb{E}_{h,k,u}$ of Definition 8.1, which is the conditional law by Proposition 8.7, and it is independent of X_0 and B. By the measurable construction of the solutions of (8.49), the probability of $\{\tau_a^{\mathrm{tr}} \le T\}$ given u is the left side of (8.5) at the axis u; it is a measurable function of u (compare Lemma 4.2). We have shown that

$$\int \mathbb{E}_{h,k,u}\Big[\mathbb{P}^H_{\nu_N}\Big(\sup_{0\le t\le e^{\gamma N}} |R(X_t,u)| \ge a\Big)\Big]\nu_N(\,\mathrm{d}u) \le C\big(N^{1-d/2} + N^{3-d}\big) + e^{-cN}. \tag{8.60}$$

Deterministic axes. Let $g \in O(N)$. The law of H_N is invariant under $H \mapsto H \circ g^{-1}$, and for the field $H \circ g^{-1}$ the quantities $H_Q(gu)$, $K_Q(gu)$, and G_{gu} equal $H_Q(u)$, $K_Q(u)$, and gG_u for the field H. By (8.3) and (8.4), the conditioned field at gu built from $H \circ g^{-1}$ is $H^{h,k,u} \circ g^{-1}$. The law of $H^{h,k,gu}$ is therefore the image of the law of $H^{h,k,u}$ under $H \mapsto H \circ g^{-1}$. If X solves (8.49) for a Hamiltonian $\widetilde{H}$ with Brownian motion B, then gX solves it for $\widetilde{H} \circ g^{-1}$ with Brownian motion gB; this is the rotation covariance of Lemma 4.2 at the level of paths. Together with the invariance of ν_N and $R(gx, gu) = R(x,u)$, it shows that the integrand in (8.60) takes the same value at u and at gu. As $O(N)$ acts transitively

on S_N, the integrand is constant, and (8.60) implies (8.5) for every $u \in S_N$. In the pure case, the same argument applies with the conditioning vector $(H_Q(u), G_u)$ and $k = ph$. Finite sums of exponentially small terms are bounded by e^{-cN} after decreasing c and increasing N_0, and this completes the proof. □

SECTION 9

Pure models: critical-point counts and Gibbs concentration on bands

In this section, we prove the hypotheses of the reduction in Section 4 on critical-point counts and on the concentration of the Gibbs measure on bands, for the pure model $\xi(r) = r^p$. The degree $p \geq 4$ may be odd or even, and $\beta > \beta_s(p)$ is fixed. Proposition 9.1 states that the expected number of critical points of the center field H_Q in a fixed window around $m_N = NE_0 - (\log N)/(2c_0)$ is bounded. It also states that this window contains the global maximum with probability close to one. Theorem 9.2 states that the bands around these critical points carry asymptotically all of the Gibbs measure, both for every small fixed width and for widths of order $N^{-1/2}$. Both results are used in Section 13, in the proofs of Theorem 1.2 and Corollary 1.3 for pure models.

We use the following results from other sections. From Section 3, we use the representation and regularity of the field (Lemma 3.2) and the concentration of free energies (Lemma 3.3). From Section 5, we use that the pure model is strictly 1RSB with a non-degenerate contact at Q, with explicit parameters (Proposition 5.20), the ground-state parameters E_0 and K_0 of Proposition 5.11, and the bound on the residual coefficient of Lemma 5.22. From Section 6, we use that every limiting directing measure is carried by the sphere of squared norm Q (Theorem 6.3). This is the only information about the overlap structure of the Gibbs measure that enters the proof. From Section 7, we use the conditioned laws and the law of the Hessian at a conditioned critical point (Definition 7.3 and Corollary 7.6), and Gaussian concentration (Lemma 7.5). We also use the moment bounds for conditioned determinants, the removal of exceptional events, and the bound on expected determinants by the logarithmic potential of the semicircle law (Propositions 7.7 and 7.9 and Corollary 7.8), as well as the marked Kac–Rice formula, the exact count formula, and the Hölder bound for determinant-weighted expectations (Theorem 7.17, Corollary 7.18, and Lemma 7.22).

The pure model has more explicit structure than a general mixture. We use several results of Subag and of Subag and Zeitouni about it. The parameters Q and x and the ground energy are given by Subag's computation of the pure free energy through the TAP approach [142]. The critical values near the maximum were studied by Subag and Zeitouni [144], who identified their extremal process; we use only the resulting tightness of the centered maximum. The law of the field conditioned on a critical point

was computed by Subag [141]. In [141, Theorem 1], Subag proved that for sufficiently large β the Gibbs measure concentrates on thin bands around the critical points of nearly maximal value. That theorem assumes that β is large, and it is not used here. Instead, we derive the concentration of the Gibbs measure on bands at every $\beta > \beta_{\mathrm{s}}(p)$ from Theorem 6.3, together with Subag's fixed-parameter estimates for free energies of replicated bands [143]. Since ξ is pure, the radial derivative is a multiple of the value, $K_Q = pH_Q$. All conditioning at a critical point is therefore on the value and the spherical gradient alone, and no covariance matrix of the value and the radial derivative is inverted.

The proof proceeds in stages. In Section 9.3, we bound the density of critical values by an exponential with a bounded prefactor on every window of sublinear width (Lemma 9.6). In Section 9.4, we show that all critical points with energy close to the ground energy are nondegenerate maxima, and that every point with such energy is close to a critical maximum (Proposition 9.12); we also prove Proposition 9.1. In Section 9.5, we show that the empirical mean of a fixed number of Gibbs samples with pairwise overlaps close to Q has energy close to the ground energy (Proposition 9.15). In Section 9.6, we combine these two facts with the directing measure and obtain a cover of the Gibbs measure by caps around critical points of near-ground energy (Proposition 9.16). In Sections 9.7 to 9.9, we pass from this cover to critical points in a fixed window and to microscopic bands, of width of order $N^{-1/2}$. The tools are second moments and lower tails of band partition functions conditioned on a critical point, a Laplace computation over latitudes, a lower bound for the partition function, and a bound on the partition functions of the bands around critical points whose value lies at least L below the centering, which decays exponentially in L.

Notation. Throughout this section, $p \geq 4$ is an integer, $\xi(r) = r^p$, $\beta > \beta_{\mathrm{s}}(p)$ is fixed, and H_N is realized by (3.2), so that it is a homogeneous polynomial of degree p on $\mathbb{R}^N$. We write π for the Gibbs measure, $Z_{N,\beta} = \int e^{\beta H_N} \, \mathrm{d}\nu_N$ for the partition function, as in Definition 3.1, and $P_N = N^{-1} \log Z_{N,\beta}$. As in Section 3, ∇ and ∇^2 denote Euclidean derivatives, and ∇_{S_N}, $\nabla^2_{S_N}$, and $\nabla^3_{S_N}$ denote the spherical gradient, Hessian, and third covariant derivative. We write d_{S_N} for the geodesic distance on S_N and $|S_N| = 2\pi^{N/2} N^{(N-1)/2} / \Gamma(N/2)$ for its surface area. Constants $C, c > 0$ depend only on p and β unless we say otherwise. All parameters other than N are fixed before $N \to \infty$, in the order stated in each proof, and Table 9.1 summarizes this order.

9.1 Parameters and statements

Thermodynamic parameters. By Proposition 5.20, the pair (ξ, β) is strictly 1RSB with a nondegenerate contact at Q (Definition 5.9), and the parameters x and Q are explicit. Let

$z_p > 0$ be the unique solution of

$$p\big[(1+z_p)\log(1+z_p) - z_p\big] = z_p^2, \tag{9.1}$$

which is the equation $v(z_p) = 1/p$ of Proposition 5.20, and let

$$\begin{gathered} q_s = \frac{z_p}{1+z_p}, \qquad e_p = \frac{p+z_p}{\sqrt{p(1+z_p)}}, \qquad e_\infty = 2\sqrt{\frac{p-1}{p}}, \\ c_p = \frac{z_p}{\sqrt{p(1+z_p)}}, \qquad w_0 = \sqrt{\frac{p-1}{1+z_p}}. \end{gathered} \tag{9.2}$$

By Proposition 5.20, we have $z_p > 2(p-2)$; the radius Q is the unique solution in $(q_s, 1)$ of

$$\beta^2 p(1+z_p)Q^{p-2}(1-Q)^2 = 1; \tag{9.3}$$

the other parameter is $x = z_p s/Q$, where $s = 1-Q$; the ground energy of H_N is $E(1) = \lim_N N^{-1}\mathbb{E}\max_{S_N} H_N = e_p$; the ground-state parameters of Proposition 5.11 are $E_0 = e_p Q^{p/2}$ and $K_0 = pE_0$; and, with the residual covariance $\psi_Q(t) = (Q+st)^p - Q^p - pQ^{p-1}st$,

$$F(\beta) = \beta e_p Q^{p/2} + \frac{1}{2}\log s + \frac{\beta^2}{2}\psi_Q(1). \tag{9.4}$$

These formulas were obtained by Subag [142, Theorems 1 and 2] from the generalized TAP representation; see Remark 5.21 for the dictionary.

We record some consequences. Since $z_p > p-2$, we have $0 < w_0 < 1$ and $q_s > (p-2)/(p-1) \geq 2/3$, and

$$e_p = \sqrt{\frac{p-1}{p}}\,(w_0 + w_0^{-1}), \qquad e_p^2 - e_\infty^2 = \frac{(z_p - p + 2)^2}{p(1+z_p)} > 0. \tag{9.5}$$

By (9.3) and the definition of x,

$$Q > q_s > \frac{2}{3}, \qquad 0 < x < 1, \qquad \beta s Q^{(p-2)/2} = \frac{1}{\sqrt{p(1+z_p)}}, \qquad c_p = \beta x Q^{p/2}. \tag{9.6}$$

From the last identity, we find that the slope $c_0 = \beta x$ of Proposition 5.11 is $c_0 = c_p Q^{-p/2}$. The center covariance parameters are $\xi(Q) = Q^p$, $a_Q = pQ^p$, and $b_Q = p(p-1)Q^p$.

Center field and windows. The center field and its radial derivative are

$$H_Q(u) = H_N(\sqrt{Q}\,u) = Q^{p/2}H_N(u), \qquad K_Q(u) = pH_Q(u), \qquad u \in S_N,$$

where the second identity is Euler's identity for homogeneous polynomials. We have $K_0 = pQ^{p/2}e_p > pQ^{p/2}e_\infty = 2\sqrt{b_Q}$ by (9.5). We fix a compact interval U that contains K_0 in its interior and satisfies $\min U > 2\sqrt{b_Q}$. We set

$$M_N = Ne_p - \frac{\log N}{2c_p}, \qquad m_N = Q^{p/2}M_N = NE_0 - \frac{\log N}{2c_0}, \tag{9.7}$$

and we let $\mathcal{C}_{N,M}$ and $B_{N,M}$ be as in Definition 4.1, with these Q, U, and m_N:

$$\mathcal{C}_{N,M} = \big\{u \in S_N : \ \nabla_{S_N} H_Q(u) = 0,\ |H_Q(u) - m_N| \le M,\ K_Q(u)/N \in U\big\}.$$

We also use the critical points of H_N in windows around M_N,

$$\mathcal{D}_{N,L} = \big\{u \in S_N : \ \nabla_{S_N} H_N(u) = 0,\ |H_N(u) - M_N| \le L\big\}.$$

The critical points of H_Q and of H_N coincide. For $u \in \mathcal{D}_{N,L}$, we have $K_Q(u)/N = pQ^{p/2}H_N(u)/N = K_0 + O((L + \log N)/N)$. Since K_0 is an interior point of U, it follows that, for every fixed $L > 0$,

$$\mathcal{C}_{N,Q^{p/2}L} \subset \mathcal{D}_{N,L} \quad \text{for every } N, \qquad \mathcal{C}_{N,Q^{p/2}L} = \mathcal{D}_{N,L} \quad \text{for all sufficiently large } N. \tag{9.8}$$

We count points of S_N; for even p, both u and $-u$ are counted.

We now state the two main results of this section.

Proposition 9.1 (Critical points in fixed windows)**.** *Let $p \ge 4$, $\beta > \beta_{\mathrm{s}}(p)$, and let U, m_N, and $\mathcal{C}_{N,M}$ be as above. For every fixed $M > 0$,*

$$\sup_{N \ge 2} \mathbb{E}|\mathcal{C}_{N,M}| < \infty. \tag{9.9}$$

Further,

$$\lim_{M\to\infty} \liminf_{N\to\infty} \mathbb{P}\Big\{\mathcal{C}_{N,M} \ne \varnothing \text{ and } \max_{S_N} H_Q \le m_N + M\Big\} = 1. \tag{9.10}$$

Finally, for every fixed $M > 0$, with probability tending to one, every $u \in \mathcal{C}_{N,M}$ is a nondegenerate local maximum of H_Q.

For $b > 0$ and $u \in S_N$, we write $B_u(b) = \{x \in S_N : |R(x,u) - \sqrt{Q}| \le b\}$ and $B_{N,M}(b) = \bigcup_{u\in\mathcal{C}_{N,M}} B_u(b)$; for a fixed b, these are the bands of Definition 4.1.

Theorem 9.2 (Gibbs concentration on bands)**.** *Let $p \ge 4$, $\beta > \beta_{\mathrm{s}}(p)$, and let U, m_N, and $\mathcal{C}_{N,M}$ be as above.*

1. *There exists $b_0 = b_0(p,\beta) \in (0, \sqrt{Q})$ such that for every fixed $b \in (0, b_0]$,*

$$\lim_{M\to\infty} \limsup_{N\to\infty} \mathbb{E}\,\pi\big(B_{N,M}(b)^c\big) = 0. \tag{9.11}$$

2. *For every $\varepsilon > 0$, there exist $M_0, w_* > 0$ such that for all fixed $M \geq M_0$ and $w \geq w_*$,*

$$\limsup_{N\to\infty} \mathbb{E}\,\pi\big(B_{N,M}(wN^{-1/2})^c\big) \leq \varepsilon. \tag{9.12}$$

Part (1) is the concentration property (4.2) for every band width $b \in (0, b_0]$. The proofs of Proposition 9.1 and Theorem 9.2 are completed in Section 9.4 and Section 9.9, respectively. Table 9.1 lists the parameters that appear in these proofs, in the order in which they are fixed.

Step	Parameter	Depends on	Where
1	$z_p, Q, s, x, e_p, c_p, E_0, K_0, c_0$	p, β	Proposition 5.20, Section 9.1
2	interval $U \ni K_0$	p, β	Section 9.1
3	cap latitude a; then b_0 and $b \in (0, b_0]$	p, β	Lemma 9.3, Theorem 9.2
4	energy deficit τ per spin and tolerance ε of the coarse cover; then $a_1, r, \tau_1, n, \eta, T$	$p, \beta, a, \tau, \varepsilon$	Proposition 9.16
5	tolerance ε in Theorem 9.2(2); then η, the window L, the width w	p, β, a, ε	Section 9.9
6	$N \to \infty$	all of the above	

Table 9.1. Parameters in the proofs of Proposition 9.1 and Theorem 9.2, in the order in which they are fixed. The window M of Proposition 9.1 and Theorem 9.2(1) is fixed after the band width b and the interval U, and before $N \to \infty$.

9.2 The exponent of the latitude integral

The bands around a critical point are weighed by a one-dimensional integral over latitudes. In this section, we choose the latitude a of the caps used below and record the properties of the exponent of that integral. As in Lemma 5.22, for $0 \leq q \leq 1$ we let

$$C_\beta(q) = \beta^2\big[1 - q^{2p} - pq^{2p-2}(1-q^2)\big], \qquad C'_\beta(q) = -2\beta^2 p(p-1)q^{2p-3}(1-q^2) \leq 0. \tag{9.13}$$

The quantity $NC_\beta(q)/\beta^2$ is the variance of the field on the latitude q around a critical point, after the conditioning (see Lemma 9.5 below). By Lemma 5.22, we have

$$C_\beta(\sqrt{Q}) = \beta^2\psi_Q(1) < \frac{159}{320} < \frac{1}{2} \tag{9.14}$$

for every $p \geq 4$ and every $\beta > \beta_{\mathrm{s}}(p)$. This bound is what makes the second moments of band partition functions finite in Proposition 9.19.

Lemma 9.3 (Choice of the cap latitude)**.** *For $0 < q < 1$, let*

$$\Lambda(q) = \beta e_p q^p + \frac{1}{2}\log(1-q^2) + \frac{1}{2}C_\beta(q).$$

There exists $a \in (1/\sqrt{2}, \sqrt{Q})$ such that

$$C_\beta(a) < \frac{1}{2}, \qquad \beta a^p > c_p, \qquad \Lambda(\sqrt{Q}) = F(\beta), \qquad \Lambda''(\sqrt{Q}) < 0, \tag{9.15}$$

Λ is strictly increasing on $[a, \sqrt{Q}]$, and Λ is strictly decreasing on $[\sqrt{Q}, 1)$. In particular, $\sqrt{Q}$ is the unique maximizer of Λ on $[a, 1)$.

We fix such an a for the rest of this section, and write $k_* = -\Lambda''(\sqrt{Q}) > 0$.

Proof. Let $f(q) = \beta q^{p-2}(1-q^2)$. Differentiating Λ and using (9.13), we find

$$\begin{aligned}\Lambda'(q) &= p\beta e_p q^{p-1} - \frac{q}{1-q^2} - \beta^2 p(p-1)q^{2p-3}(1-q^2) \\ &= \frac{q}{1-q^2}\Big(pe_p f(q) - 1 - p(p-1)f(q)^2\Big).\end{aligned}$$

The quadratic $p(p-1)f^2 - pe_p f + 1$ has the roots

$$a_\pm = \frac{e_p \pm \sqrt{e_p^2 - e_\infty^2}}{2(p-1)}, \qquad a_- = \frac{1}{\sqrt{p(1+z_p)}}, \qquad a_+ = \frac{\sqrt{1+z_p}}{\sqrt{p}\,(p-1)},$$

where we used (9.5) and $z_p > p-2$ to evaluate the square root as $(z_p - p + 2)/\sqrt{p(1+z_p)}$. Also $a_+ > a_-$, since $z_p > p-2$. Factoring the quadratic, we find

$$\Lambda'(q) = -\frac{p(p-1)q}{1-q^2}\big(f(q) - a_-\big)\big(f(q) - a_+\big). \tag{9.16}$$

Note that $f(\sqrt{Q}) = a_-$ by (9.6). Further, $f'(q) = \beta q^{p-3}[(p-2) - pq^2]$ is negative for $q^2 > (p-2)/p$, and $Q > (p-2)/p$. It follows that f is strictly decreasing on an interval $(q_1, 1)$ that contains $\sqrt{Q}$. For $q \in (\sqrt{Q}, 1)$, we have $f(q) < a_- < a_+$, and (9.16) implies that $\Lambda'(q) < 0$. For $q \in (q_1, \sqrt{Q})$ close to $\sqrt{Q}$, we have $a_- < f(q) < a_+$ and $\Lambda'(q) > 0$. Differentiating (9.16) at $\sqrt{Q}$, where the first factor in parentheses vanishes, we find

$$\Lambda''(\sqrt{Q}) = -\frac{p(p-1)\sqrt{Q}}{1-Q} f'(\sqrt{Q})\,(a_- - a_+) < 0,$$

since $f'(\sqrt{Q}) < 0$ and $a_- - a_+ < 0$. The identity $\Lambda(\sqrt{Q}) = F(\beta)$ is (9.4), because $C_\beta(\sqrt{Q}) = \beta^2 \psi_Q(1)$.

Recall from (9.14) that $C_\beta(\sqrt{Q}) < 1/2$. We also have $\beta Q^{p/2} = c_p/x > c_p$ and $\sqrt{Q} > \sqrt{2/3} > 1/\sqrt{2}$ by (9.6). All of these strict inequalities persist on a short interval just below $\sqrt{Q}$. Let q_2 be the supremum of the set $\{0\} \cup \{q \in (0, \sqrt{Q}) : \Lambda'(q) \le 0\}$, which is smaller than $\sqrt{Q}$ by the sign of Λ' established above. Choosing $a < \sqrt{Q}$ larger than q_2 and $1/\sqrt{2}$, and close enough to $\sqrt{Q}$, we obtain all the assertions. □

9.3 The conditional field and the density of critical values

Lemma 9.4 (Regularity of the pure field)**.** *There exist $C, c > 0$ such that, for every $N \ge 2$, with probability at least $1 - e^{-cN}$,*

$$\sup_{x \in S_N} \left\| \nabla^3_{S_N} H_N(x) \right\|_{\mathrm{op}} \le \frac{C}{\sqrt{N}}. \tag{9.17}$$

Moreover, $\mathbb{E} \sup_{|x| \le \sqrt{N}} |\nabla H_N(x)| \le C\sqrt{N}$ for all $N \ge 2$, and for every fixed $r > 0$ and all sufficiently large N,

$$\mathbb{P}\Big\{ \Big| \frac{1}{N} \max_{S_N} H_N - e_p \Big| > r \Big\} \le 2e^{-Nr^2/8}. \tag{9.18}$$

Proof. Let K, c be the constants of Lemma 3.2(3), and work on the event of that lemma. Let $x \in S_N$, and let z, v, w be unit tangent vectors at x. The spherical Hessian is the restriction to tangent vectors of the form $A = \nabla^2 H_N - N^{-1}(x \cdot \nabla H_N) I = \nabla^2 H_N - (p/N) H_N I$, where we used Euler's identity $x \cdot \nabla H_N = p H_N$. For tangent vector fields V, the covariant derivative on S_N is $\nabla_z V + N^{-1}(z \cdot V)x$. Differentiating $A(V, W)$ along z and subtracting the terms in which the covariant derivative falls on V or W, we obtain

$$\begin{aligned} \nabla^3_{S_N} H_N(x)[z, v, w] = \nabla^3 H_N(x)[z, v, w] &- \frac{p}{N} (\nabla H_N(x) \cdot z)(v \cdot w) \\ &- \frac{p-1}{N} \big[(z \cdot v)\, \nabla H_N(x) \cdot w + (z \cdot w)\, \nabla H_N(x) \cdot v \big]. \end{aligned}$$

Here we used $A(x, w) = \nabla^2 H_N(x)[x, w] = (p-1) \nabla H_N(x) \cdot w$ for tangent w, which is Euler's identity for the homogeneous polynomial $\nabla H_N \cdot w$ of degree $p - 1$. By (3.3), the right side is at most $K/\sqrt{N} + 3pK/\sqrt{N}$, which proves (9.17).

For the second assertion, let $Y = \sup_{|x| \le \sqrt{N}, |v| \le 1} \nabla H_N(x) \cdot v$. Writing $H_N(x) = g \cdot \Psi(x)$ as in the proof of Lemma 3.3, each $\nabla H_N(x) \cdot v$ is a linear function of g whose coefficient vector has norm at most $p|x|^{p-1}|v| N^{-(p-1)/2} \le p$. Then Y is a p-Lipschitz function of g, and the Gaussian concentration inequality of Lemma 7.5(1) implies that $\mathbb{P}\{Y \le \mathbb{E}Y - t\} \le e^{-t^2/(2p^2)}$ for $t > 0$. Recall from Lemma 3.2(3) that $\mathbb{P}\{Y > K\sqrt{N}\} \le e^{-cN}$. If $\mathbb{E}Y > K\sqrt{N} + 2p\sqrt{cN}$, we would have $1 - e^{-2cN} \le \mathbb{P}\{Y > K\sqrt{N}\} \le e^{-cN}$, which is

impossible for large N. We conclude that $\mathbb{E}Y \le (K + 2p\sqrt{c})\sqrt{N}$ for all sufficiently large N; for each of the remaining N, $\mathbb{E}Y \le p\,\mathbb{E}|g| < \infty$.

For (9.18), note that $\max_{S_N} H_N$ is a $\sqrt{N}$-Lipschitz function of g, since it is a maximum of linear functions of g whose coefficient vectors have norm $|\Psi(x)| = \sqrt{N}$. Using Proposition 5.20(3) and the definition of $E(1)$ in Theorem 5.2, we find that $|\mathbb{E}\max_{S_N} H_N/N - e_p| \le r/2$ for all sufficiently large N. For such N, Lemma 7.5(1) with $s = Nr/2$, applied to the maximum and to its negative, implies (9.18). □

Conditioned laws and the Kac–Rice formula. We use the results of Section 7 with the center radius equal to one, so that the center field is H_N itself, with covariance parameters $\xi(1) = 1$, $a_1 = p$, and $b_1 = p(p-1)$. For the pure model, the radial derivative is pH_N, and the conditioning is on the value and the spherical gradient alone. For $u \in S_N$ and $h \in \mathbb{R}$, we write $\mathbb{P}_{u;h}$ and $\mathbb{E}_{u;h}$ for the conditioned law of Definition 7.3 at u, with the normalized value h/N; under it, $H_N(u) = h$ and $\nabla_{S_N} H_N(u) = 0$. Let $n = N - 1$, let G_n be a standard GOE matrix of dimension n as in Section 7, and let

$$\varsigma_N = \Big(\frac{p(p-1)(N-1)}{N}\Big)^{1/2}, \qquad D(u) = \big|\det \nabla^2_{S_N} H_N(u)\big|.$$

Recall from Corollary 7.6 that, under $\mathbb{P}_{u;h}$, the matrix of $\nabla^2_{S_N} H_N(u)$ in an orthonormal basis of $u^\perp$ has the law of

$$\varsigma_N G_n - \frac{ph}{N}\,\mathrm{Id}_n. \tag{9.19}$$

It follows that $\mu_N(h) = \mathbb{E}_{u;h} D(u) = \mu_{n,N}(ph/N)$ does not depend on u. It is strictly positive, as noted in Section 7.1, and it is continuous in h. Indeed, the absolute determinant $|\det(\varsigma_N G_n - (ph/N)\mathrm{Id}_n)|$ is continuous in h, and for h in a compact interval it is bounded by $(\varsigma_N \|G_n\|_{\mathrm{op}} + C)^n$, which is integrable; dominated convergence applies. By Theorem 7.17, after the change of variables from the normalized value to h, we have, for every deterministic Borel set $J \subset \mathbb{R}$ and every Borel function $\varphi(H, u) \ge 0$,

$$\mathbb{E} \sum_{\substack{u:\nabla_{S_N} H_N(u)=0 \\ H_N(u)\in J}} \varphi(H,u) = \int_J \rho_N(h) \int_{S_N} \frac{\mathbb{E}_{u;h}[D(u)\,\varphi(H,u)]}{\mu_N(h)}\,\nu_N(\,\mathrm{d}u)\,\mathrm{d}h, \tag{9.20}$$

where

$$\rho_N(h) = |S_N|\,(2\pi p)^{-(N-1)/2}(2\pi N)^{-1/2} e^{-h^2/(2N)} \mu_N(h). \tag{9.21}$$

With $\varphi = 1$, (9.20) is the count formula of Corollary 7.18: ρ_N is the density of the intensity measure of the critical values $\{H_N(u) : \nabla_{S_N} H_N(u) = 0\}$, counted at points of S_N, and it is continuous.

Let $\Delta = e_p - e_\infty > 0$. By Proposition 7.7, applied with the radius equal to one, the shift $k = ph/N$, the gap $g = p\Delta/2$, and $k_+ = p(e_p + \Delta/2)$, there exists K_2, depending only on p, such that

$$\mathbb{E}_{u;h}\big[D(u)^2\big] \le \mathsf{K}_2\, \mu_N(h)^2 \qquad \text{for all } N \ge 2,\ u \in S_N \text{ and } \left|\frac{h}{N} - e_p\right| \le \frac{\Delta}{2}. \tag{9.22}$$

Indeed, for such h we have $ph/N \ge pe_\infty + p\Delta/2 = 2\sqrt{b_1} + p\Delta/2$.

The next lemma describes the whole field under $\mathbb{P}_{u;h}$, and not only its Hessian at u.

Lemma 9.5 (The field conditioned on a critical point)**.** *Let $u \in S_N$ and $h \in \mathbb{R}$. We write each $x \in S_N \setminus \{\pm u\}$ as $x = qu + \sqrt{N(1-q^2)}\,\omega$, where $q = R(x,u)$ and ω is a unit vector orthogonal to u, and let*

$$\alpha_k(q) = \binom{p}{k}^{1/2} q^{p-k}(1-q^2)^{k/2}, \qquad 0 \le k \le p.$$

Under $\mathbb{P}_{u;h}$, the field has the law of

$$H_N\big(qu + \sqrt{N(1-q^2)}\,\omega\big) = hq^p + \sqrt{N}\sum_{k=2}^{p} \alpha_k(q) X_k(\omega), \tag{9.23}$$

where $X_2, \dots, X_p$ are independent centered Gaussian fields on the unit sphere of $u^\perp$, with $\mathbb{E}X_k(\omega)X_k(\omega') = (\omega\cdot\omega')^k$. The Hessian $\nabla^2_{S_N} H_N(u) + (ph/N)\mathrm{Id}$ is a function of X_2; in particular, it is not independent of the field in (9.23)*.*

Proof. This is [141, Lemmas 14 and 15 and Corollary 17], written in our notation: Subag's pure field of degree k on the sphere of radius $\sqrt{N-1}$ in $u^\perp$, multiplied by $\sqrt{N/(N-1)}$ as in [141, (4.11)], is our $\sqrt{N}X_k(\omega)$. We indicate the computation. After a rotation, we may assume that $u = \sqrt{N}e_N$. We group the monomials in (3.2) according to the number k of indices different from N. The groups involve disjoint sets of the Gaussian variables and are independent. The group with k such indices equals $\sqrt{N}\alpha_k(q)X_k(\omega)$ for a field X_k that is a homogeneous polynomial of degree k in ω. Its covariance is identified by expanding $\mathbb{E}H_N(x)H_N(x') = N\big(qq' + \sqrt{(1-q^2)(1-q'^2)}\,\omega\cdot\omega'\big)^p$ by the binomial theorem. The group with $k = 0$ is $q^pH_N(u)$. The group with $k = 1$ is $\sqrt{N}\alpha_1(q)\,\zeta\cdot\omega$ for a standard Gaussian vector ζ in $u^\perp$, and $\nabla_{S_N}H_N(u) = \sqrt{p}\,\zeta$, since only this group contributes to the first derivatives at u. The conditioned law of Definition 7.3 is Gaussian regression on $(H_N(u), \nabla_{S_N}H_N(u))$. It replaces the first two groups by hq^p and 0 and leaves the others unchanged, which proves (9.23). In the tangent coordinate $y = \sqrt{N(1-q^2)}\,\omega$ near u, the groups with $k \ge 3$ are of order $|y|^3$, and $hq^p = h - (ph/2N)|y|^2 + O(|y|^4)$. The Hessian at u is therefore $-(ph/N)\mathrm{Id}$ plus the Hessian at $y = 0$ of the group with $k = 2$, which is a function of X_2. □

We now bound ρ_N. For $t > 2$, let $w(t) = (t - \sqrt{t^2-4})/2$, which is the Stieltjes transform $g_{\mathrm{sc}}(t)$ of the semicircle law σ_{sc} of Section 7.4, let $\Omega(t) = \int \log(t-\lambda)\,\sigma_{\mathrm{sc}}(\,\mathrm{d}\lambda)$, and let

$$\gamma_p = \sqrt{\frac{p}{p-1}}, \qquad \theta(e) = \frac{1}{2} + \frac{1}{2}\log(p-1) - \frac{e^2}{2} + \Omega(\gamma_p e), \qquad e > e_\infty.$$

Note that $\gamma_p e > 2$ if and only if $e > e_\infty$. The function $\theta(e)$ is $\Theta_p(-e)$ in the notation of [144, (2.5)], the exponential growth rate of the expected number of critical points with values above Ne, computed by Auffinger, Ben Arous, and Černý [12].

Lemma 9.6 (Density of critical values)**.** *1. The function θ is strictly concave on (e_∞, ∞), $\theta(e_p) = 0$, and $\theta'(e_p) = -c_p$.*

2. Let $\mathcal{J}$ be a compact subinterval of (e_∞, ∞). There exists C, depending only on p and $\mathcal{J}$, such that for all $N \geq 2$ with $h/\sqrt{N(N-1)} \in \mathcal{J}$,

$$\rho_N(h) \leq CN^{-1/2} \exp\Big[(N-1)\,\theta\Big(\frac{h}{\sqrt{N(N-1)}}\Big)\Big]. \tag{9.24}$$

3. Let $a_N, a_N' \geq 0$ be deterministic with $a_N + a_N' = o(N)$. There exists C, depending only on p, such that for all sufficiently large N and all $h \in [M_N - a_N, M_N + a_N']$,

$$\rho_N(h) \leq Ce^{c_p(M_N - h)}. \tag{9.25}$$

The bound (9.25) has a bounded prefactor on every window of sublinear width. This is what the summable bounds in Section 9.9, on the bands around critical points far below the centering, require.

Proof. For (1), we have $\Omega'(t) = w(t)$ and $\Omega''(t) = w'(t) = (1 - t/\sqrt{t^2-4})/2 < 0$ for $t > 2$. This implies that $\theta''(e) = -1 + \gamma_p^2 \Omega''(\gamma_p e) < 0$. For $v \in (0,1]$, we have $w(v + v^{-1}) = v$, and $\Omega(2) = 1/2$ by the closed form of Ω in [144, (2.4)]. Integrating $\Omega' = w$ along $t = v' + 1/v'$, $v' \in [v, 1]$, we find

$$\Omega(v + v^{-1}) = \frac{1}{2} + \int_1^v \Big(v' - \frac{1}{v'}\Big)\,\mathrm{d}v' = \frac{v^2}{2} - \log v.$$

By (9.5), $\gamma_p e_p = w_0 + w_0^{-1}$, and the last display gives $\Omega(\gamma_p e_p) = w_0^2/2 - \log w_0$; also $\Omega'(\gamma_p e_p) = w_0$. Substituting $w_0^2 = (p-1)/(1+z_p)$ and $e_p^2 = (p+z_p)^2/(p(1+z_p))$, we find

$$\theta(e_p) = \frac{p(1+z_p)\log(1+z_p) - pz_p - z_p^2}{2p(1+z_p)} = 0, \qquad \theta'(e_p) = -e_p + \gamma_p w_0 = -c_p,$$

where the first identity is (9.1), and the second follows from $\gamma_p w_0 = \sqrt{p/(1+z_p)}$ and (9.2).

For (2), let $e = h/\sqrt{N(N-1)}$ and $v = \gamma_p e$. Then $v = (ph/N)/\varsigma_N$, and by (9.19), $\mu_N(h) = \varsigma_N^{N-1}\mathbb{E}|\det(v\,\mathrm{Id}_n - G_n)|$, where we used that $-G_n$ has the law of G_n. Since $e^2(N-1) = h^2/N$, (9.21) becomes

$$\rho_N(h) = |S_N|\,N^{-(N-1)/2}\Big(\frac{(p-1)(N-1)}{2\pi}\Big)^{(N-1)/2}(2\pi N)^{-1/2} \times e^{-(N-1)e^2/2}\,\mathbb{E}|\det(v\,\mathrm{Id}_n - G_n)|. \quad (9.26)$$

The interval $\gamma_p\mathcal{J}$ is a compact subinterval of $(2,\infty)$, and Proposition 7.9 gives $\mathbb{E}|\det(v\,\mathrm{Id}_n - G_n)| \le Ce^{(N-1)\Omega(v)}$, with C depending only on p and $\mathcal{J}$. By Stirling's formula, $|S_N|N^{-(N-1)/2}((N-1)/(2\pi))^{(N-1)/2}(2\pi N)^{-1/2}e^{-(N-1)/2} = (1+o(1))(\pi N)^{-1/2}$. Inserting these bounds into (9.26) and using $(N-1)[\log(p-1)+1-e^2]/2 + (N-1)\Omega(\gamma_p e) = (N-1)\theta(e)$, we obtain (9.24).

For (3), let $\mathcal{J} = [e_p - \Delta/2, e_p + \Delta/2]$ and let C be as in (2). For $h \in [M_N - a_N, M_N + a'_N]$ and large N, we have $h/\sqrt{N(N-1)} \in \mathcal{J}$. Part (1) implies that $\theta(e) \le -c_p(e-e_p)$ for all $e > e_\infty$. Let $r_N = \sqrt{(N-1)/N}$ and $v = M_N - h$. Inserting the tangent bound into (9.24), we find

$$\begin{aligned}\log\rho_N(h) &\le \log C - \frac{1}{2}\log N - c_p r_N h + c_p(N-1)e_p \\ &= \log C + c_p r_N v + c_p e_p\big[(N-1) - Nr_N\big] + \frac{r_N - 1}{2}\log N,\end{aligned} \quad (9.27)$$

where we used $r_N M_N = r_N N e_p - r_N(\log N)/(2c_p)$. The last two terms are bounded, since $(N-1) - Nr_N \to -1/2$ and $(1-r_N)\log N \to 0$. If $v \ge 0$, then $r_N v \le v$. If $v < 0$, then $c_p(r_N - 1)v \le c_p|v|/N = o(1)$, since $|v| \le a'_N = o(N)$. In both cases, we obtain (9.25). □

The proof of (2) is the computation of [144, Section 4], with the Hermite asymptotics replaced by the upper bound of Proposition 7.9; only an upper bound with a bounded prefactor is needed below.

Remark 9.7 (Windows of sublinear width). For windows of fixed width, (9.25) also follows from the case of compact windows in [144, Proposition 3]. On windows of width much larger than $\sqrt{N}$, the exponent in (9.24) differs from its linearization $-c_p(h - M_N)$ by a term of order $(h - Ne_p)^2/N$, because $\theta''(e_p) < 0$. For this reason we use the upper bound (9.25), which is what the arguments below need, and not a relative asymptotic for ρ_N on such windows.

Lemma 9.8 (Tightness of the maximum)**.** *The sequence* $\max_{S_N} H_N - M_N$ *is tight. Moreover, there exists* C *such that* $\limsup_{N\to\infty}\mathbb{P}\{\max_{S_N} H_N > M_N + L\} \le Ce^{-c_p L/2}$ *for every* $L \ge 0$.

Proof. We begin with the upper tail. Let $L \ge 0$ and $J = [M_N + L, N(e_p + \Delta/4)]$. On the event $\{M_N + L < \max H_N \le N(e_p + \Delta/4)\}$, the maximizer of H_N is a critical point

with value in J. By (9.20) with $\varphi = 1$, this event has probability at most $\int_J \rho_N(h)\,\mathrm{d}h$. For $h \in J$ and large N, we have $h/\sqrt{N(N-1)} \in [e_p - \Delta/2, e_p + \Delta/2]$, and (9.27), which uses only (9.24) and the tangent bound, implies that $\rho_N(h) \le C\exp(-c_p r_N(h - M_N)) \le Ce^{-c_p(h-M_N)/2}$. Integrating, we find $\int_J \rho_N \le (2C/c_p)e^{-c_p L/2}$. The event $\{\max H_N > N(e_p + \Delta/4)\}$ has probability tending to zero by (9.18), and the bound on the upper tail follows.

The lower tail is the main content of the work of Subag and Zeitouni. They identified the extremal process of the critical values and the limiting law of the minimum of H_N [144, Theorem 1 and Corollary 2], with the centering $-Ne_p + (\log N)/(2c_p)$ up to an additive constant depending only on p [144, (1.4) and (1.5)]. Their hypothesis is $p \ge 3$, with no condition on the temperature. Their E_0, the unique zero of $\Theta_p(-\cdot)$ above E_∞, is our e_p, and their $c_p = \Theta_p'(-E_0)$ is our c_p, by Lemma 9.6(1). Applied to the field $-H_N$, which has the same law as H_N, their results imply that $\max_{S_N} H_N - M_N$ is tight. We use only this tightness, which does not depend on the additive constant in the centering or on the normalization of the limiting intensity. □

Lemma 9.9 (Marked first moments in windows of sublinear width)**.** *Let $a_N, a_N' \ge 0$ be deterministic with $a_N + a_N' = o(N)$, and let $J_N \subset [M_N - a_N, M_N + a_N']$ be deterministic Borel sets. Let $Y(H,u) \ge 0$ be a Borel function of the field and the point, and let $y_N \colon J_N \to [0,\infty]$ be a Borel function such that $\mathbb{E}_{u;h}[Y(H,u)^2] \le y_N(h)^2$ for all $u \in S_N$ and $h \in J_N$. There exists C, depending only on p, such that for all sufficiently large N,*

$$\mathbb{E} \sum_{\substack{u:\nabla_{S_N} H_N(u)=0 \\ H_N(u)\in J_N}} Y(H,u) \le C \int_{J_N} e^{c_p(M_N-h)} y_N(h)\,\mathrm{d}h. \tag{9.28}$$

Proof. For large N, every $h \in J_N$ satisfies $|h/N - e_p| \le \Delta/2$. By the Cauchy–Schwarz inequality under $\mathbb{P}_{u;h}$ and (9.22), we have $\mathbb{E}_{u;h}[D(u)Y(H,u)] \le \mathsf{K}_2^{1/2}\mu_N(h)\,y_N(h)$. Inserting this bound into (9.20) and using (9.25), we obtain (9.28). □

For an indicator $Y = \mathbf{1}_A$, the bound in the proof is the Hölder bound of Lemma 7.22 for determinant-weighted expectations, with $r = 2$. The proof keeps the determinant inside the conditional expectation; no independence between the Hessian and the weight Y is used. For latitude intervals and weights that satisfy the tameness condition of [141, Lemma 36], bounds of this type are [141, Lemmas 38 and 41], already stated for windows of sublinear width. Here we prove the bounded-prefactor envelope of Lemma 9.6(3) directly and apply the estimates to arbitrary Borel weights, including indicators of bands with small partition function and partition functions of disconnected sets of latitudes.

9.4 Localization at critical maxima

In this section, we show that every point with energy close to the ground energy lies close to a critical maximum, and complete the proof of Proposition 9.1. We write $\Delta = e_p - e_\infty$ as before.

Lemma 9.10 (A Hessian bound at all high critical points)**.** *Let $\kappa = p\Delta/4$. There exist $\tau_0 \in (0, \Delta/4]$ and $c > 0$, depending only on p, such that for all sufficiently large N, with probability at least $1 - e^{-cN}$, every critical point u of H_N with $H_N(u) \geq N(e_p - \tau_0)$ satisfies $\nabla^2_{S_N} H_N(u) \leq -\kappa \,\mathrm{Id}$.*

Proof. Let $h \in [N(e_p - \Delta/4), N(e_p + \Delta/4)]$ and $u \in S_N$, and let A be the set of pairs (H, u) for which $\nabla^2_{S_N} H_N(u)$ has an eigenvalue larger than $-\kappa$. By (9.19), under $\mathbb{P}_{u;h}$, the event A occurs if and only if $\lambda_{\max}(\varsigma_N G_n) > ph/N - \kappa$, and $ph/N - \kappa \geq p(e_p - \Delta/2) = 2\sqrt{b_1} + p\Delta/2$. The matrix $\varsigma_N G_n$ is the matrix $W_{n,N}$ of Section 7 with b_Q replaced by b_1, and it satisfies the hypothesis of Lemma 7.5(2) with $b = b_1$. Using (7.7) with $s = p\Delta/2$, we find that $\mathbb{P}_{u;h}(A) \leq 2e^{-c'N}$ with $c' = p^2\Delta^2/(16 b_1)$. The shift ph/N lies in $[2\sqrt{b_1} + 3p\Delta/4, p(e_p + \Delta/4)]$, so Corollary 7.8 with $r = 2$ gives

$$\mathbb{E}_{u;h}\big[D(u)\mathbf{1}_A(H,u)\big] \leq \mathsf{K}_2^{1/2}(2e^{-c'N})^{1/2}\mu_N(h) \leq e^{-c_1 N}\mu_N(h)$$

for large N, with $c_1 = c'/3$, where K_2 depends only on p. The constant c_1 does not change when the energy interval is shrunk below.

Let $\tau_0 \leq \Delta/4$, to be chosen. By (9.20) with $\varphi = \mathbf{1}_A$, the expected number of critical points u with $(H,u) \in A$ and $H_N(u) \in [N(e_p - \tau_0), N(e_p + \tau_0)]$ is at most $e^{-c_1 N}$ times $\int \rho_N(h)\,\mathrm{d}h$ over this interval. For such h, we have $h/\sqrt{N(N-1)} \geq e_p - \tau_0$. Using Lemma 9.6(2) with $\mathcal{J} = [e_p - \Delta/2, e_p + \Delta/2]$ and the tangent bound $\theta(e) \leq -c_p(e - e_p)$, we find $\rho_N(h) \leq CN^{-1/2}e^{(N-1)c_p\tau_0}$, and the integral is at most $2C\tau_0 N^{1/2}e^{c_p\tau_0 N}$. We choose τ_0 so small that $c_p\tau_0 \leq c_1/2$. By Markov's inequality, the probability that there exists such a critical point is at most $2C\tau_0 N^{1/2}e^{-c_1 N/2}$. Finally, a critical point with value above $N(e_p + \tau_0)$ exists only if $\max H_N > N(e_p + \tau_0)$, which has probability at most $2e^{-N\tau_0^2/8}$ by (9.18). This proves the lemma. □

The determinant weight enters through Corollary 7.8. The probability of the event A under the conditioned law alone would not control the critical points selected by the field, since critical points are counted with the weight $D(u)$.

Lemma 9.11 (Deterministic localization in superlevel sets)**.** *Let $f \in C^3(S_N)$ and $f_* = \max_{S_N} f$, and suppose that $\sup_{S_N} \|\nabla^3_{S_N} f\|_{\mathrm{op}} \leq C_3/\sqrt{N}$. Let $\delta_0, \kappa > 0$, and suppose that every critical point c of f with $f(c) > f_* - \delta_0 N$ satisfies $\nabla^2_{S_N} f(c) \leq -\kappa I$. Let $r_0 = \min(1, \kappa/(2C_3))$, with $r_0 = 1$ if $C_3 = 0$, and let $0 < \delta < \min(\delta_0, \kappa r_0^2/4)$. Then every connected component $\mathcal{V}$ of*

$\{f > f_* - \delta N\}$ contains exactly one critical point c of f, which is the maximizer of f on $\mathcal{V}$, and $\mathcal{V}$ is contained in the open geodesic ball of radius $r_0\sqrt{N}$ around c. Moreover, every $x \in \mathcal{V}$ satisfies

$$d_{S_N}(x,c) < 2\sqrt{\delta/\kappa}\,\sqrt{N}, \qquad f(c) \geq f(x). \tag{9.29}$$

Proof. The function f attains its maximum on the compact closure of $\mathcal{V}$ at some point c, and $f(c) > f_* - \delta N$ since $\mathcal{V}$ is nonempty. A small connected neighborhood of c on which $f > f_* - \delta N$ meets $\mathcal{V}$, and hence is contained in $\mathcal{V}$. We conclude that $c \in \mathcal{V}$, c is a local maximum of f, and $\nabla^2_{S_N} f(c) \leq -\kappa I$ by hypothesis.

Let γ be a unit-speed geodesic starting at c. Along γ, the Hessian, transported in parallel, changes by at most $C_3 t/\sqrt{N}$ in operator norm up to time t. For $t \leq r_0\sqrt{N}$, it is therefore at most $-\kappa I + C_3 r_0 I \leq -(\kappa/2)I$. Integrating twice along γ and using $\nabla_{S_N} f(c) = 0$, we obtain, for every y in the closed geodesic ball of radius $r_0\sqrt{N}$ around c,

$$f(y) \leq f(c) - \frac{\kappa}{4} d_{S_N}(y,c)^2 \leq f_* - \frac{\kappa}{4} d_{S_N}(y,c)^2. \tag{9.30}$$

The geodesics in this argument are unique, since the radius $r_0\sqrt{N} \leq \sqrt{N}$ is smaller than the injectivity radius $\pi\sqrt{N}$ of S_N. On the boundary sphere of the ball, (9.30) implies that $f \leq f_* - \kappa r_0^2 N/4 < f_* - \delta N$. The open set $\mathcal{V}$ is path connected, contains c, and does not meet this boundary sphere. It is therefore contained in the open ball. If $c' \neq c$ were another critical point in the ball, then along the geodesic from c to c', of length $t_1 > 0$, the derivative of f at time t_1 would be at most $-\kappa t_1/2 < 0$, which contradicts $\nabla_{S_N} f(c') = 0$. Finally, (9.30) and $f(x) > f_* - \delta N$ imply (9.29). □

Proposition 9.12 (Proximity of high-energy points to local maxima)**.** *For every fixed $r > 0$, there exist $\varepsilon_0, c > 0$ such that, for all sufficiently large N, with probability at least $1 - e^{-cN}$, every $y \in S_N$ with $H_N(y) \geq N(e_p - \varepsilon_0)$ has a nondegenerate local maximum u of H_N with*

$$|u - y| \leq r\sqrt{N}, \qquad H_N(u) \geq H_N(y). \tag{9.31}$$

The same holds with ε_0 replaced by any smaller positive number.

Proof. Let τ_0, κ be as in Lemma 9.10 and C_3 the constant in (9.17), and let $r_0 = \min(1, \kappa/(2C_3))$. We fix $\varepsilon_0 \in (0, \tau_0/6)$ so small that $3\varepsilon_0 < \kappa r_0^2/4$ and $2\sqrt{3\varepsilon_0/\kappa} < r$; every smaller ε_0 has the same properties. Let $\mathcal{E}$ be the intersection of the events of Lemmas 9.4 and 9.10 with $\{|\max H_N/N - e_p| \leq \varepsilon_0\}$. These lemmas and (9.18) imply that $\mathbb{P}(\mathcal{E}^c) \leq e^{-cN}$ for some $c > 0$.

On $\mathcal{E}$, let $f = H_N$ and $f_* = \max H_N$. Every critical point c with $f(c) > f_* - \tau_0 N/2$ satisfies $f(c) > N(e_p - \varepsilon_0 - \tau_0/2) \geq N(e_p - \tau_0)$ and $\nabla^2_{S_N} f(c) \leq -\kappa I$. We apply Lemma 9.11 with $\delta_0 = \tau_0/2$ and $\delta = 3\varepsilon_0$; note that $3\varepsilon_0 < \tau_0/2$. If $H_N(y) \geq N(e_p - \varepsilon_0)$, then $f_* - f(y) \leq 2\varepsilon_0 N < 3\varepsilon_0 N$, and y lies in a component of $\{f > f_* - 3\varepsilon_0 N\}$. Let u be its critical point. By (9.29), $H_N(u) \geq H_N(y)$ and $|u - y| \leq d_{S_N}(u,y) < 2\sqrt{3\varepsilon_0/\kappa}\,\sqrt{N} < r\sqrt{N}$, and u is a nondegenerate local maximum since $\nabla^2_{S_N} H_N(u) \leq -\kappa I$. □

Proof of Proposition 9.1. Let $M > 0$ be fixed and $L = Q^{-p/2}M$. By (9.8), $\mathcal{C}_{N,M} \subset \mathcal{D}_{N,L}$ for every N. Using (9.20) with $\varphi = 1$ and (9.25) with $a_N = a'_N = L$, we obtain, for all $N \geq N_0$,

$$\mathbb{E}|\mathcal{D}_{N,L}| = \int_{M_N - L}^{M_N + L} \rho_N(h)\,\mathrm{d}h \leq C \int_{-L}^{L} e^{-c_p t}\,\mathrm{d}t \leq 2CLe^{c_p L}.$$

For each of the finitely many $N < N_0$, the integral is finite because ρ_N is continuous. This proves (9.9).

For (9.10), let $\varepsilon > 0$. By Lemma 9.8, there exists L such that $\liminf_N \mathbb{P}\{|\max H_N - M_N| \leq L\} \geq 1 - \varepsilon$. On this event, a maximizer u_* of H_N is a critical point with $|H_N(u_*) - M_N| \leq L$. Then $u_* \in \mathcal{D}_{N,L}$, and $u_* \in \mathcal{C}_{N,Q^{p/2}L}$ for all sufficiently large N by (9.8). Further, $\max H_Q = Q^{p/2} \max H_N \leq m_N + Q^{p/2}L$. As the event in (9.10) increases with M, its probability has lim inf at least $1 - \varepsilon$ for every $M \geq Q^{p/2}L$. This proves (9.10).

Finally, let M be fixed. For all sufficiently large N, every $u \in \mathcal{C}_{N,M}$ satisfies $H_N(u) \geq M_N - Q^{-p/2}M \geq N(e_p - \tau_0)$. By Lemma 9.10, with probability at least $1 - e^{-cN}$, every such u satisfies $\nabla^2_{S_N} H_Q(u) = Q^{p/2}\nabla^2_{S_N} H_N(u) \leq -Q^{p/2}\kappa I$, and it is a nondegenerate local maximum of H_Q. □

9.5 Empirical means of fixed replica groups

In this section, we show that if n samples from the Gibbs measure, with n large but fixed, have pairwise overlaps close to Q, then their empirical mean, rescaled to squared norm NQ, has energy close to NE_0. The proof uses Subag's free energies of replicated bands. The same inputs are stated for general mixtures in Section 12.3.

Let $m \in \mathbb{R}^N$ with $|m|^2 = NQ$, so that $m = \sqrt{Q}\,u$ with $u \in S_N$, and let $n \geq 1$ and $\rho, \delta > 0$. We define

$$\begin{aligned}\mathcal{T}_N(m, n, \rho, \delta) = \big\{(\sigma^1, \dots, \sigma^n) \in S_N^n : {} & |R(\sigma^i, m) - Q| \leq \delta\sqrt{Q} \text{ for all } i, \\ & |R(\sigma^i, \sigma^j) - Q| < \rho \text{ for all } i \neq j\big\},\end{aligned}$$

and

$$\widetilde{P}_N(m, n, \rho, \delta) = \frac{1}{Nn} \log \int_{\mathcal{T}_N(m,n,\rho,\delta)} \exp\Big(\beta \sum_{i=1}^{n} \big[H_N(\sigma^i) - H_N(m)\big]\Big) \nu_N^{\otimes n}(\,\mathrm{d}\sigma). \tag{9.32}$$

The condition on $R(\sigma^i, m)$ says that $\sigma^i \in B_u(\delta)$, as $R(\sigma, m) = \sqrt{Q}\,R(\sigma, u)$. This is Subag's free energy of the replicated band [143, (1.10)], with the squared norm $q = Q$. For $N \geq n + 1$, the set $\mathcal{T}_N(m, n, \rho, \delta)$ has nonempty interior: it contains the points $\sigma^i = m + \sqrt{Ns}\,\omega_i$ for orthonormal vectors $\omega_1, \dots, \omega_n$ orthogonal to m, for which $R(\sigma^i, m) = Q$ and $R(\sigma^i, \sigma^j) = Q$, so that all the constraints hold with strict slack at this tuple. Hence $\mathcal{T}_N(m, n, \rho, \delta)$ has positive measure, and $\widetilde{P}_N$ is finite. The expectation $\mathbb{E}\widetilde{P}_N(m, n, \rho, \delta)$ does not depend on m, by rotation invariance.

Proposition 9.13 (Uniform concentration; [143, Proposition 24])**.** *Let $t, c > 0$. There exist $\delta_0, \rho_0 > 0$ and an integer n_0 such that for all $\delta \in (0, \delta_0]$, $\rho \in (0, \rho_0]$, $n \geq n_0$, and $N \geq n+1$,*

$$\mathbb{P}\Big\{ \sup_{|m|^2 = NQ} \big[\widetilde{P}_N(m,n,\rho,\delta) - \mathbb{E}\widetilde{P}_N(m,n,\rho,\delta)\big] < t \Big\} > 1 - e^{-cN}.$$

This is [143, Proposition 24, (3.1)], which holds for every mixture with exponentially decaying coefficients, in particular for the pure covariance of either parity, and for every $q \in [0,1)$ in place of Q. We restrict to $N \geq n+1$, where the replicated set has positive measure; for each fixed n, this excludes finitely many dimensions. The parameters δ, ρ, n in the proposition are fixed: they are chosen before $N \to \infty$. Subag also proved the estimates that relate $\widetilde{P}_N$ to the free energy of the residual field on the exact latitude [143, Lemmas 17–19]: a bound $L\sqrt{N}$ on the Lipschitz constant of H_N on the ball of radius $\sqrt{N}$, with probability at least $1 - e^{-c'N}$ for L large [143, Lemma 17]; the projection of a band of width δ onto its exact latitude, which raises the pair tolerance from ρ to $\rho + \sqrt{8\delta}$ and costs $N^{-1} \log \nu_N(B_u(\delta)) + \beta L\sqrt{2\delta}$ [143, Lemma 18, (2.2)]; and the removal of the field of degree one, at a cost at most a universal constant times $\beta\gamma_1\sqrt{1/n + \rho}$, where γ_1^2 is its coefficient [143, Lemma 19, (2.5)]. The proof of the next lemma follows these arguments, with fixed n, ρ, and δ, and it gives the upper bound we need directly.

Lemma 9.14 (Expected free energy of a replicated band)**.** *Let $T_Q = (\log s)/2 + \beta^2 \psi_Q(1)/2$. There exists C such that for all $\delta \in (0, \sqrt{Q}/2]$, $\rho > 0$, and $n \geq 1$,*

$$\limsup_{N\to\infty} \mathbb{E}\widetilde{P}_N(m,n,\rho,\delta) \leq T_Q + C\sqrt{\delta} + C\Big(\frac{1}{n} + \frac{\rho + \sqrt{8\delta}}{s}\Big)^{1/2}. \tag{9.33}$$

Note that $F(\beta) = \beta E_0 + T_Q$ by (9.4), since $E_0 = e_p Q^{p/2}$.

Proof. Let $u = m/\sqrt{Q}$. We bound the integral in (9.32) in three steps: we project the replicas to the exact latitude $R(\cdot, u) = \sqrt{Q}$, we separate the latitude from the direction orthogonal to u, and we bound the field on the exact latitude.

Projection to the exact latitude. We write $\sigma \in B_u(\delta)$ as $\sigma = ru + \sqrt{N(1-r^2)}\,\omega$, with $|r - \sqrt{Q}| \leq \delta$ and ω a unit vector orthogonal to u, and let $\sigma' = \sqrt{Q}\,u + \sqrt{Ns}\,\omega$. If $r = \cos\phi$ and $\sqrt{Q} = \cos\phi'$ with $\phi, \phi' \in [0, \pi]$, then

$$|\sigma - \sigma'|^2 = 4N \sin^2 \frac{\phi - \phi'}{2} \leq 4N \Big|\sin\frac{\phi + \phi'}{2} \sin\frac{\phi - \phi'}{2}\Big| = 2N|\cos\phi - \cos\phi'| \leq 2N\delta,$$

where we used $|\phi - \phi'|/2 \leq (\phi + \phi')/2 \leq \pi - |\phi - \phi'|/2$. This is the bound $|\sigma - \sigma'| \leq \sqrt{2N\delta}$ of [143, Lemma 18], and for two such points $|R(\sigma'^i, \sigma'^j) - R(\sigma^i, \sigma^j)| \leq 2\sqrt{2\delta} = \sqrt{8\delta}$. Let Lip be the Lipschitz constant of H_N on the ball of radius $\sqrt{N}$, which contains the segment between σ and σ'. Then

$$\sum_{i=1}^n \big[H_N(\sigma^i) - H_N(m)\big] \leq \sum_{i=1}^n \big[H_N(\sigma'^i) - H_N(m)\big] + n\,\mathrm{Lip}\sqrt{2N\delta}.$$

Latitude and direction. Under ν_N, the latitude $r = R(\sigma, u)$ has density $\ell_N(1-r^2)^{(N-3)/2}$ on $(-1,1)$, with

$$\ell_N = \frac{\Gamma(N/2)}{\sqrt{\pi}\,\Gamma((N-1)/2)} \le \sqrt{N}, \tag{9.34}$$

and, given r, the direction ω is uniform on the unit sphere of $u^\perp$. If $(\sigma^i) \in \mathcal{T}_N(m,n,\rho,\delta)$, then $|R(\sigma'^i,\sigma'^j) - Q| < \rho + \sqrt{8\delta}$, and since $R(\sigma'^i,\sigma'^j) = Q + s\,\omega_i\cdot\omega_j$, we have $|\omega_i\cdot\omega_j| < \rho' = (\rho+\sqrt{8\delta})/s$. The integrand after projection depends only on the directions, and it is nonnegative. Enlarging the domain of integration, we obtain

$$\int_{\mathcal{T}_N} \exp\Big(\beta\sum_i \big[H_N(\sigma^i) - H_N(m)\big]\Big)\,\mathrm{d}\nu_N^{\otimes n} \le \nu_N\big(B_u(\delta)\big)^n\, e^{n\beta\,\mathrm{Lip}\sqrt{2N\delta}} \times \int_{\{|\omega_i\cdot\omega_j|<\rho'\}} e^{\beta\sum_i X(\omega_i)}\,\mathrm{d}\omega,$$

where $X(\omega) = H_N(\sqrt{Q}\,u + \sqrt{Ns}\,\omega) - H_N(m)$ and $\mathrm{d}\omega$ is the product of uniform probability measures on the unit sphere of $u^\perp$. Since $1-r^2 \le 1-(\sqrt{Q}-\delta)^2$ on the band and $\delta \le \sqrt{Q}/2$, we have

$$\frac{1}{N}\log\nu_N\big(B_u(\delta)\big) \le \frac{1}{N}\log\big(2\delta\sqrt{N}\big) + \frac{N-3}{2N}\log\big(s + 2\sqrt{Q}\delta\big) \le \frac{1}{2}\log s + \frac{2\delta}{s} + o(1).$$

The field on the exact latitude. Since $R(\sqrt{Q}u + \sqrt{Ns}\,\omega, \sqrt{Q}u + \sqrt{Ns}\,\omega') = Q + s\,\omega\cdot\omega'$ and $R(\sqrt{Q}u + \sqrt{Ns}\,\omega, m) = Q$, the covariance of X is

$$\mathbb{E}X(\omega)X(\omega') = N\big[(Q + s\,\omega\cdot\omega')^p - Q^p\big] = N\big[pQ^{p-1}s\,\omega\cdot\omega' + \psi_Q(\omega\cdot\omega')\big].$$

Since the polynomial ψ_Q has nonnegative coefficients, we may write $X(\omega) = \sqrt{N}\gamma_1 g\cdot\omega + Y(\omega)$, with $\gamma_1 = (pQ^{p-1}s)^{1/2}$, g a standard Gaussian vector in $u^\perp$, and Y an independent centered Gaussian field with covariance $N\psi_Q(\omega\cdot\omega')$. On $\{|\omega_i\cdot\omega_j| < \rho'\}$, the Cauchy–Schwarz inequality gives

$$\Big|g\cdot\sum_{i=1}^n \omega_i\Big| \le |g|\Big|\sum_{i=1}^n\omega_i\Big| \le |g|\big(n + n(n-1)\rho'\big)^{1/2} \le n|g|\Big(\frac{1}{n} + \rho'\Big)^{1/2}.$$

After this bound, we drop the constraint on the directions, so that the remaining integral factorizes into n identical factors. By Jensen's inequality,

$$\frac{1}{N}\mathbb{E}\log\int e^{\beta Y(\omega)}\,\mathrm{d}\omega \le \frac{1}{N}\log\int \mathbb{E}e^{\beta Y(\omega)}\,\mathrm{d}\omega = \frac{\beta^2}{2}\psi_Q(1).$$

Conclusion. Combining the three steps, taking expectations, and using $\mathbb{E}|g| \le \sqrt{N}$ and $\mathbb{E}\,\mathrm{Lip} \le C\sqrt{N}$ (Lemma 9.4), we find

$$\mathbb{E}\widetilde{P}_N(m,n,\rho,\delta) \le \frac{1}{2}\log s + \frac{2\delta}{s} + o(1) + C\beta\sqrt{2\delta} + \beta\gamma_1\Big(\frac{1}{n}+\rho'\Big)^{1/2} + \frac{\beta^2}{2}\psi_Q(1).$$

Since $\delta \le \sqrt{\delta}$, this implies (9.33). The bound uses the expected Lipschitz constant; no contribution from the complement of a regularity event is lost. □

Proposition 9.15 (Empirical means of Gibbs replica groups)**.** *Let $\varepsilon > 0$. There exist an integer n_0 and $\eta, c > 0$ such that for every fixed integer $n \ge n_0$ and all sufficiently large N, depending on n,*

$$\mathbb{E}\,\pi^{\otimes n}\Big\{|R(\sigma^i,\sigma^j) - Q| < \eta \text{ for all } i \ne j,\ \ H_Q(\hat u) \le N(E_0 - \varepsilon)\Big\} \le e^{-cN}, \tag{9.35}$$

where $\bar\sigma = n^{-1}\sum_{i=1}^n \sigma^i$, $\bar q = |\bar\sigma|^2/N$, and $\hat u = \bar\sigma/\sqrt{\bar q} \in S_N$. On the event in (9.35)*, we have $\bar q > 0$, and $\hat u$ is well defined.*

The diagonal of the overlap matrix enters the mean: since $R(\sigma^i,\sigma^i) = 1$, the squared norm of $\bar\sigma$ is close to $Q + (1-Q)/n$ and not to Q. For this reason the number n of replicas is fixed and large, and it is chosen before $N \to \infty$.

Proof. We first relate $\hat u$ to the replicas, then choose the parameters, and finally bound the Gibbs probability by a union over a net.

Geometry of the mean. Let $q_n = Q + (1-Q)/n$, and suppose that $|R(\sigma^i,\sigma^j) - Q| < \eta$ for all $i \ne j$, with $\eta < Q/2$. Since $\bar q = n^{-2}\sum_{i,j} R(\sigma^i,\sigma^j)$ and $R(\sigma^i,\bar\sigma) = n^{-1}\sum_j R(\sigma^i,\sigma^j)$, with diagonal terms equal to one, we have

$$|\bar q - q_n| \le \eta, \qquad |R(\sigma^i,\bar\sigma) - \bar q| \le 2\eta.$$

In particular, $\bar q \ge Q - \eta > Q/2$. Writing $R(\sigma^i,\sqrt{Q}\hat u) - Q = \sqrt{Q/\bar q}\,[R(\sigma^i,\bar\sigma) - \bar q] + \sqrt{Q\bar q} - Q$, and using $|\sqrt{Q\bar q} - Q| \le |\bar q - Q| \le \eta + 1/n$ and $\sqrt{Q/\bar q} \le \sqrt{2}$, we obtain

$$|R(\sigma^i, \sqrt{Q}\,\hat u) - Q| \le \frac{1}{n} + 4\eta. \tag{9.36}$$

Choice of the parameters. Let L be such that the event $\{\mathrm{Lip} \le L\sqrt{N}\}$, with Lip as in the proof of Lemma 9.14, has probability at least $1 - e^{-c_2 N}$ for some $c_2 > 0$; such an L exists by Lemma 3.2(3). Let δ_0, ρ_0, n_0' be the parameters given by Proposition 9.13 with $t = \beta\varepsilon/8$ and $c = 1$. Fix $\delta \in (0, \min(\delta_0, \sqrt{Q}/2)]$, $\rho \in (0,\rho_0]$, and n_1 so that the error terms in (9.33) are at most $\beta\varepsilon/16$ for every $n \ge n_1$. Then for every fixed $n \ge n_1$ and all sufficiently large N,

$$\mathbb{E}\widetilde{P}_N(m,n,\rho,\delta) \le T_Q + \frac{\beta\varepsilon}{8}. \tag{9.37}$$

Let $\vartheta > 0$ be a net radius with $\vartheta \le \delta\sqrt{Q}/4$ and $L\vartheta \le \varepsilon/4$. The sphere $\{|m|^2 = NQ\}$ has a $\vartheta\sqrt{N}$-net $\mathcal{N}$ with at most $e^{A_\vartheta N}$ points, where $A_\vartheta = \log(1 + 2\sqrt{Q}/\vartheta)$. Let $n_0 \ge \max(n_0', n_1)$ satisfy $1/n_0 \le \delta\sqrt{Q}/4$ and $n_0\beta\varepsilon/4 \ge 2A_\vartheta$, and choose $\eta \in (0, \min(Q/2, \rho, \delta\sqrt{Q}/16))$. All these choices are made before $N \to \infty$, and they remain valid for every fixed $n \ge n_0$.

A union over the net. Let $n \ge n_0$, and let $\mathcal{E}$ be the intersection of the events $\{\mathrm{Lip} \le L\sqrt{N}\}$, $\{P_N \ge F(\beta) - \beta\varepsilon/8\}$, and the event of Proposition 9.13. By Lemma 3.3 and $\mathbb{E}P_N \to F(\beta)$, the second event has probability at least $1 - 2e^{-cN}$ for large N. We conclude that $\mathbb{P}(\mathcal{E}^c) \le e^{-c_3 N}$ for some $c_3 > 0$ and large N. Suppose that a tuple (σ^i) belongs to the event in (9.35), and let $m_j \in \mathcal{N}$ satisfy $|m_j - \sqrt{Q}\hat{u}| \le \vartheta\sqrt{N}$. Then $|R(\sigma^i, m_j) - R(\sigma^i, \sqrt{Q}\hat{u})| \le \vartheta$, and (9.36) and the choice of the parameters give $|R(\sigma^i, m_j) - Q| \le 3\delta\sqrt{Q}/4$, and $|R(\sigma^i, \sigma^k) - Q| < \eta \le \rho$ for $i \ne k$. This means that $(\sigma^i) \in \mathcal{T}_N(m_j, n, \rho, \delta)$. On $\mathcal{E}$, we also have

$$H_N(m_j) \le H_N(\sqrt{Q}\hat{u}) + L\vartheta N = H_Q(\hat{u}) + L\vartheta N \le N(E_0 - 3\varepsilon/4),$$

where we used $L\vartheta \le \varepsilon/4$. On $\mathcal{E}$, for every such net point m_j,

$$\pi^{\otimes n}\big(\mathcal{T}_N(m_j, n, \rho, \delta)\big) = \exp\Big\{Nn\Big[\frac{\beta}{N}H_N(m_j) + \widetilde{P}_N(m_j, n, \rho, \delta) - P_N\Big]\Big\} \le e^{-3Nn\beta\varepsilon/8}.$$

Indeed, on $\mathcal{E}$, the event of Proposition 9.13 and (9.37) give $\widetilde{P}_N(m_j, n, \rho, \delta) \le T_Q + \beta\varepsilon/4$, and $P_N \ge F(\beta) - \beta\varepsilon/8 = \beta E_0 + T_Q - \beta\varepsilon/8$; the exponent is therefore at most $Nn[-3\beta\varepsilon/4 + 3\beta\varepsilon/8]$. Summing over the net, we find that on $\mathcal{E}$ the Gibbs probability of the event in (9.35) is at most $e^{A_\vartheta N - 3Nn\beta\varepsilon/8} \le e^{-Nn\beta\varepsilon/8}$, where we used $n\beta\varepsilon/4 \ge 2A_\vartheta$. Taking expectations, we find that the left side of (9.35) is at most $e^{-c_3 N} + e^{-Nn\beta\varepsilon/8}$, which is at most e^{-cN} with $c = \min(c_3, \beta\varepsilon/8)/2$ for all sufficiently large N. □

9.6 A coarse cover from the directing measure

For $\tau > 0$, let

$$\mathcal{U}_N(\tau) = \bigcup_{\substack{u:\nabla_{S_N} H_N(u) = 0 \\ H_N(u) \ge N(e_p - \tau)}} \big\{x \in S_N : R(x, u) \ge a\big\},$$

where a is the cap latitude of Lemma 9.3. The set of pairs (H, x) with $x \in \mathcal{U}_N(\tau)$ is closed, by the projection argument in the proof of Lemma 4.3, and $\pi(\mathcal{U}_N(\tau)^c)$ is a Borel function of the field. In this section, we show that the caps in $\mathcal{U}_N(\tau)$ carry asymptotically all of the Gibbs measure, for every fixed $\tau > 0$.

We use the notation of Section 6.1. Let $\sigma^1, \sigma^2, \dots$ be independent samples from π, conditionally on the field, let $\mathcal{L}_N$ be the annealed law of the overlap array $(R(\sigma^i, \sigma^j))_{i,j \ge 1}$ on the compact space $\mathcal{M} = [-1, 1]^{\mathbb{N}\times\mathbb{N}}$, and let $\mathfrak{L}$ be the set of limiting overlap laws, that is, of subsequential weak limits of $(\mathcal{L}_N)$. By Proposition 5.20, Theorem 6.3 applies, with the radius Q of Section 9.1. Together with Theorem 6.2, it gives the following. Every $\mathcal{L} \in \mathfrak{L}$ has a directing measure $(G_\omega)_{\omega\in\Omega}$, so that an array with law $\mathcal{L}$ is obtained by drawing ω, then independent vectors $v_1, v_2, \dots$ in the unit ball B of ℓ^2 with law G_ω, and setting $R_{ii} = 1$ and $R_{ij} = v_i \cdot v_j$ for $i \ne j$; and for almost every ω, $G_\omega\{v : \|v\|^2 = Q\} = 1$. No other property of the directing measure is used.

Proposition 9.16 (A coarse cover)**.** *For every fixed $\tau > 0$,*

$$\lim_{N\to\infty} \mathbb{E}\, \pi\big(\mathcal{U}_N(\tau)^c\big) = 0. \tag{9.38}$$

The idea is as follows. If many replicas are drawn, the directing measure guarantees that some n of them, together with a fresh sample σ^1, have all pairwise overlaps close to Q. By Proposition 9.15, the normalized mean of these n replicas has near-ground energy, and by Proposition 9.12, it is close to a critical maximum. The fresh sample then lies in the cap of that critical point. Only finitely many replicas can be used at finite N, and their number must be chosen uniformly in N. This is where a compactness argument enters.

Proof. Let $\tau > 0$ and $\varepsilon > 0$ be fixed. The proof has four steps: the choice of the parameters, a continuous score on the overlap array, a uniform bound on the score at finite N, and the transfer to $\mathcal{U}_N(\tau)$.

Parameters. Fix $a_1 \in (a, \sqrt{Q})$ and $r \in (0, a_1 - a)$. Let ε_0 be given by Proposition 9.12 for this r, and let $\tau_1 = \min(\tau, \varepsilon_0)$. Let n_0, η, and c be the constants of Proposition 9.15 with $Q^{p/2}\tau_1$ in place of ε. Since $Q > a_1\sqrt{Q}$, we may fix an integer $n \geq n_0$ such that, with $q_n = Q + (1-Q)/n$,

$$\kappa_1 = \frac{1}{2}\big(Q - a_1\sqrt{q_n}\big) > 0. \tag{9.39}$$

A continuous score. For an array $R = (R_{ij})_{i,j\geq 1}$ in $\mathcal{M}$ and a set $I \subset \{2, 3, \dots\}$ with $|I| = n$, let

$$b_I = \frac{1}{n}\sum_{i\in I} R_{i1}, \qquad w_I = \frac{1}{n^2}\sum_{i,j\in I} R_{ij}, \qquad e_I = \max_{i,j\in I,\, i\neq j} |R_{ij} - Q|.$$

Let $\varphi_1, \varphi_2, \varphi_3 \colon \mathbb{R} \to [0,1]$ be continuous, with $\varphi_1 = 1$ on $(-\infty, \eta/2]$ and $\varphi_1 = 0$ on $[\eta, \infty)$, $\varphi_2 = 0$ on $(-\infty, Q/4]$ and $\varphi_2 = 1$ on $[Q/2, \infty)$, and $\varphi_3 = 0$ on $(-\infty, 0]$ and $\varphi_3 = 1$ on $[\kappa_1, \infty)$. We define the score

$$S_I = \varphi_1(e_I)\, \varphi_2(w_I)\, \varphi_3\big(b_I - a_1\sqrt{\max(w_I, 0)}\big) \in [0,1],$$

a continuous function of finitely many entries of the array. If $S_I > 0$, then $e_I < \eta$, $w_I > Q/4$, and $b_I > a_1\sqrt{w_I}$. For the overlaps of samples, $w_I = |\bar\sigma_I|^2/N$ and $b_I = R(\sigma^1, \bar\sigma_I)$, where $\bar\sigma_I = n^{-1}\sum_{i\in I}\sigma^i$. Thus, if $S_I > 0$, the point $\hat w_I = \bar\sigma_I/\sqrt{w_I} \in S_N$ satisfies

$$R(\sigma^1, \hat w_I) = \frac{b_I}{\sqrt{w_I}} > a_1, \tag{9.40}$$

and the samples σ^i, $i \in I$, satisfy the pair condition of Proposition 9.15, with $\hat w_I$ in the role of the point $\hat u$.

A uniform number of replicas. For $T \geq n$ and a probability measure $\mathcal{L}$ on $\mathcal{M}$, let

$$f_T(\mathcal{L}) = \mathbb{E}_{\mathcal{L}}\Big[1 - \max_{I \subset \{2,\dots,T+1\},\, |I|=n} S_I\Big].$$

Since $f_T(\mathcal{L})$ is the expectation of a bounded continuous function of finitely many coordinates, f_T is continuous in $\mathcal{L}$. It decreases in T. Let $\mathcal{L} \in \mathfrak{L}$, with a directing measure as above, and fix ω outside a null set. Almost surely, v_1 lies in the support of G_ω, since the support of a Borel probability measure on a separable metric space has full measure, and $\|v_1\|^2 = Q$. Let $r' > 0$ be small. If v_i, v_j, for distinct indices $i, j \in I$, lie in the ball $B(v_1, r')$ and have squared norm Q, then $|v_i \cdot v_j - Q| \leq 2r'\sqrt{Q}$ and $|v_i \cdot v_1 - Q| \leq r'\sqrt{Q}$, so that $e_I \leq 2r'\sqrt{Q}$, $|b_I - Q| \leq r'\sqrt{Q}$, and $|w_I - q_n| \leq 2r'\sqrt{Q}$; here we used $R_{ii} = 1$. For r' small enough, depending only on Q, η, a_1, and κ_1, these bounds and (9.39) imply that $S_I = 1$. The ball $B(v_1, r')$ has positive G_ω-mass, so, conditionally on ω and v_1, infinitely many of the vectors $v_2, v_3, \dots$ lie in it almost surely. The maximum of S_I over n-subsets I of $\{2, \dots, T+1\}$ is then equal to one for all large T, almost surely, and by dominated convergence, $f_T(\mathcal{L}) \to 0$ as $T \to \infty$.

The set $\mathfrak{L}$ is a closed subset of the compact space of Borel probability measures on $\mathcal{M}$, and it is compact. The functions f_T are continuous, decrease in T, and converge to zero pointwise on $\mathfrak{L}$. By Dini's theorem, the convergence is uniform on $\mathfrak{L}$. We fix T such that $f_T < \varepsilon$ on $\mathfrak{L}$. Then $\limsup_N f_T(\mathcal{L}_N) \leq \varepsilon$. Indeed, otherwise a subsequence would satisfy $f_T(\mathcal{L}_N) > \varepsilon$, a further subsequence would converge to some $\mathcal{L} \in \mathfrak{L}$, and continuity would give $f_T(\mathcal{L}) \geq \varepsilon$.

Transfer to caps around critical points. Let $\mathcal{I}$ be the set of $I \subset \{2, \dots, T+1\}$ with $|I| = n$ and $e_I < \eta$, and let $\mathcal{W} = \bigcup_{I \in \mathcal{I}} \{x \in S_N : R(x, \hat{w}_I) > a_1\}$. Conditionally on the field and on $\sigma^2, \dots, \sigma^{T+1}$, (9.40) shows that $\pi(\mathcal{W}^c) \leq \mathbb{E}[1 - \max_I S_I \mid H, \sigma^2, \dots, \sigma^{T+1}]$. Taking expectations, we find that the annealed probability of $\{\sigma^1 \notin \mathcal{W}\}$ is at most $f_T(\mathcal{L}_N)$.

Let $\mathcal{G}_1$ be the event that $H_Q(\hat{w}_I) > N(E_0 - Q^{p/2}\tau_1)$ for every $I \in \mathcal{I}$. Applying Proposition 9.15 to each of the $\binom{T}{n}$ subsets, we find that the annealed probability of $\mathcal{G}_1^c$ is at most $\binom{T}{n} e^{-cN}$ for large N. On $\mathcal{G}_1$, we have $H_N(\hat{w}_I) = Q^{-p/2} H_Q(\hat{w}_I) > N(e_p - \tau_1)$ for every $I \in \mathcal{I}$. Let $\mathcal{G}_2$ be the event of Proposition 9.12 for r, with $\tau_1 \leq \varepsilon_0$ in place of ε_0; it depends only on the field and has probability at least $1 - e^{-c'N}$. On $\mathcal{G}_1 \cap \mathcal{G}_2$, every $\hat{w}_I$ with $I \in \mathcal{I}$ has a critical point u_I with $|u_I - \hat{w}_I| \leq r\sqrt{N}$ and $H_N(u_I) \geq H_N(\hat{w}_I) > N(e_p - \tau)$. For every $x \in S_N$, we have $|R(x, u_I) - R(x, \hat{w}_I)| \leq |u_I - \hat{w}_I| / \sqrt{N} \leq r$, and $R(x, \hat{w}_I) > a_1$ implies that $R(x, u_I) > a_1 - r > a$. This shows that $\mathcal{W} \subset \mathcal{U}_N(\tau)$ on $\mathcal{G}_1 \cap \mathcal{G}_2$. Bounding $\pi(\mathcal{U}_N(\tau)^c)$ by one on the complement, we obtain

$$\mathbb{E}\, \pi\big(\mathcal{U}_N(\tau)^c\big) \leq f_T(\mathcal{L}_N) + \binom{T}{n} e^{-cN} + e^{-c'N}.$$

Taking $\limsup_N$, we find $\limsup_N \mathbb{E}\,\pi(\mathcal{U}_N(\tau)^c) \le \varepsilon$. The set $\mathcal{U}_N(\tau)$ does not depend on ε or on the replicas, and letting $\varepsilon \downarrow 0$ proves (9.38). The parameters $a_1, r, \tau_1, n, \eta, T$ are fixed before $N \to \infty$. □

The proof considers all subsets of a fixed number of replicas. It uses no decomposition of the Gibbs measure into pure states, no property of the atoms or weights of the directing measure, and no choice of signs. It applies to both parities.

9.7 Band partition functions

For $u \in S_N$ and a Borel set $I \subset [a, 1)$, we define

$$Z_u(I) = \int_{S_N} \mathbf{1}\{R(x,u) \in I\}\, e^{\beta H_N(x)}\, \nu_N(\mathrm{d}x), \qquad \overline{Z}_{N,h}(I) = \mathbb{E}_{u;h} Z_u(I).$$

By Tonelli's theorem, $Z_u(I)$ is a Borel function of (H, u). For $g \in O(N)$, the partition function of the field $H_N \circ g^{-1}$ around gu equals $Z_u(I)$. By rotation invariance, the law of $Z_u(I)$ under $\mathbb{P}_{u;h}$ does not depend on u, and neither does $\overline{Z}_{N,h}(I)$. Under ν_N, the latitude $q = R(x,u)$ has density $\ell_N(1-q^2)^{(N-3)/2}$, with ℓ_N as in (9.34), and the direction ω in Lemma 9.5 is uniform and independent of q. By (9.23) and Tonelli's theorem,

$$\overline{Z}_{N,h}(I) = \ell_N \int_I (1-q^2)^{(N-3)/2} \exp\Big\{\beta h q^p + \frac{N}{2} C_\beta(q)\Big\}\, \mathrm{d}q, \tag{9.41}$$

where we used $\mathbb{E}\exp(\beta\sqrt{N}\sum_{k\ge2}\alpha_k(q)X_k(\omega)) = \exp(N\beta^2\sum_{k\ge2}\alpha_k(q)^2/2)$ and the binomial identity $\sum_{k=2}^p \alpha_k(q)^2 = 1 - q^{2p} - pq^{2p-2}(1-q^2)$, so that $\beta^2\sum_{k\ge2}\alpha_k(q)^2 = C_\beta(q)$. The formula (9.41) is exact, and $\overline{Z}_{N,h}(I) > 0$ if I has positive Lebesgue measure.

Lemma 9.17 (Spherical moments)**.** *Let $m \ge 1$, let ω, ω' be independent uniform unit vectors in $\mathbb{R}^m$, and let $r = \omega \cdot \omega'$. For every integer $k \ge 0$,*

$$\mathbb{E} r^{2k} = \frac{(2k-1)!!}{m(m+2)\cdots(m+2k-2)},$$

where the right side is one for $k = 0$. For every real $\lambda < m/2$,

$$\mathbb{E} e^{\lambda r^2} \le \Big(1 - \frac{2\lambda}{m}\Big)^{-1/2}. \tag{9.42}$$

Proof. By rotation invariance, r has the law of the first coordinate of ω'. Let g be a standard Gaussian vector in $\mathbb{R}^m$. Then ω' has the law of $g/|g|$, and $g/|g|$ is independent of $|g|^2$. For $m \ge 2$, r^2 has the $\mathrm{Beta}(1/2, (m-1)/2)$ law, whose k-th moment is $\prod_{j=0}^{k-1}(2j+1)/(m+2j)$, which is the displayed formula; for $m = 1$, we have $r^2 = 1$, which agrees with it. For (9.42), let $Y = g_1^2/|g|^2$, which has the law of r^2, and let $G = g_1^2/m = (|g|^2/m)\,Y$. Since $|g|^2$ is independent of Y and $\mathbb{E}|g|^2 = m$, we have $\mathbb{E}[G \mid Y] = Y$. Using the conditional Jensen inequality, we obtain $\mathbb{E}e^{\lambda Y} \le \mathbb{E}e^{\lambda G} = (1 - 2\lambda/m)^{-1/2}$. □

The next lemma turns a bound on an exponential moment of the covariance into a lower tail bound for a Gaussian partition function. Its proof adapts the two-replica coupling interpolation and decreasing coupling path of Guerra and Toninelli [85, Section 3, (19)–(26)]. We give the argument for a general finite-dimensional Gaussian covariance, which may take negative values. A bounded second moment alone would not give such a bound uniformly.

Lemma 9.18 (Lower tail by interpolation)**.** *Let X be a centered Gaussian field on a compact space V that is a linear function of a finite-dimensional standard Gaussian vector, with continuous covariance $K(v,w)$, and let μ be a deterministic probability measure on V. Suppose that there exist $\lambda > 1$ and $C \geq 0$ with*

$$\iint e^{\lambda K(v,w)}\, \mu(\mathrm{d}v)\, \mu(\mathrm{d}w) \leq e^{C}. \tag{9.43}$$

Let $W = \int e^{X(v)-K(v,v)/2}\mu(\mathrm{d}v)$. Then

$$\mathbb{E}\log^{-} W \leq 1 + \frac{C}{2(\lambda-1)}, \qquad \mathbb{P}\{W < e^{-T}\} \leq \frac{1}{T}\Big(1 + \frac{C}{2(\lambda-1)}\Big) \quad (T>0). \tag{9.44}$$

Proof. For $0 \leq t \leq 1$, let $W_t = \int e^{\sqrt{t}X(v)-tK(v,v)/2}\mu(\mathrm{d}v)$, let $\langle\cdot\rangle_t$ be the expectation under the random probability measure proportional to $e^{\sqrt{t}X - tK/2}\mu$ and its independent copies, and let $D(t) = -\mathbb{E}\log W_t$. Then $D(0) = 0$. For $t > 0$, Gaussian integration by parts gives

$$D'(t) = \frac{1}{2}\mathbb{E}\big\langle K(v_1,v_2)\big\rangle_t,$$

and D is continuous on $[0,1]$; all integrals are finite since K is bounded. Writing $X(v) = g\cdot\Phi(v)$ with g standard Gaussian, we have $\langle K(v_1,v_2)\rangle_t = |\langle\Phi(v)\rangle_t|^2 \geq 0$, and D is nondecreasing.

Let

$$U_t = \iint \exp\Big\{\sqrt{t}\big[X(v)+X(v')\big] - \frac{t}{2}\big[K(v,v)+K(v',v')\big] + (\lambda-t)K(v,v')\Big\}\mu(\mathrm{d}v)\,\mu(\mathrm{d}v').$$

Computing the Gaussian expectation inside the integral, we find that $\mathbb{E}U_t$ is the left side of (9.43). Since $U_t = W_t^2\langle e^{(\lambda-t)K(v_1,v_2)}\rangle_t$, Jensen's inequality gives $\log U_t \geq 2\log W_t + (\lambda - t)\langle K(v_1,v_2)\rangle_t$. Taking expectations and using $\mathbb{E}\log U_t \leq \log \mathbb{E}U_t \leq C$, we obtain

$$-2D(t) + 2(\lambda - t)D'(t) \leq C.$$

This means that $t \mapsto (\lambda - t)(C + 2D(t))$ is nonincreasing, which gives $C + 2D(t) \leq C\lambda/(\lambda - t)$, that is, $D(t) \leq Ct/(2(\lambda - t))$. In particular, $D(1) \leq C/(2(\lambda-1))$. Finally, since $\mathbb{E}W = 1$ and $\log^{+} W \leq W$, we have $\mathbb{E}\log^{-} W = D(1) + \mathbb{E}\log^{+} W \leq D(1) + 1$. The second bound in (9.44) follows from Markov's inequality. □

The same conclusion holds when μ is carried by a Borel subset of V, since only integrals against μ enter the proof.

Proposition 9.19 (Band partition functions at a critical point)**.** *There exist C_2, C_- and N_0, depending only on p, β, and a, such that for all $N \ge N_0$, $h \in \mathbb{R}$, $u \in S_N$, and every Borel set $I \subset [a,1)$ of positive Lebesgue measure,*

$$\mathbb{E}_{u;h}\big[Z_u(I)^2\big] \le C_2\, \overline{Z}_{N,h}(I)^2, \tag{9.45}$$

$$\mathbb{P}_{u;h}\Big\{\frac{Z_u(I)}{\overline{Z}_{N,h}(I)} < e^{-T}\Big\} \le \frac{C_-}{T} \qquad (T>0). \tag{9.46}$$

Moreover, if $|h/N - e_p| \le \Delta/2$, then

$$\frac{\mathbb{E}_{u;h}\big[D(u)\,\mathbf{1}\{Z_u(I) < e^{-T}\overline{Z}_{N,h}(I)\}\big]}{\mu_N(h)} \le \Big(\frac{\mathsf{K}_2 C_-}{T}\Big)^{1/2} \qquad (T>0). \tag{9.47}$$

The bound (9.47) concerns the law weighted by the determinant, under which critical points are counted. It keeps the correlation between the Hessian and the residual field noted in Lemma 9.5.

Proof. Under $\mathbb{P}_{u;h}$, let $X(q,\omega) = \beta[H_N(qu + \sqrt{N(1-q^2)}\,\omega) - hq^p]$ for $q \in [a,1]$ and unit vectors $\omega \perp u$. By (9.23), X is a centered Gaussian field, linear in finitely many Gaussian variables, with covariance

$$K\big((q,\omega),(q',\omega')\big) = N\beta^2 \sum_{k=2}^{p} \alpha_k(q)\alpha_k(q')\,(\omega\cdot\omega')^k.$$

Let $r = \omega\cdot\omega'$. Since $|r|^k \le r^2$ for $k \ge 2$, the Cauchy–Schwarz inequality and the monotonicity of C_β give, for $q, q' \ge a$,

$$\big|K\big((q,\omega),(q',\omega')\big)\big| \le N\big(C_\beta(q)C_\beta(q')\big)^{1/2} r^2 \le N C_\beta(a)\, r^2. \tag{9.48}$$

Let $\mu_{N,h,I}$ be the probability measure under which q has density proportional to $\mathbf{1}_I(q)(1-q^2)^{(N-3)/2} e^{\beta h q^p + NC_\beta(q)/2}$ and, independently, ω is uniform on the unit sphere of $u^\perp$. Since $K((q,\omega),(q,\omega)) = NC_\beta(q)$, (9.41) shows that

$$W_{N,h,I} = \frac{Z_u(I)}{\overline{Z}_{N,h}(I)} = \int e^{X(v) - K(v,v)/2}\, \mu_{N,h,I}(\,\mathrm{d}v). \tag{9.49}$$

This identity is exact, for every Borel set I.

We have $C_\beta(a) < 1/2$ by (9.15). Fix $\lambda > 1$ with $2\lambda C_\beta(a) < 1$, and let $\vartheta \in (2\lambda C_\beta(a), 1)$. Let N_0 be such that $2\lambda N C_\beta(a)/(N-1) \le \vartheta$ for $N \ge N_0$. Using (9.48) and Lemma 9.17 in dimension $m = N-1$, we find, for $\lambda' \in \{1,\lambda\}$,

$$\iint e^{\lambda' K(v,v')}\, \mu_{N,h,I}(\,\mathrm{d}v)\, \mu_{N,h,I}(\,\mathrm{d}v') \le \mathbb{E} e^{\lambda N C_\beta(a) r^2} \le (1-\vartheta)^{-1/2},$$

uniformly in h, u, and I. With $\lambda' = 1$, the left side is $\mathbb{E}_{u;h} W_{N,h,I}^2$, which proves (9.45) with $C_2 = (1-\vartheta)^{-1/2}$. With $\lambda' = \lambda$, Lemma 9.18 with $C = -\log(1-\vartheta)/2$ implies (9.46) with $C_- = 1 + C/(2(\lambda-1))$. Finally, by the Cauchy–Schwarz inequality and (9.22),

$$\frac{\mathbb{E}_{u;h}\big[D(u)\mathbf{1}\{W_{N,h,I} < e^{-T}\}\big]}{\mu_N(h)} \le \frac{(\mathbb{E}_{u;h} D(u)^2)^{1/2}}{\mu_N(h)}\, \mathbb{P}_{u;h}\{W_{N,h,I} < e^{-T}\}^{1/2} \le \Big(\frac{\mathsf{K}_2 C_-}{T}\Big)^{1/2},$$

which is (9.47). The proof uses only $|r|^k \le r^2$ and does not require an even covariance. □

9.8 The annealed scale and the partition function

We define the annealed scale $\mathcal{A}_N$, which is the conditional mean of the partition function of the latitudes $[a,1)$ around a critical point with value M_N, and the microscopic latitude intervals

$$\mathcal{A}_N = \overline{Z}_{N,M_N}([a,1)), \qquad I_{N,w} = \big[\sqrt{Q} - wN^{-1/2}, \sqrt{Q} + wN^{-1/2}\big], \quad w > 0.$$

For fixed w and large N, $I_{N,w} \subset [a,1)$. Let $\Gamma(w)$ be the probability of $[-w,w]$ under the centered normal law with variance $1/k_*$, and let $\overline{\Gamma}(w) = 1 - \Gamma(w)$.

Proposition 9.20 (Laplace estimates over latitudes)**.** *We have*

$$\mathcal{A}_N = (1+o(1))\frac{s^{-3/2}}{\sqrt{k_*}}\, e^{NF(\beta)} N^{-1/(2x)}. \tag{9.50}$$

For every fixed $L, w > 0$, uniformly in $|t| \le L$,

$$\begin{gathered}\frac{\overline{Z}_{N,M_N+t}([a,1))}{\mathcal{A}_N} \to e^{\beta t Q^{p/2}}, \qquad \frac{\overline{Z}_{N,M_N+t}(I_{N,w})}{\mathcal{A}_N} \to e^{\beta t Q^{p/2}}\Gamma(w), \\ \frac{\overline{Z}_{N,M_N+t}([a,1)\setminus I_{N,w})}{\mathcal{A}_N} \to e^{\beta t Q^{p/2}}\,\overline{\Gamma}(w).\end{gathered} \tag{9.51}$$

In particular, for every fixed L, there exists $r_L > 0$ such that $\overline{Z}_{N,h}(I_{N,1}) \ge r_L \mathcal{A}_N$ for all $|h - M_N| \le L$ and all sufficiently large N. Finally, for every N and every $v \ge 0$,

$$\overline{Z}_{N,M_N - v}([a,1)) \le e^{-\beta a^p v}\mathcal{A}_N. \tag{9.52}$$

Proof. By (9.41) and (9.7),

$$\overline{Z}_{N,M_N+t}([a,1)) = \ell_N \int_a^1 (1-q^2)^{-3/2} \exp\Big\{N\Lambda(q) + \beta\Big(t - \frac{\log N}{2c_p}\Big)q^p\Big\}\,\mathrm{d}q. \tag{9.53}$$

By Stirling's formula, $\ell_N = (N/(2\pi))^{1/2}(1+O(1/N))$. Fix a closed interval $\mathcal{K} \subset (a,1)$ that contains $\sqrt{Q}$ in its interior. Since $\Lambda''(\sqrt{Q}) = -k_* < 0$ and $\Lambda(\sqrt{Q}) = F(\beta)$, there exist $c, C > 0$ with $-C(q-\sqrt{Q})^2 \le \Lambda(q) - F(\beta) \le -c(q-\sqrt{Q})^2$ on $\mathcal{K}$, after shrinking $\mathcal{K}$ if necessary.

Away from $\sqrt{Q}$. By Lemma 9.3, $\Lambda \le F(\beta) - c'$ on every compact subset of $[a,1)$ disjoint from the interior of $\mathcal{K}$, for some $c' > 0$ depending on the subset. To handle the endpoint $q = 1$, we return to (9.41). For $|t| \le L$ and large N, the exponent $\beta(M_N + t)q^p + NC_\beta(q)/2$ is at most $C_1 N$, with $C_1 = \beta e_p + \beta^2 + 1$, and on $[1-\epsilon, 1)$ we have $(1-q^2)^{(N-3)/2} \le (2\epsilon)^{(N-3)/2}$. Choosing $\epsilon > 0$ with $C_1 + \log(2\epsilon)/2 < F(\beta) - 1$, we see that the integral over $[1-\epsilon, 1)$ is at most $e^{N(F(\beta)-1/2)}$ for large N. The remaining part of $[a,1)$ outside $\mathcal{K}$ is compact, and there the integrand of (9.53) is at most $e^{N(F(\beta)-c')+C\log N}$. Both contributions are exponentially small compared with $e^{NF(\beta)}N^{-1/(2x)}$.

Near $\sqrt{Q}$. On $\mathcal{K}$, set $y = \sqrt{N}(q - \sqrt{Q})$, and divide the integrand of (9.53), multiplied by ℓ_N and by $\mathrm{d}q = N^{-1/2}\,\mathrm{d}y$, by $e^{NF(\beta)}N^{-\beta Q^{p/2}/(2c_p)}$. For bounded y, uniformly in $|t| \le L$, the result converges to

$$(2\pi)^{-1/2} s^{-3/2} e^{\beta t Q^{p/2}} e^{-k_* y^2/2}.$$

For domination, note that $|\beta(t - (\log N)/(2c_p))(q^p - Q^{p/2})| \le C \log N\, |y|/\sqrt{N} \le cy^2/2 + C'(\log N)^2/N$, and the rescaled integrand is at most $C'' e^{-cy^2/2}$ for large N. Dominated convergence gives the limits of the integrals over $\{|y| \le w\}$ and over all y with $q \in \mathcal{K}$, uniformly in $|t| \le L$. Using $\int e^{-k_* y^2/2}\,\mathrm{d}y = (2\pi/k_*)^{1/2}$ and the identity $\beta Q^{p/2}/(2c_p) = 1/(2x)$, which follows from (9.6), we obtain (9.50) with $t = 0$, and the first two limits in (9.51). The factor $N^{1/2}$ in ℓ_N cancels the width $N^{-1/2}$ of the peak at $\sqrt{Q}$. The third limit is the difference of the first two. The lower bound r_L exists because the limit in the second statement is at least $e^{-\beta L Q^{p/2}}\Gamma(1) > 0$.

Energy deficit. Replacing M_N by $M_N - v$ in (9.41) multiplies the integrand by $e^{-\beta v q^p} \le e^{-\beta v a^p}$ for $q \ge a$, which gives (9.52) at every N. □

Proposition 9.21 (A lower bound for the partition function)**.** *We have*

$$\lim_{\eta \downarrow 0} \limsup_{N \to \infty} \mathbb{P}\{Z_{N,\beta} < \eta \mathcal{A}_N\} = 0. \tag{9.54}$$

Proof. Fix $\varepsilon > 0$. By Lemma 9.8, there exists L_0 such that $\limsup_N \mathbb{P}\{|\max H_N - M_N| > L_0\} \le \varepsilon/2$. Let r_{L_0} be given by Proposition 9.20, and let $\eta \in (0, r_{L_0})$. For $u \in S_N$, let A be the set of pairs (H,u) with $Z_u(I_{N,1}) < \eta \mathcal{A}_N$. For $|h - M_N| \le L_0$ and large N, we have $\overline{Z}_{N,h}(I_{N,1}) \ge r_{L_0}\mathcal{A}_N$, so (9.46) with $T = \log(r_{L_0}/\eta)$ gives

$$\mathbb{P}_{u;h}(A) \le \mathbb{P}_{u;h}\big\{Z_u(I_{N,1}) < (\eta/r_{L_0})\overline{Z}_{N,h}(I_{N,1})\big\} \le \frac{C_-}{\log(r_{L_0}/\eta)}.$$

Using Lemma 9.9 with $J_N = [M_N - L_0, M_N + L_0]$ and $Y = \mathbf{1}_A$, we find

$$\mathbb{E}\#\big\{u \in \mathcal{D}_{N,L_0} : (H,u) \in A\big\} \le C\Big(\int_{-L_0}^{L_0} e^{-c_p t}\,\mathrm{d}t\Big)\Big(\frac{C_-}{\log(r_{L_0}/\eta)}\Big)^{1/2}. \tag{9.55}$$

We choose η so small that the right side is at most $\varepsilon/2$. By Markov's inequality, with probability at least $1-\varepsilon/2$, no $u \in \mathcal{D}_{N,L_0}$ has $(H,u) \in A$. If, in addition, $|\max H_N - M_N| \le L_0$, then a maximizer u_* of H_N belongs to $\mathcal{D}_{N,L_0}$, and $Z_{N,\beta} \ge Z_{u_*}(I_{N,1}) \ge \eta \mathcal{A}_N$. We conclude that $\limsup_N \mathbb{P}\{Z_{N,\beta} < \eta \mathcal{A}_N\} \le \varepsilon$, also for every smaller η. The parameters are chosen in the order ε, L_0, η, and then $N \to \infty$. The bound (9.55) counts every critical point in the window whose band has small partition function, and the maximizer is not selected in a way that depends on its Hessian or on its band. □

9.9 Concentration on microscopic bands

In this section, we prove Theorem 9.2. We call the bands $B_u(wN^{-1/2})$ with w fixed *microscopic bands*. We first replace the coarse cover of Proposition 9.16, which holds for every fixed energy deficit τ per spin, by a cover with a deficit τ_N that tends to zero.

Lemma 9.22 (A cover with vanishing energy deficit)**.** *There exists a deterministic sequence $\tau_N \to 0$ with $N\tau_N/(\log N)^2 \ge 1$ such that $\mathbb{E}\,\pi(\mathcal{U}_N(\tau_N)^c) \to 0$. Let $a_N = N\tau_N - (\log N)/(2c_p)$ and $J_N = [M_N - a_N, M_N + \sqrt{N}]$. Then $a_N > 0$ for large N, $a_N = o(N)$, and with probability tending to one, every critical point u with $H_N(u) \ge N(e_p - \tau_N)$ satisfies $H_N(u) \in J_N$.*

Proof. By Proposition 9.16, for each integer $j \ge 1$ there exists N_j such that $\mathbb{E}\,\pi(\mathcal{U}_N(1/j)^c) \le 1/j$ for $N \ge N_j$, and we may take $N_1 < N_2 < \cdots$. Let $\tau_N^0 = 1/j$ for $N_j \le N < N_{j+1}$, and let $\tau_N = \max(\tau_N^0, (\log N)^2/N)$. Then $\tau_N \to 0$. The set $\mathcal{U}_N(\tau)$ increases with τ, and we have $\mathbb{E}\,\pi(\mathcal{U}_N(\tau_N)^c) \le \mathbb{E}\,\pi(\mathcal{U}_N(\tau_N^0)^c) \to 0$. Further, $a_N \ge (\log N)^2 - (\log N)/(2c_p) > 0$ for large N, and $a_N \le N\tau_N = o(N)$. The lower end of J_N is $M_N - a_N = N(e_p - \tau_N)$, and $\mathbb{P}\{\max H_N > M_N + \sqrt{N}\} \to 0$ by Lemma 9.8, which proves the last assertion. □

No rate of convergence in Proposition 9.16 is needed: the sequence τ_N is chosen after the coarse cover, and it serves only to place the relevant critical points in a window of sublinear width, where Lemma 9.9 applies.

Proof of Theorem 9.2. We prove part (2) first, in the following form: for every $\varepsilon > 0$, there exist $L, w > 0$ such that

$$\limsup_{N\to\infty} \mathbb{E}\,\pi\Big(\Big[\bigcup_{u\in\mathcal{D}_{N,L}} B_u(wN^{-1/2})\Big]^c\Big) \le \varepsilon, \tag{9.56}$$

and the same holds for all larger L and w, since the union increases with both. Let τ_N, a_N, and J_N be as in Lemma 9.22. The proof has three steps: we bound the bands around critical points with a large energy deficit $M_N - H_N(u)$, we compare $\mathcal{U}_N(\tau_N)$ with the caps of the critical points in $\mathcal{D}_{N,L}$, and we pass from these caps to microscopic bands.

The argument has the structure of the proof of Proposition 4.9: a lower bound for the partition function from one band in a fixed window (Proposition 9.21), a coarse cover, and a summable bound on the bands around critical points with a large energy deficit. We give it directly because part (2) concerns bands of width $wN^{-1/2}$, which shrink as $N \to \infty$, while Proposition 4.9 concerns bands of a fixed width.

Critical points with a large energy deficit. For $L > 0$, let

$$T_{N,L} = \sum_{\substack{u:\nabla_{S_N} H_N(u)=0 \\ H_N(u)\in J_N,\ H_N(u)<M_N-L}} Z_u([a,1)).$$

Using Lemma 9.9 with $Y(H,u) = Z_u([a,1))$, (9.45), and (9.52), we find that, for all sufficiently large N,

$$\mathbb{E}T_{N,L} \le C\int_L^{\max(L,a_N)} e^{c_p v} C_2^{1/2}\, \overline{Z}_{N,M_N-v}([a,1))\,\mathrm{d}v \le C\mathcal{A}_N \int_L^\infty e^{-\delta_* v}\,\mathrm{d}v = \frac{C}{\delta_*}\mathcal{A}_N e^{-\delta_* L}, \tag{9.57}$$

where $\delta_* = \beta a^p - c_p > 0$ by (9.15), and an empty interval of integration gives zero. The constant C does not depend on L or N. This is the summable bound on the bands around critical points with a large energy deficit; it relies on the bounded prefactor in (9.25).

Caps of the fixed window. Let $\mathcal{V}_{N,L} = \bigcup_{u\in\mathcal{D}_{N,L}}\{x : R(x,u) \ge a\}$. Fix L, and let N be so large that $L < \sqrt{N}$. Suppose that $\max H_N \le M_N + L$, and that every critical point u with $H_N(u) \ge N(e_p - \tau_N)$ satisfies $H_N(u) \in J_N$. Then every $x \in \mathcal{U}_N(\tau_N) \setminus \mathcal{V}_{N,L}$ lies in the cap $\{R(\cdot,u) \ge a\}$ of a critical point u with $M_N - a_N \le H_N(u) < M_N - L$, which is counted in $T_{N,L}$. On this event and on $\{Z_{N,\beta} \ge \eta\mathcal{A}_N\}$, we have $\pi(\mathcal{V}_{N,L}^c) \le \pi(\mathcal{U}_N(\tau_N)^c) + T_{N,L}/(\eta\mathcal{A}_N)$; a point covered by several caps only increases the last term. Bounding $\pi(\mathcal{V}_{N,L}^c)$ by one off this event, and using Lemma 9.22 and (9.57), we obtain, for every $\eta > 0$,

$$\limsup_{N\to\infty} \mathbb{E}\,\pi(\mathcal{V}_{N,L}^c) \le \limsup_{N\to\infty}\Big[\mathbb{P}\{\max H_N > M_N + L\} + \mathbb{P}\{Z_{N,\beta} < \eta\mathcal{A}_N\}\Big] + \frac{C}{\eta\delta_*}e^{-\delta_* L}. \tag{9.58}$$

Microscopic bands. Fix L and w. We have $\sup_{|h-M_N|\le L}\overline{Z}_{N,h}([a,1)\setminus I_{N,w}) \le \mathcal{A}_N[e^{\beta L Q^{p/2}}\overline{\Gamma}(w) + o(1)]$ by (9.51). Using Lemma 9.9 with $Y(H,u) = Z_u([a,1)\setminus I_{N,w})$ and (9.45), we find

$$\mathbb{E}\sum_{u\in\mathcal{D}_{N,L}} Z_u\big([a,1)\setminus I_{N,w}\big) \le C_L\mathcal{A}_N\big[\overline{\Gamma}(w) + o(1)\big], \tag{9.59}$$

where C_L depends only on p, β, a and L. Here the latitude set has two components; this is allowed since Lemma 9.9 applies to Borel weights. For large N, we have $I_{N,w} \subset [a,1)$ and $B_u(wN^{-1/2}) \subset \{R(\cdot,u) \ge a\}$. The complement of $\bigcup_{u\in\mathcal{D}_{N,L}} B_u(wN^{-1/2})$ is contained

in $\mathcal{V}_{N,L}^c$ together with the sets $\{x : R(x,u) \in [a,1] \setminus I_{N,w}\}$, $u \in \mathcal{D}_{N,L}$; the point $R(x,u) = 1$ has measure zero. On $\{Z_{N,\beta} \geq \eta \mathcal{A}_N\}$, the Gibbs measure of the latter sets is at most the sum in (9.59) divided by $\eta \mathcal{A}_N$. Combining this with (9.58), we obtain

$$\limsup_{N\to\infty} \mathbb{E}\, \pi\Big(\Big[\bigcup_{u\in\mathcal{D}_{N,L}} B_u(wN^{-1/2}) \Big]^c \Big) \\ \leq \limsup_{N\to\infty} \Big[\mathbb{P}\{\max H_N > M_N + L\} + 2\mathbb{P}\{Z_{N,\beta} < \eta \mathcal{A}_N\} \Big] + \frac{C}{\eta\delta_*} e^{-\delta_* L} + \frac{C_L}{\eta} \overline{\Gamma}(w). \tag{9.60}$$

Order of the parameters. Given $\varepsilon > 0$, we first use Proposition 9.21 to fix $\eta > 0$ with $\limsup_N \mathbb{P}\{Z_{N,\beta} < \eta \mathcal{A}_N\} \leq \varepsilon/5$. Next, by Lemma 9.8, we fix L so large that $\limsup_N \mathbb{P}\{\max H_N > M_N + L\} \leq \varepsilon/5$ and $Ce^{-\delta_* L}/(\eta\delta_*) \leq \varepsilon/5$. Finally, we fix w so large that $C_L \overline{\Gamma}(w)/\eta \leq \varepsilon/5$. With these choices, (9.60) implies (9.56). Since the union in (9.56) increases with L and w, (9.56) also holds for all larger L and w.

Conclusion. Let L, w be as in (9.56), and let $M_0 = Q^{p/2}L$ and $w_* = w$. For $M \geq M_0$ and $w' \geq w_*$, we have $\mathcal{C}_{N,M} \supset \mathcal{C}_{N,M_0} = \mathcal{D}_{N,L}$ for all sufficiently large N, by (9.8), and $B_u(w'N^{-1/2}) \supset B_u(wN^{-1/2})$. Then (9.56) implies (9.12), which is part (2). For part (1), let $b_0 = \min(\sqrt{Q} - a, 1 - \sqrt{Q})/2$, and let $b \in (0, b_0]$ be fixed. For every fixed w, we have $wN^{-1/2} \leq b$ and $B_{N,M}(b) \supset B_{N,M}(wN^{-1/2})$ for large N. By part (2), for every $\varepsilon > 0$ there exists M_0 with $\limsup_N \mathbb{E}\, \pi(B_{N,M}(b)^c) \leq \varepsilon$ for all $M \geq M_0$, which is (9.11). With this choice of b_0, every band $B_u(b)$ with $b \leq b_0$ is contained in the cap $\{R(\cdot,u) \geq a\}$ and does not contain u. □

Part (1) holds for every band width in $(0, b_0]$, and the width and the interval U do not depend on the window M. This is the order required in Section 13: there, the interval U, the band width, and the time exponent are fixed before the window. In the form (9.56), the concentration is stated in the following sense: for every tolerance, there exist finite L and w, fixed during the limit $N \to \infty$. We do not claim that, for a single pair (L, w), the Gibbs mass outside the bands tends to zero with probability tending to one.

For even p, the two points u and $-u$ have the same energy and radial derivative. If critical points are counted once per antipodal pair, with the bands $\{x : ||R(x,u)| - \sqrt{Q}| \leq b\}$, the union of bands is unchanged and the counts in Proposition 9.1 are halved.

Remark 9.23 (The strict inequality $\beta > \beta_s(p)$). The bound on the bands around critical points with a large energy deficit is not uniform as $\beta \downarrow \beta_s(p)$. As $\beta \downarrow \beta_s(p)$, we have $x \to 1$, and the margin $\beta Q^{p/2} - c_p = c_p(1/x - 1)$ tends to zero. The exponent $\delta_* = \beta a^p - c_p$ in the bound (9.57) is smaller than this margin, so that bound degenerates at the static threshold.

SECTION 10

Mixed models: one-point estimates and exclusion of high-energy points at positive overlap

In this section, we study the critical points of the center field H_Q near its ground state, one point at a time. There are three main results. Theorem 10.6 identifies the centering $m_N = NE_0 - \log N/(2c_0) + O(1)$ and the limits of the expected numbers of critical points in fixed energy windows around it. The localization, Proposition 10.10, shows that with probability exponentially close to one every critical point of H_Q above energy $N(E_0 - \delta)$ is a local maximum with radial derivative close to K_0, and its Hessian is bounded above by a fixed negative multiple of the identity. Theorem 10.13 shows that, conditionally on a critical point at u whose energy and radial derivative are close to the ground-state parameters (E_0, K_0), no other critical point of comparable energy has overlap with u in $[\varepsilon, 1]$, for a prescribed $\varepsilon > 0$. We refer to this property as the exclusion of high-energy points at positive overlap, and to the event on which it holds as the selection event at u. Proposition 10.15 shows that restricting to the selection event removes an exponentially small fraction of the expected number of local maxima. Proposition 10.16 records the same statements for pure models, and Corollary 10.17 collects what the later sections use.

From Section 5, we use the 1RSB assumptions (Definition 5.9) and the identities of Proposition 5.11 that relate the 1RSB parameters x, Q to the ground-state energy E_0 and the radial derivative K_0. From Section 7, we use the Gaussian structure of the field at a point, the moments of the Hessian determinant (Proposition 7.7), and the marked Kac–Rice formula with its exact count integrals (Theorem 7.17 and Corollary 7.18). The only other outside input is the one-step upper bound of Huang and Sellke for the ground state energy, stated below as Theorem 10.11. Corollary 10.17 is used in Section 11, where the second moment of the number of local maxima on the selection event is bounded, and in Section 12, which uses Proposition 10.10 and the integral over the radial derivative in Lemma 10.5 to show that the bands carry almost all of the Gibbs mass.

The expected number of critical points of a spherical spin glass was computed rigorously with the Kac–Rice formula by Auffinger, Ben Arous, and Černý for pure models [12] and by Auffinger and Ben Arous for mixed models [8]. Ben Arous, Subag, and Zeitouni computed the complexity with a prescribed radial derivative [30, Theorem 5]. For pure models, Subag and Zeitouni identified the logarithmic correction to the centering and the extremal process of critical points [144]. The exclusion estimate of Section 10.6 follows

the route of Huang and Sellke [90, Section 3.3], who bounded a conditional ground-state energy by explicit one-step trials to show that typical critical points near the ground state have no high-energy points at intermediate overlap.

Counts in this section are ordinary counts of points of S_N. For even mixtures, both points of an antipodal pair are counted. No parity assumption is made, and no limit theorem for the extremal process is used or proved.

Standing assumptions. Throughout this section, ξ is a finite mixture with minimum degree $d \geq 4$, $\beta > 0$, and (ξ, β) is strictly 1RSB with parameters $0 < x, Q < 1$ and a nondegenerate contact at Q (Definition 5.9). We write $s = 1 - Q$ and V for the first-variation function of Definition 5.9, so that $V(0) = V(Q) = V'(Q) = 0$, $V''(Q) > 0$, and $V(r) > 0$ for $r \in (0,1) \setminus \{Q\}$. The mixture, β, x, and Q are fixed before N. Recall from Definition 3.1 that ξ is *non-pure* if at least two coefficients c_p are positive, and *pure* otherwise. Until Section 10.7, ξ is a non-pure mixture. All radial intervals U and the base interval U_0 in this section are bounded open intervals, so that their closures are compact.

Notation. As in Section 7, we set

$$\xi_Q(r) = \xi(Qr) = \sum_p c_p Q^p r^p, \tag{10.1}$$

so that $\mathbb{E} H_Q(u) H_Q(v) = N \xi_Q(R(u,v))$ for $u, v \in S_N$, and $\xi(Q) = \xi_Q(1)$, $a_Q = \xi_Q'(1)$, $b_Q = \xi_Q''(1)$. We extend H_Q to $\mathbb{R}^N$ as a polynomial, so that $K_Q(u) = u \cdot \nabla H_Q(u)$ with the Euclidean gradient. Note that $K_Q(u)$ is the derivative of $t \mapsto H_N(t\sqrt{Q}\, u)$ at $t = 1$; it is of order N, and it is not the derivative in a unit radial direction. For a point u, we use the unnormalized energy $h = H_Q(u)$, the normalized energy $e = h/N$, and the normalized radial derivative $k = K_Q(u)/N$, and we call (e,k) the normalized marks of u. In Section 7, the letter h denotes the normalized energy; it corresponds to e here. We keep h for the unnormalized energy because the fixed windows $[m_N + a, m_N + b]$ and the tail T_N below are in unnormalized units. We write

$$q(e,k) = (e,k) \Sigma_Q^{-1} (e,k)^{\mathsf{T}}, \qquad \ell(e,k) = \frac{1}{2} \partial_e q(e,k) = \frac{(a_Q + b_Q)e - a_Q k}{\det \Sigma_Q}. \tag{10.2}$$

Let $\mathrm{Crt}_N = \{u \in S_N : \nabla_{S_N} H_Q(u) = 0\}$ be the set of critical points of H_Q, and $\mathrm{Crt}_N^{\max}$ the set of $u \in \mathrm{Crt}_N$ with $\nabla^2_{S_N} H_Q(u) \prec 0$, the nondegenerate local maxima. Let $n = N - 1$. We write G_n for an $n \times n$ GOE matrix with off-diagonal entries of variance $1/n$ and diagonal entries of variance $2/n$, and set

$$b_N = \Big(\frac{b_Q n}{N}\Big)^{1/2}, \qquad W_{n,N} = b_N G_n, \qquad D_N(k) = \big|\det(kI_n - W_{n,N})\big|, \tag{10.3}$$

$$d_N^{\mathrm{all}}(k) = \mathbb{E} D_N(k), \qquad d_N^{\max}(k) = \mathbb{E}\big[D_N(k) \mathbf{1}\{\lambda_{\max}(W_{n,N}) < k\}\big]. \tag{10.4}$$

Let $\sigma_{\mathrm{sc}}(\,\mathrm{d}t) = (2\pi)^{-1}(4-t^2)^{1/2}\mathbf{1}_{[-2,2]}(t)\,\mathrm{d}t$ be the semicircle law and $\Omega(t) = \int \log|t-s|\,\sigma_{\mathrm{sc}}(\,\mathrm{d}s)$ its logarithmic potential. Finally, $|S_N| = 2\pi^{N/2}N^{(N-1)/2}/\Gamma(N/2)$ is the surface area of S_N.

Throughout, c and C denote positive constants that depend only on ξ, β, and the fixed intervals and tolerances named in the statement. They never depend on N, and they may change from line to line. A statement holds *for all large* N if it holds for $N \ge N_0$, where N_0 has the same dependence.

10.1 The inputs

Identities from Section 5. We set

$$z = \frac{xQ}{s}, \qquad y = \frac{Q}{\beta s}. \tag{10.5}$$

We use the following consequences of the 1RSB assumptions, proved in Proposition 5.11 by direct integration from $V(Q) = V'(Q) = 0$ and $V''(Q) > 0$. First,

$$y^2 = (1+z)a_Q, \qquad \frac{\xi(Q)}{a_Q} = \frac{(1+z)\log(1+z) - z}{z^2}, \qquad b_Q < y^2, \tag{10.6}$$

and the ground-state parameters are

$$E_0 = \frac{a_Q + z\xi(Q)}{y}, \qquad K_0 = y + \frac{b_Q}{y}, \qquad c_0 = \frac{z}{y} = \beta x < \beta. \tag{10.7}$$

Since $y > b_Q^{1/2}$, the value K_0 lies strictly above the edge of the spectrum: $K_0 - 2b_Q^{1/2} = (y - b_Q^{1/2})^2/y > 0$. Second, the function

$$g(r) = \xi(Q) - \xi_Q(r) - a_Q\Big[\frac{1+z}{z^2}\log\big(1+(1-r)z\big) - \frac{1-r}{z}\Big], \qquad 0 \le r \le 1, \tag{10.8}$$

equals $\beta^{-2}V(Qr)$. In particular, $g(0) = g(1) = 0$ and $g(r) > 0$ for $0 < r < 1$, since $V > 0$ on $(0, Q)$. In the language of [90], this says that the ground state of the covariance ξ_Q is strictly one-step replica symmetry breaking: the inequality $g(r) > 0$ on $(0,1)$ is (3.4) of [90, Lemma 3.5] for ξ_Q, and the ground energy $(\xi_Q'(1) + z\xi_Q(1))/y$ of that lemma is E_0. Third, multiplying out and using $y^2 = (1+z)a_Q$, we find (see also Proposition 5.13)

$$\Sigma_Q\begin{pmatrix} c_0 \\ 1/y \end{pmatrix} = \begin{pmatrix} E_0 \\ K_0 \end{pmatrix}, \qquad \text{so that} \qquad \ell(E_0, K_0) = c_0. \tag{10.9}$$

Results from Section 7. We use the notation of Section 7. The field is realized through its Gaussian coefficient vector Z as in (7.1), and every value and derivative of H_Q at a point is a linear function of Z. Here $W_{n,N}$ is the matrix of Section 7 and $d_N^{\mathrm{all}}(k) = \mu_{n,N}(k)$ in the notation there. We use the following facts.

(1) *One-point law* (Lemma 7.1). At $u \in S_N$, the pair $(H_Q(u), K_Q(u))$, the gradient $\nabla_{S_N} H_Q(u)$, and the compression $\nabla^2_\perp H_Q(u)$ of the Euclidean Hessian to $u^\perp$ are independent. The first two have covariances $N\Sigma_Q$ and $a_Q I_n$, the third has the law of $W_{n,N}$, and $\nabla^2_{S_N} H_Q(u) = \nabla^2_\perp H_Q(u) - (K_Q(u)/N) I_n$. For a non-pure mixture, Σ_Q is positive definite.

(2) *Conditioned laws* (Definition 7.3 and Lemma 7.4). For $u \in S_N$ and $(e,k) \in \mathbb{R}^2$, we write $\mathbb{P}_{e,k,u}$ and $\mathbb{E}_{e,k,u}$ for the conditioned law of Definition 7.3 at u with normalized marks (e,k); it is the law denoted $\mathbb{P}_{h,k,u}$ there, with $h = e$. Under $\mathbb{P}_{e,k,u}$, the coefficient vector is the explicit Gaussian regression (7.4), so that $(H_Q(u), K_Q(u), \nabla_{S_N} H_Q(u)) = (Ne, Nk, 0)$ surely. This law is defined for every value of the marks. Its random part is the residual $(\mathrm{Id} - \Lambda_u^* \mathcal{C}_u^{-1} \Lambda_u) Z$, a centered Gaussian vector whose covariance is an orthogonal projection and whose law does not depend on (e,k). The Riemannian Hessian $\nabla^2_{S_N} H_Q(u)$ has the law of $W_{n,N} - kI_n$. In a pure model, $K_Q = pH_Q$, and we write $\mathbb{P}_{e,u}$ and $\mathbb{E}_{e,u}$. The Hessian then has the law of $W_{n,N} - peI_n$. We write $D(u) = |\det \nabla^2_{S_N} H_Q(u)|$, so that

$$\mathbb{E}_{e,k,u} D(u) = d_N^{\mathrm{all}}(k), \qquad \mathbb{E}_{e,k,u}\big[D(u)\mathbf{1}\{\nabla^2_{S_N} H_Q(u) \prec 0\}\big] = d_N^{\max}(k). \tag{10.10}$$

(3) *Gaussian estimates* (Lemma 7.5). If f is an L-Lipschitz function of a standard Gaussian vector Y, then (7.6) holds, and applying its tail bound to $\pm f$, we have

$$\mathbb{P}\big(|f(Y) - \mathbb{E} f(Y)| > a\big) \le 2e^{-a^2/(2L^2)}, \qquad a > 0. \tag{10.11}$$

By (7.2), $W_{n,N}$ satisfies the hypothesis of Lemma 7.5(2) with $b = b_Q$. Then (7.7) reads

$$\mathbb{P}\big(\|W_{n,N}\|_{\mathrm{op}} > 2b_Q^{1/2} + t\big) \le 2e^{-Nt^2/(4b_Q)}, \qquad t > 0. \tag{10.12}$$

(4) *Determinant moments* (Proposition 7.7 and Corollary 7.8). For k in a compact subset of $(2b_Q^{1/2}, \infty)$, every fixed moment of $D(u)/d_N^{\mathrm{all}}(k)$ under $\mathbb{P}_{e,k,u}$ is bounded uniformly in N, u, and e. By the Cauchy–Schwarz inequality, for every event $\mathcal{A}$, we have $\mathbb{E}_{e,k,u}[D(u)\mathbf{1}_{\mathcal{A}}] \le \mathsf{K}_2^{1/2} \mathbb{P}_{e,k,u}(\mathcal{A})^{1/2} d_N^{\mathrm{all}}(k)$, with no assumption on the dependence between $\mathcal{A}$ and the determinant.

(5) *Kac–Rice formula* (Theorem 7.17 and Corollary 7.18). The restriction of H_Q to S_N is almost surely a Morse function with finitely many critical points, and (7.24) and (7.25) hold for nonnegative or bounded field-dependent weights.

The one-point law agrees with [30, Section 4.1, Eq. (4.1)], after a change of radius and normalization (Remark 7.2).

10.2 The one-point intensity

Lemma 10.1 (The marked one-point intensity)**.** *Let ξ be a non-pure mixture. Then Σ_Q is positive definite. Let $f(H,u) \geq 0$ be a Borel function of the field $H = H_N$ and the point $u \in S_N$, and let $A \subset \mathbb{R}^2$ be a Borel set. Then*

$$\begin{aligned} &\mathbb{E} \sum_{\substack{u \in \mathrm{Crt}_N \\ (H_Q(u), K_Q(u)/N) \in A}} f(H,u) \\ &\qquad = \int_A \mathfrak{c}_N e^{-Nq(h/N,k)/2} \int_{S_N} \mathbb{E}_{h/N,k,u}\big[D(u) f(H,u)\big] \nu_N(\,\mathrm{d}u)\,\mathrm{d}h\,\mathrm{d}k, \end{aligned} \tag{10.13}$$

where

$$\mathfrak{c}_N = \frac{|S_N|(2\pi a_Q)^{-n/2}}{2\pi(\det \Sigma_Q)^{1/2}}. \tag{10.14}$$

In particular, the expected numbers of points of Crt_N and of $\mathrm{Crt}_N^{\max}$ with $(H_Q(u), K_Q(u)/N) \in A$ are $\int_A \lambda_N^{\mathrm{all}}$ and $\int_A \lambda_N^{\max}$, where

$$\lambda_N^j(h,k) = \mathfrak{c}_N e^{-Nq(h/N,k)/2} d_N^j(k), \qquad j \in \{\mathrm{all}, \max\}. \tag{10.15}$$

Proof. The matrix Σ_Q is positive definite by Lemma 7.1. Identity (10.13) is (7.24) with $\mathcal{K} = \{(h/N,k) : (h,k) \in A\}$, written in the unnormalized energy $h = Ne$. In (7.24), the density p_N of the normalized marks carries a factor N, which is absorbed by $\mathrm{d}h = N\,\mathrm{d}e$, and this accounts for the constant (10.14). For (10.15), we take $f = 1$ and $f = \mathbf{1}\{\nabla^2_{S_N} H_Q(u) \prec 0\}$. By (10.10), the inner expectation then depends neither on u nor on h, and the integral over S_N equals one. □

For $f = 1$, (10.15) is the exact count integral of Corollary 7.18, and without the radial derivative it reduces to the formula of Auffinger and Ben Arous for the center field [8, Proposition 3].

Lemma 10.2 (Determinant estimates)**.** *Let $\Lambda(k) = \frac{1}{2}\log b_Q + \Omega(k/b_Q^{1/2})$.*

(a) Let $J \subset (2b_Q^{1/2}, \infty)$ be a compact interval. There exist $c, C > 0$ such that, for all $k \in J$ and all large N,

$$c\, e^{N\Lambda(k)} \leq d_N^{\max}(k) \leq d_N^{\mathrm{all}}(k) \leq C e^{N\Lambda(k)}, \qquad d_N^{\mathrm{all}}(k) \leq \big(1 + Ce^{-cN}\big) d_N^{\max}(k). \tag{10.16}$$

(b) Let $J' \subset \mathbb{R}$ be a compact interval and $a > 0$. There exist C_a and N_a such that $d_N^{\mathrm{all}}(k) \leq C_a e^{N(\Lambda(k)+a)}$ for all $k \in J'$ and $N \geq N_a$.

(c) There exists C such that $d_N^{\mathrm{all}}(k) \leq (|k| + C)^n$ for all $k \in \mathbb{R}$ and $N \geq 2$.

Part (a) concerns shifts outside the spectrum and follows from Section 7.4. Parts (b) and (c) are weaker upper bounds, which we need for shifts inside or near the spectrum and for large shifts.

Proof. We begin with (a). Using Corollary 7.10 with $b = b_Q$ and $m = n$, we have $c\,e^{\Lambda_{n,N}(k)} \le d_N^{\mathrm{all}}(k) \le Ce^{\Lambda_{n,N}(k)}$ uniformly for $k \in J$, and $\Lambda_{n,N}(k)$ differs by a bounded amount from $\Lambda_{N,N}(k) = N\Lambda(k)$. This proves the outer bounds in (10.16). For the comparison, let $\gamma = \min J - 2b_Q^{1/2}$. The indicator in (10.4) equals one on the event $\{\|W_{n,N}\|_{\mathrm{op}} \le 2b_Q^{1/2} + \gamma/2\}$, whose complement has probability at most $2e^{-N\gamma^2/(16b_Q)}$ by (10.12). By fact (4) above (Corollary 7.8 with $r = 2$), we have

$$d_N^{\mathrm{all}}(k) - d_N^{\max}(k) \le \mathsf{K}_2^{1/2}\big(2e^{-N\gamma^2/(16b_Q)}\big)^{1/2} d_N^{\mathrm{all}}(k) \le Ce^{-cN} d_N^{\mathrm{all}}(k),$$

and the second inequality in (10.16) follows for all large N.

Next, we prove (b). Let $\tau > 0$ and $\varphi_{t,\tau}(s) = \frac{1}{2}\log((t-s)^2 + \tau^2)$. Then $\varphi_{t,\tau}(s) \ge \log|t-s|$, and $|\partial_s \varphi_{t,\tau}| \le 1/(2\tau)$. For symmetric matrices, the Hoffman–Wielandt inequality [6, Lemma 2.1.19], $\sum_i |\lambda_i(X) - \lambda_i(Y)|^2 \le \|X - Y\|_F^2$ and the Cauchy–Schwarz inequality show that $X \mapsto \operatorname{Tr}\varphi_{t,\tau}(X)$ is $n^{1/2}/(2\tau)$-Lipschitz in the Frobenius norm. Note that the map from the independent standard Gaussian coordinates to G_n multiplies Frobenius distances by at most $(2/n)^{1/2}$. Then $\operatorname{Tr}\varphi_{t,\tau}(G_n)$ is a $(2^{1/2}\tau)^{-1}$-Lipschitz function of a standard Gaussian vector, and by (7.6) with $t = 1$ we have

$$\mathbb{E}|\det(t - G_n)| \le \mathbb{E}e^{\operatorname{Tr}\varphi_{t,\tau}(G_n)} \le \exp\big(\mathbb{E}\operatorname{Tr}\varphi_{t,\tau}(G_n) + (4\tau^2)^{-1}\big). \tag{10.17}$$

By Wigner's semicircle law [6, Theorem 2.1.1], the expected empirical spectral measure $\bar{\mu}_n = \mathbb{E}n^{-1}\sum_i \delta_{\lambda_i(G_n)}$ converges weakly to σ_{sc}. (For the GOE, this also follows from the moment method, since $n^{-1}\mathbb{E}\operatorname{Tr}G_n^{2l}$ converges to the Catalan number Cat_l and the odd moments vanish.) We have $\log\tau \le \varphi_{t,\tau}(s) \le C_t(1 + |s|)$ and $\int s^2\bar{\mu}_n(\mathrm{d}s) = 1 + n^{-1}$, which implies that the functions $\varphi_{t,\tau}$ are uniformly integrable with respect to the measures $\bar{\mu}_n$. Then $n^{-1}\mathbb{E}\operatorname{Tr}\varphi_{t,\tau}(G_n) \to \int \varphi_{t,\tau}\,\mathrm{d}\sigma_{\mathrm{sc}}$ for each t. This convergence is uniform for t in compact sets, because $t \mapsto \varphi_{t,\tau}$ is $(2\tau)^{-1}$-Lipschitz. Since the density of σ_{sc} is at most $1/\pi$ and $\int_{\mathbb{R}} \log(1 + \tau^2/s^2)\,\mathrm{d}s = 2\pi\tau$, we have

$$0 \le \int \varphi_{t,\tau}\,\mathrm{d}\sigma_{\mathrm{sc}} - \Omega(t) = \frac{1}{2}\int \log\Big(1 + \frac{\tau^2}{(t-s)^2}\Big)\sigma_{\mathrm{sc}}(\mathrm{d}s) \le \tau.$$

We now fix $\tau = a/4$, apply (10.17) with $t = k/b_N$, and multiply by b_N^n. For N large enough depending on a and J', we find $d_N^{\mathrm{all}}(k) \le b_N^n \exp(n\Omega(k/b_N) + na/2 + 4a^{-2})$. Finally, $n\log b_N = \frac{1}{2}N\log b_Q + O(1)$, and $|n\Omega(k/b_N) - N\Omega(k/b_Q^{1/2})| = O(1)$ uniformly for $k \in J'$, because Ω is bounded and Lipschitz on compact sets and $|k/b_N - k/b_Q^{1/2}| = O(N^{-1})$. This proves (b). We note that (b) is an upper bound with an arbitrarily small exponential error, not a bounded prefactor.

Finally, we prove (c). We have $D_N(k) \le (|k| + b_N\|G_n\|_{\mathrm{op}})^n$. By Minkowski's inequality, $d_N^{\mathrm{all}}(k)^{1/n} \le |k| + b_N(\mathbb{E}\|G_n\|_{\mathrm{op}}^n)^{1/n}$. We write $\|G_n\|_{\mathrm{op}} \le 2 + X$ with $X = (\|G_n\|_{\mathrm{op}} - 2)_+$. Applying Lemma 7.5(2) to G_n, we have $\mathbb{P}(X > t) \le 2e^{-nt^2/4}$, and then

$$\mathbb{E}X^n = \int_0^\infty nt^{n-1}\mathbb{P}(X > t)\,\mathrm{d}t \le n\Big(\frac{4}{n}\Big)^{n/2}\Gamma\Big(\frac{n}{2}\Big) \le C^n.$$

Using Minkowski's inequality again, we find $(\mathbb{E}\|G_n\|_{\mathrm{op}}^n)^{1/n} \le 2 + C$, which proves (c). □

10.3 The marked complexity

We define the *marked complexity*

$$\Theta(e,k) = \frac{1}{2} + \frac{1}{2}\log\frac{b_Q}{a_Q} - \frac{1}{2}q(e,k) + \Omega\big(k/b_Q^{1/2}\big), \qquad (e,k) \in \mathbb{R}^2. \tag{10.18}$$

For $k \ge 2b_Q^{1/2}$, this is the function Θ of Proposition 5.13, and for $k > 2b_Q^{1/2}$ it is the complexity of critical points of H_Q with normalized energy e and normalized radial derivative k computed in [30, Theorem 5]. We use it on all of $\mathbb{R}^2$, with Ω defined on the whole line; Lemma 10.4 below shows that $\lambda_N^{\mathrm{all}}(Ne,k)$ and $\lambda_N^{\max}(Ne,k)$ are comparable to $e^{N\Theta(e,k)}$, with bounded constants, for k in a fixed interval above $2b_Q^{1/2}$.

Lemma 10.3 (Global strong concavity)**.** *Let ξ be a non-pure mixture. The matrix*

$$M = \Sigma_Q^{-1} - \begin{pmatrix} 0 & 0 \\ 0 & (2b_Q)^{-1} \end{pmatrix} \tag{10.19}$$

is positive definite. Let $\kappa_\Theta > 0$ be its smallest eigenvalue. Then $\Theta(e,k) + \kappa_\Theta(e^2+k^2)/2$ is concave on $\mathbb{R}^2$, and

$$\Theta(E_0,K_0) = 0, \qquad \nabla\Theta(E_0,K_0) = (-c_0, 0). \tag{10.20}$$

Further, for all $(e,k) \in \mathbb{R}^2$,

$$\Theta(e,k) \le -c_0(e-E_0) - \frac{\kappa_\Theta}{2}\big[(e-E_0)^2 + (k-K_0)^2\big]. \tag{10.21}$$

In particular, K_0 is the unique maximizer of $\Theta(E_0,\cdot)$ on $\mathbb{R}$.

Proof. We recall the classical formulas for the logarithmic potential of the semicircle law:

$$\Omega(t) = \begin{cases} \dfrac{t^2}{4} - \dfrac{1}{2}, & |t| \le 2, \\ \dfrac{t^2}{4} - \dfrac{1}{2} - \dfrac{|t|(t^2-4)^{1/2}}{4} + \log\dfrac{|t| + (t^2-4)^{1/2}}{2}, & |t| > 2, \end{cases} \tag{10.22}$$

and $\Omega'(t) = t/2$ for $|t| \le 2$, $\Omega'(t) = (t - \mathrm{sgn}(t)(t^2-4)^{1/2})/2$ for $|t| > 2$. For $|t| > 2$, $\Omega'(t) = \int \sigma_{\mathrm{sc}}(\,\mathrm{d}s)/(t-s)$ is the Stieltjes transform of σ_{sc}. The function $\Omega'(t) - t/2$ is continuous, vanishes on $[-2,2]$, and is nonincreasing. This implies that $\Omega(t) - t^2/4$ is concave on $\mathbb{R}$.

We write

$$\Theta(e,k) = \frac{1}{2} + \frac{1}{2}\log\frac{b_Q}{a_Q} + \Big[\Omega\big(k/b_Q^{1/2}\big) - \frac{k^2}{4b_Q}\Big] - \frac{1}{2}(e,k)M(e,k)^{\mathsf{T}}.$$

The bracket is concave by the previous paragraph, and it suffices to show that M is positive definite. With $D = \det \Sigma_Q$, we have $M_{ee} = (a_Q + b_Q)/D > 0$, and the Schur complement of this entry is

$$M_{kk} - \frac{M_{ek}^2}{M_{ee}} = \frac{\xi(Q)}{D} - \frac{1}{2b_Q} - \frac{a_Q^2}{(a_Q + b_Q)D} = \frac{1}{a_Q + b_Q} - \frac{1}{2b_Q},$$

where we used $\xi(Q)(a_Q + b_Q) - a_Q^2 = D$. This is positive because $b_Q > a_Q$: indeed, $b_Q/a_Q = \sum_p (p-1)pc_pQ^p / \sum_p pc_pQ^p \geq d - 1 \geq 3$. This proves the concavity.

The identities (10.20) are Proposition 5.13(3). We recall the computation behind them, since it is used again below. With $w = y/b_Q^{1/2} > 1$, we have $K_0/b_Q^{1/2} = w + w^{-1}$ and $(K_0^2/b_Q - 4)^{1/2} = w - w^{-1}$, and (10.22) yields $\Omega(K_0/b_Q^{1/2}) = b_Q/(2y^2) + \log(y/b_Q^{1/2})$ and $b_Q^{-1/2}\Omega'(K_0/b_Q^{1/2}) = 1/y$. By (10.9), $q(E_0, K_0) = c_0E_0 + K_0/y$, and inserting these values into (10.18), together with the two identities in (10.6), gives $\Theta(E_0, K_0) = 0$. The gradient of Θ is $-\Sigma_Q^{-1}(e,k)^{\mathsf{T}} + (0, b_Q^{-1/2}\Omega'(k/b_Q^{1/2}))$, which equals $-(c_0, 1/y) + (0, 1/y) = (-c_0, 0)$ at (E_0, K_0).

Finally, $\phi(e,k) = \Theta(e,k) + \kappa_\Theta(e^2 + k^2)/2$ is concave and differentiable at (E_0, K_0), and it lies below its tangent plane there. We have $\nabla\phi(E_0, K_0) = (-c_0 + \kappa_\Theta E_0, \kappa_\Theta K_0)$, and the tangent-plane inequality is (10.21). □

The identities (10.20) are also the zero-complexity and radial stationarity identities of [90, Lemma 3.18 and Lemma 5.2]. The point of Lemma 10.3 is the global bound (10.21), which also covers values of the radial derivative inside the semicircle support.

Lemma 10.4 (Intensity bounds and excluded marks)**.** *Let ξ be a non-pure mixture, and let U be a bounded open interval with $K_0 \in U$ and $\overline{U} \subset (2b_Q^{1/2}, \infty)$. Let $u_U = \operatorname{dist}(K_0, \mathbb{R} \setminus U)$.*

(a) There exist $c, C > 0$ such that, for all $e \in \mathbb{R}$, $k \in \overline{U}$, and all large N,

$$c\, e^{N\Theta(e,k)} \leq \lambda_N^{\max}(Ne, k) \leq \lambda_N^{\mathrm{all}}(Ne, k) \leq Ce^{N\Theta(e,k)}. \tag{10.23}$$

(b) Let $0 < \delta \leq \kappa_\Theta u_U^2/(8c_0)$. There exist $c, C > 0$ such that

$$\mathbb{E}\#\big\{u \in \mathrm{Crt}_N : H_Q(u) \geq N(E_0 - \delta),\ K_Q(u)/N \notin U\big\} \leq Ce^{-cN}. \tag{10.24}$$

(c) For every $t > 0$, there exist $c_t, C_t > 0$ such that

$$\mathbb{E}\#\big\{u \in \mathrm{Crt}_N : H_Q(u) \geq N(E_0 + t)\big\} \leq C_t e^{-c_t N}. \tag{10.25}$$

(d) There exists C such that, for all $\delta' \geq 0$ and all large N,

$$\mathbb{E}\#\big\{u \in \mathrm{Crt}_N : H_Q(u) \geq N(E_0 - \delta'),\ K_Q(u)/N \in U\big\} \leq CNe^{c_0\delta' N}. \tag{10.26}$$

Proof. By Stirling's formula, we have

$$|S_N|(2\pi a_Q)^{-n/2} = (2a_Q)^{1/2}\big(1 + O(N^{-1})\big)e^{N(1-\log a_Q)/2}. \tag{10.27}$$

Combining this with (10.15) and (10.16) for $J = \overline{U}$ gives (a). The constants do not depend on e, because the Gaussian factor in (10.15) is exact.

For (b), let $e \geq E_0 - \delta$ and $k \notin U$. By (10.21), we have $\Theta(e,k) \leq c_0\delta - \kappa_\Theta u_U^2/2 \leq -3\kappa_\Theta u_U^2/8$. Fix $T > 0$, to be chosen below, and consider first $(e,k) \in [-T,T]^2$. Using (10.15), (10.27), and Lemma 10.2(b) with $a = \kappa_\Theta u_U^2/8$, we find $\lambda_N^{\mathrm{all}}(Ne,k) \leq Ce^{N(\Theta(e,k)+a)} \leq Ce^{-\kappa_\Theta u_U^2 N/4}$ there. Integrating over this part of the region with $\mathrm{d}h\,\mathrm{d}k = N\,\mathrm{d}e\,\mathrm{d}k$, we find a contribution of at most $CNe^{-\kappa_\Theta u_U^2 N/4}$. To discard the complement of $[-T,T]^2$, let $a_\Sigma > 0$ be the smallest eigenvalue of Σ_Q^{-1}. Similarly, (10.15), (10.27), and Lemma 10.2(c) give

$$\lambda_N^{\mathrm{all}}(Ne,k) \leq C\exp\Big(N\Big[C - \frac{a_\Sigma}{2}(e^2+k^2) + \log(|k|+C)\Big]\Big).$$

For T large enough, the bracket is at most $-a_\Sigma(e^2+k^2)/4$ outside $[-T,T]^2$, and the integral of the right side over this set, including the factor N, is at most Ce^{-cN}. This proves (10.24). The same argument, with $\Theta(e,k) \leq -c_0 t$ on $\{e \geq E_0 + t\}$ by (10.21) and $a = c_0 t/2$, proves (10.25).

For (d), we use (a) for $k \in \overline{U}$ and (10.21) in the form $\Theta(e,k) \leq c_0\delta' - \kappa_\Theta(e-E_0)^2/2$ for $e \geq E_0 - \delta'$. Integrating in $\mathrm{d}h\,\mathrm{d}k = N\,\mathrm{d}e\,\mathrm{d}k$ over $e \geq E_0 - \delta'$ and $k \in U$, we have the bound $CN|U|e^{c_0\delta' N}\int_{\mathbb{R}} e^{-N\kappa_\Theta v^2/2}\,\mathrm{d}v \leq CNe^{c_0\delta' N}$. □

The bound (10.26) is deliberately loose. It is used only to absorb conditional failure probabilities that are exponentially small. The precise normalization is proved in the next section.

10.4 Centering and expected counts in fixed energy windows

Admissible neighborhoods. For each $e \in \mathbb{R}$, the function $\Theta(e,\cdot)$ is strongly concave by Lemma 10.3 and tends to $-\infty$ at $\pm\infty$ by (10.21). Let $k_*(e)$ be its unique maximizer and $\theta(e) = \Theta(e,k_*(e))$. Then $k_*(E_0) = K_0$. Fix

$$c_- = \frac{c_0}{2}, \qquad c_+ = \frac{c_0+\beta}{2}, \qquad \text{so that} \qquad 0 < c_- < c_0 < c_+ < \beta. \tag{10.28}$$

We call a pair (U,δ), with U an open interval and $\delta > 0$, *admissible* if:

(A1) U is bounded, $K_0 \in U$, and $\overline{U} \subset (2b_Q^{1/2}, \infty)$;

(A2) $c_- \leq \ell(e,k) \leq c_+$ for all $(e,k) \in [E_0 - \delta, E_0 + \delta] \times \overline{U}$;

(A3) $\delta \leq \kappa_\Theta u_U^2/(8c_0)$, where $u_U = \mathrm{dist}(K_0, \mathbb{R}\setminus U)$;

(A4) $\operatorname{dist}(k_*(e), \mathbb{R} \setminus U) \geq u_U/2$ for all $|e - E_0| \leq \delta$.

Recall that $\ell(E_0, K_0) = c_0$ by (10.9). Since ℓ and k_* are continuous, admissible pairs exist. If (U, δ) is admissible, then (U, δ') is admissible for $0 < \delta' < \delta$, and for every open interval $U' \subset U$ containing K_0, there exists $\delta' > 0$ such that (U', δ') is admissible. The coefficient of e in ℓ is $(a_Q + b_Q)/\det \Sigma_Q > 0$, and therefore (A2) implies

$$\ell(e, k) \geq c_- \qquad \text{for all } e \geq E_0 - \delta,\ k \in \overline{U}. \tag{10.29}$$

For an admissible pair, (A3) is the hypothesis of Lemma 10.4(b). We set

$$\lambda_N(h) = \int_U \lambda_N^{\max}(h, k)\, \mathrm{d}k, \qquad T_N(h) = \int_h^\infty \lambda_N(v)\, \mathrm{d}v. \tag{10.30}$$

By Lemma 10.1, $T_N(h)$ is the expected number of points $u \in \mathrm{Crt}_N^{\max}$ with $H_Q(u) \geq h$ and $K_Q(u)/N \in U$.

Lemma 10.5 (Integral over the radial derivative and the slope in energy)**.** *Let ξ be a non-pure mixture and (U, δ) admissible.*

(a) The function k_* is smooth on $[E_0 - \delta, E_0 + \delta]$, and $\theta(E_0) = 0$, $\theta'(E_0) = -c_0$. In particular, $\theta(e) = -c_0(e - E_0) + O((e - E_0)^2)$.

(b) Uniformly for $|e - E_0| \leq \delta$,

$$cN^{-1/2} e^{N\theta(e)} \leq \lambda_N(Ne) \leq CN^{-1/2} e^{N\theta(e)}. \tag{10.31}$$

(c) Let

$$\mathcal{R}_N(e) = \int_U e^{-Nq(e,k)/2} d_N^{\max}(k)\, \mathrm{d}k, \qquad \textit{so that} \qquad \lambda_N(Ne) = \mathfrak{c}_N \mathcal{R}_N(e). \tag{10.32}$$

Then $\mathcal{R}_N(e) \asymp N^{-1/2} \exp(N[\theta(e) - \frac{1}{2} + \frac{1}{2}\log a_Q])$ uniformly for $|e - E_0| \leq \delta$.

(d) For $e \in \mathbb{R}$, let $p_{N,e}(\,\mathrm{d}k) = \mathcal{R}_N(e)^{-1} \mathbf{1}_U(k) e^{-Nq(e,k)/2} d_N^{\max}(k)\,\mathrm{d}k$ be the radial distribution of local maxima at energy Ne. Then, uniformly for $|e - E_0| \leq \delta$,

$$p_{N,e}\big(\{k : |k - k_*(e)| \geq t\}\big) \leq Ce^{-cNt^2}, \qquad t \geq 0. \tag{10.33}$$

(e) For all $h \in \mathbb{R}$,

$$\frac{\mathrm{d}}{\mathrm{d}h} \log \lambda_N(h) = -\int \ell(h/N, k)\, p_{N,h/N}(\,\mathrm{d}k). \tag{10.34}$$

In particular, $-(\log \lambda_N)'(h) \geq c_-$ for $h \geq N(E_0 - \delta)$, and $-(\log \lambda_N)'(h) \leq c_+$ for $|h - NE_0| \leq N\delta$. Moreover, for every fixed $D > 0$, uniformly for $|h - NE_0| \leq D \log N$,

$$\frac{\mathrm{d}}{\mathrm{d}h} \log \lambda_N(h) = -c_0 + O_D(N^{-1/2}). \tag{10.35}$$

Proof. On $\mathbb{R} \times (2b_Q^{1/2}, \infty)$, Θ is smooth and $\partial_k^2 \Theta \le -\kappa_\Theta$. We have $k_*(e) \in U$ for $|e - E_0| \le \delta$ by (A4). Then $\partial_k \Theta(e, k_*(e)) = 0$ there, and the implicit function theorem shows that k_* is smooth. We have $\theta'(e) = \partial_e \Theta(e, k_*(e))$, and (10.20) implies that $\theta(E_0) = 0$ and $\theta'(E_0) = -c_0$. This proves (a).

For (b), we note that $\Theta(e,k) \le \theta(e) - \kappa_\Theta (k - k_*(e))^2/2$ for all k, by the strong concavity of $\Theta(e, \cdot)$. Integrating (10.23) over $k \in U$ gives the upper bound in (10.31). For the lower bound, let C_2 be a bound for $|\partial_k^2 \Theta|/2$ on $[E_0 - \delta, E_0 + \delta] \times \overline{U}$. For $N^{-1/2} < u_U/2$, the interval $\{|k - k_*(e)| \le N^{-1/2}\}$ lies in U by (A4), and on it $\Theta(e,k) \ge \theta(e) - C_2/N$. Integrating the lower bound in (10.23) over this interval, we find $\lambda_N(Ne) \ge 2cN^{-1/2} e^{N\theta(e) - C_2}$. Part (c) follows from (b), the relation $\lambda_N(Ne) = \mathfrak{c}_N \mathcal{R}_N(e)$, and (10.27).

For (d), the same upper bound and $\int_{|v| \ge t} e^{-\kappa_\Theta N v^2/2} \, \mathrm{d}v \le C N^{-1/2} e^{-\kappa_\Theta N t^2/4}$ show that $\mathcal{R}_N(e)\, p_{N,e}(\{|k - k_*(e)| \ge t\})$ is at most $C N^{-1/2} e^{N(\theta(e) - \frac{1}{2} + \frac{1}{2}\log a_Q)} e^{-\kappa_\Theta N t^2/4}$. Dividing by the lower bound in (c), we have (10.33) with $c = \kappa_\Theta/4$.

For (e), note that $d_N^{\max}(k)$ depends on k only. By (10.2), $\partial_h [Nq(h/N, k)/2] = \ell(h/N, k)$, and differentiation under the integral over the bounded interval U is justified by dominated convergence. This proves (10.34). The two inequalities that follow are (10.29) and (A2), since $p_{N,h/N}$ is a probability measure on U. For (10.35), let $e = h/N$, so that $|e - E_0| \le D(\log N)/N$. Since ℓ is linear and $\ell(E_0, K_0) = c_0$, and $|k_*(e) - K_0| \le C|e - E_0|$ by (a), we have

$$\Big| \int \ell(e,k)\, p_{N,e}(\,\mathrm{d}k) - c_0 \Big| \le C\Big(|e - E_0| + \int |k - k_*(e)|\, p_{N,e}(\,\mathrm{d}k) \Big) \le C\Big(\frac{D \log N}{N} + N^{-1/2} \Big),$$

where in the last step we integrated the tail bound (10.33) in t. □

A version of $p_{N,e}$ that acts on functions of the whole field is also needed in Section 12. For each fixed $u \in S_N$, we define the Palm kernel of local maxima at u with energy Ne by the following formula, valid for every nonnegative Borel function $f(H,k)$:

$$\mathbb{E}^{\mathrm{Palm,max}}_{N,Ne,u} f(H,k) = \frac{1}{\mathcal{R}_N(e)} \int_U e^{-Nq(e,k)/2}\, \mathbb{E}_{e,k,u}\big[D(u) \mathbf{1}\{\nabla^2_{S_N} H_Q(u) \prec 0\} f(H,k) \big]\, \mathrm{d}k. \tag{10.36}$$

Integrating this kernel against $\nu_N(\,\mathrm{d}u)$ gives the Palm law of local maxima at energy Ne. By Lemma 10.1, the expected sum of $f(H, K_Q(u)/N)$ over the points $u \in \mathrm{Crt}_N^{\max}$ with $K_Q(u)/N \in U$ and $H_Q(u) \in B$ is $\int_B \lambda_N(h) \int_{S_N} \mathbb{E}^{\mathrm{Palm,max}}_{N,h,u} f \, \nu_N(\,\mathrm{d}u)\, \mathrm{d}h$. Here $D(u)$ is the actual determinant of the conditioned field, which is correlated with f. Only its marginal law is that of $D_N(k)$. The normalization integral $\mathcal{R}_N(e)$ and the intensity $\lambda_N(Ne) = \mathfrak{c}_N \mathcal{R}_N(e)$ differ by the factor $\mathfrak{c}_N$, and the two must not be interchanged. By (10.14) and (10.27), we have $N^{-1} \log \mathfrak{c}_N \to (1 - \log a_Q)/2$, so this factor grows exponentially if $\log a_Q < 1$ and decays exponentially if $\log a_Q > 1$.

Theorem 10.6 (Centering, expected counts, and tails in fixed energy windows)**.** *Let ξ be a non-pure mixture and (U, δ) admissible.*

(a) *For all large N, there exists a unique $m_N \in \mathbb{R}$ with $T_N(m_N) = 1$, and*

$$m_N = NE_0 - \frac{\log N}{2c_0} + O(1). \tag{10.37}$$

For the finitely many remaining N, we set $m_N = NE_0$.

(b) *For every fixed $t \in \mathbb{R}$,*

$$\lambda_N(m_N + t) \longrightarrow c_0 e^{-c_0 t}, \qquad T_N(m_N + t) \longrightarrow e^{-c_0 t}. \tag{10.38}$$

(c) *For all fixed $a < b$,*

$$\mathbb{E}\#\big\{u \in \mathrm{Crt}_N : K_Q(u)/N \in U,\ H_Q(u) - m_N \in [a,b]\big\} \longrightarrow e^{-c_0 a} - e^{-c_0 b}, \tag{10.39}$$

and the same limit holds with Crt_N replaced by $\mathrm{Crt}_N^{\max}$. In particular, the expected number of points in the window $[m_N - L, m_N + L]$ tends to $\mu_L = e^{c_0 L} - e^{-c_0 L}$, and every fixed window has expected count bounded uniformly in N.

(d) *For all large N and all $v \geq 0$, $T_N(m_N + v) \leq e^{-c_- v}$.*

(e) *For all large N and $0 \leq v \leq m_N - N(E_0 - \delta)$, $\lambda_N(m_N - v) \leq Ce^{c_+ v}$, where $c_+ < \beta$.*

(f) *For every $L \geq 0$,*

$$\limsup_{N\to\infty} \mathbb{P}\Big(\max_{S_N} H_Q > m_N + L\Big) \leq e^{-c_0 L}. \tag{10.40}$$

More precisely, with probability at least $1 - Ce^{-cN}$, every global maximizer u of H_Q with $H_Q(u) \geq N(E_0 - \delta)$ belongs to $\mathrm{Crt}_N^{\max}$ and satisfies $K_Q(u)/N \in U$.

Proof. We begin with (a). By Lemma 10.5(e), for $h \geq N(E_0 - \delta)$ and $v \geq 0$,

$$\lambda_N(h + v) \leq \lambda_N(h) e^{-c_- v}. \tag{10.41}$$

Fix $D > 1/(2c_0) + 1$ and let $|h - NE_0| \leq D \log N$. Integrating (10.41) in v, we have $T_N(h) \leq \lambda_N(h)/c_-$. The upper bound on the slope in Lemma 10.5(e) implies that $T_N(h) \geq \int_0^1 \lambda_N(h+v)\,\mathrm{d}v \geq e^{-c_+}\lambda_N(h)$. Combining these bounds with (10.31) and Lemma 10.5(a), we obtain

$$T_N(h) \asymp \lambda_N(h) \asymp N^{-1/2} e^{-c_0(h - NE_0)}, \qquad |h - NE_0| \leq D \log N, \tag{10.42}$$

where the omitted factor $\exp(O(D^2(\log N)^2/N))$ is bounded above and below. We have $c_0 D > 1/2$, and (10.42) shows that $T_N(NE_0 - D\log N) > 1 > T_N(NE_0 + D \log N)$ for all large N. The density λ_N is positive and continuous on $\mathbb{R}$, because $d_N^{\max} > 0$ on $\overline{U}$ by (10.16). Then T_N is continuous and strictly decreasing, and by the intermediate value theorem it has a unique root m_N, with $|m_N - NE_0| < D \log N$. At $h = m_N$, (10.42) reads $1 \asymp N^{-1/2} e^{-c_0(m_N - NE_0)}$, which is (10.37).

Next, we prove (b). For every fixed $v \in \mathbb{R}$, the point $m_N + v$ lies within $(D+1)\log N$ of NE_0 for large N, and integrating (10.35), with $D+1$ in place of D, between m_N and $m_N + v$ yields

$$\frac{\lambda_N(m_N+v)}{\lambda_N(m_N)} \longrightarrow e^{-c_0 v}, \tag{10.43}$$

uniformly for v in bounded sets. For $v \geq 0$, the ratio is at most $e^{-c_- v}$ by (10.41), since $m_N \geq N(E_0 - \delta)$. By dominated convergence,

$$1 = T_N(m_N) = \lambda_N(m_N) \int_0^\infty \frac{\lambda_N(m_N+v)}{\lambda_N(m_N)}\,\mathrm{d}v = \lambda_N(m_N)\big(c_0^{-1} + o(1)\big),$$

so $\lambda_N(m_N) \to c_0$, and (10.43) then implies the first limit in (10.38). Integrating the same ratios from t to ∞, we find the second limit. If $t < 0$, the additional interval $[t, 0]$ is bounded and the convergence there is uniform.

For (c), the expected number of local maxima in the window is $T_N(m_N + a) - T_N(m_N + b)$, which converges to $e^{-c_0 a} - e^{-c_0 b}$ by (b). For all critical points, the intensity λ_N^{all} differs from $\lambda_N^{\max}$ by the factor $d_N^{\mathrm{all}}(k)/d_N^{\max}(k) = 1 + O(e^{-cN})$ on $\overline{U}$, by (10.16), and the expected counts in the windows have the same limit. For the finitely many N below the threshold, the expected number of critical points is finite by Lemma 10.1, which implies that every fixed window has bounded expected count.

For (d), we integrate $\lambda_N(m_N + v + w) \leq e^{-c_- v}\lambda_N(m_N + w)$, which is (10.41), over $w \geq 0$, and find $T_N(m_N + v) \leq e^{-c_- v} T_N(m_N) = e^{-c_- v}$. For (e), integrating the upper slope bound $-(\log \lambda_N)' \leq c_+$ of Lemma 10.5(e) over $[m_N - v, m_N]$, we conclude that $\lambda_N(m_N - v) \leq e^{c_+ v}\lambda_N(m_N)$, and $\lambda_N(m_N)$ is bounded by (b).

Finally, we prove (f). By Lemma 10.4(b) and Markov's inequality, with probability at least $1 - Ce^{-cN}$ every critical point u with $H_Q(u) \geq N(E_0 - \delta)$ satisfies $K_Q(u)/N \in U$. Every global maximizer is a critical point, and since H_Q is almost surely a Morse function, it belongs to $\mathrm{Crt}_N^{\max}$. This proves the second statement. For large N, $m_N + L \geq N(E_0 - \delta)$, and the event $\{\max H_Q > m_N + L\}$ intersected with the previous one is contained in the event that some $u \in \mathrm{Crt}_N^{\max}$ has $H_Q(u) \geq m_N + L$ and $K_Q(u)/N \in U$. Using Markov's inequality and (b), we obtain

$$\mathbb{P}\Big(\max_{S_N} H_Q > m_N + L\Big) \leq T_N(m_N + L) + Ce^{-cN} \longrightarrow e^{-c_0 L},$$

which completes the proof. □

Remark 10.7 (Dependence of the centering on U). Let (U, δ) and (U', δ') be admissible, with $U' \subset U$, and let $\delta_* = \min\{\delta, \delta'\}$. We write T_N^U, $T_N^{U'}$ and m_N^U, $m_N^{U'}$ for the corresponding tails and centerings. For $h \geq N(E_0 - \delta_*)$, the difference $T_N^U(h) - T_N^{U'}(h)$ is the expected number of local maxima with $H_Q(u) \geq h$ and normalized radial derivative in $U \setminus U'$, and Lemma 10.4(b), applied to U' with the width δ', gives

$$0 \leq T_N^U(h) - T_N^{U'}(h) \leq Ce^{-cN}, \qquad h \geq N(E_0 - \delta_*).$$

Both centerings lie above $N(E_0 - \delta_*)$ for all large N. Since $T_N^{U'} \le T_N^U$, we have $m_N^{U'} \le m_N^U$. Using the display, we find that $T_N^U(m_N^{U'}) \le T_N^{U'}(m_N^{U'}) + Ce^{-cN} = 1 + Ce^{-cN}$, while $T_N^U(m_N^U - 1) \to e^{c_0} > 1$ by Theorem 10.6(b), so $m_N^{U'} > m_N^U - 1$ for large N. On $[m_N^U - 1, m_N^U]$, the density λ_N^U is bounded below by $c_0/2$ for large N, by (10.38) and the uniformity of (10.43). Integrating this lower bound, we find $1 + c_0(m_N^U - m_N^{U'})/2 \le T_N^U(m_N^{U'}) \le 1 + Ce^{-cN}$, that is, $0 \le m_N^U - m_N^{U'} \le Ce^{-cN}$. The common width δ_* is needed: the region of $U \setminus U'$ above a larger energy threshold may contain marks of positive complexity. In the pure case, the logarithmic correction in (10.37) agrees with [144, Eq. (1.4)], expressed there using minima and counting antipodal pairs once when p is even.

10.5 Conditional regularity and localization

For a field $H = H_N$, we define the derivative norm

$$\mathcal{D}_N(H) = \frac{1}{N} \sup_{S_N} |H_Q| + \frac{1}{N^{1/2}} \sup_{S_N} |\nabla_{S_N} H_Q| + \sup_{S_N} \|\nabla^2_{S_N} H_Q\|_{\mathrm{op}} + N^{1/2} \sup_{S_N} \|\nabla^3_{S_N} H_Q\|_{\mathrm{op}}, \tag{10.44}$$

where $\nabla^3_{S_N} H_Q$ is the covariant derivative of the Riemannian Hessian and $\|\cdot\|_{\mathrm{op}}$ is the norm of a multilinear form on unit tangent vectors. The suprema may be taken over a countable dense set, and $\mathcal{D}_N$ is a Borel function of the field.

Lemma 10.8 (Derivative bounds under conditioning)**.** *Let $\mathcal{K} \subset \mathbb{R}^2$ be compact. There exists $C > 0$ such that, for all $N \ge 2$, $u \in S_N$, and $(e,k) \in \mathcal{K}$,*

$$\mathbb{P}_{e,k,u}\big(\mathcal{D}_N(H) > C\big) \le Ce^{-N} \qquad \text{and} \qquad \mathbb{P}\big(\mathcal{D}_N(H) > C\big) \le Ce^{-N}. \tag{10.45}$$

In a pure model, the first bound holds under $\mathbb{P}_{e,u}$, uniformly for e in a compact interval.

Proof. By (7.1), $H_Q(v) = \sum_p (c_p Q^p N)^{1/2} T_p[(v/N^{1/2})^{\otimes p}]$, where $T_p[w_1, \dots, w_p] = \sum_{i_1,\dots,i_p} Z_{p,i_1,\dots,i_p} (w_1)_{i_1} \cdots (w_p)_{i_p}$. Under $\mathbb{P}_{e,k,u}$, the coefficient vector is $(\mathrm{Id} - \Pi_u)Z + m$ by (7.4), where $\Pi_u = \Lambda_u^* \mathcal{C}_u^{-1} \Lambda_u$ is an orthogonal projection and m is deterministic. Accordingly, $H_Q = H^{\mathrm{res}} + H^{\mathrm{mean}}$, where H^{res} is defined by the tensors T_p^{res} with coefficient vector $(\mathrm{Id} - \Pi_u)Z$, and $H^{\mathrm{mean}} = \mathbb{E}_{e,k,u} H_Q$.

We first bound the residual tensors. For unit vectors $w_1, \dots, w_p$, the variable $T_p^{\mathrm{res}}[w_1, \dots, w_p]$ is a centered Gaussian whose variance is the squared norm of the image of $w_1 \otimes \cdots \otimes w_p$ under the projection $\mathrm{Id} - \Pi_u$, hence at most one. Let $\|T\| = \sup |T[w_1, \dots, w_p]|$ over unit vectors, and let $\mathcal{N}$ be a $(2p)^{-1}$-net of the unit sphere of $\mathbb{R}^N$ with $|\mathcal{N}| \le (1 + 4p)^N$. Replacing one slot at a time by its nearest net point, we find $\|T\| \le \max_{\mathcal{N}^p} |T| + p(2p)^{-1}\|T\|$, that is, $\|T\| \le 2 \max_{\mathcal{N}^p} |T|$. Let $A_p^2 = 2p \log(1 + 4p) + 2$. Using a union bound over $\mathcal{N}^p$ and the Gaussian tail, we obtain

$$\mathbb{P}\big(\|T_p^{\mathrm{res}}\| > 2A_p N^{1/2}\big) \le 2(1 + 4p)^{pN} e^{-A_p^2 N/2} = 2e^{-N}.$$

The j-th Euclidean derivative of $(c_p Q^p N)^{1/2} T_p[(v/N^{1/2})^{\otimes p}]$ is a sum of at most p^j contractions. On the event $\|T_p^{\mathrm{res}}\| \le 2A_p N^{1/2}$, its operator norm is at most $CN^{1-j/2}$ for $|v| \le N^{1/2}$.

Next, we bound the mean. The gradient target in (7.4) is zero, and Gaussian regression shows that

$$H^{\mathrm{mean}}(v) = N\big(\xi_Q(R(v,u)), R(v,u)\xi_Q'(R(v,u))\big)\Sigma_Q^{-1}(e,k)^{\mathsf{T}} = N\sum_p c_p Q^p \lambda_p R(v,u)^p,$$

with $\lambda_p = (1,p)\Sigma_Q^{-1}(e,k)^{\mathsf{T}}$, which is bounded for $(e,k) \in \mathcal{K}$. The j-th Euclidean derivative of $R(v,u)^p = (v\cdot u/N)^p$ is $p(p-1)\cdots(p-j+1)R(v,u)^{p-j}(u/N)^{\otimes j}$, and we have $\|D^j H^{\mathrm{mean}}(v)\|_{\mathrm{op}} \le CN^{1-j/2}$ for $|v| \le N^{1/2}$. In the pure case, the mean is $NeR(v,u)^p$, and without conditioning it is zero. The same bound holds in both cases. As there are finitely many degrees, we conclude that, outside an event of probability at most Ce^{-N},

$$\sup_{|v|\le N^{1/2}} \|D^j H_Q(v)\|_{\mathrm{op}} \le CN^{1-j/2}, \qquad j = 0,1,2,3. \tag{10.46}$$

It remains to pass to the sphere. The spherical gradient is the tangential part of DH_Q. For tangent vectors X, Y at $v \in S_N$, the Riemannian Hessian is $\nabla^2_{S_N} H_Q(X,Y) = D^2H_Q(X,Y) - N^{-1}\langle X,Y\rangle\, v\cdot DH_Q(v)$, since the second fundamental form of the sphere of radius $N^{1/2}$ is $-N^{-1}\langle X,Y\rangle v$. Differentiating this identity along a tangent vector Z and using that the connection is compatible with the metric, we find

$$\nabla^3_{S_N} H_Q(Z,X,Y) = D^3H_Q(Z,X,Y) - \frac{1}{N}\Big[\langle Z,X\rangle D^2H_Q(v,Y) + \langle Z,Y\rangle D^2H_Q(v,X) + \langle X,Y\rangle\big(DH_Q(Z) + D^2H_Q(v,Z)\big)\Big].$$

For unit tangent vectors, (10.46) and $|v| = N^{1/2}$ bound each term on the right by $CN^{-1/2}$, and similarly $\|\nabla^2_{S_N} H_Q\|_{\mathrm{op}} \le C$ and $|\nabla_{S_N} H_Q| \le CN^{1/2}$. This proves (10.45). □

The next statement is deterministic.

Lemma 10.9 (No critical point near a strict maximum)**.** *Let $\kappa, C_3 > 0$ and let f be a smooth function on S_N with $\sup_{S_N}\|\nabla^3_{S_N} f\|_{\mathrm{op}} \le C_3 N^{-1/2}$. Let u be a critical point of f with $\nabla^2_{S_N} f(u) \preceq -\kappa I$. Let $t_* = \min\{\pi/4, \kappa/(2C_3)\}$ and $\eta = 1 - \cos t_*$. Then f has no critical point $v \ne u$ with $R(u,v) \ge 1-\eta$.*

Proof. Let $v \ne u$ with $R(u,v) \ge 1-\eta = \cos t_*$, and let γ be the unit-speed minimizing geodesic from u to v, of length $l_v = N^{1/2}\arccos R(u,v) \le t_* N^{1/2}$. Let $\phi(l) = f(\gamma(l))$. We have $\phi''(l) = \nabla^2_{S_N} f(\gamma(l))(\gamma'(l),\gamma'(l))$, because γ is a geodesic, and γ' is parallel along γ. Differentiating along γ, we find $|\phi''(l) - \phi''(0)| \le lC_3N^{-1/2}$. This implies that $\phi''(l) \le -\kappa + C_3 t_* \le -\kappa/2$ for $0 \le l \le t_* N^{1/2}$. Since $\phi'(0) = 0$, we conclude that $\phi'(l_v) < 0$. In particular, $\nabla_{S_N} f(v) \ne 0$. □

Base choices. We fix, once and for all, a bounded open interval U_0 with $K_0 \in U_0$ and $\overline{U}_0 \subset (2b_Q^{1/2}, \infty)$, and set

$$g_0 = \inf_{k \in \overline{U}_0} \bigl(k - 2b_Q^{1/2}\bigr) > 0, \qquad \kappa = \frac{g_0}{4}. \tag{10.47}$$

Let C_{reg} be a constant for which (10.45) holds both unconditionally and under $\mathbb{P}_{e,k,u}$, uniformly for $(e,k) \in \mathcal{K}_0 = [E_0 - 1, E_0 + 1] \times \overline{U}_0$, and set

$$t_* = \min\Bigl\{\frac{\pi}{4}, \frac{\kappa}{2C_{\mathrm{reg}}}\Bigr\}, \qquad \eta = 1 - \cos t_*. \tag{10.48}$$

These constants depend only on ξ, β, and the choice of U_0. For instance, U_0 may be the interval of Lemma 5.14(2); the number g_0 defined here is then three quarters of the number denoted g_0 there. For $k \in \overline{U}_0$, we have $k - \kappa \geq 2b_Q^{1/2} + 3g_0/4$. By (10.12) and the Hessian law, for all $u \in S_N$ and $e \in \mathbb{R}$,

$$\mathbb{P}_{e,k,u}\bigl(\nabla^2_{S_N} H_Q(u) \not\preceq -\kappa I\bigr) \leq 2e^{-c_H N}, \qquad c_H = \frac{9g_0^2}{64 b_Q}. \tag{10.49}$$

Estimates of the kind proved in part (a) of the next proposition were obtained by Sellke for all approximate ground states, not only for critical points. The field H_Q is a spherical spin glass with mixture ξ_Q, and its ground-state energy is E_0. Applied to H_Q, [138, Proposition 1.2] shows for every ξ that, for each $\varepsilon > 0$, there exists $\delta > 0$ such that, for all large N, with probability at least $1 - e^{-cN}$, every $u \in S_N$ with $H_Q(u) \geq N(E_0 - \delta)$ satisfies $|\nabla H_Q(u) - ru| \leq \varepsilon\sqrt{N}$, where r is a constant given by the zero-temperature Parisi formula for ξ_Q. At a critical point, the Euclidean gradient is $K_Q(u)u/N$, so this gives $|K_Q(u)/N - r| \leq \varepsilon$. Applying this and part (a) below to a maximizer of H_Q, we find that $r = K_0$. For even ξ, [138, Theorem 1.3] shows that the largest eigenvalue of $\nabla^2_{S_N} H_Q(u)$ at the same points is at most $2b_Q^{1/2} - r + \varepsilon$. Since $2b_Q^{1/2} - K_0 \leq -g_0 = -4\kappa$, this gives the bound on the Hessian in (10.50) for even ξ, with $\varepsilon \leq 3\kappa$. Our proof uses the Kac–Rice formula and the 1RSB assumptions, and it allows odd degrees. Xu and Yang [154, Theorem 1.3] proved convergence of the Hessian empirical measure and its extreme eigenvalues at the global minimizer for zero-temperature 1RSB models with a strict spectral gap, allowing odd degrees. Their result concerns the global minimizer; Proposition 10.10 gives uniform localization of all high critical points of the center field.

Proposition 10.10 (High critical points are strict local maxima)**.** *Let ξ be a non-pure mixture, and let U be an open interval with $K_0 \in U$ and $\overline{U} \subset U_0$. There exists $\delta_U > 0$ such that, for every $\delta \in (0, \delta_U]$, there exist $c, C > 0$ with the following property. With probability at least $1 - Ce^{-cN}$:*

(a) every $u \in \mathrm{Crt}_N$ with $H_Q(u) \geq N(E_0 - \delta)$ satisfies

$$H_Q(u) \leq N(E_0 + \delta), \qquad K_Q(u)/N \in U, \qquad \nabla^2_{S_N} H_Q(u) \preceq -\kappa I; \tag{10.50}$$

(b) $\mathcal{D}_N(H) \le C_{\mathrm{reg}}$*; in particular* $\sup_{S_N} \|\nabla^3_{S_N} H_Q\|_{\mathrm{op}} \le C_{\mathrm{reg}} N^{-1/2}$*;*
(c) no critical point v *of* H_Q *other than* u *satisfies* $R(u,v) \ge 1-\eta$*, for any* u *as in (a).*

Proof. By fact (4) above and (10.49), for $k \in \overline{U}$ we have

$$\mathbb{E}_{e,k,u}\big[D(u)\mathbf{1}\{\nabla^2_{S_N} H_Q(u) \not\preceq -\kappa I\}\big] \le \mathsf{K}_2^{1/2}\big(2e^{-c_H N}\big)^{1/2} d_N^{\mathrm{all}}(k) \le Ce^{-c_H N/2} d_N^{\mathrm{all}}(k).$$

This bound uses no independence between the determinant and the event. Let $\delta_U = \min\{\kappa_\Theta u_U^2/(8c_0), c_H/(4c_0), 1\}$ and $0 < \delta \le \delta_U$. By Lemma 10.1 with $f = \mathbf{1}\{\nabla^2_{S_N} H_Q(u) \not\preceq -\kappa I\}$ and $A = [N(E_0-\delta),\infty) \times U$, and by (10.26), the expected number of $u \in \mathrm{Crt}_N$ with $H_Q(u) \ge N(E_0-\delta)$, $K_Q(u)/N \in U$, and $\nabla^2_{S_N} H_Q(u) \not\preceq -\kappa I$ is at most

$$Ce^{-c_H N/2} \cdot CNe^{c_0 \delta N} \le CNe^{-c_H N/4}.$$

The expected number of critical points with $H_Q(u) \ge N(E_0-\delta)$ and $K_Q(u)/N \notin U$ is exponentially small by Lemma 10.4(b), whose hypothesis holds since $\delta \le \kappa_\Theta u_U^2/(8c_0)$, and the expected number with $H_Q(u) \ge N(E_0+\delta)$ is exponentially small by Lemma 10.4(c). Using Markov's inequality, we find that (a) holds with probability at least $1 - Ce^{-cN}$. Part (b) holds with probability at least $1 - C_{\mathrm{reg}} e^{-N}$ by the unconditional bound in (10.45). On the intersection of the two events, (c) follows from Lemma 10.9 with $C_3 = C_{\mathrm{reg}}$. □

Proposition 10.10 concerns all critical points, not only those counted as local maxima in advance. Part (c) excludes only the endpoint of positive overlap. It says nothing about points near $-u$, which matter for mixtures with odd degrees and are treated in Section 11.

10.6 Exclusion of high-energy points at positive overlap

This exclusion estimate is due to Huang and Sellke [90, Section 3.3]. They showed that, conditionally on a critical point u whose normalized energy and normalized radial derivative lie in a small box offset from the ground-state parameters, with probability $1 - e^{-cN}$ the field has no high point at intermediate overlap with u and no other critical point near u [90, Proposition 3.10]. Their proof combines the law of the field on a latitude of u [90, Lemma 3.11], an explicit one-step trial for the conditional ground state [90, Proposition 3.13], and the one-step upper bound below. We need the conclusion at the ground-state parameters (E_0, K_0) themselves, uniformly on a rectangle around them, with a failure probability that remains exponentially small after weighting by the Hessian determinant. Their Proposition 3.10 does not cover such a rectangle. We prove the estimate here, as Theorem 10.13. We follow their route. We compute the latitude law and the trial value again, and keep the exact slack $g(r)$ of the zero-temperature contact inequality at (E_0, K_0) (Lemma 10.12).

For a covariance $\chi(t) = \sum_{j=1}^{J} \gamma_j t^j$ with $\gamma_j \geq 0$, let H_m^χ be the centered Gaussian field on S_m with covariance $m\chi(R_m(\sigma,\sigma'))$, where $R_m(\sigma,\sigma') = \sigma\cdot\sigma'/m$. For $L > 0$, $\alpha \geq 0$, and $q_0 \in [0,1]$ such that $b(t) = L - \alpha(t-q_0)_+$ is positive on $[0,1]$, set

$$\mathcal{Q}_\chi(b) = \frac{1}{2}\Big[\chi'(0)L + \int_0^1 \chi''(t)b(t)\,\mathrm{d}t + \int_0^1 \frac{\mathrm{d}t}{b(t)}\Big]. \tag{10.51}$$

Theorem 10.11 (Huang–Sellke [90, Proposition 2.5])**.** *Let χ and b be as above. Then, for every $t > 0$,*

$$\lim_{m\to\infty} \mathbb{P}\Big(\frac{1}{m}\max_{S_m} H_m^\chi > \mathcal{Q}_\chi(b) + t\Big) = 0.$$

In [90], the trial is written as (L,α) with $\alpha(t) = \alpha\mathbf{1}\{t \geq q_0\}$ and $b(t) = L - \int_0^t \alpha$, and (10.51) is their functional [90, Eq. (2.1)]. Positivity of b is their admissibility condition $L > \int_0^1 \alpha$. They write the external field as the degree-one term of the mixture [90, footnote to Eq. (1.1)]. In particular, χ may have a linear term, and no parity or replica-symmetry assumption on χ is needed. As they explain in [90, Remark 2.6], the bound follows from the zero-temperature Crisanti–Sommers formula of Chen and Sen [48].

For $u \in S_N$ and $0 < r < 1$, we write

$$\mathrm{Lat}_r(u) = \{w \in S_N : R(u,w) = r\} = \big\{ru + (1-r^2)^{1/2}v :\ v \perp u,\ |v|^2 = N\big\}$$

for the latitude of u at overlap r.

Lemma 10.12 (Latitude law and the trial at the ground-state parameters)**.** *Let ξ be a non-pure mixture, $u \in S_N$, and $0 < r < 1$.*

(a) Under $\mathbb{P}_{e,k,u}$, the restriction of H_Q to $\mathrm{Lat}_r(u)$ equals $N\mu_{e,k}(r) + \widehat{H}_r$, where $\widehat{H}_r$ is a centered Gaussian field whose law does not depend on (e,k), with

$$\begin{aligned} \mathbb{E}\widehat{H}_r(w_1)\widehat{H}_r(w_2) &= N\tau_r\Big(\frac{v_1\cdot v_2}{N}\Big), \\ \tau_r(t) &= \xi_Q\big(r^2 + (1-r^2)t\big) - C_r - \frac{(1-r^2)\xi_Q'(r)^2}{a_Q}t, \end{aligned} \tag{10.52}$$

for $w_i = ru + (1-r^2)^{1/2}v_i$, where $C_r = (\xi_Q(r), r\xi_Q'(r))\Sigma_Q^{-1}(\xi_Q(r), r\xi_Q'(r))^{\mathsf{T}}$. The mean is

$$\mu_{e,k}(r) = \big(\xi_Q(r), r\xi_Q'(r)\big)\Sigma_Q^{-1}(e,k)^{\mathsf{T}} = \frac{r\xi_Q'(r) + z\xi_Q(r)}{y} + \varpi(r)\cdot(e - E_0, k - K_0), \tag{10.53}$$

with $\varpi(r) = \Sigma_Q^{-1}(\xi_Q(r), r\xi_Q'(r))^{\mathsf{T}}$.

(b) We have $\tau_r(0) \geq 0$, and $\bar\tau_r(t) = \tau_r(t) - \tau_r(0)$ is a polynomial in t with nonnegative coefficients and $\bar\tau_r(0) = 0$. Its linear coefficient may be positive.

(c) Let

$$D_r = 1 + (1-r)z, \qquad t_r = \frac{r}{1+r}, \qquad L_r = \frac{D_r}{(1-r^2)y}, \qquad b_r(t) = L_r - \frac{z}{y}(t-t_r)_+.$$

Then $b_r(1) = ((1-r^2)y)^{-1} > 0$, *and*

$$\mu_{E_0,K_0}(r) + \mathcal{Q}_{\bar\tau_r}(b_r) = E_0 - \frac{z}{2y}g(r) - \frac{[ra_Q - D_r\xi_Q'(r)]^2}{2ya_QD_r}. \tag{10.54}$$

(d) For $0 < \varepsilon < 1-\eta$, *let* $G_\varepsilon = (z/(2y)) \min_{[\varepsilon,1-\eta]} g$, *which is positive by* (10.8). *Then* $\mu_{E_0,K_0}(r) + \mathcal{Q}_{\bar\tau_r}(b_r) \le E_0 - G_\varepsilon$ *for all* $r \in [\varepsilon, 1-\eta]$.

Proof. We begin with (a). For $w = ru + (1-r^2)^{1/2}v$, the covariances of $H_Q(w)$ with $H_Q(u)$ and $K_Q(u)$ are $N\xi_Q(r)$ and $Nr\xi_Q'(r)$, and its covariance with $\nabla_{S_N}H_Q(u)$ is the tangent vector $\xi_Q'(r)(1-r^2)^{1/2}v$. We use Gaussian regression with the covariance $N\Sigma_Q \oplus a_Q\mathrm{Id}_{u^\perp}$ of the conditioned variables and zero gradient target. It yields the first expression in (10.53) and the covariance (10.52), since $\mathbb{E}H_Q(w_1)H_Q(w_2) = N\xi_Q(r^2 + (1-r^2)v_1\cdot v_2/N)$. The residual is the image of $(\mathrm{Id} - \Pi_u)Z$, whose law does not depend on (e,k). The second expression in (10.53) follows from (10.9), since $(\xi_Q(r), r\xi_Q'(r))(c_0, 1/y)^{\mathsf T} = (z\xi_Q(r) + r\xi_Q'(r))/y$. This is the computation of [90, Lemma 3.11], with K_Q/N as the radial derivative.

Next, we prove (b). We write $H_Q = \sum_p H_{Q,p}$ for the decomposition into homogeneous components, which are independent with $\mathbb{E}H_{Q,p}(x)H_{Q,p}(x') = Nc_pQ^p(x\cdot x'/N)^p$. Let $X_r = \sum_p r^pH_{Q,p}(u)$. Then $\operatorname{Var} X_r = N\xi_Q(r^2)$ and $\operatorname{Cov}(X_r, H_Q(u)) = N\xi_Q(r)$. By Euler's identity, $K_Q(u) = \sum_p pH_{Q,p}(u)$, which gives $\operatorname{Cov}(X_r, K_Q(u)) = Nr\xi_Q'(r)$. Each $H_{Q,p}(u)$ is uncorrelated with $\nabla_{S_N}H_Q(u)$. The variance of X_r minus its regression on $(H_Q(u), K_Q(u), \nabla_{S_N}H_Q(u))$ is therefore $N(\xi_Q(r^2) - C_r) = N\tau_r(0) \ge 0$. The coefficient of t in τ_r is $(1-r^2)[\xi_Q'(r^2) - \xi_Q'(r)^2/a_Q]$, which is nonnegative by the Cauchy–Schwarz inequality

$$\xi_Q'(r)^2 = \Big(\sum_p pc_pQ^pr^{p-1}\Big)^2 \le \Big(\sum_p pc_pQ^p\Big)\Big(\sum_p pc_pQ^pr^{2p-2}\Big) = a_Q\xi_Q'(r^2).$$

For $j \ge 2$, the coefficient of t^j is $(1-r^2)^j\xi_Q^{(j)}(r^2)/j! \ge 0$, since ξ_Q has nonnegative coefficients and $r^2 \ge 0$. This holds for every parity.

For (c), the value of $b_r(1)$ follows from $D_r - z(1-r) = 1$. We note that $b_r = L_r$ on $[0, t_r]$. After an integration by parts on $[t_r, 1]$, the first two terms of $2\mathcal{Q}_{\bar\tau_r}(b_r)$ combine to

$$\bar\tau_r'(t_r)L_r + \int_{t_r}^1 \bar\tau_r''(t)b_r(t)\,\mathrm{d}t = \bar\tau_r'(1)b_r(1) + \frac{z}{y}\big(\bar\tau_r(1) - \bar\tau_r(t_r)\big).$$

Since $r^2 + (1-r^2)t_r = r$, we have $\bar\tau_r'(1) = (1-r^2)(a_Q - \xi_Q'(r)^2/a_Q)$ and $\bar\tau_r(1) - \bar\tau_r(t_r) = \xi(Q) - \xi_Q(r) - (1-r)\xi_Q'(r)^2/a_Q$. These two terms are equal to

$$\frac{a_Q^2 + z a_Q(\xi(Q) - \xi_Q(r)) - D_r \xi_Q'(r)^2}{y a_Q}.$$

The last two terms of $2\mathcal{Q}_{\bar\tau_r}(b_r)$ are

$$\frac{t_r}{L_r} + \int_{t_r}^1 \frac{\mathrm{d}t}{b_r(t)} = \frac{r(1-r)y}{D_r} + \frac{y}{z}\log\frac{L_r}{b_r(1)} = \frac{r(1-r)y}{D_r} + \frac{y}{z}\log D_r.$$

Using the definition (10.8) of g and $y^2 = (1+z)a_Q$, we find $(y/z)\log D_r = y^{-1}[z(\xi(Q) - \xi_Q(r) - g(r)) + (1-r)a_Q]$. We now add $\mu_{E_0,K_0}(r) = (r\xi_Q'(r) + z\xi_Q(r))/y$ to half the sum of these expressions and subtract $E_0 = (a_Q + z\xi(Q))/y$. Multiplying by y, the terms containing $\xi_Q(r)$ and $\xi(Q)$ cancel, and we find

$$y\big[\mu_{E_0,K_0}(r) + \mathcal{Q}_{\bar\tau_r}(b_r) - E_0\big] = r\xi_Q'(r) - \frac{D_r\xi_Q'(r)^2}{2a_Q} - \frac{r^2 a_Q}{2D_r} - \frac{z g(r)}{2},$$

where we used $y^2 = (1+z)a_Q$ and $r(1-r)(1+z) - rD_r = -r^2$ to combine the terms $-ra_Q/2$ and $r(1-r)y^2/(2D_r)$. Completing the square in $\xi_Q'(r)$ gives (10.54). Part (d) follows from (10.54), since its last term is nonpositive and g is continuous and positive on $[\varepsilon, 1-\eta]$. □

The trial b_r is that of [90, Eq. (3.13)], and (10.54) is the computation in the proof of [90, Proposition 3.13], where the slack $g(r)$ is bounded below by the constant $6y\eta_3/z$ of [90, Eq. (3.5)].

Theorem 10.13 (Exclusion of high-energy points at positive overlap)**.** *Let ξ be a non-pure mixture, and let κ, C_{reg}, and η be the base constants* (10.47)–(10.48)*. Let $0 < \varepsilon < 1 - \eta$ and $G = G_\varepsilon$ as in Lemma 10.12(d). There exist an open interval U with $K_0 \in U$ and $\overline{U} \subset U_0$, a number $\delta \in (0, G/4)$, and $c_\varepsilon > 0$, N_ε, such that (U, δ) is admissible, $\delta \le \delta_U$ in Proposition 10.10, and the following holds with $\mathcal{K} = [E_0 - \delta, E_0 + \delta] \times \overline{U}$. For all $N \ge N_\varepsilon$, $u \in S_N$, and $(e,k) \in \mathcal{K}$, the event $\mathcal{A}_\varepsilon(u)$ on which*

(i) $\nabla^2_{S_N} H_Q(u) \preceq -\kappa I$ *and* $\mathcal{D}_N(H) \le C_{\mathrm{reg}}$*, and*

(ii) $\sup\{H_Q(w) : w \in S_N,\ \varepsilon \le R(u,w) \le 1-\eta\} \le N(E_0 - G/2)$,

satisfies $\mathbb{P}_{e,k,u}(\mathcal{A}_\varepsilon(u)) \ge 1 - Ce^{-c_\varepsilon N}$, where C depends only on ξ, β, and U_0. On $\mathcal{A}_\varepsilon(u)$, if u is a critical point:

(iii) there is no critical point $v \ne u$ with $R(u,v) \ge 1-\eta$;

(iv) there is no critical point $v \ne u$ with $H_Q(v) \ge N(E_0 - \delta)$ and $R(u,v) \ge \varepsilon$.

Parameter	Role	Depends on
U_0, κ, C_{reg}, η	base interval, Hessian margin, derivative bound, endpoint overlap	ξ, β
$\varepsilon \in (0, 1-\eta)$	overlap cutoff, chosen where the theorem is applied	free
$G = G_\varepsilon$	gap in the contact inequality on $[\varepsilon, 1-\eta]$	ε
$U \subset U_0$, $\delta < G/4$	radial interval and coarse energy width	ε
$\mathcal{T}_\varepsilon$, J_ε	finite net of latitudes and its size	ε, U, δ
c_ε, N_ε	failure exponent and dimension threshold	all of the above
a, b, t, L, M	fixed energy windows	fixed last, before $N \to \infty$

Table 10.1. Parameters of Theorem 10.13, in the order of their choice.

The same holds after replacing U by a smaller open interval containing K_0 and δ by a smaller number keeping (U, δ) admissible.

Table 10.1 lists the parameters of the theorem in the order in which they are chosen.

Proof. The proof has three steps. We first bound, for a single latitude, the conditional mean of the maximum of the residual field using Theorem 10.11. We then choose the neighborhoods and a finite net of latitudes, and use concentration at the net points. Finally, we interpolate between the net points on the event $\mathcal{D}_N(H) \le C_{\mathrm{reg}}$.

One latitude. Fix $r \in [\varepsilon, 1-\eta]$ and let $M_r = \max_{\mathrm{Lat}_r(u)} \widehat{H}_r$, with $\widehat{H}_r$ from Lemma 10.12(a). Identify $u^\perp$ with $\mathbb{R}^n$ and write $v = (N/n)^{1/2}\sigma$ with $\sigma \in S_n$, so that $v_1 \cdot v_2 / N = R_n(\sigma_1, \sigma_2)$. By Lemma 10.12(a),(b), the field $\widehat{H}_r$ has the law of $(N\tau_r(0))^{1/2}\gamma_0 + (N/n)^{1/2} H_n^{\bar\tau_r}(\sigma)$, with γ_0 a standard Gaussian variable independent of $H_n^{\bar\tau_r}$. The covariance $\bar\tau_r$ has nonnegative coefficients and b_r is positive, and Theorem 10.11 therefore applies with $\chi = \bar\tau_r$ and $b = b_r$. As $N^{-1}(N/n)^{1/2} = (n/N)^{1/2} n^{-1}$ and $N^{-1/2}\gamma_0 \to 0$ in probability, we obtain

$$\lim_{N\to\infty} \mathbb{P}\Big(\frac{M_r}{N} > \mathcal{Q}_{\bar\tau_r}(b_r) + t \Big) = 0, \qquad t > 0. \tag{10.55}$$

Next, $\widehat{H}_r(w)$ is the inner product of the residual coefficient vector with a fixed vector whose norm is the standard deviation of $\widehat{H}_r(w)$, namely $(N\tau_r(1))^{1/2} \le (N\xi(Q))^{1/2}$. (The inequality holds because a conditional variance is at most the variance.) Then M_r is a $(N\xi(Q))^{1/2}$-Lipschitz function of a standard Gaussian vector, and (10.11) implies that

$$\mathbb{P}\big(|M_r - \mathbb{E}M_r| > Nt \big) \le 2e^{-Nt^2/(2\xi(Q))}, \qquad t > 0. \tag{10.56}$$

If $N^{-1}\mathbb{E}M_r > \mathcal{Q}_{\bar\tau_r}(b_r) + 2t$ along a subsequence, then (10.56) would make the probability in (10.55) tend to one along it. We conclude that $\limsup_N N^{-1}\mathbb{E}M_r \le \mathcal{Q}_{\bar\tau_r}(b_r)$. The law of M_r depends neither on (e, k) nor on u, since the covariance (10.52) does not.

Neighborhoods and the net. The function ϖ is bounded on $[\varepsilon, 1-\eta]$. We choose an open interval U with $K_0 \in U$ and $\overline{U} \subset U_0$, and then $\delta \in (0, G/4)$, such that (U, δ) is admissible, $\delta \le \delta_U$, and

$$\sup_{r \in [\varepsilon, 1-\eta]} \bigl| \varpi(r) \cdot (e - E_0, k - K_0) \bigr| \le \frac{G}{8} \qquad \text{for } (e,k) \in \mathcal{K}. \tag{10.57}$$

By Lemma 10.12(c),(d), we then have $\mu_{e,k}(r) + \mathcal{Q}_{\bar\tau_r}(b_r) \le E_0 - 7G/8$ for $r \in [\varepsilon, 1-\eta]$ and $(e,k) \in \mathcal{K}$. Let $C_1 = C_{\mathrm{reg}}(1 - (1-\eta)^2)^{-1/2}$, and let $\mathcal{T}_\varepsilon \subset [\varepsilon, 1-\eta]$ be a finite set containing both endpoints with mesh at most $G/(8C_1)$, and let $J_\varepsilon = |\mathcal{T}_\varepsilon|$. By the first step, there exists N_ε such that $N^{-1}\mathbb{E} M_r \le \mathcal{Q}_{\bar\tau_r}(b_r) + G/8$ for all $r \in \mathcal{T}_\varepsilon$ and $N \ge N_\varepsilon$. Only finitely many latitudes enter here, and no uniformity of Theorem 10.11 in the covariance is needed. Using $H_Q = N\mu_{e,k}(r) + \widehat{H}_r$ on $\mathrm{Lat}_r(u)$, we obtain

$$\mathbb{E}_{e,k,u}\Bigl[\frac{1}{N} \max_{\mathrm{Lat}_r(u)} H_Q\Bigr] \le E_0 - \frac{3G}{4}, \qquad r \in \mathcal{T}_\varepsilon,\ N \ge N_\varepsilon.$$

Using (10.56) with $t = G/8$ and a union bound over $\mathcal{T}_\varepsilon$, we find

$$\mathbb{P}_{e,k,u}\Bigl(\exists r \in \mathcal{T}_\varepsilon : \frac{1}{N} \max_{\mathrm{Lat}_r(u)} H_Q > E_0 - \frac{5G}{8}\Bigr) \le 2J_\varepsilon e^{-b_\varepsilon N} \le e^{-b_\varepsilon N/2}, \qquad b_\varepsilon = \frac{G^2}{128\bar\xi(Q)}, \tag{10.58}$$

where the last inequality holds for $N \ge 2b_\varepsilon^{-1}\log(2J_\varepsilon)$. We enlarge N_ε accordingly. In this way the size of the net is absorbed into the threshold N_ε.

Interpolation and conclusion. For $v \perp u$ with $|v|^2 = N$, the curve $r \mapsto ru + (1-r^2)^{1/2}v$ stays on S_N and has speed $(N/(1-r^2))^{1/2}$. On the event $\mathcal{D}_N(H) \le C_{\mathrm{reg}}$, we have $|\nabla_{S_N} H_Q| \le C_{\mathrm{reg}} N^{1/2}$, which implies that $r \mapsto N^{-1}\max_{\mathrm{Lat}_r(u)} H_Q$ is C_1-Lipschitz on $[\varepsilon, 1-\eta]$. On the intersection of this event with the complement of the event in (10.58), each $r \in [\varepsilon, 1-\eta]$ is within $G/(8C_1)$ of a net point, so that $N^{-1}\max_{\mathrm{Lat}_r(u)} H_Q \le E_0 - 5G/8 + G/8 = E_0 - G/2$, which is (ii). For $(e,k) \in \mathcal{K} \subset \mathcal{K}_0$, the bounds (10.49) and (10.45) control the failure of (i) by $2e^{-c_H N} + C_{\mathrm{reg}} e^{-N}$. Altogether,

$$\mathbb{P}_{e,k,u}\bigl(\mathcal{A}_\varepsilon(u)^c\bigr) \le 2e^{-c_H N} + C_{\mathrm{reg}} e^{-N} + e^{-b_\varepsilon N/2} \le Ce^{-c_\varepsilon N}, \qquad c_\varepsilon = \min\{c_H, 1, b_\varepsilon/2\},$$

for $N \ge N_\varepsilon$, with $C = 3 + C_{\mathrm{reg}}$.

On $\mathcal{A}_\varepsilon(u)$, part (iii) follows from Lemma 10.9 with $C_3 = C_{\mathrm{reg}}$, by the choice (10.48) of η. For (iv), a critical point $v \ne u$ with $R(u,v) \ge \varepsilon$ has $R(u,v) \le 1-\eta$ by (iii). Then (ii) implies that $H_Q(v) \le N(E_0 - G/2) < N(E_0 - \delta)$, since $\delta < G/4$. Finally, every requirement on (U, δ) above is preserved when U is replaced by a smaller interval containing K_0 and δ by a smaller admissible value. □

Definition 10.14 (Retained maxima)**.** Let ε, U, δ, G, and $\mathcal{K}$ be as in Theorem 10.13. For a field H and a point $u \in S_N$, let $\mathcal{G}_\varepsilon(H,u) \in \{0,1\}$ be the indicator of the event $\mathcal{A}_\varepsilon(u)$ of Theorem 10.13. We call $\mathcal{A}_\varepsilon(u)$ the *selection event* at u. A point u is a *retained maximum* if $u \in \mathrm{Crt}_N^{\max}$, $(H_Q(u)/N, K_Q(u)/N) \in \mathcal{K}$, and $\mathcal{G}_\varepsilon(H,u) = 1$, that is, if u is a local maximum with marks in $\mathcal{K}$ on its selection event. We set

$$\lambda_N^{\mathrm{ret}}(h,k) = \mathfrak{c}_N e^{-Nq(h/N,k)/2} \int_{S_N} \mathbb{E}_{h/N,k,u}\big[D(u)\mathbf{1}\{\nabla^2_{S_N} H_Q(u) \prec 0\}\mathcal{G}_\varepsilon(H,u)\big]\nu_N(\,\mathrm{d}u). \tag{10.59}$$

The indicator $\mathcal{G}_\varepsilon$ is a bounded Borel function of the field and the point, since the suprema in its definition may be taken over countable dense sets. Lemma 10.1 implies that λ_N^{ret} is the density of the expected number of retained maxima on $\{(h,k) : (h/N,k) \in \mathcal{K}\}$.

Proposition 10.15 (Maxima discarded by the selection event)**.** *Let ξ be a non-pure mixture and let ε, U, δ, $\mathcal{K}$, and c_ε be as in Theorem 10.13.*

(a) For all large N, all $u \in S_N$, and all $(e,k) \in \mathcal{K}$,

$$\mathbb{E}_{e,k,u}\big[D(u)\big(1 - \mathcal{G}_\varepsilon(H,u)\big)\big] \le Ce^{-c_\varepsilon N/2} d_N^{\mathrm{all}}(k). \tag{10.60}$$

In particular, $0 \le 1 - \lambda_N^{\mathrm{ret}}(Ne,k)/\lambda_N^{\max}(Ne,k) \le Ce^{-c_\varepsilon N/2}$.

(b) For all fixed $a < b$ and $t \in \mathbb{R}$, the expected number of retained maxima u with $H_Q(u) - m_N \in [a,b]$ tends to $e^{-c_0 a} - e^{-c_0 b}$, and the expected number with $H_Q(u) \ge m_N + t$ tends to $e^{-c_0 t}$.

(c) Distinct retained maxima $u \ne v$ satisfy $R(u,v) < \varepsilon$.

(d) There exists $\delta' \in (0,\delta]$ such that, with the same ε and U and with δ' in place of δ, with probability at least $1 - Ce^{-cN}$ every critical point u with $H_Q(u) \ge N(E_0 - \delta')$ is a retained maximum.

Proof. For (a), fact (4) of Section 10.1 and Theorem 10.13 give

$$\mathbb{E}_{e,k,u}\big[D(u)(1-\mathcal{G}_\varepsilon)\big] \le \mathsf{K}_2^{1/2}\mathbb{P}_{e,k,u}(\mathcal{A}_\varepsilon(u)^c)^{1/2} d_N^{\mathrm{all}}(k) \le Ce^{-c_\varepsilon N/2} d_N^{\mathrm{all}}(k).$$

No independence between the selection event and the determinant is used. Since $\mathcal{G}_\varepsilon = 1$ implies $\nabla^2_{S_N} H_Q(u) \prec 0$, we have $\lambda_N^{\max} - \lambda_N^{\mathrm{ret}} = \mathfrak{c}_N e^{-Nq/2} \int \mathbb{E}_{e,k,u}[D(u)\mathbf{1}\{\nabla^2_{S_N} H_Q(u) \prec 0\}(1-\mathcal{G}_\varepsilon)]\nu_N(\,\mathrm{d}u)$. We divide by $\lambda_N^{\max}(Ne,k) = \mathfrak{c}_N e^{-Nq/2} d_N^{\max}(k)$ and use $d_N^{\mathrm{all}} \le 2 d_N^{\max}$ on $\overline{U}$, which holds for large N by (10.16). This gives the second statement.

For (b), we note that a fixed window $[m_N + a, m_N + b]$ lies in $[N(E_0 - \delta), N(E_0 + \delta)]$ for all large N, by (10.37). Integrating (a) over the window and over $k \in U$, the expected number of retained maxima differs from that of the points of $\mathrm{Crt}_N^{\max}$ with $K_Q/N \in U$ in the window by a factor $1 + O(e^{-c_\varepsilon N/2})$. The first limit then follows from Theorem 10.6(c). For the second, the retained maxima with $H_Q(u) \ge m_N + t$ are those with energy in

$[m_N + t, N(E_0 + \delta)]$, and the expected number of local maxima with $H_Q(u) \geq N(E_0 + \delta)$ is exponentially small by (10.25). The limit therefore equals that of $T_N(m_N + t)$, which is $e^{-c_0 t}$ by Theorem 10.6(b).

For (c), let $u \neq v$ be retained maxima. Then $\mathcal{A}_\varepsilon(u)$ holds, v is a critical point, and $H_Q(v) \geq N(E_0 - \delta)$. By Theorem 10.13(iv), we have $R(u, v) < \varepsilon$.

For (d), let $\delta' \leq \delta$ with $c_0 \delta' \leq c_\varepsilon/4$. The event $\mathcal{A}_\varepsilon$ does not involve δ, and by Theorem 10.13 the bound (a) holds on the smaller rectangle. By Lemma 10.1 with $f = 1 - \mathbf{1}\{\nabla^2_{S_N} H_Q(u) \prec 0\}\mathcal{G}_\varepsilon = 1 - \mathcal{G}_\varepsilon$, by (a), and by (10.26), the expected number of $u \in \mathrm{Crt}_N$ with marks in $[E_0 - \delta', E_0 + \delta'] \times U$ that are not retained is at most $Ce^{-c_\varepsilon N/2} \cdot CNe^{c_0 \delta' N} \leq CNe^{-c_\varepsilon N/4}$. The expected numbers of critical points with $H_Q(u) \geq N(E_0 - \delta')$ and $K_Q(u)/N \notin U$, or with $H_Q(u) > N(E_0 + \delta')$, are exponentially small by Lemma 10.4(b),(c). Markov's inequality completes the proof. □

10.7 Pure models

In this section, $\xi(r) = c_p r^p$ with $p \geq 4$, and (ξ, β) is strictly 1RSB with a nondegenerate contact at Q. Then $\xi_Q(r) = \xi(Q) r^p$, $a_Q = p\xi(Q)$, $b_Q = p(p-1)\xi(Q)$, and by Euler's identity $K_Q = pH_Q$. In particular, Σ_Q is singular and the marks satisfy $k = pe$. From (10.7), we have $E_0 = (p + z)\xi(Q)/y$ and $K_0 = ((1 + z)a_Q + b_Q)/y = pE_0$. We write

$$\phi(e) = \frac{1}{2} + \frac{1}{2} \log \frac{b_Q}{a_Q} - \frac{e^2}{2\xi(Q)} + \Omega\big(pe/b_Q^{1/2}\big), \tag{10.61}$$

for the complexity on the line $k = pe$.

Admissible sets in the pure case. Every statement of Sections 10.4 to 10.6 that involves the marks (e, k) is interpreted on the graph $\{(e, pe)\}$. A bounded open interval $U \ni K_0$ with $\overline{U} \subset (2b_Q^{1/2}, \infty)$ and a number $\delta > 0$ are admissible if $p[E_0 - \delta, E_0 + \delta] \subset U$ and $c'_- \leq -\phi'(e) \leq c'_+$ whenever $pe \in \overline{U}$, where $c'_- = (c_- + c_0)/2$ and $c'_+ = (c_0 + c_+)/2$. Such pairs exist because $\phi'(E_0) = -c_0$ (Proposition 10.16(a) below). The rectangle $\mathcal{K}$ is replaced by $\mathcal{K}_p = \{(e, pe) : |e - E_0| \leq \delta\}$. The conditional law $\mathbb{P}_{e,k,u}$ is replaced by $\mathbb{P}_{e,u}$. The covariance Σ_Q is singular, so q, Σ_Q^{-1}, $\mathfrak{c}_N$, the two-dimensional densities of Lemma 10.1, the radial integral $\mathcal{R}_N$, and the kernel (10.36) are not used. They are replaced by the one-dimensional intensities, the singular joint measures, and the one-dimensional Palm kernel of Proposition 10.16(a).

Proposition 10.16 (Pure specialization)**.** *Let $\xi(r) = c_p r^p$ with $p \geq 4$, and let (U, δ) be admissible in the sense just described.*

(a) Let $A_{N,p} = |S_N|(2\pi a_Q)^{-n/2}(2\pi N \xi(Q))^{-1/2}$ and, for $j \in \{\mathrm{all}, \mathrm{max}\}$,

$$\lambda^j_{N,p}(h) = \mathbf{1}_U(ph/N)\, A_{N,p}\, e^{-h^2/(2N\xi(Q))} d^j_N(ph/N). \tag{10.62}$$

For a Borel set $B \subset \mathbb{R}$, the expected numbers of points of Crt_N and of $\mathrm{Crt}_N^{\max}$ with $K_Q(u)/N \in U$ and $H_Q(u) \in B$ are $\int_B \lambda_{N,p}^{\mathrm{all}}(h)\,\mathrm{d}h$ and $\int_B \lambda_{N,p}^{\max}(h)\,\mathrm{d}h$; the corresponding expected-count measures on the pairs $(h,k) = (H_Q(u), K_Q(u)/N)$ are $\lambda_{N,p}^{j}(h)\,\mathrm{d}h\,\delta_{ph/N}(\,\mathrm{d}k)$. We write $\lambda_N = \lambda_{N,p}^{\max}$. For fixed $u \in S_N$ and $pe \in U$, the Palm kernel of local maxima at u with energy Ne is

$$\mathbb{E}_{N,Ne,u}^{\mathrm{Palm,max}} f(H,k) = \frac{\mathbb{E}_{e,u}\big[D(u)\mathbf{1}\{\nabla_{S_N}^2 H_Q(u) \prec 0\} f(H, pe)\big]}{d_N^{\max}(pe)}, \tag{10.63}$$

for nonnegative Borel f, and the expected sum of $f(H, K_Q(u)/N)$ over the points $u \in \mathrm{Crt}_N^{\max}$ with $K_Q(u)/N \in U$ and $H_Q(u) \in B$ is $\int_B \lambda_N(h) \int_{S_N} \mathbb{E}_{N,h,u}^{\mathrm{Palm,max}} f\,\nu_N(\,\mathrm{d}u)\,\mathrm{d}h$. Further, $\phi(E_0) = 0$, $\phi'(E_0) = -c_0$, $\phi(e) \le -c_0(e - E_0) - \frac{p-2}{4(p-1)\xi(Q)}(e - E_0)^2$ for all $e \in \mathbb{R}$, and $\lambda_{N,p}^{j}(Ne) \asymp N^{-1/2} e^{N\phi(e)}$ uniformly for $pe \in U$.

(b) For all large N, $(\log \lambda_N)'(h) = \phi'(h/N) + O(N^{-1})$, uniformly for $ph/N \in U$.

(c) All conclusions of Theorem 10.6 hold, with the same $c_0 = \beta x$, the same centering (10.37)*, and the same window limits.*

(d) Lemma 10.8 (with $\mathbb{P}_{e,u}$), Proposition 10.10, Theorem 10.13, Definition 10.14, and Proposition 10.15 hold with marks restricted to $\mathcal{K}_p$ and $\mathbb{P}_{e,k,u}$ replaced by $\mathbb{P}_{e,u}$, with their original neighborhood hypotheses and choices; for Theorem 10.13, the neighborhoods are shrunk for the prescribed ε as in its statement. In the pure form of Definition 10.14, the retained energy intensity is

$$\begin{aligned}\lambda_{N,p}^{\mathrm{ret}}(h) = \mathbf{1}\{|h/N - E_0| \le \delta\}\mathbf{1}_U(ph/N)\, A_{N,p}\, e^{-h^2/(2N\xi(Q))} \\ \times \int_{S_N} \mathbb{E}_{h/N,u}\big[D(u)\mathbf{1}\{\nabla_{S_N}^2 H_Q(u) \prec 0\}\mathcal{G}_\varepsilon(H,u)\big]\nu_N(\,\mathrm{d}u),\end{aligned} \tag{10.64}$$

with joint measure $\lambda_{N,p}^{\mathrm{ret}}(h)\,\mathrm{d}h\,\delta_{ph/N}(\,\mathrm{d}k)$, and in the pure form of Proposition 10.15(a) the intensity ratio is $\lambda_{N,p}^{\mathrm{ret}}(Ne)/\lambda_{N,p}^{\max}(Ne)$, for $|e - E_0| \le \delta$.

Proof. The proof follows the non-pure case. We describe the changes, which come from the singular covariance and from the energy dependence of the determinant.

The intensity and the complexity. Under $\mathbb{P}_{e,u}$, the Hessian has the law of $W_{n,N} - peI_n$. We obtain (10.62), the joint measures, and the kernel (10.63) from the pure form (7.25) of the marked Kac–Rice formula, since the density of $H_Q(u)$ is $(2\pi N\xi(Q))^{-1/2} e^{-h^2/(2N\xi(Q))}$ and $K_Q(u)/N = pH_Q(u)/N$. Next, ϕ is the function Θ_{pure} of Proposition 5.13(4), which gives $\phi(E_0) = 0$ and $\phi'(E_0) = -c_0$. (The proof is the computation in Lemma 10.3: since $pE_0 = K_0$ and $\Sigma_Q(c_0, 1/y)^{\mathsf{T}} = (E_0, K_0)^{\mathsf{T}}$ still holds, we have $E_0^2/\xi(Q) = c_0E_0 + K_0/y$ and

$\phi'(E_0) = -E_0/\xi(Q) + p/y = -c_0$.) Because $\Omega(t) - t^2/4$ is concave, $\phi(e) + \frac{1}{2}(\xi(Q)^{-1} - p^2/(2b_Q))e^2$ is concave, and

$$\frac{1}{\xi(Q)} - \frac{p^2}{2b_Q} = \frac{1}{\xi(Q)} - \frac{p}{2(p-1)\xi(Q)} = \frac{p-2}{2(p-1)\xi(Q)} > 0.$$

The tangent-plane argument at the end of the proof of Lemma 10.3 yields the quadratic bound on ϕ. By Stirling's formula (10.27) and (10.16) with $J = \overline{U}$, we have $\lambda^j_{N,p}(Ne) \asymp N^{-1/2} e^{N\phi(e)}$ for $pe \in U$, with constants uniform because (10.16) holds on $\overline{U}$. (At the endpoints of U, $\lambda^j_{N,p}$ vanishes, so the comparison cannot hold on $\overline{U}$.) The factor $N^{-1/2}$ now comes from the Gaussian density of the value, and there is no radial integral. This proves (a). The analogs of Lemma 10.4(b),(c) follow as before, with Lemma 10.2(b),(c) at $k = pe$: since $p(E_0 - \delta) \in U$, a critical point with $H_Q(u) \ge N(E_0 - \delta)$ and $pH_Q(u)/N \notin U$ has $H_Q(u) \ge N \sup U/p > N(E_0 + \delta)$. The excluded marks are covered by the bound for high energies.

The slope in energy. For $ph/N \in U$,

$$\frac{\mathrm{d}}{\mathrm{d}h} \log \lambda_N(h) = -\frac{h/N}{\xi(Q)} + \frac{p}{N} \frac{(d_N^{\max})'(ph/N)}{d_N^{\max}(ph/N)}, \tag{10.65}$$

so we must differentiate the determinant. Let $J = \overline{U}$, $W = W_{n,N}$, $\varrho = \frac{1}{2}\min_{k\in J}(k - 2b_Q^{1/2})$, and $Y_k = \sum_i \log\max\{|k - \lambda_i(W)|, \varrho\}$. Recall from Proposition 7.7 that $\mathbb{E}e^{2Y_k} \le C\, d_N^{\mathrm{all}}(k)^2$ for $k \in J$. Let $\mathcal{E} = \{\|W\|_{\mathrm{op}} \le 2b_Q^{1/2} + \varrho/2\}$, so that $\mathbb{P}(\mathcal{E}^c) \le 2e^{-N\varrho^2/(16b_Q)}$ by (10.12). On $\mathcal{E}$, we have $k - \lambda_i(W) > \varrho$ for all i. For fixed W, the function $k \mapsto \det(kI - W)\mathbf{1}\{\lambda_{\max}(W) < k\}$ is continuous, since the determinant vanishes at $k = \lambda_{\max}(W)$, and is smooth elsewhere with derivative $\partial_k \det(kI - W)\mathbf{1}\{\lambda_{\max}(W) < k\}$. By Fubini's theorem and dominated convergence,

$$(d_N^{\max})'(k) = \mathbb{E}\big[\partial_k \det(kI - W)\mathbf{1}\{\lambda_{\max}(W) < k\}\big], \qquad P_N'(k) = \mathbb{E}\,\partial_k \det(kI - W),$$

where $P_N(k) = \mathbb{E}\det(kI - W)$ is the signed expectation. Since $|\partial_k \det(kI - W)| = |\sum_i \prod_{j\ne i}(k - \lambda_j(W))| \le n\varrho^{-1}e^{Y_k}$ and $\{\lambda_{\max}(W) \ge k\} \subset \mathcal{E}^c$, the Cauchy–Schwarz inequality implies that

$$\big|P_N'(k) - (d_N^{\max})'(k)\big| \le n\varrho^{-1}\big(\mathbb{E}e^{2Y_k}\big)^{1/2}\mathbb{P}(\mathcal{E}^c)^{1/2} \le Cne^{-cN} d_N^{\mathrm{all}}(k),$$

and in the same way $|P_N(k) - d_N^{\max}(k)| \le Ce^{-cN} d_N^{\mathrm{all}}(k)$. With (10.16), we obtain

$$\big(\log d_N^{\max}\big)'(k) = \frac{P_N'(k)}{P_N(k)} + O(ne^{-cN}), \qquad k \in J. \tag{10.66}$$

We now compute P_N'/P_N. Recall from (7.13) that $P_n(t) = \mathbb{E}\det(tI - G_n) = n^{-n/2}\mathrm{He}_n(n^{1/2}t)$, and that its roots $\lambda_1, \dots, \lambda_n$ lie in $(-2, 2)$. Let $g_n(t) = P_n'(t)/(nP_n(t)) =$

$n^{-1}\sum_j (t-\lambda_j)^{-1}$ for $t>2$. In the proof of Proposition 7.9, the Riccati equation (7.14) is used to show that

$$0 \le g_{\mathrm{sc}}(t) - g_n(t) \le \frac{1}{n(t-2)^2(t^2-4)^{1/2}}, \qquad t>2, \tag{10.67}$$

where $g_{\mathrm{sc}}(t) = (t-(t^2-4)^{1/2})/2 = \Omega'(t)$. We have $P_N(k) = b_N^n P_n(k/b_N)$, and therefore $P_N'(k)/P_N(k) = (n/b_N) g_n(k/b_N)$. Inserting this into (10.66) and (10.65), and using (10.67), $n/N = 1-N^{-1}$, $b_N = b_Q^{1/2}(1+O(N^{-1}))$, and the Lipschitz continuity of g_{sc} on compact subsets of $(2,\infty)$, we obtain

$$\frac{\mathrm{d}}{\mathrm{d}h}\log\lambda_N(h) = -\frac{h/N}{\xi(Q)} + \frac{p}{b_Q^{1/2}}\Omega'\Big(\frac{ph/N}{b_Q^{1/2}}\Big) + O(N^{-1}) = \phi'(h/N) + O(N^{-1}),$$

uniformly for $ph/N \in U$. This proves (b). Since $c_- < c_-'$ and $c_+' < c_+$, the admissibility condition and (b) imply that $c_- \le -(\log\lambda_N)'(h) \le c_+$ on the support of λ_N for all large N.

Normalization. The proof of Theorem 10.6 applies with $N^{-1/2}e^{N\phi(e)}$ in place of the radial integral and with (b) in place of Lemma 10.5(e); there is no radial distribution. The density λ_N vanishes for $ph/N \notin U$. The envelope (10.41) therefore also holds when $h+v$ lies above the support, and the upper tail bound holds from m_N on. The tail T_N is constant below the support and strictly decreasing on it, and $T_N(NE_0 - D\log N) > 1 > T_N(NE_0 + D\log N)$. Then the root m_N exists, is unique, and lies in the interior of the support for all large N. This proves (c).

Localization and the exclusion estimate. Lemma 10.8 was proved for $\mathbb{P}_{e,u}$ as well. The Hessian bound (10.49) holds at $k = pe$, and the proofs of Proposition 10.10 and Proposition 10.15 apply verbatim, with Theorem 7.17 in its pure form and with the one-dimensional intensities (10.62) and (10.64) in place of the two-dimensional ones. For Theorem 10.13, we note that under $\mathbb{P}_{e,u}$ the latitude mean is $e\xi_Q(r)/\xi(Q)$ and the residual covariance is

$$\tau_r^{\mathrm{pure}}(t) = \xi_Q\big(r^2 + (1-r^2)t\big) - \frac{\xi_Q(r)^2}{\xi(Q)} - \frac{(1-r^2)\xi_Q'(r)^2}{a_Q}t.$$

Since $\xi_Q(r)^2 = \xi(Q)\xi_Q(r^2)$ and $\xi_Q'(r)^2 = a_Q\xi_Q'(r^2)$ for a pure covariance, the constant and linear coefficients of τ_r^{pure} vanish, and the higher coefficients are nonnegative. At $e = E_0$ the mean equals $(r\xi_Q'(r) + z\xi_Q(r))/y$, because $r\xi_Q'(r) = p\xi_Q(r)$ and $E_0/\xi(Q) = (p+z)/y$. The computation in Lemma 10.12(c) used only $\bar\tau_r'$ on $[t_r,1]$, the difference $\bar\tau_r(1) - \bar\tau_r(t_r)$, and the mean at (E_0,K_0), which are given by the same formulas. It follows that (10.54) holds for $\tau_r^{\mathrm{pure}} - \tau_r^{\mathrm{pure}}(0)$. The mark perturbation is $(e-E_0)\xi_Q(r)/\xi(Q)$, which replaces (10.57). With these changes, we may repeat the rest of the proof of Theorem 10.13, which uses no inverse of Σ_Q. This proves (d). □

10.8 Summary of one-point estimates

Corollary 10.17 (Summary of one-point estimates)**.** *Let ξ be a finite mixture with minimum degree $d \geq 4$, of either parity, pure or not, and let (ξ, β) be strictly 1RSB with a nondegenerate contact at Q. Let η, κ, and C_{reg} be the base constants* (10.47)–(10.48)*, and fix $\varepsilon \in (0, 1-\eta)$. Then one can choose, in this order and before N, a bounded open interval $U \ni K_0$ with $\overline{U} \subset U_0$ and a number $\delta > 0$, independent of any fixed energy window, such that, with m_N defined by Theorem 10.6 (the expected number of nondegenerate local maxima with $K_Q/N \in U$ and $H_Q \geq m_N$ equals one), the following hold. In the pure case all marks lie on the graph $k = pe$, and the conditional law is $\mathbb{P}_{e,u}$.*

(i) One-point density. *For a non-pure mixture, the joint densities are given by Lemma 10.1, with $\lambda_N^j(Ne, k) \asymp e^{N\Theta(e,k)}$ for $k \in \overline{U}$; the normalization integral $\mathcal{R}_N$ of* (10.32) *satisfies Lemma 10.5(c); and $\lambda_N(Ne) = \mathfrak{c}_N \mathcal{R}_N(e) \asymp N^{-1/2} e^{N\theta(e)}$ for $|e - E_0| \leq \delta$. For a pure model, the expected-count measures are $\lambda_{N,p}^j(h)\, \mathrm{d}h\, \delta_{ph/N}(\,\mathrm{d}k)$ of Proposition 10.16(a), with $\lambda_{N,p}^j(Ne) \asymp N^{-1/2} e^{N\phi(e)}$ for $pe \in U$. For a non-pure mixture the Palm kernel at a fixed point is* (10.36)*; for a pure model it is the one-dimensional kernel* (10.63)*.*

(ii) Normalization. *$m_N = NE_0 - \log N/(2c_0) + O(1)$ with $c_0 = \beta x$. For all fixed $a < b$ and $t \in \mathbb{R}$,*

$$\mathbb{E}\#\big\{u \in \mathrm{Crt}_N : K_Q(u)/N \in U,\ H_Q(u) - m_N \in [a,b]\big\} \longrightarrow e^{-c_0 a} - e^{-c_0 b},$$
$$\mathbb{E}\#\big\{u \in \mathrm{Crt}_N : K_Q(u)/N \in U,\ H_Q(u) \geq m_N + t\big\} \longrightarrow e^{-c_0 t},$$

and the same limits hold for nondegenerate local maxima. The density satisfies $\lambda_N(m_N - v) \leq Ce^{c_+ v}$ for $0 \leq v \leq m_N - N(E_0 - \delta)$, with $c_+ < \beta$, and $T_N(m_N + v) \leq e^{-c_- v}$ for $v \geq 0$. In particular, for every fixed M,

$$\sup_N \mathbb{E}|\mathcal{C}_{N,M}| < \infty, \quad \text{where} \quad \mathcal{C}_{N,M} = \Big\{u \in \mathrm{Crt}_N : |H_Q(u) - m_N| \leq M,\ \frac{K_Q(u)}{N} \in U\Big\}. \tag{10.68}$$

(iii) Localization. *With probability at least $1 - Ce^{-cN}$, every critical point u with $H_Q(u) \geq N(E_0 - \delta)$ satisfies $K_Q(u)/N \in U$ and $\nabla^2_{S_N} H_Q(u) \preceq -\kappa I$, and $\sup_{S_N} \|\nabla^3_{S_N} H_Q\|_{\mathrm{op}} \leq C_{\mathrm{reg}} N^{-1/2}$ (Proposition 10.10).*

(iv) Upper tail of the maximum. *$\limsup_N \mathbb{P}(\max_{S_N} H_Q > m_N + L) \leq e^{-c_0 L}$ for $L \geq 0$, and with probability at least $1 - Ce^{-cN}$ every global maximizer u with $H_Q(u) \geq N(E_0 - \delta)$ is a nondegenerate local maximum with $K_Q(u)/N \in U$.*

(v) Selection event. *The indicator $\mathcal{G}_\varepsilon(H, u)$ of the selection event (Definition 10.14) is a bounded Borel function of the field and the point. Retained maxima are nondegenerate local maxima with marks in $\mathcal{K} = [E_0 - \delta, E_0 + \delta] \times \overline{U}$ (pure: $\mathcal{K}_p$); distinct retained maxima $u \neq v$ satisfy $R(u,v) < \varepsilon$; and, uniformly for $u \in S_N$ and $(e,k) \in \mathcal{K}$,*

$$\mathbb{E}_{e,k,u}\big[D(u)\big(1 - \mathcal{G}_\varepsilon(H,u)\big)\big] \leq Ce^{-cN}\, \mathbb{E}_{e,k,u} D(u),$$

where, in a pure model, $\mathbb{E}_{e,k,u}$ is replaced by $\mathbb{E}_{e,u}$ and $(e,k) \in \mathcal{K}_p$. In the notation of Section 7, $\mathbb{E}_{e,k,u}$ is the conditioned expectation $\mathbb{E}_{h,k,u}$ with normalized energy $h = e$. The expected numbers of retained maxima in fixed windows and above $m_N + t$ have the limits in (ii). No statement is made about pairs at negative overlap.

The constants c, C depend on ξ, β, U_0, and ε; every window is fixed before $N \to \infty$.

Proof. For a non-pure mixture, choose U and δ as in Theorem 10.13 for the given ε, so that (U, δ) is admissible and $\delta \le \delta_U$. If the simultaneous retention is wanted, we then decrease δ as in Proposition 10.15(d). Each requirement is a strict inequality at (E_0, K_0) or a condition preserved under shrinking, and a single choice satisfies all of them. Part (i) is Lemma 10.1, Lemma 10.4(a), and Lemma 10.5. Part (ii) is Theorem 10.6, and (10.68) follows from the window limit for all critical points and the finiteness of the expected counts for each N. Part (iii) is Proposition 10.10, (iv) is Theorem 10.6(f), and (v) is Definition 10.14 and Proposition 10.15, where in (a) of the latter $d_N^{\mathrm{all}}(k) = \mathbb{E}_{e,k,u} D(u)$. For a pure covariance, use Proposition 10.16, with the one-dimensional intensities and kernel of its part (a) and the retained intensity (10.64). □

Remark 10.18 (Counting conventions and numerators without the selection indicator). In an even model one may count antipodal pairs once. Then the expected tail is halved, and if m_N^{proj} denotes the centering for this convention, (10.43) gives $m_N^{\mathrm{proj}} - m_N = -(\log 2)/c_0 + o(1)$. This deterministic shift must accompany any change of convention. We use ordinary counts throughout. In upper bounds for sums of a nonnegative weight over critical points, the selection indicator may be dropped from the numerator, and the Palm kernel without the selection indicator, given by (10.36) for non-pure mixtures and by (10.63) for pure models, remains the natural normalization. Proposition 10.15 compares intensities. It is not a statement about determinant moments under the conditional law restricted to the selection event.

SECTION 11

Mixed models: pair estimates and counts in fixed energy windows

In this section, we prove Theorem 11.14, which supplies the critical points of the center field H_Q used in Sections 12 and 13 for every covariance that satisfies the 1RSB assumptions of Definition 5.9. It has two parts. The first is that a fixed energy window around the centering m_N contains a bounded expected number of critical points of H_Q with normalized radial derivative near K_0. The second is that, when the window is large, with probability close to one the window contains at least one such point and the maximum of H_Q does not exceed its upper end. The first assertion follows from the one-point estimates of Section 10. The second requires an upper bound for the second moment of the number of retained maxima in the window, which is Theorem 11.13. Its proof is the main task of this section.

We split the ordered pairs of retained maxima according to their overlap r. Pairs with $r \geq \delta$ do not occur, by the separation property of the selection event of Section 10 (Theorem 10.13). Pairs with $r \leq -\delta$ are controlled deterministically by an elementary bound for spherical codes in a cap (Lemma 11.12): every retained point has overlap at most $-\delta$ with at most a fixed number of other retained points. For $|r| < \delta$, we discard the selection indicator and use the two-point Kac–Rice formula. Its integrand factorizes when $|r| \leq N^{-\alpha}$ with $1/4 < \alpha < 1/2$ (Theorem 11.10). The remaining range is suppressed by the density $f_N(r)$ of the overlap of two independent uniform points, which satisfies $f_N(r) \leq \sqrt{N}\exp(-(N-3)r^2/2)$ for $N \geq 3$. The factorization requires control of three ratios: the density of the two gradients at zero, the density of the two pairs of value and radial derivative, and the expected product of the two Hessian determinants. The first two are explicit Gaussian computations (Section 11.2). For the determinants (Section 11.3), we write each Hessian as a bulk block bordered by one row and a corner entry. Its determinant is then the bulk determinant times a Schur complement that concentrates around a deterministic value. We compare the two bulk determinants using a maximal correlation inequality for Gaussian vectors, whose constant does not depend on the dimension. No parity assumption is made, and all counts are ordinary: in an even model, both points of an antipodal pair are counted.

The joint Gaussian law of values, radial derivatives, gradients, and Hessians at two points of the sphere was computed by Ben Arous, Subag, and Zeitouni [30, Lemmas 13–15

and (4.6)–(4.10); Lemma 54 and (A.2)–(A.8)]. We derive the formulas we need directly from the covariance. Here the marks are of order N, both signs of the overlap occur, and we need the exact size of the conditional mean of the Hessian. The second moment method for critical points of spherical spin glasses goes back to Subag [140]. For general mixed models without a constant or linear component, Belius and Schmidt [22, Theorem 1.1] proved that the second moment of the number of local maxima with a prescribed radial derivative matches the square of its first moment on the exponential scale. Their count has no energy constraint and uses shrinking windows for the normalized radial derivative. The count here fixes an unnormalized energy window of bounded width and also restricts the radial derivative. For pure models, the extremal process of the critical values was identified by Subag and Zeitouni [144]. Here we prove only an upper bound for the second moment of the number of critical points in a fixed energy window, which is all that Theorem 11.14 requires.

The ground-state energy E_0 and the radial derivative K_0 come from Proposition 5.11. From Section 7, we use the conditioned laws (Definition 7.3 and Lemma 7.4), the Gaussian concentration and edge estimates (Lemma 7.5), the determinant moments of Proposition 7.7, the bordered determinant bound of Lemma 7.11, the two-point conditioned law of Section 7.4, and the Kac–Rice formulas Proposition 7.16, Theorem 7.17, and Corollary 7.18. From Section 10, we use the consequences of Corollary 10.17 listed in Section 11.1. Theorem 11.14 is used in Section 12 for the concentration of the Gibbs measure on the bands (Theorem 12.22) and in Section 13 for the proof of Theorem 1.2.

11.1 Setting, inputs, and parameters

Standing assumptions. Throughout this section, ξ is a finite mixture with minimum degree $d \geq 4$, and $\beta > 0$ is such that (ξ, β) is strictly 1RSB with a nondegenerate contact at Q (Definition 5.9), with 1RSB parameters $x, Q \in (0,1)$. The number of supported degrees and their parities are arbitrary. As in Definition 3.1, ξ is *pure* if exactly one coefficient is positive and *non-pure* otherwise. We write $\xi_Q(r) = \xi(Qr) = \sum_p c_p Q^p r^p$, so that $\xi_Q(1) = \xi(Q)$, $\xi_Q'(1) = a_Q$, and $\xi_Q''(1) = b_Q$. Because every supported degree is at least d, we have, for $0 \leq j \leq 4$ and $|r| \leq 1$,

$$\left|\xi_Q^{(j)}(r)\right| \leq C_j\,|r|^{d-j}, \qquad C_j = \sum_p c_p Q^p\, p(p-1)\cdots(p-j+1). \tag{11.1}$$

The center field and its marks. We use the polynomial extension of H_N to $\mathbb{R}^N$ (Section 7.1), the center field $H_Q(u) = H_N(\sqrt{Q}\,u)$, and the radial derivative $K_Q(u) = u \cdot \nabla H_Q(u)$, for $u \in \mathbb{R}^N$. For all $u, v \in \mathbb{R}^N$,

$$\mathbb{E} H_Q(u) H_Q(v) = N \xi_Q(R(u,v)), \tag{11.2}$$

and every covariance used below is obtained by differentiating the kernel $N\xi_Q(u\cdot v/N)$. Here ∇, ∇^j, and ∂_v denote Euclidean derivatives, while ∇_{S_N} and $\nabla^2_{S_N}$ are the spherical gradient and Hessian. At a critical point $u \in S_N$, the spherical Hessian is the compression of $\nabla^2 H_Q(u)$ to $u^\perp$ minus $(K_Q(u)/N)\mathrm{Id}$, by (7.3). A critical point u carries the normalized marks $h = H_Q(u)/N$ and $k = K_Q(u)/N$ (in Section 10, the normalized energy is denoted e, and h is the unnormalized energy). We write $D(u) = \left|\det \nabla^2_{S_N} H_Q(u)\right|$. All conditional laws are the explicit Gaussian regression laws of Section 7: $\mathbb{E}_{h,k,u}$ denotes the conditioned law of Definition 7.3 at u with marks (h,k), and $\mathbb{E}_{\mathrm{pair}}$ denotes the two-point conditioned law of Section 7.4 at a pair (u_1,u_2) with marks (h_1,k_1,h_2,k_2). These laws are defined at every value of the marks, and they describe the conditional law of the whole field.

For a non-pure mixture, the covariance matrix Σ_Q of $(H_Q(u),K_Q(u))/\sqrt{N}$ is positive definite (Lemma 7.1). In a pure model of degree p, we have $K_Q = pH_Q$ by Euler's identity, the matrix Σ_Q is singular, and the energy is the only mark; see Remark 11.3. In a pure model, every mark domain below is restricted to the compatible graph $k = ph$.

Notation. For a positive definite matrix Γ, we write p_Γ for the centered Gaussian density with covariance Γ, and $p_N(h,k) = N\,p_{\Sigma_Q}(\sqrt{N}h,\sqrt{N}k)$ for the density of the normalized marks at a point, as in (7.5). We write $\|X\|_q = (\mathbb{E}\,|X|^q)^{1/q}$ and $\|A\|$ for the operator norm of a matrix or multilinear form. As in Section 7, $W_{m,N} = \sqrt{b_Q m/N}\,G_m$, whose entries have variance b_Q/N off the diagonal and $2b_Q/N$ on it, and $\mu_{m,N}(k) = \mathbb{E}\,|\det(W_{m,N} - k\,\mathrm{Id}_m)|$. Throughout, C and c denote positive constants that may change from line to line. They depend only on ξ, β, the 1RSB parameters, the rectangle $\mathcal{K}_0$, and the gap κ fixed below, and on the moment order q, the exponent α, and the overlap cutoff ε of the selection event when these appear. They do not depend on N, on the points u_1,u_2, on the overlap r, on the marks in $\mathcal{K}_0$, or on the window half-width L, unless this is indicated.

Inputs from Section 5. By Proposition 5.11, the ground-state energy E_0, the radial derivative K_0, and the slope $c_0 = \beta x \in (0,\beta)$ are defined, and $K_0 > 2\sqrt{b_Q}$. We fix

$$\kappa = \tfrac{1}{2}\big(K_0 - 2\sqrt{b_Q}\big), \qquad I = \big[2\sqrt{b_Q}+\kappa,\, K_0+1\big], \qquad \mathcal{K}_0 = [E_0-1, E_0+1]\times I. \quad (11.3)$$

Every value $k \in I$ of the normalized radial derivative lies at distance at least κ above the edge $2\sqrt{b_Q}$ of the spectrum of $W_{N-1,N}$. All local estimates in Sections 11.2 to 11.4 are uniform over marks $(h_i,k_i) \in \mathcal{K}_0$.

Inputs from Section 10. We use the following parts of Corollary 10.17, which holds for pure and non-pure covariances.

(M1) *Selection event* (Corollary 10.17(v), from Theorem 10.13, Definition 10.14, and Proposition 10.15). Let $\eta > 0$ be the endpoint overlap of Corollary 10.17. For every $\varepsilon \in (0, 1-\eta)$, there exist an open interval U with $K_0 \in U$ and $\overline{U} \subset I$, a width

$w_\varepsilon \in (0,1]$, constants $C, c > 0$, and a bounded Borel function $G_\varepsilon(H,u) \in \{0,1\}$ of the field and of a point $u \in S_N$ (the indicator $\mathcal{G}_\varepsilon$ of the selection event in Definition 10.14), with the following properties. We write $\mathcal{K}_\varepsilon = [E_0 - w_\varepsilon, E_0 + w_\varepsilon] \times \overline{U}$; the interval U and the width w_ε (denoted δ in Section 10) may be taken as small as desired. A critical point u with marks in $\mathcal{K}_\varepsilon$ and $G_\varepsilon(H,u) = 1$ is called *retained*; it is a nondegenerate local maximum of H_Q on S_N.

(i) If $u \neq v$ are retained critical points, then $R(u,v) < \varepsilon$.

(ii) Uniformly in $u \in S_N$ and $(h,k) \in \mathcal{K}_\varepsilon$,

$$\mathbb{E}_{h,k,u}\big[D(u)\big(1 - G_\varepsilon(H,u)\big)\big] \le Ce^{-cN}\,\mathbb{E}_{h,k,u}D(u). \tag{11.4}$$

(M2) *Centering and expected counts in fixed windows* (Corollary 10.17(ii), from Theorem 10.6). The centering m_N, defined from U, satisfies $m_N = NE_0 - \log N/(2c_0) + O(1)$, and for all fixed $a < b$,

$$\mathbb{E}\#\big\{u \in S_N : \nabla_{S_N} H_Q(u) = 0,\ H_Q(u) - m_N \in [a,b],\ K_Q(u)/N \in U\big\} \to e^{-c_0 a} - e^{-c_0 b}. \tag{11.5}$$

(M3) *Upper tail of the maximum* (Corollary 10.17(iv)). For every $L \ge 0$,

$$\limsup_{N\to\infty} \mathbb{P}\Big(\max_{S_N} H_Q > m_N + L\Big) \le e^{-c_0 L}. \tag{11.6}$$

Note that the denominator in (11.4) is $\mathbb{E}_{h,k,u}D(u) = \mu_{N-1,N}(k)$, the expected absolute determinant over all critical points; in Section 10 it is denoted $d_N^{\mathrm{all}}(k)$. We also note that the expected counts (11.5) include all critical points with normalized radial derivative in U, not only local maxima.

The selection event is built in Section 10 from the zero-temperature contact inequality for the field H_Q, that is, from the strict 1RSB property of its ground state (Proposition 5.11(3)). Given this property, [90, Lemma 3.5] gives the formulas for the ground-state parameters and the strict inequality (3.4), on which the conditional ground-state trial of [90, Lemma 3.11 and Proposition 3.13] rests. Each condition defining G_ε is invariant under rotations, so that $G_\varepsilon(H \circ O^{-1}, Ou) = G_\varepsilon(H,u)$ for $O \in O(N)$. The arguments below discard the selection indicator in every pair integrand and use neither this invariance nor any two-point property of G_ε.

Parameters. Table 11.1 lists the parameters of this section in the order in which they are chosen. We fix all of them before $N \to \infty$, and send the window half-width L to infinity only after $N \to \infty$, in the last step of the proof of Theorem 11.14. Narrowing U or decreasing w_ε preserves every local estimate, because those estimates are uniform on the larger rectangle $\mathcal{K}_0$.

Parameter	Role	Depends on
ξ, β, x, Q	covariance, temperature, and 1RSB parameters	fixed
E_0, K_0, c_0	ground-state parameters and slope (Proposition 5.11)	ξ, β
$\kappa,\ I,\ \mathcal{K}_0$	edge gap and initial mark rectangle (11.3)	ξ, β
δ_0	overlap range of the Gaussian estimates (Section 11.2)	$\xi, Q, \mathcal{K}_0$
$\delta_1 \le \delta_0$	overlap range of Lemma 11.8	$\xi, Q, \mathcal{K}_0$
$\delta \le \delta_0$	local pair cutoff (Theorem 11.10)	$\xi, Q, \mathcal{K}_0$
ε	separation of retained maxima, $0 < \varepsilon < \min\{\delta^2/2, 1-\eta\}$; fixed before Corollary 10.17 is applied	δ, η
$U,\ w_\varepsilon,\ G_\varepsilon,\ m_N$	radial interval, width, selection indicator, and centering in (M1)–(M2), with $\mathcal{K}_\varepsilon \subset \mathcal{K}_0$	ε
$\alpha \in (1/4, 1/2)$	local pair exponent	free
L	window half-width	fixed as $N \to \infty$

Table 11.1. The parameters of Section 11 and the order in which they are chosen.

11.2 The two-point Gaussian structure

For the local two-point and bulk determinant estimates below, we assume $N \ge 3$. Fix $u_1, u_2 \in S_N$ with overlap $r = R(u_1, u_2) \in (-1, 1)$, and let $s_r = \sqrt{1 - r^2}$ and $n = N - 2$. Let $e_1, \dots, e_n$ be an orthonormal basis of the orthogonal complement of $\{u_1, u_2\}$, and let

$$t_1 = \frac{u_2 - r u_1}{\sqrt{N}\, s_r}, \qquad t_2 = \frac{u_1 - r u_2}{\sqrt{N}\, s_r}. \tag{11.7}$$

Then $(e_1, \dots, e_n, t_i)$ is an orthonormal basis of $u_i^{\perp}$, the vector t_i points from u_i toward the other point, and

$$t_1 \cdot u_2 = u_1 \cdot t_2 = \sqrt{N}\, s_r, \qquad t_1 \cdot t_2 = -r. \tag{11.8}$$

The direction t_2 is the negative of the direction used at u_2 in the proof of Lemma 7.13 and in [30]; this changes the signs of some off-diagonal covariances and no density or determinant. We use the in-plane gradients $g_i = \partial_{t_i} H_Q(u_i)$, the vectors $Z_i = (H_Q(u_i), K_Q(u_i))/\sqrt{N}$ and $Z = (Z_1, Z_2) \in \mathbb{R}^4$, and the following deterministic

functions of r:

$$c(r) = s_r^2 \xi_Q''(r) - r\xi_Q'(r), \qquad \omega(r) = s_r^2 \xi_Q'''(r) - 2r\xi_Q''(r), \qquad \rho(r) = \frac{\xi_Q''(r)}{b_Q},$$
$$\chi(r) = s_r \begin{pmatrix} \xi_Q'(r) \\ \xi_Q'(r) + r\xi_Q''(r) \end{pmatrix}, \qquad \Sigma_Q(r) = \begin{pmatrix} \xi_Q(r) & r\xi_Q'(r) \\ r\xi_Q'(r) & r\xi_Q'(r) + r^2\xi_Q''(r) \end{pmatrix}, \tag{11.9}$$
$$\gamma(r) = \frac{a_Q}{a_Q^2 - c(r)^2}, \qquad \tau(r) = -\frac{c(r)}{a_Q^2 - c(r)^2}, \qquad \theta(r) = 1 - \frac{a_Q\, s_r^2\, \xi_Q''(r)^2}{b_Q\big(a_Q^2 - \xi_Q'(r)^2\big)}.$$

Note that $\Sigma_Q(1) = \Sigma_Q$. We define the 4×4 matrix

$$M(r) = \begin{pmatrix} \Sigma_Q - \gamma(r)\chi(r)\chi(r)^\mathsf{T} & \Sigma_Q(r) - \tau(r)\chi(r)\chi(r)^\mathsf{T} \\ \Sigma_Q(r) - \tau(r)\chi(r)\chi(r)^\mathsf{T} & \Sigma_Q - \gamma(r)\chi(r)\chi(r)^\mathsf{T} \end{pmatrix} \tag{11.10}$$

and the row vectors $\ell_1(r), \ell_2(r) \in \mathbb{R}^4$,

$$\ell_1(r) = \Big(-s_r\omega(r)\gamma(r)\chi(r)^\mathsf{T},\ s_r^2\big(\xi_Q''(r),\, 2\xi_Q''(r) + r\xi_Q'''(r)\big) - s_r\omega(r)\tau(r)\chi(r)^\mathsf{T}\Big), \tag{11.11}$$

with $\ell_2(r)$ obtained from $\ell_1(r)$ by exchanging its first two and its last two coordinates. Using (11.1), we see that $\xi_Q'(r) = O(|r|^{d-1})$ and $c(r) = O(|r|^{d-2})$, and $M(r) \to \mathrm{diag}(\Sigma_Q, \Sigma_Q)$ as $r \to 0$. We fix $\delta_0 \in (0, 1/2]$ such that, for $|r| \le \delta_0$, we have $\left|\xi_Q'(r)\right| \le a_Q/2$ and $|c(r)| \le a_Q/2$, and, for a non-pure mixture, $\frac{1}{2}\lambda_{\min}(\Sigma_Q)\mathrm{Id} \le M(r) \le 2\left\|\Sigma_Q\right\| \mathrm{Id}$, and such that δ_0 is at most the constant ρ_0 of Lemma 7.13, applied with $g = \kappa$ and $\mathcal{K}_2 = \mathcal{K}_0 \times \mathcal{K}_0$ for a non-pure mixture, and with $\mathcal{K}_2 = (\mathcal{K}_0 \times \mathcal{K}_0) \cap \{k_1 = ph_1,\ k_2 = ph_2\}$ in a pure model. With this choice, the two-point conditioned law is defined for $|r| \le \delta_0$.

Lemma 11.1 (Two-point Gaussian structure)**.** *Let ξ be a non-pure mixture and $|r| \le \delta_0$.*

(a) The transverse gradient pairs $(\partial_{e_j} H_Q(u_1), \partial_{e_j} H_Q(u_2))$, $1 \le j \le n$, are independent of each other and of (Z, g_1, g_2), and have covariance $\left(\begin{smallmatrix} a_Q & \xi_Q'(r) \\ \xi_Q'(r) & a_Q \end{smallmatrix}\right)$. The pair (g_1, g_2) has covariance $\left(\begin{smallmatrix} a_Q & c(r) \\ c(r) & a_Q \end{smallmatrix}\right)$, and

$$\mathrm{Cov}(Z_1, g_1) = \mathrm{Cov}(Z_2, g_2) = 0, \qquad \mathrm{Cov}(Z_2, g_1) = \mathrm{Cov}(Z_1, g_2) = \chi(r). \tag{11.12}$$

Conditionally on $\nabla_{S_N} H_Q(u_1) = \nabla_{S_N} H_Q(u_2) = 0$, the vector Z is centered Gaussian with covariance $M(r)$.

(b) Let $(h_i, k_i) \in \mathcal{K}_0$ and consider the two-point conditioned law $\mathbb{E}_{\mathrm{pair}}$ with these marks. In the frame $(e_1, \dots, e_n, t_i)$,

$$-\nabla^2_{S_N} H_Q(u_i) = \begin{pmatrix} k_i \mathrm{Id}_n - W^{(i)} & -y_i \\ -y_i^\mathsf{T} & k_i - z_i \end{pmatrix}, \tag{11.13}$$

where $W^{(i)} = (\partial_{e_j}\partial_{e_l} H_Q(u_i))_{j,l\le n}$, $y_i = (\partial_{e_j}\partial_{t_i} H_Q(u_i))_{j\le n}$, *and* $z_i = \partial_{t_i}\partial_{t_i} H_Q(u_i)$. *Under* $\mathbb{E}_{\mathrm{pair}}$, *the following hold.*

(i) The pair $(W^{(1)}, W^{(2)})$ *is independent of* (y_1, y_2, z_1, z_2), *and each* $W^{(i)}$ *has the law of* $W_{n,N}$. *The standardized entries of* $W^{(1)}$ *and* $W^{(2)}$ *in the same position have covariance* $\rho(r)$, *and entries in different positions are uncorrelated.*

(ii) The pair (y_1, y_2) *is centered and independent of* (z_1, z_2), *with* $\mathrm{Cov}(y_i) = (b_Q\theta(r)/N)\,\mathrm{Id}_n$. *The coordinate pairs* $(y_{1,j}, y_{2,j})$, $1 \le j \le n$, *are independent.*

(iii) The corners satisfy $\mathbb{E}z_i = \ell_i(r)M(r)^{-1}(h_1, k_1, h_2, k_2)^{\mathsf{T}}$ *and* $\mathrm{Var}(z_i) \le 2b_Q/N$. *In particular,* $|\mathbb{E}z_i| \le C\,|r|^{d-2}$.

The corners z_1 and z_2 need not be asymptotically uncorrelated at $r = 0$: their unconditional covariance at $r = 0$ is $\xi_Q''''(0)/N$, which is nonzero when $d = 4$. Each corner has variance $O(1/N)$, and this correlation does not affect the arguments below.

Proof. All covariances are derivatives of the kernel $N\xi_Q(x \cdot y/N)$ in fixed directions v, v', w, w' of $\mathbb{R}^N$. Writing $R = R(x, y)$, we have

$$\begin{aligned} \mathrm{Cov}(H_Q(x), \partial_w H_Q(y)) &= \xi_Q'(R)\,(x \cdot w), \\ \mathrm{Cov}(\partial_v H_Q(x), \partial_w H_Q(y)) &= \xi_Q'(R)\,(v \cdot w) + \frac{\xi_Q''(R)}{N}(v \cdot y)(x \cdot w), \end{aligned} \tag{11.14}$$

and $\mathrm{Cov}(\partial_v\partial_{v'} H_Q(x), \partial_w\partial_{w'} H_Q(y))$ equals

$$\begin{aligned} &\frac{\xi_Q''(R)}{N}\big[(v \cdot w)(v' \cdot w') + (v \cdot w')(v' \cdot w)\big] + \frac{\xi_Q''''(R)}{N^3}(v \cdot y)(v' \cdot y)(x \cdot w)(x \cdot w') \\ &\quad + \frac{\xi_Q'''(R)}{N^2}(v \cdot y)\big[(v' \cdot w)(x \cdot w') + (v' \cdot w')(x \cdot w)\big] \\ &\quad + \frac{\xi_Q'''(R)}{N^2}(v' \cdot y)\big[(v \cdot w)(x \cdot w') + (v \cdot w')(x \cdot w)\big]. \end{aligned} \tag{11.15}$$

Covariances involving $K_Q(y)$ are obtained by applying $\frac{\mathrm{d}}{\mathrm{d}s}|_{s=1}$ to the kernel at (x, sy). At a single point, the law of the value and radial derivative, the tangent gradient, and the compressed Euclidean Hessian is given by Lemma 7.1: these three groups are independent, with covariances $N\Sigma_Q$, $a_Q\mathrm{Id}$, and that of $W_{N-1,N}$. We now compute the cross covariances between the two points, using (11.8) and the fact that each e_j is orthogonal to u_1, u_2, t_1, t_2.

Gradients. By (11.14), $\mathrm{Cov}(\partial_{e_j} H_Q(u_1), \partial_{e_l} H_Q(u_2)) = \xi_Q'(r)\delta_{jl}$ and $\mathrm{Cov}(g_1, g_2) = \xi_Q'(r)(t_1 \cdot t_2) + \xi_Q''(r)(t_1 \cdot u_2)(u_1 \cdot t_2)/N = c(r)$. A transverse gradient has zero covariance with g_1, g_2 and with the values and radial derivatives at both points, since every term in (11.14) and its radial analog contains a factor $e_j \cdot u_i$ or $e_j \cdot t_i$. Next, $\mathrm{Cov}(H_Q(u_2), g_1) = \xi_Q'(r)(u_2 \cdot t_1) = \sqrt{N}s_r\xi_Q'(r)$, and differentiating $(x \cdot u_2)\xi_Q'(x \cdot u_2/N)$ in the direction t_1 at $x = u_1$, we find $\mathrm{Cov}(K_Q(u_2), g_1) = \sqrt{N}s_r(\xi_Q'(r) + r\xi_Q''(r))$. After

division by $\sqrt{N}$, this is (11.12); the case of g_2 is symmetric, and the covariances at a single point vanish by Lemma 7.1.

Marks. Applying the radial derivative to the kernel, we find that the cross covariance of Z_1 and Z_2 is $\Sigma_Q(r)$; for example, $\operatorname{Cov}(K_Q(u_1), K_Q(u_2)) = \partial_s \partial_{s'} [N\tilde{\xi}_Q(ss'r)]_{s=s'=1} = N(r\xi'_Q(r) + r^2\xi''_Q(r))$. The transverse gradients do not affect the conditional law of Z, because they are uncorrelated with (Z, g_1, g_2) and are conditioned to vanish. The inverse of the covariance of (g_1, g_2) is $\left(\begin{smallmatrix} \gamma & \tau \\ \tau & \gamma \end{smallmatrix}\right)$. In view of (11.12), the covariance of (Z_1, Z_2) with (g_1, g_2) is $J = \left(\begin{smallmatrix} 0 & \chi \\ \chi & 0 \end{smallmatrix}\right)$, and a direct multiplication shows that

$$J \begin{pmatrix} \gamma & \tau \\ \tau & \gamma \end{pmatrix} J^{\mathsf{T}} = \begin{pmatrix} \gamma\chi\chi^{\mathsf{T}} & \tau\chi\chi^{\mathsf{T}} \\ \tau\chi\chi^{\mathsf{T}} & \gamma\chi\chi^{\mathsf{T}} \end{pmatrix}.$$

Subtracting this matrix from the unconditional covariance $\left(\begin{smallmatrix} \Sigma_Q & \Sigma_Q(r) \\ \Sigma_Q(r) & \Sigma_Q \end{smallmatrix}\right)$ leaves $M(r)$. This proves (a).

Bulk entries. For $v, v' \in \{e_j\}$ at $x = u_1$, the products $v \cdot y$ and $v' \cdot y$ with $y = u_2$ vanish. By (11.15), the covariance of $\partial_{e_j}\partial_{e_k} H_Q(u_1)$ with $\partial_{e_l}\partial_{e_m} H_Q(u_2)$ is $\xi''_Q(r)(\delta_{jl}\delta_{km} + \delta_{jm}\delta_{kl})/N$, and its covariance with a row entry $\partial_{e_l}\partial_{t_2} H_Q(u_2)$ or with the corner z_2 vanishes, because $e_j \cdot t_2 = e_k \cdot t_2 = 0$. By the analogous formula for three derivatives, bulk entries are uncorrelated with all gradients and marks at both points. Dividing by the variances b_Q/N and $2b_Q/N$, we find that the standardized covariance is $\rho(r)$. All variables are jointly Gaussian, and the bulk entries are uncorrelated with the conditioning variables, the rows, and the corners; this proves (i).

Rows. Let $y_{1,j} = \partial_{e_j}\partial_{t_1} H_Q(u_1)$. The only conditioning variable correlated with it is $\partial_{e_j} H_Q(u_2)$, with covariance $\xi''_Q(r)(t_1 \cdot u_2)/N = s_r \xi''_Q(r)/\sqrt{N}$. By (11.15) with $(v, v', w, w') = (e_j, t_1, e_j, t_2)$, its unconditional covariance with $y_{2,j}$ is $(s_r^2 \xi'''_Q(r) - r\xi''_Q(r))/N$, and its unconditional variance is b_Q/N. Regression on the transverse pair $(\partial_{e_j} H_Q(u_1), \partial_{e_j} H_Q(u_2))$, whose covariance has inverse $(a_Q^2 - \xi'_Q(r)^2)^{-1}\left(\begin{smallmatrix} a_Q & -\xi'_Q(r) \\ -\xi'_Q(r) & a_Q \end{smallmatrix}\right)$, subtracts $a_Q s_r^2 \xi''_Q(r)^2/(N(a_Q^2 - \xi'_Q(r)^2))$ from the variance. We conclude that $\operatorname{Cov}(y_i) = (b_Q\theta(r)/N)\mathrm{Id}_n$. Rows with different indices j, and rows and corners, are uncorrelated by the same differentiation, and conditioning on Z changes nothing, since the rows are uncorrelated with Z and with (g_1, g_2). This proves (ii).

Corners. By (11.15) and (11.14), z_1 is uncorrelated with Z_1, with g_1, and with the transverse gradients, while

$$\operatorname{Cov}(z_1, g_2) = \frac{s_r \omega(r)}{\sqrt{N}}, \qquad \operatorname{Cov}(z_1, Z_2) = \frac{s_r^2}{\sqrt{N}}\big(\xi''_Q(r),\, 2\xi''_Q(r) + r\xi'''_Q(r)\big).$$

Regressing first on (g_1, g_2), we find that the covariance of z_1 with Z given $g_1 = g_2 = 0$ is

$$\operatorname{Cov}(z_1, Z) - \operatorname{Cov}\big(z_1, (g_1, g_2)\big) \begin{pmatrix} \gamma & \tau \\ \tau & \gamma \end{pmatrix} J^{\mathsf{T}} = \frac{\ell_1(r)}{\sqrt{N}}.$$

Regressing next on Z, whose conditional covariance is $M(r)$ and whose prescribed value is $\sqrt{N}(h_1,k_1,h_2,k_2)$, we obtain $\mathbb{E}z_1 = \ell_1(r)M(r)^{-1}(h_1,k_1,h_2,k_2)^{\mathsf{T}}$: the factor $N^{-1/2}$ in the covariance cancels the size $\sqrt{N}$ of the prescribed marks. From (11.1), we have $\xi_Q''(r), r\xi_Q'''(r) = O(|r|^{d-2})$, $\omega(r) = O(|r|^{d-3})$, and $\chi(r) = O(|r|^{d-1})$. As γ, τ, and $M(r)^{-1}$ are bounded, every coordinate of $\ell_1(r)$ is $O(|r|^{d-2})$, and because the marks lie in the compact set $\mathcal{K}_0$, we conclude that $|\mathbb{E}z_1| \le C\,|r|^{d-2}$. The conditional variance of z_1 is $N^{-1}(2b_Q - s_r^2\omega(r)^2\gamma(r) - \ell_1(r)M(r)^{-1}\ell_1(r)^{\mathsf{T}})$, which is at most $2b_Q/N$. The corner z_2 is treated in the same way, with the roles of the points exchanged. This proves (iii). □

Lemma 11.2 (Small-overlap orders and density ratios)**.** *Let ξ be a non-pure mixture and $|r| \le \delta_0$. Then*

$$\begin{aligned} \big\|M_{11}(r) - \Sigma_Q\big\| \le C\,|r|^{2d-2}, \qquad \big\|M_{12}(r)\big\| \le C\,|r|^{d}, \\ |\theta(r) - 1| \le C\,|r|^{2d-4}, \qquad |\rho(r)| \le C\,|r|^{d-2}, \end{aligned} \tag{11.16}$$

where M_{11} and M_{12} are the diagonal and off-diagonal blocks of M. Let $g_N(r)$ be the density at zero of the pair of spherical gradients $(\nabla_{S_N}H_Q(u_1), \nabla_{S_N}H_Q(u_2))$, in orthonormal frames, divided by the product $(2\pi a_Q)^{-(N-1)}$ of the two one-point gradient densities at zero. Then

$$g_N(r) = \Big(1 - \frac{\xi_Q'(r)^2}{a_Q^2}\Big)^{-n/2}\Big(1 - \frac{c(r)^2}{a_Q^2}\Big)^{-1/2}, \qquad |\log g_N(r)| \le C\big(N\,|r|^{2d-2} + |r|^{2d-4}\big). \tag{11.17}$$

For $(h_i,k_i) \in \mathcal{K}_0$, let $\zeta = \sqrt{N}(h_1,k_1,h_2,k_2)$, $\zeta_i = \sqrt{N}(h_i,k_i)$, and $v_N(r;\zeta) = p_{M(r)}(\zeta)/(p_{\Sigma_Q}(\zeta_1)p_{\Sigma_Q}(\zeta_2))$. Then

$$|\log v_N(r;\zeta)| \le CN\,|r|^{d}. \tag{11.18}$$

These bounds hold for both signs of r.

Proof. By (11.1), we have $\chi(r) = O(|r|^{d-1})$, $\tau(r) = O(|r|^{d-2})$, and every entry of $\Sigma_Q(r)$ is $O(|r|^d)$. Then $\gamma\chi\chi^{\mathsf{T}} = O(|r|^{2d-2})$ and $\tau\chi\chi^{\mathsf{T}} = O(|r|^{3d-4}) = O(|r|^d)$, which is the first line of (11.16). The second line follows from the formulas for θ and ρ in (11.9).

Recall from Lemma 11.1(a) that the gradients split into n independent transverse pairs and the in-plane pair. A centered Gaussian pair with covariance $\left(\begin{smallmatrix} a_Q & b \\ b & a_Q \end{smallmatrix}\right)$ has density $(2\pi)^{-1}(a_Q^2 - b^2)^{-1/2}$ at zero, while the product of its one-dimensional marginal densities at zero is $(2\pi a_Q)^{-1}$. Taking the product over the $n+1 = N-1$ pairs, we find the formula for g_N. Using $0 \le -\log(1-x) \le 2x$ for $0 \le x \le 1/4$, and the fact that $\xi_Q'(r)^2/a_Q^2$ and $c(r)^2/a_Q^2$ are at most $1/4$ for $|r| \le \delta_0$, we find $|\log g_N| \le n\xi_Q'(r)^2/a_Q^2 + c(r)^2/a_Q^2$, and (11.1) implies the bound in (11.17).

Let $M_0 = \operatorname{diag}(\Sigma_Q, \Sigma_Q)$ and $E_r = M(r) - M_0$, so that $\|E_r\| \le C\,|r|^d$ by (11.16). The matrices $M_0 + tE_r$, $0 \le t \le 1$, are uniformly positive definite, since they are convex combinations of M_0 and $M(r)$. The resolvent identity $M(r)^{-1} - M_0^{-1} = -M(r)^{-1}E_rM_0^{-1}$ gives

$\big\|M(r)^{-1} - M_0^{-1}\big\| \le C\,\|E_r\|$, and integrating $\frac{\mathrm{d}}{\mathrm{d}t}\log\det(M_0+tE_r) = \mathrm{Tr}((M_0+tE_r)^{-1}E_r)$ over $[0,1]$, we find $|\log\det M(r) - \log\det M_0| \le C\,\|E_r\|$. Since

$$\log v_N(r;\zeta) = -\frac{1}{2}\log\frac{\det M(r)}{\det M_0} - \frac{1}{2}\zeta^{\mathsf{T}}\big(M(r)^{-1} - M_0^{-1}\big)\zeta$$

and $|\zeta|^2 \le CN$ for marks in $\mathcal{K}_0$, we obtain $|\log v_N| \le C(1+N)\,|r|^d$, which is (11.18). No step used the sign of r or the parity of the supported degrees. □

The factor N in (11.18) is the reason for the restriction $\alpha > 1/4$ below: when $d = 4$, the cross covariance of the marks is of order r^4, and it is multiplied by the squared size N of the prescribed marks.

Remark 11.3 (Pure covariances). Let $\xi_Q(r) = \xi(Q)r^p$ with $p \ge 4$, so that $K_Q = pH_Q$. The two-point conditioned law then fixes the two values and the two spherical gradients only (Section 7.4), and $k_i = ph_i$. We replace Z_i by $H_Q(u_i)/\sqrt{N}$, Σ_Q by $\xi(Q)$, $\chi(r)$ by $s_r\xi_Q'(r)$, $\Sigma_Q(r)$ by $\xi_Q(r)$, and $\ell_1(r)$ by

$$\ell_1(r) = \big(-s_r\omega(r)\gamma(r)\,s_r\xi_Q'(r),\ s_r^2\xi_Q''(r) - s_r\omega(r)\tau(r)\,s_r\xi_Q'(r)\big),$$

with ℓ_2 its reflection. The resulting 2×2 matrix $M(r)$ tends to $\xi(Q)\mathrm{Id}_2$ as $r \to 0$, and the proofs of Lemmas 11.1 and 11.2 apply without change. Marks are restricted to the admissible set $\{(h, ph)\} \cap \mathcal{K}_0$, which contains (E_0, K_0) because $K_0 = pE_0$ in a pure model, and no inverse of the singular matrix Σ_Q is taken. With these changes, every statement of Sections 11.3 and 11.4 holds for pure covariances, with the energies h_1, h_2 as the only mark variables. The intensities and integrals then involve the energies alone, with a different Jacobian; they are given in the paragraph on pure models after (11.39).

11.3 Determinants under two-point conditioning

By Lemma 7.4, under $\mathbb{P}_{h,k,u}$ the matrix $-\nabla^2_{S_N}H_Q(u)$ has the law of $k\,\mathrm{Id}_{N-1} - W_{N-1,N}$, so that $\mathbb{E}_{h,k,u}D(u) = \mu_{N-1,N}(k)$; and by Lemma 11.1(b), each bulk block $W^{(i)}$ has the law of $W_{n,N}$, $n = N-2$. For $k \ge 2\sqrt{b_Q} + \kappa$, let

$$m_{\mathrm{sc}}(k) = \frac{k - \sqrt{k^2 - 4b_Q}}{2b_Q}, \qquad w(k) = k - b_Q m_{\mathrm{sc}}(k) = \frac{k + \sqrt{k^2 - 4b_Q}}{2}. \tag{11.19}$$

In the notation of Section 7.4, $m_{\mathrm{sc}}(k) = b_Q^{-1/2}g_{\mathrm{sc}}(k/\sqrt{b_Q})$ is the Stieltjes transform at k of the semicircle law on $[-2\sqrt{b_Q}, 2\sqrt{b_Q}]$. The number $w(k) \ge k/2 \ge \sqrt{b_Q}$ is the limit of the Schur complement below. For a real symmetric matrix A with eigenvalues $\lambda_j(A)$, we define the truncated log-determinant and the truncated resolvent

$$\Phi_k(A) = \sum_j \log\max\big\{\,|k - \lambda_j(A)|\,, \kappa/2\big\}, \qquad \mathcal{R}_k(A) = \phi_k(A), \tag{11.20}$$

where $\phi_k(\lambda) = 1/\max\{k-\lambda, \kappa/2\}$ is applied to A through its spectral decomposition. The function ϕ_k is bounded by $2/\kappa$ and is Lipschitz with constant $4/\kappa^2$. If $\|A\| \le 2\sqrt{b_Q} + \kappa/2$ and $k \ge 2\sqrt{b_Q} + \kappa$, then every eigenvalue of A satisfies $k - \lambda_j(A) \ge \kappa/2$, so that $e^{\Phi_k(A)} = |\det(k-A)|$ and $\mathcal{R}_k(A) = (k-A)^{-1}$.

We use two facts from Section 7. First, by Lemma 7.5(2), if A is a centered Gaussian symmetric $m \times m$ matrix, $m \le N$, such that $\operatorname{Var}\operatorname{Tr}(AB) \le (2b_Q/N)\operatorname{Tr}(B^2)$ for every symmetric B, then

$$\mathbb{P}\big(\|A\| > 2\sqrt{b_Q} + t\big) \le 2e^{-Nt^2/(4b_Q)}, \qquad t > 0. \tag{11.21}$$

Second, by Proposition 7.7 with $g = \kappa$ and $C = 0$, for $m \in \{N-2, N-1\}$, every fixed $q \ge 1$, and all $k \in I$,

$$\mathbb{E}\Big(\frac{|\det(k - W_{m,N})|}{\mu_{m,N}(k)}\Big)^q \le \mathbb{E}\Big(\frac{e^{\Phi_k(W_{m,N})}}{\mu_{m,N}(k)}\Big)^q \le C_q. \tag{11.22}$$

Lemma 11.4 (Trace of the truncated resolvent)**.** *For every fixed $q \ge 1$, uniformly in $k \in I$,*

$$\Big\| \frac{1}{N}\operatorname{Tr}\mathcal{R}_k(W_{n,N}) - m_{\mathrm{sc}}(k) \Big\|_q \le \frac{C_q}{N}. \tag{11.23}$$

Proof. We write $W = W_{n,N}$, $m = N^{-1}\operatorname{Tr}\mathcal{R}_k(W)$, $\bar{m} = \mathbb{E}m$, and $\mathcal{E} = \{\|W\| \le 2\sqrt{b_Q} + \kappa/2\}$, so that $\mathbb{P}(\mathcal{E}^c) \le 2e^{-N\kappa^2/(16 b_Q)}$ by (11.21). The proof has three steps: concentration of m, an approximate quadratic equation for $\bar{m}$, and the selection of the correct root. We first enlarge the range of k. Fix $k_* \ge K_0 + 1$ so large that $(k_* - 2\sqrt{b_Q} - \kappa/2)^{-1} < k_*/(4b_Q)$, and let $I_* = [2\sqrt{b_Q} + \kappa, k_*] \supset I$. The first two steps hold uniformly in $k \in I_*$, with constants depending on k_*.

Concentration. In the standard Gaussian coordinates of W, one for each entry on and above the diagonal, the map to W is Lipschitz into the Frobenius norm with constant $\sqrt{2b_Q/N}$. Using the Hoffman–Wielandt inequality (see the proof of Proposition 7.7), we find that $A \mapsto \operatorname{Tr}\phi_k(A)$ is Lipschitz in the Frobenius norm with constant $(4/\kappa^2)\sqrt{n}$. Then m is a Lipschitz function of the Gaussian coordinates with constant at most $(4/\kappa^2)\sqrt{2b_Q}/N$, and the Gaussian concentration inequality of Lemma 7.5(1) yields

$$\|m - \bar{m}\|_q \le \frac{C_q}{N}, \qquad \operatorname{Var}(m) \le \frac{C}{N^2}. \tag{11.24}$$

The quadratic equation. On $\mathcal{E}$, we have $(k-W)\mathcal{R}_k(W) = \mathrm{Id}_n$, so $k\operatorname{Tr}\mathcal{R}_k(W) - \operatorname{Tr}(W\mathcal{R}_k(W)) = n$. On $\mathcal{E}^c$, the left side is at most $Cn(1+\|W\|)$ in absolute value. By the Cauchy–Schwarz inequality and $\mathbb{E}\,\|W\|^2 \le C$, we have

$$k\bar{m} = \frac{n}{N} + \frac{1}{N}\mathbb{E}\operatorname{Tr}\big(W\mathcal{R}_k(W)\big) + O(e^{-cN}). \tag{11.25}$$

We evaluate the last expectation by Gaussian integration by parts, $\mathbb{E}[W_{ij}F(W)] = \mathrm{Var}(W_{ij})\,\mathbb{E}[\partial_{W_{ij}}F(W)]$, where $W_{ij} = W_{ji}$ is treated as one variable. A scalar function with Lipschitz constant ℓ induces a matrix function with Lipschitz constant ℓ in the Frobenius norm. In particular, the entries of $\mathcal{R}_k(A)$ are Lipschitz functions of A, their weak derivatives are bounded by a constant, and on $\mathcal{E}$ they agree with the derivatives of the resolvent, $\partial_{W_{ij}}\mathcal{R}_k = \mathcal{R}_k(E^{ij}+E^{ji})\mathcal{R}_k$ for $i \neq j$ and $\partial_{W_{ii}}\mathcal{R}_k = \mathcal{R}_k E^{ii}\mathcal{R}_k$, where E^{ij} are the matrix units. For these resolvent formulas, we have the classical identity

$$\sum_{i\neq j}\frac{b_Q}{N}\big[\mathcal{R}_k(E^{ij}+E^{ji})\mathcal{R}_k\big]_{ji} + \sum_i \frac{2b_Q}{N}\big[\mathcal{R}_k E^{ii}\mathcal{R}_k\big]_{ii} = \frac{b_Q}{N}\big[(\operatorname{Tr}\mathcal{R}_k)^2 + \operatorname{Tr}\mathcal{R}_k^2\big],$$

since $\mathcal{R}_k$ is symmetric. The n^2 differences between the true derivatives and these formulas are bounded and supported on $\mathcal{E}^c$, and each is multiplied by a variance at most $2b_Q/N$. We conclude that

$$\mathbb{E}\operatorname{Tr}\big(W\mathcal{R}_k(W)\big) = \frac{b_Q}{N}\mathbb{E}\big[(\operatorname{Tr}\mathcal{R}_k)^2 + \operatorname{Tr}\mathcal{R}_k^2\big] + O(Ne^{-cN}).$$

Using $\operatorname{Tr}\mathcal{R}_k^2 \le 4n/\kappa^2$ and $\mathbb{E}[(\operatorname{Tr}\mathcal{R}_k)^2]/N^2 = \bar{m}^2 + \mathrm{Var}(m)$, and inserting this into (11.25) together with (11.24), we obtain

$$b_Q\bar{m}^2 - k\bar{m} + \frac{n}{N} = O\Big(\frac{1}{N}\Big), \qquad k \in I_*. \tag{11.26}$$

The root. The roots of $b_Q x^2 - kx + n/N$ are $m_\pm(k) = (k \pm \sqrt{k^2 - 4b_Q n/N})/(2b_Q)$, and $m_+(k) - m_-(k) \ge \sigma_0 = \sqrt{\kappa(4\sqrt{b_Q}+\kappa)}/b_Q$ for $k \in I_*$. Then (11.26) implies that $b_Q\,|\bar{m}-m_+|\,|\bar{m}-m_-| \le C/N$. As the roots are separated by σ_0, for N large $\bar{m}(k)$ lies within C'/N of one of them, where $C' = 4C/(b_Q\sigma_0)$. Let $A_\pm = \{k \in I_* : |\bar{m}(k) - m_\pm(k)| \le C'/N\}$. These sets are closed, because $k \mapsto \bar{m}(k)$ is continuous by dominated convergence, and for N large they are disjoint and cover I_*. By the connectedness of I_*, one of them is empty. At $k = k_*$, on $\mathcal{E}$ we have $\mathcal{R}_{k_*}(W) \le (k_* - 2\sqrt{b_Q} - \kappa/2)^{-1}\mathrm{Id}$, which implies $\bar{m}(k_*) < k_*/(4b_Q) + Ce^{-cN}$. Together with $m_+(k_*) \ge k_*/(2b_Q)$, this shows that $k_* \in A_-$, and therefore $A_- = I_*$. Finally, since $n/N = 1 - 2/N$ and $k^2 - 4b_Q \ge \kappa(4\sqrt{b_Q}+\kappa)$ on I_*, we have $|m_-(k) - m_{\mathrm{sc}}(k)| \le C/N$. We conclude that $|\bar{m}(k) - m_{\mathrm{sc}}(k)| \le C/N$ on I, and (11.24) gives (11.23). □

Lemma 11.5 (Approximation of the Schur complement)**.** *Let ξ be a non-pure mixture, $|r| \le \delta_0$, and $(h_i, k_i) \in \mathcal{K}_0$, and consider the two-point conditioned law $\mathbb{E}_{\mathrm{pair}}$ with these marks. Let $D_i = \Big|\det \nabla^2_{S_N} H_Q(u_i)\Big|$ and $\widetilde{W}_i = \Big|\det(k_i - W^{(i)})\Big| \,/\, \mu_{n,N}(k_i)$. Then, for every fixed $q \ge 1$,*

$$\Big\|\frac{D_i}{w(k_i)\mu_{n,N}(k_i)} - \widetilde{W}_i\Big\|_q \le C_q\big(|r|^{d-2} + N^{-1/2}\big). \tag{11.27}$$

Moreover, uniformly in $k \in I$,

$$\mu_{N-1,N}(k) = w(k)\,\mu_{n,N}(k)\big(1 + O(N^{-1/2})\big), \tag{11.28}$$

and $\mathbb{E}_{\mathrm{pair}}(D_i/\mu_{N-1,N}(k_i))^q \le C_q$.

Proof. We drop the index i. Let $W = W^{(i)}$, let $\mathcal{E} = \{\|W\| \le 2\sqrt{b_Q} + \kappa/2\}$, and define the Schur complement

$$S = k - z - y^{\mathsf{T}}\mathcal{R}_k(W)y.$$

We first show that

$$\|S - w(k)\|_q \le C_q\big(|r|^{d-2} + N^{-1/2}\big). \tag{11.29}$$

Recall from Lemma 11.1(b) that, conditionally on W, the row y is a centered Gaussian vector with covariance $\sigma^2\mathrm{Id}_n$, $\sigma^2 = b_Q\theta(r)/N$. Diagonalizing $\mathcal{R}_k(W) = O\,\mathrm{diag}(\varrho_j)O^{\mathsf{T}}$, we can write $y^{\mathsf{T}}\mathcal{R}_k(W)y - \sigma^2\operatorname{Tr}\mathcal{R}_k(W) = \sigma^2\sum_j \varrho_j(\eta_j^2 - 1)$, where $\eta = \sigma^{-1}O^{\mathsf{T}}y$ is a standard Gaussian vector independent of W and $|\varrho_j| \le 2/\kappa$. Let $s \ge 1$ be an integer. In the expansion of the $2s$-th moment of $\sum_j \varrho_j(\eta_j^2 - 1)$, every term in which some index appears exactly once vanishes. The remaining terms involve at most s distinct indices, and the moment is at most $C_s n^s$. Taking the $2s$-th root, we find that the centered quadratic form has L^{2s} norm at most $C_s\sigma^2 n^{1/2} \le C_s N^{-1/2}$, uniformly in W. Next, using $w(k) = k - b_Q m_{\mathrm{sc}}(k)$ and $\sigma^2\operatorname{Tr}\mathcal{R}_k = b_Q\theta(r)N^{-1}\operatorname{Tr}\mathcal{R}_k$, we write

$$\begin{aligned} S - w(k) = {} & -z - \big(y^{\mathsf{T}}\mathcal{R}_k(W)y - \sigma^2\operatorname{Tr}\mathcal{R}_k(W)\big) \\ & - b_Q\theta(r)\Big(\frac{1}{N}\operatorname{Tr}\mathcal{R}_k(W) - m_{\mathrm{sc}}(k)\Big) - b_Q\big(\theta(r) - 1\big)m_{\mathrm{sc}}(k). \end{aligned}$$

We have $\|z\|_q \le |\mathbb{E}z| + \|z - \mathbb{E}z\|_q \le C\,|r|^{d-2} + C_q N^{-1/2}$ by Lemma 11.1(b)(iii). The second term was bounded above, the third is at most C_q/N by Lemma 11.4, and the last is $O(|r|^{2d-4}) = O(|r|^{d-2})$ by (11.16). This proves (11.29).

On $\mathcal{E}$, the matrix $k - W$ is invertible with inverse $\mathcal{R}_k(W)$, and the Schur complement formula applied to (11.13) gives $D = |\det(k - W)|\,|S|$. Hence, on $\mathcal{E}$,

$$\frac{D}{w(k)\mu_{n,N}(k)} - \widetilde{W} = \widetilde{W}\,\frac{|S| - w(k)}{w(k)},$$

and $\big||S| - w(k)\big| \le |S - w(k)|$. Using $w(k) \ge \sqrt{b_Q}$, Hölder's inequality, (11.22) for $\widetilde{W}$, and (11.29), we find that the L^q norm of $\mathbf{1}_{\mathcal{E}}$ times this difference is at most $C_q(|r|^{d-2} + N^{-1/2})$.

On $\mathcal{E}^c$, we use the bound of Lemma 7.11(1) with a border of size one and $g = \kappa$,

$$D \le e^{\Phi_k(W)}\Big(|k - z| + \frac{2}{\kappa}\,|y|^2\Big), \tag{11.30}$$

which holds even when $k - W$ is singular. The variables z and $|y|^2$ have bounded moments of every fixed order, the normalized exponential $e^{\Phi_k(W)}/\mu_{n,N}(k)$ has bounded moments

by (11.22), and $\mathbb{P}(\mathcal{E}^c) \le Ce^{-cN}$ by (11.21). By Hölder's inequality with three factors, we conclude that the L^q norms of $\mathbf{1}_{\mathcal{E}^c} D/(w(k)\mu_{n,N}(k))$ and of $\mathbf{1}_{\mathcal{E}^c}\widetilde{W}$ are at most $C_q e^{-c_q N}$. This completes the proof of (11.27), and together with (11.22) it shows that $D/(w(k)\mu_{n,N}(k))$ has bounded moments of every fixed order.

For (11.28), consider $-\nabla^2_{S_N} H_Q(u)$ under $\mathbb{P}_{h,k,u}$, which has the law of $k\,\mathrm{Id}_{N-1} - W_{N-1,N}$. The upper left $n \times n$ block of $W_{N-1,N}$ has the law of $W_{n,N}$, the first n entries of its last column form a centered Gaussian vector with covariance $(b_Q/N)\mathrm{Id}_n$, and its corner is centered with variance $2b_Q/N$; these three pieces are independent. The argument above applies with $\theta = 1$ and a centered corner, and it shows that $\big\| D(u)/(w(k)\mu_{n,N}(k)) - \widetilde{W} \big\|_1 \le CN^{-1/2}$ under $\mathbb{P}_{h,k,u}$. Taking expectations, we find $|\mu_{N-1,N}(k)/(w(k)\mu_{n,N}(k)) - 1| \le CN^{-1/2}$, which is (11.28). Combined with the moment bound for $D_i/(w(k_i)\mu_{n,N}(k_i))$, this gives the last assertion. □

The comparison of the two bulk determinants uses the following classical inequality of Gebelein [80]. We include the short proof to show that its constant does not depend on the dimension and that it holds for negative correlations.

Lemma 11.6 (Gaussian maximal correlation)**.** *Let (X, Y) be a centered Gaussian vector in $\mathbb{R}^m \times \mathbb{R}^m$ with $\mathrm{Cov}(X) = \mathrm{Cov}(Y) = \mathrm{Id}_m$ and $\mathrm{Cov}(X, Y) = \rho\,\mathrm{Id}_m$, where $|\rho| \le 1$. Then, for all square integrable functions F and G on $\mathbb{R}^m$,*

$$|\mathrm{Cov}(F(X), G(Y))| \le |\rho| \, \big(\mathrm{Var}\, F(X)\ \mathrm{Var}\, G(Y)\big)^{1/2}. \tag{11.31}$$

Proof. Let He_j be the Hermite polynomials, defined by $e^{sx - s^2/2} = \sum_{j\ge 0} s^j \mathrm{He}_j(x)/j!$, and for a multi-index $\lambda \in \mathbb{N}^m$ let $\mathsf{h}_\lambda(x) = \prod_i \mathrm{He}_{\lambda_i}(x_i)/\sqrt{\lambda_i!}$. These polynomials form an orthonormal basis of L^2 of the standard Gaussian measure on $\mathbb{R}^m$. For $s, t \in \mathbb{R}^m$, we have

$$\mathbb{E}\big[e^{s\cdot X - |s|^2/2} e^{t\cdot Y - |t|^2/2}\big] = e^{\rho\, s\cdot t},$$

and comparing the coefficients of $s^\lambda t^\mu$ on both sides, we find $\mathbb{E}[\mathsf{h}_\lambda(X)\mathsf{h}_\mu(Y)] = \mathbf{1}_{\lambda=\mu}\rho^{|\lambda|}$, where $|\lambda| = \sum_i \lambda_i$. We expand $F = \sum_\lambda \hat{F}_\lambda \mathsf{h}_\lambda$ and $G = \sum_\lambda \hat{G}_\lambda \mathsf{h}_\lambda$ in this basis. Since $(F, G) \mapsto \mathbb{E}[F(X)G(Y)]$ is a bounded bilinear form on $L^2 \times L^2$, the identity extends from polynomials and gives $\mathrm{Cov}(F(X), G(Y)) = \sum_{\lambda\neq 0} \hat{F}_\lambda \hat{G}_\lambda \rho^{|\lambda|}$. Then (11.31) follows from the Cauchy–Schwarz inequality and $|\rho|^{|\lambda|} \le |\rho|$ for $\lambda \neq 0$. □

Lemma 11.7 (Ratio of the determinant moments)**.** *Under the assumptions of Lemma 11.5,*

$$\begin{aligned} \Big| \frac{\mathbb{E}_{\mathrm{pair}}[D_1 D_2]}{\mu_{N-1,N}(k_1)\mu_{N-1,N}(k_2)} - 1 \Big| &\le C\big(|r|^{d-2} + N^{-1/2}\big), \\ \mathbb{E}_{\mathrm{pair}}\Big(\frac{D_1 D_2}{\mu_{N-1,N}(k_1)\mu_{N-1,N}(k_2)} \Big)^2 &\le C. \end{aligned} \tag{11.32}$$

Proof. Let X and Y be the vectors of standardized entries on and above the diagonal of $W^{(1)}$ and $W^{(2)}$, of dimension $n(n+1)/2$. We see from Lemma 11.1(b)(i) that they satisfy the assumptions of Lemma 11.6 with $\rho = \rho(r)$. The variables $\widetilde{W}_1$ and $\widetilde{W}_2$ are functions of X and Y with mean one and, by (11.22), bounded second moments. Lemma 11.6 and (11.16) imply that

$$\Big|\mathbb{E}[\widetilde{W}_1\widetilde{W}_2] - 1\Big| \le C\,|\rho(r)| \le C\,|r|^{d-2}.$$

Let $\widehat{D}_i = D_i/(w(k_i)\mu_{n,N}(k_i))$. Using the Cauchy–Schwarz inequality, (11.27) with $q = 2$, and the moment bounds of Lemma 11.5, we find

$$\Big|\mathbb{E}[\widehat{D}_1\widehat{D}_2] - \mathbb{E}[\widetilde{W}_1\widetilde{W}_2]\Big| \le \Big\|\widehat{D}_1 - \widetilde{W}_1\Big\|_2 \Big\|\widehat{D}_2\Big\|_2 + \Big\|\widetilde{W}_1\Big\|_2 \Big\|\widehat{D}_2 - \widetilde{W}_2\Big\|_2 \le C\big(|r|^{d-2} + N^{-1/2}\big).$$

Finally, (11.28) replaces $w(k_i)\mu_{n,N}(k_i)$ by $\mu_{N-1,N}(k_i)$ at the cost of a factor $1 + O(N^{-1/2})$, which proves the first bound in (11.32). The second follows from the Cauchy–Schwarz inequality and the fourth moments of $D_i/\mu_{N-1,N}(k_i)$ in Lemma 11.5. □

Lemma 11.7 uses the correlation $\rho(r)$ of the bulk blocks only through (11.31), and its bound does not grow with the number $n(n+1)/2$ of entries. This is what allows $d = 4$. For comparison, the relative entropy of the joint Gaussian law of the two vectors of standardized bulk entries with respect to the product of their marginal laws is

$$-\frac{n(n+1)}{4}\log\big(1-\rho(r)^2\big) = O\big(N^2\rho(r)^2\big), \qquad \rho(r) \to 0.$$

This quantity tends to zero only if $|r| \ll N^{-1/(d-2)}$, a range that is incompatible with $\alpha < 1/2$ when $d = 4$. The comparison concerns the bulk blocks only: the corners may remain correlated even at $r = 0$, as noted after Lemma 11.1. The determinant comparison instead uses the maximal correlation inequality for the bulk and the approximation of the Schur complement in Lemma 11.5 for the border.

Lemma 11.8 (Regularity under two-point conditioning)**.** *There exist $\lambda_* > 0$, $C_* < \infty$, $\delta_1 \in (0, \delta_0]$, and $C, c > 0$ with the following property. Let $|r| \le \delta_1$ and $(h_i, k_i) \in \mathcal{K}_0$, and consider either the two-point conditioned law $\mathbb{E}_{\mathrm{pair}}$ with these marks, or the one-point conditioned law $\mathbb{E}_{h,k,u}$ with $(h,k) \in \mathcal{K}_0$. Let* Reg *be the event that the Hessian at each conditioned point satisfies $\nabla^2_{S_N} H_Q(u_i) \le -\lambda_* \mathrm{Id}$, and that*

$$\sup_{|x| \le 2\sqrt{N}} \Big\|\nabla^j H_Q(x)\Big\| \le C_* N^{1-j/2}, \qquad 0 \le j \le 3. \tag{11.33}$$

Then $\mathbb{P}(\mathrm{Reg}^c) \le Ce^{-cN}$ under either law. Moreover, for N large, $\mathbb{E}_{\mathrm{pair}}[D_1D_2] \ge \frac{1}{2}\mu_{N-1,N}(k_1)\mu_{N-1,N}(k_2)$, and

$$\mathbb{E}_{\mathrm{pair}}\big[D_1D_2\mathbf{1}_{\mathrm{Reg}^c}\big] \le Ce^{-cN}\,\mathbb{E}_{\mathrm{pair}}[D_1D_2], \qquad \mathbb{E}_{h,k,u}\big[D(u)\mathbf{1}_{\mathrm{Reg}^c}\big] \le Ce^{-cN}\,\mu_{N-1,N}(k). \tag{11.34}$$

Proof. We choose $\delta_1 \le \delta_0$ so that the constant C of Lemma 11.1(b)(iii) satisfies $C\delta_1^{d-2} \le \kappa/4$, and so that the constant of (11.32) satisfies $C\delta_1^{d-2} \le 1/4$. Then $\mathbb{E}_{\mathrm{pair}}[D_1 D_2] \ge \frac{1}{2}\mu_{N-1,N}(k_1)\mu_{N-1,N}(k_2)$ for $|r| \le \delta_1$ and N large.

The Hessians. Let T_i be the compression of $\nabla^2 H_Q(u_i)$ to $u_i^\perp$, so that $\nabla^2_{S_N} H_Q(u_i) = T_i - k_i\mathrm{Id}$ by (7.3). Before conditioning, T_i has the law of $W_{N-1,N}$, and Gaussian conditioning decreases the covariance of the centered part, as a quadratic form on symmetric matrices. Then $T_i - \mathbb{E}_{\mathrm{pair}} T_i$ satisfies the hypothesis of (11.21), and $\|T_i - \mathbb{E}_{\mathrm{pair}} T_i\| \le 2\sqrt{b_Q} + \kappa/4$ except on an event of probability at most $2e^{-N\kappa^2/(64 b_Q)}$. By Lemma 11.1(b), the conditional mean of T_i vanishes except for its corner entry, which is at most $\kappa/4$ in absolute value. Since $k_i \ge 2\sqrt{b_Q} + \kappa$, on this event we have $\nabla^2_{S_N} H_Q(u_i) \le (2\sqrt{b_Q} + \kappa/2 - k_i)\mathrm{Id} \le -(\kappa/2)\mathrm{Id}$. We take $\lambda_* = \kappa/2$. Under the one-point law, the mean vanishes and the same argument applies.

The derivatives. Recall from (7.1) that $H_Q(x) = \sum_p \sqrt{c_p Q^p}\, N^{-(p-1)/2}\langle Z_p, x^{\otimes p}\rangle$, where Z_p is the tensor of coefficients of degree p. Under the conditioned law, we write H_Q as the sum of its conditional mean $\mathfrak{m}(x)$ and a centered Gaussian polynomial whose coefficient vector has covariance at most the identity. For each degree p, let $\mathcal{T}_p$ be the symmetric p-linear form obtained by polarizing the degree-p part of the centered polynomial. For unit vectors $x^1, \dots, x^p$, we have $\mathrm{Var}\, \mathcal{T}_p(x^1, \dots, x^p) \le c_p Q^p N^{1-p}$. Let $\mathcal{N}$ be a $1/(4p)$-net of the unit sphere of $\mathbb{R}^N$ with at most $(1+8p)^N$ points. Changing one argument at a time from a point of $\mathcal{N}$ to an arbitrary unit vector, we find that the norm of $\mathcal{T}_p$ is at most its maximum over $\mathcal{N}^p$ plus one quarter of itself, hence at most twice that maximum. By a union bound over the $(1+8p)^{pN}$ points of $\mathcal{N}^p$ and the Gaussian tail bound, we see that $\|\mathcal{T}_p\| \le C_p N^{1-p/2}$ with probability at least $1 - e^{-N}$, if C_p is large enough. Evaluating $p - j$ arguments at x with $|x| \le 2\sqrt{N}$, we obtain the bound $CN^{1-j/2}$ for the j-th derivative of the centered part.

It remains to bound the conditional mean. The transverse gradients do not contribute to $\mathfrak{m}$, because they are uncorrelated with the other conditioning variables and are conditioned to vanish. The covariances of $H_Q(x)$ with $H_Q(u_i)/\sqrt{N}$, $K_Q(u_i)/\sqrt{N}$, and g_i are

$$\sqrt{N}\,\xi_Q\Big(\frac{x\cdot u_i}{N}\Big), \qquad \sqrt{N}\,\frac{x\cdot u_i}{N}\,\xi_Q'\Big(\frac{x\cdot u_i}{N}\Big), \qquad \sqrt{N}\,\frac{x\cdot t_i}{\sqrt{N}}\,\xi_Q'\Big(\frac{x\cdot u_i}{N}\Big).$$

The covariance matrix of (Z, g_1, g_2) has a uniformly bounded inverse for $|r| \le \delta_0$, since its Schur complements are the covariance of (g_1, g_2) and $M(r)$, and the prescribed value of Z is $O(\sqrt{N})$. Therefore $\mathfrak{m}(x) = N\mathcal{P}(x\cdot u_1/N,\, x\cdot u_2/N,\, x\cdot t_1/\sqrt{N},\, x\cdot t_2/\sqrt{N})$ for a polynomial $\mathcal{P}$ of bounded degree with uniformly bounded coefficients. For $|x| \le 2\sqrt{N}$, the four arguments are bounded by 2, and their derivatives in a unit direction are at most $N^{-1/2}$, which implies that $\|\nabla^j \mathfrak{m}(x)\| \le CN^{1-j/2}$. Together with the bound for the centered part, this proves (11.33), for a suitable C_*, with probability at least $1 - Ce^{-cN}$.

The weighted loss. By the Cauchy–Schwarz inequality and (11.32),

$$\begin{aligned}\mathbb{E}_{\mathrm{pair}}\big[D_1D_2\mathbf{1}_{\mathrm{Reg}^c}\big] &\le \big(\mathbb{E}_{\mathrm{pair}}[(D_1D_2)^2]\big)^{1/2}\mathbb{P}(\mathrm{Reg}^c)^{1/2}\\ &\le Ce^{-cN/2}\mu_{N-1,N}(k_1)\mu_{N-1,N}(k_2) \le 2Ce^{-cN/2}\,\mathbb{E}_{\mathrm{pair}}[D_1D_2],\end{aligned}$$

where in the last step we used the lower bound on $\mathbb{E}_{\mathrm{pair}}[D_1D_2]$. The one-point bound follows in the same way from (11.22), or from Corollary 7.8. □

11.4 The local pair intensity

Lemma 11.9 (Local two-point Kac–Rice formula)**.** *Let $\mathcal{O} = \{(u_1,u_2) \in S_N \times S_N : |R(u_1,u_2)| < \delta_0\}$. For every nonnegative Borel function b of the field and of two points,*

$$\begin{aligned}&\mathbb{E} \sum_{\substack{(u_1,u_2)\in\mathcal{O}\\ \nabla_{S_N}H_Q(u_1)=\nabla_{S_N}H_Q(u_2)=0}} b(H_N,u_1,u_2)\\ &\qquad = \int_{\mathcal{O}} p_{u_1,u_2}(0)\,\mathbb{E}^{\nabla}_{u_1,u_2}\big[D_1D_2\,b(H_N,u_1,u_2)\big]\,\mathrm{vol}(\,\mathrm{d}u_1)\,\mathrm{vol}(\,\mathrm{d}u_2),\end{aligned} \tag{11.35}$$

where $p_{u_1,u_2}(0)$ is the density at zero of $(\nabla_{S_N}H_Q(u_1), \nabla_{S_N}H_Q(u_2))$ in orthonormal frames, $\mathbb{E}^{\nabla}_{u_1,u_2}$ is the conditioned law given that both gradients vanish (defined by the regression formula (7.4)*),* vol *is the surface measure on S_N, and both sides may be infinite.*

Proof. We consider the function $f(u_1,u_2) = H_Q(u_1) + H_Q(u_2)$ on the open subset $\mathcal{O}$ of the product manifold $S_N \times S_N$. It depends linearly on the Gaussian coefficients of (7.1). Its gradient at (u_1,u_2) is the pair of spherical gradients, whose covariance is positive definite on $\mathcal{O}$ by Lemma 11.1(a): its blocks have determinants $a_Q^2 - \xi_Q'(r)^2$ and $a_Q^2 - c(r)^2$, both at least $3a_Q^2/4$. Note that the critical points of f in $\mathcal{O}$ are exactly the pairs of critical points of H_Q with overlap in $(-\delta_0,\delta_0)$, and at such a pair the Hessian of f is block diagonal, with absolute determinant D_1D_2.

We now observe that the proof of the coefficient-space formula Proposition 7.16 is local. Consider a coordinate chart on which a fixed set of coefficient coordinates has an invertible submatrix in the gradient map. On such a chart, the critical points are the zeros of an affine function of the coefficients. The area formula of Lemma 7.15 then expresses the expected sum of the mark over these zeros as an integral against the Riemannian volume, whose integrand is the gradient density at zero times the conditioned expectation of the absolute Hessian determinant times the mark. Compactness of the manifold enters that proof only through the finiteness of the atlas. The open set $\mathcal{O}$ is covered by countably many such charts. We take a countable partition of unity on $\mathcal{O}$ subordinate to this cover, apply the chart identity to the mark b times each partition function, and sum. Since all terms are nonnegative, Tonelli's theorem and monotone convergence justify the summation, with both sides allowed to be infinite. This proves (11.35). □

Under $\mathbb{E}^{\nabla}_{u_1,u_2}$, the vector Z has the nondegenerate density $p_{M(r)}$ by Lemma 11.1(a), and disintegrating with respect to Z as in Lemma 7.4(4) turns $\mathbb{E}^{\nabla}_{u_1,u_2}$ into an integral of $\mathbb{E}_{\mathrm{pair}}$ over the marks. In the normalized marks $(h_1,k_1,h_2,k_2) = Z/\sqrt{N}$, this introduces the Jacobian N^2. We define the one-point intensity

$$j_N(h,k) = |S_N|\,(2\pi a_Q)^{-(N-1)/2}\,p_N(h,k)\,\mu_{N-1,N}(k), \tag{11.36}$$

which, by Corollary 7.18, is the density with respect to $\mathrm{d}h\,\mathrm{d}k$ of the expected number of critical points of H_Q with normalized marks in $\mathrm{d}h\,\mathrm{d}k$. Two independent points with law ν_N have overlap with density

$$f_N(r) = \frac{\Gamma(N/2)}{\sqrt{\pi}\,\Gamma((N-1)/2)}(1-r^2)^{(N-3)/2}, \qquad -1 < r < 1. \tag{11.37}$$

For $|r| < \delta_0$, we define the pair intensity

$$J_N(r;h_1,k_1,h_2,k_2) = |S_N|^2\,f_N(r)\,(2\pi a_Q)^{-(N-1)}g_N(r)\,N^2 p_{M(r)}(\zeta)\,\mathbb{E}_{\mathrm{pair}}[D_1D_2], \tag{11.38}$$

where $\zeta = \sqrt{N}(h_1,k_1,h_2,k_2)$ and $\mathbb{E}_{\mathrm{pair}}$ is taken at any pair (u_1,u_2) with overlap r. Note that $(2\pi a_Q)^{-(N-1)}g_N(r)$ is the joint gradient density at zero, by (11.17). By the rotation invariance of the law of H_N, the integrand of (11.35) with $b = \mathbf{1}_A(R(u_1,u_2), H_Q(u_1)/N, K_Q(u_1)/N, H_Q(u_2)/N, K_Q(u_2)/N)$ depends on (u_1,u_2) only through r, and the image of the product surface measure under R is $|S_N|^2\,f_N(r)\,\mathrm{d}r$. Hence, for every Borel set $A \subset (-\delta_0,\delta_0)\times\mathbb{R}^4$, the expected number of ordered pairs of critical points (u_1,u_2) with $(R(u_1,u_2), H_Q(u_1)/N, K_Q(u_1)/N, H_Q(u_2)/N, K_Q(u_2)/N) \in A$ equals $\int_A J_N$. Recalling that $p_N(h,k) = N p_{\Sigma_Q}(\sqrt{N}(h,k))$ and comparing (11.38) with (11.36), we find the exact identity

$$\frac{J_N(r;h_1,k_1,h_2,k_2)}{f_N(r)\,j_N(h_1,k_1)\,j_N(h_2,k_2)} = g_N(r)\,v_N(r;\zeta)\,\frac{\mathbb{E}_{\mathrm{pair}}[D_1D_2]}{\mu_{N-1,N}(k_1)\mu_{N-1,N}(k_2)}, \tag{11.39}$$

with g_N and v_N as in Lemma 11.2. We call this quotient the *relative pair intensity*. It divides out the overlap density $f_N(r)$, which is not close to $f_N(0)$ on the whole range $|r| \le N^{-\alpha}$ when $\alpha < 1/2$.

Pure models. In a pure model of degree p, the marks lie on the graph $k = ph$, and we use the intensities in the energies alone,

$$\begin{aligned} j_N^{\mathrm{pure}}(h) &= |S_N|\,(2\pi a_Q)^{-(N-1)/2}\,p_N^{\mathrm{pure}}(h)\,\mu_{N-1,N}(ph), \\ J_N^{\mathrm{pure}}(r;h_1,h_2) &= |S_N|^2\,f_N(r)\,(2\pi a_Q)^{-(N-1)}g_N(r)\,N\,p_{M(r)}\big(\sqrt{N}(h_1,h_2)\big)\,\mathbb{E}_{\mathrm{pair}}[D_1D_2], \end{aligned} \tag{11.40}$$

where p_N^{pure} is the density of (7.5), $M(r)$ is the reduced 2×2 covariance of Remark 11.3, $k_i = ph_i$, and the factor N is the Jacobian of $(h_1,h_2) \mapsto \sqrt{N}(h_1,h_2)$. By

Theorem 7.17 and Corollary 7.18, j_N^{pure} is the density with respect to $\mathrm{d}h$ of the expected number of critical points with normalized energy in $\mathrm{d}h$, and, as above, for every Borel set $A \subset (-\delta_0, \delta_0) \times \mathbb{R}^2$ the expected number of ordered pairs of critical points with $(R(u_1,u_2), H_Q(u_1)/N, H_Q(u_2)/N) \in A$ equals $\int_A J_N^{\mathrm{pure}} \,\mathrm{d}r\,\mathrm{d}h_1\,\mathrm{d}h_2$. A pure pair has no density in the four marks (h_1,k_1,h_2,k_2); the corresponding measure on (r,h_1,k_1,h_2,k_2) is the image of $J_N^{\mathrm{pure}}\,\mathrm{d}r\,\mathrm{d}h_1\,\mathrm{d}h_2$ under $(r,h_1,h_2) \mapsto (r,h_1,ph_1,h_2,ph_2)$. Since $p_N^{\mathrm{pure}}(h) = \sqrt{N}\, p_{\xi(Q)}(\sqrt{N}h)$, the two one-point factors supply exactly the Jacobian N, and (11.39) holds with $J_N^{\mathrm{pure}}(r;h_1,h_2)/(f_N(r) j_N^{\mathrm{pure}}(h_1) j_N^{\mathrm{pure}}(h_2))$ on the left and the reduced v_N on the right. In the pure case, all subsequent statements use these intensities and these integrals in place of their counterparts for non-pure mixtures. Keeping the factor N^2 of non-pure mixtures after reducing $M(r)$ would instead give a relative pair intensity $N(1+O(N^{-1/2}))$ at $r=0$.

Theorem 11.10 (Factorization of the pair intensity at small overlap)**.** *Let ξ be a finite mixture with minimum degree $d \geq 4$, of either parity, pure or not (for pure models, with the energy-only intensities* (11.40)*, the reduced covariance of Remark 11.3, and marks $(h_i, ph_i) \in \mathcal{K}_0$), and let $1/4 < \alpha < 1/2$.*

(a) Uniformly in $|r| \leq N^{-\alpha}$ and $(h_i,k_i) \in \mathcal{K}_0$,

$$\frac{J_N(r;h_1,k_1,h_2,k_2)}{f_N(r)\, j_N(h_1,k_1)\, j_N(h_2,k_2)} = 1 + O\big(N^{1-4\alpha} + N^{-1/2}\big). \tag{11.41}$$

(b) The same holds if each of the two points is required to satisfy the regularity conditions of Lemma 11.8, with the thresholds λ_* and C_* selected there: that is, with $\mathbb{E}_{\mathrm{pair}}[D_1D_2]$ replaced by $\mathbb{E}_{\mathrm{pair}}[D_1D_2\mathbf{1}_{\mathrm{Reg}}]$ in (11.38)*, and $\mu_{N-1,N}(k)$ replaced by $\mathbb{E}_{h,k,u}[D(u)\mathbf{1}_{\mathrm{Reg}}]$ in* (11.36)*.*

(c) There exist $\delta \in (0,\delta_0]$ and $C > 0$ such that, for $N \geq 12$, $|r| < \delta$, and $(h_i,k_i) \in \mathcal{K}_0$,

$$J_N(r;h_1,k_1,h_2,k_2) \leq C\sqrt{N}\, e^{-Nr^2/4}\, j_N(h_1,k_1)\, j_N(h_2,k_2). \tag{11.42}$$

Proof. We give the proof for a non-pure mixture; by Remark 11.3, the pure case is identical. For (a), we bound the three factors in (11.39) for $|r| \leq N^{-\alpha}$. By (11.17) and $d \geq 4$,

$$|\log g_N(r)| \leq C\big(N^{1-\alpha(2d-2)} + N^{-\alpha(2d-4)}\big) \leq C\big(N^{1-6\alpha} + N^{-4\alpha}\big).$$

Next, (11.18) gives $|\log v_N(r;\zeta)| \leq CN^{1-\alpha d} \leq CN^{1-4\alpha}$, and by Lemma 11.7, the determinant ratio is $1 + O(N^{-\alpha(d-2)} + N^{-1/2}) = 1 + O(N^{-2\alpha} + N^{-1/2})$. Since $\alpha > 1/4$, we have $N^{1-6\alpha} \leq N^{1-4\alpha} \to 0$ and $N^{-4\alpha} \leq N^{-2\alpha} \leq N^{-1/2}$. Multiplying the three factors, we obtain (11.41).

For (b), Lemma 11.8 shows that inserting $\mathbf{1}_{\mathrm{Reg}}$ changes $\mathbb{E}_{\mathrm{pair}}[D_1D_2]$ and each one-point mean $\mu_{N-1,N}(k_i)$ by a factor $1 + O(e^{-cN})$, for $|r| \leq \delta_1$ and N large. For N large we have $N^{-\alpha} \leq \delta_1$, and (b) follows from (a).

For (c), the determinant ratio in (11.39) is bounded for $|r| \le \delta_0$ by Lemma 11.7, and Lemma 11.2 implies that $g_N v_N \le \exp\{C + CNr^2(|r|^{d-2} + |r|^{2d-4})\}$. By the log-convexity of the Gamma function, we have $\Gamma(x+1/2) \le \Gamma(x)^{1/2}\Gamma(x+1)^{1/2} = \sqrt{x}\,\Gamma(x)$. The normalizing constant in (11.37) is therefore at most $\sqrt{N/(2\pi)}$, and since $\log(1-r^2) \le -r^2$, we have

$$f_N(r) \le \sqrt{N}\, e^{-(N-3)r^2/2}.$$

We choose $\delta \le \delta_0$ such that $C(\delta^{d-2} + \delta^{2d-4}) \le 1/8$. For $N \ge 12$, we have $(N-3)/2 \ge 3N/8$, and the exponent is at most $-3Nr^2/8 + Nr^2/8 = -Nr^2/4$ for $|r| < \delta$. This proves (11.42). The proof used only $|r|$, and it applies to both signs of r. □

Corollary 11.11 (Integrated pair bound for retained maxima)**.** *Let $\varepsilon \in (0, 1-\eta)$, and let U, w_ε, G_ε, and m_N be as in (M1)–(M2), with $\mathcal{K}_\varepsilon \subset \mathcal{K}_0$. Let δ be as in Theorem 11.10(c), and fix $L > 0$ and $1/4 < \alpha < 1/2$. Let $\mathcal{A}_N(L)$ be the set of critical points $u \in S_N$ of H_Q with $\bigl|H_Q(u) - m_N\bigr| \le L$ and $K_Q(u)/N \in U$, and let*

$$C_N(L) = \sum_{u \in \mathcal{A}_N(L)} G_\varepsilon(H_N, u), \qquad a_N(L) = \mathbb{E}C_N(L), \qquad b_N(L) = \mathbb{E}\,|\mathcal{A}_N(L)|\,.$$

There exists $N_0(L)$ such that, for $N \ge N_0(L)$,

$$\bigl(1 - Ce^{-cN}\bigr) b_N(L) \le a_N(L) \le b_N(L), \tag{11.43}$$

and the expected number of ordered pairs (u, v) of distinct retained points of $\mathcal{A}_N(L)$ with $|R(u,v)| < \delta$ is at most $(1+\eta_N)a_N(L)^2$, where

$$\eta_N = C\bigl(N^{1-4\alpha} + N^{-1/2} + e^{-N^{1-2\alpha}/4}\bigr). \tag{11.44}$$

The constants C and c do not depend on L.

Proof. Since $m_N/N \to E_0$, there exists $N_0(L)$ such that $|h - E_0| \le w_\varepsilon$ whenever $|Nh - m_N| \le L$ and $N \ge N_0(L)$; then the marks of every point of $\mathcal{A}_N(L)$ lie in $\mathcal{K}_\varepsilon$. Then Theorem 7.17 with the bounded weight G_ε gives

$$a_N(L) = \frac{|S_N|}{(2\pi a_Q)^{(N-1)/2}} \int_{S_N} \int_{|Nh - m_N| \le L} \int_U p_N(h,k)\, \mathbb{E}_{h,k,u}\bigl[D(u) G_\varepsilon(H_N, u)\bigr]\, \mathrm{d}k\, \mathrm{d}h\, \nu_N(\,\mathrm{d}u),$$

and $b_N(L)$ is given by the same formula without G_ε. The loss bound (11.4) is uniform in $u \in S_N$ and in the marks, and it therefore implies (11.43); no rotation invariance of G_ε is needed here. In a pure model, let $W_N(L) = \{h : |Nh - m_N| \le L,\ ph \in U\}$. Then

$$a_N(L) = \frac{|S_N|}{(2\pi a_Q)^{(N-1)/2}} \int_{S_N} \int_{W_N(L)} p_N^{\mathrm{pure}}(h)\, \mathbb{E}_{h,u}\bigl[D(u) G_\varepsilon(H_N, u)\bigr]\, \mathrm{d}h\, \nu_N(\,\mathrm{d}u),$$

$b_N(L)$ is the same expression without G_ε, and the same argument applies.

For the pairs, we discard both selection indicators, which can only increase the nonnegative count. Pairs with $|R(u,v)| < \delta < 1$ are automatically distinct, and their expected number is at most

$$\int_{|r|<\delta} \int J_N(r; h_1, k_1, h_2, k_2)\, \mathrm{d}h_1\, \mathrm{d}k_1\, \mathrm{d}h_2\, \mathrm{d}k_2\, \mathrm{d}r,$$

where the inner integral is over $|Nh_i - m_N| \le L$ and $k_i \in U$; in a pure model, it is $\int_{|r|<\delta} \int_{W_N(L)^2} J_N^{\mathrm{pure}}(r; h_1, h_2)\, \mathrm{d}h_1\, \mathrm{d}h_2\, \mathrm{d}r$, and the estimates below apply without change. On $|r| \le N^{-\alpha}$, (11.41) and $\int f_N(r)\, \mathrm{d}r \le 1$ bound this by $(1 + C(N^{1-4\alpha} + N^{-1/2}))b_N(L)^2$. On $N^{-\alpha} < |r| < \delta$, (11.42) and the substitution $y = \sqrt{N}r$ bound it by

$$C b_N(L)^2 \int_{|y|>N^{1/2-\alpha}} e^{-y^2/4}\, \mathrm{d}y \le 4C\, b_N(L)^2 e^{-N^{1-2\alpha}/4},$$

where we used $\int_{|y|>T} e^{-y^2/4}\, \mathrm{d}y \le (4/T)e^{-T^2/4}$ with $T = N^{1/2-\alpha} \ge 1$. Finally, (11.43) implies that $b_N(L) \le (1 + Ce^{-cN})a_N(L)$, which completes the proof. □

Corollary 11.11 is an upper bound only: the selection indicators were discarded in the pair integrand, and the second moment argument below needs no lower bound for the pair intensity of retained maxima.

11.5 Counts in fixed energy windows

The following bound, for a set of unit vectors with small pairwise inner products, on the number of vectors that have a strongly negative inner product with a given one is the classical simplex bound for spherical codes in a cap. It is the cap form of Rankin's simplex bound [129, Theorem 1(ii) and Lemma 2], and it also follows from the reduction from a cap to the equator and the simplex discussion in [18, Theorem 5 and Section 5, item 3]. Its use here replaces the symmetry of even models under $u \mapsto -u$.

Lemma 11.12 (Vectors at strongly negative inner product)**.** *Let $0 < \delta \le 1$ and $\varepsilon < \delta^2$. Let $v_1, \dots, v_m$ be unit vectors in $\mathbb{R}^N$ with $v_i \cdot v_j \le \varepsilon$ for all $i \ne j$. Then, for every i,*

$$\#\{j \ne i : v_i \cdot v_j \le -\delta\} \le K_{\delta,\varepsilon}, \qquad K_{\delta,\varepsilon} = \frac{1-\varepsilon}{\delta^2 - \varepsilon}. \tag{11.45}$$

In particular, the number of ordered pairs (i,j) with $i \ne j$ and $v_i \cdot v_j \le -\delta$ is at most $K_{\delta,\varepsilon} m$.

Proof. Since $\varepsilon < \delta^2 \le 1$, the constant $K_{\delta,\varepsilon}$ is positive, and there is nothing to prove if no $j \ne i$ satisfies $v_i \cdot v_j \le -\delta$. Otherwise, let J be the set of the $k \ge 1$ indices $j \ne i$ with

$v_i \cdot v_j \le -\delta$, and let $s = \sum_{j \in J} v_j$. Then $v_i \cdot s \le -k\delta$, and by the Cauchy–Schwarz inequality and the pairwise bound, we have

$$k^2\delta^2 \le |s|^2 = k + \sum_{\substack{j,l \in J \\ j \ne l}} v_j \cdot v_l \le k + k(k-1)\varepsilon.$$

Dividing by k and rearranging, we find $k(\delta^2 - \varepsilon) \le 1 - \varepsilon$, which is (11.45). Summing over i gives the bound for ordered pairs. □

We note that the lemma assumes no lower bound on the pairwise inner products, and that exactly antipodal pairs are included in the count.

Theorem 11.13 (Second moment of the number of retained maxima)**.** *Let ξ and β satisfy the standing assumptions of this section, with ξ pure or not. Let δ be as in Theorem 11.10(c), fix $\varepsilon \in (0, \delta^2/2)$ with $\varepsilon < 1 - \eta$, and let U, w_ε, G_ε, and m_N be as in (M1)–(M2), with $\mathcal{K}_\varepsilon \subset \mathcal{K}_0$. Fix $1/4 < \alpha < 1/2$ and $L > 0$, and let $C_N(L)$ and $a_N(L)$ be as in Corollary 11.11. Then*

$$a_N(L) \longrightarrow e^{c_0 L} - e^{-c_0 L}, \tag{11.46}$$

and, for $N \ge N_0(L)$, with η_N as in (11.44)*,*

$$\mathbb{E}C_N(L)^2 \le (1 + \eta_N) a_N(L)^2 + (1 + K_{\delta,\varepsilon}) a_N(L), \tag{11.47}$$

$$\operatorname{Var} C_N(L) \le \eta_N a_N(L)^2 + (1 + K_{\delta,\varepsilon}) a_N(L). \tag{11.48}$$

Proof. By (11.5) with $[a, b] = [-L, L]$, we have $b_N(L) \to e^{c_0 L} - e^{-c_0 L}$, and (11.43) gives (11.46).

We have $C_N(L)^2 = C_N(L) + P_N$, where P_N is the number of ordered pairs (u, v) of distinct retained points of $\mathcal{A}_N(L)$. We split P_N according to whether $R(u, v) \ge \delta$, $|R(u, v)| < \delta$, or $R(u, v) \le -\delta$; these three cases are disjoint and exhaustive. Let $N \ge N_0(L)$, so that the marks of all points of $\mathcal{A}_N(L)$ lie in $\mathcal{K}_\varepsilon$. By (M1)(i), two distinct retained points of $\mathcal{A}_N(L)$ have overlap less than $\varepsilon < \delta$, and the first case does not occur. For the third case, we apply Lemma 11.12 to the unit vectors $u/\sqrt{N}$, for u retained in $\mathcal{A}_N(L)$, whose pairwise inner products are at most $\varepsilon < \delta^2$; recall that $\delta \le \delta_0 < 1$. The number of such pairs is at most $K_{\delta,\varepsilon} C_N(L)$, deterministically. Finally, Corollary 11.11 bounds the expected number of pairs in the second case by $(1 + \eta_N) a_N(L)^2$. Taking expectations, we have

$$\mathbb{E}C_N(L)^2 \le a_N(L) + K_{\delta,\varepsilon} a_N(L) + (1 + \eta_N) a_N(L)^2,$$

which is (11.47). Subtracting $a_N(L)^2$ gives (11.48). □

Observe that no Kac–Rice estimate near the singular overlaps $r = \pm 1$ was used: pairs with $r \ge \delta$ are excluded by the selection event, and pairs with $r \le -\delta$ are counted by Lemma 11.12.

We now state and prove the bound on counts in fixed energy windows under the 1RSB assumptions.

Theorem 11.14 (Critical-point counts in fixed energy windows)**.** *Let ξ be a finite mixture with minimum degree $d \geq 4$, of either parity, and let $\beta > 0$ be such that (ξ, β) is strictly 1RSB with a nondegenerate contact at Q (Definition 5.9), with 1RSB parameters x, Q. Let δ, ε, U, and $m_N = NE_0 - \log N/(2c_0) + O(1)$ be as in Theorem 11.13, and for $M > 0$ let*

$$\mathcal{C}_{N,M} = \big\{u \in S_N : \nabla_{S_N} H_Q(u) = 0,\ \big|H_Q(u) - m_N\big| \leq M,\ K_Q(u)/N \in U\big\},$$

counted as an ordinary set of points; in an even model, both points of an antipodal pair belong to it. Then

$$\sup_N \mathbb{E}\,|\mathcal{C}_{N,M}| < \infty \qquad \text{for every fixed } M > 0, \tag{11.49}$$

$$\lim_{M\to\infty} \liminf_{N\to\infty} \mathbb{P}\Big(\mathcal{C}_{N,M} \neq \varnothing,\ \max_{S_N} H_Q \leq m_N + M\Big) = 1. \tag{11.50}$$

More precisely, let $K = K_{\delta,\varepsilon}$, and let $C_N(L)$ and $a_N(L)$ be as in Theorem 11.13. For every fixed $L > 0$,

$$\limsup_{N\to\infty} \mathbb{P}\big(C_N(L) = 0\big) \leq \frac{1+K}{e^{c_0 L} - e^{-c_0 L}}, \tag{11.51}$$

$$\limsup_{N\to\infty} \mathbb{P}\big(C_N(L) < a_N(L)/2\big) \leq \frac{4(1+K)}{e^{c_0 L} - e^{-c_0 L}}, \tag{11.52}$$

$$\limsup_{N\to\infty} \mathbb{P}\Big(\Big|\max_{S_N} H_Q - m_N\Big| > L\Big) \leq e^{-c_0 L} + \frac{1+K}{e^{c_0 L} - e^{-c_0 L}}. \tag{11.53}$$

Proof. For (11.49), fix M. Applying (11.5) with $[a, b] = [-M, M]$, we see that the expectation $\mathbb{E}\,|\mathcal{C}_{N,M}|$ converges to $e^{c_0 M} - e^{-c_0 M}$; in particular, it is bounded for N large. For each of the finitely many remaining N, the expected number of critical points of H_Q on S_N is finite by Theorem 7.17 and Corollary 7.18.

For (11.51), note that $a_N(L) > 0$ for N large by (11.46). By Chebyshev's inequality and (11.48), we have

$$\mathbb{P}\big(C_N(L) = 0\big) \leq \frac{\operatorname{Var} C_N(L)}{a_N(L)^2} \leq \eta_N + \frac{1+K}{a_N(L)},$$

and (11.51) follows by letting $N \to \infty$ with L fixed, since $\eta_N \to 0$ and $a_N(L) \to e^{c_0 L} - e^{-c_0 L}$. The same argument with the deviation $a_N(L)/2$ proves (11.52). Note that if $\max_{S_N} H_Q < m_N - L$, then no critical point has energy in $[m_N - L, m_N + L]$, and $C_N(L) = 0$. Together with (11.6) and (11.51), this proves (11.53).

Finally, since every retained point of $\mathcal{A}_N(M)$ belongs to $\mathcal{C}_{N,M}$, we see that the event in (11.50) contains $\{C_N(M) \geq 1\} \cap \{\max_{S_N} H_Q \leq m_N + M\}$. Combining (11.51) and (11.6), we obtain

$$\liminf_{N\to\infty} \mathbb{P}\Big(\mathcal{C}_{N,M} \neq \varnothing,\ \max_{S_N} H_Q \leq m_N + M\Big) \geq 1 - e^{-c_0 M} - \frac{1+K}{e^{c_0 M} - e^{-c_0 M}}.$$

The constant K does not depend on M, and letting $M \to \infty$ proves (11.50). □

The order of limits in (11.50) matters: the window half-width M is fixed while $N \to \infty$, and only then sent to infinity. We obtain no estimate for windows whose half-width grows with N, and do not identify the limiting point process of the critical values.

SECTION 12

Mixed models: Gibbs concentration on bands

In this section, we prove Theorem 12.22, the concentration of the Gibbs measure on the bands around the critical points in a fixed energy window. Let (ξ, β) be strictly 1RSB with a nondegenerate contact at Q, as in Definition 1.1; the minimum degree is $d \geq 4$, and the degrees may have either parity. We show that, in expectation, the union of the bands $B_u = \{\sigma \in S_N : |R(\sigma, u) - \sqrt{Q}| \leq b\}$ around the critical points u of the center field H_Q whose energy lies in a fixed window $[m_N - M, m_N + M]$ carries almost all of the Gibbs mass. Here $N \to \infty$ first and $M \to \infty$ afterward. This is the concentration hypothesis of Proposition 4.5, and Section 13 uses it in the proof of Theorem 1.2. For pure models, the same conclusion is proved in Section 9 (Theorem 9.2). The argument below also covers them. We note the changes where the covariance of the value and the radial derivative of H_Q is singular. For generic mixtures that are strictly one-step replica symmetry breaking, Dembo and Subag proved a decomposition of the Gibbs measure into narrow bands around critical points whose energy is maximal up to $o(N)$ [59, Proposition 3.1]; we compare it with Theorem 12.22 in Section 1.6.

We use the following results of earlier sections. From Section 5, we use the static consequences of the 1RSB assumptions (Proposition 5.11), and from Section 6, the constant squared norm of the directing measure (Theorem 6.3). From Section 7, we use the determinant moments and prefactors (Proposition 7.7 and Corollary 7.10) and the marked Kac–Rice formula (Theorem 7.17 and Corollary 7.18). From Sections 10 and 11, we use the centering and the counts in fixed energy windows (Theorems 10.6 and 11.14) and the localization of all high critical points (Propositions 10.10 and 10.16). We deduce the theorem from the reduction in Proposition 4.9, and use the measurability statement of Lemma 4.3. From Section 3, we need the regularity of the field (Lemma 3.2), the concentration of free energies (Lemma 3.3), and the lower tail of normalized partition functions (Lemma 3.19), which is the theorem of Huang, Montanari, and Pham [93, Lemma 3.2 and Remark 3.3]. The only further outside input is Subag's estimates for replicated bands with a fixed number of replicas [143, Lemmas 17–19 and Proposition 24].

Outline. The proof has three steps, one for each hypothesis of Proposition 4.9.

The first step (Section 12.3) shows that the bands around all critical points of H_Q with radial derivative in a fixed interval U and energy at least $N(E_0 - \tau)$ carry almost all of the Gibbs mass, for every fixed $\tau > 0$ (Proposition 12.10). By Theorem 6.3, the

directing measure of every limit of the overlap array is carried by vectors of squared norm Q. This implies that a Gibbs sample σ^0 is typically accompanied, among finitely many further samples, by a group of n samples whose mutual overlaps and overlaps with σ^0 are all close to Q. Using Subag's replicated band estimates, we show that the normalized empirical mean of such a group has energy close to the ground-state energy NE_0 of H_Q. Here n is large but fixed before N, and the squared norm of the empirical mean is close to $Q + (1-Q)/n$, not to Q. A deterministic localization lemma then moves the empirical mean to a nearby local maximum of H_Q, and σ^0 lies in its band.

The second step (Sections 12.4 to 12.6) reduces the coarse family to a fixed window. We compare the sum of the partition functions of the bands around coarse centers with energy below $m_N - M$ with the partition function of a single band in a fixed window, on the common deterministic scale D_N of (12.5). In Proposition 12.15, we show that the partition function of the band around a critical point of energy E has Palm expectation at most $CD_N e^{\beta(E-m_N)+CN(E/N-E_0)^2}$. In Lemma 12.16, we show that the density of critical points at energy $m_N - u$ is at most $Ce^{c_+ u}$ with $c_+ < \beta$. The product of these bounds is summable in u, which gives Corollary 12.17. The upper bound must hold without any power of N, because M is fixed while $N \to \infty$. This is the main technical point of the section. The Palm expectation is the ratio of a Laplace integral in two variables, the radial derivative and the latitude, to a Laplace integral in the radial derivative alone. The latitude width $N^{-1/2}$ that remains in the ratio cancels the growth of the coarea constant $c_N^{\mathrm{lat}} \approx (N/2\pi)^{1/2}$, and the determinant prefactors are bounded by Lemma 12.11.

The third step (Section 12.7) shows that the band around a critical point of energy E has partition function at least $\delta D_N e^{\beta(E-m_N)}$ with Palm probability close to one, uniformly in a fixed window (Proposition 12.20). At the exact latitude $\sqrt{Q}$, the field on the band is the prescribed energy plus a residual field with covariance $N\psi_Q$. Since the residual mixture is strictly replica symmetric with $\beta^2\psi_Q''(0) < 1$, the theorem of Huang, Montanari, and Pham implies that the normalized partition function of this field is tight away from zero. Jensen's inequality over a symmetric latitude interval of width $N^{-1/2}$ passes from the exact latitude to the band. The Palm law weights the field by the absolute Hessian determinant, which is correlated with the band. We transfer the estimates to the Palm law with the normalized second moment of the determinant.

12.1 Notation and inputs

Standing assumptions. Throughout this section, $\xi(r) = \sum_p c_p r^p$ is a finite mixture with $c_p \geq 0$ and minimum degree $d \geq 4$, and $\beta > 0$ is such that (ξ, β) is strictly 1RSB with parameters $x, Q \in (0,1)$ and a nondegenerate contact at Q (Definition 1.1). Recall from Definition 3.1 that the mixture is non-pure if at least two degrees have positive coefficients,

and *pure* if $\xi(r) = c_p r^p$. As in Section 5, we write

$$s = 1 - Q, \qquad C = s + xQ, \qquad z = \frac{xQ}{s}, \qquad y = \frac{Q}{\beta s}, \qquad c_0 = \beta x,$$

$$a_Q = Q\xi'(Q), \qquad b_Q = Q^2\xi''(Q), \qquad \Sigma_Q = \begin{pmatrix} \xi(Q) & a_Q \\ a_Q & a_Q + b_Q \end{pmatrix},$$

and $\psi_Q(t) = \xi(Q+st) - \xi(Q) - s\xi'(Q)t$ for the residual mixture. For a non-pure mixture, Σ_Q is positive definite: with the weights $w_p = c_p Q^p/\xi(Q)$, its determinant is $\xi(Q)^2$ times the variance of the degree under w. We set

$$\rho = \beta^2 s^2 \xi''(Q) = \frac{b_Q}{y^2}, \qquad q(e,k) = (e,k)\,\Sigma_Q^{-1}\begin{pmatrix} e \\ k \end{pmatrix}, \qquad \ell(e,k) = \frac{1}{2}\partial_e q(e,k), \tag{12.1}$$

as in Section 10. That is, $\ell(e,k)$ is the first coordinate of $\Sigma_Q^{-1}(e,k)^{\mathsf{T}}$, and $-\ell(e,k)$ is the derivative in e of the exponent $-\frac{1}{2}q(e,k)$ of the Gaussian density of the value and the radial derivative. Constants $C, c > 0$ depend only on ξ, β, and the fixed parameters introduced below, never on N or on energies and marks in the stated ranges, and their values may change from line to line. The constant $C = s + xQ$ of Definition 5.9 appears only in the identity $a_Q/y^2 = s/C$ below and in the computations in the proof of Lemma 12.13 and after (12.42), where no generic constant occurs. The letter x denotes the 1RSB parameter. In this section, we denote points of S_N by σ.

The center field and its marks. Recall that $H_Q(u) = H_N(\sqrt{Q}\,u)$ for $u \in S_N$, and that the radial derivative is $K_Q(u) = u \cdot \nabla H_Q(u)$, where ∇ is the Euclidean gradient of the polynomial extension of H_Q to $\mathbb{R}^N$. A critical point of H_Q is a point $u \in S_N$ with $\nabla_{S_N} H_Q(u) = 0$. We write $e = H_Q(u)/N$ and $k = K_Q(u)/N$ for the normalized marks of a center u, as in Section 10; in Sections 7 and 11, the normalized energy is denoted h. In a pure model, $K_Q = pH_Q$ identically, by Euler's identity.

Latitudes and bands. Fix $u \in S_N$. Every $\sigma \in S_N \setminus \{\pm u\}$ can be written uniquely as

$$\sigma(a,\omega) = au + (1-a^2)^{1/2}\omega, \qquad a = R(\sigma,u) \in (-1,1), \quad \omega \in S_u^{\perp}, \tag{12.2}$$

where $S_u^{\perp} = \{\omega \in \mathbb{R}^N : \omega \cdot u = 0,\ |\omega|^2 = N\}$. We write $\nu_u^{\perp}$ for the uniform probability measure on the sphere $S_u^{\perp}$. Under ν_N, the latitude a and the point ω are independent, ω has law $\nu_u^{\perp}$, and a has density $c_N^{\mathrm{lat}}(1-a^2)^{(N-3)/2}$ on $(-1,1)$, where

$$c_N^{\mathrm{lat}} = \frac{\Gamma(N/2)}{\sqrt{\pi}\,\Gamma((N-1)/2)}, \qquad \lim_{N\to\infty} \frac{c_N^{\mathrm{lat}}}{\sqrt{N}} = \frac{1}{\sqrt{2\pi}}. \tag{12.3}$$

For a band width $b > 0$, we set $I_b = [\sqrt{Q} - b, \sqrt{Q} + b]$ and $B_u = \{\sigma \in S_N : R(\sigma, u) \in I_b\}$. For a Borel set $B \subset S_N$, we write $Z_H(B) = \int_B e^{\beta H_N} \, \mathrm{d}\nu_N$, so that $Z_{N,\beta} = Z_H(S_N)$ and $\pi(B) = Z_H(B)/Z_{N,\beta}$. By the coarea formula, we have

$$Z_H(B_u) = c_N^{\mathrm{lat}} \int_{I_b \cap (-1,1)} (1-a^2)^{(N-3)/2} Z_u(a) \, \mathrm{d}a, \qquad Z_u(a) = \int e^{\beta H_N(\sigma(a,\omega))} \, \nu_u^{\perp}(\, \mathrm{d}\omega). \tag{12.4}$$

We also define the deterministic reference scale

$$D_N = e^{\beta m_N} \frac{c_N^{\mathrm{lat}}}{\sqrt{N}} s^{(N-3)/2} e^{N\beta^2 \psi_Q(1)/2}, \tag{12.5}$$

where m_N is the centering of Theorem 11.14.

Static inputs. We use the following consequences of the 1RSB assumptions, which are proved in Proposition 5.11. The ground-state parameters are

$$E_0 = \frac{a_Q + z\xi(Q)}{y}, \qquad K_0 = y + \frac{b_Q}{y}, \tag{12.6}$$

the stationarity relation $y^2 = (1+z)a_Q$ holds, and $\rho < 1$. Note that the last inequality implies that $y > \sqrt{b_Q}$ and $K_0 > 2\sqrt{b_Q}$, and that the stationarity relation gives $a_Q/y^2 = 1/(1+z) = s/C$. For a non-pure mixture,

$$\Sigma_Q^{-1} \begin{pmatrix} E_0 \\ K_0 \end{pmatrix} = \begin{pmatrix} c_0 \\ 1/y \end{pmatrix}; \tag{12.7}$$

indeed, $\Sigma_Q(c_0, 1/y)^{\mathsf{T}} = ((z\xi(Q) + a_Q)/y, ((1+z)a_Q + b_Q)/y)^{\mathsf{T}} = (E_0, K_0)^{\mathsf{T}}$ by $c_0 = z/y$ and the stationarity relation. The residual mixture is strictly replica symmetric in the sense of Definition 3.16,

$$\beta^2 \psi_Q(t) + t + \log(1-t) < 0 \qquad (0 < t < 1), \tag{12.8}$$

and the free energy satisfies the TAP identity

$$F(\beta) = \beta E_0 + F_{\mathrm{band}}, \qquad F_{\mathrm{band}} = \frac{1}{2} \log s + \frac{\beta^2}{2} \psi_Q(1). \tag{12.9}$$

Finally $c_0 = \beta x < \beta$, since $x < 1$.

Counts and localization. Let $\widetilde{U}$ and $m_N = NE_0 - \log N/(2c_0) + O(1)$ be the open interval and the centering of Theorem 11.14. They are chosen in Corollary 10.17, and $\overline{\widetilde{U}} \subset U_0$, where U_0 is the base interval of Section 10. For a closed interval U and a Borel set $I \subset \mathbb{R}$, we write

$$\begin{gathered} \mathcal{C}(I) = \mathcal{C}_U(I) = \{u \in S_N : \nabla_{S_N} H_Q(u) = 0, \ H_Q(u) \in I, \ K_Q(u)/N \in U\}, \\ \mathcal{C}_{N,M} = \mathcal{C}([m_N - M, m_N + M]). \end{gathered} \tag{12.10}$$

Let $\kappa > 0$ and $C_3 = C_{\mathrm{reg}}$ be the Hessian margin and the derivative bound of Section 10. By Proposition 10.10, and by Proposition 10.16(d) in a pure model, applied to the interior of a closed interval $U \subset \widetilde{U}$ with K_0 in its interior, there exist $\delta_U, c > 0$ such that, with probability at least $1 - Ce^{-cN}$,

$$K_Q(u)/N \in U \quad \text{and} \quad \nabla^2_{S_N} H_Q(u) \preceq -\kappa \,\mathrm{Id} \qquad \text{if } \nabla_{S_N} H_Q(u) = 0,\ H_Q(u) \geq N(E_0 - \delta_U), \tag{12.11}$$

and $\sup_{S_N} \|\nabla^3_{S_N} H_Q\|_{\mathrm{op}} \leq C_3 N^{-1/2}$.

Determinants and the Kac–Rice formula. Fix $u \in S_N$. For normalized marks (e,k), we write $\mathbb{P}_{e,k,u}$ and $\mathbb{E}_{e,k,u}$ for the canonical conditional law of the whole field H_N given $(H_Q(u), K_Q(u), \nabla_{S_N} H_Q(u)) = (Ne, Nk, 0)$, as in Section 7. In a pure model, the conditioning is on $(H_Q(u), \nabla_{S_N} H_Q(u)) = (Ne, 0)$, and $k = pe$. Under $\mathbb{P}_{e,k,u}$, the intrinsic Hessian is $\nabla^2_{S_N} H_Q(u) = W - k\,\mathrm{Id}$, where W has the law of $W_{N-1,N} = (b_Q(N-1)/N)^{1/2} G_{N-1}$ and G_n is a normalized $n \times n$ GOE matrix (off-diagonal variance $1/n$, diagonal variance $2/n$). We set

$$D_u = |\det \nabla^2_{S_N} H_Q(u)|, \qquad d_N(k) = \mu_{N-1,N}(k) = \mathbb{E}|\det(W_{N-1,N} - k\,\mathrm{Id})|; \tag{12.12}$$

for notational convenience, we abbreviate the mean determinant $\mu_{N-1,N}(k)$ of Section 7 by $d_N(k)$ (it is $d_N^{\mathrm{all}}(k)$ in Section 10). By Proposition 7.7, for every compact set $\mathcal{K} \subset (2\sqrt{b_Q}, \infty)$ and every $r \geq 1$, there exists $\mathsf{K}_r < \infty$ such that $\mathbb{E}_{e,k,u}[D_u^r] \leq \mathsf{K}_r d_N(k)^r$ for all $k \in \mathcal{K}$, all e, and all large N.

Definition 12.1 (Palm laws and the intensity)**.** Let $U \subset (2\sqrt{b_Q}, \infty)$ be a nondegenerate closed interval and fix $u \in S_N$. For marks (e,k) with $k \in U$ and a nonnegative Borel function Y of the field, the *Palm expectation* is, as in Remark 7.19,

$$\mathbb{E}^{\mathrm{Palm}}_{e,k,u}[Y] = \frac{\mathbb{E}_{e,k,u}[D_u Y]}{d_N(k)}.$$

For a non-pure mixture and $E \in \mathbb{R}$, we set $w_N(E,k) = e^{-Nq(E/N,k)/2} d_N(k)$ for $k \in U$, and define the *Palm expectation at energy E* and the *intensity of all critical points* by

$$\mathbb{E}^{\mathrm{Palm}}_{E,u}[Y] = \frac{\int_U w_N(E,k)\, \mathbb{E}^{\mathrm{Palm}}_{E/N,k,u}[Y]\, \mathrm{d}k}{\int_U w_N(E,k)\, \mathrm{d}k}, \qquad \lambda_N^{\mathrm{all}}(E) = \mathfrak{c}_N \int_U w_N(E,k)\, \mathrm{d}k,$$

where

$$\mathfrak{c}_N = \frac{|S_N|}{(2\pi a_Q)^{(N-1)/2}\, 2\pi (\det \Sigma_Q)^{1/2}}, \qquad |S_N| = \frac{2\pi^{N/2} N^{(N-1)/2}}{\Gamma(N/2)},$$

and $|S_N|$ is the surface area of S_N. For a pure mixture, $\mathbb{E}^{\mathrm{Palm}}_{E,u} = \mathbb{E}^{\mathrm{Palm}}_{E/N,pE/N,u}$ and

$$\lambda_N^{\mathrm{all}}(E) = |S_N| (2\pi a_Q)^{-(N-1)/2} (2\pi N \xi(Q))^{-1/2} e^{-E^2/(2N\xi(Q))} d_N(pE/N) \mathbf{1}\{pE/N \in U\}.$$

Let $Y(H,u) \geq 0$ be a Borel function which is rotation covariant, in the sense that $Y(H \circ g^{-1}, gu) = Y(H,u)$ for $g \in O(N)$. Since the law of H_N is invariant under rotations (Lemma 3.2), the Palm expectations of $Y(\cdot,u)$ do not depend on u. Applying Theorem 7.17 to the bounded field-dependent weights $Y \wedge j$, changing variables $E = Ne$, and letting $j \to \infty$, we obtain, by monotone convergence,

$$\mathbb{E} \sum_{u \in \mathcal{C}(I)} Y(H,u) = \int_I \lambda_N^{\mathrm{all}}(E)\, \mathbb{E}_{E,u}^{\mathrm{Palm}}[Y(\cdot,u)]\, \mathrm{d}E \tag{12.13}$$

for every Borel set $I \subset \mathbb{R}$. With $Y = 1$, this is the count formula of Corollary 7.18: $\mathbb{E}|\mathcal{C}(I)| = \int_I \lambda_N^{\mathrm{all}}(E)\, \mathrm{d}E$. Thus λ_N^{all} is the density, in the energy variable, of the expected number of critical points with radial derivative in U. The density λ_N of Section 10 counts only nondegenerate local maxima. All weights used below, such as $Z_H(B_u)$ and indicators of events defined by overlaps with u and by $H_Q(u)$, are rotation covariant, and their measurability is checked in Lemma 12.9.

The interval U. We fix once and for all a closed interval $U \subset \widetilde{U}$, with K_0 in its interior, and numbers $b_0, \varepsilon_0 > 0$, chosen as in Section 12.5. The interval U is small enough for (12.11), and $\min U > 2\sqrt{b_Q}$. We write $\mathcal{K}_* = [E_0 - \varepsilon_0, E_0 + \varepsilon_0] \times U$ for the rectangle of marks on which the Palm estimates hold. In a pure model, $\mathcal{K}_*$ stands for the compatible marks $\{(e, pe) : |e - E_0| \leq \varepsilon_0\}$. The following lemma transfers the counts of Theorem 11.14 from $\widetilde{U}$ to U, with the same centering m_N.

Lemma 12.2 (Counts in the smaller interval)**.** *For every fixed $M > 0$, we have $\sup_N \mathbb{E}|\mathcal{C}_{N,M}| < \infty$, and*

$$\lim_{M\to\infty} \liminf_{N\to\infty} \mathbb{P}\Big(\mathcal{C}_{N,M} \neq \varnothing,\ \max_{S_N} H_Q \leq m_N + M\Big) = 1. \tag{12.14}$$

Proof. The set $\mathcal{C}_{N,M}$ is contained in the corresponding set for $\widetilde{U}$, and the first claim follows from Theorem 11.14. For the second, fix M. As $m_N/N \to E_0$, we have $m_N - M \geq N(E_0 - \delta_U)$ for all large N. On the event in (12.11), every critical point u with $|H_Q(u) - m_N| \leq M$ has $K_Q(u)/N \in U$, and the sets $\mathcal{C}_{N,M}$ defined with U and with $\widetilde{U}$ coincide. This event has probability at least $1 - Ce^{-cN}$, and (12.14) follows from the corresponding statement in Theorem 11.14. □

Parameters. Table 12.1 lists the parameters of this section and the order in which they are chosen. All auxiliary parameters are fixed before $N \to \infty$, and every window half-width M or L is fixed during this limit. The coarse cutoff $N(E_0 - \tau)$ and the rectangle $\mathcal{K}_*$ of normalized marks define energy ranges of order N.

Parameter	Role	Chosen
U, b_0, ε_0	radial interval, largest band width, energy window of the Palm estimates	first, in Section 12.5
$b \in (0, b_0]$	band width	after b_0
τ	lower energy cutoff $N(E_0 - \tau)$ of the coarse family	after U, ε_0 (Corollary 12.17)
$\varepsilon, n, \theta, J$	localization tolerance, replica number, overlap tolerance, number of additional replicas	after b and τ, inside the proof of Proposition 12.10
L, δ, M	half-width of the existence window, lower threshold for band partition functions, half-width of the final window	in this order, after all of the above (Proposition 4.9)

Table 12.1. The parameters of Section 12 and the order of their choice.

12.2 Regularity and regression at a center

We realize H_N by independent Gaussian tensors, as in Lemma 3.2:

$$H_N(v) = \sum_p \sqrt{c_p}\, N^{-(p-1)/2}\, T_p[v, \dots, v], \qquad v \in \mathbb{R}^N, \tag{12.15}$$

where T_p is the p-linear form whose entries are the independent standard Gaussian variables $g^{(p)}_{i_1 \dots i_p}$ of Lemma 3.2, indexed by ordered p-tuples. The right side is the polynomial extension of H_N to $\mathbb{R}^N$, with covariance $\mathbb{E} H_N(v) H_N(v') = N\xi(v \cdot v'/N)$. For a p-linear form T, we write $\|T\|_{\mathrm{op}} = \sup\{T[v_1, \dots, v_p] : |v_1| = \cdots = |v_p| = 1\}$.

Lemma 12.3 (Regularity under the conditioning at a critical point)**.** *Let $\mathcal{K}_0 \subset \mathbb{R}^2$ be a compact set of marks (in a pure model, of compatible marks (e, pe)), and let $I \subset (-1, 1)$ be a compact interval. There exist $C, c > 0$ such that, for every $u \in S_N$ and every $(e, k) \in \mathcal{K}_0$,*

$$\mathbb{P}_{e,k,u}\Big(\sup_{a \in I,\ \omega \in S_u^\perp} \big|\partial_a^2 H_N(\sigma(a, \omega))\big| > CN \Big) \le Ce^{-cN}. \tag{12.16}$$

Moreover, the Lipschitz constant Lip_N of H_N on the ball $\{v \in \mathbb{R}^N : |v| \le \sqrt{N}\}$ satisfies $\mathbb{P}(\mathrm{Lip}_N > C\sqrt{N}) \le Ce^{-cN}$ and $\mathbb{E}\,\mathrm{Lip}_N \le C\sqrt{N}$.

Proof. We begin with a net bound. Let T be a p-linear form on $\mathbb{R}^N$ and let $\mathcal{N}_p$ be a $1/(2p)$-net of the unit sphere with $|\mathcal{N}_p| \le (1 + 4p)^N$. Replacing the arguments of T by net points one at a time changes $T[v_1, \dots, v_p]$ by at most $p \cdot (2p)^{-1} \|T\|_{\mathrm{op}}$. This

gives $\|T\|_{\mathrm{op}} \le 2\max_{\mathcal{N}_p^p} |T|$. If every evaluation $T[v_1,\dots,v_p]$ at unit vectors is a centered Gaussian variable with variance at most one, then the union bound implies that

$$\mathbb{P}\big(\|T\|_{\mathrm{op}} > 2t\big) \le 2(1+4p)^{pN} e^{-t^2/2}, \qquad t > 0. \tag{12.17}$$

For T_p, every such evaluation is standard Gaussian. Taking $t = C_p\sqrt{N} + t'$ with C_p large in (12.17) and integrating the tail in t', we obtain $\mathbb{P}(\|T_p\|_{\mathrm{op}} > C\sqrt{N}) \le Ce^{-cN}$ and $\mathbb{E}\|T_p\|_{\mathrm{op}} \le C\sqrt{N}$. The gradient of the degree-p term of (12.15) at $|v| \le \sqrt{N}$ has norm at most $p\sqrt{c_p}\,\|T_p\|_{\mathrm{op}}$, and the assertions on Lip_N follow, since there are finitely many degrees. (The bound in probability is also part of Lemma 3.2.)

Under $\mathbb{P}_{e,k,u}$, the vector of all tensor entries is Gaussian. By the explicit form of the canonical conditional law in Section 7, its centered part is the image of a standard Gaussian vector under an orthogonal projection, and its covariance is dominated by the identity. Then every evaluation of the centered conditioned forms T_p^{c} at unit vectors has variance at most one, and (12.17) implies that $\|T_p^{\mathrm{c}}\|_{\mathrm{op}} \le C\sqrt{N}$ for all p, except on an event of probability Ce^{-cN}. Let $\hat\sigma(a) = \sigma(a,\omega)/\sqrt{N}$. Then $|\hat\sigma(a)| = 1$, and $|\hat\sigma'(a)|$ and $|\hat\sigma''(a)|$ are bounded on I, uniformly in ω. The centered conditioned field at $\sigma(a,\omega)$ equals $\sum_p \sqrt{c_p}\sqrt{N}\,T_p^{\mathrm{c}}[\hat\sigma(a),\dots,\hat\sigma(a)]$. Differentiating twice in a, we find that its second latitude derivative is at most $C\sqrt{N}\sum_p \|T_p^{\mathrm{c}}\|_{\mathrm{op}} \le CN$ on the preceding event. The conditional mean of the field at $\sigma(a,\omega)$ is $Nm_{e,k}(a)$, with $m_{e,k}$ given in Lemma 12.4 below. It depends only on a, and $|m_{e,k}''|$ is bounded on $I \times \mathcal{K}_0$. Adding the mean proves (12.16). □

The conditional mean must be bounded separately: the domination of the conditional covariance by the identity controls only the centered part of the field.

We now describe the field on the latitude spheres around a fixed center. The conditional law of the field on a band of fixed latitude, given the value, the radial derivative, and a vanishing gradient at the center, was computed by Huang and Sellke [90, Lemma 3.11]; see also [30] for the one-point law at a critical point. We need, in addition, the decomposition of the field by the number of tangential coordinates, which shows how the band is correlated with the Hessian at the center. For $a \in (-1,1)$, let

$$b(a) = \begin{pmatrix} \xi(\sqrt{Q}\,a) \\ \sqrt{Q}\,a\,\xi'(\sqrt{Q}\,a) \end{pmatrix}, \qquad d(a) = \frac{(1-a^2)\,\xi'(\sqrt{Q}\,a)^2}{\xi'(Q)}, \qquad v_2(a) = \frac{1}{2}(1-a^2)^2\xi''(a^2). \tag{12.18}$$

For a non-pure mixture, let $m_{e,k}(a) = b(a)^{\mathsf{T}}\Sigma_Q^{-1}(e,k)^{\mathsf{T}}$ and $v(a) = \xi(1) - b(a)^{\mathsf{T}}\Sigma_Q^{-1}b(a) - d(a)$. For a pure mixture $\xi(r) = c_p r^p$, let $m_{e,k}(a) = e(a/\sqrt{Q})^p$ and $v(a) = c_p\{1 - a^{2p} - pa^{2p-2}(1-a^2)\}$.

Lemma 12.4 (Decomposition by tangential degree and Gaussian regression)**.** *Fix $u \in S_N$ and marks (e,k). Under $\mathbb{P}_{e,k,u}$, for $a \in (-1,1)$ and $\omega \in S_u^{\perp}$,*

$$H_N(\sigma(a,\omega)) = Nm_{e,k}(a) + Y_{\le 1}(a,\omega) + Y_2(a,\omega) + Y_{\ge 3}(a,\omega), \tag{12.19}$$

where the following hold.

1. *The processes $Y_{\le 1}$ and $Y_{\ge 3}$ and the pair (Y_2, W) are independent and centered Gaussian, where $\nabla^2_{S_N} H_Q(u) = W - k\,\mathrm{Id}$ and W has the law of $W_{N-1,N}$.*
2. *For all a and $\omega, \omega' \in S_u^\perp$, we have $\mathbb{E}\, Y_2(a,\omega) Y_2(a,\omega') = N v_2(a) R(\omega,\omega')^2$ and*

$$\mathrm{Var}\; Y_{\le 1}(a,\omega) + \mathrm{Var}\; Y_{\ge 3}(a,\omega) = N\big(v(a) - v_2(a)\big).$$

   In particular, $H_N(\sigma(a,\omega))$ has conditional mean $N m_{e,k}(a)$ and conditional variance $N v(a)$.
3. *At the latitude $a = \sqrt{Q}$, we have $Y_{\le 1}(\sqrt{Q}, \cdot) = 0$. Hence*

$$H_N(\sqrt{Q}\,u + \sqrt{s}\,\omega) = Ne + G_u^{\mathrm{res}}(\omega), \qquad \mathbb{E}\, G_u^{\mathrm{res}}(\omega) G_u^{\mathrm{res}}(\omega') = N\psi_Q(R(\omega,\omega')), \tag{12.20}$$

   where $G_u^{\mathrm{res}} = Y_2(\sqrt{Q}, \cdot) + Y_{\ge 3}(\sqrt{Q}, \cdot)$, and the law of G_u^{res} does not depend on (e,k).

Finally, $2\beta^2 v_2(\sqrt{Q}) = \rho < 1$.

Proof. By rotation invariance, we may take $u = \sqrt{N} e_1$. In (12.15), group the entries of T_p according to the number j of positions in the index tuple that differ from 1, and let $\mathcal{Y}_{p,j}$ be the corresponding part of the degree-p term, which we call its component of tangential degree j. The processes $\mathcal{Y}_{p,j}$ are independent, because they involve disjoint families of entries. For $\sigma = \sigma(a,\omega)$ and $\sigma' = \sigma(a',\omega')$, counting the choices of the j positions, we find

$$\mathbb{E}\, \mathcal{Y}_{p,j}(\sigma) \mathcal{Y}_{p,j}(\sigma') = N c_p \binom{p}{j} (aa')^{p-j} \big((1-a^2)(1-a'^2)\big)^{j/2} R(\omega,\omega')^j. \tag{12.21}$$

Let $\mathcal{Y}_j = \sum_p \mathcal{Y}_{p,j}$. The value $H_Q(u)$ and the radial derivative $K_Q(u)$ are functionals of the components $\mathcal{Y}_{p,0}$, the tangential gradient $\nabla_{S_N} H_Q(u)$ is a linear functional of the components $\mathcal{Y}_{p,1}$, and the Euclidean tangential Hessian of H_Q at u is a linear functional of the components $\mathcal{Y}_{p,2}$. Its law is that of $W_{N-1,N}$, and the intrinsic Hessian is obtained by subtracting $(K_Q(u)/N)\,\mathrm{Id} = k\,\mathrm{Id}$ (see Section 7). Then the conditioning leaves the components with $j \ge 2$ unchanged, and it replaces $\mathcal{Y}_0 + \mathcal{Y}_1$ by its conditional mean plus a centered Gaussian process $Y_{\le 1}$ independent of the other components. We set $Y_2 = \mathcal{Y}_2$ and $Y_{\ge 3} = \sum_{j \ge 3} \mathcal{Y}_j$. This proves (1).

To identify the conditional mean and variance, we compute the covariances of $H_N(\sigma(a,\omega))$ with the conditioned variables. Recall that $(H_Q(u), K_Q(u))$ has covariance matrix $N\Sigma_Q$. Since $R(\sigma(a,\omega), \sqrt{Q}\,u) = \sqrt{Q}\,a$, differentiating $N\xi(tR(\sigma, \sqrt{Q}u))$ at $t = 1$, we find that the covariance of $H_N(\sigma(a,\omega))$ with $(H_Q(u), K_Q(u))$ is $Nb(a)$. For a unit tangent vector w at u, the covariance with $w \cdot \nabla H_Q(u)$ is $\sqrt{Q}\,\xi'(\sqrt{Q}\,a)(1-a^2)^{1/2}\,\omega \cdot w$, and the tangential gradient has covariance $a_Q\,\mathrm{Id}$. The sum of the squares of these covariances over an orthonormal basis of tangent vectors, divided by a_Q, is $Nd(a)$. By Gaussian regression, we have that the conditional mean is $Nb(a)^{\mathsf{T}} \Sigma_Q^{-1}(e,k)^{\mathsf{T}} = N m_{e,k}(a)$ and that

the conditional variance is $N\{\xi(1) - b(a)^{\mathsf{T}}\Sigma_Q^{-1}b(a) - d(a)\} = Nv(a)$. In a pure model, the same computation with the single observation $H_Q(u)$ yields the mean $Ne(a/\sqrt{Q})^p$ and, since $d(a) = c_p p(1-a^2)a^{2p-2}$, the variance $Nv(a)$. Summing (12.21) over p with $j = 2$ and $a = a'$ gives $Nv_2(a)R(\omega,\omega')^2$, since $\sum_p c_p\binom{p}{2}t^{p-2} = \xi''(t)/2$. The variance of $Y_{\le 1} + Y_{\ge 3}$ is the difference $N(v(a) - v_2(a))$, by independence. This proves (2).

At $a = \sqrt{Q}$, the component $\mathcal{Y}_0(\sqrt{Q},\cdot) = \sum_p H_{N,p}(\sqrt{Q}\,u)$ equals $H_Q(u) = Ne$. The component $\mathcal{Y}_1$ is the part of the polynomial that is linear in the tangential component, so $\mathcal{Y}_1(\sqrt{Q},\omega) = \sqrt{s}\,\omega\cdot\nabla H_N(\sqrt{Q}\,u) = (s/Q)^{1/2}\,\omega\cdot\nabla_{S_N}H_Q(u)$, which vanishes under the conditioning, so that $Y_{\le 1}(\sqrt{Q},\cdot) = 0$. By (12.21), the covariance of $\sum_{j\ge 2}\mathcal{Y}_j$ at $a = a' = \sqrt{Q}$ is

$$N\sum_{j\ge 2}\sum_p c_p\binom{p}{j}Q^{p-j}s^jR(\omega,\omega')^j = N\sum_{j\ge 2}\frac{s^j\xi^{(j)}(Q)}{j!}R(\omega,\omega')^j = N\psi_Q(R(\omega,\omega')),$$

which proves (3). Since the components with $j \ge 2$ are not affected by the conditioning, the law of G_u^{res} does not depend on (e,k). Note that G_u^{res} is correlated with the Hessian through Y_2. Finally, $2\beta^2 v_2(\sqrt{Q}) = \beta^2 s^2\xi''(Q) = \rho$. □

12.3 Fixed replicas and the coarse cover

In this section, we prove Proposition 12.10. We first state Subag's estimates for replicated bands in our normalization. For $m \in \mathbb{R}^N$ with $|m|^2 = NQ$, an integer $n \ge 1$, and $\alpha,\theta > 0$, let

$$\mathcal{A}(m;n,\alpha,\theta) = \big\{(\sigma^1,\dots,\sigma^n) \in S_N^n :\ |R(\sigma^i,m) - Q| < \alpha,\ \ |R(\sigma^i,\sigma^j) - Q| < \theta\ (i \ne j)\big\},$$

and define the replicated free energy

$$\widetilde{F}_N(m;n,\alpha,\theta) = \frac{1}{Nn}\log\int_{\mathcal{A}(m;n,\alpha,\theta)}\exp\Big\{\beta\sum_{i=1}^n\big(H_N(\sigma^i) - H_N(m)\big)\Big\}\,\nu_N^{\otimes n}(\,\mathrm{d}\sigma^1\cdots\mathrm{d}\sigma^n). \tag{12.22}$$

In the notation of [143], this is $\widetilde{F}_{N,\beta}(m,n,\rho,\delta)$ with $\rho = \theta$ and $\delta = \alpha/\sqrt{Q}$: Subag's band of width δ around m is the set of σ with $|R(\sigma,m) - Q| \le \delta\sqrt{Q}$. The boundaries of these sets have measure zero, so strict and non-strict inequalities give the same free energy. Subag's results hold for general mixtures with nonnegative coefficients, and impose no parity condition. For pure models, the same input is Proposition 9.13, written for $\widetilde{P}_N(m,n,\rho,\delta)$ with $\rho = \theta$ and $\delta = \alpha/\sqrt{Q}$, and Lemmas 12.6 to 12.8 and Proposition 12.10 below are the counterparts of Lemmas 9.11 and 9.14 and Propositions 9.15 and 9.16.

Theorem 12.5 ([143, Proposition 24])**.** *For all $t, c > 0$, there exist $\alpha_0,\theta_0 > 0$ and an integer n_0 such that, for all fixed $n \ge n_0$, $\alpha \in (0,\alpha_0]$, and $\theta \in (0,\theta_0]$, and all sufficiently large N,*

$$\mathbb{P}\Big(\sup_{|m|^2=NQ}\big(\widetilde{F}_N(m;n,\alpha,\theta) - \mathbb{E}\widetilde{F}_N(m;n,\alpha,\theta)\big) < t\Big) > 1 - e^{-cN}. \tag{12.23}$$

The statement in [143] is made for every N. When $N \le n$, the set $\mathcal{A}(m; n, \alpha, \theta)$ can be empty: for $N = 2$, $Q = 0.9$, and $\alpha = \theta = 0.001$, the band condition confines each σ^i to one of two short arcs, two of any three points lie on the same arc, and their overlap exceeds 0.9999, which is larger than $Q + \theta$. Then both sides of (12.23) involve $-\infty$. For $N \ge n + 1$, there exist orthogonal vectors $\omega_1, \dots, \omega_n$ in $S_u^\perp$, where $u = m/\sqrt{Q}$. The points $\sigma^i = \sqrt{Q}\, u + \sqrt{s}\, \omega_i$ satisfy $R(\sigma^i, m) = Q$ and $R(\sigma^i, \sigma^j) = Q$, and $\mathcal{A}(m; n, \alpha, \theta)$ contains an open neighborhood of this tuple, which has positive measure. This restriction is harmless, because n is fixed and only large N matter below.

The next lemma is the bound on the expectation in (12.23) that we need. It holds for each fixed n. We derive it from [143, Lemmas 18 and 19], rather than from a limit in which n grows with N.

Lemma 12.6 (Expected replicated free energy)**.** *The expectation $\mathbb{E}\widetilde{F}_N(m; n, \alpha, \theta)$ does not depend on m with $|m|^2 = NQ$. There exist $C < \infty$ and $\alpha_1 > 0$ such that, for every integer $n \ge 1$ and all $\alpha \in (0, \alpha_1]$ and $\theta > 0$,*

$$\limsup_{N \to \infty} \mathbb{E}\widetilde{F}_N(m; n, \alpha, \theta) \le F_{\mathrm{band}} + C\sqrt{\alpha} + C\Big(\frac{1}{n} + \frac{\theta + C\sqrt{\alpha}}{s}\Big)^{1/2}. \tag{12.24}$$

Proof. The expectation does not depend on m, by the rotation invariance of the law of H_N. We write $m = \sqrt{Q}\, u$ and $\delta = \alpha/\sqrt{Q}$. The exact band $\{\sigma : R(\sigma, m) = Q\}$ is the sphere $\{\sqrt{Q}\, u + \sqrt{s}\, \omega : \omega \in S_u^\perp\}$, with uniform probability measure the image of $\nu_u^\perp$. For $\theta' > 0$, let $\widetilde{F}_N^\circ(m; n, \theta')$ be defined as in (12.22), with ν_N replaced by this measure and with only the constraints $|R(\sigma^i, \sigma^j) - Q| < \theta'$.

We first pass to the exact band. Subag projects each point of the band $\{|R(\sigma, m) - Q| < \alpha\}$ along its meridian to the exact band. The projection moves points by at most $(2N\delta)^{1/2}$, changes overlaps by at most $(8\delta)^{1/2}$, and maps the uniform measure on the band to the uniform measure on the exact band. With this projection, the argument of [143, Lemma 18, Eq. (2.2)] gives

$$\widetilde{F}_N(m; n, \alpha, \theta) \le \widetilde{F}_N^\circ\big(m; n, \theta + (8\delta)^{1/2}\big) + \frac{1}{N} \log \nu_N\big(|R(\sigma, m) - Q| < \alpha\big) + \beta(2\delta)^{1/2} \frac{\mathrm{Lip}_N}{\sqrt{N}}. \tag{12.25}$$

Indeed, the lemma in [143] is stated on the event $\mathrm{Lip}_N \le L\sqrt{N}$, with L in place of $\mathrm{Lip}_N/\sqrt{N}$, and its proof uses the Lipschitz bound only through $|H_N(\sigma) - H_N(\sigma')| \le \mathrm{Lip}_N |\sigma - \sigma'|$. It therefore gives (12.25) for every field. We have $\mathbb{E}\,\mathrm{Lip}_N \le C\sqrt{N}$ by Lemma 12.3; the corresponding bound in probability is [143, Lemma 17].

Next, the band condition reads $|a - \sqrt{Q}| < \delta$ in terms of the latitude $a = R(\sigma, u)$. By (12.3), its probability is at most $2\delta c_N^{\mathrm{lat}} \max(1 - a^2)^{(N-3)/2}$ over $|a - \sqrt{Q}| < \delta$, and the maximum of $\frac{1}{2}\log(1 - a^2)$ over this interval differs from $\frac{1}{2}\log s$ by at most $C\alpha$ for small α. We obtain

$$\frac{1}{N} \log \nu_N\big(|R(\sigma, m) - Q| < \alpha\big) \le \frac{1}{2} \log s + C\alpha + O\Big(\frac{\log N}{N}\Big).$$

On the exact band, we have $R(\sigma^i,\sigma^j) = Q + sR(\omega_i,\omega_j)$, and the constraint becomes $|R(\omega_i,\omega_j)| < \theta'/s$. The field $\omega \mapsto H_N(\sqrt{Q}\,u + \sqrt{s}\,\omega) - H_N(m)$ is centered Gaussian with covariance $N\{\xi(Q+sR(\omega,\omega')) - \xi(Q)\} = N\{s\xi'(Q)R(\omega,\omega') + \psi_Q(R(\omega,\omega'))\}$. It has the law of the sum of two independent fields, a linear field with covariance $Ns\xi'(Q)R$ and a field h^{res} with covariance $N\psi_Q(R)$. After identifying $S_u^\perp$ with the sphere of radius $\sqrt{N}$ in $\mathbb{R}^{N-1}$, as in [143, Eq. (2.4)], [143, Lemma 19, Eq. (2.5)] removes the linear field: with the one-spin coefficient $\gamma_1^2 = Ns\xi'(Q)/(N-1)$, the expectation of $\widetilde{F}_N^\circ(m;n,\theta')$ exceeds the corresponding expectation with h^{res} alone by at most $C\beta\gamma_1(1/n+\theta'/s)^{1/2}$. Dropping all constraints and using Jensen's inequality, we find that the expectation with h^{res} alone is at most $N^{-1}\log \mathbb{E}\int e^{\beta h^{\mathrm{res}}}\,\mathrm{d}\nu_u^\perp = \beta^2\psi_Q(1)/2$.

Inserting these bounds into (12.25), with $\theta' = \theta + (8\delta)^{1/2}$ and $\delta = \alpha/\sqrt{Q}$, yields (12.24), where we absorbed $C\alpha$ into $C\sqrt{\alpha}$. □

The next lemma is the step in which the diagonal of the overlap matrix matters. For $\sigma^1,\dots,\sigma^n \in S_N$, we write

$$\bar\sigma = \frac{1}{n}\sum_{i=1}^n \sigma^i, \qquad \bar q = \frac{|\bar\sigma|^2}{N}, \qquad \widehat{m} = \Big(\frac{Q}{\bar q}\Big)^{1/2}\bar\sigma \quad (\bar q > 0), \qquad q_n = Q + \frac{s}{n}. \tag{12.26}$$

If $|R(\sigma^i,\sigma^j) - Q| < \theta \le Q/2$ for all $i \neq j$, then, since $R(\sigma^i,\sigma^i) = 1$,

$$|\bar q - q_n| \le \theta, \qquad |R(\sigma^i,\bar\sigma) - q_n| \le \theta, \qquad |R(\sigma^i,\widehat{m}) - Q| \le C_Q\Big(\frac{1}{n} + \theta\Big). \tag{12.27}$$

Indeed, $\bar q = n^{-2}\sum_{i,j} R(\sigma^i,\sigma^j)$ and $R(\sigma^i,\bar\sigma) = n^{-1}\sum_j R(\sigma^i,\sigma^j)$ consist of the diagonal term $1/n$ and the average of $n-1$ off-diagonal terms. For the last bound, $R(\sigma^i,\widehat{m}) = (Q/\bar q)^{1/2}R(\sigma^i,\bar\sigma)$ is a Lipschitz function of $(\bar q, R(\sigma^i,\bar\sigma))$ on $[Q/2,2]\times[-1,1]$, and its value at (q_n,q_n) is $(Qq_n)^{1/2}$, which exceeds Q by at most $s/(2n)$. The squared norm of the empirical mean is close to $q_n = Q + (1-Q)/n$, and we never replace it by Q.

Lemma 12.7 (Near-ground replica means)**.** *Let $\varepsilon > 0$. There exists an integer n_1 such that, for every integer $n \ge n_1$, there exist $\theta_n > 0$ and $c_n > 0$ such that, for every $\theta \in (0,\theta_n]$ and all sufficiently large N,*

$$\mathbb{E}\,\pi^{\otimes n}\Big(|R(\sigma^i,\sigma^j) - Q| < \theta \text{ for all } i \neq j,\ \ H_N(\widehat{m}) < N(E_0 - \varepsilon)\Big) \le e^{-c_n N}. \tag{12.28}$$

Proof. The parameters are chosen in the following order: first the tolerances α,θ_1 and a lower bound n_2 for the replica number, for which the published estimates hold; then the mesh h of a net; then n_1; and finally θ_n.

By Theorem 12.5 with $t = \beta\varepsilon/32$ and $c = 1$, and by Lemma 12.6, there exist $\alpha,\theta_1 > 0$ and n_2 such that, for every fixed $n \ge n_2$, the following holds with probability at least $1 - e^{-N}$ for all large N:

$$\sup_{|m|^2 = NQ} \widetilde{F}_N(m;n,\alpha,\theta_1) \le F_{\mathrm{band}} + \frac{\beta\varepsilon}{8}. \tag{12.29}$$

Indeed, let α_0, θ_0, n_0 be given by Theorem 12.5. We choose $\alpha \le \min\{\alpha_0, \alpha_1\}$, $\theta_1 \le \theta_0$, and $n_2 \ge n_0$ so that the error terms on the right side of (12.24) sum to at most $\beta\varepsilon/32$ for all $n \ge n_2$; this is possible, because they decrease in n and increase in θ. Then $\mathbb{E}\widetilde{F}_N(m; n, \alpha, \theta_1) \le F_{\mathrm{band}} + \beta\varepsilon/16$ for large N, and (12.23) gives (12.29). By Lemma 3.3 with $B = S_N$ and $h = 0$, we have $\mathbb{P}(|\log Z_{N,\beta} - \mathbb{E}\log Z_{N,\beta}| \ge tN) \le 2e^{-t^2N/(2\beta^2\xi(1))}$. Since $N^{-1}\mathbb{E}\log Z_{N,\beta} \to F(\beta)$, we obtain

$$\mathbb{P}\Big(\frac{1}{N}\log Z_{N,\beta} < F(\beta) - \frac{\beta\varepsilon}{8}\Big) \le e^{-cN} \tag{12.30}$$

for all large N. By Lemma 12.3, there exists L with $\mathbb{P}(\mathrm{Lip}_N > L\sqrt{N}) \le Ce^{-cN}$.

Let $h = \min\{\varepsilon/(4L), \alpha/4\}$, and let $\mathcal{N}$ be an $h\sqrt{N}$-net of the sphere $\{|m|^2 = NQ\}$ with $|\mathcal{N}| \le e^{Na_h}$, where $a_h = \log(1 + 2\sqrt{Q}/h)$. We fix $n_1 \ge n_2$ with $C_Q/n_1 < \alpha/4$ and $n_1\beta\varepsilon > 4a_h$, and for $n \ge n_1$ we set $\theta_n = \min\{\theta_1, \alpha/(4C_Q), Q/2\}$.

Fix $n \ge n_1$ and $\theta \le \theta_n$, and suppose that (12.29), the bound $N^{-1}\log Z_{N,\beta} \ge F(\beta) - \beta\varepsilon/8$, and $\mathrm{Lip}_N \le L\sqrt{N}$ hold. Let $(\sigma^1, \dots, \sigma^n)$ belong to the event in (12.28), and let $m' \in \mathcal{N}$ satisfy $|\widehat{m} - m'| \le h\sqrt{N}$. Then

$$H_N(m') \le H_N(\widehat{m}) + LhN < N\Big(E_0 - \frac{3\varepsilon}{4}\Big), \qquad |R(\sigma^i, m') - Q| \le |R(\sigma^i, \widehat{m}) - Q| + h < \alpha,$$

where in the second bound we used (12.27), $C_Q(1/n + \theta) < \alpha/2$, and $h \le \alpha/4$. We conclude that the tuple belongs to $\mathcal{A}(m'; n, \alpha, \theta_1)$ for a net point m' with $H_N(m') < N(E_0 - 3\varepsilon/4)$. For such m', using (12.29), the bound on $Z_{N,\beta}$, and the TAP identity (12.9), we have

$$\begin{aligned}\pi^{\otimes n}\big(\mathcal{A}(m'; n, \alpha, \theta_1)\big) &= Z_{N,\beta}^{-n}\exp\big\{n\beta H_N(m') + Nn\widetilde{F}_N(m'; n, \alpha, \theta_1)\big\} \\ &\le \exp\Big\{nN\Big[\beta\Big(E_0 - \frac{3\varepsilon}{4}\Big) + F_{\mathrm{band}} + \frac{\beta\varepsilon}{4} - F(\beta)\Big]\Big\} = e^{-nN\beta\varepsilon/2}.\end{aligned}$$

Summing over the net, we find that the Gibbs probability of the event in (12.28) is at most $e^{Na_h - nN\beta\varepsilon/2} \le e^{-Na_h}$. The three events on the field that we assumed fail with probability at most Ce^{-cN}, and taking expectations proves (12.28). □

The next lemma is deterministic. It moves a point of high energy to a nearby local maximum. Localization of deep connected components around critical points also appears in [30, Proposition 39] and [138, Corollary 1.5 and its proof]; we give the deterministic estimate used here.

Lemma 12.8 (Proximity of high-energy points to local maxima)**.** *Let f be a C^3 function on S_N, and let $E_0 \in \mathbb{R}$ and $\varepsilon_{\mathrm{loc}}, \kappa > 0$, $C_3 \ge 0$. Suppose that every critical point c of f with $f(c) \ge N(E_0 - \varepsilon_{\mathrm{loc}})$ satisfies $\nabla^2_{S_N} f(c) \preceq -\kappa\,\mathrm{Id}$, and that $\sup \|\nabla^3_{S_N} f\|_{\mathrm{op}} \le C_3 N^{-1/2}$. Let $r_0 = \min\{1, \kappa/(2C_3)\}$ (with $r_0 = 1$ if $C_3 = 0$), and suppose that $\max f \le N(E_0 + \varepsilon)$, where $3\varepsilon < \min\{\varepsilon_{\mathrm{loc}}, \kappa r_0^2/4\}$. Then every $v \in S_N$ with $f(v) \ge N(E_0 - \varepsilon)$ lies within geodesic distance $(8\varepsilon/\kappa)^{1/2}\sqrt{N}$ of a critical point c of f with $f(c) \ge f(v)$ and $\nabla^2_{S_N} f(c) \preceq -\kappa\,\mathrm{Id}$.*

Proof. Let K be the connected component of the closed set $\{f \geq N(E_0 - 2\varepsilon)\}$ that contains v. Since K is compact, f attains its maximum on K at some point c, and $f(c) \geq f(v) > N(E_0 - 2\varepsilon)$. A point at which f exceeds $N(E_0 - 2\varepsilon)$ has a connected open neighborhood in the superlevel set, which is contained in K. Then c is a local maximum of f on S_N, and c is a critical point with $f(c) \geq N(E_0 - \varepsilon) > N(E_0 - \varepsilon_{\mathrm{loc}})$. By assumption, $\nabla^2_{S_N} f(c) \preceq -\kappa \,\mathrm{Id}$.

Along a unit-speed geodesic γ from c, the Hessian $\nabla^2_{S_N} f(\gamma(t))[\gamma'(t), \gamma'(t)]$ changes by at most $C_3 N^{-1/2} t$, which is at most $\kappa/2$ for $t \leq r_0\sqrt{N}$. As $r_0 \leq 1$, the geodesic ball of radius $r_0\sqrt{N}$ about c lies within the injectivity radius $\pi\sqrt{N}$. Integrating twice along radial geodesics, we obtain

$$f(c) - f(w) \geq \frac{\kappa}{4} \operatorname{dist}(c, w)^2 \qquad \text{for } \operatorname{dist}(c, w) \leq r_0\sqrt{N}.$$

On the sphere $\operatorname{dist}(c, w) = r_0\sqrt{N}$, this gives $f(w) \leq N(E_0 + \varepsilon) - \kappa r_0^2 N/4 < N(E_0 - 2\varepsilon)$. The connected set K contains c and does not meet this sphere, and therefore K lies in the open ball of radius $r_0\sqrt{N}$ about c. In particular v lies in this ball, and $\kappa \operatorname{dist}(c, v)^2/4 \leq f(c) - f(v) \leq 2\varepsilon N$, as desired. □

We also record the measurability of the unions of bands and of the weights used below, which is needed to apply (12.13). We identify the field H with its finite vector of tensor entries, as in Section 7.

Lemma 12.9 (Measurability)**.** *Let $U' \subset \mathbb{R}$ and $I \subset \mathbb{R}$ be closed and $b > 0$. The set of pairs (H, σ) such that $\sigma \in B_u$ for some $u \in \mathcal{C}_{U'}(I)$ is closed in the product of coefficient space and S_N. The function $(H, u) \mapsto Z_H(B_u)$ is continuous. In particular, $\pi(S_N \setminus \bigcup_{u \in \mathcal{C}_{U'}(I)} B_u)$ is a Borel function of H, and $(H, u) \mapsto \mathbf{1}\{Z_H(B_u) < t D_N e^{\beta(H_Q(u) - m_N)}\}$ is Borel for every $t > 0$.*

Proof. The value $H_Q(u)$, the radial derivative $K_Q(u)$, and the gradient $\nabla_{S_N} H_Q(u)$ are continuous in (H, u). The set of triples (H, σ, u) with $\nabla_{S_N} H_Q(u) = 0$, $K_Q(u)/N \in U'$, $H_Q(u) \in I$, and $|R(\sigma, u) - \sqrt{Q}| \leq b$ is closed. Its projection onto (H, σ) is closed, because S_N is compact: if $(H_i, \sigma_i) \to (H, \sigma)$ with witnessing centers u_i, a subsequence of (u_i) converges, and its limit is a witnessing center for (H, σ). This is the argument of Lemma 4.3. The boundary $\{|R(\sigma, u) - \sqrt{Q}| = b\}$ has ν_N-measure zero, and dominated convergence implies that $Z_H(B_u)$ is continuous in (H, u). The remaining assertions follow, since $Z_{N,\beta} > 0$ is continuous in H. □

For $\tau > 0$, the *coarse family* is $\mathcal{C}(I_\tau)$ with $I_\tau = [N(E_0 - \tau), \infty)$, the set of critical points of H_Q with radial derivative in U and energy at least $N(E_0 - \tau)$, and the coarse union is

$$B_N^{\mathrm{coarse}}(\tau) = \bigcup_{u \in \mathcal{C}(I_\tau)} B_u.$$

Proposition 12.10 (Coarse cover). *For every fixed $b > 0$ and $\tau > 0$,*

$$\lim_{N\to\infty} \mathbb{E}\,\pi\big(S_N \setminus B_N^{\mathrm{coarse}}(\tau)\big) = 0. \tag{12.31}$$

Proof. The proof has four steps. We first fix a localization tolerance ε; then we choose a replica number n and an overlap tolerance θ, so that a sample whose overlaps with the members of a group are close to Q, where all mutual overlaps in the group are close to Q, lies in the band of the local maximum near the empirical mean of the group; then, along a subsequence, we use the directing measure to produce such a group among J additional independent replicas; and finally we average over the samples.

Localization. Let δ_U, κ, C_3 be as in (12.11), and let $r_0 = \min\{1, \kappa/(2C_3)\}$. Fix $\varepsilon \in (0, \tau]$ with $3\varepsilon < \min\{\delta_U, \kappa r_0^2/4\}$ and $(8\varepsilon/\kappa)^{1/2} < b/4$. Let $\mathcal{G}_N$ be the event that the conclusions of (12.11) hold and $\max H_Q \le N(E_0 + \varepsilon)$. Note that $m_N + M \le N(E_0 + \varepsilon)$ for every fixed M and all large N; together with (12.11) and Lemma 12.2, this implies that $\mathbb{P}(\mathcal{G}_N) \to 1$. On $\mathcal{G}_N$, Lemma 12.8 applies to $f = H_Q$ with $\varepsilon_{\mathrm{loc}} = \delta_U$, and it shows that every $v \in S_N$ with $H_Q(v) \ge N(E_0 - \varepsilon)$ lies within geodesic distance $(b/4)\sqrt{N}$ of a critical point c with $H_Q(c) \ge N(E_0 - \varepsilon)$. By (12.11), we have $K_Q(c)/N \in U$, and $c \in \mathcal{C}(I_\tau)$ because $\varepsilon \le \tau$.

The replica group. Let n_1 be given by Lemma 12.7 for this ε, and fix $n \ge n_1$ with $|Q/q_n^{1/2} - \sqrt{Q}| < b/4$, which is possible since $q_n \to Q$. Then fix $\theta \in (0, \theta_n]$ so small that the following holds: if $\sigma^1, \dots, \sigma^n$ satisfy $|R(\sigma^i, \sigma^j) - Q| < \theta$ for $i \ne j$ and σ^0 satisfies $|R(\sigma^0, \sigma^i) - Q| < \theta$ for all i, then $u = \bar\sigma/\bar q^{1/2} \in S_N$ satisfies $|R(\sigma^0, u) - Q/q_n^{1/2}| < b/4$. This is possible because $R(\sigma^0, \bar\sigma)$ is the average of the overlaps $R(\sigma^0, \sigma^i)$ and $|\bar q - q_n| \le \theta$ by (12.27). Note that $H_Q(u) = H_N(\widehat m)$. Suppose now that, in addition, $\mathcal{G}_N$ holds and $H_N(\widehat m) \ge N(E_0 - \varepsilon)$. Let c be the critical point given by the localization step for $v = u$. Since $|R(\sigma^0, c) - R(\sigma^0, u)| \le |c - u|/\sqrt{N}$ and the Euclidean distance is at most the geodesic distance, we obtain

$$|R(\sigma^0, c) - \sqrt{Q}| \le |R(\sigma^0, c) - R(\sigma^0, u)| + |R(\sigma^0, u) - Q/q_n^{1/2}| + |Q/q_n^{1/2} - \sqrt{Q}| < \frac{3b}{4}.$$

We conclude that $\sigma^0 \in B_c \subset B_N^{\mathrm{coarse}}(\tau)$.

Additional replicas. Let $(\sigma^l)_{l\ge 0}$ be independent samples from π, and consider their overlap array under the annealed law $\mathbb{E}\,\pi^{\otimes\infty}$, as in Section 6. Every subsequence of N has a further subsequence along which this law converges to a limiting overlap law $\mathcal{L}$. We fix such a subsequence, and a directing measure $(G_\omega)_{\omega\in\Omega}$ of $\mathcal{L}$ (Definition 6.1 and Theorem 6.2). Under $\mathcal{L}$, we draw ω and then independent vectors $v^0, v^1, \dots$ with law G_ω, and $R_{ll'} = v^l \cdot v^{l'}$ for $l \ne l'$. By Theorem 6.3, G_ω is carried by $\{|v|^2 = Q\}$ for almost every ω. The vector v^0 lies almost surely in the support of G_ω, and $G_\omega(\{v : |v - v^0| < \theta/4\}) > 0$ almost surely. If $|v|^2 = |v'|^2 = Q$ and $|v - v^0|, |v' - v^0| < \theta/4$,

then $|v \cdot v' - Q| < \theta/2$ and $|v \cdot v^0 - Q| < \theta/2$, since all vectors have norm at most one. Conditionally on (ω, v^0), the number of $l \in \{1, \dots, J\}$ with $|v^l - v^0| < \theta/4$ is binomial with J trials and positive success probability. Given $\varepsilon' > 0$, it follows that there exists J such that, with limiting probability at least $1 - \varepsilon'$, at least n of the vectors $v^1, \dots, v^J$ lie in this ball. Let $\mathcal{O}_J$ be the event that there is an n-element set $S \subset \{1, \dots, J\}$ with $|R_{ll'} - Q| < \theta$ for all distinct $l, l' \in S \cup \{0\}$. It is an open subset of the space of finite arrays, and the limiting array belongs to it with probability at least $1 - \varepsilon'$. By the portmanteau theorem, $\liminf \mathbb{E}\, \pi^{\otimes(J+1)}(\mathcal{O}_J) \geq 1 - \varepsilon'$ along the subsequence.

Averaging. By Lemma 12.9, $B_N^{\mathrm{coarse}}(\tau)$ is a Borel function of the field, and $\mathbb{E}\, \pi(S_N \setminus B_N^{\mathrm{coarse}}(\tau))$ is the annealed probability that $\sigma^0 \notin B_N^{\mathrm{coarse}}(\tau)$. By the second step, this event is contained in the union of $\mathcal{O}_J^c$, of $\mathcal{G}_N^c$, and of the events that some n-element set $S \subset \{1, \dots, J\}$ has all mutual overlaps within θ of Q and has $H_N(\widehat{m}_S) < N(E_0 - \varepsilon)$, where $\widehat{m}_S$ is the normalized empirical mean of $(\sigma^l)_{l \in S}$. Using Lemma 12.7 and a union bound over the $\binom{J}{n}$ sets S, we find that the last events have total annealed probability at most $\binom{J}{n} e^{-c_n N}$. We conclude that

$$\limsup \mathbb{E}\, \pi\big(S_N \setminus B_N^{\mathrm{coarse}}(\tau)\big) \leq \varepsilon'$$

along the subsequence. Since ε' is arbitrary, the limit along the subsequence is zero. Every subsequence has such a further subsequence, which proves (12.31). □

The number J of additional replicas depends on the subsequence through its limiting directing measure, while the coarse union depends only on the field. The argument uses neither atoms of the directing measure nor a description of its weights.

12.4 Determinant prefactors and the quadratic component

Let $\sigma_{\mathrm{sc}}(\,\mathrm{d}r) = (2\pi)^{-1}(4 - r^2)^{1/2} \mathbf{1}_{[-2,2]}(r)\,\mathrm{d}r$ be the semicircle law. Recall the logarithmic potential Ω of Sections 7 and 10. For $t > 2$, we have

$$\Omega(t) = \int \log(t - r)\, \sigma_{\mathrm{sc}}(\,\mathrm{d}r), \qquad \Omega'(t) = \frac{t - \sqrt{t^2 - 4}}{2}, \qquad \Omega''(t) = \frac{1}{2}\Big(1 - \frac{t}{\sqrt{t^2 - 4}}\Big) < 0. \tag{12.32}$$

The formula for Ω' is the Stieltjes transform of the semicircle law. The next lemma shows that the mean determinant $d_N(k)$ equals its exponential rate up to a bounded factor. The absence of any power of N in this comparison is used in Proposition 12.15.

Lemma 12.11 (Bounded determinant prefactors)**.** *For every compact set $\mathcal{K} \subset (2\sqrt{b_Q}, \infty)$, there exist $0 < c \leq C < \infty$ such that, for all large N and all $k \in \mathcal{K}$,*

$$d_N(k) = \chi_N(k) \exp\Big\{N\Big[\frac{1}{2}\log b_Q + \Omega\big(k/\sqrt{b_Q}\big)\Big]\Big\}, \qquad c \leq \chi_N(k) \leq C. \tag{12.33}$$

Proof. We use Corollary 7.10 with $b = b_Q$. With $m = N-1$, we have $b_m = b_Q(N-1)/N$ and $\sqrt{b_m}\, G_{N-1} = W_{N-1,N}$, and the corollary implies that $e^{\Lambda_{N-1,N}(k)} \le d_N(k) \le Ce^{\Lambda_{N-1,N}(k)}$ uniformly for $k \in \mathcal{K}$. With $m = N$, we have $b_m = b_Q$ and $\Lambda_{N,N}(k) = N[\frac{1}{2}\log b_Q + \Omega(k/\sqrt{b_Q})]$, and the comparison of the two dimensions in the corollary yields $|\Lambda_{N-1,N}(k) - \Lambda_{N,N}(k)| \le C$. Combining these bounds, we obtain (12.33). □

The band around a center is correlated with the Hessian at the center through the quadratic component Y_2 of Lemma 12.4. The next lemma shows that this correlation costs only a bounded factor in the Palm expectation of the partition function $Z_u(a)$ of the latitude sphere at a, defined in (12.4). For marks (e,k) and $a \in (-1,1)$, let

$$\Xi(e,k,a) = \beta m_{e,k}(a) + \frac{\beta^2}{2} v(a), \tag{12.34}$$

so that $\mathbb{E}_{e,k,u} e^{\beta H_N(\sigma(a,\omega))} = e^{N\Xi(e,k,a)}$ for every ω, by Lemma 12.4.

Lemma 12.12 (Correlation with the quadratic component)**.** *Let $b_0 > 0$ and $\varepsilon_2 \in (0,1)$ satisfy $I_{b_0} \subset (0,1)$ and $2\beta^2 v_2(a) \le 1-\varepsilon_2$ for $a \in I_{b_0}$, and let $\mathcal{K}_0$ be a compact set of marks with k in a compact subset of $(2\sqrt{b_Q}, \infty)$. There exists C such that, for all $N \ge 8/\varepsilon_2$, $a \in I_{b_0}$, and $(e,k) \in \mathcal{K}_0$,*

$$\mathbb{E}_{e,k,u}\big[D_u Z_u(a)\big] \le C d_N(k) e^{N\Xi(e,k,a)}. \tag{12.35}$$

Proof. Fix a and (e,k). Lemma 12.4 implies that D_u is a function of W, and that (W, Y_2) is independent of $Y_{\le 1}$ and $Y_{\ge 3}$. Using Fubini's theorem and this independence, we find

$$\begin{aligned}\mathbb{E}_{e,k,u}\big[D_u Z_u(a)\big] &= e^{\beta N m_{e,k}(a)} \int \mathbb{E}\big[e^{\beta(Y_{\le 1}+Y_{\ge 3})(a,\omega)}\big]\, \mathbb{E}\big[D_u e^{\beta Y_2(a,\omega)}\big]\, \nu_u^{\perp}(\,\mathrm{d}\omega) \\ &= e^{\beta N m_{e,k}(a) + \beta^2 N(v(a)-v_2(a))/2}\, \mathbb{E}\big[D_u Z_2\big],\end{aligned}$$

where $Z_2 = \int e^{\beta Y_2(a,\omega)} \nu_u^{\perp}(\,\mathrm{d}\omega)$, and we used that the variance of $(Y_{\le 1} + Y_{\ge 3})(a,\omega)$ does not depend on ω. By the Cauchy–Schwarz inequality and Proposition 7.7,

$$\mathbb{E}[D_u Z_2] \le \big(\mathbb{E} D_u^2\big)^{1/2} \big(\mathbb{E} Z_2^2\big)^{1/2} \le \mathsf{K}_2^{1/2} d_N(k) \big(\mathbb{E} Z_2^2\big)^{1/2}.$$

It remains to show that $\mathbb{E} Z_2^2 \le C(\mathbb{E} Z_2)^2$, where $\mathbb{E} Z_2 = e^{\beta^2 N v_2(a)/2}$. Using rotation invariance and the covariance of Y_2, we have

$$\frac{\mathbb{E} Z_2^2}{(\mathbb{E} Z_2)^2} = \int e^{\beta^2 N v_2(a) R(\omega,\omega')^2}\, \nu_u^{\perp}(\,\mathrm{d}\omega') = c_{N-1}^{\mathrm{lat}} \int_{-1}^{1} e^{\beta^2 N v_2(a) r^2} (1-r^2)^{(N-4)/2}\,\mathrm{d}r,$$

since the overlap $r = R(\omega,\omega')$ of a fixed point with a uniform point of the $(N-2)$-dimensional sphere $S_u^{\perp}$ has density $c_{N-1}^{\mathrm{lat}}(1-r^2)^{(N-4)/2}$. Since $\log(1-r^2) \le -r^2$ and $2\beta^2 v_2(a) \le 1-\varepsilon_2$, the exponent in the last integral is at most $Nr^2(1-\varepsilon_2)/2 - (N-4)r^2/2 \le -N\varepsilon_2 r^2/4$ for $N \ge 8/\varepsilon_2$. Then $c_{N-1}^{\mathrm{lat}} \le C\sqrt{N}$ and $\int e^{-N\varepsilon_2 r^2/4}\,\mathrm{d}r \le C/\sqrt{N}$ imply that the ratio is bounded. This proves (12.35). □

The proof uses a second moment of the quadratic part of the field on the latitude sphere only, not of the full residual partition function.

12.5 The Palm expectation of the band partition function

In this section, we prove Proposition 12.15. For a non-pure mixture, define

$$\begin{aligned} \Theta(e,k) &= \frac{1}{2} + \frac{1}{2}\log\frac{b_Q}{a_Q} + \Omega\big(k/\sqrt{b_Q}\big) - \frac{1}{2}q(e,k), \\ g(e,k,a) &= \Theta(e,k) + \Xi(e,k,a) + \frac{1}{2}\log(1-a^2). \end{aligned} \tag{12.36}$$

Here Θ is the marked complexity of Section 10, and g adds to it the exponent of the annealed contribution of the latitude a to the band partition function. By (12.33), we have

$$w_N(Ne,k) = \chi_N(k)\, e^{N(\frac{1}{2}\log a_Q - \frac{1}{2})}\, e^{N\Theta(e,k)}. \tag{12.37}$$

Lemma 12.13 (The Laplace exponent at the ground-state parameters)**.** *Let the mixture be non-pure. At $(e,k,a) = (E_0, K_0, \sqrt{Q})$,*

$$\partial_e\Theta = -c_0, \qquad \partial_k\Theta = 0, \qquad \partial_k^2\Theta = -\frac{1}{y^2(1-\rho)} - (\Sigma_Q^{-1})_{22}, \tag{12.38}$$

$$\partial_e g = \beta - c_0, \quad \partial_k g = \partial_a g = \partial_e\partial_a g = 0, \quad \partial_k\partial_a g = \frac{\beta}{\sqrt{Q}}, \quad \partial_a^2 g = -\frac{(1-\rho)(1+Q-C\rho)}{s^2}, \tag{12.39}$$

and the Hessian of g in the variables (k,a) is negative definite. Moreover, for all (e,k),

$$g(e,k,\sqrt{Q}) = \Theta(e,k) + \beta e + F_{\mathrm{band}}. \tag{12.40}$$

Proof. Let $t_0 = K_0/\sqrt{b_Q}$. Since $K_0 = y + b_Q/y$ and $y > \sqrt{b_Q}$, we have $(t_0^2-4)^{1/2} = (y^2-b_Q)/(y\sqrt{b_Q})$. Using (12.32), we find $\Omega'(t_0) = \sqrt{b_Q}/y$ and $\Omega''(t_0) = -b_Q/(y^2-b_Q)$. Together with (12.7), this implies (12.38), since $\partial_k\Theta = \Omega'(k/\sqrt{b_Q})/\sqrt{b_Q} - (\Sigma_Q^{-1}(e,k)^{\mathsf{T}})_2$ and $b_Q/y^2 = \rho$.

For the derivatives of Ξ, note that $b(\sqrt{Q}) = \Sigma_Q e_1$ and $b'(\sqrt{Q}) = Q^{-1/2}\Sigma_Q e_2$, where e_1, e_2 are the standard basis vectors of $\mathbb{R}^2$. Also, $d(\sqrt{Q}) = s\xi'(Q)$ and $d'(\sqrt{Q}) = 2\sqrt{Q}(s\xi''(Q) - \xi'(Q))$. Using (12.7), we find that, at (E_0, K_0),

$$b'(\sqrt{Q})^{\mathsf{T}}\Sigma_Q^{-1}\begin{pmatrix} E_0 \\ K_0 \end{pmatrix} = \frac{K_0}{\sqrt{Q}}, \quad b'(\sqrt{Q})^{\mathsf{T}}\Sigma_Q^{-1}b(\sqrt{Q}) = \sqrt{Q}\,\xi'(Q), \quad b(\sqrt{Q})^{\mathsf{T}}\Sigma_Q^{-1} = e_1^{\mathsf{T}}.$$

The last identity implies that $\partial_e g = -c_0 + \beta$, and it also gives (12.40), since at $a = \sqrt{Q}$ the mean is $m_{e,k}(\sqrt{Q}) = e$ and $v(\sqrt{Q}) = \xi(1) - \xi(Q) - s\xi'(Q) = \psi_Q(1)$. Since Ξ is linear in

(e,k) with coefficients $\beta b(a)^{\mathsf T}\Sigma_Q^{-1}$, we have $\partial_e\partial_a g = \beta(e_2^{\mathsf T}e_1)/\sqrt{Q} = 0$ and $\partial_k\partial_a g = \beta/\sqrt{Q}$, and $\partial_k g = \partial_k\Theta = 0$. For the latitude derivative,

$$\partial_a g = \frac{\beta K_0}{\sqrt{Q}} - \beta^2\sqrt{Q}\,\xi'(Q) - \frac{\beta^2}{2}d'(\sqrt{Q}) - \frac{\sqrt{Q}}{s} = \frac{\beta K_0}{\sqrt{Q}} - \beta^2 s\sqrt{Q}\,\xi''(Q) - \frac{\sqrt{Q}}{s} = 0,$$

where in the last step we used $\beta y = Q/s$ and $\beta b_Q/y = \beta^2 s\,Q\,\xi''(Q)$ in $\beta K_0 = \beta y + \beta b_Q/y$.

For the second latitude derivative, a direct computation gives

$$b''(\sqrt{Q}) = \frac{1}{Q}\begin{pmatrix} b_Q \\ 2b_Q + Q^3\xi'''(Q)\end{pmatrix}, \qquad d''(\sqrt{Q}) = -\frac{2a_Q + 8b_Q}{Q} + \frac{2s}{Q^2}\Big(\frac{b_Q^2}{a_Q} + Q^3\xi'''(Q)\Big),$$

and, since $\Sigma_Q^{-1}(E_0,K_0)^{\mathsf T} = (c_0, 1/y)^{\mathsf T}$ and $\Sigma_Q^{-1}b(\sqrt{Q}) = e_1$,

$$\partial_a^2 g = -\frac{1+Q}{s^2} + \beta\, b''(\sqrt{Q})^{\mathsf T}\begin{pmatrix} c_0-\beta \\ 1/y\end{pmatrix} - \frac{\beta^2(a_Q+b_Q)}{Q} - \frac{\beta^2}{2}d''(\sqrt{Q}).$$

The terms containing $\xi'''(Q)$ are $\beta Q^2\xi'''(Q)/y = \beta^2 sQ\,\xi'''(Q)$ from the second term, where we used $1/y = \beta s/Q$, and $-\beta^2 sQ\,\xi'''(Q)$ from the last term; they cancel. Collecting the remaining terms, we obtain

$$\partial_a^2 g = -\frac{1+Q}{s^2} + \frac{\beta b_Q}{Q}\Big(c_0 + \frac{2}{y}\Big) + \frac{2\beta^2 b_Q}{Q} - \frac{\beta^2 s\, b_Q^2}{Q^2 a_Q}.$$

Substituting $b_Q = \rho y^2$, $y = Q/(\beta s)$, $c_0 = \beta x$, and $y^2/a_Q = 1 + z = C/s$, we find

$$\begin{aligned}\partial_a^2 g &= \frac{-(1+Q) + xQ\rho + 2s\rho + 2Q\rho - C\rho^2}{s^2} \\ &= \frac{-(1+Q) + (2+xQ)\rho - C\rho^2}{s^2} = -\frac{(1-\rho)(1+Q-C\rho)}{s^2}.\end{aligned}$$

It remains to show that the (k,a)-Hessian is negative definite. Its diagonal entry $\partial_k^2 g = \partial_k^2\Theta$ is negative. For the inverse of a positive definite 2×2 matrix, we have $(\Sigma_Q^{-1})_{22} - (\Sigma_Q^{-1})_{12}^2/(\Sigma_Q^{-1})_{11} = 1/(\Sigma_Q)_{22} = 1/(a_Q+b_Q)$. It follows that $(\Sigma_Q^{-1})_{22} \ge 1/(a_Q+b_Q)$ and $|\partial_k^2 g| \ge h$, where

$$h = \frac{1}{y^2(1-\rho)} + \frac{1}{a_Q+b_Q} = \frac{1}{y^2}\Big(\frac{1}{1-\rho} + \frac{C}{s+C\rho}\Big) = \frac{1+s/C}{y^2(1-\rho)(s/C+\rho)},$$

where we used $a_Q + b_Q = y^2(s/C+\rho)$. Then the Schur complement of $\partial_k^2 g$ in the (k,a)-Hessian satisfies

$$\begin{aligned}\partial_a^2 g - \frac{(\partial_k\partial_a g)^2}{\partial_k^2 g} \le \partial_a^2 g + \frac{\beta^2}{Qh} &= -\frac{1-\rho}{s^2}\Big[1+Q-C\rho - \frac{Q(s+C\rho)}{C+s}\Big] \\ &= -\frac{1-\rho}{s^2}\Big[(1-C) + \frac{(1-\rho)C(C+1)}{C+s}\Big],\end{aligned}$$

where we used $\beta^2 y^2/Q = Q/s^2$. The last equality is checked by multiplying by $C+s$ and using $s+Q=1$. The right side is negative, because $\rho<1$ and $C = 1-Q(1-x)<1$. This proves the claim. □

In a pure model, the radial derivative is determined by the energy, and only the latitude is integrated. Let $\xi(r) = c_p r^p$ and define

$$L_p(e,a) = \beta e\Big(\frac{a}{\sqrt{Q}}\Big)^p + \frac{1}{2}\log(1-a^2) + \frac{\beta^2}{2}v(a), \qquad \theta_p(e) = -\frac{e^2}{2\xi(Q)} + \Omega\big(pe/\sqrt{b_Q}\big). \tag{12.41}$$

We have $a_Q = p\xi(Q)$, $b_Q = p(p-1)\xi(Q)$, $K_0 = pE_0$, and $E_0 = \xi(Q)(p+z)/y$. Differentiating, we find

$$\begin{gathered}\partial_a L_p(E_0,\sqrt{Q}) = 0, \qquad \partial_a^2 L_p(E_0,\sqrt{Q}) = -\frac{(1-\rho)(pQ-(p-2))}{s^2},\\ L_p(e,\sqrt{Q}) = \beta e + F_{\text{band}}, \qquad \theta_p'(E_0) = -c_0.\end{gathered} \tag{12.42}$$

Indeed, $\partial_a L_p(E_0,\sqrt{Q}) = Q^{-1/2}(\beta K_0 - Q/s - \beta^2 s b_Q/Q) = 0$ by $\beta K_0 = Q/s + \beta^2 s b_Q/Q$. For the second derivative, we have $\beta K_0(p-1)/Q = (p-1)(1+\rho)/s$, $v''(\sqrt{Q}) = 2b_Q((2p-1)Q-(2p-3))/Q^2$, and $\beta^2 b_Q/Q^2 = \rho/s^2$, and collecting terms gives the formula. The value at $\sqrt{Q}$ follows from $v(\sqrt{Q}) = \psi_Q(1)$, and $\theta_p'(E_0) = -E_0/\xi(Q) + p/y = -z/y = -c_0$, by $\Omega'(t_0) = \sqrt{b_Q}/y$. Since $\rho = b_Q/y^2 = (p-1)a_Q/y^2 = (p-1)s/C < 1$ and $C<1$, we have $s < 1/(p-1)$. This implies that $pQ-(p-2) = 2-ps > 0$ and $\partial_a^2 L_p(E_0,\sqrt{Q}) < 0$.

Choice of the neighborhoods

We now choose U, b_0, and ε_0. Let $c_- = c_0/2$ and $c_+ = (c_0+\beta)/2$, so that $0<c_-<c_0<c_+<\beta$. The first requirement below holds for every sufficiently small closed interval U about K_0. The other three are strict inequalities that hold at $(E_0,K_0,\sqrt{Q})$, by $\rho<1$ and Lemma 12.4, by Lemma 12.13 and (12.42), and by (12.7) and (12.42), respectively. By continuity, there exist a closed interval U with K_0 in its interior and $b_0,\varepsilon_1,\varepsilon_2,c_*>0$ such that:

1. $U \subset \widetilde{U}$, $\min U > 2\sqrt{b_Q}$, and U is small enough for (12.11);
2. $I_{b_0} \subset (0,1)$ and $2\beta^2 v_2(a) \le 1-\varepsilon_2$ for $a \in I_{b_0}$;
3. for a non-pure mixture, $\partial_k^2\Theta \le -c_*$ and the (k,a)-Hessian of g is at most $-c_*\mathrm{Id}$ on $[E_0-\varepsilon_1,E_0+\varepsilon_1]\times U\times I_{b_0}$; for a pure mixture, $\partial_a^2 L_p \le -c_*$ on $[E_0-\varepsilon_1,E_0+\varepsilon_1]\times I_{b_0}$;
4. for a non-pure mixture, $c_- \le \ell(e,k) \le c_+$ on $[E_0-\varepsilon_1,E_0+\varepsilon_1]\times U$; for a pure mixture, $p[E_0-\varepsilon_1,E_0+\varepsilon_1]$ lies in the interior of U and $c_- \le -\theta_p' \le c_+$ on $[E_0-\varepsilon_1,E_0+\varepsilon_1]$.

By the third requirement, for $|e-E_0|\le\varepsilon_1$, the function $\Theta(e,\cdot)$ has a unique maximizer $k_*(e)$ on U, and $g(e,\cdot,\cdot)$ has a unique maximizer $(\widehat{k}(e),\widehat{a}(e))$ on $U\times I_{b_0}$; in a pure model,

$L_p(e,\cdot)$ has a unique maximizer $a_*(e)$ on I_{b_0}. At $e = E_0$ these maximizers are K_0, $(K_0, \sqrt{Q})$, and $\sqrt{Q}$, which are interior critical points. By the implicit function theorem, we can choose $\varepsilon_0 \in (0, \varepsilon_1]$ such that, for $|e - E_0| \le \varepsilon_0$, the maximizers are smooth functions of e and lie at distance at least $r_* > 0$ from the boundaries of U, $U \times I_{b_0}$, and I_{b_0}, respectively. This fixes U, b_0, and ε_0. We set

$$\theta(e) = \Theta(e, k_*(e)), \qquad \bar g(e) = g(e, \widehat k(e), \widehat a(e)), \qquad L_*(e) = L_p(e, a_*(e)).$$

Lemma 12.14 (Laplace bounds without polynomial loss)**.** *Let the mixture be non-pure. There exist $c, C, C_2 > 0$ such that, for $|e - E_0| \le \varepsilon_0$ and all large N,*

$$c\,\frac{e^{N\theta(e)}}{\sqrt{N}} \le \mathcal{I}_N(e) = \int_U \chi_N(k) e^{N\Theta(e,k)}\,\mathrm{d}k \le C\,\frac{e^{N\theta(e)}}{\sqrt{N}}, \tag{12.43}$$

$$\mathcal{J}_N(e) = \int_{I_{b_0}}\int_U \chi_N(k)(1-a^2)^{-3/2} e^{Ng(e,k,a)}\,\mathrm{d}k\,\mathrm{d}a \le C\,\frac{e^{N\bar g(e)}}{N}, \tag{12.44}$$

$$\bar g(e) - \theta(e) \le F(\beta) + \beta(e - E_0) + C_2(e - E_0)^2. \tag{12.45}$$

In a pure model, $\int_{I_{b_0}} (1-a^2)^{-3/2} e^{NL_p(e,a)}\,\mathrm{d}a \le Ce^{NL_*(e)}/\sqrt{N}$ and $L_*(e) \le F(\beta) + \beta(e - E_0) + C_2(e - E_0)^2$.

Proof. By the third requirement of Section 12.5, integrating $\partial_k^2\Theta \le -c_*$ twice from the maximizer, we obtain $\Theta(e,k) \le \theta(e) - c_*(k - k_*(e))^2/2$ on U, and, since $\partial_k^2\Theta$ is bounded on the compact rectangle, $\Theta(e,k) \ge \theta(e) - C(k - k_*(e))^2$. Recall that $c \le \chi_N \le C$ by Lemma 12.11. The upper bound in (12.43) follows from the Gaussian integral, and the lower bound follows by integrating over $|k - k_*(e)| \le N^{-1/2}$, which lies in U for large N because $k_*(e)$ has distance at least r_* from the boundary. Similarly, $g(e,k,a) \le \bar g(e) - c_*\{(k - \widehat k(e))^2 + (a - \widehat a(e))^2\}/2$ on $U \times I_{b_0}$, and $\chi_N(k)(1-a^2)^{-3/2}$ is bounded there; integration over $\mathbb{R}^2$ gives (12.44). These inequalities hold at every N, and no error term of the Laplace method is suppressed. The integral $\int_U w_N(Ne, k)\,\mathrm{d}k$ over the radial derivative equals $e^{N(\frac12 \log a_Q - \frac12)}\mathcal{I}_N(e)$ by (12.37). After removal of this common factor, (12.43) is the bound of Lemma 10.5(c), written with the mean absolute determinant in place of the mean determinant restricted to local maxima.

By (12.40), we have $g(E_0, K_0, \sqrt{Q}) = \Theta(E_0, K_0) + \beta E_0 + F_{\text{band}}$, and (12.9) implies that $\bar g(E_0) - \theta(E_0) = \beta E_0 + F_{\text{band}} = F(\beta)$. The maximizers are interior and smooth in e, and we have $\theta'(e) = \partial_e\Theta(e, k_*(e))$ and $\bar g'(e) = \partial_e g(e, \widehat k(e), \widehat a(e))$. We have $\theta'(E_0) = -c_0$ and $\bar g'(E_0) = \beta - c_0$ by Lemma 12.13, and therefore $(\bar g - \theta)'(E_0) = \beta$. The second derivatives of θ and $\bar g$ are bounded on $[E_0 - \varepsilon_0, E_0 + \varepsilon_0]$, and Taylor's theorem yields (12.45). The pure case is the same computation in the variable a, with (12.42): $L_*(E_0) = F(\beta)$ and $L_*'(E_0) = \partial_e L_p(E_0, \sqrt{Q}) = \beta$. □

In the passage from the exponent $\bar g$ of the joint integral to the Palm expectation, the derivative in the energy changes from $\partial_e\bar g(E_0) = \beta - c_0$ to β, because the exponent θ of

the normalizing integral over the radial derivative has slope $-c_0$. The energy variable is held fixed in both Laplace integrals.

Proposition 12.15 (Palm expectation of the band partition function)**.** *There exist $C_1, C_2 < \infty$ such that, for every $b \in (0, b_0]$, every E with $|E/N - E_0| \le \varepsilon_0$, and all large N,*

$$\mathbb{E}^{\mathrm{Palm}}_{E,u}\big[Z_H(B_u)\big] \le C_1 D_N \exp\Big\{\beta(E - m_N) + C_2 N\Big(\frac{E}{N} - E_0\Big)^2\Big\}. \tag{12.46}$$

Proof. The band partition function $Z_H(B_u)$ increases with b, and it suffices to take $b = b_0$. Let $e = E/N$ and let the mixture be non-pure. Using (12.4), Fubini's theorem, and Lemma 12.12, we find, for $k \in U$,

$$\begin{aligned}\mathbb{E}^{\mathrm{Palm}}_{e,k,u}\big[Z_H(B_u)\big] &= \frac{c_N^{\mathrm{lat}}}{d_N(k)} \int_{I_{b_0}} (1-a^2)^{(N-3)/2}\, \mathbb{E}_{e,k,u}\big[D_u Z_u(a)\big]\, \mathrm{d}a \\ &\le C c_N^{\mathrm{lat}} \int_{I_{b_0}} (1-a^2)^{-3/2} e^{N\Xi(e,k,a) + \frac{N}{2}\log(1-a^2)}\, \mathrm{d}a.\end{aligned}$$

We integrate against the radial density in Definition 12.1. By (12.37), the factor $e^{N(\frac{1}{2}\log a_Q - \frac{1}{2})}$ is common to the numerator and the denominator, and we obtain

$$\mathbb{E}^{\mathrm{Palm}}_{E,u}\big[Z_H(B_u)\big] \le C c_N^{\mathrm{lat}}\, \frac{\mathcal{J}_N(e)}{\mathcal{I}_N(e)} \le C\, \frac{c_N^{\mathrm{lat}}}{\sqrt{N}}\, e^{N(\bar{g}(e) - \theta(e))} \le C\, \frac{c_N^{\mathrm{lat}}}{\sqrt{N}}\, e^{N[F(\beta) + \beta(e - E_0) + C_2(e-E_0)^2]}, \tag{12.47}$$

where in the second and third inequalities we used Lemma 12.14. The joint integral contributes the two widths $N^{-1/2}$ in the radial derivative and in the latitude, and the normalizing integral over the radial derivative contributes one width $N^{-1/2}$. Their ratio leaves the latitude width $N^{-1/2}$, which cancels the growth of c_N^{lat}. On the other hand, by (12.5) and (12.9), we have the exact identity

$$D_N e^{\beta(E - m_N)} = s^{-3/2}\, \frac{c_N^{\mathrm{lat}}}{\sqrt{N}}\, e^{N[F(\beta) + \beta(e - E_0)]}. \tag{12.48}$$

Combining (12.47) and (12.48), we obtain (12.46). The centering error $m_N - NE_0 = O(\log N)$ enters only through the deterministic factor $e^{\beta(E - m_N)}$, and it is not a variable of the Laplace integrals.

In a pure model, the Palm law at energy E is $\mathbb{E}^{\mathrm{Palm}}_{e,pe,u}$, and there is no radial integral. Using Lemma 12.12, with $\Xi(e, pe, a) + \frac{1}{2}\log(1 - a^2) = L_p(e, a)$, and Lemma 12.14, we obtain

$$\mathbb{E}^{\mathrm{Palm}}_{e,pe,u}\big[Z_H(B_u)\big] \le C c_N^{\mathrm{lat}} \int_{I_{b_0}} (1-a^2)^{-3/2} e^{N L_p(e,a)}\, \mathrm{d}a \le C\, \frac{c_N^{\mathrm{lat}}}{\sqrt{N}}\, e^{N[F(\beta) + \beta(e - E_0) + C_2(e-E_0)^2]},$$

and (12.48) again implies (12.46). No inverse of the singular matrix Σ_Q is used. □

12.6 The intensity and the bands below the window

Lemma 12.16 (Exponential bound on the intensity)**.** *There exists C such that, for all large N,*

$$\lambda_N^{\mathrm{all}}(m_N - u) \le Ce^{c_+ u} \qquad \text{for } 0 \le u \le m_N - N(E_0 - \varepsilon_0). \tag{12.49}$$

Proof. Let the mixture be non-pure. For fixed k, the determinant $d_N(k)$ does not depend on E, and

$$\partial_E \log w_N(E,k) = -\ell(E/N,k).$$

Recall that $\ell(E_0,K_0) = c_0$ by (12.7), so that this derivative equals $-c_0$ at the ground-state parameters. By the fourth requirement of Section 12.5, its negative $\ell(E/N,k)$ lies in $[c_-,c_+]$ for $E/N \in [E_0-\varepsilon_0, E_0+\varepsilon_0]$ and $k \in U$. Differentiating under the integral over the compact interval U, which is justified by the Gaussian form of w_N, we find that $\partial_E \log \lambda_N^{\mathrm{all}}(E)$ is a weighted average of the values $-\ell(E/N,k)$, $k \in U$. We conclude that

$$-c_+ \le \partial_E \log \lambda_N^{\mathrm{all}}(E) \le -c_- \qquad \text{for } |E/N - E_0| \le \varepsilon_0. \tag{12.50}$$

Integrating the lower bound from $m_N - u$ to m_N, we obtain $\lambda_N^{\mathrm{all}}(m_N-u) \le \lambda_N^{\mathrm{all}}(m_N)e^{c_+u}$. It remains to bound $\lambda_N^{\mathrm{all}}(m_N)$. Since the interval $[m_N, m_N+1]$ lies in the range of (12.50) for large N, we have

$$\lambda_N^{\mathrm{all}}(m_N) \int_0^1 e^{-c_+ v}\,\mathrm{d}v \le \int_{m_N}^{m_N+1} \lambda_N^{\mathrm{all}}(E)\,\mathrm{d}E \le \mathbb{E}|\mathcal{C}_{N,1}|,$$

and the right side is bounded by Lemma 12.2.

In a pure model, the determinant depends on the energy through its shift pE/N. We do not differentiate it. By Lemma 12.11, for $E/N \in [E_0-\varepsilon_0, E_0+\varepsilon_0]$,

$$\lambda_N^{\mathrm{all}}(E) = |S_N|(2\pi a_Q)^{-(N-1)/2}(2\pi N\xi(Q))^{-1/2}\, b_Q^{N/2}\, \chi_N(pE/N)\, e^{N\theta_p(E/N)},$$

with $c \le \chi_N \le C$. By the fourth requirement of Section 12.5, we have $-\theta_p' \le c_+$ on this range, which implies that $\lambda_N^{\mathrm{all}}(m_N-u) \le (C/c)\lambda_N^{\mathrm{all}}(m_N)e^{c_+u}$. The bound $\lambda_N^{\mathrm{all}}(m_N) \le C$ follows as before, from $\lambda_N^{\mathrm{all}}(m_N+v) \ge (c/C)e^{-c_+v}\lambda_N^{\mathrm{all}}(m_N)$ for $v \in [0,1]$ and Lemma 12.2. □

Corollary 10.17(ii) contains the same exponential bound for the density of nondegenerate local maxima. The lemma above concerns all critical points, which is the count used in (12.13).

Corollary 12.17 (Band partition functions below the window)**.** *Let $\tau_0 = \min\{\varepsilon_0, (\beta - c_+)/(4C_2)\}$ and $a_0 = (\beta - c_+)/2$. There exists C such that, for every $b \in (0, b_0]$, every $\tau \in (0,\tau_0]$, every $M \ge 0$, and all large N,*

$$\mathbb{E} \sum_{u \in \mathcal{C}([N(E_0-\tau), m_N - M))} Z_H(B_u) \le CD_N e^{-a_0 M}. \tag{12.51}$$

The bound also holds when an additional Borel indicator of (H,u) is inserted in the sum.

Proof. By (12.13) with $Y = Z_H(B_u)$, the left side equals $\int \lambda_N^{\mathrm{all}}(E)\mathbb{E}_{E,u}^{\mathrm{Palm}}[Z_H(B_u)]\,\mathrm{d}E$ over $N(E_0-\tau) \le E < m_N - M$; the interval is empty, and the sum is zero, when $m_N - M < N(E_0-\tau)$. We write $E = m_N - u$ and $h_N = m_N - NE_0$, so that $|h_N| \le C\log N$ and u ranges over $(M, \tau N + h_N]$. For large N, we have $3|h_N| \le \tau N$, and then, for $0 \le u \le \tau N + h_N$,

$$N\Big(\frac{E}{N} - E_0\Big)^2 = \frac{(h_N-u)^2}{N} \le \frac{h_N^2 + 2|h_N|u + u(\tau N + h_N)}{N} \le \frac{h_N^2}{N} + 2\tau u.$$

The energies in the integral satisfy $|E/N - E_0| \le \varepsilon_0$ for large N, since $\tau \le \varepsilon_0$ and $m_N/N \to E_0$. Combining Proposition 12.15 and Lemma 12.16, we obtain

$$\frac{1}{D_N}\mathbb{E}\sum_{u\in\mathcal{C}([N(E_0-\tau),m_N-M))} Z_H(B_u) \le Ce^{C_2h_N^2/N}\int_M^\infty e^{-(\beta-c_+-2C_2\tau)u}\,\mathrm{d}u \le \frac{2C}{a_0}e^{-a_0M},$$

where we used $2C_2\tau \le (\beta - c_+)/2$ and $e^{C_2h_N^2/N} \le 2$ for large N. An additional indicator only decreases the integrand. □

The strict inequality $c_0 < \beta$ enters the proof of Theorem 12.22 only here, through the choice $c_+ < \beta$.

12.7 The lower tail of the band partition function

In this section, we prove Proposition 12.20. Let G_u^{res} be the residual field of (12.20), and let

$$L_N^{\mathrm{res}} = e^{-N\beta^2\psi_Q(1)/2}\int e^{\beta G_u^{\mathrm{res}}(\omega)}\,\nu_u^\perp(\,\mathrm{d}\omega) \tag{12.52}$$

be its normalized partition function. Its law does not depend on u or on the marks (e,k), by Lemma 12.4.

Lemma 12.18 (Lower tail of the residual partition function)**.** *We have*

$$\lim_{t\downarrow 0}\limsup_{N\to\infty}\mathbb{P}\big(L_N^{\mathrm{res}} < t\big) = 0. \tag{12.53}$$

Proof. Let $\eta = \beta^2\psi_Q$. Its Taylor coefficients $\beta^2 s^j\xi^{(j)}(Q)/j!$, $j \ge 2$, are nonnegative, because $Q > 0$ and $c_p \ge 0$. This holds also when ξ has odd degrees. It has no constant or linear term, $\eta''(0) = \rho < 1$, and $\eta(t) + t + \log(1-t) < 0$ for $0 < t < 1$ by (12.8). Thus η satisfies (3.22), which is the condition [93, Eq. (2.21)] under which [93, Lemma 3.2] is proved, by [93, Remark 3.3]. A quadratic component of η is allowed. For $n \ge 2$, let h_n be a centered Gaussian field on S_n with covariance $n\psi_Q(R)$, and let $L_n = e^{-n\beta^2\psi_Q(1)/2}\int e^{\beta h_n}\,\mathrm{d}\nu_n$. By Lemma 3.19(1), applied with this η and the field βh_n,

$$\lim_{t\downarrow 0}\limsup_{n\to\infty}\mathbb{P}(L_n < t) = 0. \tag{12.54}$$

The field G_u^{res} lives on a sphere in an ambient space of dimension $n = N-1$, but its covariance carries the factor N. Identify the tangent space at u with $\mathbb{R}^n$ and rescale ω to $\omega' = (n/N)^{1/2}\omega \in S_n$; overlaps are unchanged. Then $G_u^{\mathrm{res}}(\omega)$ has the law of $(N/n)^{1/2}h_n(\omega')$. By Lemma 3.2, applied to the finite mixture ψ_Q, we have $\sup|h_n| \le C_0 n$ except on an event of probability e^{-cn}. On this event,

$$\beta\big((N/n)^{1/2}-1\big)\sup|h_n| \le \frac{\beta C_0}{(1+1/n)^{1/2}+1} \le \frac{\beta C_0}{2},$$

and the normalizations differ by the fixed factor $e^{\beta^2\psi_Q(1)/2}$, since $N-n=1$. It follows that $e^{-K}L_n \le L_N^{\mathrm{res}} \le e^K L_n$ on this event, with $K = \beta C_0/2 + \beta^2\psi_Q(1)/2$. Then $\mathbb{P}(L_N^{\mathrm{res}} < t) \le e^{-cn} + \mathbb{P}(L_n < e^K t)$, and (12.54) implies (12.53). □

The proof uses only the tightness of the limit away from zero. It does not identify the limiting law of L_N^{res}, and it does not require a version of the theorem of Huang, Montanari, and Pham that is uniform in the dimension or in the covariance.

Lemma 12.19 (Symmetric latitude averaging)**.** *Let $b \in (0, b_0]$ and $C_0 > 0$. There exists $K < \infty$ such that the following holds for all N with $[\sqrt{Q}-N^{-1/2}, \sqrt{Q}+N^{-1/2}] \subset I_b$. If $|\partial_a^2 H_N(\sigma(a,\omega))| \le C_0 N$ for all $a \in I_b$ and $\omega \in S_u^{\perp}$, then*

$$Z_H(B_u) \ge \frac{2}{\sqrt{N}}\, c_N^{\mathrm{lat}}\, s^{(N-3)/2} e^{-K}\, Z_u(\sqrt{Q}). \tag{12.55}$$

Proof. For fixed ω, let $J(a) = \beta H_N(\sigma(a,\omega)) + \frac{N-3}{2}\log(1-a^2)$, so that $Z_H(B_u) = c_N^{\mathrm{lat}} \int_{I_b}\int e^{J(a)} v_u^{\perp}(\,\mathrm{d}\omega)\,\mathrm{d}a$. The second derivative of the entropy term is $-(N-3)(1+a^2)/(1-a^2)^2$. Since its absolute value is at most $C_b N$ on I_b, we have $|J''| \le C'N$ on I_b, with $C' = \beta C_0 + C_b$. Let avg denote the uniform average over $a \in [\sqrt{Q}-N^{-1/2}, \sqrt{Q}+N^{-1/2}]$. We expand J around $\sqrt{Q}$ by Taylor's theorem. The linear term has average zero and $\mathrm{avg}(a-\sqrt{Q})^2 = 1/(3N)$, and we find

$$\mathrm{avg}\, J(a) \ge J(\sqrt{Q}) - \frac{C'N}{2}\cdot\frac{1}{3N} = J(\sqrt{Q}) - \frac{C'}{6}.$$

By Jensen's inequality, $\int_{\sqrt{Q}-N^{-1/2}}^{\sqrt{Q}+N^{-1/2}} e^{J(a)}\,\mathrm{d}a \ge 2N^{-1/2} e^{\mathrm{avg}\, J} \ge 2N^{-1/2} e^{J(\sqrt{Q})-C'/6}$. Integrating in ω, multiplying by c_N^{lat}, and using $e^{J(\sqrt{Q})} = s^{(N-3)/2} e^{\beta H_N(\sigma(\sqrt{Q},\omega))}$, we obtain (12.55) with $K = C'/6$. □

The averaging is symmetric, and no bound on the first latitude derivative is needed.

Proposition 12.20 (Lower tail of the band partition function under the Palm law)**.** *For every fixed $b \in (0, b_0]$,*

$$\lim_{\delta\downarrow 0}\limsup_{N\to\infty}\sup_{(e,k)\in\mathcal{K}_*} \mathbb{P}_{e,k,u}^{\mathrm{Palm}}\Big(Z_H(B_u) < \delta D_N e^{\beta(Ne-m_N)}\Big) = 0. \tag{12.56}$$

The same holds for the Palm laws at energy E, uniformly for $|E/N - E_0| \le \varepsilon_0$.

Proof. Let $(e,k) \in \mathcal{K}_*$ and write $\widehat{D}_u = D_u/d_N(k)$. For every event A of the whole field, Proposition 7.7 and the Cauchy–Schwarz inequality give

$$\mathbb{P}^{\mathrm{Palm}}_{e,k,u}(A) = \mathbb{E}_{e,k,u}\big[\widehat{D}_u \mathbf{1}_A\big] \le \mathsf{K}_2^{1/2}\, \mathbb{P}_{e,k,u}(A)^{1/2}. \tag{12.57}$$

This holds even when A depends on the quadratic component, which is correlated with the determinant. Let A_1 be the event in (12.16) with $I = I_{b_0}$ and $\mathcal{K}_0 = \mathcal{K}_*$; by (12.57), we have $\mathbb{P}^{\mathrm{Palm}}_{e,k,u}(A_1) \le Ce^{-cN/2}$. On A_1^c, Lemma 12.19 applies with the constant C_0 of (12.16). By (12.20), we have $Z_u(\sqrt{Q}) = e^{\beta N e} e^{N\beta^2 \psi_Q(1)/2} L_N^{\mathrm{res}}$. Together with (12.55) and (12.5), this gives

$$\frac{Z_H(B_u)}{D_N e^{\beta(Ne - m_N)}} \ge 2e^{-K} L_N^{\mathrm{res}} \qquad \text{on } A_1^c.$$

Applying (12.57) to the event $\{L_N^{\mathrm{res}} < \delta e^K/2\}$, whose $\mathbb{P}_{e,k,u}$-probability does not depend on (e,k), we obtain

$$\sup_{(e,k)\in\mathcal{K}_*} \mathbb{P}^{\mathrm{Palm}}_{e,k,u}\Big(Z_H(B_u) < \delta D_N e^{\beta(Ne-m_N)}\Big) \le Ce^{-cN/2} + \mathsf{K}_2^{1/2}\, \mathbb{P}\big(L_N^{\mathrm{res}} < \delta e^K/2\big)^{1/2}.$$

Letting $N \to \infty$ and then $\delta \downarrow 0$, we obtain (12.56) from Lemma 12.18. The Palm law at energy E is an average of the laws $\mathbb{P}^{\mathrm{Palm}}_{E/N,k,u}$ over $k \in U$, and the uniform bound passes to it. □

The determinant weight is never removed from the Palm law: both the regularity failure and the lower tail of the residual are transferred to the Palm law through (12.57).

Corollary 12.21 (Few light bands, and one heavy band)**.** *Let $b \in (0, b_0]$. We call a critical point u of H_Q δ-*light *if $Z_H(B_u) < \delta D_N e^{\beta(H_Q(u) - m_N)}$.*

1. *For every fixed $L > 0$,*

$$\lim_{\delta\downarrow 0} \limsup_{N\to\infty} \mathbb{E}\,\#\{u \in \mathcal{C}_{N,L} : u \text{ is } \delta\text{-light}\} = 0. \tag{12.58}$$

2. *For every $\varepsilon' > 0$, there exist $L < \infty$ and $\delta > 0$ such that $\limsup_N \mathbb{P}(Z_{N,\beta} < \delta e^{-\beta L} D_N) \le 2\varepsilon'$.*

Proof. Fix L. For large N, the window $[m_N - L, m_N + L]$ lies in $N[E_0 - \varepsilon_0, E_0 + \varepsilon_0]$. By Lemma 12.9, the indicator that u is δ-light is a Borel weight, and it is rotation covariant. Then (12.13) implies that

$$\begin{aligned}\mathbb{E}\,\#\{u \in \mathcal{C}_{N,L} : u \text{ is } \delta\text{-light}\} &= \int_{m_N - L}^{m_N + L} \lambda_N^{\mathrm{all}}(E)\, \mathbb{P}^{\mathrm{Palm}}_{E,u}\Big(Z_H(B_u) < \delta D_N e^{\beta(E - m_N)}\Big)\, \mathrm{d}E \\ &\le q_N(\delta)\, \mathbb{E}|\mathcal{C}_{N,L}|,\end{aligned}$$

where $q_N(\delta)$ is the supremum in (12.56). We have $\mathbb{E}|\mathcal{C}_{N,L}| \le K_L < \infty$ by Lemma 12.2, and (1) follows from Proposition 12.20.

For (2), we choose L by (12.14) so that $\limsup_N \mathbb{P}(\mathcal{C}_{N,L} = \varnothing) \le \varepsilon'$, and then δ by (1) so that the limit superior of the expected number of δ-light centers in $\mathcal{C}_{N,L}$ is at most ε'. By Markov's inequality, the probability that some center of $\mathcal{C}_{N,L}$ is δ-light has limit superior at most ε'. Outside these two events, there is a center $u \in \mathcal{C}_{N,L}$ that is not δ-light, and

$$Z_{N,\beta} \ge Z_H(B_u) \ge \delta D_N e^{\beta(H_Q(u) - m_N)} \ge \delta e^{-\beta L} D_N. \qquad \square$$

Part (2) uses a single band and the existence of one center in the window. It needs neither many centers nor any independence between the bands of different centers. If existence is proved for a subfamily of centers selected by an additional event, the expected number of light centers in the subfamily is bounded by the count in (1) without the selection, since the indicator of the selection event can be dropped from the nonnegative integrand in (12.13). No Palm law normalized after an additional selection is needed.

12.8 Gibbs concentration on bands in a fixed window

Theorem 12.22 (Gibbs concentration on bands in a fixed window)**.** *Let ξ be a finite mixture with nonnegative coefficients and minimum degree at least four, with any set of degrees of either parity, and let $\beta > 0$ be such that (ξ, β) is strictly 1RSB with parameters $x, Q \in (0,1)$ and a nondegenerate contact at Q (Definition 1.1). Let m_N be the centering of Theorem 11.14. There exist a closed interval $U \subset (2\sqrt{b_Q}, \infty)$ containing K_0 in its interior and $b_0 \in (0, \sqrt{Q})$ with the following property. Let $b \in (0, b_0]$, and let*

$$\mathcal{C}_{N,M} = \big\{u \in S_N : \nabla_{S_N} H_Q(u) = 0,\ |H_Q(u) - m_N| \le M,\ K_Q(u)/N \in U\big\},$$
$$B_{N,M} = \bigcup_{u \in \mathcal{C}_{N,M}} \big\{\sigma \in S_N : |R(\sigma, u) - \sqrt{Q}| \le b\big\}.$$

Then the following hold.

1. *For every fixed $M > 0$, $\sup_N \mathbb{E}|\mathcal{C}_{N,M}| < \infty$, and* (12.14) *holds.*
2. *We have*
$$\lim_{M \to \infty} \limsup_{N \to \infty} \mathbb{E}\, \pi\big(S_N \setminus B_{N,M}\big) = 0. \tag{12.59}$$

Proof. Let U, b_0, and ε_0 be as in Section 12.5; then $b_0 < \sqrt{Q}$ since $I_{b_0} \subset (0,1)$. Part (1) is Lemma 12.2. For part (2), fix $b \in (0, b_0]$ and $\tau = \tau_0$, the number of Corollary 12.17, and consider the coarse family $\mathcal{C}(I_\tau)$ of Section 12.3. It consists of critical points of H_Q with radial derivative in U, and its bands are the bands B_u of the statement. Its graph $\{(H, u) : u \in \mathcal{C}(I_\tau)\}$ is closed, by the argument of Lemma 12.9. It is finite whenever H_Q is a Morse function on S_N, hence almost surely by Theorem 7.17. The set of fields for which

it is infinite is Borel, by the argument of Lemma 4.3; on this null set we replace the coarse family by the empty set, which preserves the Borel graph and changes no expectation. We verify the hypotheses of Proposition 4.9, with the deterministic scale D_N of (12.5):

- the existence of centers and the upper tightness of $\max_{S_N} H_Q - m_N$ hold by part (1);
- the bands of the coarse family carry almost all of the Gibbs mass, by Proposition 12.10;
- for every fixed L, the expected number of δ-light centers in $\mathcal{C}_{N,L}$ tends to zero in the limit $N \to \infty$ followed by $\delta \downarrow 0$, by Corollary 12.21(1);
- for every fixed M, the coarse centers with energy below $m_N - M$ satisfy $\limsup_N D_N^{-1} \mathbb{E} \sum_u Z_H(B_u) \le Ce^{-a_0 M}$, by Corollary 12.17, since $\mathcal{C}(I_\tau) \cap \{H_Q < m_N - M\} = \mathcal{C}([N(E_0 - \tau), m_N - M))$.

Proposition 4.9 gives (12.59). The sets $B_{N,M}$ increase with M, so the limit in M exists. □

Remark 12.23 (How the inputs combine). For the reader's convenience, we recall how the four inputs give (12.59). This is the proof of Proposition 4.9, specialized to the coarse family $\mathcal{C}(I_\tau)$. Fix $\varepsilon' > 0$, and choose L and δ by Corollary 12.21(2); the light-band estimate enters only here. Then choose M. All three parameters are fixed before $N \to \infty$. On the event $\{\max H_Q \le m_N + M\}$, every coarse center is either in $\mathcal{C}_{N,M}$ or has energy below $m_N - M$. On the intersection of this event with $\{Z_{N,\beta} \ge \delta e^{-\beta L} D_N\}$, a union bound gives

$$\pi\big(S_N \setminus B_{N,M}\big) \le \pi\big(S_N \setminus B_N^{\mathrm{coarse}}(\tau)\big) + \frac{e^{\beta L}}{\delta D_N} \sum_{u \in \mathcal{C}([N(E_0-\tau), m_N - M))} Z_H(B_u);$$

overlaps between bands only increase the sum. Bounding the Gibbs mass by one outside these events and using Proposition 12.10 and Corollary 12.17, we obtain

$$\limsup_{N\to\infty} \mathbb{E}\, \pi\big(S_N \setminus B_{N,M}\big) \le 2\varepsilon' + \limsup_{N\to\infty} \mathbb{P}\Big(\max_{S_N} H_Q > m_N + M\Big) + \frac{C}{\delta} e^{\beta L - a_0 M}.$$

The right side is at most $3\varepsilon'$ for all large M by (12.14). Since ε' is arbitrary, (12.59) follows.

Remark 12.24 (Even models and antipodal centers). In an even model, $H_N(-\sigma) = H_N(\sigma)$, the critical points of H_Q come in antipodal pairs, and we count both points of each pair. The bands B_u and B_{-u} are reflections of each other and have equal partition functions. For an antipodal class $[u] = \{u, -u\}$, let $B_{[u]} = \{\sigma : ||R(\sigma, u)| - \sqrt{Q}| \le b\}$. Since $b < \sqrt{Q}$, this is the disjoint union $B_u \cup B_{-u}$, and $Z_H(B_{[u]}) = 2Z_H(B_u)$. Keeping the same centering m_N, counting classes once leaves $B_{N,M}$ unchanged, halves the unweighted center counts, and leaves the sums of band partition functions in Corollary 12.17 unchanged. A class is δ-light precisely when either representative is $\delta/2$-light, so the number of light classes is one half of the number of oriented $\delta/2$-light centers. Thus Corollary 12.21 and (12.59) hold with the same sequential limits. In a model with odd degrees, the proof uses only the bands of positive latitude.

Remark 12.25 (What the proof uses). The upper band estimate, Proposition 12.15, combines the static identities of Proposition 5.11 with the bounded GOE determinant prefactors of Lemma 12.11, conditional Gaussian regression, and the second-moment estimate for the quadratic component in Lemma 12.12. The lower band estimate, Proposition 12.20, uses the conditional regularity of Lemma 12.3, symmetric latitude averaging (Lemma 12.19), and the lower-tail tightness of the strictly replica-symmetric residual partition function (Lemma 12.18), obtained from the theorem of Huang, Montanari, and Pham stated in Lemma 3.19. The determinant moments of Proposition 7.7 transfer these lower-tail estimates to the Palm law. The coarse cover uses in addition the directing measure (Theorem 6.3), Subag's replicated band estimates, and the localization (12.11). The strict inequality $c_0 < \beta$ is used only in Corollary 12.17. The proof does not use a limit theorem for the extremal process of the critical points, an independence between a band and the Hessian at its center, a second moment of the full residual partition function, or an evenness assumption.

SECTION 13

Proofs of the condensed-phase theorems

In this section, we prove Theorem 1.2, Corollary 1.3, and Theorem 1.4. The proofs combine the reduction of Section 4 with the inputs proved in Sections 5 to 12. The one new estimate is Proposition 13.1. It states that, for the critical points of the center field in any fixed energy window, the expected probability that the diffusion started from ν_N reaches $C_a(u)$ for one of them before the time $e^{\gamma N}$ is at most a negative power of N, where γ is chosen before the window. Table 13.1 lists the section that supplies each input, and Table 13.2 lists the parameters and the order in which they are fixed.

We use the notation of Section 4.1, in particular Definition 4.1. Recall that $a_Q = Q\xi'(Q)$ and $b_Q = Q^2\xi''(Q)$. Throughout, ξ is a nonzero finite mixture with $c_p \geq 0$ and minimum degree $d \geq 4$, and we set

$$\alpha = \frac{d-2}{2}.$$

13.1 A common time horizon for all fixed energy windows

The next proposition combines the hitting estimate for the caps under critical-point conditioning of Theorem 8.2, in the uniform form of Corollary 8.3, which concerns the ordinary Gaussian conditioning, with the transfer from Gaussian conditioning to the Palm law in Theorem 7.23. Its point is the order of the parameters: the set $C_a(u)$ and the time $e^{\gamma N}$ are chosen once, from a compact set of marks that contains every fixed window, and only the constants depend on the window half-width M.

Proposition 13.1 (A common time horizon for all fixed energy windows)**.** *Let ξ be a finite mixture with $c_p \geq 0$ and minimum degree $d \geq 4$, let $\beta > 0$ and $Q \in (0,1)$. Let U be a compact interval with $\min U > 2\sqrt{b_Q}$, let $E_0 \in \mathbb{R}$, let (m_N) be a deterministic sequence with $m_N/N \to E_0$, and let $b \in (0, \sqrt{Q})$. Define $\mathcal{C}_{N,M}$ and $B_{N,M}$ by Definition 4.1, and suppose that*

$$A_M = \limsup_{N\to\infty} \mathbb{E}_H|\mathcal{C}_{N,M}(H)| < \infty \qquad \text{for every fixed } M > 0.$$

Then there exist $a \in (0, \sqrt{Q} - b)$ and $\gamma > 0$, depending only on ξ, β, Q, U, E_0, and b, such that the following holds with $T_N = e^{\gamma N}$ and with θ_N as in Proposition 4.5. For every fixed $M > 0$ and $r > 1$, there exists $C_{M,r} < \infty$ such that, for all sufficiently large N,

$$\mathbb{E}_H \sup_{0\leq t\leq T_N} (\nu_N P_t^H)\big(B_{N,M}(H)\big) \leq \mathbb{E}_H \sum_{u\in\mathcal{C}_{N,M}(H)} \theta_N(H,u) \leq C_{M,r} N^{-\alpha(1-1/r)}. \tag{13.1}$$

If, in addition, the bands carry almost all of the Gibbs mass in the sense of (4.2)*, then*

$$\inf_{0\le t\le T_N} \left\| \nu_N P_t^H - \pi^H \right\|_{\mathrm{TV}} \longrightarrow 1 \qquad \text{in } \mathbb{P}_H\text{-probability.}$$

Note that (13.1) bounds the endpoint mass $q_{N,M} = \mathbb{E}_H(\nu_N P_{T_N}^H)(B_{N,M}(H))$ of Section 4.5, and the hitting probability $\mathbb{E}_H \mathbb{P}_{\nu_N}^H\{X_t \in C_a(u) \text{ for some } t \le T_N \text{ and } u \in \mathcal{C}_{N,M}(H)\}$, which lies between the two sides of (13.1) by (4.8). The supremum over t in (13.1) is a Borel function of H: by reversibility, the density of $\nu_N P_t^H$ with respect to π^H is $P_t^H(\,\mathrm{d}\nu_N/\,\mathrm{d}\pi^H)$, which is continuous in t in $L^2(\pi^H)$, so the supremum may be taken over rational t.

Proof. We fix a compact set of marks that contains every fixed window, choose a and γ from the conditioned hitting estimate on this set, and then insert the determinant weight using Theorem 7.23.

The compact set of marks. Let $g = \min U - 2\sqrt{b_Q} > 0$. In a pure model $\xi(r) = cr^p$ we have $K_Q = pH_Q$ identically, and only marks on the line $\{k = ph\}$ occur. We therefore set $\mathcal{L}_\xi = \mathbb{R}^2$ if ξ is not pure and $\mathcal{L}_\xi = \{(h,k) : k = ph\}$ if $\xi(r) = cr^p$, and define

$$\mathcal{K} = \big([E_0 - 1, E_0 + 1] \times U\big) \cap \mathcal{L}_\xi, \qquad \mathcal{W}_{N,M} = \big([(m_N - M)/N, (m_N + M)/N] \times U\big) \cap \mathcal{L}_\xi.$$

The set $\mathcal{K}$ is compact and is fixed before M, and $\mathcal{C}_{N,M}(H)$ is exactly the set of critical points of H_Q whose normalized marks $(H_Q(u)/N, K_Q(u)/N)$ lie in $\mathcal{W}_{N,M}$. Fix $M > 0$. Since $m_N/N \to E_0$, there exists $N_1(M)$ such that $\mathcal{W}_{N,M} \subset \mathcal{K}$ for $N \ge N_1(M)$. If $\mathcal{K} = \varnothing$, then $\mathcal{C}_{N,M}(H) = \varnothing$ for every fixed M and all $N \ge N_1(M)$, so (13.1) holds with, for example, $a = (\sqrt{Q} - b)/2$ and $\gamma = 1$; then (4.2) cannot hold, and the last assertion is vacuous. We may therefore assume that $\mathcal{K} \ne \varnothing$. Every point $(h,k) \in \mathcal{K}$ satisfies $k \ge 2\sqrt{b_Q} + g$, which is the strict gap required in Proposition 7.7 and Theorem 7.23, and in the pure case these results and Corollary 8.3 use conditioning on the value and the gradient only.

The time scale. Corollary 8.3, applied at the radius $\sqrt{Q}$ to the compact set $\mathcal{K}$, provides $a_* > 0$. We fix $a \in (0, \min\{a_*, \sqrt{Q} - b\})$. For this a, by Corollary 8.3 there exist $\gamma, c, C > 0$ and N_0, depending only on ξ, β, Q, $\mathcal{K}$, and a, such that, for $N \ge N_0$, every $u \in S_N$ and every $(h,k) \in \mathcal{K}$,

$$\mathbb{E}_{h,k,u}\, \theta_N(H,u) \le C\big(N^{1-d/2} + N^{3-d}\big) + e^{-cN}, \qquad T_N = e^{\gamma N}. \tag{13.2}$$

Here $\mathbb{E}_{h,k,u}$ is the expectation with respect to the law of the whole field conditioned on $H_Q(u) = Nh$, $K_Q(u) = Nk$, and $\nabla_{S_N} H_Q(u) = 0$, as in Theorem 8.2, and the event in θ_N is $\{\sup_{t\le T_N} |R(X_t,u)| \ge a\}$. Since $d \ge 4$, we have $3 - d \le 1 - d/2 = -\alpha$, and the right side of (13.2) is at most $C_1 N^{-\alpha}$, where $C_1 = 2C + \sup_{N\ge 1} N^\alpha e^{-cN}$.

The determinant weight. Recall from Lemma 4.3(b) that θ_N is a Borel function on $\mathcal{H}_N \times S_N$ with values in $[0,1]$. Fix $r > 1$. For $N \geq \max\{N_0, N_1(M)\}$, Theorem 7.23, applied to the Borel window $\mathcal{W}_{N,M} \subset \mathcal{K}$ with the bound (13.2), implies that

$$\mathbb{E}_H \sum_{u \in \mathcal{C}_{N,M}(H)} \theta_N(H,u) \leq \mathsf{K}_r^{1/r} \big(C_1 N^{-\alpha}\big)^{1-1/r} \, \mathbb{E}_H |\mathcal{C}_{N,M}(H)|,$$

where K_r is the bound on the rth moment of the normalized determinant over $\mathcal{K}$ from Proposition 7.7; it does not depend on N or M. This is the argument of Corollary 7.24. Since $\mathbb{E}_H|\mathcal{C}_{N,M}| \leq A_M + 1$ for all sufficiently large N, we obtain the second inequality in (13.1) with

$$C_{M,r} = \mathsf{K}_r^{1/r} C_1^{1-1/r} (A_M + 1).$$

The first inequality in (13.1) follows from (4.8), since $B_{N,M}(H) \subset \bigcup_{u \in \mathcal{C}_{N,M}(H)} C_a(u)$ for $a < \sqrt{Q} - b$.

Separation. The bound just proved, with any fixed $r > 1$, is hypothesis (iii) of Proposition 4.5 with $q_{N,M} = \mathsf{K}_r^{1/r}(C_1 N^{-\alpha})^{1-1/r}$. Hypothesis (ii) is assumed, and hypothesis (i) is the assumption (4.2). We conclude the last assertion from Proposition 4.5. The constants $C_{M,r}$ and the threshold for N depend on M and r, but a and γ do not. □

13.2 Proof of the main condensed-phase theorem

Table 13.1 lists the inputs to the proofs below. The estimates for the concentration of the Gibbs measure on bands are proved by different arguments for pure models and for non-pure mixtures, and the table has a column for each.

Proof of Theorem 1.2. Let (ξ, β) be strictly 1RSB with parameters $x, Q \in (0,1)$ and a non-degenerate contact at Q (Definition 1.1), and let $d \geq 4$ be the minimum degree of ξ.

Pure models. Suppose first that $\xi(r) = cr^p$ with $c > 0$ and $p \geq 4$, and set $\beta' = \beta\sqrt{c}$. Then $H_N = \sqrt{c}\, H_N^{(p)}$, where $H_N^{(p)}$ has covariance $NR(x,y)^p$, and $\beta H_N = \beta' H_N^{(p)}$. Hence the Gibbs measure $\pi_{N,\beta}^{H_N}$ and the generator L_{β,H_N} are those of the model r^p at inverse temperature β', and so is the distance $\big\|\nu_N P_t^{H_N} - \pi^{H_N}\big\|_{\mathrm{TV}}$ for every t. The function V of Definition 1.1 depends on ξ and β only through $\beta^2\xi = (\beta')^2 r^p$, so (r^p, β') satisfies the 1RSB assumptions with the same x and Q. Lemma 5.10 implies that $\beta > \beta_{\mathrm{s}}(cr^p) = \beta_{\mathrm{s}}(p)/\sqrt{c}$, that is, $\beta' > \beta_{\mathrm{s}}(p)$. It therefore suffices to prove the theorem for the model r^p at the inverse temperature $\beta' > \beta_{\mathrm{s}}(p)$, and in the pure case we work with (r^p, β') in the rest of the proof. This is the setting of Section 9, whose results are stated for $\xi(r) = r^p$ and $\beta > \beta_{\mathrm{s}}(p)$; we apply them with β' in place of β.

Input	Pure models	Non-pure mixtures
Static consequences of the 1RSB assumptions: Q, E_0, $K_0 > 2\sqrt{b_Q}$; $\beta > \beta_s(\xi)$	Proposition 5.11 and Lemma 5.10 (Section 5)	Proposition 5.11 and Lemma 5.10 (Section 5)
Directing measure carried by vectors of squared norm Q (used inside the proofs of concentration on bands)	Theorem 6.3 (Section 6)	Theorem 6.3 (Section 6)
Determinant moments, marked Kac–Rice formula, determinant-weighted transfer	Proposition 7.7 and Theorems 7.17 and 7.23 (Section 7)	Proposition 7.7 and Theorems 7.17 and 7.23 (Section 7)
Conditioned hitting estimate for the caps	Theorem 8.2 and Corollary 8.3 (Section 8)	Theorem 8.2 and Corollary 8.3 (Section 8)
Bounded expected counts and tightness of the maximum	Proposition 9.1 (Section 9)	Theorem 11.14 (Sections 10 and 11)
Gibbs concentration on the bands	Theorem 9.2 (Section 9)	Theorem 12.22 (Section 12)
Separation, rotations, initial laws	Section 4	Section 4

Table 13.1. The inputs to the proofs of Theorems 1.2 and 1.4. The membership statements Propositions 5.17, 5.18, 5.20, 5.23, and 5.24 enter only Corollary 1.3.

The 1RSB parameters. By Proposition 5.11, applied to (ξ, β) for a non-pure mixture and to (r^p, β') in the pure case, $x\delta_0 + (1-x)\delta_Q$ is the unique minimizer of the Crisanti–Sommers functional. In particular, x and Q are determined by the model, and in the pure case Q is the overlap used in Section 9. We also obtain from Proposition 5.11 the ground-state parameters E_0 and K_0, with $K_0 > 2\sqrt{b_Q}$.

Counts and concentration on bands. In the pure case, we fix a compact interval U with K_0 in its interior and $\min U > 2\sqrt{b_Q}$ and let $m_N = NE_0 + O(\log N)$ be the centering of Section 9. By Proposition 9.1 and Theorem 9.2, applied to r^p at inverse temperature β', the critical sets $\mathcal{C}_{N,M}$ of Definition 4.1 satisfy $\limsup_N \mathbb{E}_H|\mathcal{C}_{N,M}| < \infty$ for each fixed M, and there exists $b_0 \in (0, \sqrt{Q})$ such that the bands of every width $b \in (0, b_0]$ carry almost all of the Gibbs mass in the sense of (4.2). If ξ is not pure, let $m_N = NE_0 - \log N/(2c_0) + O(1)$ be the centering of Theorem 11.14. Then Theorem 12.22 provides a closed interval $U \subset (2\sqrt{b_Q}, \infty)$ with K_0 in its interior and $b_0 \in (0, \sqrt{Q})$ with the same two properties, at the radius Q of Proposition 5.11. In both cases we fix $b = b_0$. When the concentration is stated for antipodal classes of an even model, with the bands $\{x : ||R(x,u)| - \sqrt{Q}| \le b\}$,

the union of these bands is $B_{N,M}(H)$, since $\mathcal{C}_{N,M}(H) = -\mathcal{C}_{N,M}(H)$ and $B_{-u} = -B_u$; the expected counts differ by a factor of two.

Conclusion. The hypotheses of Proposition 13.1 hold for these Q, U, E_0, (m_N), and b, since $m_N/N \to E_0$; in the pure case we apply that proposition to r^p at inverse temperature β'. From that proposition, we obtain a and $\gamma > 0$ such that $\inf_{0 \le t \le e^{\gamma N}} \left\| \nu_N P_t^H - \pi^H \right\|_{\mathrm{TV}} \to 1$ in $\mathbb{P}_H$-probability, and by the first paragraph this is the conclusion for (ξ, β). The objects Q, U, (m_N), and b are determined by (ξ, β) through the results cited above, and therefore $\gamma = \gamma(\xi, \beta)$. The argument uses no restriction on the number of supported degrees, on their parity, or on the sizes of the coefficients. □

The parameters in this proof, and the moment order r optimized for Theorem 1.4(ii), are fixed in the order of Table 13.2. Every entry is fixed before $N \to \infty$.

Step	Parameter	Depends on	Source
1	x, Q; ground-state parameters E_0, K_0	ξ, β	Proposition 5.11
2	compact interval $U \ni K_0$, $\min U > 2\sqrt{b_Q}$	ξ, β	Proposition 9.1, or Theorems 11.14 and 12.22
3	centering $m_N = NE_0 + O(\log N)$	ξ, β, U	same
4	band width $b \in (0, b_0]$	ξ, β, U	Theorem 9.2 or Theorem 12.22
5	compact set $\mathcal{K} = ([E_0 - 1, E_0 + 1] \times U) \cap \mathcal{L}_\xi$	E_0, U	Proposition 13.1
6	$a \in (0, \min\{a_*, \sqrt{Q} - b\})$	steps 1–5	Corollary 8.3
7	time scale $T_N = e^{\gamma N}$	steps 1–6	Corollary 8.3
8	moment order $r > 1$ ($r = 2$ for separation; optimized for prescribed-law selection)	d, κ	Proposition 13.1
9	tolerances ε, δ; window half-width M	steps 1–8	Lemma 4.4
10	$N \ge N_0(M, r)$, then $N \to \infty$	steps 1–9	

Table 13.2. The parameters in the proofs of Theorems 1.2 and 1.4, in the order in which they are fixed. Only the constants and the threshold N_0 depend on the window half-width M; the time scale does not.

13.3 Proofs of the corollary and of the theorem on initial laws

Proof of Corollary 1.3. In each case, we check that (ξ, β) is strictly 1RSB with a nondegenerate contact at Q and that the minimum degree is at least four. Then Theorem 1.2 implies the conclusion.

For (a), let $\xi(r) = r^p$ with $p \geq 4$ and $\beta > \beta_s(p)$. By Proposition 5.20, the pair (ξ, β) satisfies the 1RSB assumptions, for p of either parity.

For (b), let ξ have minimum degree $d \geq 4$, let $r \mapsto \xi''(r)^{-1/2}$ be strictly convex on $(0,1]$, and let $\beta > \beta_s(\xi)$. Proposition 5.17 implies that (ξ, β) satisfies the 1RSB assumptions. If the smallest and largest supported degrees $d \geq 4$ and P satisfy $6(d-2)(P-2) > (d+P-5)^2$, then by Proposition 5.18 the function $\xi''^{-1/2}$ is strictly convex on $(0,1]$, and the same proposition shows that this inequality holds whenever all supported degrees lie in $\{4, \dots, 11\}$.

For (c), let $d \geq 4$ be even, let $\beta > \beta_s(d)$, and let $\varepsilon_0 = \varepsilon_0(d, P, \beta)$ be given by Proposition 5.23. For every fixed $\varepsilon \in (0, \varepsilon_0)$ and every nonnegative vector (b_p), indexed by the even p with $d < p \leq P$ and with sum one, that proposition shows that $\xi(r) = r^d + \varepsilon \sum_p b_p r^p$ satisfies the 1RSB assumptions at β. Its minimum degree is d.

For (d), Proposition 5.24 shows that the explicit covariance $r^4 + br^{12}$ satisfies the 1RSB assumptions at its temperature, and that $\xi''^{-1/2}$ is not convex on $(0,1]$; this case is not contained in (b). □

For the proof of Theorem 1.4, let (ξ, β) be covered by Theorem 1.2. We keep the objects Q, U, (m_N), b, a, and γ chosen in its proof, and set $T_N = e^{\gamma N}$. That proof shows that the band unions $B_{N,M}(H)$ carry almost all of the Gibbs mass, and they have Borel graphs and satisfy $B_{N,M}(H^g) = gB_{N,M}(H)$ by Lemma 4.3(a). Further, Proposition 13.1 implies that, for each fixed $M > 0$ and $r > 1$,

$$q_{N,M} = \mathbb{E}_H(\nu_N P_{T_N}^H)\big(B_{N,M}(H)\big) \leq C_{M,r} N^{-\alpha(1-1/r)} \qquad \text{for all sufficiently large } N. \quad (13.3)$$

Proof of Theorem 1.4. For (i), the inequality between the expected deficits, for every deterministic μ and $T \geq 0$, is (4.14) of Theorem 4.11. Theorem 1.2 shows that $\inf_{0 \leq t \leq T_N} \|\nu_N P_t^H - \pi^H\|_{\mathrm{TV}} \to 1$ in probability, and the last part of Theorem 4.11 now yields the conclusion at the same time T_N, uniformly over deterministic initial laws, and for every random initial law independent of H_N.

For (ii), let $K_N \leq C_0 N^\kappa$ with $0 \leq \kappa < \alpha$. We choose a fixed $r > \alpha/(\alpha - \kappa)$, so that $\kappa < \alpha(1 - 1/r)$. Using (13.3), we find that, for each fixed M,

$$K_N q_{N,M} \leq C_0 C_{M,r} N^{\kappa - \alpha(1-1/r)} \longrightarrow 0.$$

Then Corollary 4.14 implies that the conclusion holds simultaneously for all convex mixtures of the prescribed laws, with weights that may depend on the field. The order r is fixed before N, depending only on d and κ, and the time T_N does not depend on it.

For (iii), we take $r = 2$ in (13.3) and find that (4.18) holds with $\zeta = \alpha/2 = (d-2)/4$. By Corollary 4.15, the conclusion holds for every Borel kernel $H \mapsto \mu_N^H$ with $D(\mu_N^H \| \nu_N)/\log N \to 0$ in probability. □

Remark 13.2 (What the theorems do not assert). The conclusion of Theorem 1.2 is convergence in $\mathbb{P}_H$-probability; we do not show that the probability of failure is exponentially small in N. The covariance and the temperature are fixed before $N \to \infty$. We do not claim that γ stays bounded below as β decreases to $\beta_{\mathrm{s}}(\xi)$. The minimum degree is at least four throughout. For a component of degree two or three, the error term N^{3-d} in the estimate of Theorem 8.2 would not tend to zero, and our proof does not cover models with quadratic or cubic components, including the pure cubic model. In Theorem 1.4(ii), the restriction $\kappa < (d-2)/2$ is strict: the transfer in Proposition 13.1 loses the factor $1 - 1/r$, and the endpoint $\kappa = (d-2)/2$ is not covered.

PART III

Below the static threshold

SECTION 14

Constrained pressures

In this section, we prove the static estimates on which Part III rests. Throughout the section, ξ is a finite mixture and $0 < \beta < \beta_s$; in Sections 14.3 and 14.4, we also assume that $\beta > \beta_d$. The central object is the constrained pressure $F_{\mathrm{FP}}(r)$, the free energy per site of the configurations whose overlap with a reference point drawn from the Gibbs measure is close to r (Definition 14.4). It is the Franz–Parisi potential [73], written as a pressure instead of a free-energy cost.

The section has five main statements. In Proposition 14.7, we express $F_{\mathrm{FP}}(r)$ through the Parisi formula for an auxiliary mixture. In Proposition 14.11, we show that F_{FP} lies below an explicit trial pressure Ψ, with equality at q_*; this identifies the maximum of F_{FP} on the cap and gives a strict gap on the collar. Proposition 14.13 turns these limits into quenched estimates, which hold with probability exponentially close to one and simultaneously for all reference points in a set whose Gibbs measure is nearly one. Finally, we record in Proposition 14.14 and Lemma 14.15 that the Gibbs mass of every cap of fixed positive latitude is exponentially small, uniformly over its axis (Corollary 14.16).

From Section 3, we use the representation and regularity of the field (Lemma 3.2), the concentration of free energies (Lemma 3.3), the Crisanti–Sommers functional, the Parisi formula, and the Crisanti–Sommers optimality criterion (Definition 3.13, Theorem 3.14, and Lemma 3.15), the thresholds and the replica-symmetric regime (Definition 3.16 and Lemma 3.18), and the overlap estimate of Huang, Montanari, and Pham (Lemma 3.20). The results of this section are used as follows. The proof of Theorem 1.5 in Sections 15 and 16 uses Proposition 14.11 to choose an admissible collar with positive exponents, and then Proposition 14.13 and Corollary 14.16 for the estimates on the Gibbs measure; the corollary follows from Proposition 14.14 and Lemma 14.15. The proof of Theorem 1.6 in Section 17 also uses the planted law of Section 14.1, the cross-section formula, and the equality at q_*. Section 18 develops the infinite-mixture inputs for Theorem 1.7.

We follow the route of El Alaoui, Montanari, and Sellke, who computed the constrained pressure of pure spherical models through a planted model [65, Propositions 3.7 and 3.8] and bounded it by the trial value of Lemma 14.9 below [65, Proposition 3.9]. Exponential transfer from the planted law to the Gibbs-reference law was proved in [65, Lemma 3.5] and, for general mixtures, in [13, Lemma 2.3]. Planted models have a longer history in the study of random constraint satisfaction [1, 105]. For general finite mixtures, we derive the cross-section formula from these ingredients and the Parisi formula of [46].

The mixed-model cross-section formula is also covered by the Gaussian-conditioning approach of Dembo and Subag [59, Section 2.2, Corollary 2.1]. The equality $F_{\mathrm{FP}}(q_*) = \Psi(q_*)$ requires an additional argument, based on the monotonicity of the ratio $r\xi''(r)/\xi'(r)$, the logarithmic derivative of ξ' with respect to $\log r$.

Notation. We write $\mathbb{P}$ and $\mathbb{E}$ for the law of the field $H = H_N$ and its expectation, $\pi = \pi^H$ for the Gibbs measure, and $Z_{N,\beta} = Z_{N,\beta}(H) = \int e^{\beta H}\,\mathrm{d}\nu_N$ for the partition function, as in Definition 3.1. We set

$$F = \frac{\beta^2 \xi(1)}{2}. \tag{14.1}$$

Then $\mathbb{E}Z_{N,\beta} = e^{NF}$ by (3.1), and $F = F(\beta)$ for $\beta < \beta_{\mathrm{s}}$ by Lemma 3.18(1). We realize the field by the Gaussian tensors of (3.2), and write g for the vector of their coefficients, which has $\sum_{p=2}^{P} N^p$ entries. As in the proof of Lemma 3.3, for every nonzero finite deterministic Borel measure μ on S_N and every bounded deterministic Borel function φ, the function $g \mapsto \log \int e^{\beta H + \varphi}\,\mathrm{d}\mu$ is Lipschitz with constant $\beta(N\xi(1))^{1/2} = (2FN)^{1/2}$, because its gradient is β times an average of vectors of norm $(N\xi(1))^{1/2}$. We use this observation together with the Gaussian concentration inequality in the form [111, Proposition 4.1 and (4.1)], which also underlies Lemma 3.3: if Θ is a Lipschitz function of g with constant ℓ, then

$$\mathbb{P}\big(|\Theta - \mathbb{E}\Theta| > s\big) \le 2e^{-s^2/(2\ell^2)}, \qquad s > 0. \tag{14.2}$$

Finally, let K and c_0 be the constants K and c of Lemma 3.2, so that, for every $N \ge 2$,

$$\mathbb{P}\Big(\sup_{S_N} |H| \le KN,\ \sup_{S_N} |\nabla_{S_N} H| \le K\sqrt{N}\Big) \ge 1 - e^{-c_0 N}. \tag{14.3}$$

14.1 The planted and Gibbs-reference laws

Definition 14.1 (Planted and Gibbs-reference laws)**.** The *Gibbs-reference law* Q_N is the joint law of a field H with law $\mathbb{P}$ and a reference point $u \in S_N$ which, conditionally on H, has law π^H:

$$\mathsf{Q}_N(dH, du) = \mathbb{P}(dH)\,\pi^H(du).$$

The *planted law* P_N is

$$\mathsf{P}_N(dH, du) = e^{\beta H(u) - NF}\,\mathbb{P}(dH)\,\nu_N(du).$$

For every $u \in S_N$, we write $\mathsf{P}_N(\cdot \mid u)$ for the conditional law of H given u defined by the kernel

$$\mathsf{P}_N(dH \mid u) = e^{\beta H(u) - NF}\,\mathbb{P}(dH).$$

We call

$$L_N(H) = \frac{Z_{N,\beta}(H)}{\mathbb{E}Z_{N,\beta}} = Z_{N,\beta}(H)e^{-NF}$$

the *likelihood ratio*.

The following lemma collects the basic properties of the planted law. They are standard (see [13, Eq. (2.4)] and [65, Proposition 3.7]).

Lemma 14.2 (The planted law)**.** *The measure P_N is a probability measure, and the law of u under P_N is ν_N. For every $u \in S_N$, under $\mathsf{P}_N(\cdot \mid u)$,*

$$H(x) = \widetilde{H}(x) + \beta N \xi(R(x,u)), \qquad x \in \mathbb{R}^N, \tag{14.4}$$

where $\widetilde{H}$ has law $\mathbb{P}$. Finally, P_N and Q_N are mutually absolutely continuous, and

$$\frac{d\mathsf{P}_N}{d\mathsf{Q}_N}(H,u) = L_N(H). \tag{14.5}$$

Proof. Fix $u \in S_N$ and let $X = \beta H(u)$, a centered Gaussian variable with variance $2NF$. Then $\mathbb{E}e^X = e^{NF}$, which implies that P_N is a probability measure whose u-marginal is ν_N, and that the kernel $\mathsf{P}_N(\cdot \mid u)$ of Definition 14.1 is a version of the conditional law of H given u. By definition, $\mathsf{P}_N(\cdot \mid u)$ has density e^{X-NF} with respect to $\mathbb{P}$. For points $x_1, \dots, x_k \in \mathbb{R}^N$ and real numbers $\lambda_1, \dots, \lambda_k$, let $Y = \sum_i \lambda_i H(x_i)$. Because (X, Y) is a centered Gaussian vector, we have

$$\mathbb{E}\big[e^{X-NF}e^{Y}\big] = \exp\Big(\frac{1}{2}\operatorname{Var} Y + \operatorname{Cov}(X,Y)\Big) = \exp\Big(\frac{1}{2}\operatorname{Var} Y + \sum_i \lambda_i \beta N \xi(R(x_i,u))\Big),$$

where we used $\operatorname{Var} X = 2NF$ and $\mathbb{E}H(x)H(u) = N\xi(R(x,u))$ from Lemma 3.2. The right side is the Laplace transform of $(H(x_i))_{i\le k}$ shifted by $(\beta N\xi(R(x_i,u)))_{i\le k}$. Hence the finite-dimensional distributions of H under $\mathsf{P}_N(\cdot \mid u)$ are those of $\widetilde{H} + \beta N\xi(R(\cdot,u))$, where $\widetilde{H}$ has law $\mathbb{P}$. As H is a random polynomial of bounded degree, these finite-dimensional distributions determine its law (see the proof of Lemma 3.2(2)), which proves (14.4). Finally, $\mathsf{Q}_N(dH, du) = \mathbb{P}(dH)\, Z_{N,\beta}(H)^{-1} e^{\beta H(u)} \nu_N(du)$. Comparing this with the definition of P_N, we see that (14.5) holds. The two laws are mutually absolutely continuous because $0 < L_N < \infty$. □

Note that (14.4) describes an exponential tilt of the field in the direction of u, and that the value $H(u)$ is not conditioned. The likelihood ratio depends on H only, and its logarithm concentrates at exponential scale by Lemma 3.18(3). As a consequence, events that are exponentially rare under the planted law are also exponentially rare under the Gibbs-reference law. The next lemma is the quantitative form of [65, Lemma 3.5] and [13, Lemma 2.3]. We include the short proof to record the rate.

Lemma 14.3 (Exponential likelihood transfer)**.** *Let ξ be a finite mixture and $0 < \beta < \beta_{\mathrm{s}}$.*

(a) For every $\varepsilon > 0$, for all sufficiently large N,

$$\mathbb{P}\big(|N^{-1}\log Z_{N,\beta} - F| > \varepsilon\big) \le 2e^{-\varepsilon^2 N/(16F)}. \tag{14.6}$$

(b) Let $c > 0$, and let E_N be Borel sets of pairs (H, u) such that $\mathsf{P}_N(E_N) \le e^{-cN}$ for all sufficiently large N. Then, for all sufficiently large N,

$$\mathsf{Q}_N(E_N) \le 3e^{-c'N}, \qquad c' = \min\Big\{\frac{c}{2}, \frac{c^2}{64F}\Big\}. \tag{14.7}$$

Proof. Part (a) is (3.21) in Lemma 3.18(3), since $2F = \beta^2\xi(1)$. For (b), fix $\tau = c/2$. Since $d\mathsf{Q}_N/d\mathsf{P}_N = L_N^{-1}$ by (14.5), we have

$$\begin{aligned}\mathsf{Q}_N(E_N) &\le \mathsf{Q}_N\big(L_N < e^{-\tau N}\big) + \int_{E_N \cap \{L_N \ge e^{-\tau N}\}} L_N^{-1}\, \mathrm{d}\mathsf{P}_N \\ &\le \mathbb{P}\big(N^{-1}\log Z_{N,\beta} < F - \tau\big) + e^{\tau N}\mathsf{P}_N(E_N).\end{aligned}$$

In the second line, we used that the law of H under Q_N is $\mathbb{P}$. By (a) with $\varepsilon = \tau$, the first term is at most $2e^{-\tau^2 N/(16F)} = 2e^{-c^2 N/(64F)}$ for large N, and the second is at most $e^{-cN/2}$. Their sum is at most $3e^{-c'N}$. □

14.2 The cross-section formula

For a Borel set $\mathcal{A} \subset [-1, 1]$ and a pair (H, u), we define the restricted partition function

$$Z_{N,\beta}(\mathcal{A}; H, u) = \int_{\{x \in S_N\colon R(x,u) \in \mathcal{A}\}} e^{\beta H(x)}\, \nu_N(dx),$$

so that $\pi\{x\colon R(x,u) \in \mathcal{A}\} = Z_{N,\beta}(\mathcal{A}; H, u)/Z_{N,\beta}(H)$. When $\mathcal{A}$ is an interval, the boundary of $\{x\colon R(x,u) \in \mathcal{A}\}$ is ν_N-null, and $(H, u) \mapsto Z_{N,\beta}(\mathcal{A}; H, u)$ is continuous by dominated convergence.

Definition 14.4 (Constrained pressure)**.** For $r \in (-1, 1)$, the *constrained pressure* is

$$F_{\mathrm{FP}}(r) = \lim_{w \downarrow 0} \lim_{N \to \infty} \frac{1}{N}\, \mathbb{E}_{\mathsf{Q}_N} \log Z_{N,\beta}\big([r - w, r + w]; H, u\big),$$

whenever both limits exist.

We take the limit in N first, with a fixed band width $2w$. We do not normalize the restricted partition function. Since $N^{-1}\log Z_{N,\beta} \to F$, the difference $F_{\mathrm{FP}}(r) - F$ is the exponential rate of the Gibbs mass of the band around a typical reference point. In the convention of Franz and Parisi [73, Section 3, Eqs. (12)–(14)], the potential corresponds to the free-energy cost $(F - F_{\mathrm{FP}}(r))/\beta$.

Sections of the sphere. Fix $u \in S_N$ and let $n = N - 1$. We identify the orthogonal complement $u^\perp$ with $\mathbb{R}^n$, and write $S_n \subset u^\perp$ for the sphere of radius $\sqrt{n}$, ν_n for its uniform probability measure, and $R_n(y, z) = y \cdot z/n$. For $r \in (-1, 1)$ and $y \in S_n$, we set

$$x_r(y) = ru + \Big(\frac{N(1 - r^2)}{n}\Big)^{1/2} y. \tag{14.8}$$

Because $y \perp u$, $|u|^2 = N$, and $|y|^2 = n$, we have

$$|x_r(y)|^2 = N, \qquad R(x_r(y), u) = r, \qquad R(x_r(y), x_r(z)) = r^2 + (1 - r^2) R_n(y, z). \tag{14.9}$$

For $N \geq 3$, we define the section partition function and the section pressure

$$\mathcal{Z}_N(r; H, u) = \int_{S_n} e^{\beta H(x_r(y))}\, \nu_n(dy),$$
$$f_N(r; H, u) = \frac{1}{N} \log \mathcal{Z}_N(r; H, u) + \frac{N-3}{2N} \log(1 - r^2).$$

Auxiliary mixtures. For a covariance polynomial η in the sense of Definition 3.13, not identically zero, we write H_n^η for the centered Gaussian process on S_n with covariance $\mathbb{E} H_n^\eta(y) H_n^\eta(z) = n\eta(R_n(y, z))$, and set

$$p_n(\eta) = \frac{1}{n}\, \mathbb{E} \log \int_{S_n} e^{\beta H_n^\eta}\, \mathrm{d}\nu_n.$$

Recall that a covariance polynomial may have a term of degree one. By the Parisi formula Theorem 3.14, which covers this case through [46, Theorem 1], we have $p_n(\eta) \to \mathcal{P}_\beta(\eta)$, where $\mathcal{P}_\beta(\eta)$ is the infimum of the Crisanti–Sommers functional in Definition 3.13. For $r \in (-1, 1)$, the auxiliary mixture of the section at latitude r is

$$\eta_r(t) = \xi\big(r^2 + (1 - r^2)t\big) - \xi(r^2) = \sum_{k \geq 1} \frac{(1 - r^2)^k \xi^{(k)}(r^2)}{k!}\, t^k. \tag{14.10}$$

Lemma 14.5 (Spherical sections)**.** *Fix $N \geq 3$ and $u \in S_N$.*

(a) For every field H and every Borel set $\mathcal{A} \subset (-1, 1)$,

$$Z_{N,\beta}(\mathcal{A}; H, u) = C_N \int_{\mathcal{A}} e^{N f_N(r; H, u)}\, \mathrm{d}r, \qquad C_N = \frac{\Gamma(N/2)}{\sqrt{\pi}\, \Gamma((N-1)/2)}, \tag{14.11}$$

and $N^{-1} \log C_N \to 0$.

(b) For every $r \in (-1, 1)$, the polynomial η_r is a covariance polynomial, and its coefficient of t is $(1 - r^2)\xi'(r^2)$.

(c) For every $r \in (-1, 1)$, under $\mathsf{P}_N(\cdot \mid u)$, the process $(H(x_r(y)))_{y \in S_n}$ has the law of $(\beta N \xi(r) + W + Y(y))_{y \in S_n}$, where W is a centered Gaussian variable with variance $N\xi(r^2)$ and Y is independent of W and has the law of $H_n^{(N/n)\eta_r}$.

Proof. For (a), recall that if ω is uniform on the unit sphere of $\mathbb{R}^N$ and e is a unit vector, then $\omega \cdot e$ has density $C_N(1 - t^2)^{(N-3)/2}$ on $(-1, 1)$, and, conditionally on $\omega \cdot e = t$, the vector $\omega - te$ is uniform on the sphere of radius $(1 - t^2)^{1/2}$ in $e^\perp$. Applying this to $x = \sqrt{N}\, \omega$ and $e = u/\sqrt{N}$, we see that $R(x, u)$ has density $C_N(1 - r^2)^{(N-3)/2}$ under ν_N

and that, conditionally on $R(x,u) = r$, the point x has the law of $x_r(y)$ with $y \sim \nu_n$. It follows that

$$Z_{N,\beta}(\mathcal{A}; H, u) = C_N \int_{\mathcal{A}} (1-r^2)^{(N-3)/2}\, \mathcal{Z}_N(r; H, u)\, \mathrm{d}r,$$

which is (14.11). By Stirling's formula, $C_N \sim (N/2\pi)^{1/2}$.

For (b), the second expression in (14.10) is the Taylor expansion of the polynomial ξ at $r^2 \geq 0$. Every derivative of ξ is nonnegative on $[0,1]$, because the coefficients of ξ are.

For (c), we have $H(x_r(y)) = \widetilde{H}(x_r(y)) + \beta N \xi(r)$ by (14.4) and (14.9). We use Lemma 3.2(1) and (14.9) to compute the covariance of the centered process $\widetilde{H}(x_r(\cdot))$:

$$N\xi\big(r^2 + (1-r^2)R_n(y,z)\big) = N\xi(r^2) + n \cdot \frac{N}{n}\, \eta_r(R_n(y,z)),$$

which is the covariance of $W + Y$. Note that Y is well defined, since $(N/n)\eta_r$ is a covariance polynomial by (b). Centered Gaussian processes with continuous paths and the same covariance have the same law. □

We see that, under planting, the section at latitude r is a spherical spin glass in dimension n with mixture close to η_r, shifted by the constant $\beta N \xi(r) + W$. In the next lemma, we make the resulting convergence of f_N uniform in r, using a fixed finite net of latitudes. The lemma holds at every temperature.

Lemma 14.6 (Uniform latitude concentration)**.** *Let ξ be a finite mixture and $\beta > 0$. For $r \in (-1,1)$, we define*

$$f(r) = \beta^2 \xi(r) + \frac{1}{2}\log(1-r^2) + \mathcal{P}_\beta(\eta_r). \tag{14.12}$$

(a) For every n and all covariance polynomials η, θ, not identically zero,

$$|p_n(\eta) - p_n(\theta)| \leq \beta^2 \sup_{|t|\leq 1} |\eta(t) - \theta(t)|,$$

and the same bound holds for $\mathcal{P}_\beta$.

(b) For every $\rho \in (0,1)$, the function f is Lipschitz on $[-\rho,\rho]$ with constant

$$L_f(\rho) = 7\beta^2 \xi'(1) + \frac{\rho}{1-\rho^2}.$$

(c) Let $K_1 = K + \beta\xi'(1)$. For every $\rho \in (0,1)$ and $\varepsilon > 0$, there exists N_0 such that, for all $N \geq N_0$ and $u \in S_N$,

$$\mathsf{P}_N\big(\Omega_N(\rho,\varepsilon)^c \mid u\big) \leq e^{-cN}, \qquad c = \frac{1}{2}\min\Big\{c_0, \frac{\varepsilon^2}{64F}\Big\},$$

where $\Omega_N(\rho,\varepsilon)$ is the event that

$$\sup_{S_N} |H| \leq K_1 N \qquad \text{and} \qquad \sup_{|r|\leq\rho} \big|f_N(r; H, u) - f(r)\big| \leq \varepsilon.$$

Proof. For (a), we use Gaussian covariance interpolation [147, Section 1.3, Lemma 1.3.1, and Eqs. (1.55)–(1.60)]. Let H_n^η and H_n^θ be independent, and for $s \in [0,1]$ let $\phi(s) = n^{-1}\mathbb{E}\log\int e^{\beta H_s}\,\mathrm{d}\nu_n$, where $H_s = \sqrt{s}\,H_n^\eta + \sqrt{1-s}\,H_n^\theta$. By Gaussian integration by parts, we have

$$\phi'(s) = \frac{\beta^2}{2}\,\mathbb{E}\Big\langle (\eta-\theta)(1) - (\eta-\theta)\big(R_n(y^1,y^2)\big)\Big\rangle_s,$$

where $\langle\cdot\rangle_s$ is the average over two independent samples y^1, y^2 from the Gibbs measure of H_s. Because $|R_n| \le 1$, this implies that $|\phi'(s)| \le \beta^2 \sup_{|t|\le 1}|\eta(t)-\theta(t)|$. Integrating over $s \in [0,1]$ proves the bound for p_n, and letting $n \to \infty$ proves it for $\mathcal{P}_\beta$.

For (b), note that $|\xi'(v)| \le \xi'(1)$ for $|v| \le 1$. For $|t| \le 1$, the point $r^2 + (1-r^2)t$ lies in $[-1,1]$, and

$$\big|\partial_r \eta_r(t)\big| = 2|r|\,\big|(1-t)\xi'\big(r^2+(1-r^2)t\big) - \xi'(r^2)\big| \le 6\xi'(1).$$

Using (a) and the mean value theorem, we find that $r \mapsto \mathcal{P}_\beta(\eta_r)$ is Lipschitz with constant $6\beta^2\xi'(1)$. The derivatives of $\beta^2\xi(r)$ and $\frac{1}{2}\log(1-r^2)$ are bounded on $[-\rho,\rho]$ by $\beta^2\xi'(1)$ and $\rho/(1-\rho^2)$, respectively.

For (c), fix ρ and ε, and fix $u \in S_N$. The proof has four steps.

Field bounds. Under $\mathsf{P}_N(\cdot \mid u)$, the field $\widetilde{H}$ in (14.4) has law $\mathbb{P}$. The function $x \mapsto \beta N\xi(R(x,u))$ is bounded on S_N by $\beta N\xi(1) \le \beta N\xi'(1)$, and its Euclidean gradient $\beta\xi'(R(x,u))u$ has norm at most $\beta\xi'(1)\sqrt{N}$. By (14.3), with probability at least $1 - e^{-c_0 N}$ under $\mathsf{P}_N(\cdot \mid u)$, we have

$$\sup_{S_N}|H| \le K_1 N, \qquad \sup_{S_N}|\nabla_{S_N} H| \le K_1\sqrt{N}. \tag{14.13}$$

On this event, $f_N(\cdot\,; H, u)$ is Lipschitz on $[-\rho,\rho]$. Indeed, the vector $\partial_r x_r(y) = u - r(1-r^2)^{-1/2}(N/n)^{1/2}y$ is tangent to S_N and has norm $\sqrt{N}(1-r^2)^{-1/2}$. Together with (14.13), this gives $|\partial_r H(x_r(y))| \le K_1 N(1-\rho^2)^{-1/2}$. Differentiating under the integral, we find

$$\big|\partial_r f_N(r; H, u)\big| \le L_H(\rho) = \frac{\beta K_1}{(1-\rho^2)^{1/2}} + \frac{\rho}{1-\rho^2}, \qquad |r| \le \rho. \tag{14.14}$$

The second term bounds the derivative of $\frac{N-3}{2N}\log(1-r^2)$.

Means. Fix $r \in (-1,1)$. Using Lemma 14.5(c) and $\mathbb{E}W = 0$, we find

$$\mathbb{E}_{\mathsf{P}_N(\cdot|u)} f_N(r; H, u) = \beta^2\xi(r) + \frac{N-3}{2N}\log(1-r^2) + \frac{n}{N}\,p_n\Big(\frac{N}{n}\,\eta_r\Big).$$

This does not depend on u. Part (a) gives $|p_n((N/n)\eta_r) - p_n(\eta_r)| \le \beta^2 n^{-1}\sup_{|t|\le 1}|\eta_r(t)| \le 2\beta^2\xi(1)/n$, and $p_n(\eta_r) \to \mathcal{P}_\beta(\eta_r)$. We conclude that $\mathbb{E}_{\mathsf{P}_N(\cdot|u)} f_N(r; H, u) \to f(r)$ for each fixed r.

Fluctuations. For fixed r, the variable $f_N(r; H, u)$ is, up to a deterministic constant, $N^{-1}\log\int e^{\beta\widetilde{H}}\,\mathrm{d}\mu_r$, where μ_r is the image of ν_n under x_r. As noted before (14.2), this is a

Lipschitz function of the Gaussian coordinates of $\widetilde{H}$ with constant $(2F/N)^{1/2}$. By (14.2), we have

$$\mathsf{P}_N\Big(\big|f_N(r;H,u) - \mathbb{E}_{\mathsf{P}_N(\cdot\,|\,u)} f_N(r;H,u)\big| > \frac{\varepsilon}{4} \;\Big|\; u\Big) \le 2e^{-\varepsilon^2 N/(64F)}.$$

A finite net. Let $\kappa = \varepsilon/(2(L_H(\rho) + L_f(\rho)))$, which does not depend on N, and let $r_1, \dots, r_m \in [-\rho,\rho]$ be points such that every $r \in [-\rho,\rho]$ is within κ of some r_i; we may take $m \le 2\rho/\kappa + 1$. Let N_0 be such that $|\mathbb{E}_{\mathsf{P}_N(\cdot\,|\,u)} f_N(r_i;H,u) - f(r_i)| \le \varepsilon/4$ for $i \le m$ and $N \ge N_0$. On the intersection of the event (14.13) with the m fluctuation events, for $r \in [-\rho,\rho]$ and r_i within κ of r, we have

$$\begin{aligned} |f_N(r) - f(r)| &\le |f_N(r) - f_N(r_i)| + |f_N(r_i) - \mathbb{E} f_N(r_i)| \\ &\quad + |\mathbb{E} f_N(r_i) - f(r_i)| + |f(r_i) - f(r)| \le \varepsilon, \end{aligned}$$

where $\mathbb{E}$ is the expectation under $\mathsf{P}_N(\cdot \mid u)$; for the first and last terms, we used (14.14) and (b). The failure probability of this intersection is at most $e^{-c_0 N} + 2me^{-\varepsilon^2 N/(64F)} \le (2m+1)e^{-2cN}$, which is at most e^{-cN} after enlarging N_0. □

The net $r_1, \dots, r_m$ and its size are fixed before $N \to \infty$, so that no estimate with a width shrinking in N is used. We can now identify the constrained pressure.

Proposition 14.7 (Cross-section formula)**.** *Let ξ be a finite mixture and $0 < \beta < \beta_{\mathrm{s}}$. For every $r \in (-1,1)$, the limits in Definition 14.4 exist, and*

$$F_{\mathrm{FP}}(r) = \beta^2 \xi(r) + \frac{1}{2}\log(1-r^2) + \mathcal{P}_\beta(\eta_r). \tag{14.15}$$

In particular, $F_{\mathrm{FP}} = f$ is Lipschitz on $[-\rho,\rho]$ with constant $L_f(\rho)$ for every $\rho \in (0,1)$. Moreover, for every $r \in (-1,1)$ and $\varepsilon > 0$, there exists $w_0 > 0$ such that, for every $w \in (0, w_0)$, there exist $c > 0$ and N_0 with

$$\mathsf{Q}_N\Big(\Big|\frac{1}{N}\log Z_{N,\beta}\big([r-w, r+w]; H, u\big) - F_{\mathrm{FP}}(r)\Big| > \varepsilon\Big) \le e^{-cN}, \qquad N \ge N_0. \tag{14.16}$$

Proof. Fix $r \in (-1,1)$ and $w > 0$ with $[r-w, r+w] \subset (-1,1)$, and let $\rho < 1$ satisfy $[r-w, r+w] \subset [-\rho,\rho]$. Let $m_w = \max_{[r-w,r+w]} f$ and $X_N = N^{-1}\log Z_{N,\beta}([r-w,r+w];H,u)$. Fix $\varepsilon' > 0$, and let $\ell > 0$ be the length of an interval inside $[r-w,r+w]$ on which $f \ge m_w - \varepsilon'$; such an interval exists because f is continuous. On the event $\Omega_N(\rho,\varepsilon')$ of Lemma 14.6, (14.11) implies that

$$C_N\, \ell\, e^{N(m_w - 2\varepsilon')} \le Z_{N,\beta}\big([r-w,r+w];H,u\big) \le C_N\, 2w\, e^{N(m_w+\varepsilon')}.$$

Then $|X_N - m_w| \le 3\varepsilon'$ on $\Omega_N(\rho,\varepsilon')$ for all large N, because $N^{-1}\log C_N \to 0$. We have $\mathsf{P}_N(|X_N - m_w| > 3\varepsilon') \le e^{-c_1 N}$ for some $c_1 > 0$ by Lemma 14.6(c), and Lemma 14.3(b) yields

$$\mathsf{Q}_N\big(|X_N - m_w| > 3\varepsilon'\big) \le 3e^{-c_2 N} \tag{14.17}$$

for some $c_2 > 0$ and all large N.

To pass to expectations, note that

$$|X_N| \le \frac{\beta}{N} \sup_{S_N} |H| + \Big| \frac{1}{N} \log \nu_N\{x\colon R(x,u) \in [r-w, r+w]\} \Big|.$$

The second term does not depend on u and is bounded in N by (14.11). The law of H under Q_N is $\mathbb{P}$. By the Cauchy–Schwarz inequality in the representation (3.2), $\sup_{S_N} |H| \le |g|\,(N\xi(1))^{1/2}$, and therefore $\mathbb{E}(\sup_{S_N} |H|)^2 \le N\xi(1) \sum_{p=2}^{P} N^p$. We find that $\mathbb{E}_{\mathsf{Q}_N} |X_N - m_w|^2 \le CN^k$ for some C and k independent of N. By the Cauchy–Schwarz inequality and (14.17),

$$\big|\mathbb{E}_{\mathsf{Q}_N} X_N - m_w\big| \le 3\varepsilon' + \big(\mathbb{E}_{\mathsf{Q}_N} |X_N - m_w|^2\big)^{1/2} \big(3e^{-c_2 N}\big)^{1/2},$$

and the last term tends to zero. As ε' is arbitrary, we conclude that $N^{-1}\mathbb{E}_{\mathsf{Q}_N} \log Z_{N,\beta}([r-w, r+w]; H, u) \to m_w$. Letting $w \downarrow 0$ and using the continuity of f, we obtain (14.15). The Lipschitz bound is Lemma 14.6(b).

For (14.16), let $\rho = (1+|r|)/2$ and $w_0 = \min\{(1-|r|)/4, \varepsilon/(2L_f(\rho))\}$; then $[r-w_0, r+w_0] \subset [-\rho, \rho]$. For $w < w_0$, we have $|m_w - F_{\mathrm{FP}}(r)| \le L_f(\rho) w \le \varepsilon/2$, and (14.17) with $\varepsilon' = \varepsilon/6$ gives (14.16) after decreasing c and enlarging N_0. □

Remark 14.8. For the pure p-spin model, $\eta_r(t) = (r^2 + (1-r^2)t)^p - r^{2p}$, Lemma 14.5(c) is [65, Proposition 3.7] and (14.15) is [65, Proposition 3.8]. By Lemma 14.5(b), the coefficient of t in η_r is positive for $r \neq 0$, even when ξ has no degree-one term. This is why we use the Parisi formula for covariance polynomials with a degree-one term. The planted computation in Lemma 14.6 holds for every $\beta > 0$. Only the identification with the Gibbs-reference law uses $\beta < \beta_{\mathrm{s}}$, through Lemma 14.3. For an alternative derivation of the mixed-model formula, including different temperatures for the reference point and the constrained measure, see [59, Corollary 2.1 and Eqs. (2.4)–(2.7)].

14.3 The trial pressure and the equality at q_*

In this section, we assume $\beta_{\mathrm{d}} < \beta < \beta_{\mathrm{s}}$. We define the trial pressure

$$\Psi(r) = F + \frac{1}{2}\Phi_\beta(r) = \frac{1}{2}\Big(\beta^2\xi(1) + \beta^2\xi(r) + r + \log(1-r)\Big), \qquad r \in [0,1). \tag{14.18}$$

We first evaluate the Crisanti–Sommers functional of the auxiliary mixture η_r at a one-atom measure, which gives the trial bound $F_{\mathrm{FP}} \le \Psi$ (Lemma 14.9). We then show, using the Crisanti–Sommers optimality criterion of Lemma 3.15(3), that this measure is the minimizer at $r = q_*$ (Proposition 14.11).

We use the notation of Definition 3.13: $\mathscr{M}$ is the set of admissible measures, $\mathcal{P}_\beta(\zeta;\eta)$ is the Crisanti–Sommers functional, and $\hat{x}_\zeta(t) = \int (1 - \max(t,s))\,\zeta(ds)$. We call the first

term in the brackets of (3.16) the energy term and the sum of the other two the entropy term. We also use the first variation f_ζ of (3.18), its maximizing set S_ζ, and its derivative

$$f'_\zeta(s) = \beta^2\eta'(s) - \int_0^s \frac{dt}{\hat{x}_\zeta(t)^2}, \qquad s \in [0,1). \tag{14.19}$$

Lemma 14.9 (Trial bound)**.** *Let ξ be a finite mixture and $0 < \beta < \beta_s$. For $r \in [0,1)$, let $r_0 = r/(1+r)$. Then*

$$\mathcal{P}_\beta(\delta_{r_0}; \eta_r) = \frac{1}{2}\Big\{\beta^2\big[\xi(1) - \xi(r)\big] + r - \log(1+r)\Big\} \tag{14.20}$$

and

$$F_{\mathrm{FP}}(r) \le \Psi(r), \tag{14.21}$$

with equality if and only if $\mathcal{P}_\beta(\eta_r) = \mathcal{P}_\beta(\delta_{r_0}; \eta_r)$.

For the pure p-spin model, (14.21) is [65, Proposition 3.9], with the same choice $r_0 = r/(1+r)$.

Proof. Note that $1 - r_0 = 1/(1+r)$ and $r^2 + (1-r^2)r_0 = r$. We find from (14.10) that the energy term of $\mathcal{P}_\beta(\delta_{r_0}; \eta_r)$ is

$$\beta^2\big[\eta_r(1) - \eta_r(r_0)\big] = \beta^2\big[\xi(1) - \xi(r)\big].$$

Since $\hat{x}_{\delta_{r_0}}(t) = 1 - r_0$ for $t \le r_0$ and $1 - t$ for $t \ge r_0$, the entropy term with any $\hat{q} \in (r_0, 1)$ is

$$\frac{r_0}{1-r_0} + \log(1-r_0) - \log(1-\hat{q}) + \log(1-\hat{q}) = r - \log(1+r).$$

This proves (14.20). By Proposition 14.7 and the definition of $\mathcal{P}_\beta(\eta_r)$ as an infimum,

$$F_{\mathrm{FP}}(r) \le \beta^2\xi(r) + \frac{1}{2}\log(1-r^2) + \frac{1}{2}\Big\{\beta^2\big[\xi(1) - \xi(r)\big] + r - \log(1+r)\Big\} = \Psi(r),$$

where we used $\log(1-r^2) = \log(1-r) + \log(1+r)$. Equality holds exactly when the infimum defining $\mathcal{P}_\beta(\eta_r)$ is attained at δ_{r_0}. □

The endpoint and the collar. Recall from Definition 3.16 that $A(r) = (1-r)\xi'(r)/r$, and that $q_* = \sup\{r \in (0,1)\colon \beta^2 A(r) > 1\}$. As ξ has no terms of degree below two, $A(r) = (1-r)\sum_p p\, c_p r^{p-2}$ is a polynomial. By (3.20), we have

$$\Psi'(r) = \frac{1}{2}\Phi'_\beta(r) = \frac{r}{2(1-r)}\big[\beta^2 A(r) - 1\big], \qquad r \in (0,1). \tag{14.22}$$

Definition 14.10 (Admissible collars, caps, and core bands)**.** An *admissible collar* is a pair (q_1, q_2) with $0 < q_1 < q_2 < q_*$ such that $\beta^2 A(r) > 1$ for every $r \in (q_1, q_*)$. For $u \in S_N$, $s \in (-1, 1)$, and $w > 0$, we call $\{x \in S_N \colon R(x,u) \geq s\}$ the *cap* of latitude s around u, $\{x \in S_N \colon q_1 \leq R(x,u) \leq q_2\}$ the *collar* around u, and $\{x \in S_N \colon |R(x,u) - q_*| \leq w\}$ the *core band* of half-width w around u. We define the core exponent J and the collar gap G by

$$J = -\frac{1}{2}\Phi_\beta(q_*), \qquad G = \frac{1}{2}\big[\Phi_\beta(q_*) - \Phi_\beta(q_2)\big]. \tag{14.23}$$

The core exponent depends only on ξ and β, while the collar gap also depends on q_2. Note that $\Psi(q_*) = F - J$ and $\Psi(q_2) = F - J - G$.

Proposition 14.11 (Optimality of the replica-symmetric trial at q_*)**.** *Let ξ be a finite mixture and $\beta_{\mathrm{d}} < \beta < \beta_{\mathrm{s}}$.*

(a) We have $q_* \in (0,1)$, $\beta^2 A(q_*) = 1$, $\beta^2 A(r) \leq 1$ for every $r \in [q_*, 1)$, and $A'(q_*) \leq 0$. There exists $\varepsilon_0 > 0$ such that $\beta^2 A(r) > 1$ for every $r \in (q_* - \varepsilon_0, q_*)$. In particular, for every $b \in (0, q_*^2)$, there exists an admissible collar with $q_1 > \sqrt{b}$.

(b) For every $r \in [0,1)$, we have $F_{\mathrm{FP}}(r) \leq \Psi(r)$, and

$$F_{\mathrm{FP}}(q_*) = \Psi(q_*) = F - J. \tag{14.24}$$

(c) We have $J > 0$. For every admissible collar (q_1, q_2), we have $G > 0$ and

$$\sup_{q_1 \leq r < 1} F_{\mathrm{FP}}(r) = F_{\mathrm{FP}}(q_*) = F - J, \qquad \sup_{q_1 \leq r \leq q_2} F_{\mathrm{FP}}(r) \leq \Psi(q_2) = F - J - G. \tag{14.25}$$

The proof of the equality in (b) has four steps. We first record the properties of q_* in (a). Next, we show that $\xi'(v)/(v - q_*^2)$ is nonincreasing on $(q_*^2, q_*]$, using the monotonicity of $v\xi''(v)/\xi'(v)$. This determines the sign of f'_ζ below r_0, for $\zeta = \delta_{r_0}$ and the auxiliary mixture η_{q_*}. Above r_0, the sign of f'_ζ is that of $\beta^2 A - 1$ to the right of q_*. Then f_ζ is maximized at r_0, and Lemma 3.15(3) applies.

Proof. The endpoint q_*. Because $\beta > \beta_{\mathrm{d}}$ and $\beta_{\mathrm{d}}^{-2} = \sup_{(0,1)} A$, the set $\{r \in (0,1) \colon \beta^2 A(r) > 1\}$ is nonempty, and $q_* > 0$. As $A(1) = 0$ and A is continuous, this set is bounded away from 1, which gives $q_* < 1$. There exist points $r \uparrow q_*$ with $\beta^2 A(r) > 1$, while $\beta^2 A(r) \leq 1$ for every $r \in (q_*, 1)$. Then $\beta^2 A(q_*) = 1$ and $\beta^2 A \leq 1$ on $[q_*, 1)$, by continuity. It follows that the right difference quotients $(A(v) - A(q_*))/(v - q_*)$, $v > q_*$, are nonpositive, and $A'(q_*) \leq 0$. The polynomial $\beta^2 A - 1$ equals -1 at $r = 1$. It is therefore not identically zero, and its zeros are isolated. Let $\varepsilon_0 > 0$ be such that it has no zero in $(q_* - \varepsilon_0, q_*)$. On this interval it has a constant sign, which is positive, because it is positive at points arbitrarily close to q_* from the left. For $b \in (0, q_*^2)$, any $q_1 < q_2$ in $(\max\{\sqrt{b}, q_* - \varepsilon_0\}, q_*)$ form an admissible collar. This proves (a).

Monotonicity of $v\xi''(v)/\xi'(v)$. For the rest of the proof, we write $q = q_*$, $d = 1 - q^2$, and $r_0 = q/(1+q)$. For $v \in (0,1]$, let

$$w_p(v) = \frac{p\, c_p v^{p-1}}{\xi'(v)}, \qquad \chi(v) = \frac{v\xi''(v)}{\xi'(v)} = \sum_p (p-1) w_p(v).$$

The weights $w_p(v)$ are nonnegative and sum to one, and $w_p'(v) = v^{-1} w_p(v)[(p-1) - \chi(v)]$. Multiplying by $p-1$ and summing over p, we obtain

$$\chi'(v) = \frac{1}{v}\Big[\sum_p (p-1)^2 w_p(v) - \chi(v)^2\Big] = \frac{1}{v}\operatorname{Var}_{w(v)}(p-1) \geq 0, \tag{14.26}$$

where $\operatorname{Var}_{w(v)}$ is the variance of $p-1$ when p is distributed according to the weights $w_p(v)$. This is the standard identity for the second derivative of a log-partition function, here of $s \mapsto \log \xi'(e^s)$ (see [152, Proposition 3.1(a)]). For $v \in (q^2, q]$, the two factors in $(v-q^2)\xi''(v)/\xi'(v) = \chi(v)(1 - q^2/v)$ are nonnegative and nondecreasing in v, which implies that

$$(v-q^2)\frac{\xi''(v)}{\xi'(v)} \leq q(1-q)\frac{\xi''(q)}{\xi'(q)} \leq 1, \qquad v \in (q^2, q]. \tag{14.27}$$

For the last inequality, we used

$$\frac{A'(q)}{A(q)} = \frac{\xi''(q)}{\xi'(q)} - \frac{1}{q(1-q)}$$

together with $A(q) > 0$ and $A'(q) \leq 0$. Using (14.27), we see that the derivative of $\log(\xi'(v)/(v-q^2))$, which equals $\xi''(v)/\xi'(v) - 1/(v-q^2)$, is nonpositive on $(q^2, q]$. Comparing with the value at $v = q$, we obtain

$$\xi'(v) \geq \frac{v-q^2}{q(1-q)}\,\xi'(q), \qquad v \in (q^2, q]. \tag{14.28}$$

The first variation below r_0. Let $\eta = \eta_q$ and $\zeta = \delta_{r_0}$. Then $\hat{x}_\zeta(t) = 1 - r_0 = (1+q)^{-1}$ for $t \leq r_0$, and $\hat{x}_\zeta(t) = 1 - t$ for $t \geq r_0$. For $t \in [0,1)$, we write $v = q^2 + dt$; then $\eta_q'(t) = d\,\xi'(v)$, and $v = q$ at $t = r_0$. Since $\beta^2 A(q) = 1$, we have

$$\beta^2 d\,\xi'(q) = (1+q)\cdot \beta^2(1-q)\xi'(q) = q(1+q). \tag{14.29}$$

For $t \in (0, r_0]$, we have $f_\zeta'(t) = \beta^2 d\,\xi'(v) - (1+q)^2 t$ by (14.19), and $v \in (q^2, q]$. Using (14.28), (14.29), and $v - q^2 = dt$, we find

$$\beta^2 d\,\xi'(v) \geq \frac{dt}{q(1-q)}\, q(1+q) = (1+q)^2 t.$$

Then $f_\zeta' \geq 0$ on $(0, r_0]$, and $f_\zeta'(0) = \beta^2 d\,\xi'(q^2) > 0$. Note that $f_\zeta'(r_0) = 0$ by (14.29).

The first variation above r_0. For $t \in [r_0, 1)$, we have $v \in [q, 1)$, $1 - v = d(1 - t)$, and

$$\int_0^t \frac{ds}{\hat{x}_\zeta(s)^2} = \frac{r_0}{(1-r_0)^2} + \frac{1}{1-t} - \frac{1}{1-r_0} = q(1+q) + \frac{1}{1-t} - (1+q) = \frac{d}{1-v} - d = \frac{dv}{1-v}.$$

Inserting this into (14.19), we obtain

$$f'_\zeta(t) = \beta^2 d\, \xi'(v) - \frac{dv}{1-v} = \frac{dv}{1-v}\big[\beta^2 A(v) - 1\big] \le 0,$$

where the inequality follows from (a).

Conclusion. The function f_ζ is nondecreasing on $[0, r_0]$ and nonincreasing on $[r_0, 1)$. It therefore attains its maximum at r_0, so that $r_0 \in S_\zeta$ and $\zeta(S_\zeta) = 1$. Using Lemma 3.15(3), we find that δ_{r_0} is the unique minimizer of $\mathcal{P}_\beta(\cdot\,; \eta_q)$, and that $\mathcal{P}_\beta(\eta_q) = \mathcal{P}_\beta(\delta_{r_0}; \eta_q)$. The equality case of Lemma 14.9 then gives $F_{\mathrm{FP}}(q) = \Psi(q)$. Together with (14.21), this proves (b).

The cap and the collar. By Lemma 3.18(2), the condition $\beta < \beta_{\mathrm{s}}$ implies that $\Phi_\beta < 0$ on $(0, 1)$, and therefore $J > 0$. Let (q_1, q_2) be an admissible collar. By (14.22), Ψ is strictly increasing on $[q_1, q_*]$ and nonincreasing on $[q_*, 1)$. In particular, $G = \Psi(q_*) - \Psi(q_2) > 0$, $\Psi \le \Psi(q_*)$ on $[q_1, 1)$, and $\Psi \le \Psi(q_2)$ on $[q_1, q_2]$. Combining these facts with (b), we obtain (14.25). □

Remark 14.12. The proof uses only $A'(q_*) \le 0$, so it covers the marginal case $A'(q_*) = 0$, in which q_* is a multiple root of $\beta^2 A = 1$. The point q_* need not be the largest root of $\beta^2 A = 1$, since $\beta^2 A$ may touch 1 from below to the right of q_*. The proof uses only $\beta^2 A \le 1$ on $[q_*, 1)$. The equality (14.24) is a statement about the global minimizer of the auxiliary Crisanti–Sommers problem at $r = q_*$, not a local stability statement. We never need monotonicity of F_{FP} itself: the trial bound and the equality at q_* suffice for (14.25).

14.4 Caps, core bands, and collars around reference points

In this section, we fix a finite mixture ξ, $\beta_{\mathrm{d}} < \beta < \beta_{\mathrm{s}}$, and an admissible collar (q_1, q_2), and write q_*, J, and G as in Definition 14.10. The next proposition converts Proposition 14.11 into estimates for the Gibbs measure that hold simultaneously for all reference points in a set $\mathcal{G}$ whose Gibbs measure is nearly one. Table 14.1 lists its parameters in the order in which they are chosen.

Proposition 14.13 (Masses of caps, core bands, and collars)**.** *Let ξ be a finite mixture, $\beta_{\mathrm{d}} < \beta < \beta_{\mathrm{s}}$, and let (q_1, q_2) be an admissible collar. Fix δ with $0 < \delta < \min\{J, G\}/8$. There exist $w \in (0, (q_* - q_2)/4]$ with $q_* + w < 1$, $a > 0$, and N_0, depending only on ξ, β, q_1, q_2, and δ, such that the following holds for every $N \ge N_0$. There exists an event Ω_N*

Parameter	Depends on	Role
ξ, β	—	model and temperature, $\beta_{\mathrm{d}} < \beta < \beta_{\mathrm{s}}$
q_*, F, J	ξ, β	endpoint, annealed pressure, core exponent
K, c_0	ξ	field bounds (14.3)
$q_1 < q_2$	ξ, β	admissible collar
G	ξ, β, q_2	collar gap
δ	J, G	pressure tolerance, $0 < \delta < \min\{J, G\}/8$
w	ξ, β, q_2, δ	half-width of the core band
ρ	ξ, β, K	cutoff near overlap one
net $r_1, \dots, r_m$	$\xi, \beta, K, \rho, \delta$	latitudes in Lemma 14.6(c)
c_P, c_Q, a	c_0, F, δ	exponential rates
N_0	all of the above	threshold for N

Table 14.1. The parameters in Proposition 14.13, in the order of choice. All are fixed before $N \to \infty$.

with $\mathbb{P}(\Omega_N) \geq 1 - 5e^{-aN}$ *on which* $|N^{-1} \log Z_{N,\beta} - F| \leq \delta$ *and there exists a Borel set* $\mathcal{G} = \mathcal{G}_N(H) \subset S_N$ *with* $\pi(\mathcal{G}^c) \leq e^{-aN}$ *such that, for every* $u \in \mathcal{G}$,

$$\pi\{x \colon R(x,u) \geq q_1\} \leq e^{-N(J-2\delta)}, \tag{14.30}$$

$$\pi\{x \colon |R(x,u) - q_*| \leq w\} \geq e^{-N(J+2\delta)}, \tag{14.31}$$

$$\pi\{x \colon q_1 \leq R(x,u) \leq q_2\} \leq e^{-N(J+G-2\delta)}. \tag{14.32}$$

We call the points of $\mathcal{G}$ *good reference points*, and $\mathcal{G}$ the *good set*. In words, the proposition states that around every good reference point the cap of latitude q_1 has Gibbs mass at most $e^{-N(J-2\delta)}$, the core band has mass at least $e^{-N(J+2\delta)}$, and the collar is smaller than the core band by the factor $e^{-N(G-4\delta)}$. The core band lies inside the cap and is disjoint from the collar, because $w \leq (q_* - q_2)/4$. The reference points are not restricted to a net: by rotational invariance, the planted estimates hold uniformly in u, and likelihood transfer and Markov's inequality turn them into the quenched statement.

Proof. The proof has four steps. We first choose the half-width w and a cutoff ρ near overlap one. We then prove the three estimates for the restricted partition functions under the planted law, transfer the failure event to the Gibbs-reference law, and finally divide by $Z_{N,\beta}$.

Choice of the parameters. Let $\rho_0 = (1 + q_*)/2$ and

$$w = \min\Big\{ \frac{q_* - q_2}{4}, \frac{1 - q_*}{4}, \frac{\delta}{8L_f(\rho_0)} \Big\},$$

where L_f is the Lipschitz constant of Lemma 14.6(b). Then $[q_* - w, q_* + w] \subset (q_2, \rho_0)$. Since F_{FP} is Lipschitz on $[-\rho_0, \rho_0]$ with constant $L_f(\rho_0)$ by Proposition 14.7, and $F_{\mathrm{FP}}(q_*) = F - J$ by (14.24), we have

$$F_{\mathrm{FP}}(r) \geq F - J - \frac{\delta}{8}, \qquad |r - q_*| \leq w. \tag{14.33}$$

Let $K_1 = K + \beta\xi'(1)$ as in Lemma 14.6(c), and let

$$\rho = \max\Big\{\rho_0,\ \big(1 - e^{-2(\beta K_1 + |F-J|+2)}\big)^{1/2}\Big\} < 1,$$

so that

$$\beta K_1 + \frac{1}{2}\log(1-\rho^2) \leq F - J - 2. \tag{14.34}$$

By Lemma 14.6(c) with this ρ and $\varepsilon = \delta/4$, there exists N_0 such that $\mathsf{P}_N(\Omega_N(\rho, \delta/4)^c \mid u) \leq e^{-c_P N}$ for $N \geq N_0$ and every $u \in S_N$, where

$$c_P = \frac{1}{2}\min\Big\{c_0, \frac{\delta^2}{1024F}\Big\}.$$

Planted estimates. Fix $u \in S_N$, and suppose that $\Omega_N(\rho, \delta/4)$ holds. Recall from Proposition 14.7 that $f = F_{\mathrm{FP}}$ in Lemma 14.6. We bound the cap in two parts. On $[q_1, \rho] \subset [-\rho, \rho]$, we have $f_N \leq F_{\mathrm{FP}} + \delta/4 \leq F - J + \delta/4$ by (14.25). Inserting this into (14.11), we obtain

$$Z_{N,\beta}\big([q_1, \rho]; H, u\big) \leq C_N e^{N(F-J+\delta/4)}.$$

Near overlap one, we use $\sup_{S_N} |H| \leq K_1 N$ and (14.11) with $H = 0$:

$$Z_{N,\beta}\big((\rho, 1]; H, u\big) \leq e^{\beta K_1 N}\, \nu_N\{x\colon R(x,u) > \rho\} \leq e^{\beta K_1 N}\, C_N (1-\rho^2)^{(N-3)/2} \leq e^{N(F-J-1)}$$

for all large N, where in the last step we used (14.34) and $N^{-1}\log C_N \to 0$. Adding the two bounds, we find, for all large N,

$$Z_{N,\beta}\big([q_1, 1]; H, u\big) \leq e^{N(F-J+\delta)}. \tag{14.35}$$

For the core band, we use (14.11), (14.33), and $f_N \geq F_{\mathrm{FP}} - \delta/4$ on $[q_* - w, q_* + w] \subset [-\rho, \rho]$:

$$Z_{N,\beta}\big([q_* - w, q_* + w]; H, u\big) \geq C_N\, 2w\, e^{N(F-J-3\delta/8)} \geq e^{N(F-J-\delta)} \tag{14.36}$$

for all large N, since w does not depend on N. For the collar, (14.25) and $f_N \leq F_{\mathrm{FP}} + \delta/4$ yield

$$Z_{N,\beta}\big([q_1, q_2]; H, u\big) \leq C_N e^{N(F-J-G+\delta/4)} \leq e^{N(F-J-G+\delta)} \tag{14.37}$$

for all large N. Let E_N be the set of pairs (H, u) for which at least one of (14.35)–(14.37) fails. It is a Borel set, because the restricted partition functions in these three bounds

are continuous in (H,u). After enlarging N_0, we have shown that $E_N \subset \Omega_N(\rho,\delta/4)^c$ for $N \ge N_0$. Integrating over $u \sim \nu_N$, we find

$$\mathsf{P}_N(E_N) = \int \mathsf{P}_N(E_N \mid u)\,\nu_N(du) \le e^{-c_P N}.$$

Transfer and disintegration. By Lemma 14.3(b), for all large N,

$$\mathsf{Q}_N(E_N) \le 3e^{-c_Q N}, \qquad c_Q = \min\Big\{\frac{c_P}{2}, \frac{c_P^2}{64F}\Big\}.$$

For each field H, let $\mathcal{G}_N(H) = \{u \in S_N \colon (H,u) \notin E_N\}$. By the definition of Q_N, we have $\mathbb{E}\,\pi^H(\mathcal{G}_N(H)^c) = \mathsf{Q}_N(E_N)$, and Markov's inequality gives

$$\mathbb{P}\big(\pi^H(\mathcal{G}_N(H)^c) > e^{-c_Q N/2}\big) \le 3e^{-c_Q N/2}.$$

Normalization. Let $a = \min\{c_Q/2, \delta^2/(16F)\}$, and let Ω_N be the event that $\pi^H(\mathcal{G}_N(H)^c) \le e^{-c_Q N/2}$ and $|N^{-1}\log Z_{N,\beta} - F| \le \delta$. Using the previous step and Lemma 14.3(a) with $\varepsilon = \delta$, we have $\mathbb{P}(\Omega_N^c) \le 3e^{-aN} + 2e^{-aN}$ for all large N. On Ω_N, we set $\mathcal{G} = \mathcal{G}_N(H)$ and note that $\pi(\mathcal{G}^c) \le e^{-aN}$. For $u \in \mathcal{G}$, we divide (14.35)–(14.37) by $Z_{N,\beta} \in [e^{N(F-\delta)}, e^{N(F+\delta)}]$ and obtain (14.30)–(14.32). This completes the proof. □

14.5 Replica concentration and the Gibbs mass of caps

The last statements of this section concern the overlap of two independent samples from π. The first is the estimate of Huang, Montanari, and Pham recorded in Lemma 3.20, stated here for the temperatures of Part III.

Proposition 14.14 (Replica concentration)**.** *Let ξ be a finite mixture and $0 < \beta < \beta_{\mathrm{s}}$. For every $s \in (0,1)$, there exist $c > 0$ and N_0, depending on ξ, β, and s, such that, for every $N \ge N_0$, with probability at least $1 - e^{-cN}$,*

$$(\pi \otimes \pi)\big\{(x,y) \colon |R(x,y)| > s\big\} \le 2e^{-cN}. \tag{14.38}$$

The rate c depends on s, and the proposition gives no rate that is uniform as $s \downarrow 0$. It applies to any fixed finite set of thresholds, by intersecting the corresponding events.

Proof. By Lemma 3.18(2), the pair (ξ,β) is strictly replica symmetric with $\beta^2\xi''(0) < 1$, that is, $\Phi_\beta < 0$ on $(0,1)$ and $\beta^2\xi''(0) < 1$. These are the hypotheses [93, Eq. (7.9)] for the mixture $\beta^2\xi$ with zero external field; see (3.22). Then (3.24) in Lemma 3.20 is (14.38) with $c = c_s$. □

The next lemma is deterministic. It bounds the mass of a cap by the mass of a set of pairs, which converts replica concentration into an exponential bound on the Gibbs

mass of every cap. The bound holds for every probability measure and every axis. In particular, the axis may depend on the field and on any other data, and no union bound over axes is needed. It is an elementary barycenter estimate, and we make no claim of novelty for it.

Lemma 14.15 (Masses of caps from overlaps of two replicas)**.** *Let π be a Borel probability measure on S_N and $s \in (0,1)$. Then*

$$\sup_{v \in S_N} \pi\{x \colon R(x,v) \geq s\} \leq \Big[\frac{2-s^2}{s^2}\,(\pi\otimes\pi)\Big\{(x,y)\colon R(x,y) > \frac{s^2}{2}\Big\}\Big]^{1/2}. \tag{14.39}$$

Proof. Fix $v \in S_N$, let $B = \{x \colon R(x,v) \geq s\}$ and $\alpha = \pi(B)$. If $\alpha = 0$, there is nothing to prove. Otherwise, let $\mu = \pi(\cdot \mid B)$ and $m = \int U \,\mathrm{d}\mu$. Since $U(x)\cdot U(v) = R(x,v) \geq s$ on B and $|U(v)| = 1$, we have $|m| \geq m \cdot U(v) \geq s$. For independent $x, y \sim \mu$, we find

$$\mathbb{E}_{\mu\otimes\mu} R(x,y) = \mathbb{E}_{\mu\otimes\mu} U(x)\cdot U(y) = |m|^2 \geq s^2.$$

Let $t = s^2/2$ and $p = (\mu\otimes\mu)\{R > t\}$. As $R \leq t$ on the complementary event and $R \leq 1$ everywhere, we have $s^2 \leq t(1-p) + p$. Rearranging, we find

$$p \geq \frac{s^2 - t}{1 - t} = \frac{s^2}{2 - s^2}.$$

Then $(\pi\otimes\pi)\{R > t\} \geq (\pi\otimes\pi)\big(\{R > t\} \cap (B\times B)\big) = \alpha^2 p \geq \alpha^2 s^2/(2-s^2)$, which is (14.39) for this v. □

Combining Proposition 14.14 at the threshold $s^2/2$ with Lemma 14.15, we arrive at the uniform cap bound used in Sections 15 and 16.

Corollary 14.16 (Uniform bound on the Gibbs masses of caps)**.** *Let ξ be a finite mixture and $0 < \beta < \beta_{\mathrm{s}}$. For every $s \in (0,1)$, there exist $c > 0$ and N_0, depending on ξ, β, and s, such that, for every $N \geq N_0$, with probability at least $1 - e^{-cN}$,*

$$\sup_{v \in S_N} \pi\{x \colon R(x,v) \geq s\} \leq \frac{2}{s}\, e^{-cN} \qquad \text{and} \qquad |\pi(U)|^2 \leq s + e^{-cN}. \tag{14.40}$$

Proof. Let $c_1 > 0$ be the smaller of the two rates given by Proposition 14.14 for the thresholds $s^2/2$ and s, and let $c = c_1/2$. For large N, the intersection of the two events has probability at least $1 - 2e^{-c_1 N} \geq 1 - e^{-cN}$. On this intersection, we have by Lemma 14.15 that

$$\sup_{v \in S_N} \pi\{x \colon R(x,v) \geq s\} \leq \Big[\frac{2-s^2}{s^2}\cdot 2e^{-c_1 N}\Big]^{1/2} \leq \frac{2}{s}\, e^{-cN}.$$

For the second bound, recall that $|\pi(U)|^2 = \mathbb{E}_{\pi\otimes\pi} R(x,y)$. As $R \leq 1$, we have $|\pi(U)|^2 \leq s + (\pi\otimes\pi)\{R > s\} \leq s + 2e^{-c_1 N} \leq s + e^{-cN}$ for large N. □

SECTION 15

Counting slow modes and the overlap-weighted Gibbs mean

This section collects the deterministic estimates that turn the pressure estimates of Section 14 into the conclusions of Theorem 1.5. Everything here concerns a fixed dimension N and a fixed smooth field; no randomness enters. In Lemma 15.1, we bound from below the number of eigenvalues of $-L$ below a cutoff, for any family of nonnegative trial functions with small averaged energy and small overlaps. In Lemma 15.2, we apply this bound to translated bumps of a single vector observable B. This observable is aligned with the spin U on most of the Gibbs measure and has small vector energy. When the energy term is small, the number of slow modes is then at least a constant times the inverse of the sum of two masses: the Gibbs mass of the set where the alignment fails, and the largest Gibbs mass of a spherical cap of fixed latitude. We show in Lemmas 15.3 and 15.5 that if the centered spin and the centered observable have a fixed positive inner product and the energy errors are small, then a fixed part of the centered spin lies on slow modes and the stationary connected correlation of the spin stays above a plateau. The overlap-weighted Gibbs mean of Definition 15.6 is the observable to which these lemmas are applied. Its basic properties are recorded in Lemma 15.7. It is the mean of U under a reweighted Gibbs measure, its norm is at most one, and its gradient is controlled by a weighted variance. Finally, we show in Lemma 15.8 that an aligned observable with small vector energy keeps the law of the diffusion started from most Gibbs points far from π, in total variation and in relative entropy, provided that the alignment fails only on a set of small Gibbs mass and that all caps of a fixed latitude have small Gibbs mass.

The lemmas use only the spectral facts in Lemmas 3.5 and 3.6 and the contraction and data processing facts in Lemmas 3.8 and 3.9. In Section 16, they are combined with the estimates of Section 14 to prove Theorem 1.5. Section 17 uses Lemmas 15.5 and 15.7, and Section 18 uses Lemma 15.1 through Remark 16.1.

For pure spherical models in a narrower range of temperatures, exponentially many exponentially small eigenvalues and an exponential-time plateau of the stationary spin correlation were proved by El Alaoui, Montanari, and Sellke [65, Corollaries 2.8 and 2.9]. Their argument counts eigenvalues with smoothed indicators of the clusters of a shattering decomposition, which have disjoint supports, together with the variational characterization of eigenvalues. We construct no clusters. Instead, we replace disjoint supports

by the averaged counting bound of Lemma 15.1, which needs only an upper bound on overlaps. We apply this bound to translates of one observable. More broadly, the relation between metastable sets and small eigenvalues of reversible dynamics is a central theme of the potential-theoretic approach to metastability [37, Theorem 1.3], [39, Theorem 1.2]; see also the monograph [34]. For the reversible diffusions of [39, Theorem 1.2], the dimension is fixed and the asymptotics concern a noise parameter tending to zero. Here, the dimension tends to infinity and the number of slow modes is exponential in N.

15.1 Standing notation

Throughout this section, $N \geq 2$, $\beta > 0$, and a smooth function H_N on S_N are fixed, and π, L, $\mathcal{E}$, and P_t are the Gibbs measure, generator, Dirichlet form, and semigroup of Definition 3.4. All statements are deterministic; in Section 16 they are applied to each realization of a finite mixture.

We use the spectral notation of Lemma 3.5(3): the eigenvalues $0 = \lambda_0 < \lambda_1 \leq \lambda_2 \leq \cdots$ of $-L$, listed with multiplicity, an orthonormal basis $(\phi_j)_{j\geq 0}$ of $L^2(\pi)$ of real smooth eigenfunctions with $\phi_0 = 1$, the form domain $\mathcal{D}(\mathcal{E})$, and the coefficients $\hat{f}_j = \langle f, \phi_j \rangle$ of $f \in L^2(\pi)$. Recall that for all $f \in \mathcal{D}(\mathcal{E})$,

$$\|f\|_2^2 = \sum_{j\geq 0} \hat{f}_j^2, \qquad \mathcal{E}(f,f) = \sum_{j\geq 0} \lambda_j \hat{f}_j^2. \tag{15.1}$$

A vector observable is a map $F = (F_1, \dots, F_n) \colon S_N \to \mathbb{R}^n$ with components in $L^2(\pi)$; usually $n = N$. As in Section 3.3, inner products, norms, and operators such as P_t act componentwise and are summed over components, so that $\langle F, G \rangle = \int F \cdot G \, \mathrm{d}\pi$, and we write $\pi(F) = \int F \, \mathrm{d}\pi \in \mathbb{R}^n$. If every component of F lies in $\mathcal{D}(\mathcal{E})$, then

$$\mathcal{E}_{\mathrm{vec}}(F) = \sum_{i=1}^{n} \mathcal{E}(F_i, F_i) = \int \|\nabla F\|_{\mathrm{HS}}^2 \, \mathrm{d}\pi, \qquad \|\nabla F\|_{\mathrm{HS}}^2 = \sum_{i=1}^{n} |\nabla F_i|^2 ,$$

where $\nabla = \nabla_{S_N}$ is the spherical gradient. Note that $\mathcal{E}_{\mathrm{vec}}(F - \pi(F)) = \mathcal{E}_{\mathrm{vec}}(F)$, since constants have zero energy. For $\Lambda > 0$, we write

$$\Pi_\Lambda = \mathbf{1}_{(0,\Lambda]}(-L)$$

for the spectral projection onto the eigenfunctions with eigenvalues in $(0, \Lambda]$, as in Lemma 3.6(4); it annihilates constants. The number $\#\{j \geq 0 : \lambda_j \leq \Lambda\}$ counts eigenvalues with multiplicity and includes $\lambda_0 = 0$. If it is at least $m + 1$, then $\lambda_m \leq \Lambda$. Because λ_0 is simple, the eigenvalues $\lambda_1, \dots, \lambda_m$ are then positive and at most Λ.

Recall that $U(x) = x/\sqrt{N}$, so that $|U(x)| = 1$ and $U(x) \cdot U(y) = R(x,y)$ for $x, y \in S_N$. For fixed $u \in S_N$, the spherical gradient of the overlap is

$$\nabla_x R(x,u) = \frac{u - R(x,u)x}{N}, \qquad |\nabla_x R(x,u)|^2 = \frac{1 - R(x,u)^2}{N}. \tag{15.2}$$

Indeed, the Euclidean gradient of $x \mapsto x \cdot u/N$ is u/N, its projection onto the tangent space $x^{\perp}$ is $(u - R(x,u)x)/N$, and $|u - Rx|^2 = N - 2R\,(x \cdot u) + R^2 N = N(1 - R^2)$ with $R = R(x,u)$.

We fix once and for all a smooth nondecreasing function $\chi\colon \mathbb{R} \to [0,1]$ with $\chi = 0$ on $(-\infty, 0]$, $\chi = 1$ on $[1,\infty)$, and $\|\chi'\|_\infty \le 2$. For instance, $\chi(t) = \int_0^t \varphi / \int_0^1 \varphi$ works for any smooth $\varphi\colon \mathbb{R} \to [0,1]$ supported in $[0,1]$ and equal to one on $[1/8, 7/8]$, since then $\int_0^1 \varphi \ge 3/4$.

15.2 Counting with nonnegative trial functions

Lemma 15.1 (Counting with nonnegative trial functions)**.** *Let $(\mathcal{V}, \mu)$ be a finite measure space, and let $(u,x) \mapsto f_u(x)$ be a nonnegative measurable function on $\mathcal{V} \times S_N$ such that $f_u \in \mathcal{D}(\mathcal{E})$ for every $u \in \mathcal{V}$. Suppose that there exist $r \in [0,\infty)$ and $\mathsf{R}, \mathsf{C} \in (0,\infty)$ such that*

$$0 < \mathsf{T} = \int \|f_u\|_2^2 \, \mu(\mathrm{d}u) < \infty, \qquad \int \mathcal{E}(f_u, f_u)\, \mu(\mathrm{d}u) \le r\,\mathsf{T}, \tag{15.3}$$

and

$$\pi(f_u) \le \mathsf{R} \quad \text{for } \mu\text{-almost every } u, \qquad \int f_u(x)\, \mu(\mathrm{d}u) \le \mathsf{C} \quad \text{for } \pi\text{-almost every } x. \tag{15.4}$$

Then for every $\Lambda > r$,

$$\#\{j \ge 0 : \lambda_j \le \Lambda\} \ge \Big(1 - \frac{r}{\Lambda}\Big)\frac{\mathsf{T}}{\mathsf{R}\,\mathsf{C}}.$$

The count includes the eigenvalue $\lambda_0 = 0$. The finiteness of T is used at the end of the proof, where $r\mathsf{T}/\Lambda$ is subtracted from T; without it, the energy hypothesis could hold in the form $\infty \le \infty$ while the conclusion fails.

Proof. The spectral step below is the averaged variational principle [66, Theorem 2.1]. The bound on b_j is the integral analog of Schur's row–column estimate [135, Section 2, part I, p. 6]; we give the argument under the stated hypotheses.

For $j \ge 0$, we define

$$b_j = \int \langle f_u, \phi_j \rangle^2 \, \mu(\mathrm{d}u).$$

The map $u \mapsto \langle f_u, \phi_j \rangle$ is measurable by Tonelli's theorem, applied to the positive and negative parts of $f_u \phi_j$. Using $f_u \ge 0$ and the Cauchy–Schwarz inequality for the measure $f_u\,\mathrm{d}\pi$, we have

$$\langle f_u, \phi_j \rangle^2 = \Big(\int f_u \phi_j \,\mathrm{d}\pi \Big)^2 \le \pi(f_u) \int f_u \phi_j^2 \,\mathrm{d}\pi.$$

Integrating in u, using the first bound in (15.4), and then Tonelli's theorem and the second bound, we obtain

$$b_j \le \mathsf{R} \int \phi_j(x)^2 \Big(\int f_u(x)\, \mu(\mathrm{d}u) \Big) \pi(\mathrm{d}x) \le \mathsf{R}\,\mathsf{C}, \tag{15.5}$$

where in the last step we used $\|\phi_j\|_2 = 1$. Next, (15.1), applied to each f_u, and Tonelli's theorem for series give

$$\sum_{j\ge 0} b_j = \int \|f_u\|_2^2 \, \mu(\mathrm{d}u) = \mathsf{T}, \qquad \sum_{j\ge 0} \lambda_j b_j = \int \mathcal{E}(f_u, f_u)\, \mu(\mathrm{d}u) \le r\,\mathsf{T}.$$

Here $u \mapsto \mathcal{E}(f_u, f_u)$ is measurable, as a series of nonnegative measurable functions. Bounding λ_j from below by Λ on the indices with $\lambda_j > \Lambda$, we have

$$\Lambda \sum_{j:\lambda_j>\Lambda} b_j \le \sum_{j\ge 0} \lambda_j b_j \le r\,\mathsf{T}.$$

Subtracting this from $\sum_j b_j = \mathsf{T}$, which is finite, we find

$$\sum_{j:\lambda_j\le\Lambda} b_j \ge \mathsf{T} - \frac{r\,\mathsf{T}}{\Lambda} = \Big(1 - \frac{r}{\Lambda}\Big)\mathsf{T}.$$

The left side is at most $\mathsf{R}\,\mathsf{C}\cdot \#\{j \ge 0 : \lambda_j \le \Lambda\}$ by (15.5), which proves the lemma. □

Lemma 15.1 is a specialization of the averaged variational principle of Harrell and Stubbe [86, Theorem 3.1]; see [66, Theorem 2.1] for its form for elliptic operators on manifolds, which covers $-L$. In our notation, that principle, applied to the whole family (f_u) with cutoff Λ, implies that $(\Lambda - r)\mathsf{T} \le \sum_j (\Lambda - \lambda_j)_+ b_j$, and the counting bound follows from (15.5). Positivity of the trial functions is used only in (15.5), which is an upper frame bound of Schur type. Averaged families of trial functions were used earlier by Kröger [104] to bound Neumann eigenvalues. We have included the short proof because the hypotheses of Lemma 15.1 are exactly the ones verified below.

15.3 Many slow modes from one aligned observable

Lemma 15.2 (One aligned observable gives many slow modes)**.** *Let $B\colon S_N \to \mathbb{R}^N$ be smooth with $|B| \le 1$, and suppose that $U(x)\cdot B(x) \ge q > 0$ for every x in a measurable set $\mathcal{G} \subset S_N$. Fix $0 < \rho < q/2$ and set*

$$M_\rho = \pi(\mathcal{G}^c) + \sup_{v\in S_N} \pi\{y \in S_N : R(y, v) \ge q - 2\rho\}.$$

Let $\alpha\colon \mathbb{R} \to [0,1]$ be smooth, equal to zero on $(-\infty, 0]$ and to one on $[1,\infty)$. Then for every $\Lambda > 0$,

$$\#\{j \ge 0 : \lambda_j \le \Lambda\} \ge \frac{\frac{1}{2} - 4\,\|\alpha'\|_\infty^2\, N^2 \rho^{-2} \mathcal{E}_{\mathrm{vec}}(B)/\Lambda}{M_\rho}. \tag{15.6}$$

The count includes the constant eigenfunction. Note that $M_\rho > 0$. Indeed, if $\mathcal{G}$ is empty, then $\pi(\mathcal{G}^c) = 1$. Otherwise, the alignment and $|B| \le 1$ imply that $q \le 1$, so $q - 2\rho < 1$, and every cap of latitude $q - 2\rho$ has positive Gibbs mass, because π has a positive density.

Proof. We use Lemma 15.1 with the bumps $x \mapsto \alpha(N[1 - |B(x) - z|^2/\rho^2])$, indexed by all centers z in a ball of $\mathbb{R}^N$. A bump is supported where B lies within ρ of z. On $\mathcal{G}$, its support lies in a spherical cap, and averaging over z keeps the overlaps of the bumps under control.

For $s > 0$ and $z \in \mathbb{R}^N$, let $\mathbb{B}_s(z) \subset \mathbb{R}^N$ be the closed Euclidean ball of radius s centered at z, write $\mathbb{B}_s = \mathbb{B}_s(0)$, and let $|\cdot|$ also denote Lebesgue measure on $\mathbb{R}^N$. We use the finite index measure and the trial functions

$$\mu(\mathrm{d}z) = \frac{\mathbf{1}_{\mathbb{B}_{1+\rho}}(z)\,\mathrm{d}z}{|\mathbb{B}_\rho|}, \qquad f_z(x) = \alpha\big(N[1 - |B(x) - z|^2/\rho^2]\big), \qquad z \in \mathbb{R}^N,\ x \in S_N.$$

The function $(z, x) \mapsto f_z(x)$ is smooth with values in $[0, 1]$; in particular, it is measurable and each f_z lies in $\mathcal{D}(\mathcal{E})$. As α vanishes on $(-\infty, 0]$ and equals one on $[1, \infty)$, we have

$$f_z(x) > 0 \ \Rightarrow\ |B(x) - z| < \rho, \qquad |B(x) - z| \le \rho\,(1 - 1/N)^{1/2} \ \Rightarrow\ f_z(x) = 1. \tag{15.7}$$

Column bound and total mass. Fix $x \in S_N$. The ball $\mathbb{B}_\rho(B(x))$ is contained in $\mathbb{B}_{1+\rho}$, because $|B(x)| \le 1$, and on $\mathbb{B}_{1+\rho}$ the measure μ is Lebesgue measure divided by $|\mathbb{B}_\rho|$. By the first implication in (15.7) and $f_z \le 1$, we have

$$\int f_z(x)\,\mu(\mathrm{d}z) \le \mu\big(\mathbb{B}_\rho(B(x))\big) = 1.$$

Similarly, the second implication gives

$$\int f_z(x)^2\,\mu(\mathrm{d}z) \ge \mu\big(\mathbb{B}_{\rho(1-1/N)^{1/2}}(B(x))\big) = (1 - 1/N)^{N/2} \ge \frac{1}{2}.$$

For the last inequality, note that the function $F(t) = (t/2)\log(1 - 1/t)$ satisfies $F(2) = -\log 2$ and

$$F'(t) = \frac{1}{2}\Big[\frac{1}{t-1} - \log\Big(1 + \frac{1}{t-1}\Big)\Big] \ge 0 \qquad (t > 1),$$

since $\log(1 + y) \le y$. Integrating over x and using $f_z^2 \le f_z$, we obtain

$$\frac{1}{2} \le \mathsf{T} = \int \|f_z\|_2^2\,\mu(\mathrm{d}z) \le 1, \qquad \int f_z(x)\,\mu(\mathrm{d}z) \le 1 \quad (x \in S_N). \tag{15.8}$$

Energy. By the chain rule,

$$\nabla f_z(x) = -\frac{2N}{\rho^2}\,\alpha'\big(N[1-|B(x)-z|^2/\rho^2]\big)\sum_{i=1}^{N}(B_i(x)-z_i)\nabla B_i(x).$$

The right side vanishes unless $|B(x)-z|\le\rho$, because α' vanishes outside $[0,1]$. Using the Cauchy–Schwarz inequality in i, we have

$$\begin{aligned}|\nabla f_z(x)|^2 &\le \frac{4N^2}{\rho^4}\,\|\alpha'\|_\infty^2\,|B(x)-z|^2\,\|\nabla B(x)\|_{\mathrm{HS}}^2\,\mathbf{1}\{|B(x)-z|\le\rho\}\\ &\le \frac{4\,\|\alpha'\|_\infty^2\,N^2}{\rho^2}\,\|\nabla B(x)\|_{\mathrm{HS}}^2\,\mathbf{1}\{z\in\mathbb{B}_\rho(B(x))\}.\end{aligned}$$

Integrating first over z, using $\mu(\mathbb{B}_\rho(B(x)))=1$, and then over x yields

$$\int \mathcal{E}(f_z,f_z)\,\mu(\mathrm{d}z) \le 4\,\|\alpha'\|_\infty^2\,N^2\rho^{-2}\mathcal{E}_{\mathrm{vec}}(B) = E_\alpha, \tag{15.9}$$

where the last equality defines E_α.

Row bound. Fix z and let $S_z=\{x\in S_N : |B(x)-z|<\rho\}$. By (15.7) and $f_z\le 1$, we have $\pi(f_z)\le\pi(\mathcal{G}^c)+\pi(\mathcal{G}\cap S_z)$. If $\mathcal{G}\cap S_z$ is empty, then $\pi(f_z)\le M_\rho$. Otherwise, choose a point $x_z\in\mathcal{G}\cap S_z$. For $y\in\mathcal{G}\cap S_z$, we have $|B(y)-B(x_z)|<2\rho$ and $|U(y)|=1$, which give

$$U(y)\cdot B(x_z)\ge U(y)\cdot B(y)-2\rho\ge q-2\rho>0.$$

Further, $|B(x_z)|\ge U(x_z)\cdot B(x_z)\ge q>0$. With $v_z=\sqrt{N}\,B(x_z)/\,|B(x_z)|\in S_N$, we have

$$R(y,v_z)=\frac{U(y)\cdot B(x_z)}{|B(x_z)|}\ge q-2\rho,$$

where we used $0<|B(x_z)|\le 1$ and $q-2\rho>0$. We conclude that $\mathcal{G}\cap S_z$ lies in the cap $\{R(\cdot,v_z)\ge q-2\rho\}$, and

$$\pi(f_z)\le M_\rho \qquad (z\in\mathbb{R}^N). \tag{15.10}$$

The point x_z is used only to prove (15.10) for each fixed z, so no measurable choice of x_z is needed.

Conclusion. If $\frac12 - E_\alpha/\Lambda \le 0$, there is nothing to prove. Otherwise, let $r=\mathsf{T}^{-1}\int\mathcal{E}(f_z,f_z)\,\mu(\mathrm{d}z)$. Combining (15.8) and (15.9), we obtain

$$\Big(1-\frac{r}{\Lambda}\Big)\mathsf{T}=\mathsf{T}-\frac{1}{\Lambda}\int\mathcal{E}(f_z,f_z)\,\mu(\mathrm{d}z)\ge\frac12-\frac{E_\alpha}{\Lambda}>0,$$

so $r<\Lambda$, and Lemma 15.1, with $\mathsf{R}=M_\rho$ and $\mathsf{C}=1$ from (15.8) and (15.10), implies that

$$\#\{j\ge 0:\lambda_j\le\Lambda\}\ge\Big(1-\frac{r}{\Lambda}\Big)\frac{\mathsf{T}}{M_\rho}\ge\frac{\frac12-E_\alpha/\Lambda}{M_\rho},$$

which is (15.6). □

Averaging translates of a fixed trial function over a parameter space is a standard way of applying averaged variational principles; compare [66, Section 4]. Compared with counting smoothed indicators of disjoint clusters, as in [65, proof of Corollary 2.8], Lemma 15.2 needs no decomposition of the Gibbs measure: the supports of the bumps overlap, and the overlaps are controlled by the column bound $\mathsf{C} = 1$ and by the Gibbs masses of caps.

15.4 Correlation and spectral projection estimates

The next two lemmas turn an aligned observable with small energy into a lower bound on the projection of the centered spin onto slow modes. The first works directly with the semigroup, and the second with spectral projections.

Lemma 15.3 (Direct correlation estimate)**.** *Let $Z, \overline{B} \in L^2(\pi; \mathbb{R}^N)$ be centered vector observables with $\|Z\|_2 \le 1$ and $\|\overline{B}\|_2 \le 1$, and suppose that every component of $\overline{B}$ lies in $\mathcal{D}(\mathcal{E})$. Set $a = \langle Z, \overline{B}\rangle$ and $E = \mathcal{E}_{\mathrm{vec}}(\overline{B})$. Then for every $t \ge 0$,*

$$\langle Z, P_t Z\rangle \ge \bigl[a - (tE/2)^{1/2}\bigr]_+^2.$$

Proof. Because the operators P_t are self-adjoint and $P_t = P_{t/2}P_{t/2}$, we have $\langle Z, P_t Z\rangle = \|P_{t/2}Z\|_2^2$ and

$$\langle P_{t/2}Z, \overline{B}\rangle = \langle Z, P_{t/2}\overline{B}\rangle = a - \langle Z, (1 - P_{t/2})\overline{B}\rangle.$$

By Lemma 3.6(2), applied to each component of $\overline{B}$, we have $\|(1 - P_{t/2})\overline{B}\|_2^2 \le tE/2$; this rests on the inequality $(1 - e^{-x})^2 \le x$ for $x \ge 0$. Using the Cauchy–Schwarz inequality and $\|Z\|_2, \|\overline{B}\|_2 \le 1$, we obtain

$$\|P_{t/2}Z\|_2 \ge \langle P_{t/2}Z, \overline{B}\rangle \ge a - (tE/2)^{1/2}.$$

The left side is nonnegative, and squaring the positive part of the right side proves the lemma. □

Remark 15.4 (Choice of the exponents). Suppose that for each N, the observables Z_N and $\overline{B}_N$ satisfy the hypotheses of Lemma 15.3 with $a_N \ge a_0 > 0$ and $E_N \le Ce^{-c_E N}$, where a_0, C, and c_E do not depend on N. Then for each fixed $b < a_0^2$ and $\tau < c_E$, we have $\inf_{0 \le t \le e^{\tau N}} \langle Z_N, P_t Z_N\rangle \ge b$ for all sufficiently large N, since $tE_N \le Ce^{-(c_E - \tau)N}$ on this interval. A fixed margin is needed here: a strict inequality $b < a_N^2$ that depends on N would not suffice. The same plateau also follows from Lemma 15.5 below. Let $\tau < \sigma < c_E$ and $\Lambda = e^{-\sigma N}$. We have $E_N/\Lambda \to 0$, and $\Lambda t \le e^{-(\sigma - \tau)N}$ for $t \le e^{\tau N}$. The spectral cutoff and the time exponent can be chosen independently; there is no need to impose $\sigma = 2\tau$.

Lemma 15.5 (Spectral projection estimate)**.** *Let $Z \in L^2(\pi; \mathbb{R}^N)$ with $\pi(Z) = 0$, let $B\colon S_N \to \mathbb{R}^N$ have components in $\mathcal{D}(\mathcal{E})$, and set $\overline{B} = B - \pi(B)$. Then for every $\Lambda > 0$,*

$$\left\|\overline{B} - \Pi_\Lambda \overline{B}\right\|_2^2 \le \Lambda^{-1} \mathcal{E}_{\mathrm{vec}}(B), \tag{15.11}$$

$$\|\Pi_\Lambda Z\|_2 \left\|\overline{B}\right\|_2 \ge \langle Z, \overline{B}\rangle - \|Z\|_2 \left(\mathcal{E}_{\mathrm{vec}}(B)/\Lambda\right)^{1/2}, \tag{15.12}$$

and for every $t \ge 0$,

$$\langle Z, P_t Z\rangle \ge e^{-\Lambda t} \|\Pi_\Lambda Z\|_2^2. \tag{15.13}$$

Proof. Each component of $\overline{B}$ has zero coefficient on $\phi_0 = 1$, because $\overline{B}$ is centered. As $\lambda_0 = 0$ is simple, we have $\overline{B} - \Pi_\Lambda \overline{B} = \mathbf{1}_{(\Lambda,\infty)}(-L)\overline{B}$, and by (15.1),

$$\left\|\overline{B} - \Pi_\Lambda \overline{B}\right\|_2^2 = \sum_{i=1}^{N} \sum_{j:\lambda_j > \Lambda} \langle \overline{B}_i, \phi_j\rangle^2 \le \frac{1}{\Lambda} \sum_{i=1}^{N} \sum_{j \ge 0} \lambda_j \langle \overline{B}_i, \phi_j\rangle^2 = \frac{\mathcal{E}_{\mathrm{vec}}(B)}{\Lambda},$$

which is (15.11); this is also the last assertion of Lemma 3.6(4), applied componentwise. Using that Π_Λ is an orthogonal projection, we write

$$\langle Z, \overline{B}\rangle = \langle \Pi_\Lambda Z, \overline{B}\rangle + \langle Z, \overline{B} - \Pi_\Lambda \overline{B}\rangle,$$

and (15.12) follows from the Cauchy–Schwarz inequality and (15.11). Finally, (15.13) is Lemma 3.6(4) applied to each component of the centered observable Z: the components have zero coefficient on ϕ_0, all terms of $\langle Z, P_t Z\rangle = \sum_i \sum_{j \ge 1} e^{-t\lambda_j} \langle Z_i, \phi_j\rangle^2$ are nonnegative, and the terms with $0 < \lambda_j \le \Lambda$ are at least $e^{-\Lambda t}\langle Z_i, \phi_j\rangle^2$. □

In particular, if $\|Z\|_2 \le 1$ and $\left\|\overline{B}\right\|_2 \le 1$, then (15.12) implies that

$$\|\Pi_\Lambda Z\|_2 \ge \langle Z, \overline{B}\rangle - \left(\mathcal{E}_{\mathrm{vec}}(B)/\Lambda\right)^{1/2}.$$

These estimates are standard spectral calculus; see [17, Section 1.7].

15.5 The overlap-weighted Gibbs mean

Definition 15.6 (Overlap-weighted Gibbs mean)**.** Let $K > 0$, and let $h\colon [-1, 1] \to [0, K]$ be smooth and nonincreasing. For $x, u \in S_N$, we set $\psi_u(x) = e^{-Nh(R(x,u))}$, and we define

$$d_N(x) = \int \psi_u(x)^2 \, \pi(\mathrm{d}u), \qquad B_N(x) = \frac{1}{d_N(x)} \int U(u)\, \psi_u(x)^2 \, \pi(\mathrm{d}u).$$

We call B_N the overlap-weighted Gibbs mean associated with h. For $x \in S_N$, we also let w_x be the probability measure $w_x(\mathrm{d}u) = \psi_u(x)^2 \pi(\mathrm{d}u)/d_N(x)$ on S_N, so that $B_N(x) = \int U \, \mathrm{d}w_x$.

In words, $B_N(x)$ is the mean of $U(u)$ when u is drawn from the Gibbs measure reweighted by $\psi_u(x)^2$; we call w_x the overlap-reweighted Gibbs measure seen from x. The weight penalizes references u at small overlap with x, as h is nonincreasing. All references $u \in S_N$ enter the definition, including those at which the estimates of Section 14 fail; this matters in Section 16, where those estimates are available only for x in a set of large Gibbs mass. For vectors $a \in \mathbb{R}^N$ and tangent vectors e, we write $a \otimes e$ for the matrix with rows $a_i e$, so that $\|a \otimes e\|_{\mathrm{HS}} = |a|\,|e|$.

Lemma 15.7 (Properties of the overlap-weighted Gibbs mean)**.** *In the setting of Definition 15.6, the following hold.*

1. *The function B_N is smooth, $|B_N(x)| \le 1$, and $d_N(x) \ge e^{-2KN}$ for all $x \in S_N$.*
2. *For all $x \in S_N$, we have $U(x) \cdot B_N(x) = \int R(x,u)\, w_x(\mathrm{d}u)$ and $|B_N(x)|^2 = \iint R(u,u')\, w_x(\mathrm{d}u)\, w_x(\mathrm{d}u')$.*
3. *The gradient of B_N is the w_x-mean*

$$\nabla B_N(x) = 2 \int \big(U(u) - B_N(x)\big) \otimes \nabla \log \psi_u(x)\, w_x(\mathrm{d}u), \tag{15.14}$$

and for every $x \in S_N$,

$$\|\nabla B_N(x)\|_{\mathrm{HS}}^2 \le \frac{4}{d_N(x)} \int |\nabla \psi_u(x)|^2 \, \pi(\mathrm{d}u) \le 4N \, \|h'\|_\infty^2 . \tag{15.15}$$

4. *Suppose that $-1 < q_1 < q_2 \le 1$, that $h = K$ on $[-1, q_1]$, and that $h = 0$ on $[q_2, 1]$. Then for every $x \in S_N$,*

$$d_N(x) \ge \pi\{u : R(x,u) \ge q_2\}, \tag{15.16}$$

$$\|\nabla B_N(x)\|_{\mathrm{HS}}^2 \le 4N \, \|h'\|_\infty^2 \, \frac{\pi\{u : q_1 \le R(x,u) \le q_2\}}{d_N(x)}, \tag{15.17}$$

$$U(x) \cdot B_N(x) \ge q_1 - (1 + q_1)\, \frac{e^{-2KN}}{d_N(x)}. \tag{15.18}$$

Proof. We may differentiate under the integral sign, since the integrands in the definition of d_N and B_N are smooth functions of (x,u) on the compact manifold $S_N \times S_N$. We have $d_N \ge e^{-2KN} > 0$, as $\psi_u \ge e^{-KN}$ and π is a probability measure; in particular, B_N is smooth. Further, $|B_N(x)| \le \int |U| \, \mathrm{d}w_x = 1$. This proves (1). Part (2) follows from $U(x) \cdot U(u) = R(x,u)$ and from $|B_N(x)|^2 = \iint U(u) \cdot U(u')\, w_x(\mathrm{d}u)\, w_x(\mathrm{d}u')$.

For (3), write $B_N = n/d_N$ with $n(x) = \int U(u) \psi_u(x)^2 \, \pi(\mathrm{d}u)$. Differentiating, we have $\nabla n = 2 \int U(u) \otimes \psi_u \nabla \psi_u \, \pi(\mathrm{d}u)$ and $\nabla d_N = 2 \int \psi_u \nabla \psi_u \, \pi(\mathrm{d}u)$, and therefore

$$\nabla B_N = \frac{\nabla n}{d_N} - B_N \otimes \frac{\nabla d_N}{d_N} = \frac{2}{d_N} \int \big(U(u) - B_N\big) \otimes \psi_u \nabla \psi_u \, \pi(\mathrm{d}u).$$

This is (15.14), because $\psi_u \nabla \psi_u = \psi_u^2 \nabla \log \psi_u$. Using the triangle inequality for the integral and then the Cauchy–Schwarz inequality for w_x, we obtain

$$\begin{aligned}\|\nabla B_N\|_{\mathrm{HS}} &\le 2 \int |U(u) - B_N| \, |\nabla \log \psi_u| \, \mathrm{d}w_x \\ &\le 2 \Big(\int |U - B_N|^2 \, \mathrm{d}w_x \Big)^{1/2} \Big(\int |\nabla \log \psi_u|^2 \, \mathrm{d}w_x \Big)^{1/2}.\end{aligned}$$

Recall that $B_N(x)$ is the w_x-mean of U and that $|U| = 1$, which implies that $\int |U - B_N|^2 \, \mathrm{d}w_x = 1 - |B_N|^2 \le 1$. Also, $\int |\nabla \log \psi_u|^2 \, \mathrm{d}w_x = d_N^{-1} \int |\nabla \psi_u|^2 \, \pi(\mathrm{d}u)$. Inserting these two facts into the last display, we obtain the first inequality in (15.15). For the second, we note that $\nabla \psi_u(x) = -N h'(R(x,u)) \psi_u(x) \nabla_x R(x,u)$, and by (15.2),

$$|\nabla \psi_u(x)|^2 = N h'(R(x,u))^2 \big(1 - R(x,u)^2\big) \psi_u(x)^2 \le N \, \|h'\|_\infty^2 \, \psi_u(x)^2. \tag{15.19}$$

Integrating over u gives $\int |\nabla \psi_u(x)|^2 \, \pi(\mathrm{d}u) \le N \, \|h'\|_\infty^2 \, d_N(x)$.

For (4), the bound (15.16) holds because $\psi_u(x) = 1$ when $R(x,u) \ge q_2$. Next, h' vanishes outside $[q_1, q_2]$ and $\psi_u \le 1$, and (15.19) implies that $\int |\nabla \psi_u(x)|^2 \, \pi(\mathrm{d}u) \le N \, \|h'\|_\infty^2 \, \pi\{u : q_1 \le R(x,u) \le q_2\}$. Inserting this into (15.15) yields (15.17). Finally, $\psi_u(x)^2 = e^{-2KN}$ when $R(x,u) \le q_1$, which implies that $p = w_x\{u : R(x,u) \le q_1\} \le e^{-2KN}/d_N(x)$. On the complement of this set we have $R(x,u) > q_1$, and $R \ge -1$ everywhere. By part (2), we have $U(x) \cdot B_N(x) \ge q_1(1-p) - p = q_1 - (1+q_1)p$, which proves (15.18). □

The bound $4N \, \|h'\|_\infty^2$ in (15.15) holds at every point, even where d_N is exponentially small. In Section 16, this is what lets us treat the points where the pressure estimates fail through their Gibbs mass alone.

15.6 Slow equilibration from Gibbs-typical points

Lemma 15.8 (Slow equilibration from Gibbs-typical points)**.** *Let $B \colon S_N \to \mathbb{R}^N$ be smooth, and let $T > 0$, $\rho > 0$, and $\eta \in (0,1)$. Let $(X_t)_{t \ge 0}$ be the stationary diffusion, with $X_0 \sim \pi$.*

1. *We have $\mathbb{E}_\pi \, |B(X_T) - B(X_0)|^2 = 2\langle B, (1 - P_T) B \rangle \le 2T \mathcal{E}_{\mathrm{vec}}(B)$.*
2. *The set $\mathcal{Y} = \{x \in S_N : P_T(x, \{y : |B(y) - B(x)| > \rho\}) \le \eta\}$ is measurable, and*

$$\pi(\mathcal{Y}^c) \le \frac{2T \mathcal{E}_{\mathrm{vec}}(B)}{\rho^2 \eta}.$$

3. *Suppose in addition that $|B| \le 1$ on S_N and that $U \cdot B \ge q$ on a measurable set $\mathcal{G}$, where $q > \rho$. Set*

$$w = \pi(\mathcal{G}^c) + \sup_{v \in S_N} \pi\{y \in S_N : R(y,v) \ge q - \rho\}.$$

For $x \in \mathcal{Y} \cap \mathcal{G}$, let $D_x = \{y \in S_N : |B(y) - B(x)| \le \rho\}$. Then $P_T(x, D_x) \ge 1 - \eta$ and $\pi(D_x) \le w$, and

$$\inf_{0 \le t \le T} \|P_t(x, \cdot) - \pi\|_{\mathrm{TV}} \ge 1 - \eta - w, \tag{15.20}$$

$$\inf_{0 < t \le T} D\big(P_t(x, \cdot) \,\big\|\, \pi\big) \ge (1 - \eta) \log(1/w) - \log 2. \tag{15.21}$$

Proof. Part (1) is Lemma 3.6(3), applied to each component of B and summed. In words, both endpoints have law π, so the left side equals $2\,\|B\|_2^2 - 2\langle B, P_T B\rangle$, and in the spectral expansion of $\langle B, (1 - P_T)B\rangle$ we use $1 - e^{-T\lambda_j} \le T\lambda_j$.

For (2), let $F(x) = P_T(x, \{y : |B(y) - B(x)| > \rho\})$. By Lemma 3.5, the kernel $P_T(x, \mathrm{d}y)$ has a density with respect to π that is continuous in (x, y), and F is measurable by Tonelli's theorem. By Markov's inequality and (1), we have

$$\int F \,\mathrm{d}\pi = \mathbb{P}_\pi\big(\,|B(X_T) - B(X_0)| > \rho\big) \le \frac{2T\mathcal{E}_{\mathrm{vec}}(B)}{\rho^2}.$$

Applying Markov's inequality once more, we find $\pi\{F > \eta\} \le 2T\mathcal{E}_{\mathrm{vec}}(B)/(\rho^2\eta)$.

For (3), fix $x \in \mathcal{Y} \cap \mathcal{G}$, so that $P_T(x, D_x) = 1 - F(x) \ge 1 - \eta$. For $y \in D_x \cap \mathcal{G}$, we have

$$U(y) \cdot B(x) \ge U(y) \cdot B(y) - |B(y) - B(x)| \ge q - \rho > 0.$$

Further, $|B(x)| \ge U(x) \cdot B(x) \ge q > 0$. With $v_x = \sqrt{N}\, B(x)/\,|B(x)| \in S_N$, we find $R(y, v_x) = U(y) \cdot B(x)/\,|B(x)| \ge q - \rho$, since $0 < |B(x)| \le 1$. We conclude that $D_x \subset \mathcal{G}^c \cup \{R(\cdot, v_x) \ge q - \rho\}$ and $\pi(D_x) \le w$.

Testing the event D_x, we have $\|P_T(x, \cdot) - \pi\|_{\mathrm{TV}} \ge P_T(x, D_x) - \pi(D_x) \ge 1 - \eta - w$. For $0 \le t \le T$, we have $P_T(x, \cdot) = P_t(x, \cdot)P_{T-t}$ and $\pi P_{T-t} = \pi$, and the contraction of total variation in Lemma 3.8 implies that $\|P_T(x, \cdot) - \pi\|_{\mathrm{TV}} \le \|P_t(x, \cdot) - \pi\|_{\mathrm{TV}}$. This proves (15.20); it is also the case $\nu = \delta_x$ of Proposition 3.10.

For (15.21), we may assume $w < 1$. The set D_x contains a neighborhood of x, because B is continuous, and $\pi(D_x) > 0$ because π has a positive density. By binary data processing (Lemma 3.9), applied to $P_T(x, \cdot)$, π, and the event D_x, we have

$$D\big(P_T(x, \cdot) \,\big\|\, \pi\big) \ge P_T(x, D_x) \log \frac{1}{\pi(D_x)} - \log 2 \ge (1 - \eta) \log \frac{1}{w} - \log 2.$$

For $0 < t \le T$, we have $D(P_T(x, \cdot)\|\pi) \le D(P_t(x, \cdot)\|\pi)$, by the data processing inequality for relative entropy under the kernel P_{T-t} and $\pi P_{T-t} = \pi$ (Lemma 3.8). This proves (15.21). □

Lemma 15.8 bounds the law of X_t at each time $t \le T$. It does not bound the probability that a trajectory started at x leaves D_x and later returns, and the proof needs no estimate of that kind.

SECTION 16

Proof of the main theorem below the static threshold

In this section, we prove Theorem 1.5. Recall the setting: ξ is a finite mixture with $\beta_{\mathrm{d}} < \beta_{\mathrm{s}}$, the temperature satisfies $\beta_{\mathrm{d}} < \beta < \beta_{\mathrm{s}}$, and $q_* = \sup\{r \in (0,1) : \beta^2 A(r) > 1\}$. Fix $0 < b < q_*^2$. The theorem asserts that there exist $\kappa, \gamma, a, c, C > 0$ such that for all sufficiently large N, the following five conclusions hold with probability at least $1 - Ce^{-aN}$.

(i) $\lambda_{\lfloor e^{\kappa N} \rfloor} \le e^{-2\gamma N}$.

(ii) $\|\Pi_N Z_N\|_2^2 \ge b$, where $Z_N = U - \pi(U)$ is the centered spin and $\Pi_N = \mathbf{1}_{(0, e^{-2\gamma N}]}(-L)$ acts componentwise.

(iii) $\mathbb{E}_\pi R(X_0, X_t) - |\pi(U)|^2 \ge b$ for the stationary diffusion and all $0 \le t \le e^{\gamma N}$.

(iv) There exists a set $\mathcal{X}_N$ with $\pi(\mathcal{X}_N) \ge 1 - Ce^{-aN}$ such that $\|P_t(x, \cdot) - \pi\|_{\mathrm{TV}} \ge 1 - Ce^{-aN}$ for all $x \in \mathcal{X}_N$ and $0 \le t \le e^{\gamma N}$.

(v) $I(X_0; X_t) \ge cN$ for the stationary diffusion in the fixed field and all $0 < t \le e^{\gamma N}$.

The proof combines the deterministic lemmas of Section 15 with three inputs from Section 14: the core lower bound and the collar upper bound of Proposition 14.13, and the uniform bound on the Gibbs mass of caps in Corollary 14.16, which combines replica concentration (Proposition 14.14) with Lemma 14.15. The cap upper bound of Proposition 14.13 is not used in the proof; it enters only in Remark 16.1. All five conclusions come from one observable, the overlap-weighted Gibbs mean of Definition 15.6, built from a cutoff that suppresses references at overlap below q_1 with the argument. Its vector energy is exponentially small because, seen from a Gibbs-typical point, the collar carries exponentially less Gibbs mass than the core. It is aligned with the spin because, after the reweighting, references at overlap below q_1 carry only an exponentially small part of the mass. The starting set $\mathcal{X}_N$ in (iv) is measured by π, and (iv) gives no information about starting points that are chosen independently of the disorder.

In Section 17, conclusions (ii) and (iii) are improved to every $b < q_*$, with probability tending to one, while (i), (iv), and (v) are inherited from the present proof. Section 18 uses Remark 16.1.

16.1 Parameters

The proof involves a number of constants, all fixed before $N \to \infty$. Table 16.1 lists them in the order in which they are chosen. Each depends only on ξ, β, b, and the entries above it, and none depends on N or on the realization of the field.

Table 16.1. Parameters in the proof of Theorem 1.5, in the order in which they are fixed.

Parameter	Choice	Input
q_1, q_2	admissible collar with $\sqrt{b} < q_1$	Definition 14.10 and Proposition 14.11
J, G	$J = -\Phi_\beta(q_*)/2$, $G = [\Phi_\beta(q_*) - \Phi_\beta(q_2)]/2$	Definition 14.10
δ	$0 < \delta < \min\{J, G\}/8$	Proposition 14.13
$a_{\mathcal{G}}$	rate of the good set	Proposition 14.13
K, h	$K = (J + 2\delta + 1)/2$, $h(r) = K[1 - \chi((r - q_1)/(q_2 - q_1))]$	
q_0, ε, s_0	$q_0 = (\sqrt{b} + q_1)/2$, $\varepsilon = (q_1 - q_0)^2/16$, $s_0 = q_0/2$	
c_ε, c_0	rates of the cap and magnetization bounds at ε and s_0	Corollary 14.16
c_E, C_E	$c_E = \min\{G - 4\delta, a_{\mathcal{G}}\}/2$, $C_E = 8\,\|h'\|_\infty^2/(ec_E)$	
γ	$\gamma = c_E/4$	
θ, C_M, C_D	$\theta = \min\{a_{\mathcal{G}}, c_0\}$, $C_M = 1 + 2/s_0$, $C_D = 32C_E/q_1^2$	
κ, c, a, C	$\kappa = \theta/2$, $c = \theta/4$, $a = \min\{c_\varepsilon, \theta, \gamma\}$, $C = \max\{7, 1 + C_D, 1 + C_M\}$	
N_0	large enough for finitely many inequalities in the proof	

16.2 Proof of the theorem

Proof of Theorem 1.5. We fix the parameters of Table 16.1 and work on an event Ω_N^*, the intersection of the event of Proposition 14.13 with two events given by Corollary 14.16. On Ω_N^* the argument is deterministic. We first show that the overlap-weighted Gibbs mean B_N has exponentially small vector energy and is aligned with U on the good set $\mathcal{G}$. Conclusion (i) then follows from Lemma 15.2, (ii) and (iii) from Lemma 15.5, and (iv) and (v) from Lemma 15.8. At the end we collect the constants.

The collar and the good set. Fix ξ, β, and b. By Proposition 14.11(a), there exists an admissible collar (q_1, q_2) in the sense of Definition 14.10 with $q_1 > \sqrt{b}$; that is, $\sqrt{b} < q_1 < q_2 < q_*$ and $\beta^2 A(r) > 1$ for all $r \in (q_1, q_*)$. We fix such a collar. Proposition 14.11(c) implies that the core exponent $J = -\Phi_\beta(q_*)/2$ and the collar gap $G = [\Phi_\beta(q_*) - \Phi_\beta(q_2)]/2$ are positive. Fix δ with $0 < \delta < \min\{J, G\}/8$, and let $a_{\mathcal{G}} > 0$, $w_{\mathcal{G}} > 0$, and $N_{\mathcal{G}}$

be the constants a, w, and N_0 of Proposition 14.13. For $N \ge N_{\mathcal{G}}$, with probability at least $1 - 5e^{-a_{\mathcal{G}} N}$, there exists a Borel set $\mathcal{G} \subset S_N$ with $\pi(\mathcal{G}^c) \le e^{-a_{\mathcal{G}} N}$ such that for all $x \in \mathcal{G}$,

$$\pi\{u : R(x,u) \ge q_2\} \ge e^{-N(J+2\delta)}, \qquad \pi\{u : q_1 \le R(x,u) \le q_2\} \le e^{-N(J+G-2\delta)}. \tag{16.1}$$

The first bound follows from the core lower bound (14.31), because the core band $\{u : |R(x,u) - q_*| \le w_{\mathcal{G}}\}$ lies in $\{u : R(x,u) > q_2\}$, and the second is the collar upper bound (14.32). We have written both with x as the reference point and u as the integration variable, which is possible because $R(x,u) = R(u,x)$.

The overlap-weighted Gibbs mean and the cap bounds. Let $K = (J + 2\delta + 1)/2$, so that

$$2K - J - 2\delta = 1, \tag{16.2}$$

and let $h(r) = K[1 - \chi((r - q_1)/(q_2 - q_1))]$ for $r \in [-1, 1]$, with χ as in Section 15.1. Then h is smooth and nonincreasing with values in $[0, K]$, it equals K on $[-1, q_1]$ and 0 on $[q_2, 1]$, and $\|h'\|_\infty \le 2K/(q_2 - q_1)$. Let B_N be the overlap-weighted Gibbs mean associated with h (Definition 15.6). Next, set

$$q_0 = \frac{\sqrt{b} + q_1}{2}, \qquad \varepsilon = \frac{(q_1 - q_0)^2}{16}, \qquad s_0 = \frac{q_0}{2},$$

so that $\sqrt{b} < q_0 < q_1$ and $q_0^2 > b$. By Corollary 14.16, applied at the two fixed latitudes s_0 and ε, there exist $c_0, c_\varepsilon > 0$ and N_R such that for $N \ge N_R$, with probability at least $1 - e^{-c_0 N} - e^{-c_\varepsilon N}$,

$$\sup_{v \in S_N} \pi\{y : R(y,v) \ge s_0\} \le \frac{2}{s_0}\, e^{-c_0 N}, \qquad |\pi(U)|^2 \le \varepsilon + e^{-c_\varepsilon N}. \tag{16.3}$$

Let Ω_N^* be the intersection of the event of Proposition 14.13 with the event on which (16.3) holds. Then

$$\mathbb{P}\big((\Omega_N^*)^c\big) \le 5e^{-a_{\mathcal{G}} N} + e^{-c_0 N} + e^{-c_\varepsilon N}. \tag{16.4}$$

For the rest of the proof, $N \ge \max\{N_{\mathcal{G}}, N_R\}$ and we work on Ω_N^*. All further statements are deterministic.

Energy and alignment of B_N. Fix $x \in \mathcal{G}$. By (15.16) and the first bound in (16.1), we have $d_N(x) \ge e^{-N(J+2\delta)}$. Inserting this and the second bound in (16.1) into (15.17), we obtain

$$\|\nabla B_N(x)\|_{\mathrm{HS}}^2 \le 4N \,\|h'\|_\infty^2\, e^{-N(G-4\delta)} \qquad (x \in \mathcal{G}).$$

For $x \notin \mathcal{G}$, we use the global bound $4N\,\|h'\|_\infty^2$ from (15.15). Integrating over x and using $\pi(\mathcal{G}^c) \le e^{-a_{\mathcal{G}} N}$, we find

$$\mathcal{E}_{\mathrm{vec}}(B_N) \le 4N \,\|h'\|_\infty^2 \left[e^{-N(G-4\delta)} + e^{-a_{\mathcal{G}} N}\right] \le 8N \,\|h'\|_\infty^2\, e^{-2c_E N} \le C_E e^{-c_E N}, \tag{16.5}$$

where $c_E = \min\{G - 4\delta, a_{\mathcal{G}}\}/2$, $C_E = 8\,\|h'\|_\infty^2/(ec_E)$, and in the last step we used $Ne^{-c_E N} \le (ec_E)^{-1}$. The points outside $\mathcal{G}$ enter (16.5) only through their Gibbs mass, and no comparison between $a_{\mathcal{G}}$ and J is needed.

Next, combining (15.18) with the bound $d_N(x) \ge e^{-N(J+2\delta)}$ and (16.2), we find

$$U(x) \cdot B_N(x) \ge q_1 - (1+q_1)e^{-N} \qquad (x \in \mathcal{G}). \tag{16.6}$$

Let N_1 be such that $(1+q_1)e^{-N} \le \min\{q_1 - q_0, q_1/4\}$ for $N \ge N_1$. For such N, we have $U \cdot B_N \ge q_0$ and $U \cdot B_N \ge 3q_1/4$ on $\mathcal{G}$. We note that (16.5) and (16.6) hold on the event of Proposition 14.13 alone, for every admissible collar (q_1, q_2) and every $0 < \delta < \min\{J, G\}/8$; the condition $q_1 > \sqrt{b}$ enters only through the choice of q_0.

The exponents. We set

$$\gamma = \frac{c_E}{4}, \qquad \Lambda_N = e^{-2\gamma N}, \qquad T_N = e^{\gamma N}.$$

Then (16.5) yields

$$\frac{\mathcal{E}_{\mathrm{vec}}(B_N)}{\Lambda_N} \le C_E e^{-2\gamma N}, \qquad T_N\,\mathcal{E}_{\mathrm{vec}}(B_N) \le C_E e^{-3\gamma N}. \tag{16.7}$$

The cutoff Λ_N is used for (i) and (ii), and the time T_N for (iii), (iv), and (v). As explained in Remark 15.4, the relation between Λ_N and T_N is a convenience.

Conclusion (i). Let $N \ge N_1$. We use Lemma 15.2 with $B = B_N$, $q = q_0$, the set $\mathcal{G}$, $\rho = q_0/8$, $\alpha = \chi$, and $\Lambda = \Lambda_N$. Note that $q_0 - 2\rho = 3q_0/4 \ge s_0$. By the cap bound in (16.3) and $\pi(\mathcal{G}^c) \le e^{-a_{\mathcal{G}}N}$, we have

$$M_{q_0/8} \le e^{-a_{\mathcal{G}}N} + \frac{2}{s_0}\,e^{-c_0 N} \le C_M e^{-\theta N}, \qquad \theta = \min\{a_{\mathcal{G}}, c_0\}, \quad C_M = 1 + \frac{2}{s_0}.$$

Using $\|\chi'\|_\infty \le 2$ and (16.7), we find that the numerator in (15.6) is at least

$$\frac{1}{2} - \frac{1024\,C_E N^2}{q_0^2}\,e^{-2\gamma N},$$

which is at least $1/4$ for $N \ge N_2$. We conclude that

$$\#\{j \ge 0 : \lambda_j \le \Lambda_N\} \ge \frac{e^{\theta N}}{4C_M}.$$

Let $\kappa = \theta/2$, and let N_3 be such that $e^{\theta N}/(4C_M) \ge e^{\kappa N} + 1$ for $N \ge N_3$. For such N, at least $\lfloor e^{\kappa N} \rfloor + 1$ eigenvalues, counted from $\lambda_0 = 0$, lie in $[0, \Lambda_N]$, and therefore

$$\lambda_{\lfloor e^{\kappa N} \rfloor} \le e^{-2\gamma N}.$$

The eigenvalues $\lambda_1, \dots, \lambda_{\lfloor e^{\kappa N} \rfloor}$ are positive, as λ_0 is simple. This is (i).

Conclusions (ii) and (iii). We define $\mathfrak{m}_N = \pi(U)$, $Z_N = U - \mathfrak{m}_N$, and $\overline{B}_N = B_N - \pi(B_N)$. Integrating (16.6) over $\mathcal{G}$, and using $U \cdot B_N \geq -1$ on $\mathcal{G}^c$, we have

$$\langle U, B_N \rangle \geq q_1 - (1 + q_1)\big(e^{-N} + e^{-a_{\mathcal{G}} N}\big). \tag{16.8}$$

By the magnetization bound in (16.3),

$$|\mathfrak{m}_N| \leq \sqrt{\varepsilon} + e^{-c_\varepsilon N/2}. \tag{16.9}$$

Further, $\|Z_N\|_2^2 = 1 - |\mathfrak{m}_N|^2 \leq 1$ and $\big\|\overline{B}_N\big\|_2^2 = \pi(|B_N|^2) - |\pi(B_N)|^2 \leq 1$. Using $|\pi(B_N)| \leq 1$, (16.8), and (16.9), we find

$$\langle Z_N, \overline{B}_N \rangle = \langle U, B_N \rangle - \mathfrak{m}_N \cdot \pi(B_N) \geq q_1 - \sqrt{\varepsilon} - \eta_N,$$

where $\eta_N = (1 + q_1)(e^{-N} + e^{-a_{\mathcal{G}} N}) + e^{-c_\varepsilon N/2}$. We now take $\Lambda = \Lambda_N$ in Lemma 15.5, so that $\Pi_{\Lambda_N} = \Pi_N$. By (15.12), the two norm bounds, and (16.7), we have

$$\|\Pi_N Z_N\|_2 \geq q_1 - \sqrt{\varepsilon} - \eta_N - C_E^{1/2} e^{-\gamma N}.$$

Recall that $\sqrt{\varepsilon} = (q_1 - q_0)/4$, and let N_4 be such that $\eta_N + C_E^{1/2} e^{-\gamma N} \leq 3(q_1 - q_0)/4$ for $N \geq N_4$. For such N, we have $\|\Pi_N Z_N\|_2 \geq q_0$, and

$$\|\Pi_N Z_N\|_2^2 \geq q_0^2 > b. \tag{16.10}$$

This is (ii).

For (iii), consider the stationary diffusion. Lemma 3.6(3) gives $\mathbb{E}_\pi R(X_0, X_t) = \mathbb{E}_\pi[U(X_0) \cdot U(X_t)] = \langle U, P_t U \rangle$. We have $P_t \mathfrak{m}_N = \mathfrak{m}_N$ and $\pi(Z_N) = 0$, which implies that $\langle U, P_t U \rangle = \langle Z_N, P_t Z_N \rangle + |\mathfrak{m}_N|^2$; that is, the connected correlation in (iii) equals $\langle Z_N, P_t Z_N \rangle$. For $0 \leq t \leq T_N$, we have $\Lambda_N t \leq e^{-\gamma N}$, and by (15.13) and (16.10),

$$\mathbb{E}_\pi R(X_0, X_t) - |\pi(U)|^2 = \langle Z_N, P_t Z_N \rangle \geq e^{-e^{-\gamma N}} q_0^2 \geq b$$

for $N \geq N_5$, where N_5 depends on the fixed margin $q_0^2 - b > 0$. This is (iii). Applying Lemma 15.3 to Z_N and $\overline{B}_N$ would give the same conclusion without spectral projections.

Conclusion (iv). Let $N \geq N_1$. We use Lemma 15.8 with $B = B_N$, $T = T_N$, $\rho = q_1/4$, $\eta = e^{-\gamma N}$, and with $q = 3q_1/4$ on the set $\mathcal{G}$. Part (2) of that lemma and (16.7) give

$$\pi(\mathcal{Y}_N^c) \leq \frac{2 T_N \mathcal{E}_{\mathrm{vec}}(B_N)}{(q_1/4)^2 e^{-\gamma N}} \leq C_D e^{-2\gamma N}, \qquad C_D = \frac{32 C_E}{q_1^2},$$

where $\mathcal{Y}_N$ is the set $\mathcal{Y}$ of Lemma 15.8. We define $\mathcal{X}_N = \mathcal{G} \cap \mathcal{Y}_N$. Then

$$\pi(\mathcal{X}_N^c) \leq e^{-a_{\mathcal{G}} N} + C_D e^{-2\gamma N}. \tag{16.11}$$

In part (3) of Lemma 15.8, we have $q - \rho = q_1/2 \geq s_0$, and by (16.3),

$$w \leq e^{-a_{\mathcal{G}} N} + \frac{2}{s_0}\, e^{-c_0 N} \leq C_M e^{-\theta N}. \tag{16.12}$$

By (15.20), for all $x \in \mathcal{X}_N$,

$$\inf_{0 \leq t \leq T_N} \|P_t(x, \cdot) - \pi\|_{\mathrm{TV}} \geq 1 - e^{-\gamma N} - C_M e^{-\theta N}. \tag{16.13}$$

The test sets D_x may contain points outside $\mathcal{G}$. Their Gibbs mass is included in w, and no separate bound on transitions into $\mathcal{G}^c$ is needed.

Conclusion (v). By (15.21) and (16.12), we have for all $x \in \mathcal{X}_N$ and $0 < t \leq T_N$ that

$$D\big(P_t(x, \cdot) \,\big\|\, \pi\big) \geq \big(1 - e^{-\gamma N}\big)\big(\theta N - \log C_M\big) - \log 2 \geq \frac{\theta N}{2}$$

for $N \geq N_6$. For the stationary diffusion in the fixed field and $t > 0$, we have $I(X_0; X_t) = \int D(P_t(x, \cdot) \| \pi)\, \pi(\mathrm{d}x)$ by (3.13) in Lemma 3.12, applied with $K = P_t$. By (16.11), there exists N_7 such that $\pi(\mathcal{X}_N) \geq 1/2$ for $N \geq N_7$. Integrating over $\mathcal{X}_N$, we conclude that

$$I(X_0; X_t) \geq \frac{\theta N}{4} \qquad (0 < t \leq T_N).$$

This is (v) with $c = \theta/4$. At $t = 0$ the mutual information is infinite, since the law of (X_0, X_0) is carried by the diagonal, which is a null set for $\pi \otimes \pi$.

The constants. We set $a = \min\{c_\varepsilon, \theta, \gamma\}$ and

$$C = \max\{7,\ 1 + C_D,\ 1 + C_M\}. \tag{16.14}$$

As $\theta \leq a_{\mathcal{G}}$ and $\theta \leq c_0$, (16.4) implies that $\mathbb{P}((\Omega_N^*)^c) \leq 7e^{-aN} \leq Ce^{-aN}$. Similarly, (16.11) implies that $\pi(\mathcal{X}_N) \geq 1 - Ce^{-aN}$, because $a \leq \min\{a_{\mathcal{G}}, \gamma\}$, and by (16.13), the total variation distance in (iv) is at least $1 - Ce^{-aN}$. Let N_0 be the largest of $N_{\mathcal{G}}, N_R, N_1, \ldots, N_7$. For $N \geq N_0$, conclusions (i)–(v) hold on Ω_N^*, whose probability is at least $1 - Ce^{-aN}$, with constants κ, γ, a, c, C that depend only on ξ, β, and b. This completes the proof. □

16.3 A direct route to slow modes

The proof above obtains all five conclusions from one observable. The following remark applies Lemma 15.1 directly to cap indicators. This gives better spectral exponents and is used in Section 18.

Remark 16.1 (Slow modes from cap indicators). The energy bound (16.5) contains the exponent $a_{\mathcal{G}}$ of the exceptional references, and the spectral exponent $2\gamma = c_E/2$ of (i) can

be much smaller than the collar gap G. Taking indicators of caps as trial functions avoids this loss. Let (q_1, q_2) be an admissible collar (Definition 14.10); the condition $q_1 > \sqrt{b}$ is not needed here. Let $c_1 > 0$ be the rate given by Corollary 14.16 at the latitude q_1, and fix $c_1' \in (0, c_1)$. Fix δ with $0 < \delta < \min\{J, G\}/8$ and $4\delta < \min\{G, c_1'\}$, and let $\mathcal{G}$ be the good set of Proposition 14.13, with $m_{\mathcal{G}} = \pi(\mathcal{G})$. Let $\alpha(r) = \chi((r - q_1)/(q_2 - q_1))$, so that $\alpha = 0$ on $(-\infty, q_1]$, $\alpha = 1$ on $[q_2, \infty)$, and $|\alpha'| \le 2/(q_2 - q_1)$. We use Lemma 15.1 with

$$\mathcal{V} = S_N, \qquad \mu(\mathrm{d}u) = \mathbf{1}_{\mathcal{G}}(u)\, \pi(\mathrm{d}u), \qquad f_u(x) = \alpha(R(x, u)).$$

On the event of Proposition 14.13, we have

$$\mathsf{T} \ge m_{\mathcal{G}}\, e^{-N(J+2\delta)}, \qquad \mathsf{R} = e^{-N(J-2\delta)}, \qquad \int \mathcal{E}(f_u, f_u)\, \mu(\mathrm{d}u) \le \frac{4 m_{\mathcal{G}}}{N(q_2 - q_1)^2}\, e^{-N(J+G-2\delta)}.$$

The first bound uses the core lower bound (14.31), since $f_u(x) = 1$ when $R(x, u) \ge q_2$. The second uses the cap upper bound (14.30), since $f_u \le \mathbf{1}\{R(\cdot, u) \ge q_1\}$. For the third, we have $|\nabla f_u|^2 = \alpha'(R)^2(1 - R^2)/N$ with $R = R(\cdot, u)$ by (15.2), which is at most $4/(N(q_2 - q_1)^2)$ and vanishes outside the collar $\{q_1 \le R(\cdot, u) \le q_2\}$; we then use the collar upper bound (14.32). Dividing the third bound by the first, before replacing $m_{\mathcal{G}}$ by $1/2$, we may take

$$r = \frac{4}{N(q_2 - q_1)^2}\, e^{-N(G-4\delta)}.$$

The factor $m_{\mathcal{G}}$ cancels, and the exponent $a_{\mathcal{G}}$ does not enter r. On the event of Corollary 14.16 at the latitude q_1, we have for every $x \in S_N$ that

$$\int f_u(x)\, \mu(\mathrm{d}u) \le \pi\{u : R(u, x) \ge q_1\} \le \frac{2}{q_1}\, e^{-c_1 N} \le e^{-c_1' N} = \mathsf{C}$$

for all sufficiently large N. It follows that $\mathsf{T}/(\mathsf{RC}) \ge m_{\mathcal{G}} e^{N(c_1' - 4\delta)} \ge e^{N(c_1' - 4\delta)}/2$ once $m_{\mathcal{G}} \ge 1/2$. Now fix $\sigma \in (0, G - 4\delta)$ and $\kappa \in (0, c_1' - 4\delta)$, and let $\Lambda = e^{-\sigma N}$. Then $r/\Lambda \to 0$, and Lemma 15.1 implies that

$$\#\{j \ge 0 : \lambda_j \le e^{-\sigma N}\} \ge \frac{1}{4} e^{N(c_1' - 4\delta)} \ge \lfloor e^{\kappa N} \rfloor + 1, \qquad \text{so} \qquad \lambda_{\lfloor e^{\kappa N} \rfloor} \le e^{-\sigma N},$$

for all sufficiently large N, with probability at least $1 - 5e^{-a_{\mathcal{G}} N} - e^{-c_1 N}$. The exponent $a_{\mathcal{G}}$ affects only this probability. Since $c_1' < c_1$ and $\delta > 0$ may be chosen freely, every spectral exponent $\sigma < G$ and every count exponent $\kappa < c_1$ are available for the chosen collar. Unlike the proof of Theorem 1.5, this argument uses the cap upper bound of Proposition 14.13, and it gives no information about the spin.

SECTION 17

Correlation bounds at every level below q_*

In this section, we prove Theorem 1.6. Its conclusions (i), (iv), and (v) do not involve the plateau b; they follow from Theorem 1.5. The new statement concerns (ii) and (iii). It asserts that for every fixed $b < q_*$, with probability tending to one, the slow part of the centered spin $Z_N = U - \pi(U)$ has squared norm at least b and the connected stationary correlation of the spin stays above b up to time $e^{\gamma N}$. In Theorem 1.5, these two conclusions hold with probability at least $1 - Ce^{-aN}$, but only for $b < q_*^2$.

We first explain why the proof of Theorem 1.5 stops at q_*^2. That proof detects slow modes of the spin with the overlap-weighted Gibbs mean B_N of Definition 15.6, and shows that $\langle U, B_N \rangle \geq q_1 - o(1)$, where $q_1 < q_*$ may be taken arbitrarily close to q_*. This lower bound is close to optimal. Indeed, $U(x) \cdot B_N(x)$ is the mean of $R(x,u)$ under the reweighted Gibbs measure w_x of Definition 15.6. We will see that w_x charges mostly points u with $R(x,u)$ close to q_*. The loss occurs in the next step. Let $\overline{B}_N = B_N - \pi(B_N)$. By Lemma 15.5, we have

$$\|\Pi_\Lambda Z_N\|_2 \, \|\overline{B}_N\|_2 \geq \langle Z_N, \overline{B}_N \rangle - o(1),$$

and the proof of Theorem 1.5 uses only $\|\overline{B}_N\|_2 \leq 1$, which leads to $\|\Pi_\Lambda Z_N\|_2^2 \geq q_1^2 - o(1)$. However, by Lemma 15.7, $|B_N(x)|^2$ is the mean of $R(u,u')$ for two independent samples u, u' from w_x. If the overlap of two such samples concentrates at q_*, then $|B_N(x)|^2$ is close to q_*, and the same inequality implies that $\|\Pi_\Lambda Z_N\|_2^2 \geq q_1^2/q_* - o(1)$, which is close to q_*. The main step of this section is therefore the estimate

$$\int |B_N|^2 \, \mathrm{d}\pi \longrightarrow q_* \qquad \text{in probability,}$$

which is Proposition 17.1 below.

To prove it, we work under the planted law, in which x is uniform on S_N and the field is a centered Gaussian field plus the drift $\beta N \xi(R(x,\cdot))$. We write each $u \in S_N$ through its latitude $r = R(x,u)$ and a transverse part orthogonal to x. We show, using the constrained pressure of Section 14, that the latitude concentrates at q_* under w_x. For two samples, $R(u,u')$ equals rr' plus $((1-r^2)(1-r'^2))^{1/2}$ times the overlap of the transverse parts. It suffices to show that the transverse overlap concentrates at

$$t_0 = \frac{q_*}{1+q_*},$$

because $q_*^2 + (1 - q_*^2)t_0 = q_*$. By the equality at q_* (Proposition 14.11), the Parisi measure of the transverse spin glass at latitude q_* is the point mass at t_0. Its covariance contains the monomials t and t^2 with positive coefficients, even when ξ has no quadratic term. We multiply the corresponding two Gaussian components of the field by amplitudes s_1 and s_2, and differentiate the expected logarithm of the partition function in these amplitudes. By Gaussian integration by parts, the two derivatives at $s_1 = s_2 = 1$ are affine functions of the first and the second moment of the transverse overlap. Their limits can be computed from the Crisanti–Sommers formula by the envelope theorem, and convexity in the amplitudes lets us pass from the convergence of the pressures to the convergence of their derivatives. It follows that the two moments are $t_0 + o(1)$ and $t_0^2 + o(1)$, so that the variance of the transverse overlap tends to zero. Finally, since (ξ, β) is strictly replica symmetric with $\beta^2\xi''(0) < 1$, the planted law and the Gibbs-reference law are mutually contiguous [93, Lemma 3.2, Remark 3.3, and Corollary 3.5], which transfers the estimate to the original disorder. Contiguity gives convergence in probability but no rate, and this is why (ii) and (iii) hold only with probability tending to one in Theorem 1.6.

The proof uses Definition 15.6 and Lemmas 15.5 and 15.7 from Section 15. From Section 14, it uses the planted law (Definition 14.1 and Lemma 14.2), the sections of the sphere and the uniform latitude estimates (Lemmas 14.5 and 14.6), the trial bound and the equality at q_* (Lemma 14.9 and Proposition 14.11), the good reference points of Proposition 14.13, and the uniform cap bound of Corollary 14.16. From Section 3, it uses the representation of the field (Lemma 3.2), the concentration of free energies (Lemma 3.3), the Crisanti–Sommers functional, the Parisi formula, and the Crisanti–Sommers optimality criterion (Definition 3.13, Theorem 3.14, and Lemma 3.15), and the limit law of the normalized partition function (Lemma 3.19). No later section depends on this one.

Differentiating a free energy in the coefficients of its covariance in order to identify moments of the overlap is a classical device in the theory of mean-field spin glasses; see [122], and [13, Proposition 2.2] for the derivative in the coefficients of a spherical mixture. Here it is applied to the transverse model under planting, whose components of degree one and two need not come from terms of ξ of the same degree. The equation $\beta^2(1 - q)\xi'(q) = q$ is the stationary plateau equation of [19, Eq. (14)] and [70, Eq. (A2)]. In the pure models of [19], the predicted plateau is its largest positive solution, which is q_*. For general finite mixtures, we keep the definition of q_* as the right endpoint of the rightmost component of $\{\beta^2 A > 1\}$, which excludes isolated solutions farther to the right at which $\beta^2 A$ touches 1 from below. The lower bounds of Theorem 1.6 approach q_*; we do not prove a matching upper bound.

17.1 Setting and parameters

Throughout this section, ξ is a finite mixture with $\beta_{\mathrm{d}} < \beta_{\mathrm{s}}$, the inverse temperature $\beta \in (\beta_{\mathrm{d}}, \beta_{\mathrm{s}})$ is fixed, and $q_* = \sup\{r \in (0,1) \colon \beta^2 A(r) > 1\}$. Sums over p run over the finitely many degrees with $c_p > 0$. As in Section 14, we write $F = \beta^2 \xi(1)/2$ and use the trial pressure Ψ, the admissible collars (q_1, q_2), the core exponent J, and the collar gap G of Definition 14.10. Recall from Proposition 14.11 that $q_* \in (0,1)$, that $J > 0$ and $\Psi(q_*) = F - J$, that $\beta^2 A \le 1$ on $[q_*, 1)$, and that there exists $\varepsilon_0 > 0$ such that $\beta^2 A > 1$ on $(q_* - \varepsilon_0, q_*)$. Let (q_1, q_2) be an admissible collar. Then (14.22) shows that Ψ is strictly increasing on $[q_1, q_*]$. Since $\beta^2 A \le 1$ on $[q_*, 1)$ and the polynomial $\beta^2 A - 1$ has finitely many zeros, Ψ is strictly decreasing on $[q_*, 1)$.

We set $t_0 = q_*/(1 + q_*)$, which is the value of $r_0 = r/(1+r)$ in Lemma 14.9 at $r = q_*$. Note that

$$q_*^2 + (1 - q_*^2)\, t_0 = q_*^2 + (1 - q_*) q_* = q_*. \tag{17.1}$$

The overlap-weighted Gibbs mean. Let (q_1, q_2) be an admissible collar and $K > 0$, and let $h \colon [-1,1] \to [0, K]$ be smooth and nonincreasing, with $h = K$ on $[-1, q_1]$ and $h = 0$ on $[q_2, 1]$. As in Definition 15.6, we set $\psi_u(x) = e^{-N h(R(x,u))}$ and $d_N(x) = \int \psi_u(x)^2 \, \pi(\mathrm{d}u)$, and we let B_N be the overlap-weighted Gibbs mean associated with h. The reweighted Gibbs measure is $w_x(\mathrm{d}u) = \psi_u(x)^2 \pi(\mathrm{d}u)/d_N(x)$, so that $B_N(x) = \int U \, \mathrm{d}w_x$. Then Lemma 15.7(2) gives

$$U(x) \cdot B_N(x) = \int R(x,u)\, w_x(\mathrm{d}u), \qquad |B_N(x)|^2 = \iint R(u,u')\, w_x(\mathrm{d}u)\, w_x(\mathrm{d}u'). \tag{17.2}$$

The measure w_x and the mean B_N depend on the field H through $\pi = \pi^H$. When this dependence matters, we use the notation w_x^H and B_N^H.

Parameters. Table 17.1 lists the constants in the proof of Theorem 1.6, in the order in which they are fixed. Each depends only on ξ, β, b, and the entries above it, and all of them are fixed before $N \to \infty$. The proof of Proposition 17.1 in Sections 17.3 to 17.6 uses further auxiliary constants, which are also chosen before $N \to \infty$.

Because $b' < q_*$, we have $(b' q_*)^{1/2} < q_*$, and q_1 can be chosen as required. Then (q_1, q_2) is an admissible collar. The squared inequality defining ϑ is strict at $\vartheta = 0$, because $q_1^2 > b' q_*$. We may therefore choose $0 < \vartheta < \min\{1, q_1/3\}$ so small that it holds. The condition $\vartheta < q_1/3$ gives $q_1 - 3\vartheta > 0$, which is used in the proof of (ii) below.

17.2 Reduction to the norm of the overlap-weighted Gibbs mean

The following proposition is the main step. It is proved in Sections 17.3 to 17.6.

Table 17.1. Parameters in the proof of Theorem 1.6, in the order in which they are fixed.

Parameter	Choice	Input
b'	$b' = (b + q_*)/2$, so that $b < b' < q_*$	
q_1, q_2	$q_* - \varepsilon_0 < q_1 < q_2 < q_*$ and $q_1^2 > b' q_*$	Proposition 14.11
δ	$0 < \delta < \min\{J, G\}/8$	Proposition 14.13
K, h	$K = (J + 2\delta + 1)/2$, and h as in Section 17.1	
ϑ	$0 < \vartheta < \min\{1, q_1/3\}$ and $(q_1 - 3\vartheta)^2 > b'(q_* + \vartheta)$	
c_E, a_E, C_E	energy and alignment constants	Lemma 17.2
c_ϑ	rate of the cap bound at $s = \vartheta^2$	Corollary 14.16
$\kappa', \gamma', a', c', C'$	constants for the plateau $q_*^2/2$	Theorem 1.5
γ	$\gamma = \min\{\gamma', c_E/4\}$	
κ, a, c, C	$\kappa = \kappa'$, $a = a'$, $c = c'$, $C = C'$	

Proposition 17.1 (The norm of the overlap-weighted Gibbs mean)**.** *Let (q_1, q_2) be an admissible collar, let $K > J/2$, and let h and B_N be as in Section 17.1. Then*

$$\int \left(|B_N(x)|^2 - q_* \right)^2 \pi(\mathrm{d}x) \longrightarrow 0 \qquad \textit{in probability.}$$

In particular, $\int |B_N|^2 \, \mathrm{d}\pi \to q_*$ in probability.

The next lemma records the two other properties of B_N that we need. They hold with probability exponentially close to one, and they are obtained in the proof of Theorem 1.5.

Lemma 17.2 (Energy and alignment)**.** *Let q_1, q_2, δ, K, h be as in Table 17.1, and let B_N be the overlap-weighted Gibbs mean associated with h. There exist $c_E, a_E, C_E > 0$ such that, for all sufficiently large N, with probability at least $1 - C_E e^{-a_E N}$,*

$$\mathcal{E}_{\mathrm{vec}}(B_N) \le e^{-c_E N} \qquad \textit{and} \qquad \langle U, B_N \rangle \ge q_1 - 2(1 + q_1) e^{-a_E N}.$$

Proof. The argument in the proof of Theorem 1.5 that gives these two bounds uses only that (q_1, q_2) is an admissible collar, that $8\delta < \min\{J, G\}$, and that $2K - J - 2\delta = 1$. We recall it here. By Proposition 14.13, there exist $a_{\mathcal{G}} > 0$ and $N_{\mathcal{G}}$ such that, for $N \ge N_{\mathcal{G}}$, with probability at least $1 - 5e^{-a_{\mathcal{G}} N}$, there exists a Borel set $\mathcal{G} \subset S_N$ with $\pi(\mathcal{G}^c) \le e^{-a_{\mathcal{G}} N}$ such that, for every $x \in \mathcal{G}$,

$$\pi\{u \colon R(x,u) \ge q_2\} \ge e^{-N(J+2\delta)}, \qquad \pi\{u \colon q_1 \le R(x,u) \le q_2\} \le e^{-N(J+G-2\delta)}.$$

The first bound follows from (14.31), since the core band lies in $(q_2, 1)$, and the second is (14.32). Here we used $R(x,u) = R(u,x)$. Fix $x \in \mathcal{G}$. We have $d_N(x) \ge e^{-N(J+2\delta)}$ by (15.16). Inserting this into (15.17) and (15.18), we obtain

$$\begin{aligned} &\|\nabla B_N(x)\|_{\mathrm{HS}}^2 \le 4N \, \|h'\|_\infty^2 \, e^{-N(G-4\delta)}, \\ &U(x) \cdot B_N(x) \ge q_1 - (1 + q_1) e^{-N(2K-J-2\delta)} = q_1 - (1 + q_1) e^{-N}. \end{aligned}$$

For $x \notin \mathcal{G}$, we use the bound $\|\nabla B_N(x)\|_{\mathrm{HS}}^2 \le 4N\,\|h'\|_\infty^2$ of (15.15) and $U(x)\cdot B_N(x) \ge -1$. Integrating over x, we find

$$\mathcal{E}_{\mathrm{vec}}(B_N) \le 4N\,\|h'\|_\infty^2\,\big(e^{-N(G-4\delta)} + e^{-a_{\mathcal{G}}N}\big), \qquad \langle U, B_N\rangle \ge q_1 - (1+q_1)\big(e^{-N} + e^{-a_{\mathcal{G}}N}\big).$$

The lemma follows with $c_E = \min\{G-4\delta, a_{\mathcal{G}}\}/2$, $a_E = \min\{1, a_{\mathcal{G}}\}$, and $C_E = 5$, after enlarging N so that $8N\,\|h'\|_\infty^2\, e^{-c_E N} \le 1$. □

Proof of Theorem 1.6. Fix $0 < b < q_*$, and fix the parameters of Table 17.1 in the order listed. We first treat conclusions (i), (iv), and (v), and then (ii) and (iii).

Conclusions (i), (iv), and (v). Theorem 1.5 applies with the plateau $q_*^2/2 < q_*^2$. Let $\kappa', \gamma', a', c', C'$ be the resulting constants, so that, for all sufficiently large N, with probability at least $1 - C'e^{-a'N}$, conclusions (i), (iv), and (v) hold with these constants. Let $\gamma = \min\{\gamma', c_E/4\}$, $\kappa = \kappa'$, $a = a'$, $c = c'$, and $C = C'$. Since $\gamma \le \gamma'$, we have $e^{-2\gamma' N} \le e^{-2\gamma N}$ and $[0, e^{\gamma N}] \subset [0, e^{\gamma' N}]$. Then (i), (iv), and (v) also hold with the constants κ, γ, a, c, C, on the same event.

Conclusions (ii) and (iii). Let Ω_N be the intersection of the events of Lemma 17.2 and of Corollary 14.16 with $s = \vartheta^2$, on which $|\pi(U)|^2 \le \vartheta^2 + e^{-c_\vartheta N}$, with the event $\{\int |B_N|^2\,\mathrm{d}\pi \le q_* + \vartheta\}$. Then $\mathbb{P}(\Omega_N) \to 1$ by Lemma 17.2, Corollary 14.16, and Proposition 17.1. We work on Ω_N for the rest of the proof.

Let $\mathfrak{m}_N = \pi(U)$, $Z_N = U - \mathfrak{m}_N$, $\overline{B}_N = B_N - \pi(B_N)$, and $\Lambda_N = e^{-2\gamma N}$, so that $\Pi_{\Lambda_N} = \Pi_N$. Then $\|Z_N\|_2^2 = 1 - |\mathfrak{m}_N|^2 \le 1$ and

$$\big\|\overline{B}_N\big\|_2^2 = \pi\big(|B_N|^2\big) - |\pi(B_N)|^2 \le q_* + \vartheta.$$

Using $|\pi(B_N)| \le 1$, $|\mathfrak{m}_N| \le \vartheta + e^{-c_\vartheta N/2}$, and the alignment bound of Lemma 17.2, we find

$$\langle Z_N, \overline{B}_N\rangle = \langle U, B_N\rangle - \mathfrak{m}_N\cdot\pi(B_N) \ge q_1 - 2(1+q_1)e^{-a_E N} - \vartheta - e^{-c_\vartheta N/2}.$$

The energy bound of Lemma 17.2 and $4\gamma \le c_E$ imply that $(\mathcal{E}_{\mathrm{vec}}(B_N)/\Lambda_N)^{1/2} \le e^{-(c_E/2-\gamma)N} \le e^{-c_E N/4}$. By (15.12) and $\|Z_N\|_2 \le 1$, we have

$$\|\Pi_N Z_N\|_2\,\big\|\overline{B}_N\big\|_2 \ge \langle Z_N, \overline{B}_N\rangle - \big(\mathcal{E}_{\mathrm{vec}}(B_N)/\Lambda_N\big)^{1/2} \ge q_1 - \vartheta - \varepsilon_N,$$

where $\varepsilon_N = 2(1+q_1)e^{-a_E N} + e^{-c_\vartheta N/2} + e^{-c_E N/4}$. Let N_1 be such that $\varepsilon_N \le 2\vartheta$ for $N \ge N_1$. For such N, the right side is at least $q_1 - 3\vartheta > 0$. In particular, $\big\|\overline{B}_N\big\|_2 > 0$, and we have

$$\|\Pi_N Z_N\|_2^2 \ge \frac{(q_1 - 3\vartheta)^2}{q_* + \vartheta} > b',$$

where in the last step we used the choice of ϑ. Since $b' > b$, this proves (ii).

For (iii), recall from the proof of Theorem 1.5 that $\mathbb{E}_\pi R(X_0, X_t) - |\pi(U)|^2 = \langle Z_N, P_t Z_N \rangle$ for the stationary diffusion: indeed, $\mathbb{E}_\pi R(X_0, X_t) = \langle U, P_t U \rangle$ by the Markov property, $P_t \mathfrak{m}_N = \mathfrak{m}_N$, and $\pi(Z_N) = 0$. For $0 \le t \le e^{\gamma N}$, we have $\Lambda_N t \le e^{-\gamma N}$, and (15.13) implies that

$$\mathbb{E}_\pi R(X_0, X_t) - |\pi(U)|^2 \ge e^{-\Lambda_N t} \, \|\Pi_N Z_N\|_2^2 \ge e^{-e^{-\gamma N}} \, b' \ge b$$

for $N \ge N_2$, where N_2 depends on the fixed margin $b' - b > 0$. This proves (iii).

For $N \ge \max\{N_1, N_2\}$, conclusions (ii) and (iii) hold on Ω_N, whose probability tends to one, and conclusions (i), (iv), and (v) hold on an event of probability at least $1 - Ce^{-aN}$, all with the constants κ, γ, a, c, C. In particular, all five conclusions hold simultaneously with probability tending to one. This completes the proof. □

17.3 The planted law and the transverse decomposition

In this section and the next two, we prove a version of Proposition 17.1 under the planted law; Section 17.6 transfers it to the original disorder. Throughout these sections, (q_1, q_2), K, and h are as in Proposition 17.1. We fix a degree p_0 with $c_{p_0} > 0$. Constants may depend on ξ, β, q_1, q_2, K, h, and p_0, but not on N.

The planted law. We use the planted law P_N and the Gibbs-reference law Q_N of Definition 14.1, with the reference point called x. Recall from Lemma 14.2 that under P_N, the point x has law ν_N and that, conditionally on x, the field has the law of $G + \beta N \xi(R(x, \cdot))$, where G has the law of H_N. The quantity $\left|B_N^H(x)\right|^2$ is invariant under a simultaneous rotation of the field and of x. Indeed, for $g \in O(N)$, the Gibbs measure of $H \circ g^{-1}$ is the image of π^H under g, so $w_{gx}^{H \circ g^{-1}}$ is the image of w_x^H and $B_N^{H \circ g^{-1}}(gx) = g B_N^H(x)$. As $H_N \circ g^{-1}$ has the law of H_N by Lemma 3.2(2), and $\xi(R(gx, \cdot)) = \xi(R(x, \cdot)) \circ g^{-1}$, we conclude that the conditional law of $\left|B_N^H(x)\right|^2$ given x does not depend on x. The following proposition is the planted form of Proposition 17.1.

Proposition 17.3 (The norm of B_N under the planted law)**.** *Fix $x \in S_N$, let G have the law of H_N, and let $H = G + \beta N \xi(R(x, \cdot))$. Then*

$$\lim_{N \to \infty} \mathbb{E}\Big(\Big|B_N^H(x)\Big|^2 - q_*\Big)^2 = 0.$$

By the rotation invariance just explained, Proposition 17.3 implies that $\mathbb{E}_{\mathsf{P}_N}(\left|B_N^H(x)\right|^2 - q_*)^2 \to 0$, where now x is the reference point of the planted law.

For the rest of this section and in the next two, $x \in S_N$ is fixed, $N \ge 3$, and $n = N - 1$. As in Section 14.2, we identify $x^\perp$ with $\mathbb{R}^n$ and write $S_n \subset x^\perp$ for the sphere of radius

$\sqrt{n}$, ν_n for its uniform probability measure, and $R_n(y,y') = y\cdot y'/n$. For $r\in(-1,1)$ and $y\in S_n$, we set

$$u_r(y) = rx + \Big(\frac{N(1-r^2)}{n}\Big)^{1/2} y; \tag{17.3}$$

this is the point of (14.8), with the reference point called x. Then $u_r(y)\in S_N$ and $R(x,u_r(y)) = r$. Since $y,y'\perp x$, we have, for $r,r'\in(-1,1)$ and $y,y'\in S_n$,

$$R\big(u_r(y),u_{r'}(y')\big) = rr' + \big((1-r^2)(1-r'^2)\big)^{1/2} R_n(y,y'). \tag{17.4}$$

Every $u\in S_N\setminus\{x,-x\}$ equals $u_r(y)$ for exactly one pair (r,y), namely $r = R(x,u)$ and $y = (n/(N(1-r^2)))^{1/2}(u-rx)$. We call r the latitude and y the transverse part of u. The two points $\pm x$ form a null set for all the measures on S_N below, which have densities with respect to ν_N.

Independent Gaussian components. For $0\le k\le p$ and $r\in[-1,1]$, let

$$b_{p,k} = \Big(c_p\binom{p}{k}\Big)^{1/2}, \qquad a_{p,k}(r) = b_{p,k}\, r^{p-k}(1-r^2)^{k/2},$$

so that $\big|a_{p,k}\big|\le b_{p,k}$ on $[-1,1]$. We note that, by the binomial theorem, the auxiliary mixture η_r of (14.10) is

$$\eta_r(t) = \xi\big(r^2+(1-r^2)t\big) - \xi(r^2) = \sum_p\sum_{k=1}^{p} a_{p,k}(r)^2 t^k, \qquad r\in(-1,1). \tag{17.5}$$

Lemma 17.4 (Independent Gaussian components)**.** *Let $(G_{p,k})$, for $c_p>0$ and $0\le k\le p$, be independent centered Gaussian fields on $x^\perp$ such that $G_{p,k}$ is a homogeneous polynomial of degree k and*

$$\mathbb{E}G_{p,k}(w)G_{p,k}(w') = N\Big(\frac{w\cdot w'}{n}\Big)^k, \qquad w,w'\in x^\perp.$$

For $y\in\mathbb{R}^N$, let $P^\perp y$ be the orthogonal projection of y onto $x^\perp$, and define

$$G(y) = \sum_p\sum_{k=0}^{p} b_{p,k}\Big(\frac{n}{N}\Big)^{k/2} R(x,y)^{p-k}\, G_{p,k}(P^\perp y). \tag{17.6}$$

1. *For $y,y'\in\mathbb{R}^N$, we have $\mathbb{E}G(y)G(y') = N\xi(R(y,y'))$. In particular, the restriction of G to S_N has the law of H_N.*
2. *For $r\in(-1,1)$ and $y\in S_n$, we have $G(u_r(y)) = \sum_p\sum_{k=0}^{p} a_{p,k}(r)G_{p,k}(y)$.*
3. *There exist $K_1,c_1>0$ such that, for all sufficiently large N, the event $\mathcal{F}_N$ that $\sup_{S_n}\big|G_{p,k}\big|\le K_1N$ for all p and k has probability at least $1-e^{-c_1N}$. Moreover, $\mathbb{E}(\sup_{S_n}\big|G_{p,k}\big|)^2\le N^{k+1}$ for all p, k, and N.*

Proof. Such fields exist. Let $e_1, \dots, e_n$ be an orthonormal basis of $x^\perp$, and let $g^{(p,k)}_{i_1 \cdots i_k}$ be independent standard Gaussian variables. Then

$$G_{p,k}(w) = N^{1/2} n^{-k/2} \sum_{i_1,\dots,i_k=1}^{n} g^{(p,k)}_{i_1 \cdots i_k} (w \cdot e_{i_1}) \cdots (w \cdot e_{i_k}) \tag{17.7}$$

has covariance $N n^{-k} (w \cdot w')^k$. For $k = 0$, the field $G_{p,0} = N^{1/2} g^{(p,0)}$ is constant.

For (1), let $R = R(x,y)$, $R' = R(x,y')$, and $\varrho = (P^\perp y \cdot P^\perp y')/N$. Since $x/\sqrt{N}$ is a unit vector, we have $R(y,y') = RR' + \varrho$. Using the independence of the fields and the binomial theorem, we find

$$\begin{aligned} \mathbb{E} G(y) G(y') &= \sum_p c_p \sum_{k=0}^{p} \binom{p}{k} \Big(\frac{n}{N}\Big)^k (RR')^{p-k} N \Big(\frac{N\varrho}{n}\Big)^k \\ &= N \sum_p c_p (RR' + \varrho)^p = N \xi(R(y,y')). \end{aligned}$$

Two centered Gaussian processes on S_N with continuous paths and the same covariance have the same law.

For (2), note that $P^\perp u_r(y) = (N(1-r^2)/n)^{1/2} y$ and $R(x, u_r(y)) = r$. By the homogeneity of $G_{p,k}$, we have $(n/N)^{k/2} G_{p,k}(P^\perp u_r(y)) = (1-r^2)^{k/2} G_{p,k}(y)$, and (2) follows from (17.6).

For (3), it suffices to treat each of the finitely many pairs (p,k). For $k = 0$, the bound follows from $\mathbb{P}(|g| > K_1 N^{1/2}) \le e^{-K_1^2 N/2}$ for a standard Gaussian variable g. For $k = 1$, we have $G_{p,1}(y) = (N/n)^{1/2}\, y \cdot \tilde{g}$, where $\tilde{g} = \sum_i g_i^{(p,1)} e_i$ is a standard Gaussian vector in $x^\perp$, and $\sup_{S_n} |G_{p,1}| = N^{1/2} |\tilde{g}|$. The function $\tilde{g} \mapsto |\tilde{g}|$ is 1-Lipschitz and $\mathbb{E}\,|\tilde{g}| \le n^{1/2}$, and the Gaussian concentration inequality recalled in the proof of Lemma 3.3 gives $\mathbb{P}(|\tilde{g}| > 2N^{1/2}) \le 2e^{-N/2}$. For $k \ge 2$, the field $(n/N)^{1/2} G_{p,k}$ on S_n has covariance $n R_n(y,y')^k$, that is, it is the pure k-spin field in dimension n, and Lemma 3.2 applied to the mixture t^k in dimension n gives $K, c > 0$ with $\mathbb{P}(\sup_{S_n} |G_{p,k}| > (N/n)^{1/2} K n) \le e^{-cn}$ for large n. As $(N/n)^{1/2} n \le N$, this proves the first statement of (3). For the second, applying the Cauchy–Schwarz inequality to (17.7), we find $|G_{p,k}(y)| \le N^{1/2} n^{-k/2} |y|^k \left\| g^{(p,k)} \right\| = N^{1/2} \left\| g^{(p,k)} \right\|$ for $y \in S_n$, where $\left\| g^{(p,k)} \right\|$ is the Euclidean norm of the array, and $\mathbb{E} \left\| g^{(p,k)} \right\|^2 = n^k \le N^k$. □

We record a consequence of (3). For every event $\mathcal{A}$, splitting $\mathcal{A}$ according to $\mathcal{F}_N$ and using the Cauchy–Schwarz inequality on $\mathcal{F}_N^c$, we find, for all sufficiently large N,

$$\mathbb{E}\Big[\frac{1}{N} \sup_{S_n} |G_{p,k}| \; ; \mathcal{A}\Big] \le K_1 \mathbb{P}(\mathcal{A}) + N^{(k-1)/2} e^{-c_1 N/2}. \tag{17.8}$$

Amplitudes. By Lemma 17.4(1), we may and do realize the field H of Proposition 17.3 as $H = G + \beta N\xi(R(x,\cdot))$, with G given by (17.6); then $\mathbb{E}$ denotes the expectation over the components $G_{p,k}$. For $s = (s_1, s_2) \in [1/2, 3/2]^2$, let $s_{p_0,1} = s_1$, $s_{p_0,2} = s_2$, and $s_{p,k} = 1$ for all other pairs (p,k); since $p_0 \geq 2$, the components $G_{p_0,1}$ and $G_{p_0,2}$ exist. We define G_s by inserting the factor $s_{p,k}$ in each term of (17.6), and

$$\begin{aligned} \mathcal{Z}_N(s) &= \int_{S_N} \exp\Big\{\beta G_s(u) + \beta^2 N\xi(R(x,u)) - 2Nh(R(x,u))\Big\}\, \nu_N(\mathrm{d}u), \\ \mathrm{Y}_N(s) &= \frac{1}{N}\mathbb{E}\log \mathcal{Z}_N(s). \end{aligned} \tag{17.9}$$

The partition function $\mathcal{Z}_N(s)$ is not normalized. Let μ_s be the probability measure on S_N whose density with respect to ν_N is proportional to the integrand in (17.9), and let $\mu_s^{\otimes 2}$ be the law of two independent samples from μ_s. Only the Gaussian components are multiplied by amplitudes; the drift $\beta^2 N\xi(R(x,u))$ and the penalty $2Nh(R(x,u))$ do not change. At $s = (1,1)$, we have $G_s = G$ and $\beta G + \beta^2 N\xi(R(x,\cdot)) = \beta H$. Using the identity $e^{\beta H(u) - 2Nh(R(x,u))}\nu_N(\mathrm{d}u) = Z_{N,\beta}(H)\,\psi_u(x)^2\,\pi^H(\mathrm{d}u)$, where $Z_{N,\beta}(H) = \int e^{\beta H}\,\mathrm{d}\nu_N$ is the partition function of Definition 3.1, we find

$$\mu_{(1,1)} = w_x^H, \qquad \mathcal{Z}_N(1,1) = Z_{N,\beta}(H)\, d_N(x). \tag{17.10}$$

For $r \in (-1,1)$ and $s \in [1/2,3/2]^2$, we define the perturbed auxiliary mixture

$$\eta_{r;s}(t) = \sum_p \sum_{k=1}^{p} s_{p,k}^2\, a_{p,k}(r)^2\, t^k = \eta_r(t) + \sum_{k=1}^{2} (s_k^2 - 1)\, a_{p_0,k}(r)^2\, t^k. \tag{17.11}$$

It is a covariance polynomial in the sense of Definition 3.13, and $\eta_{r;(1,1)} = \eta_r$ by (17.5). We deduce from Lemma 17.4(2) that, for fixed $r \in (-1,1)$ and $y \in S_n$,

$$\begin{aligned} G_s(u_r(y)) &= G^{(0)}(r) + X_{r;s}(y), \qquad \text{where} \\ G^{(0)}(r) &= \sum_p a_{p,0}(r) G_{p,0}, \qquad X_{r;s}(y) = \sum_p \sum_{k=1}^{p} s_{p,k} a_{p,k}(r) G_{p,k}(y). \end{aligned} \tag{17.12}$$

The variable $G^{(0)}(r)$ is centered Gaussian and independent of the field $X_{r;s}$, whose covariance is

$$\mathbb{E}X_{r;s}(y)X_{r;s}(y') = N\eta_{r;s}\big(R_n(y,y')\big) = n \cdot \frac{N}{n}\,\eta_{r;s}\big(R_n(y,y')\big).$$

Thus $X_{r;s}$ has the law of $H_n^{(N/n)\eta_{r;s}}$, in the notation of Section 14.2.

Finally, for a Borel set $D \subset (-1,1)$, let $\mathcal{Z}_N^D(s)$ be the integral in (17.9) restricted to $\{u\colon R(x,u) \in D\}$. Using the coarea identity in the proof of Lemma 14.5(a), we have

$$\begin{aligned} \mathcal{Z}_N^D(s) &= C_N \int_D (1-r^2)^{(N-3)/2} e^{N(\beta^2\xi(r) - 2h(r))}\, \mathcal{Y}_N(r;s)\,\mathrm{d}r, \\ \mathcal{Y}_N(r;s) &= \int_{S_n} e^{\beta G_s(u_r(y))}\, \nu_n(\mathrm{d}y), \end{aligned} \tag{17.13}$$

where C_N is the constant of (14.11), and $N^{-1}\log C_N \to 0$.

17.4 The limiting pressure

For $s \in [1/2, 3/2]^2$, we define the profile

$$\Theta_s(r) = \Theta^0(r) + \mathcal{P}_\beta(\eta_{r;s}), \qquad \Theta^0(r) = \beta^2\xi(r) - 2h(r) + \frac{1}{2}\log(1-r^2), \qquad r \in (-1,1), \tag{17.14}$$

and $\Upsilon(s) = \sup_{-1<r<1} \Theta_s(r)$. At $s = (1,1)$, we have $\Theta_{(1,1)} = f - 2h$, where f is the function (14.12); by Proposition 14.7, $f = F_{\mathrm{FP}}$. We first show that q_* is the only maximizer of $f - 2h$, and then that $\Upsilon_N \to \Upsilon$.

Lemma 17.5 (The latitude profile)**.** *The following hold.*

1. *The measure δ_{t_0} is the unique minimizer of $\mathcal{P}_\beta(\cdot\,; \eta_{q_*})$.*
2. *For every $r \in (-1,1)$, we have $f(r) \le F$.*
3. *The function $f - 2h$ attains its supremum over $(-1,1)$ only at q_*, and $f(q_*) - 2h(q_*) = F - J$. For every $\varepsilon > 0$,*

$$\Delta_\varepsilon = F - J - \sup\bigl\{f(r) - 2h(r) \colon r \in (-1,1),\ |r - q_*| \ge \varepsilon\bigr\} > 0.$$

Proof. By Proposition 14.11(b) and Proposition 14.7, we have $f(q_*) = F_{\mathrm{FP}}(q_*) = \Psi(q_*)$. In view of the equality case of Lemma 14.9, this means that $\mathcal{P}_\beta(\delta_{t_0}; \eta_{q_*}) = \mathcal{P}_\beta(\eta_{q_*})$, so δ_{t_0} is a minimizer. It is the only one by Lemma 3.15(3). This proves (1); it is also the conclusion of the proof of Proposition 14.11.

For (2), let $r \in [0,1)$. Then Proposition 14.11(b) implies that $f(r) = F_{\mathrm{FP}}(r) \le \Psi(r) = F + \Phi_\beta(r)/2$, and $\Phi_\beta(r) \le 0$ since (ξ, β) is strictly replica symmetric (Lemma 3.18) and $\Phi_\beta(0) = 0$. For $r = -v \in (-1,0)$, we have $\eta_{-v} = \eta_v$, since η_r depends on r only through r^2, and $\xi(-v) \le \xi(v)$, since the coefficients of ξ are nonnegative. We conclude that $f(-v) \le f(v) \le F$.

For (3), let $r \in [q_1, 1)$. Using $h \ge 0$ and Proposition 14.11(b), we have $f(r) - 2h(r) \le \Psi(r)$, and by the monotonicity of Ψ recalled in Section 17.1, $\Psi(r) < \Psi(q_*) = F - J$ for $r \ne q_*$. For $r \in (-1, q_1)$, we have $h(r) = K$, and (2) implies that $f(r) - 2h(r) \le F - 2K$, which is smaller than $F - J$ because $2K > J$. Finally, $h(q_*) = 0$ because $q_* > q_2$, and therefore $f(q_*) - 2h(q_*) = \Psi(q_*) = F - J$. This proves the first statement of (3). For the second, the same three bounds and the monotonicity of Ψ imply that

$$\Delta_\varepsilon \ge \min\bigl\{2K - J,\ \Psi(q_*) - \Psi(q_* - \varepsilon),\ \Psi(q_*) - \Psi(q_* + \varepsilon)\bigr\} > 0,$$

where the second term is omitted when $q_* - \varepsilon < q_1$, and the third when $q_* + \varepsilon \ge 1$. □

The next lemma is the analog of Lemma 14.6 and Proposition 14.7 for the perturbed fields, with the penalty $2h$ and with all latitudes in $(-1,1)$.

Lemma 17.6 (The limiting pressure)**.** *The following hold.*

1. *For every $s \in [1/2, 3/2]^2$, the supremum $\mathrm{Y}(s)$ is attained, and $\mathrm{Y}_N(s) \to \mathrm{Y}(s)$.*
2. *For every $\varepsilon > 0$, we have $\mathbb{E}\, w_x^H\{u\colon |R(x,u) - q_*| \geq \varepsilon\} \to 0$.*

Proof. The proof has five steps. We first record continuity and growth properties of the profile, and then bounds on the fields. Next, we compare $N^{-1}\log \mathcal{Y}_N(r;s)$ with $\mathcal{P}_\beta(\eta_{r;s})$ uniformly on a compact set of latitudes. Then we integrate over the latitude, and finally we take expectations.

The profile. By Lemma 14.6(a) and (17.11), we have, for $r, r' \in (-1,1)$ and $s, s' \in [1/2,3/2]^2$,

$$\bigl|\mathcal{P}_\beta(\eta_{r;s}) - \mathcal{P}_\beta(\eta_{r;s'})\bigr| \leq \beta^2 \sum_{k=1}^{2} \bigl|s_k^2 - s_k'^2\bigr|\, a_{p_0,k}(r)^2, \tag{17.15}$$

$$\bigl|\mathcal{P}_\beta(\eta_{r;s}) - \mathcal{P}_\beta(\eta_{r';s})\bigr| \leq \beta^2 L_\eta \,|r - r'|\,, \qquad L_\eta = \frac{9}{2}\sum_p p\, 2^p c_p. \tag{17.16}$$

For (17.16), we used $s_{p,k}^2 \leq 9/4$, $|t| \leq 1$, $\sum_k \binom{p}{k} = 2^p$, and the bound $\bigl|\partial_r (r^{2(p-k)}(1-r^2)^k)\bigr| \leq 2(p-k) + 2k = 2p$ for $|r| \leq 1$. Next, for every covariance polynomial η, we have $0 \leq \mathcal{P}_\beta(\eta) \leq \beta^2\eta(1)/2$. The lower bound follows from Jensen's inequality, which gives $p_n(\eta) \geq n^{-1}\mathbb{E}\int \beta H_n^\eta \,\mathrm{d}\nu_n = 0$. The upper bound follows from $\mathcal{P}_\beta(\eta) \leq \mathcal{P}_\beta(\delta_0;\eta) = \beta^2\eta(1)/2$, see Definition 3.13 and the remark after it. Since $\eta_{r;s}(1) \leq (9/4)\eta_r(1) \leq (9/4)\xi(1)$, $\beta^2\xi(r) \leq \beta^2\xi(1)$, and $h \geq 0$, we obtain

$$\Theta_s(r) \leq \frac{17}{8}\beta^2\xi(1) + \frac{1}{2}\log(1-r^2), \qquad \mathrm{Y}(s) \geq \Theta_s(0) \geq -2K. \tag{17.17}$$

Note that Θ_s is continuous on $(-1,1)$ by (17.16), and that $\Theta_s(r) \to -\infty$ as $|r| \to 1$ by (17.17). Then the supremum $\mathrm{Y}(s)$ is attained.

Bounds on the fields. Let $K_2 = (3/2)K_1 \sum_p \sum_k p\, b_{p,k}$. Fix $\rho \in (0,1)$. For $|r| \leq \rho$, we have $\bigl|a'_{p,k}(r)\bigr| \leq b_{p,k}\, p\, (1-\rho^2)^{-1/2}$. On the event $\mathcal{F}_N$ of Lemma 17.4(3), we deduce from (17.12) and $s_{p,k} \leq 3/2$ that

$$\sup_{S_N} |G_s| \leq K_2 N, \qquad \bigl|\partial_r G_s(u_r(y))\bigr| = \Bigl|\sum_p \sum_{k=0}^{p} s_{p,k} a'_{p,k}(r) G_{p,k}(y)\Bigr| \leq \frac{K_2 N}{(1-\rho^2)^{1/2}},$$

where the second bound holds for $|r| \leq \rho$ and $y \in S_n$. For the first bound, note that every $u \in S_N \setminus \{\pm x\}$ is of the form $u_r(y)$, and that $G_s(\pm x) = \sum_p (\pm 1)^p b_{p,0} G_{p,0}$. The derivative $\partial_r \log \mathcal{Y}_N(r;s)$ is the average of $\beta\,\partial_r G_s(u_r(y))$ under a probability measure in y. On $\mathcal{F}_N$, we therefore have

$$\Bigl|\partial_r \frac{1}{N}\log \mathcal{Y}_N(r;s)\Bigr| \leq \frac{\beta K_2}{(1-\rho^2)^{1/2}} \quad (|r| \leq \rho), \qquad \mathcal{Y}_N(r;s) \leq e^{\beta K_2 N} \quad (-1 < r < 1). \tag{17.18}$$

Uniform comparison with the Parisi formula. Fix $r \in (-1,1)$. By (17.12) and $\mathbb{E}G^{(0)}(r) = 0$, we have $N^{-1}\mathbb{E}\log \mathcal{Y}_N(r;s) = (n/N)\, p_n((N/n)\eta_{r;s})$. Further, Lemma 14.6(a) implies that $|p_n((N/n)\eta_{r;s}) - p_n(\eta_{r;s})| \le \beta^2 n^{-1}\eta_{r;s}(1)$, and $p_n(\eta_{r;s}) \to \mathcal{P}_\beta(\eta_{r;s})$ by Theorem 3.14, since $\eta_{r;s}$ is a covariance polynomial; it may have a term of degree one. We conclude that

$$\lim_{N\to\infty} \frac{1}{N}\mathbb{E}\log \mathcal{Y}_N(r;s) = \mathcal{P}_\beta(\eta_{r;s}). \tag{17.19}$$

In the Gaussian coordinates of all the components, see (17.7), $G_s(u)$ is a linear function whose coefficient vector has squared norm $\mathbb{E}G_s(u)^2 \le 9N\xi(1)/4$ for $u \in S_N$. The proof of Lemma 3.3 therefore applies to $\log \mathcal{Y}_N(r;s)$, with $\xi(1)$ replaced by $9\xi(1)/4$: this function is Lipschitz in the coordinates with constant $(3/2)\beta(N\xi(1))^{1/2}$, and since $\beta^2\xi(1) = 2F$, we find

$$\mathbb{P}\Big(\Big|\frac{1}{N}\log \mathcal{Y}_N(r;s) - \frac{1}{N}\mathbb{E}\log \mathcal{Y}_N(r;s)\Big| > \varepsilon'\Big) \le 2e^{-\varepsilon'^2 N/(9F)}, \qquad \varepsilon' > 0. \tag{17.20}$$

Now fix $\varepsilon' \in (0,1)$, and choose $\rho \in (0,1)$ such that

$$\frac{17}{8}\beta^2\xi(1) + \beta K_2 + \frac{1}{2}\log(1-\rho^2) \le -2K - 1. \tag{17.21}$$

Let $L = \beta K_2(1-\rho^2)^{-1/2} + \beta^2 L_\eta$, and choose points $r_1, \dots, r_m \in [-\rho,\rho]$ such that every $r \in [-\rho,\rho]$ is within ε'/L of some r_i. Let $\mathcal{A}_N$ be the intersection of $\mathcal{F}_N$ with the m events $\{\big|N^{-1}\log \mathcal{Y}_N(r_i;s) - N^{-1}\mathbb{E}\log \mathcal{Y}_N(r_i;s)\big| \le \varepsilon'\}$. By Lemma 17.4(3) and (17.20),

$$\mathbb{P}(\mathcal{A}_N^c) \le e^{-c_1 N} + 2me^{-\varepsilon'^2 N/(9F)} \longrightarrow 0. \tag{17.22}$$

In view of (17.19), there exists N_0 such that $\big|N^{-1}\mathbb{E}\log \mathcal{Y}_N(r_i;s) - \mathcal{P}_\beta(\eta_{r_i;s})\big| \le \varepsilon'$ for $i \le m$ and $N \ge N_0$. For $r \in [-\rho,\rho]$, let r_i be within ε'/L of r. On $\mathcal{A}_N$, for $N \ge N_0$, (17.18) and (17.16) bound the changes from r to r_i by $L\,|r - r_i| \le \varepsilon'$ in total, and we obtain

$$\Big|\frac{1}{N}\log \mathcal{Y}_N(r;s) - \mathcal{P}_\beta(\eta_{r;s})\Big| \le 3\varepsilon', \qquad r \in [-\rho,\rho]. \tag{17.23}$$

The constants ρ, m, and N_0 depend on s and ε', and they are fixed before $N \to \infty$.

Integration over the latitude. We show that, on $\mathcal{A}_N$ and for $N \ge N_0$, every Borel set $D \subset (-1,1)$ satisfies

$$\frac{1}{N}\log \mathcal{Z}_N^D(s) \le \max\Big\{\sup_{D\cap[-\rho,\rho]} \Theta_s,\ \Upsilon(s) - 1\Big\} + 3\varepsilon' + \frac{1}{N}\log\big(4C_N(1-\rho^2)^{-3/2}\big), \tag{17.24}$$

and that there exists $w > 0$, independent of N, such that

$$\frac{1}{N}\log \mathcal{Z}_N(s) \ge \Upsilon(s) - 4\varepsilon' + \frac{1}{N}\log(2wC_N). \tag{17.25}$$

For (17.24), we split the integral in (17.13) into $|r| \le \rho$ and $|r| > \rho$. For $|r| \le \rho$, we use $(1-r^2)^{(N-3)/2} \le (1-\rho^2)^{-3/2} e^{(N/2)\log(1-r^2)}$, and (17.23) shows that the integrand is at most $C_N(1-\rho^2)^{-3/2} e^{N(\Theta_s(r)+3\varepsilon')}$. For $|r| > \rho$, we use $(1-r^2)^{(N-3)/2} \le (1-\rho^2)^{(N-3)/2}$, the bound $\beta^2\xi(r) - 2h(r) \le \beta^2\xi(1)$, and (17.18). Then (17.21) and (17.17) imply that the integrand is at most

$$C_N(1-\rho^2)^{-3/2} \exp\Big\{ N\Big[\beta^2\xi(1) + \beta K_2 + \frac{1}{2}\log(1-\rho^2)\Big]\Big\} \le C_N(1-\rho^2)^{-3/2} e^{N(\Upsilon(s)-1)}.$$

Each part is integrated over a set of length at most 2, which gives (17.24). For (17.25), let r_* be a maximizer of Θ_s. By (17.17) and (17.21), we have $\Theta_s(r) \le -2K - 1 < \Upsilon(s)$ for $|r| \ge \rho$, which implies that $|r_*| < \rho$. As Θ_s is continuous, there exists $w > 0$ with $[r_* - w, r_* + w] \subset (-\rho, \rho)$ and $\Theta_s \ge \Upsilon(s) - \varepsilon'$ on this interval. Using $(1-r^2)^{-3/2} \ge 1$ and (17.23), we find that the integrand in (17.13) is at least $C_N e^{N(\Upsilon(s)-4\varepsilon')}$ on $[r_* - w, r_* + w]$, which gives (17.25).

Expectations. For every realization of the components, $\xi(r) \ge -\xi(1)$ and $0 \le h \le K$ give

$$\Big|\frac{1}{N}\log \mathcal{Z}_N(s)\Big| \le \frac{\beta}{N}\sup_{S_N}|G_s| + \beta^2\xi(1) + 2K, \qquad \sup_{S_N}|G_s| \le \frac{3}{2}\sum_p \sum_k b_{p,k} \sup_{S_n}\big|G_{p,k}\big|. \tag{17.26}$$

The expectation of the right side of the first bound on $\mathcal{A}_N^c$ tends to zero by (17.8) and (17.22). Combining this with (17.24) for $D = (-1,1)$ and with (17.25), and recalling $N^{-1}\log C_N \to 0$, we obtain

$$\Upsilon(s) - 4\varepsilon' \le \liminf_{N\to\infty} \Upsilon_N(s) \le \limsup_{N\to\infty} \Upsilon_N(s) \le \Upsilon(s) + 3\varepsilon'.$$

Since ε' is arbitrary, this proves (1).

For (2), let $s = (1,1)$ and $D = \{r \in (-1,1) \colon |r - q_*| \ge \varepsilon\}$. Recall from Lemma 17.5(3) that $\Upsilon(1,1) = F - J$ and $\sup_D \Theta_{(1,1)} \le F - J - \Delta_\varepsilon$. We run the construction above with $\varepsilon' = \min\{\Delta_\varepsilon, 1\}/16$. By (17.10), we have $w_x^H\{u \colon |R(x,u) - q_*| \ge \varepsilon\} = \mathcal{Z}_N^D(1,1)/\mathcal{Z}_N(1,1)$. On $\mathcal{A}_N$, we find from (17.24) and (17.25) that, for all sufficiently large N,

$$w_x^H\{u \colon |R(x,u) - q_*| \ge \varepsilon\} \le \exp\Big\{-N\big(\min\{\Delta_\varepsilon, 1\} - 7\varepsilon'\big) + o(N)\Big\} \le e^{-N\min\{\Delta_\varepsilon,1\}/2}.$$

Since this probability is at most one, we conclude that $\mathbb{E}\, w_x^H\{u \colon |R(x,u) - q_*| \ge \varepsilon\} \le e^{-N\min\{\Delta_\varepsilon,1\}/2} + \mathbb{P}(\mathcal{A}_N^c)$, which tends to zero by (17.22). □

17.5 Two amplitude derivatives

In this section, we compute the derivatives of Υ in s_1 and s_2 at $(1,1)$, we express the derivatives of Υ_N through the overlaps of μ_s, and we combine the two to prove Proposition 17.3.

We use the notation of Definition 3.13 and Lemma 3.15: $\mathscr{M}$ is the set of admissible measures, $\hat{x}_\zeta$ and $\mathcal{P}_\beta(\zeta;\eta)$ are as in (3.16), and f_ζ and S_ζ are the first variation (3.18) and its maximizing set, with derivative f'_ζ given by (14.19). Recall from Lemma 3.15(3), which is [90, Lemma 1.2], that for every covariance polynomial $\eta \neq 0$ the functional $\mathcal{P}_\beta(\cdot\,;\eta)$ has a unique minimizer $\zeta \in \mathscr{M}$ and that $\zeta(S_\zeta) = 1$.

Lemma 17.7 (Minimizers)**.** *For $r \in (-1,1)$ and $s \in [1/2,3/2]^2$, let $\zeta_{r;s}$ be the minimizer of $\mathcal{P}_\beta(\cdot\,;\eta_{r;s})$.*

1. *The measure $\zeta_{r;s}$ is concentrated on $[0,\bar{t}\,]$, where $\bar{t} = 9\beta^2\xi'(1)/(4+9\beta^2\xi'(1)) < 1$.*
2. *If $(r_j,s_j) \to (q_*,(1,1))$, then $\zeta_{r_j;s_j} \to \delta_{t_0}$ weakly.*

Proof. For (1), let $\zeta = \zeta_{r;s}$, and let t_1 be the largest point of the support of ζ; it is smaller than one because $\zeta \in \mathscr{M}$. We may assume $t_1 > 0$. The set S_ζ is compact by Lemma 3.15(1) and has full ζ-measure by Lemma 3.15(3). It therefore contains the support of ζ, and $t_1 \in (0,1)$ is a point at which f_ζ attains its supremum over $[0,1)$. Since f_ζ is continuously differentiable on $[0,1)$, we have $f'_\zeta(t_1) = 0$. Using (14.19) and $\hat{x}_\zeta(t) \le 1-t$, we find

$$\beta^2\eta'_{r;s}(t_1) = \int_0^{t_1} \frac{\mathrm{d}t}{\hat{x}_\zeta(t)^2} \ge \int_0^{t_1} \frac{\mathrm{d}t}{(1-t)^2} = \frac{t_1}{1-t_1}.$$

For $t \in [0,1]$, we have $\eta'_{r;s}(t) \le (9/4)\eta'_r(t) = (9/4)(1-r^2)\xi'(r^2+(1-r^2)t) \le (9/4)\xi'(1)$ by (17.11) and (17.5). Combining the two bounds, we obtain $t_1/(1-t_1) \le 9\beta^2\xi'(1)/4$, that is, $t_1 \le \bar{t}$.

For (2), let $\eta_j = \eta_{r_j;s_j}$ and $\zeta_j = \zeta_{r_j;s_j}$. The argument for (17.16), together with (17.11), shows that $\sup_{0\le t\le 1}\bigl|\eta_j(t) - \eta_{q_*}(t)\bigr| \to 0$. Part (1) implies that all ζ_j are probability measures on $[0,\bar{t}\,]$, and this set of measures is compact for weak convergence. Let ζ_∞ be the weak limit of a subsequence, and fix $Q \in (\bar{t},1)$. For each $t \in [0,Q]$, we have $\hat{x}_{\zeta_j}(t) \to \hat{x}_{\zeta_\infty}(t)$ along the subsequence, since $v \mapsto 1-\max\{t,v\}$ is bounded and continuous, and $\hat{x}_{\zeta_j}(t) \ge 1-Q$. By dominated convergence, $\int_0^Q \hat{x}_{\zeta_j}^{-1} \to \int_0^Q \hat{x}_{\zeta_\infty}^{-1}$, and the uniform convergence of η_j gives $\int \eta_j\,\mathrm{d}\zeta_j \to \int \eta_{q_*}\,\mathrm{d}\zeta_\infty$. We conclude that $\mathcal{P}_\beta(\zeta_j;\eta_j) \to \mathcal{P}_\beta(\zeta_\infty;\eta_{q_*})$ along the subsequence. On the other hand, for every $\zeta \in \mathscr{M}$, we have $\mathcal{P}_\beta(\zeta_j;\eta_j) \le \mathcal{P}_\beta(\zeta;\eta_j)$, and the right side tends to $\mathcal{P}_\beta(\zeta;\eta_{q_*})$, since only the energy term in (3.16) depends on η. It follows that ζ_∞ minimizes $\mathcal{P}_\beta(\cdot\,;\eta_{q_*})$, and $\zeta_\infty = \delta_{t_0}$ by Lemma 17.5(1). Since every subsequence has a further subsequence converging to δ_{t_0}, the whole sequence converges to δ_{t_0}. □

For $k \in \{1,2\}$, we abbreviate $a_k = a_{p_0,k}$ and note that

$$a_k(q_*)^2 = c_{p_0}\binom{p_0}{k} q_*^{2(p_0-k)}(1-q_*^2)^k > 0, \tag{17.27}$$

since $c_{p_0} > 0$, $k \le 2 \le p_0$, and $0 < q_* < 1$. Let e_1, e_2 be the standard unit vectors of $\mathbb{R}^2$.

Lemma 17.8 (Derivatives of the limiting pressure)**.** *Let $k \in \{1,2\}$. The function $\tau \mapsto \mathrm{Y}((1,1)+\tau e_k)$, defined for $|\tau| \le 1/2$, is differentiable at $\tau = 0$, and its derivative there is $\beta^2 a_k(q_*)^2(1-t_0^k)$.*

This is an instance of the envelope theorem for a family of optimization problems [118, Theorem 1 and Corollary 4], applied first to the minimum over ζ and then to the maximum over r. We give the short argument, which needs only the convergence of the optimizers and never differentiates them. It uses no nondegeneracy of the maximum at q_*.

Proof. Let $s(\tau) = (1,1)+\tau e_k$, so that $s(\tau)_k^2 - 1 = \tau(2+\tau)$. Then $\eta_{r;s(\tau)} = \eta_r + \tau(2+\tau)a_k(r)^2 t^k$ by (17.11). Since only the energy term of (3.16) depends on η, we have, for every $\zeta \in \mathscr{M}$,

$$\mathcal{P}_\beta(\zeta;\eta_{r;s(\tau)}) = \mathcal{P}_\beta(\zeta;\eta_r) + \frac{\beta^2}{2}\,\tau(2+\tau)\,a_k(r)^2\Big(1-\int t^k\,\zeta(\mathrm{d}t)\Big). \tag{17.28}$$

Let r_τ be a maximizer of $\Theta_{s(\tau)}$, which exists by Lemma 17.6(1). We first show that $r_\tau \to q_*$ as $\tau \to 0$. By (17.15), $\sup_r \left|\Theta_{s(\tau)}(r) - \Theta_{(1,1)}(r)\right| \le \omega(\tau)$ with $\omega(\tau) = \beta^2\,|\tau(2+\tau)|\,c_{p_0}\binom{p_0}{k}$. Then

$$\Theta_{(1,1)}(r_\tau) \ge \Theta_{s(\tau)}(r_\tau) - \omega(\tau) \ge \Theta_{s(\tau)}(q_*) - \omega(\tau) \ge \Theta_{(1,1)}(q_*) - 2\omega(\tau),$$

and Lemma 17.5(3) shows that $|r_\tau - q_*| < \varepsilon$ as soon as $2\omega(\tau) < \Delta_\varepsilon$.

Recall that $\mathrm{Y}(1,1) = \Theta_{(1,1)}(q_*) = \Theta^0(q_*) + \mathcal{P}_\beta(\eta_{q_*})$. For the upper bound, we test the minimum defining $\mathcal{P}_\beta(\eta_{r_\tau;s(\tau)})$ with $\zeta_{r_\tau;(1,1)}$. Using (17.28) and $\Theta^0(r_\tau) + \mathcal{P}_\beta(\eta_{r_\tau}) \le \mathrm{Y}(1,1)$, we obtain

$$\mathrm{Y}(s(\tau)) = \Theta^0(r_\tau) + \mathcal{P}_\beta(\eta_{r_\tau;s(\tau)}) \le \mathrm{Y}(1,1) + \frac{\beta^2}{2}\,\tau(2+\tau)\,a_k(r_\tau)^2\Big(1-\int t^k\,\zeta_{r_\tau;(1,1)}(\mathrm{d}t)\Big).$$

For the lower bound, we evaluate $\Theta_{s(\tau)}$ at q_* and write $\mathcal{P}_\beta(\eta_{q_*;s(\tau)}) = \mathcal{P}_\beta(\zeta_{q_*;s(\tau)};\eta_{q_*;s(\tau)})$. Using (17.28) and $\mathcal{P}_\beta(\zeta_{q_*;s(\tau)};\eta_{q_*}) \ge \mathcal{P}_\beta(\eta_{q_*})$, we obtain

$$\mathrm{Y}(s(\tau)) \ge \Theta^0(q_*) + \mathcal{P}_\beta(\eta_{q_*;s(\tau)}) \ge \mathrm{Y}(1,1) + \frac{\beta^2}{2}\,\tau(2+\tau)\,a_k(q_*)^2\Big(1-\int t^k\,\zeta_{q_*;s(\tau)}(\mathrm{d}t)\Big).$$

As $\tau \to 0$, we have $r_\tau \to q_*$ and $a_k(r_\tau) \to a_k(q_*)$, and Lemma 17.7(2) implies that $\zeta_{r_\tau;(1,1)} \to \delta_{t_0}$ and $\zeta_{q_*;s(\tau)} \to \delta_{t_0}$. Since $t \mapsto t^k$ is bounded and continuous on $[0,1]$, both integrals tend to t_0^k, and the two bounds read

$$\frac{\beta^2}{2}\,\tau(2+\tau)\big[\ell_k + o(1)\big] \le \mathrm{Y}(s(\tau)) - \mathrm{Y}(1,1) \le \frac{\beta^2}{2}\,\tau(2+\tau)\big[\ell_k + o(1)\big],$$

where $\ell_k = a_k(q_*)^2(1-t_0^k)$, with two different $o(1)$ terms. Dividing by τ, which reverses both inequalities when $\tau < 0$, and letting $\tau \to 0$, we conclude that the derivative at $\tau = 0$ exists and equals $\beta^2 a_k(q_*)^2(1-t_0^k)$. □

For $u \in S_N \setminus \{\pm x\}$, we write $r = R(x,u)$ for its latitude and y for its transverse part, as in Section 17.3, and for a second point u', we write r' and y'.

Lemma 17.9 (Gaussian integration by parts)**.** *For every N, the function Y_N is convex and differentiable on $(1/2,3/2)^2$, and for $k \in \{1,2\}$,*

$$\partial_{s_k}\mathrm{Y}_N(s) = \beta^2 s_k\,\mathbb{E}\Big[\mu_s\big(a_k(r)^2\big) - \mu_s^{\otimes 2}\big(a_k(r)a_k(r')R_n(y,y')^k\big)\Big]. \tag{17.29}$$

Proof. We use the covariance-amplitude differentiation method behind [122, Theorem 1] and [13, Proposition 2.2], retaining the latitude-dependent factors. For $k \in \{1,2\}$, let $Y_k(u) = b_{p_0,k}(n/N)^{k/2}R(x,u)^{p_0-k}G_{p_0,k}(P^\perp u)$ be the (p_0,k) term of (17.6), so that $G_s = G + (s_1-1)Y_1 + (s_2-1)Y_2$. For every u and every realization of the components, the exponent in (17.9) is an affine function of s, so $\log \mathcal{Z}_N(s)$ is convex in s by Hölder's inequality, and so is Y_N. Differentiating under the integral sign, we have $\partial_{s_k}\log\mathcal{Z}_N(s) = \beta\mu_s(Y_k)$, which is bounded by $\beta\sup_{S_N}|Y_k|$ uniformly in s. This bound is integrable by Lemma 17.4(3), and by dominated convergence, Y_N is differentiable, with $\partial_{s_k}\mathrm{Y}_N(s) = N^{-1}\beta\,\mathbb{E}\mu_s(Y_k)$.

Using (17.7), we may write $Y_k = \sum_\alpha g_\alpha v_\alpha$ as a finite sum, where the g_α are the independent standard Gaussian coordinates of the component $G_{p_0,k}$ and the v_α are deterministic polynomials. By Lemma 17.4(2), for $u,u' \in S_N \setminus \{\pm x\}$,

$$\sum_\alpha v_\alpha(u)v_\alpha(u') = \mathbb{E}Y_k(u)Y_k(u') = a_k(r)a_k(r')\,\mathbb{E}G_{p_0,k}(y)G_{p_0,k}(y') = Na_k(r)a_k(r')R_n(y,y')^k. \tag{17.30}$$

The exponent in (17.9) depends on g_α only through the term $\beta s_k g_\alpha v_\alpha(u)$, and the other components are independent of $G_{p_0,k}$. Differentiating, we find $\partial_{g_\alpha}\mu_s(v_\alpha) = \beta s_k[\mu_s(v_\alpha^2) - \mu_s(v_\alpha)^2]$, and $\mu_s(v_\alpha)$ is a bounded Lipschitz function of g_α when the other coordinates are fixed. By Gaussian integration by parts in each coordinate, $\mathbb{E}[g_\alpha F(g_\alpha)] = \mathbb{E}[F'(g_\alpha)]$, we find

$$\mathbb{E}\mu_s(Y_k) = \sum_\alpha \mathbb{E}\big[g_\alpha\mu_s(v_\alpha)\big] = \beta s_k\,\mathbb{E}\Big[\mu_s\Big(\sum_\alpha v_\alpha^2\Big) - \mu_s^{\otimes 2}\Big(\sum_\alpha v_\alpha(u)v_\alpha(u')\Big)\Big].$$

Inserting (17.30) and multiplying by β/N proves (17.29). □

At $s = (1,1)$, (17.10) identifies μ_s with w_x^H, so (17.29) expresses the derivatives of Y_N through the overlaps of two independent samples from the reweighted Gibbs measure.

Proof of Proposition 17.3. The proof has three steps. We first show that the derivatives of Y_N at $(1,1)$ converge to those of Y. We then deduce that the first two moments of the transverse overlap converge to t_0 and t_0^2, and finally we pass to the full overlap and to B_N.

Convergence of the derivatives. Let $k \in \{1,2\}$ and $0 < \tau \le 1/2$. By the convexity of Y_N in s_k (Lemma 17.9),

$$\frac{\mathrm{Y}_N(1,1) - \mathrm{Y}_N((1,1)-\tau e_k)}{\tau} \le \partial_{s_k}\mathrm{Y}_N(1,1) \le \frac{\mathrm{Y}_N((1,1)+\tau e_k) - \mathrm{Y}_N(1,1)}{\tau}.$$

We let $N \to \infty$, using Lemma 17.6(1) at the three points, and then $\tau \downarrow 0$, using Lemma 17.8. This gives

$$\lim_{N\to\infty} \partial_{s_k} \mathrm{Y}_N(1,1) = \beta^2 a_k(q_*)^2 (1 - t_0^k). \tag{17.31}$$

This is the one-dimensional convex derivative squeeze, which requires differentiability of the limit only at the point in question; see [131, Theorem 24.5], applied to the restrictions $\tau \mapsto \mathrm{Y}_N((1,1) + \tau e_k)$. In statistical mechanics it is often called Griffiths' lemma. It gives no rate of convergence.

Two moments of the transverse overlap. By (17.10) and (17.29) at $s = (1,1)$,

$$\partial_{s_k} \mathrm{Y}_N(1,1) = \beta^2 \,\mathbb{E}\Big[w_x^H\big(a_k(r)^2\big) - (w_x^H)^{\otimes 2}\big(a_k(r)a_k(r')R_n(y,y')^k\big)\Big]. \tag{17.32}$$

Recall that $|a_k| \le b_{p_0,k}$ on $[-1,1]$. For $\varepsilon > 0$, let

$$\omega_k(\varepsilon) = \sup\big\{|a_k(r) - a_k(q_*)| : \ |r - q_*| < \varepsilon\big\}.$$

On the event $\{|r - q_*| < \varepsilon,\ |r' - q_*| < \varepsilon\}$, we have $\big|a_k(r)a_k(r') - a_k(q_*)^2\big| \le 2b_{p_0,k}\omega_k(\varepsilon)$ and $\big|a_k(r)^2 - a_k(q_*)^2\big| \le 2b_{p_0,k}\omega_k(\varepsilon)$; both differences are at most $2b_{p_0,k}^2$ everywhere. Since $|R_n(y,y')| \le 1$, we obtain

$$\Big|\partial_{s_k} \mathrm{Y}_N(1,1) - \beta^2 a_k(q_*)^2\big(1 - M_{N,k}\big)\Big| \le 4\beta^2\Big(b_{p_0,k}\omega_k(\varepsilon) + 2b_{p_0,k}^2\, \mathbb{E}\, w_x^H\{u : \ |r - q_*| \ge \varepsilon\}\Big),$$

where $M_{N,k} = \mathbb{E}\,(w_x^H)^{\otimes 2}(R_n(y,y')^k)$. The last expectation tends to zero by Lemma 17.6(2), and $\omega_k(\varepsilon) \to 0$ as $\varepsilon \to 0$ since a_k is continuous. Combining this with (17.31) and (17.27), we find that $M_{N,k} \to t_0^k$ for $k = 1, 2$, and therefore

$$\mathbb{E}\,(w_x^H)^{\otimes 2}\big((R_n(y,y') - t_0)^2\big) = M_{N,2} - 2t_0 M_{N,1} + t_0^2 \longrightarrow 0. \tag{17.33}$$

The overlap and B_N. By (17.4) and (17.1), for $u, u' \in S_N \setminus \{\pm x\}$,

$$\begin{aligned} R(u,u') - q_* = \big(rr' - q_*^2\big) + \Big[\big((1-r^2)(1-r'^2)\big)^{1/2} - (1 - q_*^2)\Big] R_n(y,y') \\ + (1 - q_*^2)\big(R_n(y,y') - t_0\big). \end{aligned}$$

The first two terms are bounded by two in absolute value, and they are bounded by $\omega(\varepsilon)$ when $|r - q_*| < \varepsilon$ and $|r' - q_*| < \varepsilon$, where $\omega(\varepsilon) \to 0$ as $\varepsilon \to 0$. Using $(\alpha_1 + \alpha_2 + \alpha_3)^2 \le 3(\alpha_1^2 + \alpha_2^2 + \alpha_3^2)$, Lemma 17.6(2), and (17.33), we find

$$\begin{aligned} \mathbb{E}\,(w_x^H)^{\otimes 2}\big((R(u,u') - q_*)^2\big) \le 6\,\omega(\varepsilon)^2 + 48\,\mathbb{E}\, w_x^H\{u : \ |r - q_*| \ge \varepsilon\} \\ + 3\,\mathbb{E}\,(w_x^H)^{\otimes 2}\big((R_n(y,y') - t_0)^2\big), \end{aligned}$$

and the right side tends to $6\omega(\varepsilon)^2$ as $N \to \infty$. Since ε is arbitrary, the left side tends to zero. Finally, by (17.2) and the Cauchy–Schwarz inequality,

$$\Big(\Big|B_N^H(x)\Big|^2 - q_*\Big)^2 = \Big((w_x^H)^{\otimes 2}\big(R(u,u') - q_*\big)\Big)^2 \le (w_x^H)^{\otimes 2}\big((R(u,u') - q_*)^2\big).$$

Taking expectations proves Proposition 17.3. □

17.6 Transfer to the original disorder

Proof of Proposition 17.1. Let $\phi_N(H,x) = (\left|B_N^H(x)\right|^2 - q_*)^2$. Note that $0 \le \phi_N \le 1$, since $\left|B_N^H(x)\right|^2 \in [0,1]$ and $q_* \in (0,1)$. Proposition 17.3 and the rotation invariance explained before it imply that $\mathbb{E}_{\mathsf{P}_N}\phi_N \to 0$. Let $L_N(H) = Z_{N,\beta}(H)/\mathbb{E}Z_{N,\beta}$ be the likelihood ratio of Definition 14.1, and fix $\varepsilon, \theta > 0$. By (14.5),

$$\mathsf{Q}_N(\phi_N > \theta) \le \mathsf{Q}_N(L_N < \varepsilon) + \int_{\{\phi_N > \theta,\ L_N \ge \varepsilon\}} \frac{L_N}{\varepsilon}\,\mathrm{d}\mathsf{Q}_N \le \mathbb{P}(L_N < \varepsilon) + \frac{\mathbb{E}_{\mathsf{P}_N}\phi_N}{\varepsilon\theta}.$$

In the second step, we used that the law of H under Q_N is $\mathbb{P}$, and Markov's inequality under P_N. Recall that (ξ,β) is strictly replica symmetric with $\beta^2\xi''(0) < 1$, by $\beta < \beta_{\mathrm{s}}$ and Lemma 3.18(2). We apply Lemma 3.19 with $n = N$, $\eta = \beta^2\xi$, and $G_N = \beta H_N$, so that $W_N = L_N$. It shows that L_N converges in distribution to $e^{\sigma g - \sigma^2/2}$ with $\sigma^2 = -\log(1 - \beta^2\xi''(0))/2$, where $\sigma = 0$ when $c_2 = 0$, and its part (1) gives $\lim_{\varepsilon\downarrow 0} \limsup_{N\to\infty} \mathbb{P}(L_N < \varepsilon) = 0$. Letting $N \to \infty$ and then $\varepsilon \downarrow 0$, we obtain $\mathsf{Q}_N(\phi_N > \theta) \to 0$ for every $\theta > 0$. The order of the limits matters here: ε is fixed while $N \to \infty$.

As $0 \le \phi_N \le 1$, we have $\mathbb{E}_{\mathsf{Q}_N}\phi_N \le \theta + \mathsf{Q}_N(\phi_N > \theta)$ for every $\theta > 0$. We conclude that $\mathbb{E}_{\mathsf{Q}_N}\phi_N \to 0$. By the definition of Q_N, we have $\mathbb{E}_{\mathsf{Q}_N}\phi_N = \mathbb{E}V_N$, where

$$V_N = \int \left(|B_N(x)|^2 - q_*\right)^2 \pi(\mathrm{d}x).$$

Markov's inequality then implies that $V_N \to 0$ in probability. Finally, we have $\left|\int |B_N|^2\,\mathrm{d}\pi - q_*\right|^2 \le V_N$ by the Cauchy–Schwarz inequality, which completes the proof. □

Remark 17.10 (What the proof uses). The probability in (ii) and (iii) of Theorem 1.6 tends to one, but the proof gives no rate, for two reasons. The convergence of the derivatives in (17.31) comes from convexity and has no rate, and the transfer in Section 17.6 uses only the distributional limit of L_N. The exponential transfer of Lemma 14.3 would require an exponentially small failure probability under the planted law, which the derivative argument does not provide. The proof uses no nondegeneracy at q_*: the marginal case $A'(q_*) = 0$ is allowed, and the maximum of the profile at q_* need not be quadratic, since Lemma 17.8 compares values of the profile and never differentiates the maximizer. Finally, the proof uses no decomposition of the Gibbs measure into bands around critical points.

SECTION 18

Eigenvalues near the dynamical threshold

In this section, we prove Theorem 1.7: as β decreases to β_{d}, exponentially many eigenvalues remain below $\exp[-(1-\vartheta)C_d\varepsilon^{1+1/(2m)}N]$, where $\varepsilon = \beta^2/\beta_{\mathrm{d}}^2 - 1$, the integer m is determined by the order of contact of the function A with its maximum, and C_d is explicit. We end with two examples, Examples 18.9 and 18.10.

The proof uses the constrained-pressure results of Section 14 and the counting lemma Lemma 15.1. We first establish the inputs for infinite mixtures: the construction of the field and dynamics at C^2 regularity (Lemmas 18.2 and 18.3), pressure limits (Lemma 18.4), good reference points (Lemma 18.6), and replica concentration with an exponent uniform in the temperature (Lemma 18.5).

Notation. Throughout this section, $\xi(r) = \sum_{p\geq 2} c_p r^p$ with $c_p \geq 0$ and $\xi \neq 0$, but ξ need not be a finite mixture. We write $\pi_\beta = \pi^H_{N,\beta}$, $L_\beta = L_{\beta,H_N}$, and $\lambda_j(\beta) = \lambda_j(-L_{\beta,H_N})$, since the inverse temperature varies. We recall that

$$A(r) = \frac{(1-r)\xi'(r)}{r}, \qquad \Phi_\beta(r) = \beta^2\xi(r) + r + \log(1-r),$$

and we write $F_\beta = \beta^2\xi(1)/2$ and $\Psi_\beta(r) = F_\beta + \Phi_\beta(r)/2$ for the annealed pressure and the trial pressure (14.18) at inverse temperature β. We write $D_\beta = \{r \in (0,1) : \beta^2 A(r) > 1\}$ for the strict superlevel set, so that $q_*(\beta) = \sup D_\beta$ for $\beta_{\mathrm{d}} < \beta < \beta_{\mathrm{s}}$, and $J(\beta) = -\Phi_\beta(q_*(\beta))/2$ for the core exponent of Definition 14.10. We also write $\ell(r) = -r - \log(1-r)$, so that $\beta_{\mathrm{s}}^2 = \inf_{0<r<1} \ell(r)/\xi(r)$. Finally, $B_N = \{x \in \mathbb{R}^N : |x| \leq \sqrt{N}\}$, and ∇ and ∇^2 denote Euclidean derivatives. For an integer $P \geq 2$, we set

$$\xi_{\leq P}(r) = \sum_{p\leq P} c_p r^p, \qquad \xi_{>P} = \xi - \xi_{\leq P}, \qquad s_P = \xi'_{>P}(1), \qquad v_P = \xi_{>P}(1).$$

18.1 Infinite mixtures

We consider covariance functions satisfying

$$\sum_{p\geq 2} p^{5/2}\sqrt{c_p} < \infty, \tag{18.1}$$

and we write $S_j = \sum_{p\geq 2} p^j\sqrt{c_p}$ for $j \in \{1/2, 3/2, 5/2\}$; these sums are finite under (18.1). Since $\sum_p a_p^2 \leq (\sum_p a_p)^2$ for nonnegative numbers a_p, (18.1) implies that

$\sum_p p^5 c_p \le S_{5/2}^2 < \infty$, and the power series of ξ and of its first five derivatives converge absolutely and uniformly on $[-1,1]$. In particular, ξ is real analytic on $(-1,1)$, the numbers $\xi(1), \xi'(1), \dots, \xi^{(4)}(1)$ are finite, and $|\xi^{(k)}(v)| \le \xi^{(k)}(1)$ for $|v| \le 1$ and $k \le 4$. The function $A(r) = (1-r)\sum_p p c_p r^{p-2}$ is real analytic on $(-1,1)$ and continuous on $[0,1]$, with $A(0) = \xi''(0) = 2c_2$ and $A(1) = 0$. We define β_{d}, β_{s}, and strict replica symmetry by the formulas of Definition 3.16. If $\beta < \beta_{\mathrm{s}}$, then $\beta^2 \xi(r) < \ell(r)$ for $r \in (0,1)$, that is, $\Phi_\beta < 0$ on $(0,1)$.

For an infinite mixture, the field H_N is twice continuously differentiable but need not be smooth. We construct it in Lemma 18.2 and show in Lemma 18.3 that the generator, the Dirichlet form, the spectrum, the semigroup, and the diffusion of Definition 3.4 are defined for every field of class C^2, with the properties used in Sections 15 and 16.

Infinite-mixture inputs. The pressure and spectral arguments needed for Theorem 1.7 use the following inputs:

(I1) the regularity of the field and the construction of the dynamics (Lemmas 3.2 and 3.5);

(I2) the limit $F_N(\beta) \to F_\beta$ of the quenched free energy and the Gaussian concentration of free energies (Lemmas 3.3 and 3.18), which enter through Lemmas 14.3 and 14.6 and Proposition 14.13;

(I3) the convergence of the expected planted section pressure at each fixed latitude, in the proof of Lemma 14.6(c), which rests on the Parisi formula for finite auxiliary mixtures;

(I4) the analytic properties of ξ, and the first-variation facts of Lemma 3.15 for the auxiliary mixture η_{q_*}, used in Proposition 14.11;

(I5) replica concentration (Proposition 14.14), which was deduced from Huang et al. [93, Lemma 7.7].

We establish (I1) in Lemmas 18.2 and 18.3, (I2) and (I3) in Lemma 18.4, and (I5) in Lemma 18.5. Lemma 18.6 checks (I4) and assembles the constrained-pressure inputs. The inputs (I2) and (I3) are proved by truncating the mixture at a fixed degree P, applying the finite-mixture result to $\xi_{\le P}$, and letting $P \to \infty$ after $N \to \infty$; the tail $\xi_{>P}$ is controlled by a Gaussian comparison. We prove (I5) directly from the planted estimates, the trial bound of Lemma 14.9, and likelihood transfer, without the result of Huang, Montanari, and Pham. We state Lemma 18.5 with an exponent that is uniform over a range of temperatures, which we need for the eigenvalue bounds near the dynamical threshold in Section 18.3.

Construction of the field

We first record derivative bounds for homogeneous polynomials.

Lemma 18.1 (Derivatives of homogeneous polynomials)**.** *Let P be a homogeneous real polynomial of degree $p \geq 2$ on $\mathbb{R}^N$, and set $M = \sup_{|x|=\sqrt{N}} |P(x)|$. Then*

$$\sup_{x \in B_N} |\nabla P(x)| \leq \frac{2pM}{\sqrt{N}}, \qquad \sup_{x \in B_N} \left\| \nabla^2 P(x) \right\|_{\mathrm{op}} \leq \frac{4p^2 M}{N}. \tag{18.2}$$

Proof. Let $Q(y) = P(\sqrt{N}y)$, so that $M = \sup_{|y|=1} |Q(y)|$, $\nabla P(x) = N^{-1/2} \nabla Q(x/\sqrt{N})$, and $\nabla^2 P(x) = N^{-1} \nabla^2 Q(x/\sqrt{N})$. It suffices to show that $|\nabla Q(y)| \leq \sqrt{2}\, pM$ and $\|\nabla^2 Q(y)\|_{\mathrm{op}} \leq 2p^2 M$ for $|y| \leq 1$.

Fix a unit vector x and a unit vector v orthogonal to x, and let $f(\theta) = Q(x \cos\theta + v \sin\theta)$. Note that f is a real trigonometric polynomial of degree at most p, since each monomial of degree p in $\cos\theta$ and $\sin\theta$ is one, and that $|f| \leq M$. Applying Bernstein's inequality [128, Theorem 1.1] to f and then to f', we obtain $\sup |f'| \leq pM$ and $\sup |f''| \leq p^2 M$. Writing $\gamma(\theta) = x \cos\theta + v \sin\theta$ and using $\gamma'' = -\gamma$, we find

$$f'(0) = DQ(x)[v], \qquad f''(0) = D^2 Q(x)[v, v] - DQ(x)[x].$$

We have $DQ(y)[y] = pQ(y)$ for all y by Euler's identity. Differentiating this identity in a direction w, we find that $D^2 Q(y)[y, w] = (p-1) DQ(y)[w]$. Taking $y = x$, we obtain

$$\begin{aligned} &|DQ(x)[v]| \leq pM, && DQ(x)[x] = pQ(x), \\ &D^2 Q(x)[x, x] = p(p-1)Q(x), && D^2 Q(x)[x, v] = (p-1) DQ(x)[v], \end{aligned} \tag{18.3}$$

and $|D^2 Q(x)[v, v]| \leq |f''(0)| + p|Q(x)| \leq (p^2 + p)M$.

The gradient $\nabla Q(x)$ splits into its component along x, which equals $pQ(x)$, and its tangential component, whose norm is $\sup_v DQ(x)[v] \leq pM$. This implies that $|\nabla Q(x)|^2 \leq 2p^2 M^2$. For the Hessian, we write a unit vector as $z = \alpha x + \alpha' v$, with $v \perp x$ a unit vector and $\alpha^2 + \alpha'^2 = 1$. By (18.3),

$$\begin{aligned} |D^2 Q(x)[z, z]| &\leq M\big[p(p-1)\alpha^2 + 2p(p-1)|\alpha\alpha'| + (p^2 + p)\alpha'^2\big] \\ &\leq M\big[p(p-1) + p(p-1)\alpha^2 + p(p+1)\alpha'^2\big] \leq 2p^2 M, \end{aligned}$$

where we used $2|\alpha\alpha'| \leq 1$ in the second line. As the matrix $\nabla^2 Q(x)$ is symmetric, its operator norm is at most $2p^2 M$.

Finally, for $0 < \varrho \leq 1$ and $|x| = 1$, we have $\nabla Q(\varrho x) = \varrho^{p-1} \nabla Q(x)$ and $\nabla^2 Q(\varrho x) = \varrho^{p-2} \nabla^2 Q(x)$ by homogeneity, and the bounds extend to $y = 0$ by continuity. As $\sqrt{2}\, p \leq 2p$ and $2p^2 \leq 4p^2$, this proves (18.2). □

The proof gives the constants $\sqrt{2}\, p$ and $2p^2$ in place of $2p$ and $4p^2$. The tangential bound alone does not give the gradient bound pM, since the radial component $pQ(x)$ must also be controlled. Szegő's strengthening of Bernstein's inequality,

$f'(\theta)^2 + p^2 f(\theta)^2 \le p^2 M^2$ [128, Theorem 3.7], controls both components at once and gives $|\nabla Q| \le pM$ on the unit ball; see [45, (32) and Theorem 4.3]. We do not need these sharper constants.

Let $(g^{(p)}_{i_1 \cdots i_p})$, for $p \ge 2$ and $i_1, \dots, i_p \in \{1, \dots, N\}$, be independent standard Gaussian variables, and define the homogeneous polynomials

$$H_{N,p}(x) = N^{-(p-1)/2} \sum_{i_1,\dots,i_p=1}^{N} g^{(p)}_{i_1 \cdots i_p} x_{i_1} \cdots x_{i_p}, \qquad x \in \mathbb{R}^N. \tag{18.4}$$

Then $\mathbb{E} H_{N,p}(x) H_{N,p}(y) = N^{-(p-1)} (x \cdot y)^p$, which equals $N R(x,y)^p$ for $x, y \in S_N$. Once H_N is defined, we write $H_{N,\le P} = \sum_{p \le P} \sqrt{c_p} H_{N,p}$ and $H_{N,>P} = H_N - H_{N,\le P}$.

Lemma 18.2 (Construction and regularity of the field)**.** *Let ξ satisfy* (18.1).

1. *For every N and p, we have $\mathbb{E} \sup_{S_N} |H_{N,p}| \le 2N\sqrt{p}$. Almost surely, the series $H_N = \sum_{p \ge 2} \sqrt{c_p} H_{N,p}$ converges in $C^2(B_N)$ for every N. The limit is a centered Gaussian field with $\mathbb{E} H_N(x) H_N(y) = N \xi(x \cdot y / N)$ for $x, y \in B_N$, and $H_N \circ g^{-1}$ has the same law as H_N for every $g \in O(N)$.*
2. *There exist $K, c_F > 0$, depending only on ξ, such that for every $N \ge 1$, with probability at least $1 - 4e^{-c_F N}$,*

$$\sup_{S_N} |H_N| \le KN, \qquad \sup_{B_N} |\nabla H_N| \le K\sqrt{N}, \qquad \sup_{B_N} \left\| \nabla^2 H_N \right\|_{\mathrm{op}} \le K. \tag{18.5}$$

   On this event, $\sup_{S_N} |\nabla_{S_N} H_N| \le K\sqrt{N}$ and $\sup_{S_N} \left\| \nabla^2_{S_N} H_N \right\|_{\mathrm{op}} \le K$. Moreover, $\mathbb{E}(\sup_{S_N} |H_N|)^2 \le CN^2$ for a constant C depending only on ξ.
3. *For every $P \ge 2$ and $N \ge 1$, we have $\mathbb{E} \sup_{S_N} |H_{N,>P}| \le 2N\sqrt{s_P}$. For every $h > 0$, there exists P_0 such that for all $P \ge P_0$ and $N \ge 1$,*

$$\mathbb{P}\Big(\sup_{S_N} |H_{N,>P}| > hN \Big) \le 2 \exp\Big(-\frac{h^2 N}{8 v_P} \Big), \tag{18.6}$$

   where the right side is zero if $v_P = 0$.

Proof. We use the Sudakov–Fernique inequality [3, Theorem 2.2.3]: if X and Y are centered, almost surely bounded Gaussian processes on a set T with $\mathbb{E}(X_s - X_t)^2 \le \mathbb{E}(Y_s - Y_t)^2$ for all $s, t \in T$, then $\mathbb{E} \sup_T X \le \mathbb{E} \sup_T Y$. We also use the Borell–TIS inequality [3, Theorem 2.1.1]: $\mathbb{P}(\sup_T X > \mathbb{E} \sup_T X + t) \le e^{-t^2/(2\sigma_T^2)}$ for $t > 0$, where $\sigma_T^2 = \sup_{t \in T} \mathbb{E} X_t^2$.

A comparison bound. Let X be a centered continuous Gaussian field on S_N with covariance $N\theta(R(x,y))$, where $\theta(r) = \sum_p b_p r^p$ has nonnegative coefficients and $\theta'(1) < \infty$. Extend X by $X(0) = 0$. For $r \in [-1,1]$, we have $1 - r^p = (1-r)\sum_{k<p} r^k \le p(1-r)$, which implies that $\theta(1) - \theta(r) \le \theta'(1)(1-r)$. For $x, y \in S_N$, we obtain

$$\mathbb{E}(X(x) - X(y))^2 = 2N(\theta(1) - \theta(R(x,y))) \le 2N\theta'(1)(1 - R(x,y)) = \theta'(1)|x-y|^2.$$

Further, $\mathbb{E}X(x)^2 = N\theta(1) \le N\theta'(1)$. Let $Y(x) = \sqrt{\theta'(1)}\, g \cdot x$ with g a standard Gaussian vector in $\mathbb{R}^N$. Then $\mathbb{E}(Y(x) - Y(y))^2 = \theta'(1)|x-y|^2$ and $\mathbb{E}Y(x)^2 = N\theta'(1)$. Thus the hypothesis of the Sudakov–Fernique inequality holds on $T = S_N \cup \{0\}$, and since $X(0) = Y(0) = 0$, we have

$$\mathbb{E}\max\Big(0, \sup_{S_N} X\Big) \le \mathbb{E}\sup_T Y = \sqrt{\theta'(1)N}\,\mathbb{E}|g| \le N\sqrt{\theta'(1)}. \tag{18.7}$$

The same bound holds for $-X$. Using $\sup|X| \le \max(0, \sup X) + \max(0, \sup(-X))$, we conclude that $\mathbb{E}\sup_{S_N}|X| \le 2N\sqrt{\theta'(1)}$.

Part (1). The first claim is the comparison bound with $\theta(r) = r^p$. Let $M_p = \sup_{S_N}|H_{N,p}|$. By homogeneity, we have $\sup_{B_N}|H_{N,p}| = M_p$, and Lemma 18.1 implies that $\sup_{B_N}|\nabla H_{N,p}| \le 2pM_p/\sqrt{N}$ and $\sup_{B_N}\|\nabla^2 H_{N,p}\|_{\mathrm{op}} \le 4p^2 M_p/N$. Using $\mathbb{E}M_p \le 2N\sqrt{p}$, we find

$$\begin{aligned}\sum_{p\ge 2}\sqrt{c_p}\,\mathbb{E}\Big[\sup_{B_N}|H_{N,p}| + \sup_{B_N}|\nabla H_{N,p}| + \sup_{B_N}\Big\|\nabla^2 H_{N,p}\Big\|_{\mathrm{op}}\Big] \\ \le \sum_{p\ge 2}\sqrt{c_p}\,2N\sqrt{p}\,(1+2p+4p^2) \le 14NS_{5/2}.\end{aligned}$$

By Tonelli's theorem, the corresponding random series is finite almost surely, and the partial sums of $\sum_p \sqrt{c_p}H_{N,p}$ converge in $C^2(B_N)$. As there are countably many N, this holds for all N simultaneously. The limit is a centered Gaussian field, and by independence its covariance is $\sum_p c_p N^{-(p-1)}(x\cdot y)^p = N\xi(x\cdot y/N)$. The covariance depends only on $x\cdot y$, and a continuous centered Gaussian field is determined in law by its covariance, which gives the invariance under $O(N)$.

Part (2). By the comparison bound with $\theta = \xi$, we have $\mathbb{E}\sup_{S_N}(\pm H_N) \le N\sqrt{\xi'(1)}$. The Borell–TIS inequality with $\sigma_T^2 = N\xi(1)$ and $t = N\sqrt{\xi(1)}$ then gives

$$\mathbb{P}\Big(\sup_{S_N}|H_N| > K_0 N\Big) \le 2e^{-N/2}, \qquad K_0 = \sqrt{\xi'(1)} + \sqrt{\xi(1)}.$$

Integrating the same inequality over t, we also obtain $\mathbb{E}(\sup_{S_N}|H_N|)^2 \le CN^2$. For the gradient, consider the process $(x,v) \mapsto v\cdot\nabla H_N(x)$ on $B_N \times S^{N-1}$, whose supremum is

$\sup_{B_N} |\nabla H_N|$. By Part (1), its expected supremum is at most $\sum_p \sqrt{c_p}(2p/\sqrt{N})\, 2N\sqrt{p} = 4\sqrt{N}S_{3/2}$. Differentiating the covariance $N\xi(x\cdot y/N)$ once in x and once in y, in the direction v, and setting $y = x$, we find

$$\mathbb{E}(v\cdot\nabla H_N(x))^2 = \xi''(|x|^2/N)\frac{(x\cdot v)^2}{N} + \xi'(|x|^2/N) \le \xi''(1) + \xi'(1),$$

where we used $(x\cdot v)^2 \le |x|^2 \le N$ and the monotonicity of ξ' and ξ'' on $[0,1]$. The Borell–TIS inequality with $t = (N(\xi'(1)+\xi''(1)))^{1/2}$ gives

$$\mathbb{P}\Big(\sup_{B_N} |\nabla H_N| > K_1\sqrt{N}\Big) \le e^{-N/2}, \qquad K_1 = 4S_{3/2} + \sqrt{\xi'(1)+\xi''(1)}.$$

For the Hessian, consider the process $(x,v,w) \mapsto D^2H_N(x)[v,w]$ on $B_N \times S^{N-1} \times S^{N-1}$, whose supremum is $\sup_{B_N} \|\nabla^2 H_N\|_{\mathrm{op}}$. Its expected supremum is at most $\sum_p \sqrt{c_p}(4p^2/N)\, 2N\sqrt{p} = 8S_{5/2}$. To bound its variance, we differentiate $N^{-(p-1)}(x\cdot y)^p$ twice in x, in the directions v and w, and twice in y, in the same directions, and set $y = x$. With $|x|^2 = \varrho^2 N \le N$, $a_1 = x\cdot v$, $a_2 = x\cdot w$, and $a_3 = v\cdot w$, this gives

$$\frac{p(p-1)}{N^{p-1}}\Big[(p-2)(p-3)(\varrho^2N)^{p-4}a_1^2a_2^2 + (p-2)(\varrho^2N)^{p-3}(a_1^2+a_2^2+2a_1a_2a_3) + (\varrho^2N)^{p-2}(1+a_3^2)\Big],$$

where terms with a vanishing coefficient are omitted. Using $|a_1|, |a_2| \le \varrho\sqrt{N}$, $|a_3| \le 1$, and $\varrho \le 1$, we bound this by $p(p-1)[(p-2)(p-3)+4(p-2)+2]/N$. Summing over p with weights c_p, we find that the variance is at most D_2/N, where $D_2 = \xi^{(4)}(1) + 4\xi'''(1) + 2\xi''(1)$. The Borell–TIS inequality with $t = 1$ gives

$$\mathbb{P}\Big(\sup_{B_N} \big\|\nabla^2 H_N\big\|_{\mathrm{op}} > K_2\Big) \le e^{-N/(2D_2)}, \qquad K_2 = 8S_{5/2} + 1.$$

We set $K = K_0 + K_1 + K_2$ and $c_F = \min\{1/2, 1/(2D_2)\}$. On the intersection of the three events, (18.5) holds, and $|\nabla_{S_N} H_N| \le |\nabla H_N|$. For tangent vectors v, w at $x \in S_N$, we have $\nabla^2_{S_N} H_N(x)[v,w] = \nabla^2 H_N(x)[v,w] - N^{-1}(x\cdot\nabla H_N(x))\, v\cdot w$, as in the proof of Lemma 3.2. On the same event, this gives $\big\|\nabla^2_{S_N} H_N\big\|_{\mathrm{op}} \le K_2 + K_1 \le K$.

Part (3). The field $H_{N,>P}$ has covariance $N\xi_{>P}(R(x,y))$ on S_N. The comparison bound with $\theta = \xi_{>P}$ gives $\mathbb{E}\max(0, \sup_{S_N}(\pm H_{N,>P})) \le N\sqrt{s_P}$ and $\mathbb{E}\sup_{S_N}|H_{N,>P}| \le 2N\sqrt{s_P}$. Note that $s_P \to 0$, because $\xi'(1) < \infty$, and let P_0 be such that $\sqrt{s_P} \le h/2$ for $P \ge P_0$. The Borell–TIS inequality with $\sigma_T^2 = Nv_P$ and $t = hN/2$, applied to $H_{N,>P}$ and to $-H_{N,>P}$, gives (18.6). □

Lemma 18.2 implies that the bounds of Lemma 3.2(3) on H_N and its first two derivatives hold for every ξ satisfying (18.1), with $1 - 4e^{-c_F N}$ in place of $1 - e^{-cN}$. The bound (14.3) holds with $c_0 = c_F$ and with the same constant K. Indeed, let $\sigma_1^2 = \xi'(1) + \xi''(1) = \sum_p p^2 c_p$. Then $\sqrt{\xi(1)} \le \sum_p \sqrt{c_p} \le S_{5/2}$ and $\sigma_1 \le \sum_p p\sqrt{c_p} \le S_{5/2}$, so the definition of K in the proof of Lemma 18.2 gives

$$K - \sqrt{\xi'(1)} \ge 2\sqrt{\xi(1)}, \qquad K - 4S_{3/2} \ge 2\sigma_1.$$

The Borell–TIS inequality at the thresholds KN for $\pm H_N$ and $K\sqrt{N}$ for the gradient process, applied as in that proof, bounds the failure probability in (14.3) by $3e^{-2N}$, which is at most $e^{-c_F N}$ for $N \ge 1$, since $c_F \le 1/2$. The bound on the third derivative in Lemma 3.2(3) is not used in Sections 14 to 16.

Lemma 18.3 (Langevin dynamics for fields of class C^2)**.** *Let $\beta \ge 0$ and let H be a function of class C^2 on S_N. Then the generator, the Dirichlet form, the semigroup, and the diffusion of Definition 3.4 are defined, and the conclusions of Lemmas 3.5 and 3.6 hold, with the following changes: the eigenfunctions are of class C^2 but need not be smooth, the operator $-L$ is essentially self-adjoint on $C^2(S_N)$ instead of on smooth functions, the transition density $p_t(x,y)$ is continuous on $(0,\infty) \times S_N \times S_N$ instead of smooth, and in the min–max principle of Lemma 3.6(5), the minimum is attained at a subspace of $C^2(S_N)$, so that V may range over subspaces of $C^2(S_N)$ instead of subspaces of smooth functions. If V ranges over subspaces of smooth functions, the formula of Lemma 3.6(5) holds with the minimum replaced by an infimum. In Lemma 3.5(5), the drift is extended to a compactly supported vector field of class C^1 instead of a smooth one.*

Proof. By Remark 3.7, Lemma 3.5(1)–(3) hold for H, except for the regularity of the eigenfunctions and the essential self-adjointness. Further, Lemma 3.6(1), (2), and (4) hold, and so does the min–max principle of Lemma 3.6(5), with V ranging over the subspaces of the form domain $\mathcal{D}(\mathcal{E}) = H^1(S_N)$. Integration by parts gives $Af = -Lf$ for $f \in C^2(S_N)$. In this proof, ∇ and Δ are taken along S_N. We prove the remaining assertions in four steps. The classical inputs are those of the proof of Lemma 3.5, together with the Kato–Rellich theorem, the Duhamel formula, and the Trotter product formula.

A unitary transformation. Let $\rho = \mathrm{d}\pi / \mathrm{d}\nu_N$, which is proportional to $e^{\beta H}$, and let $\omega = \rho^{1/2}$. Both are of class C^2 and bounded above and below by positive constants. The map $Uf = \omega f$ is unitary from $L^2(\pi)$ to $L^2(\nu_N)$, and it maps $C^2(S_N)$ onto itself. Since $L(g/\omega) = \omega^{-2} \operatorname{div}(\omega^2 \nabla(g/\omega)) = \omega^{-1}\Delta g - \omega^{-2} g \Delta\omega$ for $g \in C^2(S_N)$, we have

$$U(-L)U^{-1} g = -\Delta g + Wg, \qquad W = \frac{\Delta\omega}{\omega} = \frac{\beta}{2}\Delta H + \frac{\beta^2}{4}|\nabla H|^2,$$

where we used $\nabla\omega = \frac{\beta}{2}\omega\nabla H$. The potential W is continuous and bounded. The Laplace–Beltrami operator is essentially self-adjoint on $C^\infty(S_N)$, with domain $H^2(S_N)$, and the

Kato–Rellich theorem shows that $-\Delta + W$ is self-adjoint on $H^2(S_N)$, with core $C^\infty(S_N)$. The operator $U^{-1}(-\Delta + W)U$ is therefore self-adjoint, with core $U^{-1}C^\infty(S_N) \subset C^2(S_N)$, and on this core it agrees with $-L$, hence with A. Since A is closed, it extends the closure $U^{-1}(-\Delta + W)U$ of this restriction, and since a self-adjoint operator has no proper self-adjoint extension, we obtain $A = U^{-1}(-\Delta + W)U$. In particular, $C^2(S_N)$ is a core for A, which is the essential self-adjointness of $-L$ on $C^2(S_N)$.

The transition kernel. Let $k_t^0(x,y)$ be the heat kernel of Δ with respect to ν_N. It is smooth, nonnegative, and symmetric, by Lemma 3.5(4) with $H = 0$. By the Duhamel formula, $e^{-t(-\Delta+W)} = \sum_{n\ge 0} T_t^{(n)}$, where $T_t^{(n)}$ has the kernel

$$k_t^{(n)}(x,y) = (-1)^n \int_{0<s_1<\cdots<s_n<t} \int k_{s_1}^0(x,z_1) W(z_1) k_{s_2-s_1}^0(z_1,z_2) \cdots \\ \cdots W(z_n) k_{t-s_n}^0(z_n,y)\, \mathrm{d}z\, \mathrm{d}s,$$

with $\mathrm{d}z = \nu_N(\mathrm{d}z_1)\cdots\nu_N(\mathrm{d}z_n)$ and $\mathrm{d}s = \mathrm{d}s_1 \cdots \mathrm{d}s_n$. Since $k^0 \ge 0$, the Chapman–Kolmogorov equations for k^0 give

$$|k_t^{(n)}(x,y)| \le \frac{t^n \, \|W\|_\infty^n}{n!} \, k_t^0(x,y). \tag{18.8}$$

We claim that each $k_t^{(n)}(x,y)$ is continuous in (t,x,y) for $t > 0$. Indeed, the integral over the part of the simplex where all the increments $s_1, s_2 - s_1, \dots, t - s_n$ exceed $\eta > 0$ is continuous, by the smoothness of k^0 on $[\eta,\infty) \times S_N \times S_N$. By the argument that gives (18.8), the remaining part is at most $\|W\|_\infty^n$ times the measure of the excluded part of the simplex times $k_t^0(x,y)$, which tends to zero as $\eta \downarrow 0$, locally uniformly in $t > 0$. The series $k_t = \sum_n k_t^{(n)}$ converges uniformly on $[t_0,t_1] \times S_N \times S_N$ for $0 < t_0 < t_1$ by (18.8), and k_t is a continuous kernel for $e^{-t(-\Delta+W)}$. It is symmetric, because this operator is self-adjoint, and nonnegative, because by the Trotter product formula, $e^{-t(-\Delta+W)}$ is the strong limit of $(e^{t\Delta/m}e^{-tW/m})^m$, which preserves nonnegative functions. We set

$$p_t(x,y) = \frac{k_t(x,y)}{\omega(x)\omega(y)}.$$

Since $e^{-tA} = U^{-1}e^{-t(-\Delta+W)}U$ and $\pi(\mathrm{d}y) = \omega(y)^2\nu_N(\mathrm{d}y)$, we have $e^{-tA}f(x) = \int p_t(x,y)f(y)\,\pi(\mathrm{d}y)$ for $f \in L^2(\pi)$. The kernel p_t is continuous, nonnegative, and symmetric, and $\int p_t(x,y)\,\pi(\mathrm{d}y) = 1$ because $A1 = 0$. The kernels $P_t(x,\mathrm{d}y) = p_t(x,y)\,\pi(\mathrm{d}y)$ therefore form a Markov semigroup that maps continuous functions to continuous functions, and π is reversible and invariant. This proves Lemma 3.5(4), with a continuous density. Further, (18.8) implies that $|k_t - k_t^0| \le (e^{t\|W\|_\infty} - 1)k_t^0$, and $e^{t\Delta}g \to g$ uniformly for continuous g. It follows that $e^{-tA}f \to f$ uniformly as $t \downarrow 0$, for every continuous f.

Regularity of the eigenfunctions. Let ϕ be an eigenfunction of A with eigenvalue λ. Then $\phi \in H^1(S_N)$ satisfies $\Delta\phi = -\lambda\phi - \beta\nabla H \cdot \nabla\phi$ weakly, and the right side lies in L^p whenever $\nabla\phi$ does. Iterating the L^p estimates for the Laplace–Beltrami operator and the Sobolev embedding, we find that $\phi \in W^{2,p}(S_N)$ for every $p < \infty$, so $\nabla\phi$ is Hölder continuous. Since ∇H is of class C^1, the right side is then Hölder continuous, and the Schauder estimates give $\phi \in C^{2,\alpha}(S_N)$ for some $\alpha > 0$. The eigenfunctions are of class C^2, and $L\phi = -\lambda\phi$ holds pointwise. In particular, for every $k \geq 0$, the span of $\phi_0, \dots, \phi_k$, at which the minimum in Lemma 3.6(5) is attained, is a subspace of $C^2(S_N)$.

It remains to show that the infimum of the maximal Rayleigh quotient over the $(k+1)$-dimensional subspaces of smooth functions equals λ_k. These subspaces lie in $\mathcal{D}(\mathcal{E})$, so the infimum is at least λ_k. Since smooth functions are dense in $H^1(S_N)$, there exist smooth functions $\psi_0^n, \dots, \psi_k^n$ such that $\psi_i^n \to \phi_i$ in $H^1(S_N)$ as $n \to \infty$, for each $i \leq k$. Since the inner product of $L^2(\pi)$ and the form $\mathcal{E}$ are continuous on $H^1(S_N)$, the matrices $G_n = (\langle \psi_i^n, \psi_l^n \rangle)_{i,l \leq k}$ and $J_n = (\mathcal{E}(\psi_i^n, \psi_l^n))_{i,l \leq k}$ converge to the identity and to the diagonal matrix with entries $\lambda_0, \dots, \lambda_k$, respectively. For large n, the matrix G_n is invertible, and the maximum of the Rayleigh quotient over the span of $\psi_0^n, \dots, \psi_k^n$ is the largest eigenvalue of $G_n^{-1/2} J_n G_n^{-1/2}$, which tends to λ_k. This proves the assertion on smooth subspaces.

The diffusion. We extend the tangent vector field $b = \beta\nabla H$ to $\mathbb{R}^N$ by $\tilde{b}(y) = \psi(y) b(\sqrt{N}y/|y|)$, where ψ is smooth, supported in $\{\sqrt{N}/2 < |y| < 2\sqrt{N}\}$, and equal to one near S_N. Then $\tilde{b}$ is C^1 with compact support, hence Lipschitz, and $y \cdot \tilde{b}(y) = 0$ for all y. The proof of Lemma 3.5(5) applies to (3.6) with this drift. The equation has a unique strong solution for every $x \in S_N$. The term $2y \cdot \tilde{b}(y)$ in the drift of $|X_t|^2 - N$ vanishes identically, the remaining terms are those treated in that proof, and the solution stays on S_N. Further, Itô's formula (3.7) holds for every F of class C^2. Applying it to $e^{\lambda_j s}\phi_j(X_s)$, we see that this process is a martingale, and therefore $\mathbb{E}_x \phi_j(X_t) = e^{-\lambda_j t}\phi_j(x) = \int p_t(x,y)\phi_j(y)\,\pi(\mathrm{d}y)$ for every x. Finite linear combinations of eigenfunctions are uniformly dense in the continuous functions. Indeed, for continuous f and $t > 0$, we have $e^{-tA} f = e^{-tA/2}(e^{-tA/2} f)$; the operator $e^{-tA/2}$ maps $L^2(\pi)$ boundedly into the continuous functions, since $\sup_x \int p_{t/2}(x,y)^2\,\pi(\mathrm{d}y) < \infty$; so the eigenfunction expansion of $e^{-tA} f$ converges uniformly, and $e^{-tA} f \to f$ uniformly as $t \downarrow 0$. Hence the law of X_t under $\mathbb{P}_x$ and $P_t(x,\cdot)$ agree on a uniformly dense set of continuous functions, and they coincide. This proves Lemma 3.5(5). Finally, the proof of Lemma 3.6(3) uses only Lemma 3.5(4) and (5). □

The proofs in Sections 15 and 16 use the dynamics only through Lemmas 3.5 and 3.6, and they do not use the smoothness of eigenfunctions or of the transition density beyond the measurability of $x \mapsto P_t(x,E)$, which holds for continuous kernels. The same proofs apply to every field of class C^2.

Pressure inputs

We use the planted law P_N, the Gibbs-reference law Q_N, and the likelihood ratio $L_N = Z_N e^{-NF_\beta}$ of Definition 14.1, the section partition function $\mathcal{Z}_N$ and the section pressure f_N of Section 14.2, and the auxiliary mixture η_r of (14.10). They are defined by the same formulas for every ξ satisfying (18.1). For $r \in (-1,1)$, the function η_r is now a power series in t; its coefficients $(1-r^2)^k \xi^{(k)}(r^2)/k!$ are nonnegative, and $\eta_r(1) = \xi(1) - \xi(r^2) < \infty$. The coarea formula (14.11) holds for every field, and $C_N \le \sqrt{N}$, by the log-convexity of Γ, which gives $\Gamma(x+1/2) \le \sqrt{x}\,\Gamma(x)$.

The Crisanti–Sommers functional for general covariances. Let $\eta(t) = \sum_{k\ge 1} a_k t^k$ with $a_k \ge 0$ and $\eta(1) < \infty$. For $\zeta \in \mathscr{M}$, we define $\mathcal{P}_\beta(\zeta;\eta)$ by the formula (3.16), and we set

$$\mathcal{P}_\beta(\eta) = \inf_{\zeta \in \mathscr{M}} \mathcal{P}_\beta(\zeta;\eta).$$

For covariance polynomials, this is the definition in Definition 3.13. For two such functions η and θ, the difference $\mathcal{P}_\beta(\zeta;\eta) - \mathcal{P}_\beta(\zeta;\theta) = \frac{\beta^2}{2}[(\eta-\theta)(1) - \int (\eta-\theta)\,\mathrm{d}\zeta]$ is bounded in absolute value by $\beta^2 \sup_{[0,1]} |\eta - \theta|$. If moreover $\eta - \theta$ has nonnegative coefficients, then $\eta - \theta$ is nonnegative and nondecreasing on $[0,1]$, and the difference lies in $[0, \frac{\beta^2}{2}(\eta-\theta)(1)]$. Taking infima over ζ, we obtain

$$|\mathcal{P}_\beta(\eta) - \mathcal{P}_\beta(\theta)| \le \beta^2 \sup_{[0,1]} |\eta - \theta|, \tag{18.9}$$

and, if $\eta - \theta$ has nonnegative coefficients,

$$0 \le \mathcal{P}_\beta(\eta) - \mathcal{P}_\beta(\theta) \le \frac{\beta^2}{2}(\eta-\theta)(1). \tag{18.10}$$

We define $f(r)$ by (14.12) with this $\mathcal{P}_\beta$.

Lemma 18.4 (Pressure inputs for infinite mixtures)**.** *Let ξ satisfy* (18.1)*, and let $\beta > 0$.*

1. *Let μ be a finite measure on S_N with $\mu(S_N) > 0$, and let h be a bounded Borel function on S_N, both deterministic. Then $Y = \log \int e^{\beta H_N + h}\,\mathrm{d}\mu$ is integrable and, for every $s > 0$,*

$$\mathbb{P}\big(|Y - \mathbb{E}Y| \ge s\big) \le 2\exp\Big(-\frac{s^2}{2\beta^2 N \xi(1)}\Big). \tag{18.11}$$

   In particular, Lemma 3.3 holds for ξ.
2. *If $\beta < \beta_{\mathrm{s}}$, then $F_N(\beta) \to F_\beta$, the bound* (3.21) *holds for every $\varepsilon > 0$ and all sufficiently large N, and Lemmas 14.2 and 14.3 hold for ξ.*
3. *For every $r \in (-1,1)$ and $u \in S_N$, we have $\mathbb{E}_{\mathsf{P}_N(\cdot|u)} f_N(r;H,u) \to f(r)$ as $N \to \infty$.*

4. *Lemma 14.6(b) and (c) hold for ξ, with the constants K and c_0 of* (14.3) *replaced by the constants K and c_F of Lemma 18.2.*

Proof. Throughout, P is at least the smallest degree p with $c_p > 0$, so that $\xi_{\le P} \neq 0$. By Lemma 18.2(3), we have

$$\mathbb{E} \sup_{S_N} |H_{N,>P}| \le 2N\sqrt{s_P}, \qquad s_P \to 0 \quad (P \to \infty). \tag{18.12}$$

Part (1). Let Y_P be defined as Y with H_N replaced by the polynomial field $H_{N,\le P}$. The proof of Lemma 3.3, which applies to any nonzero finite measure μ (as noted at the beginning of Section 14), shows that Y_P is a Lipschitz function of the finitely many Gaussian coordinates of $H_{N,\le P}$ with constant $\beta(N\xi_{\le P}(1))^{1/2} \le \beta(N\xi(1))^{1/2}$, and (18.11) holds for Y_P. We have $|Y - Y_P| \le \beta \sup_{S_N} |H_{N,>P}|$, which tends to zero in L^1 by (18.12) and almost surely by Lemma 18.2(1). For $0 < s' < s$, we have $\mathbb{P}(|Y - \mathbb{E}Y| \ge s) \le \mathbb{P}(|Y_P - \mathbb{E}Y_P| \ge s') + \mathbb{P}(|Y - Y_P| + |\mathbb{E}Y - \mathbb{E}Y_P| \ge s - s')$, and letting $P \to \infty$ and then $s' \uparrow s$ gives (18.11). Integrability follows from $|Y - \log \mu(S_N)| \le \|h\|_\infty + \beta \sup_{S_N} |H_N|$.

Part (2). Since $\xi_{\le P} \le \xi$ on $(0,1)$, the definition of β_{s} gives $\beta_{\mathrm{s}}(\xi_{\le P}) \ge \beta_{\mathrm{s}}(\xi) > \beta$. By Lemma 3.18(1) and Theorem 3.14, applied to the finite mixture $\xi_{\le P}$, we have $N^{-1}\mathbb{E} \log Z_{N,\beta}(H_{N,\le P}) \to \beta^2 \xi_{\le P}(1)/2$, where $Z_{N,\beta}(H_{N,\le P})$ is the partition function of the truncated field. We also have $|F_N(\beta) - N^{-1}\mathbb{E} \log Z_{N,\beta}(H_{N,\le P})| \le 2\beta\sqrt{s_P}$ by (18.12), and $F_\beta - \beta^2 \xi_{\le P}(1)/2 = \beta^2 v_P/2$. Combining these facts, we obtain

$$\limsup_{N\to\infty} |F_N(\beta) - F_\beta| \le 2\beta\sqrt{s_P} + \tfrac{1}{2}\beta^2 v_P,$$

and letting $P \to \infty$ gives $F_N(\beta) \to F_\beta$. This limit and Part (1), for $\mu = \nu_N$ and $h = 0$, imply (3.21), as in the proof of Lemma 3.18(3). For Lemma 14.2, note that $\mathbb{E}e^{\beta H_N(u)} = e^{NF_\beta}$, and that under the probability measure with density $e^{\beta H_N(u) - NF_\beta}$, the field $H_N - \beta N \xi(R(\cdot, u))$ is a centered Gaussian field with the covariance of H_N; this follows from the finite-dimensional computation in the proof of Lemma 14.2, applied to the Gaussian vector $(H_N(x_1), \dots, H_N(x_k), H_N(u))$ for arbitrary points $x_1, \dots, x_k$. The rest of the proof of Lemma 14.2 is unchanged. The proof of Lemma 14.3 uses only (3.21).

Part (3). Fix $r \in (-1, 1)$ and $u \in S_N$. Recall from (14.4) that under $\mathsf{P}_N(\cdot \mid u)$, the field is $H = \widetilde{H} + \beta N \xi(R(\cdot, u))$. We write $H = H^{(P)} + \widetilde{H}_{>P} + \beta N \xi_{>P}(R(\cdot, u))$, where $\widetilde{H} = \widetilde{H}_{\le P} + \widetilde{H}_{>P}$ is the decomposition of Lemma 18.2 and

$$H^{(P)} = \widetilde{H}_{\le P} + \beta N \xi_{\le P}(R(\cdot, u)).$$

By (14.4) for the finite mixture $\xi_{\le P}$, the field $H^{(P)}$ has the conditional law, given u, of the field under the planted law of $\xi_{\le P}$ at inverse temperature β. On the latitude $\{R(x, u) = r\}$, the difference $H - H^{(P)}$ equals $\widetilde{H}_{>P} + \beta N \xi_{>P}(r)$. Using $|\xi_{>P}(r)| \le v_P$, we find

$$|f_N(r; H, u) - f_N(r; H^{(P)}, u)| \le \frac{\beta}{N} \sup_{S_N} |\widetilde{H}_{>P}| + \beta^2 v_P,$$

and taking expectations and using (18.12), we obtain $|\mathbb{E} f_N(r; H, u) - \mathbb{E} f_N(r; H^{(P)}, u)| \le 2\beta\sqrt{s_P} + \beta^2 v_P$. By the convergence of the means in the proof of Lemma 14.6(c), which holds for every finite mixture and every $\beta > 0$ (see Remark 14.8), we have

$$\mathbb{E} f_N(r; H^{(P)}, u) \to f^{(P)}(r) = \beta^2 \xi_{\le P}(r) + \tfrac{1}{2}\log(1 - r^2) + \mathcal{P}_\beta(\eta_r^{(P)}),$$

where $\eta_r^{(P)}(t) = \xi_{\le P}(r^2 + (1 - r^2)t) - \xi_{\le P}(r^2)$. The difference $\eta_r - \eta_r^{(P)}$ is the Taylor series of $\xi_{>P}$ at $r^2 \ge 0$, evaluated at $(1 - r^2)t$, minus its constant term. It has nonnegative coefficients, and its value at $t = 1$ is at most v_P. We obtain $|f(r) - f^{(P)}(r)| \le \beta^2 v_P + \frac{1}{2}\beta^2 v_P$ from (18.10). Combining these bounds, we find

$$\limsup_{N\to\infty} |\mathbb{E} f_N(r; H, u) - f(r)| \le 2\beta\sqrt{s_P} + 3\beta^2 v_P,$$

and letting $P \to \infty$ proves Part (3).

Part (4). The proof of Lemma 14.6(b) uses only the bound of Lemma 14.6(a) for $\mathcal{P}_\beta$, which is (18.9) (with a supremum over $[0, 1]$, which is at most the supremum over $[-1, 1]$), and the bound $|\xi'(v)| \le \xi'(1)$ for $|v| \le 1$. The proof of Lemma 14.6(c) uses four inputs: the field bounds (14.3), which hold with $c_0 = c_F$ by Lemma 18.2; Lemma 14.6(b); the convergence of the means at each fixed latitude, which is Part (3); and the Lipschitz constant $(2F_\beta/N)^{1/2}$ of $f_N(r; H, u)$ in the Gaussian coordinates of $\widetilde{H}$, which gives the concentration bound used there by Part (1), with μ the image of ν_n under x_r and h the constant $\beta^2 N\xi(r)$. With these inputs, the proof applies without change. □

Replica concentration, uniformly in the temperature

The next lemma replaces Proposition 14.14 and Corollary 14.16 for infinite mixtures. It also holds for finite mixtures, where its new content is that the exponent can be chosen uniformly over the temperatures $\beta \le \bar{\beta}$ for a fixed $\bar{\beta} < \beta_s$. The proof uses the planted estimates of Lemma 14.6, the trial bound of Lemma 14.9, and likelihood transfer.

Lemma 18.5 (Uniform replica concentration)**.** *Let ξ satisfy* (18.1). *Fix $s \in (0, 1)$ and $\bar{\beta} \in (0, \beta_s)$. There exists $a > 0$, depending only on ξ, s, and $\bar{\beta}$, with the following property. For every $\beta \in (0, \bar{\beta}]$, there exist $C, N_0 > 0$ such that for all $N \ge N_0$, with probability at least $1 - Ce^{-aN}$,*

$$(\pi_\beta \otimes \pi_\beta)\{(x, y) : |R(x, y)| > s\} \le e^{-aN} \tag{18.13}$$

and

$$\sup_{v \in S_N} \pi_\beta\{x : R(x, v) \ge s\} \le e^{-aN}. \tag{18.14}$$

Proof. We first prove, for every fixed $t \in (0,1)$, that (18.13) holds at the threshold t with an exponent $a_1(t) > 0$ depending only on ξ, t, and $\bar\beta$. Then (18.14) follows from Lemma 14.15, which gives the deterministic bound

$$\sup_{v \in S_N} \pi_\beta\{R(x,v) \ge s\} \le \Big[\frac{2-s^2}{s^2}\,(\pi_\beta \otimes \pi_\beta)\{R > s^2/2\}\Big]^{1/2}$$

for every axis, applied with (18.13) at the threshold $s^2/2$; for large N, the right side is at most $e^{-a_1(s^2/2)N/4}$. We may then take $a = \min\{a_1(s), a_1(s^2/2)/4\}$.

Fix $t \in (0,1)$ and $\beta \in (0,\bar\beta]$, and write $\bar F = \bar\beta^2\xi(1)/2 \ge F_\beta$. All bounds below are stated with $\bar F$ in place of F_β, which only decreases the exponents.

Overlaps near one. Let K and c_F be as in Lemma 18.2, and let Ω_1 be the event (18.5). Let $r_c \in (t,1)$ satisfy $\frac12\log(1-r_c^2) \le -2\bar\beta K - 2$. On Ω_1, we have $\operatorname{osc} H_N \le 2KN$, which implies that $\pi_\beta(E) \le e^{2\bar\beta KN}\nu_N(E)$ for every Borel set E. By (14.11) with $H = 0$, we have $\nu_N\{R(x,u) \ge r_c\} \le C_N(1-r_c^2)^{(N-3)/2}$, and the same bound holds for $\{R(x,u) \le -r_c\}$. Hence, on Ω_1, for every $u \in S_N$,

$$\pi_\beta\{|R(x,u)| \ge r_c\} \le 2\sqrt{N}(1-r_c^2)^{-3/2}e^{-2N} \le e^{-N} \tag{18.15}$$

for all $N \ge N_1$, where N_1 depends only on ξ, t, and $\bar\beta$.

A uniform pressure gap. Recall that $\Phi_{\bar\beta} < 0$ on $(0,1)$, because $\bar\beta < \beta_{\mathrm{s}}$. We define

$$g = \min\Big\{1, \min_{t \le r \le r_c}\big(-\tfrac12\Phi_{\bar\beta}(r)\big)\Big\} > 0.$$

We have $\Phi_\beta \le \Phi_{\bar\beta}$ on $[0,1)$, because $\xi \ge 0$ there. For $r \in [0,1)$, we have $\mathcal{P}_\beta(\eta_r) \le \mathcal{P}_\beta(\delta_{r/(1+r)}; \eta_r)$, and (14.20), whose proof is a computation valid for every ξ, shows as in the proof of Lemma 14.9 that $f(r) \le \Psi_\beta(r)$. We also have $f(-r) \le f(r)$ for $r \ge 0$, because $\eta_{-r} = \eta_r$ and $\xi(-r) \le \xi(r)$. We conclude that

$$f(r) \le \Psi_\beta(|r|) \le F_\beta + \tfrac12\Phi_{\bar\beta}(|r|) \le F_\beta - g \qquad (t \le |r| \le r_c). \tag{18.16}$$

The planted bound. By Lemma 18.4(4), Lemma 14.6(c) holds with $\rho = r_c$ and $\varepsilon = g/4$: for $N \ge N_2$ and every $u \in S_N$, the event $\Omega_N(r_c, g/4)$ of that lemma has conditional probability at least $1 - e^{-c_1 N}$ under $\mathsf{P}_N(\cdot \mid u)$, where

$$c_1 = \frac12\min\Big\{c_F, \frac{g^2}{1024\bar F}\Big\}.$$

On this event, we have $f_N(r; H, u) \le F_\beta - 3g/4$ for $t \le |r| \le r_c$ by (18.16), and (14.11) implies that $Z_{N,\beta}(\{r : t \le |r| \le r_c\}; H, u) \le 2\sqrt{N}e^{N(F_\beta - 3g/4)} \le e^{N(F_\beta - g/2)}$ for large N. Let

$$E_N = \Big\{(H,u) : Z_{N,\beta}\big(\{r : t \le |r| \le r_c\}; H, u\big) > e^{N(F_\beta - g/2)}\Big\}.$$

Integrating over $u \sim \nu_N$, we have shown that $\mathsf{P}_N(E_N) \le e^{-c_1 N}$ for large N.

Transfer and disintegration. By Lemma 14.3, which holds for ξ by Lemma 18.4(2), we have $\mathsf{Q}_N(E_N) \le 3e^{-c_2 N}$ for large N, where $c_2 = \min\{c_1/2, c_1^2/(64\bar{F})\}$. Let $b_N(H) = \pi_\beta^H\{u : (H,u) \in E_N\}$. We have $\mathbb{E} b_N = \mathsf{Q}_N(E_N)$ by the definition of Q_N, and Markov's inequality gives $\mathbb{P}(b_N > e^{-c_2 N/2}) \le 3e^{-c_2 N/2}$. Finally, let $\Omega_2 = \{|N^{-1}\log Z_{N,\beta} - F_\beta| \le g/4\}$. Then (3.21), which holds for ξ by Lemma 18.4(2), gives $\mathbb{P}(\Omega_2^c) \le 2e^{-g^2 N/(256\bar{F})}$ for large N.

Conclusion. On $\Omega_1 \cap \Omega_2 \cap \{b_N \le e^{-c_2 N/2}\}$, every u with $(H,u) \notin E_N$ satisfies

$$\pi_\beta\{t \le |R(x,u)| \le r_c\} \le \frac{e^{N(F_\beta - g/2)}}{e^{N(F_\beta - g/4)}} = e^{-gN/4}.$$

Combining this with (18.15), we find that $\pi_\beta\{|R(x,u)| \ge t\} \le 2e^{-gN/4}$ for these u. Integrating over $u \sim \pi_\beta$, and bounding the contribution of the remaining reference points by their mass, we obtain

$$(\pi_\beta \otimes \pi_\beta)\{|R| > t\} \le 2e^{-gN/4} + e^{-c_2 N/2}.$$

The failure probability of the three events is at most $4e^{-c_F N} + 2e^{-g^2 N/(256\bar{F})} + 3e^{-c_2 N/2}$. We conclude that (18.13) holds at the threshold t for large N, with

$$a_1(t) = \frac{1}{2}\min\Big\{\frac{g}{4}, \frac{c_2}{2}, \frac{g^2}{256\bar{F}}\Big\},$$

where we used $c_2 \le c_1/2 \le c_F/4$. The numbers r_c, g, c_1, c_2, and $a_1(t)$ depend only on ξ, t, and $\bar{\beta}$. The thresholds for N depend also on β, through the convergence of the means in Lemma 18.4(3) and of $F_N(\beta)$. □

Good reference points for infinite mixtures

Lemma 18.6 (Constrained-pressure inputs for infinite mixtures)**.** *Let ξ satisfy* (18.1)*, suppose $\beta_{\mathrm{d}} < \beta_{\mathrm{s}}$, and fix $\beta \in (\beta_{\mathrm{d}}, \beta_{\mathrm{s}})$. Then Propositions 14.11 and 14.13 hold for ξ, with the constrained pressure defined by the infinite-mixture cross-section formula.*

Proof. We check the field, pressure, and first-variation inputs.

Step 1: the field and the dynamics. By Lemma 18.2, almost surely the field H_N is of class C^2 on B_N for every N, and (14.3) holds with $c_0 = c_F$. Lemma 18.3 implies that the dynamics is defined for every realization, with the properties in Lemmas 3.5 and 3.6 that are used in Sections 15 and 16. This proves (I1).

Step 2: the planted law and the cross-section formula. By Lemma 18.4, the convergence $F_N(\beta) \to F_\beta$, the bound (3.21), Lemma 3.3, Lemma 14.2, Lemma 14.3, and Lemma 14.6(b) and (c) hold for ξ, with $\mathcal{P}_\beta$ defined as above. This proves (I2) and (I3). The proof of

Proposition 14.7 uses these results, the coarea formula (14.11), and, for the uniform integrability of $N^{-1}\log Z_{N,\beta}([r-w,r+w];H,u)$ under Q_N, a bound on $\mathbb{E}(\sup_{S_N}|H|)^2$ that is polynomial in N, obtained there from the finite tensor representation. For an infinite mixture, we use instead the bound $\mathbb{E}(\sup_{S_N}|H_N|)^2 \le CN^2$ of Lemma 18.2(2). The rest of the proof is unchanged, so Proposition 14.7 holds for ξ.

Step 3: the equality at q_*. The proof of Lemma 14.9 uses only Proposition 14.7, the definition of $\mathcal{P}_\beta(\eta_r)$ as an infimum, and the computation (14.20), all of which are available for ξ. The proof of Proposition 14.11 applies Lemma 3.15(3) to the auxiliary series η_{q_*}, and it uses only the sufficient condition there: if $\zeta(S_\zeta)=1$, then ζ minimizes $\mathcal{P}_\beta(\cdot\,;\eta_{q_*})$. The proof of this condition, through Lemma 3.15(1) and (2), uses only that η is continuous on $[0,1]$ with $\eta(0)=0$: the energy term of (3.16) is affine in ζ and finite, the entropy term does not involve η, and $f_\zeta(s)\to-\infty$ as $s\to 1$ because η is bounded on $[0,1]$. The uniqueness of the minimizer follows from the strict convexity of the entropy term, which also does not involve η. We conclude that these facts hold for η_{q_*}, which is continuously differentiable on $[0,1]$ with $\eta'_{q_*}(t)=(1-q_*^2)\xi'(q_*^2+(1-q_*^2)t)$. The proof of Proposition 14.11 also uses the following properties of ξ, which hold under (18.1):

(a) A is continuous on $[0,1]$ with $A(1)=0$, and real analytic on $(0,1)$. As β^2A-1 tends to -1 at 1, it is not identically zero on $(0,1)$, and its zeros in $(0,1)$ are isolated. This replaces the observation that β^2A-1 is a polynomial.

(b) The identity (14.26) for the ratio $\chi(v)=v\xi''(v)/\xi'(v)$ holds on $(0,1)$. Indeed, the series $\sum_p p^k c_p v^p$ converge for $v<1$ and every k, which allows us to differentiate the weights $w_p(v)$ termwise on compact subsets of $(0,1)$.

(c) The quantities $\xi(v)$, $\xi'(v)$, and $\xi''(v)$ that appear in the computation of f'_ζ for $\zeta=\delta_{r_0}$ are finite for $v\in[0,1)$.

With these changes, the proof of Proposition 14.11 applies, and so does the proof of Proposition 14.13, which uses Lemmas 14.3 and 14.6 and Propositions 14.7 and 14.11 and (14.3). This proves (I4), and Propositions 14.11 and 14.13 hold for ξ.

□

Remark 18.7 (The summability and threshold-gap hypotheses). For a finite mixture, the sum in (18.1) is finite. If the power series of ξ converges at some $\varrho>1$, then $c_p\le C\varrho^{-p}$ for some C, and (18.1) holds. For the mixture with $c_2=0$ and $c_p=Cp^{-8}$ for $p\ge 3$, we have $\sum_p p^{5/2}\sqrt{c_p}=\sqrt{C}\sum_{p\ge3}p^{-3/2}<\infty$, and the radius of convergence is one because $c_p^{1/p}\to 1$.

It remains to prove that $\beta_{\mathrm{d}}<\beta_{\mathrm{s}}$ for this mixture. We show more generally that $\beta_{\mathrm{d}}<\beta_{\mathrm{s}}$ whenever $c_2=0$. Let $A_{\max}=\sup_{(0,1)}A=\beta_{\mathrm{d}}^{-2}$. We have $\xi'=A\ell'\le A_{\max}\ell'$ on $(0,1)$, where $\ell'(r)=r/(1-r)$. As $A(0)=2c_2=0<A_{\max}$ and A is continuous, this inequality is strict near zero. Integrating from 0, we obtain $\xi(r)<A_{\max}\ell(r)$ for

$r \in (0,1)$. As $c_2 = 0$, we have $\xi(r) = O(r^3)$ and $\ell(r) \geq r^2/2$ as $r \downarrow 0$, while $\ell(r) \to \infty$ and $\xi(r) \to \xi(1)$ as $r \uparrow 1$. Then ℓ/ξ tends to infinity at both ends of $(0,1)$, its infimum is attained at some $r_0 \in (0,1)$, and

$$\beta_{\mathrm{s}}^2 = \frac{\ell(r_0)}{\xi(r_0)} > \frac{1}{A_{\max}} = \beta_{\mathrm{d}}^2.$$

For $C = 1$, a numerical evaluation gives $\beta_{\mathrm{d}}^2 \approx 8126$ and $\beta_{\mathrm{s}}^2 \approx 8892$.

18.2 Counting slow modes

The proof of Theorem 1.7 uses the following deterministic form of the argument with cap indicators in Remark 16.1. The bound on the Gibbs mass of caps is a hypothesis, so that its exponent can be chosen independently of the temperature.

Lemma 18.8 (Counting slow modes with cap indicators)**.** *Let $N \geq 2$ and $\beta \geq 0$, let H be a function of class C^2 on S_N, and write $\pi = \pi_{\beta,H}$ and $\lambda_j = \lambda_j(-L_{\beta,H})$. Let $0 < q_1 < q_2 < 1$ and $J, G, \delta, c > 0$. Suppose that there exists a Borel set $\mathcal{G} \subset S_N$ with $\pi(\mathcal{G}) \geq 1/2$ such that for every $u \in \mathcal{G}$,*

$$\begin{aligned} \pi\{x : R(x,u) \geq q_1\} \leq e^{-N(J-2\delta)}, \qquad \pi\{x : R(x,u) \geq q_2\} \geq e^{-N(J+2\delta)}, \\ \pi\{x : q_1 \leq R(x,u) \leq q_2\} \leq e^{-N(J+G-2\delta)}, \end{aligned} \tag{18.17}$$

and that $\sup_{x \in S_N} \pi\{u : R(x,u) \geq q_1\} \leq e^{-cN}$. Let $r_N = 4N^{-1}(q_2 - q_1)^{-2} e^{-N(G-4\delta)}$. Then for every $\Lambda > r_N$,

$$\#\{j \geq 0 : \lambda_j \leq \Lambda\} \geq \Big(1 - \frac{r_N}{\Lambda}\Big)\frac{1}{2} e^{N(c-4\delta)}. \tag{18.18}$$

Proof. Let $\alpha(r) = \chi((r - q_1)/(q_2 - q_1))$, with χ as in Section 15.1, so that $\alpha = 0$ on $(-\infty, q_1]$, $\alpha = 1$ on $[q_2, \infty)$, and $|\alpha'| \leq 2/(q_2 - q_1)$. We apply Lemma 15.1, which holds for fields of class C^2 by Lemma 18.3, with

$$\mathcal{V} = S_N, \qquad \mu(\mathrm{d}u) = \mathbf{1}_{\mathcal{G}}(u)\, \pi(\mathrm{d}u), \qquad f_u(x) = \alpha(R(x,u)).$$

The function $(u,x) \mapsto f_u(x)$ is continuous and nonnegative, f_u is smooth, and $\mathbf{1}\{R(\cdot,u) \geq q_2\} \leq f_u \leq \mathbf{1}\{R(\cdot,u) \geq q_1\}$. We have $\mathsf{T} \geq \pi(\mathcal{G}) e^{-N(J+2\delta)} > 0$ by the second bound in (18.17), and $\mathsf{T} \leq 1$ because $0 \leq f_u \leq 1$. The first bound gives $\pi(f_u) \leq \mathsf{R} = e^{-N(J-2\delta)}$ for $u \in \mathcal{G}$, and the hypothesis on caps gives $\int f_u(x)\, \mu(\mathrm{d}u) \leq \mathsf{C} = e^{-cN}$ for every x. Recall from (15.2) that $|\nabla_{S_N} R(\cdot,u)|^2 \leq 1/N$. Then $|\nabla_{S_N} f_u|^2 \leq 4N^{-1}(q_2 - q_1)^{-2} \mathbf{1}\{q_1 \leq R(\cdot,u) \leq q_2\}$, and the third bound implies that

$$\int \mathcal{E}(f_u, f_u)\, \mu(\mathrm{d}u) \leq \pi(\mathcal{G}) \frac{4}{N(q_2 - q_1)^2} e^{-N(J+G-2\delta)} \leq r_N \mathsf{T}.$$

By Lemma 15.1, for every $\Lambda > r_N$,

$$\#\{j \geq 0 : \lambda_j \leq \Lambda\} \geq \Big(1 - \frac{r_N}{\Lambda}\Big)\frac{\mathsf{T}}{\mathsf{RC}} \geq \Big(1 - \frac{r_N}{\Lambda}\Big)\frac{\pi(\mathcal{G})e^{-N(J+2\delta)}}{e^{-N(J-2\delta)}e^{-cN}},$$

Inserting $\pi(\mathcal{G}) \geq 1/2$, we obtain (18.18). □

The core exponent J cancels in $\mathsf{T}/(\mathsf{RC})$, and the good-reference mass $\pi(\mathcal{G})$ cancels in the energy ratio r_N. The spectral exponent $G - 4\delta$ and the mode-count exponent $c - 4\delta$ can therefore be chosen independently. If ξ satisfies (18.1), $\beta_{\mathrm{d}} < \beta < \beta_{\mathrm{s}}$, (q_1, q_2) is an admissible collar, and $0 < \delta < \min\{J, G\}/8$, then the hypotheses (18.17) hold on an event of probability at least $1 - 5e^{-a_{\mathcal{G}}N}$ by Proposition 14.13, with $J = J(\beta)$ and G the collar gap of Definition 14.10; the second bound follows from the core lower bound (14.31), since the core band lies in $\{R > q_2\}$.

18.3 Eigenvalues near the dynamical threshold

In this section, we prove Theorem 1.7. We first recall its setting. Let ξ satisfy (18.1) with $\beta_{\mathrm{d}} < \beta_{\mathrm{s}}$. The function A is continuous on $[0,1]$, and we write $A_{\max} = \sup_{(0,1)} A = \beta_{\mathrm{d}}^{-2}$. Step 1 of the proof below shows that the set of global maximizers of A on $[0,1]$ is a finite subset of $(0,1)$; let q_d be its largest element. Step 1 also shows that there exist an integer $m \geq 1$ and $a > 0$ such that

$$A(q_d + y) = A_{\max}\big[1 - ay^{2m} + O(y^{2m+1})\big] \qquad (y \to 0). \tag{18.19}$$

For $\beta > \beta_{\mathrm{d}}$, we set $\varepsilon = \beta^2 A_{\max} - 1$, so that $\beta = \beta_{\mathrm{d}}(1+\varepsilon)^{1/2}$, and

$$C_d = \frac{2mq_d}{(2m+1)(1-q_d)}\, a^{-1/(2m)}, \qquad D_\varepsilon = C_d \varepsilon^{1+1/(2m)}.$$

Theorem 1.7 states that there exists $\kappa_* > 0$ such that for every $\vartheta \in (0,1)$, there exists $\varepsilon_0(\vartheta) > 0$ such that for every fixed $\varepsilon \in (0, \varepsilon_0(\vartheta))$, there exist $a', C > 0$ with

$$\lambda_{\lfloor e^{\kappa_* N} \rfloor}(\beta) \leq e^{-(1-\vartheta)D_\varepsilon N}$$

with probability at least $1 - Ce^{-a'N}$, for all sufficiently large N. We write a' for the exponent in the probability bound, to avoid a clash with the Taylor coefficient a. The constants a' and C and the threshold for N may depend on ε and ϑ; the exponent κ_* does not.

Proof of Theorem 1.7. Steps 1–4 are deterministic. They locate the final component of the strict superlevel set D_β near q_d, compute the barrier of the trial pressure Ψ_β, and place the collar. Step 5 fixes the mode-count exponent κ_* with Lemma 18.5, before ϑ and ε are chosen. Step 6 applies Lemma 18.8. The parameters are fixed in the order of Table 18.1.

Parameter	Choice	Depends on
q_d, m, a, φ	Step 1	ξ
ρ_0, ε_1, L_0	Step 2	ξ
s, $\bar\beta$	$s = q_d/2$, $\bar\beta = (\beta_{\mathrm{d}} + \beta_{\mathrm{s}})/2$	ξ
a_s, κ_*	a_s from Lemma 18.5, $\kappa_* = a_s/4$	ξ
ϑ	given in $(0,1)$	—
$\varepsilon_0(\vartheta)$	Steps 4 and 6	ξ, ϑ
ε, β	given $\varepsilon \in (0, \varepsilon_0)$; $\beta = \beta_{\mathrm{d}}(1+\varepsilon)^{1/2}$	—
$q_\pm$, q_1, q_2, G, J	Steps 2–4	ξ, ε
δ	$\frac{1}{64}\min\{a_s, \vartheta D_\varepsilon, J, 1\}$	$\xi, \vartheta, \varepsilon$
a', C, N_0	from Proposition 14.13 and Lemma 18.5	$\xi, \vartheta, \varepsilon$

Table 18.1. Parameters in the proof of Theorem 1.7, in the order in which they are fixed. All are fixed before $N \to \infty$.

Step 1: the global maximizers of A. If $c_2 > 0$, then $\ell(r)/\xi(r) \to 1/(2c_2)$ as $r \downarrow 0$, and the definition of β_{s} gives $\beta_{\mathrm{s}}^{-2} \geq 2c_2 = A(0)$; if $c_2 = 0$, then $A(0) = 0$. As $\beta_{\mathrm{d}} < \beta_{\mathrm{s}}$, we have in both cases

$$A_{\max} = \beta_{\mathrm{d}}^{-2} > \max\{\beta_{\mathrm{s}}^{-2}, 0\} \geq A(0), \qquad A(1) = 0 < A_{\max}.$$

By the continuity of A on $[0,1]$, its supremum is attained, and the set of global maximizers is a compact subset of $(0,1)$. The function $A - A_{\max}$ is real analytic on $(0,1)$ and not identically zero, since it tends to $-A_{\max}$ at 1. Its zeros in a compact subset of $(0,1)$, and in particular the global maximizers, are therefore finite in number. Let q_d be the largest. The Taylor series of $A - A_{\max}$ at q_d has a first nonzero coefficient, and since $A - A_{\max} \leq 0$ near q_d, its order is even, say $2m$, and the coefficient is negative. We can then write $A(q_d + y) = A_{\max}[1 - a y^{2m} \varphi(y)]$, where $a = -A^{(2m)}(q_d)/((2m)!\, A_{\max}) > 0$ and φ is real analytic near 0 with $\varphi(0) = 1$. This proves (18.19). We have not assumed that q_d is the only global maximizer, or that $m = 1$.

Step 2: the final component of D_β. Differentiating, we have $A'(q_d + y) = -2ma A_{\max} y^{2m-1}(1 + O(y))$. Let $\rho_0 > 0$ be such that $[q_d - \rho_0, q_d + \rho_0] \subset (0,1)$ contains no other global maximizer, $A' > 0$ on $[q_d - \rho_0, q_d)$, $A' < 0$ on $(q_d, q_d + \rho_0]$, and $\varphi \geq 1/2$ on $[-\rho_0, \rho_0]$. Since the compact set $[q_d + \rho_0, 1] \cup \{q_d - \rho_0\}$ contains no global maximizer, we have

$$g_1 = A_{\max} - \max\Big\{ \max_{[q_d + \rho_0, 1]} A,\ A(q_d - \rho_0) \Big\} > 0.$$

Let $\varepsilon_1 > 0$ be such that $A_{\max}/(1+\varepsilon) > A_{\max} - g_1$ for $\varepsilon < \varepsilon_1$, and fix $\varepsilon \in (0, \varepsilon_1)$. Then $r \in D_\beta$ if and only if $A(r) > A_{\max}/(1+\varepsilon)$, and in particular $D_\beta \cap [q_d + \rho_0, 1) = \emptyset$.

On $[q_d-\rho_0, q_d+\rho_0]$, the function A increases strictly up to $A(q_d)=A_{\max}$ and then decreases strictly, and its values at both endpoints are below $A_{\max}/(1+\varepsilon)$. It follows that $D_\beta \cap (q_d-\rho_0, q_d+\rho_0) = (q_-, q_+)$, where $q_- < q_d < q_+$ are the two solutions of $A(r) = A_{\max}/(1+\varepsilon)$ in this interval. We conclude that

$$q_*(\beta) = \sup D_\beta = q_+,$$

and that (q_-, q_+) is the connected component of D_β that ends at q_*. Any other component of D_β lies in $(0, q_d-\rho_0]$ and does not affect q_*.

Let $h = (\varepsilon/a)^{1/(2m)}$, and write $q_\pm = q_d + hz_\pm$. Using $ah^{2m} = \varepsilon$, we can write the equation $A(q_\pm) = A_{\max}/(1+\varepsilon)$ as $z_\pm^{2m}\varphi(hz_\pm) = (1+\varepsilon)^{-1}$. As $|hz_\pm| < \rho_0$ and $\varphi \ge 1/2$ there, we have $|z_\pm| \le 2$. Further, $\varphi(hz) = 1 + O(h)$ uniformly for $|z| \le 2$, and $(1+\varepsilon)^{-1} = 1 + O(h^{2m})$. We obtain $z_\pm = \pm 1 + O(h)$, that is,

$$q_\pm = q_d \pm h + O(h^2). \tag{18.20}$$

Step 3: the trial barrier. By (14.22) at inverse temperature β, we have

$$\Psi'_\beta(r) = \frac{r}{2(1-r)}\big(\beta^2 A(r) - 1\big). \tag{18.21}$$

For $r = q_d + hz$ with $|z| \le 2$, we find

$$\begin{aligned}\beta^2 A(r) - 1 &= (1+\varepsilon)\big(1 - \varepsilon z^{2m}\varphi(hz)\big) - 1 \\ &= \varepsilon(1-z^{2m}) - \varepsilon z^{2m}\big(\varphi(hz)-1\big) - \varepsilon^2 z^{2m}\varphi(hz) = \varepsilon(1-z^{2m}) + O(\varepsilon h),\end{aligned}$$

where we used $\varphi(hz) - 1 = O(h)$ and $\varepsilon = O(h)$. Also $r/(2(1-r)) = q_d/(2(1-q_d)) + O(h)$. Changing variables $r = q_d + hz$ in the integral of (18.21) over $[q_-, q_+]$, and using $z_\pm = \pm 1 + O(h)$ and that the integrand is $O(\varepsilon)$, we obtain

$$\Psi_\beta(q_+) - \Psi_\beta(q_-) = \frac{q_d}{2(1-q_d)}\, h\varepsilon \int_{-1}^{1} (1-z^{2m})\, \mathrm{d}z + O(\varepsilon h^2).$$

Inserting $\int_{-1}^{1}(1-z^{2m})\,\mathrm{d}z = 4m/(2m+1)$ and $h\varepsilon = a^{-1/(2m)}\varepsilon^{1+1/(2m)}$, we find

$$\Psi_\beta(q_+) - \Psi_\beta(q_-) = D_\varepsilon\big(1 + O(\varepsilon^{1/(2m)})\big). \tag{18.22}$$

Step 4: the collar. Let $q_1 = q_- + \varepsilon/2$ and $q_2 = q_- + \varepsilon$. Since $\varepsilon/h = a^{1/(2m)}\varepsilon^{1-1/(2m)} \to 0$ and $q_+ - q_- = 2h + O(h^2)$ by (18.20), there exists $\varepsilon_2 \in (0, \varepsilon_1]$ such that $q_- < q_1 < q_2 < q_+$ for $\varepsilon < \varepsilon_2$. Then (q_1, q_2) is an admissible collar at β, since $(q_1, q_*) \subset (q_-, q_+) \subset D_\beta$. On (q_-, q_+), we have $0 < \beta^2 A - 1 \le \varepsilon$, since $A \le A_{\max}$, and $r/(2(1-r)) \le L_0 = (q_d + \rho_0)/(2(1-q_d-\rho_0))$. Integrating (18.21) over $[q_-, q_2]$, we find $0 \le \Psi_\beta(q_2) - \Psi_\beta(q_-) \le L_0\varepsilon^2$. As $\varepsilon^2/D_\varepsilon$ is a constant multiple of $\varepsilon^{1-1/(2m)}$, we deduce from (18.22) that the collar gap of Definition 14.10 satisfies

$$G = \Psi_\beta(q_*) - \Psi_\beta(q_2) = D_\varepsilon\big(1 + O(\varepsilon^{1/(2m)} + \varepsilon^{1-1/(2m)})\big). \tag{18.23}$$

Step 5: a mode-count exponent independent of ε. Let $s = q_d/2$ and $\bar\beta = (\beta_{\mathrm{d}} + \beta_{\mathrm{s}})/2$, and let $a_s > 0$ be the exponent of Lemma 18.5 for these s and $\bar\beta$; it depends only on ξ. For every $\beta \in (0, \bar\beta]$, there exist C_2, N_2 such that for $N \geq N_2$, with probability at least $1 - C_2 e^{-a_s N}$,

$$\sup_{v \in S_N} \pi_\beta\{x : R(x,v) \geq s\} \leq e^{-a_s N}. \tag{18.24}$$

We set $\kappa_* = a_s/4$.

Step 6: counting slow modes. Fix $\vartheta \in (0,1)$. Let $\varepsilon_0(\vartheta) \leq \min\{\varepsilon_2, (\bar\beta/\beta_{\mathrm{d}})^2 - 1\}$ be such that for all $\varepsilon \in (0, \varepsilon_0(\vartheta))$, we have $q_1 \geq s$ and, by (18.23), $G \geq (1 - \vartheta/4) D_\varepsilon$. Fix $\varepsilon \in (0, \varepsilon_0(\vartheta))$, so that $\beta = \beta_{\mathrm{d}}(1+\varepsilon)^{1/2} \in (\beta_{\mathrm{d}}, \bar\beta]$. Let $J = J(\beta) > 0$ and

$$\delta = \tfrac{1}{64} \min\{a_s, \vartheta D_\varepsilon, J, 1\}.$$

Then $8\delta \leq \vartheta D_\varepsilon/8 < G$ and $8\delta < J$, that is, $\delta < \min\{J, G\}/8$. By Proposition 14.13 at β, with the collar (q_1, q_2) and tolerance δ, which holds for ξ by Lemma 18.6, there exist $a_1 > 0$ and an event of probability at least $1 - 5e^{-a_1 N}$ on which there is a Borel set $\mathcal{G}$ with $\pi_\beta(\mathcal{G}^c) \leq e^{-a_1 N}$ such that (18.17) holds for every $u \in \mathcal{G}$. We work on the intersection of this event with the event of (18.24). As $q_1 \geq s$ and R is symmetric, (18.24) implies that

$$\sup_{x \in S_N} \pi_\beta\{u : R(x,u) \geq q_1\} \leq \sup_{v \in S_N} \pi_\beta\{R(\cdot, v) \geq s\} \leq e^{-a_s N}.$$

For N large enough that $\pi_\beta(\mathcal{G}) \geq 1/2$, we apply Lemma 18.8 with $c = a_s$ and $\Lambda = e^{-(1-\vartheta)D_\varepsilon N}$. We have $r_N = 16 N^{-1} \varepsilon^{-2} e^{-N(G - 4\delta)}$, since $q_2 - q_1 = \varepsilon/2$. Using $4\delta \leq \vartheta D_\varepsilon/16$, we find

$$G - 4\delta - (1-\vartheta) D_\varepsilon \geq \Big(\vartheta - \frac{\vartheta}{4} - \frac{\vartheta}{16}\Big) D_\varepsilon \geq \frac{\vartheta}{2} D_\varepsilon,$$

which implies that $r_N/\Lambda \leq 16 N^{-1} \varepsilon^{-2} e^{-\vartheta D_\varepsilon N/2} \leq 1/2$ for large N. As $4\delta \leq a_s/16$, (18.18) implies that at least $\frac{1}{4} e^{15 a_s N/16}$ eigenvalues, counted with multiplicity and including $\lambda_0 = 0$, lie in $[0, \Lambda]$. For large N, this number exceeds $\lfloor e^{\kappa_* N} \rfloor + 1$, which gives $\lambda_{\lfloor e^{\kappa_* N} \rfloor}(\beta) \leq \Lambda$. The probability of the event is at least $1 - 5e^{-a_1 N} - C_2 e^{-a_s N}$, which proves the theorem with $a' = \min\{a_1, a_s\}$. □

In Lemma 18.8, the core exponent J enters both T and R and cancels. This is why κ_* does not depend on ε, although the spectral exponent $(1-\vartheta)D_\varepsilon$ tends to zero with ε.

The trial barrier. By Proposition 14.11, the trial pressure Ψ_β is an upper bound for the constrained pressure F_{FP}, with equality at $q_* = q_+$. Thus the actual gap between the supremum of F_{FP} over the cap $[q_1, 1)$ and its supremum over the collar $[q_1, q_2]$ is at least G. The quantity $D_\varepsilon = C_d \varepsilon^{1 + 1/(2m)}$ is the leading order of the barrier $\Psi_\beta(q_+) - \Psi_\beta(q_-)$ of the trial pressure, and (18.23) shows that it is also the leading order of G. It is therefore the leading order of a lower bound for the cap-to-collar gap of the constrained pressure. We

do not identify the leading order of that gap itself, and the theorem bounds eigenvalues from above without identifying their asymptotics. For a nondegenerate maximum, $m = 1$ and the exponent is $3/2$. In the physics literature, the Franz–Parisi potential near the dynamical transition has the cubic normal form of a spinodal [72, Section 5.2.1]; see also [150] for the classical spinodal. In this normal form, the difference in height between the local minimum and the saddle point scales as $|T - T_d|^{3/2}$. The case $m = 1$ of Theorem 1.7 is consistent with this scaling, and the exponent $1 + 1/(2m)$ is the same computation for a maximum of order $2m$. The mode-count exponent κ_* stays bounded below as $\varepsilon \downarrow 0$, while the spectral exponent tends to zero.

Example 18.9 (The pure cubic model)**.** Let $\xi(r) = r^3$. Then $A(r) = 3r(1-r)$, $q_d = 1/2$, $A_{\max} = 3/4$, and $\beta_{\mathrm{d}}^2 = 4/3$. We have $A(1/2 + y) = \frac{3}{4}(1 - 4y^2)$, which gives $m = 1$, $a = 4$, and $C_d = 1/3$. Minimizing $\ell(r)/r^3$ numerically gives $\beta_{\mathrm{s}}^2 \approx 1.456$, attained near $r \approx 0.645$. Then $\beta_{\mathrm{d}} < \beta_{\mathrm{s}}$, and Theorem 1.7 applies.

For this model, the trial barrier can be computed exactly. Let $u = (\varepsilon/(1+\varepsilon))^{1/2}$. Then $\beta^2 = 4/(3(1-u^2))$, the equation $\beta^2 A(r) = 1$ reads $(r - 1/2)^2 = u^2/4$, and $D_\beta = (q_-, q_+)$ with $q_\pm = (1 \pm u)/2$. Using $q_+^3 - q_-^3 = (3u + u^3)/4$, $q_+ - q_- = u$, and $\log((1-q_+)/(1-q_-)) = \log((1-u)/(1+u)) = -2\operatorname{atanh} u$, we find

$$\begin{aligned}\Psi_\beta(q_+) - \Psi_\beta(q_-) &= \frac{1}{2}\Big[\beta^2(q_+^3 - q_-^3) + q_+ - q_- + \log\frac{1-q_+}{1-q_-}\Big] \\ &= \frac{u(3-u^2)}{3(1-u^2)} - \operatorname{atanh} u.\end{aligned}$$

Expanding $u(3-u^2)/(3(1-u^2)) = u + \frac{2}{3}\sum_{k\ge 1} u^{2k+1}$ and $\operatorname{atanh} u = \sum_{k\ge 0} u^{2k+1}/(2k+1)$, we obtain the convergent series

$$\Psi_\beta(q_+) - \Psi_\beta(q_-) = \sum_{k\ge 1} \frac{4k-1}{3(2k+1)}\, u^{2k+1} = \frac{u^3}{3} + \frac{7u^5}{15} + \frac{11u^7}{21} + O(u^9).$$

Substituting $u^2 = \varepsilon/(1+\varepsilon)$ and expanding $(1+\varepsilon)^{-3/2}$, $(1+\varepsilon)^{-5/2}$, and $(1+\varepsilon)^{-7/2}$ to the needed order gives

$$\Psi_\beta(q_+) - \Psi_\beta(q_-) = \frac{\varepsilon^{3/2}}{3} - \frac{\varepsilon^{5/2}}{30} - \frac{\varepsilon^{7/2}}{56} + O(\varepsilon^{9/2}),$$

in agreement with $D_\varepsilon = \varepsilon^{3/2}/3$. Replacing ξ by $c_3\xi$ with $c_3 > 0$ multiplies A by c_3 and leaves q_d, m, a, C_d, and the barrier, as a function of ε, unchanged. As explained above, this is the barrier of the trial pressure; it is not a formula for the barrier of the constrained pressure.

Example 18.10 (Separated thresholds are needed)**.** Let $\xi(r) = r^2/2 + r^4$. Then

$$A(r) = (1-r)(1+4r^2) = 1 - 4r\big(r - \tfrac{1}{2}\big)^2,$$

so $A \le 1$ on $[0,1]$, with equality exactly at $r = 0$ and $r = 1/2$. Hence $\beta_{\mathrm{d}} = 1$, and $1/2$ is the unique global maximizer of A in $(0,1)$. The function $\ell - \xi$ vanishes at 0, and

$$(\ell - \xi)'(r) = \frac{r}{1-r} - r - 4r^3 = \frac{r^2(2r-1)^2}{1-r} \ge 0,$$

with equality only at $r \in \{0, 1/2\}$. Hence $\ell(r)/\xi(r) > 1$ for $r \in (0,1)$. Since $\ell(r)/\xi(r) \to 1/(2c_2) = 1$ as $r \downarrow 0$, we obtain $\beta_{\mathrm{s}}^2 = \inf_{(0,1)} \ell/\xi = 1$. Thus $\beta_{\mathrm{d}} = \beta_{\mathrm{s}} = 1$, the interval $(\beta_{\mathrm{d}}, \beta_{\mathrm{s}})$ between the dynamical and static thresholds is empty, and for every $\varepsilon > 0$ the inverse temperature $\beta_{\mathrm{d}}(1+\varepsilon)^{1/2}$ lies above β_{s}. Step 1 of the proof of Theorem 1.7 uses the hypothesis $\beta_{\mathrm{d}} < \beta_{\mathrm{s}}$ to exclude 0 from the global maximizers of A; this example shows that the hypothesis does not follow from the uniqueness of an interior maximizer.

In Theorem 1.7, ε is fixed before $N \to \infty$. The result bounds eigenvalues only; it does not give the conclusions (ii)–(v) of Theorem 1.5 with time exponents of order D_ε.

APPENDIX A

Exact rational bounds for the nonconvex example

In this appendix, we prove the two finite statements used in the proof of Proposition 5.24: an enclosure of the parameter t of (5.38) between two rational numbers (Lemma A.1), and the sign pattern of the polynomial P_t of (5.41) for all t in this enclosure (Lemma A.3). We also record the two rational comparisons used at the end of that proof. Every step is a finite computation with rational numbers. The sign pattern follows from the standard Bernstein-coefficient enclosure [68, 115], which we recall in Lemma A.2.

The computations can be repeated with the ancillary file `verify-thermal-example.py`, which is distributed with this paper. It uses only the Python standard library, with integer and rational arithmetic (`fractions.Fraction`); it uses no floating-point arithmetic, sampling, or root finding. It checks the inequalities of Lemma A.1, computes all 126 Bernstein coefficients in Table A.1 at both endpoints of the enclosure, checks the four sign-and-margin bounds of Lemma A.3, checks the endpoint identities (A.3) as a control of the conversion, and checks the comparisons in (A.5).

Throughout, we write

$$t_- = \frac{4539}{100000}, \qquad t_+ = \frac{4541}{100000}, \qquad \mathcal{L}(u) = \frac{45 + 85u}{24 + 72u}.$$

Lemma A.1 (Enclosure of the parameter)**.** *Let* $t = (45 - 24\log 6)/(72\log 6 - 85)$. *Then* $t_- < t < t_+$, *and*

$$\frac{4885985}{2726952} = \mathcal{L}(t_+) < \log 6 < \mathcal{L}(t_-) = \frac{1628605}{908936}. \tag{A.1}$$

Proof. Let $y_0 = 5/7$. Since $(1 + y_0)/(1 - y_0) = 6$, we have $\log 6 = 2\,\mathrm{atanh}(y_0) = 2\sum_{k\geq 0} y_0^{2k+1}/(2k+1)$, a series with positive terms. Let

$$S_{13} = 2\sum_{k=0}^{12} \frac{y_0^{2k+1}}{2k+1}, \qquad R_{13} = \frac{2y_0^{27}}{27(1 - y_0^2)}.$$

The omitted terms satisfy $2y_0^{2k+1}/(2k+1) \leq 2y_0^{2k+1}/27$ for $k \geq 13$, and by summing the geometric series we find $S_{13} < \log 6 < S_{13} + R_{13}$. The rational numbers S_{13} and R_{13} satisfy

$$\frac{85}{72} < S_{13}, \qquad \mathcal{L}(t_+) < S_{13} < S_{13} + R_{13} < \mathcal{L}(t_-).$$

For orientation, $S_{13} \approx 1.7917434$, $R_{13} \approx 1.7 \times 10^{-5}$, $\mathcal{L}(t_+) \approx 1.7917385$, and $\mathcal{L}(t_-) \approx 1.7917708$; the margins are $S_{13} - \mathcal{L}(t_+) \approx 4.9 \times 10^{-6}$ and $\mathcal{L}(t_-) - S_{13} - R_{13} \approx 1.0 \times 10^{-5}$.

These decimals are not part of the proof, which compares the rational numbers exactly. This proves (A.1), and also $\log 6 > 85/72$. The function $\ell \mapsto (45 - 24\ell)/(72\ell - 85)$ is the inverse of $\mathcal{L}$ on $\ell > 85/72$, and its derivative is $-1200/(72\ell - 85)^2 < 0$. Its value at $\ell = \log 6$ is t, and its values at $\mathcal{L}(t_\pm)$ are $t_\pm$. Then (A.1) implies that $t_- < t < t_+$. □

Lemma A.2 (Bernstein coefficients)**.** *Let $p(r) = \sum_{i=0}^{n} c_i r^i$ be a real polynomial of degree at most n, and let $a < b$. For $0 \le k \le n$, define*

$$B_k(p; [a,b]) = \sum_{j=0}^{k} \frac{\binom{k}{j}}{\binom{n}{j}} (b-a)^j \sum_{i=j}^{n} c_i \binom{i}{j} a^{i-j}. \tag{A.2}$$

Then $p(a + (b-a)v) = \sum_{k=0}^{n} B_k(p; [a,b]) \binom{n}{k} v^k (1-v)^{n-k}$ for all $v \in \mathbb{R}$. In particular,

$$B_0(p; [a,b]) = p(a), \qquad B_n(p; [a,b]) = p(b), \tag{A.3}$$

and $\min_k B_k(p; [a,b]) \le p(r) \le \max_k B_k(p; [a,b])$ for every $r \in [a,b]$.

Proof. Expanding $p(a+h)$ in powers of h, we find that the coefficient of v^j in $p(a + (b-a)v)$ is $(b-a)^j \sum_{i=j}^{n} c_i \binom{i}{j} a^{i-j}$. By the binomial theorem applied to $(v + (1-v))^{n-j}$, and the identity $\binom{n}{k}\binom{k}{j} = \binom{n}{j}\binom{n-j}{k-j}$, we have

$$v^j = \sum_{k=j}^{n} \frac{\binom{k}{j}}{\binom{n}{j}} \binom{n}{k} v^k (1-v)^{n-k}, \qquad 0 \le j \le n.$$

Substituting this into the expansion of $p(a + (b-a)v)$ and collecting the coefficient of $\binom{n}{k} v^k (1-v)^{n-k}$, we obtain the first assertion. Evaluating at $v = 0$ and $v = 1$ gives (A.3). For $v \in [0,1]$, the functions $\binom{n}{k} v^k (1-v)^{n-k}$ are nonnegative and sum to one, so $p(a + (b-a)v)$ is a convex combination of the coefficients $B_k(p; [a,b])$. □

For $t \in \mathbb{R}$, the polynomial of (5.41) is

$$P_t(r) = 24 + 72t - 432r^2 + 720r^3 - 300r^4 - 4752tr^{10} + 7920tr^{11} - 3300tr^{12}, \tag{A.4}$$

of degree at most 12, and its derivative P_t' has degree at most 11. Each coefficient of P_t and of P_t' is an affine function of t. It follows that each Bernstein coefficient $B_k(P_t; [a,b])$ and $B_k(P_t'; [a,b])$ is an affine function of t, and a strict bound on it that holds at $t = t_-$ and at $t = t_+$ holds for all $t \in [t_-, t_+]$.

Lemma A.3 (Sign pattern of P_t)**.** *For every $t \in [t_-, t_+]$, we have*

$$\begin{aligned}
&P_t(r) > 2/25 && \text{for } r \in [0, 357/1000] \cup [892/1000, 1],\\
&P_t(r) < -1/25 && \text{for } r \in [359/1000, 890/1000],\\
&P_t'(r) < -86 && \text{for } r \in [357/1000, 359/1000],\\
&P_t'(r) > 73 && \text{for } r \in [890/1000, 892/1000].
\end{aligned}$$

In particular, P_t has exactly two zeros $a_0 < b_0$ in $[0,1]$, both simple, with $a_0 \in (357/1000, 359/1000)$ and $b_0 \in (890/1000, 892/1000)$, and $P_t > 0$ on $[0,a_0) \cup (b_0,1]$ and $P_t < 0$ on (a_0,b_0).

Polynomial	Interval	Sign ς	Lower bounds at t_-, t_+	Coefficients
P_t	[0, 357/1000]	$+1$	93, 94	26
P_t	[359/1000, 890/1000]	-1	56, 57	26
P_t	[892/1000, 1]	$+1$	92, 92	26
P_t'	[357/1000, 359/1000]	-1	87410, 87410	24
P_t'	[890/1000, 892/1000]	$+1$	74538, 74529	24

Table A.1. Bounds on the Bernstein coefficients of P_t and P_t'. In each row, every Bernstein coefficient B_k of the polynomial on the interval, computed by (A.2) with $n = 12$ for P_t and $n = 11$ for P_t', satisfies $1000\,\varsigma B_k > L$, where L is the first listed integer at $t = t_-$ and the second at $t = t_+$. The last column counts the coefficients over both endpoints; the total is 126.

Proof. An exact computation of the coefficients (A.2) at $t = t_-$ and $t = t_+$ gives the bounds in Table A.1. The lower bounds $93, 94, 92$ exceed $80 = 1000 \cdot 2/25$, the bounds $56, 57$ exceed $40 = 1000 \cdot 1/25$, and the bounds 87410 and $74538, 74529$ exceed 86000 and 73000. The rest of the proof uses only these weaker margins, which are the ones checked by the ancillary script; the integer bounds in the table are stronger and record the size of the margins. By the remark before Lemma A.3, the same strict bounds on the Bernstein coefficients hold for all $t \in [t_-, t_+]$, and Lemma A.2 transfers them to the polynomials on the intervals. This proves the four displayed bounds.

The five intervals cover $[0,1]$. On the three long intervals, P_t has no zero. On $[357/1000, 359/1000]$, P_t is strictly decreasing, positive at the left endpoint, and negative at the right endpoint, so it has exactly one zero a_0 there, which is simple because $P_t'(a_0) < 0$. Similarly, P_t has exactly one zero $b_0 \in (890/1000, 892/1000)$, and $P_t'(b_0) > 0$. The sign pattern follows. □

Finally, the proof of Proposition 5.24 uses the following rational comparisons:

$$\frac{12 + 132t_+}{24 + 72t_+} = \frac{149951}{227246} < \frac{2}{3}, \qquad \frac{1}{4} < 11t_-\Big(\frac{25}{27}\Big)^8, \qquad 11t_+\Big(\frac{25}{27}\Big)^8 < \frac{3}{10}. \tag{A.5}$$

The first is checked by cross-multiplication, since $3 \cdot 149951 = 449853 < 454492 = 2 \cdot 227246$. For the other two, $11t_\pm(25/27)^8$ are explicit rational numbers, approximately 0.26975 and 0.26987, and the two inequalities are again checked by cross-multiplication.

References

[1] Dimitris Achlioptas and Amin Coja-Oghlan. Algorithmic barriers from phase transitions. In *2008 49th Annual IEEE Symposium on Foundations of Computer Science*, pages 793–802. IEEE, 2008. doi: 10.1109/FOCS.2008.11. URL `https://arxiv.org/abs/0803.2122`.

[2] Arka Adhikari, Christian Brennecke, Changji Xu, and Horng-Tzer Yau. Spectral gap estimates for mixed p-spin models at high temperature. *Probability Theory and Related Fields*, 189(3-4):879–907, 2024. doi: 10.1007/s00440-024-01261-9.

[3] Robert J. Adler and Jonathan E. Taylor. *Random Fields and Geometry*. Springer Monographs in Mathematics. Springer, 2007. doi: 10.1007/978-0-387-48116-6.

[4] Michael Aizenman and Pierluigi Contucci. On the stability of the quenched state in mean-field spin-glass models. *Journal of Statistical Physics*, 92(5-6):765–783, 1998. doi: 10.1023/A:1023080223894.

[5] Nima Anari, Vishesh Jain, Frederic Koehler, Huy Tuan Pham, and Thuy-Duong Vuong. Universality of spectral independence with applications to fast mixing in spin glasses. In *Proceedings of the 2024 Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 5029–5056. SIAM, 2024. doi: 10.1137/1.9781611977912.181.

[6] Greg W. Anderson, Alice Guionnet, and Ofer Zeitouni. *An Introduction to Random Matrices*, volume 118 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, 2010. doi: 10.1017/CBO9780511801334.

[7] Kendall Atkinson and Weimin Han. *Spherical Harmonics and Approximations on the Unit Sphere: An Introduction*, volume 2044 of *Lecture Notes in Mathematics*. Springer, Heidelberg, 2012. doi: 10.1007/978-3-642-25983-8.

[8] Antonio Auffinger and Gérard Ben Arous. Complexity of random smooth functions on the high-dimensional sphere. *The Annals of Probability*, 41(6):4214–4247, 2013. doi: 10.1214/13-AOP862.

[9] Antonio Auffinger and Wei-Kuo Chen. On properties of Parisi measures. *Probability Theory and Related Fields*, 161:817–850, 2015. doi: 10.1007/s00440-014-0563-y.

[10] Antonio Auffinger and Wei-Kuo Chen. On the energy landscape of spherical spin glasses. *Advances in Mathematics*, 330:553–588, 2018. doi: 10.1016/j.aim.2018.03.028.

[11] Antonio Auffinger and Yuxin Zhou. The spherical $p + s$ spin glass at zero temperature. *Communications in Mathematical Physics*, 406(12):314, 2025. doi:

10.1007/s00220-025-05487-z. URL https://link.springer.com/article/10.1007/s00220-025-05487-z. Article 314.

[12] Antonio Auffinger, Gérard Ben Arous, and Jiří Černý. Random matrices and complexity of spin glasses. *Communications on Pure and Applied Mathematics*, 66(2): 165–201, 2013. doi: 10.1002/cpa.21422.

[13] Antonio Auffinger, Ahmed El Alaoui, and Mark Sellke. On the discontinuous breaking of replica symmetry and shattering in mean-field spin glasses, 2025. URL https://arxiv.org/abs/2506.02238v1.

[14] Jean-Marc Azaïs and Mario Wschebor. *Level Sets and Extrema of Random Processes and Fields*. John Wiley & Sons, 2009. doi: 10.1002/9780470434642. URL https://www.math.univ-toulouse.fr/~azais/styles/other/student/level.pdf. Author draft.

[15] Masoud Badiei Khuzani. Spectral gap bounds for Langevin dynamics in mixed spherical spin glasses, 2026. URL https://arxiv.org/abs/2609.15157v3. Version 3, October 2, 2026.

[16] Dominique Bakry and Michel Émery. Diffusions hypercontractives. In *Séminaire de Probabilités XIX, 1983/84*, volume 1123 of *Lecture Notes in Mathematics*, pages 177–206. Springer, Berlin, 1985. doi: 10.1007/BFb0075847.

[17] Dominique Bakry, Ivan Gentil, and Michel Ledoux. *Analysis and Geometry of Markov Diffusion Operators*, volume 348 of *Grundlehren der mathematischen Wissenschaften*. Springer, 2014. doi: 10.1007/978-3-319-00227-9.

[18] Alexander Barg and Oleg R. Musin. Codes in spherical caps. *Advances in Mathematics of Communications*, 1(1):131–149, 2007. doi: 10.3934/amc.2007.1.131. ArXiv:math/0606734v2.

[19] A. Barrat, R. Burioni, and M. Mézard. Dynamics within metastable states in a mean-field spin glass. *Journal of Physics A: Mathematical and General*, 29(5):L81–L87, 1996. doi: 10.1088/0305-4470/29/5/001.

[20] A. Barrat, S. Franz, and G. Parisi. Temperature evolution and bifurcations of metastable states in mean-field spin glasses, with connections with structural glasses. *Journal of Physics A: Mathematical and General*, 30(16):5593–5612, 1997. doi: 10.1088/0305-4470/30/16/006.

[21] Roland Bauerschmidt and Thierry Bodineau. A very simple proof of the lsi for high temperature spin systems. *Journal of Functional Analysis*, 276(8):2582–2588, 2019. doi: 10.1016/j.jfa.2019.01.007.

[22] David Belius and Marius A. Schmidt. Complexity of local maxima of given radial derivative for mixed p-spin Hamiltonians, 2022. URL https://arxiv.org/abs/2207.14361v1.

[23] Gérard Ben Arous and Alice Guionnet. Large deviations for langevin spin glass dynamics. *Probability Theory and Related Fields*, 102(4):455–509, 1995. doi: 10.1007/BF01198846.

[24] Gérard Ben Arous and Alice Guionnet. Symmetric langevin spin glass dynamics. *The Annals of Probability*, 25(3):1367–1422, 1997. doi: 10.1214/aop/1024404517.

[25] Gérard Ben Arous and Aukosh Jagannath. Spectral gap estimates in mean field spin glasses. *Communications in Mathematical Physics*, 361(1):1–52, 2018. doi: 10.1007/s00220-018-3152-6.

[26] Gérard Ben Arous and Aukosh Jagannath. Shattering versus metastability in spin glasses. *Communications on Pure and Applied Mathematics*, 77(1):139–176, 2024. doi: 10.1002/cpa.22133.

[27] Gérard Ben Arous, Amir Dembo, and Alice Guionnet. Aging of spherical spin glasses. *Probability Theory and Related Fields*, 120(1):1–67, 2001. doi: 10.1007/PL00008774.

[28] Gérard Ben Arous, Amir Dembo, and Alice Guionnet. Cugliandolo–kurchan equations for dynamics of spin-glasses. *Probability Theory and Related Fields*, 136(4):619–660, 2006. doi: 10.1007/s00440-005-0491-y.

[29] Gérard Ben Arous, Reza Gheissari, and Aukosh Jagannath. Bounding flows for spherical spin glass dynamics. *Communications in Mathematical Physics*, 373:1011–1048, 2020. doi: 10.1007/s00220-019-03649-4.

[30] Gérard Ben Arous, Eliran Subag, and Ofer Zeitouni. Geometry and temperature chaos in mixed spherical spin glasses at low temperature: the perturbative regime. *Communications on Pure and Applied Mathematics*, 73(8):1732–1828, 2020. doi: 10.1002/cpa.21875. URL `https://arxiv.org/abs/1804.10573v1`. ArXiv version 1.

[31] Rajendra Bhatia and Chandler Davis. A better bound on the variance. *The American Mathematical Monthly*, 107(4):353–357, 2000. doi: 10.1080/00029890.2000.12005203.

[32] G. Biroli. Dynamical TAP approach to mean field glassy systems. *Journal of Physics A: Mathematical and General*, 32(48):8365–8388, 1999. doi: 10.1088/0305-4470/32/48/301.

[33] G. Biroli and J. Kurchan. Metastable states in glassy systems. *Physical Review E*, 64:016101, 2001. doi: 10.1103/PhysRevE.64.016101.

[34] Anton Bovier and Frank den Hollander. *Metastability: A Potential-Theoretic Approach*, volume 351 of *Grundlehren der mathematischen Wissenschaften*. Springer, 2015. doi: 10.1007/978-3-319-24777-9.

[35] Anton Bovier, Michael Eckhoff, Véronique Gayrard, and Markus Klein. Metastability and small eigenvalues in markov chains. *Journal of Physics A: Mathematical and General*, 33(46):L447–L451, 2000. doi: 10.1088/0305-4470/33/46/102.

[36] Anton Bovier, Michael Eckhoff, Véronique Gayrard, and Markus Klein. Metastability in stochastic dynamics of disordered mean-field models. *Probability Theory and Related Fields*, 119(1):99–161, 2001. doi: 10.1007/PL00012740.

[37] Anton Bovier, Michael Eckhoff, Véronique Gayrard, and Markus Klein. Metastability and low lying spectra in reversible markov chains. *Communications in Mathematical Physics*, 228(2):219–255, 2002. doi: 10.1007/s002200200609.

[38] Anton Bovier, Michael Eckhoff, Véronique Gayrard, and Markus Klein. Metastability in reversible diffusion processes i. sharp asymptotics for capacities and exit times. *Journal of the European Mathematical Society*, 6(4):399–424, 2004. doi: 10.4171/JEMS/14.

[39] Anton Bovier, Véronique Gayrard, and Markus Klein. Metastability in reversible diffusion processes ii. precise asymptotics for small eigenvalues. *Journal of the European Mathematical Society*, 7(1):69–99, 2005. doi: 10.4171/JEMS/22.

[40] T. Castellani and A. Cavagna. Spin-glass theory for pedestrians. *Journal of Statistical Mechanics: Theory and Experiment*, 2005(05):P05012, 2005. doi: 10.1088/1742-5468/2005/05/P05012.

[41] A. Cavagna, I. Giardina, and G. Parisi. An investigation of the hidden structure of states in a mean-field spin-glass model. *Journal of Physics A: Mathematical and General*, 30(20):7021–7038, 1997. doi: 10.1088/0305-4470/30/20/009.

[42] A. Cavagna, I. Giardina, and G. Parisi. Barriers between metastable states in the p-spin spherical model, 1997. URL `https://arxiv.org/abs/cond-mat/9702069`.

[43] A. Cavagna, I. Giardina, and G. Parisi. Stationary points of the thouless-anderson-palmer free energy. *Physical Review B*, 57:11251–11257, 1998. doi: 10.1103/PhysRevB.57.11251.

[44] Patrick Charbonneau, Giampaolo Folena, Enrico M. Malatesta, Tommaso Rizzo, and Francesco Zamponi. Rare trajectories in a prototypical mean-field disordered model: Insights into landscape and instantons. *Physical Review E*, 113(3):034107, 2026. doi: 10.1103/f4dj-d2sh. URL `https://arxiv.org/abs/2505.00107v3`. Article 034107; arXiv version 3, February 9, 2026.

[45] M. Chatzakou and Y. Sarantopoulos. Bernstein and Markov-type inequalities for polynomials on $L_p(\mu)$ spaces. *Dolomites Research Notes on Approximation*, 12:16–28, 2019. doi: 10.14658/pupj-drna-2019-Special_Issue-4.

[46] Wei-Kuo Chen. The Aizenman–Sims–Starr scheme and Parisi formula for mixed p-spin spherical models. *Electronic Journal of Probability*, 18(94):1–14, 2013. doi: 10.1214/EJP.v18-2580. URL `https://arxiv.org/abs/1204.5115v1`. ArXiv version 1.

[47] Wei-Kuo Chen and Dmitry Panchenko. On the TAP free energy in the mixed p-spin models. *Communications in Mathematical Physics*, 362(1):219–252, 2018. doi: 10.1007/s00220-018-3143-7.

[48] Wei-Kuo Chen and Arnab Sen. Parisi formula, disorder chaos and fluctuation for the ground state energy in the spherical mixed p-spin models. *Communications in Mathematical Physics*, 350(1):129–173, 2017. doi: 10.1007/s00220-016-2808-3. URL `https://arxiv.org/abs/1512.08492v2`. ArXiv version 2.

[49] Wei-Kuo Chen, Dmitry Panchenko, and Eliran Subag. The generalized TAP free energy II. *Communications in Mathematical Physics*, 381(1):257–291, 2021. doi: 10.1007/s00220-020-03887-x.

[50] Wei-Kuo Chen, Dmitry Panchenko, and Eliran Subag. Generalized TAP free energy. *Communications on Pure and Applied Mathematics*, 76(7):1329–1415, 2023. doi: 10.1002/cpa.22040.

[51] Vincenzo Citro and Federico Ricci-Tersenghi. Strong ergodicity breaking in dynamical mean-field equations for mixed p-spin glasses. *Physical Review Letters*, 135(24): 247102, 2025. doi: 10.1103/lvmp-pydk. URL `https://arxiv.org/abs/2504.12367v3`. Article 247102; arXiv version 3, December 10, 2025.

[52] A. Crisanti and H.-J. Sommers. The spherical p-spin interaction spin glass model: The statics. *Zeitschrift für Physik B Condensed Matter*, 87:341–354, 1992. doi: 10.1007/BF01309287.

[53] A. Crisanti and H.-J. Sommers. Thouless–Anderson–Palmer approach to the spherical p-spin spin glass model. *Journal de Physique I*, 5(7):805–813, 1995. doi: 10.1051/jp1:1995164.

[54] A. Crisanti, H. Horner, and H.-J. Sommers. The spherical p-spin interaction spin-glass model: The dynamics. *Zeitschrift für Physik B Condensed Matter*, 92:257–271, 1993. doi: 10.1007/BF01312184.

[55] Imre Csiszár. Information-type measures of difference of probability distributions and indirect observations. *Studia Scientiarum Mathematicarum Hungarica*, 2:299–318, 1967.

[56] L. F. Cugliandolo and J. Kurchan. Analytical solution of the off-equilibrium dynamics of a long-range spin-glass model. *Physical Review Letters*, 71:173–176, 1993. doi: 10.1103/PhysRevLett.71.173.

[57] Amir Dembo and Eliran Subag. Dynamics for spherical spin glasses: disorder dependent initial conditions. *Journal of Statistical Physics*, 181(2):465–514, 2020. doi: 10.1007/s10955-020-02587-z.

[58] Amir Dembo and Eliran Subag. Dynamics for spherical spin glasses: Gibbs distributed initial conditions, 2025. URL `https://arxiv.org/abs/2503.23342v3`. Version 3, revised July 8, 2026.

[59] Amir Dembo and Eliran Subag. Disordered Gibbs measures and Gaussian conditioning. *Probability Theory and Related Fields*, 2026. doi: 10.1007/s00440-026-01526-5. URL `https://arxiv.org/abs/2409.19453v3`. Published online August 22, 2026; arXiv version 3.

[60] Amir Dembo, Alice Guionnet, and Christian Mazza. Limiting dynamics for spherical models of spin glasses at high temperature. *Journal of Statistical Physics*, 128(4): 847–881, 2007. doi: 10.1007/s10955-006-9239-z.

[61] Amir Dembo, Eyal Lubetzky, and Ofer Zeitouni. Universality for langevin-like spin glass dynamics. *The Annals of Applied Probability*, 31(6):2864–2880, 2021. doi: 10.1214/21-AAP1665.

[62] DLMF. NIST digital library of mathematical functions. `https://dlmf.nist.gov/`, Release 1.2.8 of 2026-09-15, 2026. Section 18.5(iii), equation (18.5.10).

[63] L. N. Dovbysh and V. N. Sudakov. On the Gram–de Finetti matrices. *Zapiski Nauchnykh Seminarov LOMI*, 119:77–86, 1982. URL `https://www.mathnet.ru/eng/znsl3986`.

[64] Semyon Dyatlov. Lecture notes for 18.155: distributions, elliptic regularity, and applications to PDEs, 2026. URL `https://math.mit.edu/~dyatlov/18.155/155-notes.pdf`. Version dated October 2, 2026.

[65] Ahmed El Alaoui, Andrea Montanari, and Mark Sellke. Shattering in pure spherical spin glasses. *Communications in Mathematical Physics*, 406(5):111, 2025. doi: 10.1007/s00220-025-05243-3. URL `https://arxiv.org/abs/2307.04659v4`. Article 111; arXiv version 4.

[66] Ahmad El Soufi, Evans M. Harrell, II, Saïd Ilias, and Joachim Stubbe. On sums of eigenvalues of elliptic operators on manifolds. *Journal of Spectral Theory*, 7(4): 985–1022, 2017. doi: 10.4171/JST/183.

[67] Ronen Eldan, Frederic Koehler, and Ofer Zeitouni. A spectral condition for spectral gap: fast mixing in high-temperature ising models. *Probability Theory and Related Fields*, 182(3-4):1035–1051, 2022. doi: 10.1007/s00440-021-01085-x.

[68] Rida T. Farouki. The Bernstein polynomial basis: A centennial retrospective. *Computer Aided Geometric Design*, 29(6):379–419, 2012. doi: 10.1016/j.cagd.2012.03.001.

[69] Herbert Federer. *Geometric Measure Theory*, volume 153 of *Grundlehren der mathematischen Wissenschaften*. Springer, 1969. doi: 10.1007/978-3-642-62010-2. Reprinted in Classics in Mathematics, 1996.

[70] G. Folena, S. Franz, and F. Ricci-Tersenghi. Rethinking mean-field glassy dynamics and its relation with the energy landscape: The surprising case of the spherical mixed p-spin model. *Physical Review X*, 10:031045, 2020. doi: 10.1103/PhysRevX.10.031045.

[71] Giampaolo Folena and Francesco Zamponi. On weak ergodicity breaking in mean-field spin glasses. *SciPost Physics*, 15(3):109, 2023. doi: 10.21468/SciPostPhys.15.3.109. URL `https://arxiv.org/abs/2303.00026v3`. Article 109; arXiv version 3, June 10, 2023.

[72] S. Franz. First steps of a nucleation theory in disordered systems. *Journal of Statistical Mechanics: Theory and Experiment*, 2005(04):P04001, 2005. doi: 10.1088/1742-5468/2005/04/P04001.

[73] S. Franz and G. Parisi. Recipes for metastable states in spin glasses. *Journal de Physique I*, 5(11):1401–1415, 1995. doi: 10.1051/jp1:1995201.

[74] S. Franz and G. Parisi. Phase diagram of coupled glassy systems: A mean-field study. *Physical Review Letters*, 79:2486–2489, 1997. doi: 10.1103/PhysRevLett.79.2486.

[75] Mark I. Freidlin and Alexander D. Wentzell. *Random Perturbations of Dynamical Systems*, volume 260 of *Grundlehren der mathematischen Wissenschaften*. Springer, 3 edition, 2012. doi: 10.1007/978-3-642-25847-3.

[76] Masatoshi Fukushima, Yoichi Oshima, and Masayoshi Takeda. *Dirichlet Forms and Symmetric Markov Processes*, volume 19 of *De Gruyter Studies in Mathematics*. De Gruyter, 2 edition, 2011. doi: 10.1515/9783110218091.

[77] David Gamarnik and Aukosh Jagannath. The overlap gap property and approximate message passing algorithms for p-spin models. *The Annals of Probability*, 49(1):180–205, 2021. doi: 10.1214/20-AOP1448.

[78] David Gamarnik, Aukosh Jagannath, and Alexander S. Wein. Hardness of random optimization problems for boolean circuits, low-degree polynomials, and langevin dynamics. *SIAM Journal on Computing*, 53(1):1–46, 2024. doi: 10.1137/22M150263X.

[79] Bernard Gaveau and Lawrence S. Schulman. Theory of nonequilibrium first-order phase transitions for stochastic dynamics. *Journal of Mathematical Physics*, 39(3):1517–1533, 1998. doi: 10.1063/1.532394.

[80] Hans Gebelein. Das statistische Problem der Korrelation als Variations- und Eigenwertproblem und sein Zusammenhang mit der Ausgleichsrechnung. *Zeitschrift für Angewandte Mathematik und Mechanik*, 21(6):364–379, 1941. doi: 10.1002/zamm.19410210604.

[81] Reza Gheissari and Aukosh Jagannath. On the spectral gap of spherical spin glass dynamics. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 55(2):756–776, 2019. doi: 10.1214/18-AIHP897.

[82] Reza Gheissari, Curtis Grant, and Tianmin Yu. Mixing times of langevin dynamics for spiked matrix models, 2026. URL `https://arxiv.org/abs/2604.20008v2`. Version 2.

[83] Stefano Ghirlanda and Francesco Guerra. General properties of overlap probability distributions in disordered spin systems. towards Parisi ultrametricity. *Journal of*

Physics A: Mathematical and General, 31(46):9149–9155, 1998. doi: 10.1088/0305-4470/31/46/006.

[84] Leonard Gross. Logarithmic Sobolev inequalities. *American Journal of Mathematics*, 97(4):1061–1083, 1975. doi: 10.2307/2373688.

[85] Francesco Guerra and Fabio Lucio Toninelli. Quadratic replica coupling in the Sherrington–Kirkpatrick mean field spin glass model. *Journal of Mathematical Physics*, 43(7):3704–3716, 2002. doi: 10.1063/1.1483378. URL https://arxiv.org/abs/cond-mat/0201091v2. ArXiv version 2, Section 3.

[86] Evans M. Harrell, II and Joachim Stubbe. On sums of graph eigenvalues. *Linear Algebra and its Applications*, 455:168–186, 2014. doi: 10.1016/j.laa.2014.05.001. Corrigendum: Linear Algebra and its Applications 458 (2014), 699–700, https://doi.org/10.1016/j.laa.2014.06.020.

[87] Bernard Helffer and Johannes Sjöstrand. Multiple wells in the semi-classical limit i. *Communications in Partial Differential Equations*, 9(4):337–408, 1984. doi: 10.1080/03605308408820335.

[88] Bernard Helffer, Markus Klein, and Francis Nier. Quantitative analysis of metastability in reversible diffusion processes via a witten complex approach. *Matemática Contemporânea*, 26:41–85, 2004. doi: 10.21711/231766362004/rmc265.

[89] Alan J. Hoffman and Helmut W. Wielandt. The variation of the spectrum of a normal matrix. *Duke Mathematical Journal*, 20(1):37–39, 1953. doi: 10.1215/S0012-7094-53-02004-3.

[90] Brice Huang and Mark Sellke. A constructive proof of the spherical parisi formula, 2024. URL https://arxiv.org/abs/2311.15495v3. Version 3; first posted in 2023.

[91] Brice Huang and Mark Sellke. Tight lipschitz hardness for optimizing mean field spin glasses. *Communications on Pure and Applied Mathematics*, 78(1):60–119, 2025. doi: 10.1002/cpa.22222.

[92] Brice Huang and Mark Sellke. Strong topological trivialization of multi-species spherical spin glasses. *The Annals of Probability*, 54(2):1034–1107, 2026. doi: 10.1214/25-AOP1785.

[93] Brice Huang, Andrea Montanari, and Huy Tuan Pham. Sampling from spherical spin glasses in total variation via algorithmic stochastic localization, 2024. URL https://arxiv.org/abs/2404.15651v1.

[94] Brice Huang, Sidhanth Mohanty, Amit Rajaraman, and David X. Wu. Weak poincaré inequalities, simulated annealing, and sampling from spherical spin glasses. In *Proceedings of the 57th Annual ACM Symposium on Theory of Computing*, pages 915–923. ACM, 2025. doi: 10.1145/3717823.3718236. URL https://arxiv.org/abs/2411.09075v2. ArXiv version 2.

[95] Aukosh Jagannath. Dynamics of mean field spin glasses on short and long timescales. *Journal of Mathematical Physics*, 60(8):083305, 2019. doi: 10.1063/1.5094173.

[96] Aukosh Jagannath and Ian Tobasco. Low temperature asymptotics of spherical mean field spin glasses. *Communications in Mathematical Physics*, 352(3):979–1017, 2017. doi: 10.1007/s00220-017-2864-3.

[97] Aukosh Jagannath and Ian Tobasco. Bounds on the complexity of replica symmetry breaking for spherical spin glasses. *Proceedings of the American Mathematical Society*, 146(7):3127–3142, 2018. doi: 10.1090/proc/13875.

[98] Alexander S. Kechris. *Classical Descriptive Set Theory*, volume 156 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1995. doi: 10.1007/978-1-4612-4190-4.

[99] Taegyun Kim. Non-shattering at and above the dynamical temperature in the spherical pure p-spin model, 2026. URL `https://arxiv.org/abs/2608.14369v2`. Version 2, revised August 21, 2026.

[100] T. R. Kirkpatrick and D. Thirumalai. Dynamics of the structural glass transition and the p-spin–interaction spin-glass model. *Physical Review Letters*, 58:2091–2094, 1987. doi: 10.1103/PhysRevLett.58.2091.

[101] T. R. Kirkpatrick and P. G. Wolynes. Stable and metastable states in mean-field potts and structural glasses. *Physical Review B*, 36:8552–8564, 1987. doi: 10.1103/PhysRevB.36.8552.

[102] T. R. Kirkpatrick, D. Thirumalai, and P. G. Wolynes. Scaling concepts for the dynamics of viscous liquids near an ideal glassy state. *Physical Review A*, 40:1045–1054, 1989. doi: 10.1103/PhysRevA.40.1045.

[103] Anouar Kouraich and Simone Warzel. Lower bound on the mixing time of p-spin glasses, 2026. URL `https://arxiv.org/abs/2605.14681v1`.

[104] Pawel Kröger. Upper bounds for the neumann eigenvalues on a bounded domain in euclidean space. *Journal of Functional Analysis*, 106(2):353–357, 1992. doi: 10.1016/0022-1236(92)90052-K.

[105] F. Krzakala and L. Zdeborová. Hiding quiet solutions in random constraint satisfaction problems. *Physical Review Letters*, 102:238701, 2009. doi: 10.1103/PhysRevLett.102.238701.

[106] F. Krzakala, A. Montanari, F. Ricci-Tersenghi, G. Semerjian, and L. Zdeborová. Gibbs states and the set of solutions of random constraint satisfaction problems. *Proceedings of the National Academy of Sciences*, 104(25):10318–10323, 2007. doi: 10.1073/pnas.0703685104.

[107] J. Kurchan and L. Laloux. Phase space geometry and slow dynamics. *Journal of Physics A: Mathematical and General*, 29(9):1929–1948, 1996. doi: 10.1088/0305-4470/29/9/009.

[108] J. Kurchan, G. Parisi, and M. A. Virasoro. Barriers and metastable states as saddle points in the replica approach. *Journal de Physique I*, 3(8):1819–1838, 1993. doi: 10.1051/jp1:1993217.

[109] Johannes Lang, Subir Sachdev, and Sebastian Diehl. Numerical renormalization of glassy dynamics. *Physical Review Letters*, 135(24):247101, 2025. doi: 10.1103/z64g-nqs6. URL `https://arxiv.org/abs/2504.06849v2`. Article 247101; arXiv version 2, December 11, 2025.

[110] Johannes Lang, Vincenzo Citro, Luca Leuzzi, and Federico Ricci-Tersenghi. Aging phase diagram and exact asymptotic energies of mixed spherical spin glasses, 2026. URL `https://arxiv.org/abs/2608.20623v1`. Version 1, August 20, 2026.

[111] Michel Ledoux. Semigroup proofs of the isoperimetric inequality in Euclidean and Gauss space. *Bulletin des Sciences Mathématiques*, 118(6):485–510, 1994. Author manuscript.

[112] Michel Ledoux. *The Concentration of Measure Phenomenon*, volume 89 of *Mathematical Surveys and Monographs*. American Mathematical Society, 2001. doi: 10.1090/surv/089.

[113] Erich L. Lehmann and Joseph P. Romano. *Testing Statistical Hypotheses*. Springer, 3 edition, 2005. doi: 10.1007/0-387-27605-X.

[114] A. V. Lopatin and L. B. Ioffe. Barriers in the p-spin interacting spin-glass model: The dynamical approach. *Physical Review Letters*, 84(18):4208–4211, 2000. doi: 10.1103/PhysRevLett.84.4208.

[115] George G. Lorentz. *Bernstein Polynomials*. Chelsea Publishing Company, New York, 2 edition, 1986. ISBN 0828403236.

[116] Léo Mathis and Michele Stecconi. Expectation of a random submanifold: the zonoid section. *Annales Henri Lebesgue*, 7:903–967, 2024. doi: 10.5802/ahl.214. URL `https://www.numdam.org/articles/10.5802/ahl.214/`.

[117] M. Mézard and A. Montanari. Reconstruction on trees and spin glass transition. *Journal of Statistical Physics*, 124:1317–1350, 2006. doi: 10.1007/s10955-006-9162-3.

[118] Paul Milgrom and Ilya Segal. Envelope theorems for arbitrary choice sets. *Econometrica*, 70(2):583–601, 2002. doi: 10.1111/1468-0262.00296.

[119] R. Monasson. Structural glass transition and the entropy of the metastable states. *Physical Review Letters*, 75:2847–2850, 1995. doi: 10.1103/PhysRevLett.75.2847.

[120] Andrea Montanari and Guilhem Semerjian. Rigorous inequalities between length and time scales in glassy systems. *Journal of Statistical Physics*, 125(1):23–54, 2006. doi: 10.1007/s10955-006-9175-y.

[121] Carl E. Mueller and Fred B. Weissler. Hypercontractivity for the heat semigroup for ultraspherical polynomials and on the n-sphere. *Journal of Functional Analysis*, 48(2):252–283, 1982. doi: 10.1016/0022-1236(82)90069-6.

[122] Dmitry Panchenko. On differentiability of the Parisi formula. *Electronic Communications in Probability*, 13:241–247, 2008. doi: 10.1214/ECP.v13-1365.

[123] Dmitry Panchenko. On the Dovbysh–Sudakov representation result. *Electronic Communications in Probability*, 15:330–338, 2010. doi: 10.1214/ECP.v15-1562.

[124] Dmitry Panchenko. The Ghirlanda–Guerra identities for mixed p-spin model. *Comptes Rendus Mathematique*, 348(3-4):189–192, 2010. doi: 10.1016/j.crma.2010.02.004.

[125] Dmitry Panchenko. *The Sherrington–Kirkpatrick Model*. Springer Monographs in Mathematics. Springer, New York, 2013. doi: 10.1007/978-1-4614-6289-7.

[126] Dmitry Panchenko. The Parisi ultrametricity conjecture. *Annals of Mathematics*, 177 (1):383–393, 2013. doi: 10.4007/annals.2013.177.1.8.

[127] Yury Polyanskiy and Yihong Wu. *Information Theory: From Coding to Learning*. Cambridge University Press, 2025. doi: 10.1017/9781108966351. URL `https://people.lids.mit.edu/yp/homepage/data/itbook-export.pdf`. Author draft dated August 16, 2024.

[128] Hervé Queffélec and Rachid Zarouf. On Bernstein's inequality for polynomials. *Analysis and Mathematical Physics*, 9(3):1181–1207, 2019. doi: 10.1007/s13324-019-00294-x. URL `https://arxiv.org/abs/1903.10801v1`. ArXiv:1903.10801v1.

[129] Robert A. Rankin. The closest packing of spherical caps in n dimensions. *Proceedings of the Glasgow Mathematical Association*, 2(3):139–144, 1955. doi: 10.1017/S2040618500033219.

[130] T. Rizzo. Path integral approach unveils role of complex energy landscape for activated dynamics of glassy systems. *Physical Review B*, 104:094203, 2021. doi: 10.1103/PhysRevB.104.094203.

[131] R. Tyrrell Rockafellar. *Convex Analysis*, volume 28 of *Princeton Mathematical Series*. Princeton University Press, 1970. doi: 10.1515/9781400873173.

[132] Valentina Ros, Giulio Biroli, and Chiara Cammarota. Complexity of energy barriers in mean-field glassy systems. *EPL (Europhysics Letters)*, 126(2):20003, 2019. doi: 10.1209/0295-5075/126/20003. Article 20003.

[133] Valentina Ros, Giulio Biroli, and Chiara Cammarota. Dynamical instantons and activated processes in mean-field glass models. *SciPost Physics*, 10(1):002, 2021. doi: 10.21468/SciPostPhys.10.1.002. URL `https://arxiv.org/abs/2006.08399v3`. Article 002; arXiv version 3, December 15, 2020.

[134] Arthur Sard. The measure of the critical values of differentiable maps. *Bulletin of the American Mathematical Society*, 48(12):883–890, 1942. doi: 10.1090/S0002-9904-1942-07811-6.

[135] Issai Schur. Bemerkungen zur theorie der beschränkten bilinearformen mit unendlich vielen veränderlichen. *Journal für die reine und angewandte Mathematik*, 140: 1–28, 1911. doi: 10.1515/crll.1911.140.1. URL https://eudml.org/doc/149352.

[136] Mark Sellke. The threshold energy of low temperature langevin dynamics for pure spherical spin glasses. *Communications on Pure and Applied Mathematics*, 77(11): 4065–4099, 2024. doi: 10.1002/cpa.22197.

[137] Mark Sellke. Exponentially slow mixing of the low temperature sk model, 2025. URL https://arxiv.org/abs/2511.22621v2. Version 2.

[138] Mark Sellke. On marginal stability in low temperature spherical spin glasses. *Communications in Mathematical Physics*, 406(9):227, 2025. doi: 10.1007/s00220-025-05410-6. URL https://arxiv.org/abs/2409.15728v4. Article 227; arXiv version 4, August 23, 2026.

[139] Michele Stecconi. Kac–rice formula for transverse intersections. *Analysis and Mathematical Physics*, 12(2):44, 2022. doi: 10.1007/s13324-022-00654-0. URL https://arxiv.org/abs/2103.10853v1. Article 44; arXiv version 1, Theorem 29, Section 4.

[140] Eliran Subag. The complexity of spherical p-spin models: a second moment approach. *The Annals of Probability*, 45(5):3385–3450, 2017. doi: 10.1214/16-AOP1139.

[141] Eliran Subag. The geometry of the Gibbs measure of pure spherical spin glasses. *Inventiones Mathematicae*, 210(1):135–209, 2017. doi: 10.1007/s00222-017-0726-4. URL https://arxiv.org/abs/1604.00679v2. ArXiv version 2.

[142] Eliran Subag. The free energy of spherical pure p-spin models: computation from the TAP approach. *Probability Theory and Related Fields*, 186(3–4):715–734, 2023. doi: 10.1007/s00440-023-01200-0.

[143] Eliran Subag. Free energy landscapes in spherical spin glasses. *Duke Mathematical Journal*, 173(7):1291–1357, 2024. doi: 10.1215/00127094-2023-0033. URL https://arxiv.org/abs/1804.10576v6. ArXiv version 6.

[144] Eliran Subag and Ofer Zeitouni. The extremal process of critical points of the pure p-spin spherical spin glass model. *Probability Theory and Related Fields*, 168(3-4): 773–820, 2017. doi: 10.1007/s00440-016-0724-2. URL https://arxiv.org/abs/1509.03098v2. ArXiv version 2.

[145] Y. Sun, A. Crisanti, F. Krzakala, L. Leuzzi, and L. Zdeborová. Following states in temperature in the spherical $s + p$-spin glass model. *Journal of Statistical Mechanics: Theory and Experiment*, 2012(07):P07002, 2012. doi: 10.1088/1742-5468/2012/07/P07002.

[146] Michel Talagrand. Free energy of the spherical mean field model. *Probability Theory and Related Fields*, 134(3):339–382, 2006. doi: 10.1007/s00440-005-0433-8.

[147] Michel Talagrand. *Mean Field Models for Spin Glasses. Volume I: Basic Examples*, volume 54 of *Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge*. Springer, Berlin, Heidelberg, 2011. doi: 10.1007/978-3-642-15202-3. URL `https://michel.talagrand.net/challenge/volume1.pdf`. Author manuscript dated September 14, 2010.

[148] S. Tănase-Nicola and J. Kurchan. Metastable states, transitions, basins and borders at finite temperatures. *Journal of Statistical Physics*, 116:1201–1245, 2004. doi: 10.1023/B:JOSS.0000041739.53068.6a.

[149] D. J. Thouless, P. W. Anderson, and R. G. Palmer. Solution of 'Solvable model of a spin glass'. *Philosophical Magazine*, 35(3):593–601, 1977. doi: 10.1080/14786437708235992.

[150] C. Unger and W. Klein. Nucleation theory near the classical spinodal. *Physical Review B*, 29:2698–2708, 1984. doi: 10.1103/PhysRevB.29.2698.

[151] S. R. S. Varadhan. Stochastic processes, Chapter 4: Stochastic differential equations, 2000. URL `https://math.nyu.edu/~varadhan/processes/chapter4.pdf`. Lecture notes, Courant Institute, fall 2000; author-hosted chapter.

[152] Martin J. Wainwright and Michael I. Jordan. Graphical models, exponential families, and variational inference. *Foundations and Trends in Machine Learning*, 1(1–2):1–305, 2008. doi: 10.1561/2200000001.

[153] Edward Witten. Supersymmetry and morse theory. *Journal of Differential Geometry*, 17(4):661–692, 1982. doi: 10.4310/jdg/1214437492.

[154] Hao Xu and Haoran Yang. Hessian spectrum at the global minimum of the spherical pure-like mixed p-spin glasses. *Journal of Mathematical Physics*, 66(9):093301, 2025. doi: 10.1063/5.0251208. URL `https://arxiv.org/abs/2408.15849v3`. ArXiv version 3.