

hexagon:2610.00019v1

The Negative Spherical Perceptron

*Capacity, Isostatic Jamming, and
Full Replica Symmetry Breaking*

P. M. Aronow

Patrick Lopatto

AUTHORS' NOTE

This manuscript is the result of a methodological experiment in using large language models to develop the rigorous mathematical theory of the negative perceptron, which has received significant attention in the physics literature as an accessible model for studying jamming and related phenomena.

With the exception of this note, the entire document is machine written. We have taken care to supervise this writing to ensure that the proofs meet a minimum standard of readability and that previous literature is appropriately cited. However, the resulting document still leaves much to be desired in terms of exposition and overall coherence. We are nonetheless releasing this manuscript on Hexagon to disseminate the results as quickly as possible, because we believe they may be of interest to the mathematics and physics communities.

We plan to take some time to write the results in this manuscript carefully for publication. While we focus on writing, we do not intend to produce further results for the negative perceptron. However, we invite readers to do so, because we believe there is much more to be done. Notably, it remains to establish the scaling solution of Franz, Parisi, Sevelev, Urbani, and Zamponi [59] for the Parisi minimizers as $\kappa \uparrow \kappa_c$ and the predictions that follow from it. We welcome emails from anyone who obtains such results and will update the references accordingly. We also welcome suggestions concerning mathematical corrections, omitted citations, and attribution of ideas.

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

Contents

Notation	vi
PART I Parisi formula, capacity, and isostaticity for the negative spherical perceptron	1
1 Introduction	2
1.1 Model and the Parisi functional	5
1.2 Main results	7
1.3 Ideas of the proof.	13
2 The Parisi functional	20
2.1 Order parameters and the Cole–Hopf recursion	21
2.2 Continuity in the order parameter for smooth activations	25
2.3 The hard-wall functional at a finite cutoff	30
3 The upper bound	37
3.1 A change of measure on a marked cascade	38
3.2 Gaussian cascades	41
3.3 The interpolation and its derivatives	46
3.4 Concentration and identities at a contact	52
3.5 Synchronization	55
3.6 The spherical term	56
3.7 The comparison	59
3.8 Hard constraints	61
3.9 Volume from one feasible point	63
4 The lower bound	67
4.1 Spherical coordinates, the row count, and the three systems	68
4.2 The overlap representation theorem	71
4.3 Perturbations, continuity, and the precision identity	76
4.4 The two cavity terms	83
4.5 The cavity quantities on cascades	88
4.6 Rotation invariance	90
4.7 Proof of the lower bound	91
4.8 Proofs of the perturbation, continuity, and precision lemmas.	95

5	Proof of the Parisi formula	109
5.1	Mollification	109
5.2	Truncation near $q = 1$	110
5.3	The two bounds and attainment of the infimum	111
6	The variational problem at hard margins	114
6.1	A value identity and existence of minimizers	116
6.2	The slope in the margin.	120
6.3	Weak top stability and finiteness of the critical margin.	121
6.4	Divergence at the critical margin	125
6.5	Pinning of the top and collapse of minimizers.	127
6.6	Proof of the theorem on the variational problem.	128
7	The critical margin near $\alpha = 2$	129
8	From soft free energy to hard feasible volume	135
8.1	Repairing a small violation vector.	136
8.2	Transferring volume by a tangent Gaussian perturbation	139
8.3	The hard-volume lower bound	142
8.4	Proof of the capacity theorem	144
9	Gap statistics	146
9.1	Contacts at optimal configurations	147
9.2	The finite recursion and the Parisi chain.	150
9.3	A quadratic source estimate.	153
9.4	Gap averages on sets of positive volume	156
9.5	Gaps of typical feasible configurations	158
9.6	Configurations near the optimal margin.	159
9.7	Gaps of nearly optimal configurations.	161
PART II	Full replica symmetry breaking in the negative spherical perceptron	165
1	Introduction	166
1.1	The Parisi variational problem	169
1.2	Main results	171
1.3	Ideas of the proof.	176
2	The Parisi functional and its optimality conditions	181
2.1	Gaussian functions and Mills ratios	182
2.2	Order parameters and the finite-cutoff problem	185
2.3	Regularity, constant blocks, and the diffusion	187
2.4	Generator identities and the sandwich inequality	193
2.5	First variation and the obstacle conditions.	197
2.6	The value identity, the slope in the margin, and weak top stability	201

2.7	The critical margin and existence of minimizers	203
2.8	The zero-temperature functional	204
2.9	Logarithmic divergence at the critical margin	209
3	Replica-symmetric order parameters, the de Almeida–Thouless line, and the ordering of thresholds	212
3.1	Gardner’s prediction near $\alpha = 2$	212
3.2	Replica-symmetric order parameters	213
3.3	The stationary curve and the crossing	215
3.4	The zero-temperature functional and a dual certificate	222
3.5	Ordering of the thresholds	224
4	Rescaled variables and a priori bounds for minimizers	229
4.1	Rescaled variables	229
4.2	The value formula and the exact decomposition	231
4.3	The stationary bottom	236
4.4	A priori bounds for global minimizers	238
5	Bounded depth of minimizers	241
5.1	The balance law	242
5.2	Signed-area cancellation	243
5.3	Proof of the bounded-depth theorem	244
6	Compactness of minimizers	247
6.1	Conventions and inputs	248
6.2	Approximation by step profiles at fixed e	249
6.3	The jamming identity	250
6.4	The jamming coefficient near $\alpha = 2$	254
6.5	The mass lower bound	257
6.6	The top-jump bound	258
6.7	The top-gap bound	261
6.8	Proof that minimizers lie in a compact class	262
7	The flat limit	263
7.1	Compact classes and the flat functional	263
7.2	The compact-class expansion	265
7.3	The inverse parametrization	266
7.4	Monotonicity of the flat density	268
7.5	Convexity along ordered rays	271
7.6	The first variation at the replica-symmetric profile	273
7.7	Uniqueness of the flat minimizer	276
7.8	Stability at the crossing	278

8	Uniqueness at the crossing and full replica symmetry breaking	280
8.1	KKT points and the local curvature	281
8.2	The replica-symmetric obstacle at the crossing.	283
8.3	Natural-scale estimates in the crossing window	284
8.4	Support localization in a window of margins	285
8.5	Window rigidity	289
8.6	A uniform window around the crossing.	291
9	Proofs of the main theorems	294
9.1	The logical order	294
9.2	Proof of the threshold theorem	295
9.3	Proof of the uniform compactness theorem	295
9.4	Proof of the flat-limit theorem.	295
9.5	Proof of the full replica symmetry breaking theorem	295
APPENDICES TO PART II		
A	Regularity of the Parisi equation	297
A.1	One Gaussian step	298
A.2	Step order parameters	299
A.3	Continuity in the order parameter and the general case	301
A.4	Itô's formula along the diffusion	303
A.5	The control representation	304
A.6	The expansion at $-\infty$	305
B	Estimates for the top-jump bound	307
B.1	Gaussian facts	307
B.2	The last constant piece in scale-free form	310
B.3	Uniform bounds on the scale-free functions	310
B.4	Proof of the breakpoint bound	314
C	Proofs for the flat limit	318
C.1	Tails of V and its derivatives	318
C.2	Dilation and uniform tails	319
C.3	Proof of the compact-class expansion	322
C.4	Differentiability along ordered rays and the second variation	324
D	Proofs for the local analysis at the crossing	327
D.1	Flat constants.	327
D.2	The replica-symmetric kernel, obstacle, and family	328
D.3	Natural-scale continuity	334
D.4	The bottom comparison.	337
	References	342

Notation

Shared notation. Both Parts use the following. G is a standard Gaussian variable, ϕ and Φ are its density and distribution function, $R = \phi/\Phi$, $\bar{R}(z) = z + R(z)$, $V = R\bar{R}$, $m(k) = \mathbb{E}(k - G)_+^2$, and $m_1(k) = \mathbb{E}(k - G)_+$. An order parameter is a right-continuous nondecreasing $\gamma : [0, 1] \rightarrow [0, 1]$ with $\gamma = 1$ near one; μ_γ is the measure with distribution function γ , Q_γ (often written Q) is the top of its support, $\delta = 1 - Q_\gamma$, and $\lambda_\gamma(q) = \int_q^1 \gamma$; $\mathbf{1}_{[q,1]}$ is a replica-symmetric order parameter. $\mathcal{U}$ and $\mathcal{U}_r$ are the classes of order parameters, $\mathcal{P}_\kappa$ is the Parisi functional, $\mathcal{P}_*(\kappa)$ its infimum (the *Parisi value* in Part I), $\mathcal{P}_\kappa^{\text{RS}}(q) = \mathcal{P}_\kappa(\mathbf{1}_{[q,1]})$ Gardner's replica-symmetric functional, u the solution of the Parisi equation, X the Parisi diffusion, D , S , and $\mathcal{H}_\gamma$ the functions of the first-order conditions, $\mathcal{T}_{m,s}$ one Gaussian step of the Parisi recursion, m_j the values of a step order parameter, $K = (1 - r)^{-1}$ the constant of the class $\mathcal{U}_r$, and $b = \partial_x u$. The margins are the critical margin κ_c , Gardner's prediction κ_{RS} , and the de Almeida–Thouless margin κ_d ; near $\alpha = 2$ we write $\alpha = 2 + e$. The *free energy* is $N^{-1} \log Z_N(U)$ for a soft activation U (Part I) and $N^{-1} \log V_N(\kappa)$ at a hard margin κ . In Part II the *rescaled profile* of γ is $\zeta(s) = \gamma(1 - e^2 s)$.

The table lists the letters whose meaning differs between the two Parts. Many letters, for example $\tau, \vartheta, \rho, Y, \varepsilon, \eta$, and $c_0, c_1, \dots$, also carry local meanings that are fixed where they appear.

Symbol	Part I	Part II
ℓ	level of the soft penalty $-\beta(\ell - x)_+^2$; width of the top layer near $\alpha = 2$	$\ell(s), \ell_0$: rescaled masses λ_γ/e^2
β	leaves of the cascade tree; strength of the soft penalty and of the source term	$\beta = \sqrt{(1-q)/q}$ for replica-symmetric order parameters; also δ/λ
J	$J_\ell(k) = J^k(\ell^{-1}\mathbf{1}_{[1-\ell,1)})$; the level at which two replicas branch	J^k : the zero-temperature functional; $\mathcal{J}$: the jamming coefficient
C_0	$-\log \Phi(-1)$; the growth constant $2L$ of an activation	a growth constant for terminal data; the constant in the a priori bound $s_0 \leq C_0 e^{-2/3}$
U	the activation	$U = U_\zeta$, the solution in rescaled variables
A	a constant bounding an activation; the pattern matrix	$A(\zeta)$, the entropy term of the flat limit; a function $A(k)$ of the margin
$K, \mathcal{K}$	$\ U''\ _\infty$; the number of atoms of a cascade; an operator-norm constant, C_{op} ; $\mathcal{K}$ a cone or convex set; K_N a polyhedron	K_e , a constant; $K = \gamma/\lambda$; $\mathcal{K}(d_1, S)$, the compact class of profiles
S, s	$S(q) = \alpha \lambda_\gamma(q)^2 \mathbb{E} \partial_{xx} u_\gamma(q, X_q)^2$; s a sparsity level	S_γ , the same; the depth parameter S in $\mathcal{K}(d_1, S)$ and the bound S_* ; $s = (1-q)/e^2$, the rescaled overlap
m	$m_N(\sigma)$, the minimal margin of σ	m_* , a bound on γ below the top of the support
ζ	ζ_l , conditional means of the response overlap; the limiting overlap law	the rescaled profile
b	also the rows $b_a = g_a/\sqrt{N}$ and the top precision b of a cascade	only $b = \partial_x u$

PART I

Parisi formula, capacity, and isostaticity for the negative spherical perceptron

We study the spherical perceptron with Gaussian patterns at negative margins, a mean-field model of jamming in which the set of solutions is not convex. We prove the Parisi formula for its free energy for bounded-above C^2 activations with bounded second derivative and either bounded first derivative or concavity, without any monotonicity assumption on the activation. For every constraint density $\alpha > 2$, we show that the normalized logarithm of the volume of solutions converges in probability to the Parisi value at every margin below a critical margin $\kappa_c(\alpha) < 0$, that this value tends to $-\infty$ continuously as the margin increases to $\kappa_c(\alpha)$, and that the largest achievable margin converges in probability to $\kappa_c(\alpha)$. At sufficiently negative margins, Gardner's replica-symmetric formula for the free energy is exact. For α slightly above two, the limit $\kappa_c(\alpha)$ lies below Gardner's prediction by at least a constant times $(\alpha - 2)^3$; with Part II, this order is exact. At the optimal margin, with probability exponentially close to one, every optimal configuration has exactly N contacts: the jamming point is isostatic. The empirical gap measures of optimal configurations, normalized by N , converge weakly in probability to a deterministic measure with an atom of mass exactly one at zero. At a fixed margin below $\kappa_c(\alpha)$, the gap statistics of all but an exponentially small fraction of the feasible set are given by the Parisi diffusion.

Introduction

The perceptron is the simplest model of a neural network and a basic example of a random constraint satisfaction problem. Given M random patterns $g_1, \dots, g_M \in \mathbb{R}^N$ and a margin $\kappa \in \mathbb{R}$, the spherical perceptron asks for a vector σ on the sphere of radius $\sqrt{N}$ such that

$$\frac{g_a \cdot \sigma}{\sqrt{N}} \geq \kappa \quad \text{for every } a \leq M.$$

Two questions are classical: how large the set of such vectors is, and what the largest achievable margin is when $M/N \rightarrow \alpha$. Gardner and Derrida computed the answers with the replica method, under the assumption of replica symmetry [62, 63]. In particular, Gardner's formula predicts that the largest achievable margin converges to the solution $\kappa_{\text{RS}}(\alpha)$ of

$$\alpha \mathbb{E}(\kappa_{\text{RS}} - G)_+^2 = 1,$$

where G is a standard Gaussian variable.

For $\kappa \geq 0$ each constraint cuts out a geodesically convex spherical cap, and replica symmetry is expected to hold throughout the satisfiable phase. In this regime Gardner's formula for the free energy was proved by Shcherbina and Tirozzi [115] and by Talagrand [128, Chapter 8], and Stojnic gave another proof of the formula for the capacity [118]. For $\kappa < 0$, which is the relevant regime when $\alpha > 2$, each constraint removes a cap smaller than a hemisphere, and the feasible set is a nonconvex intersection of the complements of such caps. Franz and Parisi proposed this *negative spherical perceptron* as the simplest model of jamming [58]. Franz, Parisi, Sevelev, Urbani, and Zamponi argued that its satisfiability threshold belongs to the same universality class as the jamming transition of hard spheres in infinite dimension [59]. The same model describes linear classification with a negative margin, a basic example of an overparametrized learning problem [60, 90].

In the physical picture, replica symmetry is lost at negative margins before the satisfiability threshold is reached. The stability of the replica-symmetric solution of the perceptron was studied in [23, 63]. Close to $\kappa = 0$ replica symmetry is lost at a de Almeida–Thouless line [45], through a continuous transition to full replica symmetry breaking in the sense of Parisi [104, 105], and the whole jamming line lies in the broken phase [58, 59]. The physics literature therefore predicts that Gardner's formula for the critical margin is exact only for $\kappa \geq 0$. At the critical margin, jammed configurations are predicted to be *isostatic*: they are in contact with as many constraints as there are degrees

of freedom. Franz and Parisi predicted that jamming is hypostatic in the convex regime $\kappa > 0$ and isostatic in the nonconvex regime $\kappa < 0$ [58]; for the isostatic counting in jammed packings see [2, 91]. The gaps and contact forces of jammed configurations are predicted to have power-law distributions with universal exponents [29, 30, 58, 59, 107], related to marginal stability [106, 132] and to the jamming of soft spheres [95]. Annesi, Malatesta, and Zamponi computed the satisfiability threshold numerically from the equations of full replica symmetry breaking, and located the region of the phase diagram in which the overlap distribution of typical solutions has connected support [3].

Rigorous results for negative margins have so far been partial. In [118], Stojnic used Gordon’s comparison inequality [66] to show that Gardner’s formula gives an upper bound on the capacity at every margin. He also proved a sharper bound by applying Gordon’s inequality to an exponential of the cost; its improvement over the replica-symmetric bound was shown numerically [117]. El Alaoui and Sellke conjectured that the Parisi variational problem for the hard-margin perceptron, in the form used below, gives the free energy and the capacity [54, Conjecture 1]; they trace this conjecture to [59]. Assuming that some Parisi minimizer γ has no overlap gap, and given oracle access to it, their algorithm constructs vectors of norm close to $\sqrt{Q_\gamma N}$ with a small vector of constraint violations, where $Q_\gamma < 1$ is the top of the support of γ [54, Theorem 1]; see also [55]. Montanari, Zhong, and Zhou proved upper and lower bounds on the capacity that agree to leading order as $\kappa \rightarrow -\infty$, and studied its algorithmic tractability [90]. Their upper bound [90, Theorem 4.2] shows that Gardner’s prediction overestimates the capacity when α is large. Near $\alpha = 2$, where $\kappa_{\text{RS}}(\alpha)$ is close to zero, the cited bounds do not quantify this separation. Stojnic later proposed a characterization of the capacity through fully lifted random duality theory, and computed the resulting values numerically at finite levels of a stationarized lifting [119]. Huang, Sellke, and Sun characterized the empirical distributions of the margins that algorithms with dimension-free Lipschitz dependence on the Gaussian data can achieve [73].

For the Ising perceptron, where $\sigma \in \{-1, 1\}^N$, the capacity predicted by Krauth and Mézard [83] was studied by Ding and Sun and by Huang, who proved lower and upper bounds matching the prediction, subject in part to numerical conditions [48, 72]. Xu proved a sharp threshold for the Ising perceptron with Bernoulli disorder [133]. At small constraint densities, Talagrand proved that the free energy of the Ising perceptron is given by the replica-symmetric formula, for a class of activations that includes the one-sided threshold [123], [127, Chapter 2], [128, Chapter 9]; his treatment of the Shcherbina–Tirozzi model, whose spins are continuous, is in [127, Chapter 3]. Bolthausen, Nakajima, Sun, and Xu gave another proof of this replica-symmetric formula, for a larger class of activations [20]. In the Bayes-optimal (teacher–student) setting, replica-symmetric formulas for perceptron-type generalized linear models were established by Barbier, Krzakala, Macris, Miolane, and Zdeborová [13].

The Parisi formula for mean-field spin glasses was predicted in [104, 105] (see also [31, 88]) and proved by Guerra and Talagrand for the Sherrington–Kirkpatrick model [68, 124]. Talagrand and W.-K. Chen proved the Crisanti–Sommers formula [44] for spherical models, the latter through the Aizenman–Sims–Starr scheme [1, 37, 125], and Panchenko proved the formula for general mixed p -spin models, building on the ultrametricity theorem [97–99]. The perceptron differs from these models in a basic way. Its Hamiltonian $\sum_a U(g_a \cdot \sigma / \sqrt{N})$ is a nonlinear function of Gaussian fields, and it is not itself a Gaussian process in σ . Structurally the model is closer to multi-species or bipartite spin glasses, with one species formed by the coordinates and the other by the constraints. For such models Guerra’s interpolation gives an upper bound when the interaction is convex [14, 16, 100]. The matching lower bound by the Aizenman–Sims–Starr scheme requires synchronization of the overlaps of the species [100, 101], and it needs no convexity, for Ising spins [100] and for spherical spins [16]; see also [15, 17]. The interaction of the bipartite model, which the perceptron resembles, is not convex. For spherical bipartite models see [6, 82, 122]. Mourrat proved that the solution of an infinite-dimensional Hamilton–Jacobi equation bounds the free energy of bipartite models from above, and conjectured that it is the limit [92]; see also [49]. Our upper bound follows the proof of the upper bound in [92]. That work also contains a quantitative synchronization estimate for pairs of overlap arrays, which is an input to both bounds below. Mourrat later proved an upper bound of the same kind for vector spin glasses whose energy is a Gaussian field [94]; the energy $\sum_a U(g_a \cdot \sigma / \sqrt{N})$ is not of this form. For nonconvex vector models, H.-B. Chen and Mourrat showed that the limit of the free energy, if it exists, is a critical value of a Parisi-type functional [34], and H.-B. Chen proved the same for nonconvex multi-species models [33]. H.-B. Chen, Issa, and Mourrat identified the free energy of nonconvex multi-species models with centered Ising spins [36]. H.-B. Chen and Mourrat then identified the limit of the free energy of all multi-species spherical spin glasses, with no convexity assumption [35]. Ho proved that the enriched free energy of vector spin glasses converges without convexity, so the critical-point representation of H.-B. Chen and Mourrat holds without assuming convergence [69].

Our contribution is a Parisi formula for the free energy, an identification of the limit of the largest achievable margin, and a description of the gap statistics and contacts of optimal configurations for the negative spherical perceptron. We prove the Parisi formula for smooth activations, including activations that are neither monotone nor concave (Theorem 1.1). For $\alpha > 2$ we identify the exponential volume of the feasible set at every margin below the critical margin $\kappa_c(\alpha)$ of the Parisi functional, and we prove that the largest achievable margin converges to $\kappa_c(\alpha)$ (Theorem 1.2). For α slightly above two, we show that $\kappa_c(\alpha)$ lies below Gardner’s prediction by at least a constant times $(\alpha - 2)^3$ (Theorem 1.4). Combined with the matching upper bound on $\kappa_{RS} - \kappa_c$ (Part II, Theorem 1.1), this order is exact (Corollary 1.6).

Our main result on the geometry of solutions concerns the configurations of maximal margin. With probability exponentially close to one, every optimal configuration has exactly N contacts. All configurations whose margin is close to optimal share a deterministic limiting distribution of gaps, and this distribution has an atom of mass exactly one at zero (Theorem 1.8). Thus, in the limit, no macroscopic set of constraints beyond the N contacts accumulates near contact. This is a rigorous form of the predicted isostaticity of the jamming point recalled above. The link between the random model and the limiting gap law is a collapse of Parisi minimizers. As the margin increases to $\kappa_c(\alpha)$, the minimal value of the Parisi functional tends to $-\infty$ continuously, and the mass of every minimizer (defined in §1.1) tends to zero at a power rate (Theorem 1.3). This is a rigorous counterpart of the physical picture of jamming, in which the overlap of solutions tends to one and the weight of the overlaps below any fixed level tends to zero [58].

1.1 Model and the Parisi functional

Fix $\alpha > 0$ and integers $M = M_N$ with $M_N/N \rightarrow \alpha$. Let $\mathbf{G} = (g_{ai})_{a \leq M, i \leq N}$ have independent standard Gaussian entries, write g_a for its a -th row, and put $b_a = g_a/\sqrt{N}$. Let μ_N be the uniform probability measure on

$$S_N = \{\sigma \in \mathbb{R}^N : \|\sigma\|^2 = N\}.$$

For $\sigma \in S_N$, the number $b_a \cdot \sigma$ is the *margin* of the constraint a at σ ; for each fixed $\sigma \in S_N$ it is a standard Gaussian variable. For an activation $U : \mathbb{R} \rightarrow [-\infty, \infty)$ that is bounded above, define

$$\begin{aligned} Z_N(U) &= \int_{S_N} \exp \left\{ \sum_{a=1}^M U(b_a \cdot \sigma) \right\} \mu_N(d\sigma), \\ f_N(U) &= \frac{1}{N} \log Z_N(U), \quad F_N(U) = \mathbb{E} f_N(U). \end{aligned} \tag{1.1}$$

For a margin $\kappa \in \mathbb{R}$, the feasible set, its volume, and the largest achievable margin are

$$\begin{aligned} S_N(\kappa) &= \{\sigma \in S_N : b_a \cdot \sigma \geq \kappa \text{ for all } a \leq M\}, \\ V_N(\kappa) &= \mu_N(S_N(\kappa)), \quad \kappa_N = \max_{\sigma \in S_N} \min_{a \leq M} b_a \cdot \sigma. \end{aligned} \tag{1.2}$$

We use the convention $\log 0 = -\infty$. Thus $V_N(\kappa) = Z_N(H_\kappa)$ for the hard constraint $H_\kappa(x) = \log \mathbf{1}_{\{x \geq \kappa\}}$. For $\sigma \in S_N$ let $m_N(\sigma) = \min_{a \leq M} b_a \cdot \sigma$, so that $\kappa_N = \max_{S_N} m_N$. The *gaps* of σ are the numbers $b_a \cdot \sigma - m_N(\sigma) \geq 0$, and the *empirical gap measure* of σ is

$$\nu_{N,\sigma} = \frac{1}{N} \sum_{a=1}^M \delta_{b_a \cdot \sigma - m_N(\sigma)},$$

a measure on $[0, \infty)$ of total mass M/N . A constraint with gap zero is called a *contact* of σ .

An *order parameter* is a nondecreasing, right-continuous function $\gamma : [0, 1] \rightarrow [0, 1]$ that is identically one on $[\bar{q}, 1]$ for some $\bar{q} < 1$. We write $\mathcal{U}$ for the set of order parameters and

$$Q_\gamma = \inf\{q \in [0, 1] : \gamma(q) = 1\} < 1, \quad \lambda_\gamma(q) = \int_q^1 \gamma(t) dt.$$

Thus γ is the distribution function of a probability measure μ_γ on $[0, Q_\gamma]$, and Q_γ is the top of its support. We call $\lambda_\gamma(0)$ the *mass* of γ . Given $\gamma \in \mathcal{U}$ and an activation U , let $u_\gamma = u_\gamma(q, x; U)$ solve the *Parisi equation*

$$\partial_q u_\gamma + \frac{1}{2} \partial_{xx} u_\gamma + \frac{\gamma(q)}{2} (\partial_x u_\gamma)^2 = 0, \quad u_\gamma(1, x; U) = U(x). \quad (1.3)$$

When γ is a step function, (1.3) is solved by an explicit backward recursion of Gaussian integrals, the Cole–Hopf transformation [42, 71] on each step; for general γ , and for the terminal data used in this part, the solution is defined by approximation. Both are recalled in Chapter 2. For the Sherrington–Kirkpatrick model these constructions are in [52, 67, 68, 105] and [76]. The *Crisanti–Sommers entropy* [44, 125] and the *Parisi functional* are

$$\begin{aligned} \text{CS}(\gamma) &= \frac{1}{2} \int_0^{\bar{q}} \frac{dq}{\lambda_\gamma(q)} + \frac{1}{2} \log(1 - \bar{q}), \\ \mathcal{P}_U(\gamma) &= \alpha u_\gamma(0, 0; U) + \text{CS}(\gamma), \quad \mathcal{P}_*(U) = \inf_{\gamma \in \mathcal{U}} \mathcal{P}_U(\gamma), \end{aligned} \quad (1.4)$$

where $\bar{q}$ is any number in $[Q_\gamma, 1)$. The value of $\text{CS}(\gamma)$ does not depend on this choice, since $\lambda_\gamma(q) = 1 - q$ on $[Q_\gamma, 1]$. The first term in $\mathcal{P}_U(\gamma)$ accounts for the constraints and the second for the spherical entropy.

For the hard constraint, let ϕ and Φ denote the standard Gaussian density and distribution function. For $\gamma \in \mathcal{U}$, $\bar{q} \in [Q_\gamma, 1)$ and $\kappa \in \mathbb{R}$, let $u_\gamma(q, x; \kappa)$ solve (1.3) on $[0, \bar{q}] \times \mathbb{R}$ with the terminal datum

$$u_\gamma(\bar{q}, x; \kappa) = \log \mathbb{P}\{x + \sqrt{1 - \bar{q}} G \geq \kappa\} = \log \Phi\left(\frac{x - \kappa}{\sqrt{1 - \bar{q}}}\right), \quad (1.5)$$

where G is a standard Gaussian variable, and set

$$\begin{aligned} \mathcal{P}_\kappa(\gamma) &= \alpha u_\gamma(0, 0; \kappa) + \text{CS}(\gamma), \quad \mathcal{P}_*(\kappa) = \inf_{\gamma \in \mathcal{U}} \mathcal{P}_\kappa(\gamma), \\ \kappa_c(\alpha) &= \sup\{\kappa \in \mathbb{R} : \mathcal{P}_*(\kappa) > -\infty\}. \end{aligned} \quad (1.6)$$

Again the value does not depend on $\bar{q} \in [Q_\gamma, 1)$. The terminal datum (1.5) is what one obtains by running the equation with $U = H_\kappa$ from time 1 to time $\bar{q}$ with $\gamma = 1$, which

avoids a singular terminal problem. This variational problem appears in the physics literature [58, 59], and it was stated in the form (1.5)–(1.6) by El Alaoui and Sellke [54]. For the terminal datum (1.5), they constructed a classical solution of (1.3), with bounds on its derivatives, and proved that its derivatives depend continuously on γ [54, Section 4]. We call $\mathcal{P}_*(\kappa)$ the *Parisi value* and $\kappa_c(\alpha)$ the *critical margin*.

The replica-symmetric order parameters $\mathbf{1}_{[q,1]}$, $q \in [0,1]$, give Gardner's replica-symmetric functional:

$$\mathcal{P}_\kappa^{\text{RS}}(q) = \mathcal{P}_\kappa(\mathbf{1}_{[q,1]}) = \alpha \mathbb{E} \log \Phi\left(\frac{\sqrt{q}G - \kappa}{\sqrt{1-q}}\right) + \frac{q}{2(1-q)} + \frac{1}{2} \log(1-q). \quad (1.7)$$

Recall that $\kappa_{\text{RS}}(\alpha)$ is the unique solution of $\alpha \mathbb{E}(\kappa_{\text{RS}} - G)_+^2 = 1$. It is Gardner's prediction for the limit of κ_N [62], and $\kappa_{\text{RS}}(\alpha) < 0$ if and only if $\alpha > 2$, since $\mathbb{E}G_-^2 = \frac{1}{2}$. The value $\alpha = 2$ at $\kappa = 0$ is Cover's capacity of the perceptron [43].

Finally, for $\gamma \in \mathcal{U}$ and $\kappa \in \mathbb{R}$ we define the *Parisi diffusion* $X = X^{\gamma, \kappa}$ as follows. On $[0, Q_\gamma]$ it is the unique strong solution of

$$dX_q = \gamma(q) \partial_x u_\gamma(q, X_q; \kappa) dq + dB_q, \quad X_0 = 0, \quad (1.8)$$

for a standard Brownian motion B . Conditionally on $(X_q)_{q \leq Q_\gamma}$, the endpoint X_1 has the law of $X_{Q_\gamma} + \sqrt{1 - Q_\gamma} G$ conditioned on the event $\{X_{Q_\gamma} + \sqrt{1 - Q_\gamma} G \geq \kappa\}$, where G is a standard Gaussian variable independent of B . The equation (1.8) is well posed by Lemmas 2.6 and 2.7. For mixed p -spin models this diffusion appears in the work of Auffinger and W.-K. Chen [7, Theorem 3 and (17)], and for the terminal datum (1.5) it appears in [54, (2.8)]. For step order parameters, X observed at the breakpoints of γ and at time 1 is a Markov chain with explicit tilted Gaussian transitions (Lemma 9.3). The variable $X_1 - \kappa \geq 0$ plays the role of the gap of a typical constraint.

1.2 Main results

Our first result is the Parisi formula for smooth activations. We consider two classes of activations:

- (A) $U \in C^2(\mathbb{R})$, $\sup U < \infty$, and $\|U'\|_\infty + \|U''\|_\infty < \infty$;
- (B) $U \in C^2(\mathbb{R})$ is concave, $\sup U < \infty$, and $\|U''\|_\infty < \infty$.

No sign condition is imposed on U' in class (A), and the activations in class (B) may tend to $-\infty$ quadratically.

Theorem 1.1. *Let $\alpha > 0$ and let U belong to class (A) or class (B). Then*

$$\lim_{N \rightarrow \infty} F_N(U) = \mathcal{P}_*(U), \quad f_N(U) \longrightarrow \mathcal{P}_*(U) \quad \text{in probability.}$$

Moreover, the infimum defining $\mathcal{P}_*(U)$ is attained, and every minimizer γ satisfies $Q_\gamma \leq r_U$, where $r_U < 1$ is an explicit number depending only on α and U , defined in (5.1). If U belongs to class (A), then $r_U \leq \alpha \|U'\|_\infty^2 / (1 + \alpha \|U'\|_\infty^2)$.

Class (A) contains activations that are neither monotone nor concave. This generality is used for the gaps: adding a bounded source term to a smooth approximation of the hard wall destroys monotonicity and concavity. Class (B) contains smooth concave penalties comparable to $-\beta(\ell - x)_+^2$, which we use to pass to hard constraints. The bound on Q_γ says that the minimizers for a smooth activation stay a fixed distance away from overlap one. For the hard wall this fails as the margin approaches the critical margin; see Theorem 1.3(iv).

For $\alpha > 2$ the critical margin is negative, and the feasible sets $S_N(\kappa)$ with $\kappa < 0$ are not convex. Our second result identifies their exponential volume and the largest achievable margin.

Theorem 1.2. *Let $\alpha > 2$. Then $-\infty < \kappa_c(\alpha) < 0$, and the following statements hold.*

- (i) *For every $\kappa < \kappa_c(\alpha)$, $N^{-1} \log V_N(\kappa) \rightarrow \mathcal{P}_*(\kappa)$ in probability.*
- (ii) *$\kappa_N \rightarrow \kappa_c(\alpha)$ in probability.*
- (iii) *For every $\kappa \geq \kappa_c(\alpha)$ and every $C < \infty$, $\mathbb{P}\{N^{-1} \log V_N(\kappa) > -C\} \rightarrow 0$.*

Part (iii) states that the normalized log volume tends to $-\infty$ at the critical margin itself; it does not assert that $S_N(\kappa_c)$ is empty. The assumption $\alpha > 2$ enters only through the inequality $\kappa_c(\alpha) < 0$; for every $\alpha > 0$, part (i) holds at the margins $\kappa < \min\{\kappa_c(\alpha), 0\}$ (see the end of §8.4). For $\alpha > 2$, parts (i) and (iii) confirm the free-energy part of a conjecture of El Alaoui and Sellke [54, Conjecture 1], and part (ii) confirms its capacity part in the form $\kappa_N \rightarrow \kappa_c(\alpha)$.

The proofs of Theorem 1.2 and of the gap theorems below rest on the following properties of the variational problem, which hold for every $\alpha > 0$. For $\kappa \in \mathbb{R}$ put $C_\kappa = \frac{1}{2}(\kappa + \sqrt{\kappa^2 + 4})$, and let $C_0 = -\log \Phi(-1)$.

Theorem 1.3. *Let $\alpha > 0$.*

- (i) *If $\mathcal{P}_*(\kappa) > -\infty$, then the infimum defining $\mathcal{P}_*(\kappa)$ is attained, every minimizer γ satisfies $1 - Q_\gamma > e^{2\mathcal{P}_*(\kappa)}$, and the set of minimizers is compact in $L^1([0, 1])$.*
- (ii) *$-\infty < \kappa_c(\alpha) \leq \kappa_{\text{RS}}(\alpha) < \infty$. If $\alpha > 2$, then $\kappa_c(\alpha) < 0$.*
- (iii) *$\mathcal{P}_*$ is finite, nonincreasing, concave, and continuous on $(-\infty, \kappa_c)$, and $\mathcal{P}_*(\kappa) = -\infty$ for $\kappa \geq \kappa_c$. More precisely, with $C_* = C_{\kappa_c}$, for all sufficiently small $\eta > 0$,*

$$\mathcal{P}_*(\kappa_c - \eta) \leq \frac{1}{2} \log \eta + \frac{1}{2} \log(2\alpha C_*) + \frac{1}{2}.$$

(iv) If $\mathcal{P}_*(\kappa) > -\infty$ and γ is a minimizer, then

$$\mathcal{P}_*(\kappa) \geq \frac{1+\alpha}{2} \log(1 - Q_\gamma) - \alpha(C_0 + \log(1 + C_\kappa)).$$

Consequently there are $C, \eta_0 > 0$, depending only on α , such that for $0 < \eta \leq \eta_0$ every minimizer γ of $\mathcal{P}_{\kappa-\eta}$ satisfies

$$1 - Q_\gamma \leq C\eta^{1/(1+\alpha)}, \quad \lambda_\gamma(0) \leq C\eta^{1/(2+2\alpha)}.$$

(v) There is $\kappa_0(\alpha) < \kappa_c(\alpha)$ such that, for every $\kappa \leq \kappa_0(\alpha)$, every minimizer of $\mathcal{P}_\kappa$ equals $\mathbf{1}_{[q,1]}$ for some $q \in (0,1)$. In particular $\mathcal{P}_*(\kappa) = \min_{q \in [0,1]} \mathcal{P}_\kappa^{\text{RS}}(q)$. If $\alpha \leq 1$, the same holds for every $\kappa < \kappa_c(\alpha)$.

Part (iii) says that the jamming transition is continuous on the exponential scale: the Parisi value tends to $-\infty$ as the margin increases to κ_c , and it does so at least logarithmically fast. Part (iv) shows that a very negative value pins the top of the support of every minimizer close to one. Together, (iii) and (iv) give the collapse of minimizers at the critical margin: their tops tend to one and their masses tend to zero, at power rates. This collapse is the input for the gap theorems. By part (v) and Theorem 1.2(i), for $\alpha > 2$ and $\kappa \leq \kappa_0(\alpha)$,

$$\frac{1}{N} \log V_N(\kappa) \longrightarrow \min_{q \in [0,1]} \mathcal{P}_\kappa^{\text{RS}}(q) \quad \text{in probability,}$$

which is Gardner's replica-symmetric formula for the free energy [62]. Thus Gardner's formula holds at sufficiently negative margins; the margin $\kappa_0(\alpha)$ is not explicit. Theorem 1.3 is proved in §6.6, and part (v) is Corollary 6.9. Two structural facts behind it hold for every order parameter: an exact value identity (Lemma 6.2) and the universal slope bound $0 < \lambda_\gamma(0) \partial_x u_\gamma(0,0;\kappa) < C_\kappa$ (Lemma 6.4).

Our next result concerns constraint densities slightly above two. There $\kappa_{\text{RS}}(\alpha)$ is close to zero, and Gardner's prediction is closest to being correct. We write $\alpha = 2 + e$.

Theorem 1.4. *There are numerical constants $e_0 > 0$ and $c > 0$ such that, for $0 < e < e_0$,*

$$\kappa_c(2+e) \leq \kappa_{\text{RS}}(2+e) - ce^3. \quad (1.9)$$

Theorem 1.4 is a statement about the variational problem alone, and it is proved in Chapter 7. Combined with Theorem 1.2(ii), it gives the following.

Corollary 1.5. *Let e_0 and c be as in Theorem 1.4, let $0 < e < e_0$, and put $\alpha = 2 + e$. Then $\kappa_N \rightarrow \kappa_c(\alpha)$ in probability, and $\kappa_c(\alpha) \leq \kappa_{\text{RS}}(\alpha) - ce^3$. In particular,*

$$\lim_{N \rightarrow \infty} \mathbb{P} \left\{ \kappa_N \geq \kappa_{\text{RS}}(\alpha) - \frac{c}{2} e^3 \right\} = 0.$$

For large α , comparing the asymptotics as $\kappa \rightarrow -\infty$ of the upper bound of Montanari, Zhong, and Zhou, recalled above, with the relation $\alpha \mathbb{E}(\kappa_{\text{RS}} - G)_+^2 = 1$ shows that, for all sufficiently large α , with probability tending to one no configuration has margin at least $\kappa_{\text{RS}}(\alpha) - 1/|\kappa_{\text{RS}}(\alpha)|$. Together with Theorem 1.2(ii), this gives $\kappa_c(\alpha) < \kappa_{\text{RS}}(\alpha)$ for all sufficiently large α . Corollary 1.5 shows that the failure of Gardner's formula extends to α close to two, where $\kappa_{\text{RS}}(2 + e) = -\frac{\sqrt{2\pi}}{8}e + O(e^2)$ is itself close to zero (see (7.3)). The proof of Theorem 1.4 evaluates the functional at an explicit order parameter with two support points near one.

In Part II we study the minimizers of $\mathcal{P}_\kappa$ for α slightly above two. We show that the replica-symmetric solution loses its stability at a de Almeida–Thouless margin $\kappa_d(\alpha) < \kappa_c(\alpha)$, and that immediately beyond $\kappa_d(\alpha)$ every minimizer exhibits full replica symmetry breaking: its Parisi measure is supported on a nondegenerate interval, with a positive continuous density in the interior and atoms at both endpoints (Part II, Theorem 1.2). We also determine the distances between the three margins $\kappa_d < \kappa_c < \kappa_{\text{RS}}$ (Part II, Theorem 1.1). We now restate the results of that part that we use.

Let $R = \phi/\Phi$ and $V(z) = R(z)(z + R(z))$, so that $0 < V < 1$, and put

$$c_0 = \int_{\mathbb{R}} V(z)(1 - V(z)) \, dz, \quad \theta_d = \frac{(2\pi)^{3/2}}{128 c_0^2}.$$

By Part II, Lemma 3.5 and Lemma 3.3(c), for $e > 0$ sufficiently small and every $\kappa < \kappa_{\text{RS}}(\alpha)$ the function $\mathcal{P}_\kappa^{\text{RS}}$ of (1.7) has a unique minimizer $q^*(\kappa)$ on $[0, 1]$, and $q^*(\kappa) \in (0, 1)$. By Part II, Proposition 3.8, there is also a numerical constant $c_2 > 0$ such that exactly one margin $\kappa \in (\kappa_{\text{RS}}(\alpha) - c_2, \kappa_{\text{RS}}(\alpha))$ satisfies

$$\alpha \mathbb{E} V(Z^*)^2 = 1, \quad Z^* = \frac{\sqrt{q^*} G - \kappa}{\sqrt{1 - q^*}}, \quad q^* = q^*(\kappa).$$

This margin is the de Almeida–Thouless margin $\kappa_d(\alpha)$. The displayed condition says that the replicon eigenvalue of the replica-symmetric solution vanishes (see [23, 59]); on $(\kappa_d, \kappa_{\text{RS}})$ this eigenvalue is negative, and the replica-symmetric solution is unstable (Part II, Proposition 3.8). The ordering of thresholds (Part II, Theorem 1.1) states that there are numerical constants $e_0, c, C > 0$ such that, for $0 < e < e_0$ and $\alpha = 2 + e$,

$$\kappa_{\text{RS}}(\alpha) - \kappa_d(\alpha) = \theta_d e^2 + O(e^3 \log(1/e)), \quad c e^3 \leq \kappa_{\text{RS}}(\alpha) - \kappa_c(\alpha) \leq C e^3. \quad (1.10)$$

The lower bound on $\kappa_{\text{RS}} - \kappa_c$ is Theorem 1.4, which Part II proves again through the zero-temperature functional (Part II, Theorem 3.13). The upper bound rests on a dual certificate for this zero-temperature version of the variational problem (Part II, Theorems 3.10 and 3.11). Zero-temperature Parisi functionals were introduced for the ground state energy of mixed p -spin models by Auffinger and W.-K. Chen [9], and for spherical

models by W.-K. Chen and Sen [40] and by Jagannath and Tobasco [78]; see also [75]. Franz and Parisi described the jamming limit of the perceptron through the boundary condition that the replica equations take as the overlap tends to one [58, (15)]. Part II also shows that for $\kappa \in (\kappa_d(\alpha), \kappa_c(\alpha))$ no minimizer of $\mathcal{P}_\kappa$ is replica symmetric (Part II, Corollary 3.14). Combining Theorems 1.2 and 1.3 with Part II, Theorem 1.1, and with this absence of replica-symmetric minimizers gives the following.

Corollary 1.6. *Let e_0, c, C be as in (1.10), let $0 < e < e_0$, and put $\alpha = 2 + e$. Then the following hold.*

(i) $\kappa_N \rightarrow \kappa_c(\alpha)$ in probability, $c e^3 \leq \kappa_{RS}(\alpha) - \kappa_c(\alpha) \leq C e^3$, and

$$\kappa_c(\alpha) - \kappa_d(\alpha) = \theta_d e^2 + O(e^3 \log(1/e)). \quad (1.11)$$

(ii) *For every $\kappa \in (\kappa_d(\alpha), \kappa_{RS}(\alpha))$, Gardner's replica-symmetric formula $\min_{q \in [0,1]} \mathcal{P}_\kappa^{RS}(q)$ is finite, and there is $\varepsilon > 0$ such that*

$$\mathbb{P} \left\{ \frac{1}{N} \log V_N(\kappa) \leq \min_{q \in [0,1]} \mathcal{P}_\kappa^{RS}(q) - \varepsilon \right\} \longrightarrow 1.$$

Part (i) sharpens Corollary 1.5 to exact order: the largest achievable margin converges to a limit that lies below Gardner's prediction by an amount of exact order $(\alpha - 2)^3$, and above the de Almeida–Thouless margin by an amount of order $(\alpha - 2)^2$. The thermodynamic limit is taken at fixed e , and part (i) describes the subsequent behavior of the limit $\kappa_c(2 + e)$ as $e \downarrow 0$. Part (ii) shows that Gardner's formula for the free energy fails on the whole interval $(\kappa_d(\alpha), \kappa_{RS}(\alpha))$. By Theorem 1.3(v) it holds at sufficiently negative margins, and by Theorem 1.2(i) and Part II, Theorem 1.2(i), it holds at the margins in $[\kappa_d(\alpha) - c'e^2, \kappa_d(\alpha)]$ for a numerical constant $c' > 0$. The corollary follows directly from the results above. Part (i) is Theorem 1.2(ii) together with (1.10), since $\kappa_c - \kappa_d = (\kappa_{RS} - \kappa_d) - (\kappa_{RS} - \kappa_c)$. For part (ii), the minimum of $\mathcal{P}_\kappa^{RS}$ is attained at $q^*(\kappa)$, so it is finite. Let $\kappa \in (\kappa_d, \kappa_c)$. By Theorem 1.3(i),(iii), minimizers of $\mathcal{P}_\kappa$ exist, and by Part II, Corollary 3.14, none of them is replica symmetric. Hence $\mathcal{P}_*(\kappa) < \mathcal{P}_\kappa^{RS}(q^*(\kappa))$, and Theorem 1.2(i) gives the claim. For $\kappa \in [\kappa_c, \kappa_{RS})$ the claim follows from Theorem 1.2(iii).

Part II also proves a lower bound on the Parisi value that matches Theorem 1.3(iii) up to the constant in front of the logarithm, through a comparison with a zero-temperature functional (Part II, Proposition 2.24): for every $\alpha > 2$ there are $C, \eta_0 > 0$, depending only on α , such that

$$\alpha \log \eta - C \leq \mathcal{P}_*(\kappa_c - \eta) \leq \frac{1}{2} \log \eta + C \quad (0 < \eta \leq \eta_0). \quad (1.12)$$

Together with Theorem 1.2(i), this logarithmic divergence gives the following.

Corollary 1.7. *Let $\alpha > 2$. There are $C, \eta_0 > 0$, depending only on α , such that for every $0 < \eta \leq \eta_0$,*

$$\mathbb{P}\left\{\alpha \log \eta - C \leq \frac{1}{N} \log V_N(\kappa_c - \eta) \leq \frac{1}{2} \log \eta + C\right\} \longrightarrow 1.$$

Equivalently, with probability tending to one, $e^{-CN} \eta^{\alpha N} \leq V_N(\kappa_c - \eta) \leq e^{CN} \eta^{N/2}$. On the exponential scale the feasible volume thus vanishes polynomially in the distance η to the critical margin. The upper bound uses only Theorems 1.2 and 1.3, and we do not claim that either exponent is optimal.

We next describe the configurations that nearly achieve the maximal margin, through their empirical gap measures $\nu_{N,\sigma}$.

Theorem 1.8. *Let $\alpha > 2$. There is a deterministic measure ν_α on $[0, \infty)$ with $\nu_\alpha([0, \infty)) = \alpha$ and $\nu_\alpha(\{0\}) = 1$, whose positive part satisfies*

$$\nu_\alpha((0, t]) \leq \frac{\alpha \log 3}{\log(\sqrt{2\pi}/t)} \quad (0 < t < \sqrt{2\pi}), \quad (1.13)$$

such that the following statements hold.

(i) *For every bounded Lipschitz $\psi : [0, \infty) \rightarrow \mathbb{R}$ and every $\varepsilon > 0$ there is $\rho > 0$ such that*

$$\mathbb{P}\left\{\sup_{\sigma \in S_N: m_N(\sigma) \geq \kappa_c - \rho} \left| \int \psi d\nu_{N,\sigma} - \int \psi d\nu_\alpha \right| \leq \varepsilon\right\} \longrightarrow 1.$$

(ii) *For every $c < (\alpha - 2)^2 / (2\alpha)$ and all sufficiently large N , with probability at least $1 - e^{-cN}$ every maximizer σ_* of m_N has exactly N contacts. For any choice of maximizers, $\nu_{N,\sigma_*} \rightarrow \nu_\alpha$ weakly in probability, and*

$$\frac{1}{N} \#\{a \leq M : 0 < b_a \cdot \sigma_* - \kappa_N \leq t\} \longrightarrow \nu_\alpha((0, t]) \quad \text{in probability}$$

at every continuity point $t > 0$ of ν_α .

(iii) *For every choice of margins $k \uparrow \kappa_c(\alpha)$ and minimizers γ_k of $\mathcal{P}_k$, the measures $\alpha \text{Law}(X_1^{\gamma_k, k} - k)$ converge weakly to ν_α .*

Here $\nu_{N,\sigma_*} \rightarrow \nu_\alpha$ weakly in probability means that $\int \psi d\nu_{N,\sigma_*} \rightarrow \int \psi d\nu_\alpha$ in probability for every bounded continuous ψ , uniformly over the maximizers. The count N in part (ii) is isostatic in the sense recalled above: it equals the $N - 1$ degrees of freedom on the sphere plus one for the optimized margin. The identity $\nu_\alpha(\{0\}) = 1$ and the bound (1.13) show that, in the limit, no additional macroscopic set of constraints has small positive gaps: the mass $\nu_\alpha((0, t])$ of the small positive gaps tends to zero as $t \downarrow 0$, at least as fast as a constant times $1/\log(1/t)$. Nearly optimal configurations need not have exactly N

contacts, but by part (i) their gap statistics are close to ν_α , and by part (iii) the limiting gap law is identified through Parisi minimizers approaching the critical margin. The theorem does not identify the behavior of ν_α near zero, where physics predicts a power law.

The last theorem identifies the gap statistics of typical feasible configurations at a fixed margin below the critical one.

Theorem 1.9. *Let $\alpha > 2$ and $\kappa < \kappa_c(\alpha)$, let γ be a minimizer of $\mathcal{P}_\kappa$, and let $\psi : [0, \infty) \rightarrow \mathbb{R}$ be bounded and L -Lipschitz with $L > 0$. For every $\varepsilon > 0$, with probability tending to one, $V_N(\kappa) > 0$ and*

$$\begin{aligned} \mu_N \left(\left\{ \sigma \in S_N(\kappa) : \left| \frac{1}{N} \sum_{a=1}^M \psi(b_a \cdot \sigma - \kappa) - \alpha \mathbb{E} \psi(X_1^{\gamma, \kappa} - \kappa) \right| > \varepsilon \right\} \right) \\ \leq \exp \left\{ -\frac{\varepsilon^2 N}{8\alpha L^2 \lambda_\gamma(0)} \right\} V_N(\kappa). \end{aligned} \quad (1.14)$$

Consequently, writing $\langle \cdot \rangle_{N, \kappa}$ for expectation under the uniform probability measure on $S_N(\kappa)$,

$$\left\langle \frac{1}{N} \sum_{a=1}^M \psi(b_a \cdot \sigma - \kappa) \right\rangle_{N, \kappa} \longrightarrow \alpha \mathbb{E} \psi(X_1^{\gamma, \kappa} - \kappa) \quad \text{in probability,} \quad (1.15)$$

and all minimizers of $\mathcal{P}_\kappa$ induce the same law of $X_1^{\gamma, \kappa}$.

Applied to finitely many test functions at once, the theorem shows that all but an exponentially small fraction of the feasible set has the gap statistics of the Parisi diffusion. The rate in (1.14) depends on the minimizer only through its mass $\lambda_\gamma(0)$, and any minimizer may be used. We do not prove that the minimizers of $\mathcal{P}_\kappa$ are unique, and none of our arguments requires it; the last assertion of the theorem shows that they all induce the same terminal law. For mixed p -spin models, uniqueness follows from strict convexity of the Parisi functional [7, 76], which is not known for $\mathcal{P}_\kappa$. For generic mixed p -spin models, Auffinger and Jagannath expressed spin statistics through solutions of partial differential equations [10].

Throughout, C and c denote positive constants. A numerical constant depends on nothing; otherwise the dependence of a constant is indicated in the statement where it appears, as in “depending only on α ”. Their values may change from one occurrence to the next. Limits in probability refer to the joint law of the patterns and of any auxiliary randomness introduced in the proofs.

1.3 Ideas of the proof

The proof has three parts. Steps 1 and 2 below prove the Parisi formula for smooth activations. Steps 3 and 4 analyze the deterministic variational problem for the hard wall.

Steps 5 and 6 transfer these results to the hard constraints of the random model and to the gaps. Figure 1.1 shows how the steps depend on one another.

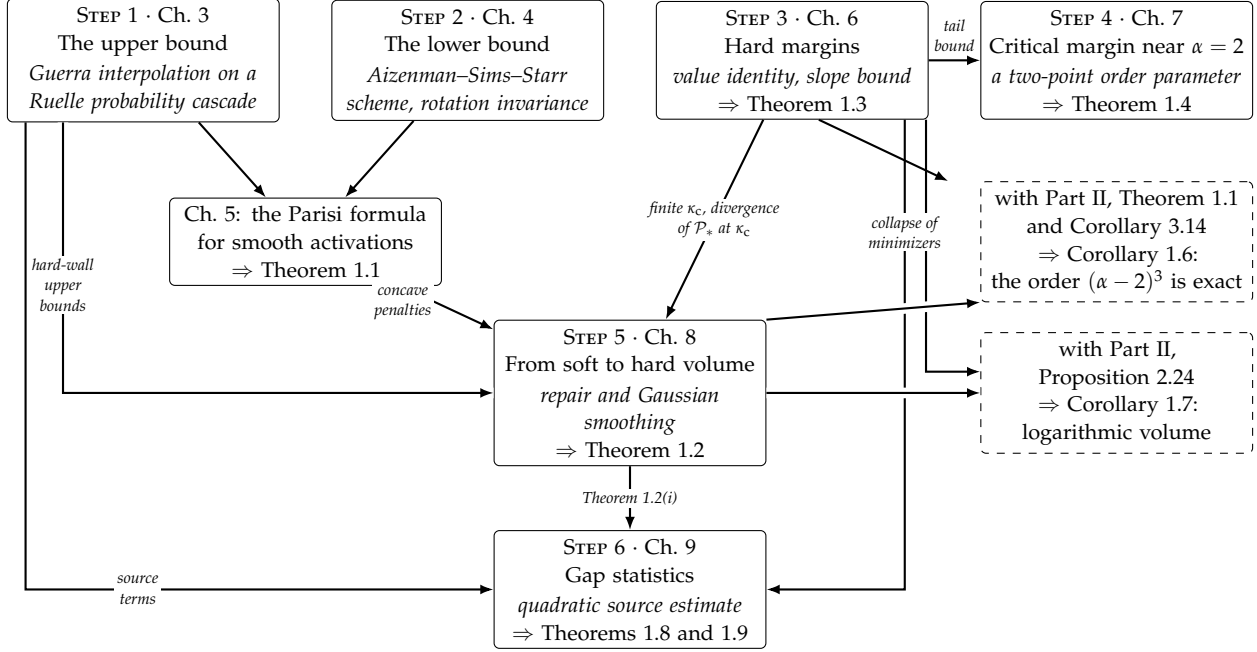


Figure 1.1. Selected dependencies among the steps of the proof. Each solid box names a step or chapter and, where there is one, the main result it proves; an arrow runs from a box to a box that uses it. The dashed boxes are the corollaries that also use results of Part II. Corollary 1.5 combines Theorems 1.2 and 1.4 and is not drawn, and neither are the analytic properties of the Parisi functional collected in Chapter 2, which are used throughout. Among them, the lemmas of §2.3 are derived from Part II, §§2.1–2.5 and Chapter A, which do not use Part I.

Step 1: The upper bound. The argument follows the proof of [92, Theorem 4.1], a step in Mourrat’s upper bound for bipartite models; the opening of Chapter 3 compares our argument with his. We use a Guerra-type interpolation [68] between the perceptron and a decoupled system on a Ruelle probability cascade [19, 112]. Since $\mathcal{P}_U$ is continuous in the order parameter, it suffices to treat order parameters γ_h with finitely many atoms $h_0 \leq \dots \leq h_k$ of equal mass. As the interpolation time t runs from 0 to 1, the argument of each activation moves from a Gaussian field on the leaves of the cascade, with covariance h_l between two leaves that branch at level l , to $b_a \cdot \sigma$. The coordinates are coupled to a second cascade field, whose levels $p = (p_0, \dots, p_k)$ are free parameters. The derivative in t of the interpolating free energy u_N is expressed through two overlaps: the spin overlap $R_{12} = \sigma^1 \cdot \sigma^2 / N$, and the *response overlap* $X_{12} = M^{-1} \sum_a U'(S_a^1) U'(S_a^2)$, where S_a^l is the interpolated field of constraint a for replica l . Overlaps of the derivatives of the activation enter in a similar way in Talagrand’s treatment of the perceptron [127, Chapters 2 and 3],

[128, Chapter 8]. Let J be the level at which the leaves of two replicas branch, w_l the probability that $J = l$, and r_l and ζ_l the conditional means of R_{12} and of $\alpha_N X_{12}$ given $J = l$, where $\alpha_N = M/N$. Then

$$\partial_t u_N = -\frac{1}{2} \sum_{l=0}^k w_l \left[(r_l - h_l) \zeta_l + \text{Cov}(R_{12}, \alpha_N X_{12} \mid J = l) \right].$$

Neither term has a sign. We control the products through the choice of p . At a maximizer over p of $u_N + \frac{1}{2} \sum_l w_l h_l p_l$, the first-order conditions, together with the monotonicity $0 \leq \zeta_0 \leq \dots \leq \zeta_k$, make the total contribution of the products nonpositive. This monotonicity holds for every activation in class (A); we derive it from an exact change of measure on a cascade with marks. For Gaussian fields the corresponding monotonicity is [128, Proposition 14.3.2] and [92, Lemma 2.4]. A maximum principle in t alone then bounds the free energy by the maximum over p at $t = 0$, up to the covariance terms. To control these, small Gaussian perturbations of the Hamiltonian, of the form used in [92, Section 4], enforce approximate Ghirlanda–Guerra identities [64] for the pair of overlap arrays at the points where the maximum principle is applied. Mourrat’s quantitative synchronization [92], which builds on Panchenko’s synchronization mechanism [100, 101], then makes the conditional variance of R_{12} small, and the Cauchy–Schwarz inequality makes the covariances small whatever the sign of U' . This is why no sign condition on U' is needed. At $t = 0$ the system splits into the row term $\alpha_N u_{\gamma_h}(0, 0; U)$ and a spherical term. An exact computation with Gaussian cascades bounds the spherical term plus $\frac{1}{2} \sum_l w_l h_l p_l$, maximized over p , by $\text{CS}(\gamma_h) + o(1)$; a computation of this kind for spherical spin glasses is in [125]. This is carried out in Chapter 3.

Step 2: The lower bound. We use the Aizenman–Sims–Starr scheme [1] in the spherical form of [37]. For multi-species spherical models, Bates and Sohn combined W.-K. Chen’s cavity computation with Panchenko’s synchronization [100, 101] to prove the lower bound [16]. We follow the same route, with the response overlap in the role of the overlap of the second species. The scheme bounds the free energy from below by the increment of the expected log partition function when n coordinates, and about αn constraints, are added to a system with N coordinates. We write the sphere of dimension $N + n$ exactly as a product of the sphere of dimension N and a density for the n new coordinates (compare [47]). The increment then splits into a row term and a coordinate term. In the limit the row term is at least $\alpha n u_{\zeta}(0, 0; U)$, where ζ is the limiting overlap law. After a Gaussian interpolation, the coordinate term becomes a Gaussian integral over the new coordinates $\varepsilon \in \mathbb{R}^n$. After a perturbation (see [98] and [100, (24)–(28)]), the joint law of the spin overlap and the response overlap satisfies the Ghirlanda–Guerra identities. By synchronization and the characterization of such arrays [49, 92, 97, 100, 101], it is a limit of Ruelle probability cascades with paired nondecreasing levels. On such a cascade the

coordinate term is an explicit Gaussian integral, which for general paired levels is not the entropy of any order parameter.

Two identities of the system determine the result. First, a Gaussian integration by parts in the rows expresses the mean effective precision of the new coordinates through the two overlaps. In the limit, this *precision identity* shows that the smallest precision is at least one, so that the Gaussian integral is finite, and it cancels the constant terms. Second, the law of the patterns is invariant under rotations. Under a uniformly random rotation, the overlap $\varepsilon^1 \cdot \varepsilon^2 / n$ of the new coordinates of two replicas is close to their full overlap, with mean squared error at most $3/n$ in the limit. Further, $\mathbb{E}\|\varepsilon\|^2/n = 1$, and this norm identity turns the coordinate term into $n \text{CS}(\gamma)$, where γ is the law of the mean overlap of the new coordinates at the branching level, and the overlap estimate makes γ close to ζ . Hence, as $N \rightarrow \infty$, the increment per added coordinate is at least

$$\mathcal{P}_U(\gamma) - O(n^{-1/2}) \geq \mathcal{P}_*(U) - O(n^{-1/2}).$$

Rotation invariance thus forces the cavity term to equal the Crisanti–Sommers entropy, and no minimizer or critical point of a cavity functional has to be identified. The main technical issue is the unbounded Gaussian integral over ε . We compute it on balls, pass to the limit $N \rightarrow \infty$ there, evaluate the limit on finite cascades, and remove the cutoff on these cascades, where the precision identity gives uniform control. This is carried out in Chapter 4. Chapter 5 combines the two bounds. There a truncation argument, which moves the mass of an order parameter above the level r_U down to r_U and so decreases $\mathcal{P}_U$, gives the attainment of the infimum and the bound $Q_\gamma \leq r_U$ for minimizers.

Step 3: The variational problem at hard margins. For the hard wall we prove an exact *value identity*, valid for every order parameter:

$$\mathcal{P}_\kappa(\gamma) = \frac{1}{2} \log \delta + \alpha \mathbb{E} h(Z) - \frac{\delta}{2} D(Q) - \frac{1}{2} \int_0^Q \gamma(q) D(q) dq.$$

Here $Q = Q_\gamma$, $\delta = 1 - Q$, $Z = (X_Q - \kappa)/\sqrt{\delta}$ is the normalized top position of the Parisi diffusion, $h(z) = \log \Phi(z) + \frac{1}{2} R(z)^2 < 0$, and $D(q) = \alpha \mathbb{E} \partial_x u_\gamma(q, X_q; \kappa)^2 - \int_0^q \lambda_\gamma^{-2}$ is twice the density of the first variation of $\mathcal{P}_\kappa$. El Alaoui and Sellke computed this first variation at a minimizer whose support is an interval, in directions supported on that interval [54, Proposition 16] (see the opening of Chapter 2). Here it is used at every order parameter and in every direction that keeps the top of the support below a fixed level; Chapter 6 recalls the first-order conditions known for mixed p -spin models. At a minimizer over the order parameters whose top is at most a fixed level $r < 1$, the last two terms are nonpositive, so $\mathcal{P}_\kappa(\gamma) < \frac{1}{2} \log \delta$. A finite value therefore keeps the top of every minimizer away from one, which gives existence and compactness of minimizers. The second tool is the *universal slope bound*

$$0 < \lambda_\gamma(0) \partial_x u_\gamma(0, 0; \kappa) < C_\kappa,$$

valid for every order parameter and proved by a maximum principle for an explicit combination of $\partial_x u_\gamma$ and $\partial_{xx} u_\gamma$. The map $\kappa \mapsto \mathcal{P}_\kappa(\gamma)$ is concave, by Prékopa's theorem [108, Theorem 6], with derivative $-\alpha \partial_x u_\gamma(0, 0; \kappa)$, and $\lambda_\gamma(0) \geq \delta$, so the two tools transport values between margins. Take a constrained minimizer γ with a prescribed value w at a margin $k > \kappa_c$, where the unconstrained value is $-\infty$. The value identity gives $1 - Q_\gamma > e^{2w}$, so its slope in the margin is at most $\alpha C_k e^{-2w}$. Concavity and the limit $k \downarrow \kappa_c$ give $\mathcal{P}_*(\kappa_c - \eta) \leq w + \alpha \eta C_{\kappa_c} e^{-2w}$, and the choice $w = \frac{1}{2} \log(2\alpha C_{\kappa_c} \eta)$ gives Theorem 1.3(iii).

For a minimizer over all order parameters, the value identity holds without the last two terms. Since $h(z)$ is bounded below by $-\log(1 + z_-)$ up to an additive constant, and the slope bound controls the negative part of Z , Jensen's inequality gives the pinning bound of Theorem 1.3(iv). Combined with (iii) at the margin $\kappa_c - \eta$, it forces $1 - Q_\gamma \leq C\eta^{1/(1+\alpha)}$. Finally, weak top stability, a lower bound on the quantity $S(Q) = \alpha \delta^2 \mathbb{E} \partial_{xx} u_\gamma(Q, X_Q; \kappa)^2$ for constrained minimizers with more than one support point, gives $\lambda_\gamma(0) \leq C(1 - Q_\gamma)^{1/2}$. At very negative margins the same lower bound cannot hold, so every minimizer is replica symmetric; this gives $\kappa_c > -\infty$ and Theorem 1.3(v). No zero-temperature functional is needed. This is carried out in Chapter 6.

Step 4: The critical margin near $\alpha = 2$. To go below Gardner's prediction we evaluate $\mathcal{P}_\kappa$ at an explicit order parameter with two support points near one,

$$\gamma_\delta = a \mathbf{1}_{[p, Q]} + \mathbf{1}_{[Q, 1]}, \quad Q = 1 - \delta, \quad p = Q(1 - \ell), \quad a = \frac{\delta}{Q\ell},$$

which splits the atom of a replica-symmetric order parameter and places a thin layer of relative width ℓ below the top. The homogeneity and the monotonicity of the Cole-Hopf step and a Gaussian tail bound give $\delta \mathcal{P}_\kappa(\gamma_\delta) \leq Q J_\ell(\kappa/\sqrt{Q}) + \frac{1}{2} \delta \log \delta$ for an explicit Gaussian integral J_ℓ , and

$$J_\ell(k) = \frac{1 - \alpha \mathbb{E}(k - G)_+^2}{4} - \frac{\ell}{2} \left(\log 2 - \frac{1}{2} \right) (\alpha \Phi(k) - 1) + O(\ell^{3/2}).$$

If $J_\ell(k) < 0$, then $\mathcal{P}_\kappa(\gamma_\delta) \rightarrow -\infty$ as $\delta \downarrow 0$ for every $\kappa > k$, so that $\kappa_c \leq k$. At $k = \kappa_{\text{RS}}$ the first term vanishes, while $\alpha \Phi(\kappa_{\text{RS}}) - 1$ is of exact order $e = \alpha - 2$. The layer therefore gains an amount of order $e\ell$ at a cost of order $\ell^{3/2}$, and the choice $\ell \asymp e^2$ makes $J_\ell(\kappa_{\text{RS}})$ negative, of order e^3 . Lowering the margin by a small multiple of e^3 keeps it negative. The argument uses no minimizer and no regularity theory beyond the recursion for this single order parameter. Unlike the bounds of Stojnic and of Montanari, Zhong, and Zhou recalled at the beginning of this chapter, the bound comes from the Parisi functional at an explicit order parameter, and the thin layer of width $\ell \asymp e^2$ gives the order e^3 near $\alpha = 2$. An explicit trial order parameter was used in the same way by Toninelli to show that the Sherrington–Kirkpatrick model is not replica symmetric beyond the de Almeida–Thouless line [129]. This is carried out in Chapter 7.

Step 5: From soft free energy to hard feasible volume. The upper bounds on the volume and on κ_N come from Step 1, applied to smooth approximations of the hard wall from above, together with a lemma showing that a single feasible configuration forces an exponentially nonnegligible feasible volume after a fixed decrease of the margin. For the lower bound, the Parisi formula for a concave penalty comparable to $-\beta(\ell - x)_+^2$ bounds from below the volume of the configurations whose total squared violation $\sum_a (k_0 - b_a \cdot \sigma)_+^2$, at a level k_0 slightly below ℓ , is at most tN , for small $t > 0$. Two geometric facts turn this into feasible volume. First, a small violation can be repaired by a small displacement. Using the smallest singular value of the pattern matrix over sparse sets of rows (see [130]), a separation argument moves such a configuration by at most $4\sqrt{tN}$ to a point that satisfies every constraint at a slightly smaller margin. The opening of Chapter 8 compares this repair with that of El Alaoui and Sellke [54, Lemma 11]. Second, the repair map may collapse volume, so we do not use its image. Instead we add to the configuration a Gaussian vector tangent to the sphere, with variance ρ^2 in each direction, and project back to S_N . This preserves μ_N . Translating the Gaussian vector by a displacement d lowers the probability of a symmetric set by at most the factor $e^{-\|d\|^2/(2\rho^2)}$, and Šidák's inequality [116] shows that, with probability at least $e^{-N\epsilon}$ for a small $\epsilon > 0$, the Gaussian vector moves no constraint by more than a small amount. Hence the perturbed point is feasible at a slightly smaller margin with probability at least $e^{-N\epsilon'}$, where ϵ' is small when t is small compared with ρ^2 . The upper bounds are proved in §§3.8 and 3.9 and the lower bound in Chapter 8.

Step 6: Gap statistics. Gap averages enter through the free energy of the feasible set with a source term $\beta \sum_a \psi(b_a \cdot \sigma - k)$, for a bounded Lipschitz test function ψ . Adding the source to a smooth approximation of the hard wall destroys monotonicity, and this is where the generality of class (A) is used. We prove a *quadratic source estimate*: the Parisi functional at γ with the source exceeds its linearization in β , whose slope is $\alpha \mathbb{E} \psi(X_1^{\gamma,k} - k)$, by at most $\frac{1}{2} \alpha \text{Lip}(\psi)^2 \beta^2 \lambda_\gamma(0)$. For a step order parameter the effect of the source is an iterated logarithmic moment along the Parisi diffusion observed at the breakpoints of γ . The transitions of this Markov chain are Gaussian laws tilted by log-concave weights. Like Gaussian laws, they have sub-Gaussian linear statistics and map Lipschitz functions to Lipschitz functions with the same constant (for log-concave tilts of Gaussian laws see [25, 27]), and a backward induction with a second-order expansion at each step gives the estimate. Combined with Jensen's inequality and the upper bound of Step 1, it yields one deviation inequality. With probability tending to one, simultaneously for every set $T \subseteq S_N(k)$ of positive volume, the average over T of $N^{-1} \sum_a \psi(b_a \cdot \sigma - k)$ differs from $\alpha \mathbb{E} \psi(X_1^{\gamma,k} - k)$ by at most

$$\frac{\mathcal{P}_k(\gamma) - N^{-1} \log \mu_N(T)}{\beta} + \frac{\alpha \text{Lip}(\psi)^2 \beta}{2} \lambda_\gamma(0) + \epsilon.$$

For Theorem 1.9 we apply it at a fixed margin $\kappa < \kappa_c$ to the set of feasible configurations whose gap average deviates by more than ε , and Theorem 1.2(i) shows that this set is an exponentially small fraction of $S_N(\kappa)$. For Theorem 1.8, a single configuration of nearly maximal margin is surrounded by a set of configurations that are feasible at a margin smaller by η , have almost the same gaps, and have volume at least $e^{-N(\log(1/\eta)+C)}$. Optimizing over β leaves an error of order $\eta + \{\lambda_\gamma(0)(1 + \log(1/\eta))\}^{1/2}$, which tends to zero by the collapse of minimizers in Theorem 1.3(iv).

The exact count of contacts does not use the variational problem. By Wendel's theorem [131], $\kappa_N < 0$ with probability exponentially close to one when $\alpha > 2$. The maximizers of m_N then correspond to the points of the polytope $\{y \in \mathbb{R}^N : g_a \cdot y \geq -1 \text{ for all } a\}$ farthest from the origin. These are vertices, and Gaussian general position gives exactly N contacts. A union bound over sets of N rows gives a finite- N form of (1.13) (Proposition 9.2). This is carried out in Chapter 9.

Chapter 2 collects the analytic properties of the Parisi functional used throughout: the Cole–Hopf recursion for step order parameters, Lipschitz continuity in the order parameter, and, for the hard wall, a priori bounds, signs of derivatives, and the first variation. Theorem 1.1 is proved in Chapter 5, Theorem 1.3 in Chapter 6, Theorem 1.4 in Chapter 7, Theorem 1.2 in Chapter 8, and Theorems 1.8 and 1.9 in Chapter 9. The corollaries follow from these theorems and, for Corollaries 1.6 and 1.7, from Part II, as explained next to their statements. The lemmas of §2.3, parts (i)–(iv) of Lemma 6.1, one Gaussian bound in the proof of Lemma 6.6 and the expansions of Chapter 7 are derived from Part II, §§2.1–2.5, Lemma 3.1 and Chapter A, which do not use Part I. Through them, Theorems 1.2–1.4, 1.8, and 1.9 also rest on Part II; Theorem 1.1 does not.

The Parisi functional

This chapter collects the deterministic facts about the Parisi functional that are used in the rest of this part. §2.1 introduces the Cole–Hopf step $\mathcal{T}_{m,s}$ and the backward recursion that solves the Parisi equation (1.3) when the order parameter is a step function. It records the monotonicity of the recursion in its terminal datum, which is used in Chapters 3, 5, 8, and 9, and the homogeneity of the Cole–Hopf step, which is used in Chapter 7. It also shows that neither the Crisanti–Sommers entropy nor the hard-wall value depends on the cutoff used to define it. §2.2 proves that, for activations in class (A) or class (B), the value $u_\gamma(0,0;U)$ is Lipschitz in γ for the L^1 distance, with an explicit constant (Lemmas 2.3 and 2.4). This defines u_γ for general order parameters, and it lets the upper bound (Chapter 3), the lower bound (Chapter 4), the truncation argument of Chapter 5 and the passage to hard constraints in Chapter 8 move between step order parameters and general ones. These continuity estimates rest on one identity, (2.15) in Lemma 2.2, which writes the difference of two solutions with the same terminal datum as an integral along a diffusion. Its only input is a formula for the derivatives of one Cole–Hopf step (Lemma 2.1). §2.3 treats the hard-wall functional $\mathcal{P}_\kappa$ on order parameters that equal one above a fixed level $r < 1$. There we state a priori bounds on the solution and on the associated diffusion (Lemma 2.6), Lipschitz continuity in the order parameter (Lemma 2.7), signs and a martingale (Lemma 2.8), and the first variation (Lemma 2.9). These are the inputs for the analysis of the variational problem in Chapter 6 and for the gap statistics in Chapter 9. Most of them are derived from Part II, as explained at the beginning of §2.3.

The hard-wall functional $\mathcal{P}_\kappa$ was studied by El Alaoui and Sellke in [54, Section 4]. For step order parameters they solved the Parisi equation with the terminal datum (1.5) by the Cole–Hopf recursion [54, (4.16)] and proved a priori bounds on the solution [54, Proposition 12]. They proved Lipschitz continuity in the order parameter [54, Proposition 13], extended the solution to general order parameters [54, (4.12)], and proved that it has one-sided derivatives in time [54, Lemma 14]. They also computed the first variation of $\mathcal{P}_\kappa$ in terms of the Parisi diffusion [54, (2.8), (4.13), and (4.15)], at a minimizer whose support is an interval, in directions supported on that interval, and deduced the first-order conditions $D = 0$ and $S = 1$ on that interval [54, Proposition 16], with D and S as in (2.29). Their arguments build on work on the Sherrington–Kirkpatrick and mixed p -spin models: Guerra’s Lipschitz bound in the order parameter [67], the Feynman–Kac

proof of it by Jagannath and Tobasco [76, Lemma 14 and Remark 15], the diffusion and the regularity estimates of Auffinger and W.-K. Chen [7, 8], and the first-variation formulas of W.-K. Chen [38] and of El Alaoui, Montanari, and Sellke [55, Proposition 6.8]. This chapter and Part II, Chapter 2 follow the same route, with the following differences. Concavity of the solution is obtained from Prékopa’s theorem (Lemma 2.1) rather than from an explicit computation along the recursion, which uses that the order parameter is nondecreasing [54, proof of (4.6)]. The bound $-\partial_{xx}u_\gamma \leq (1 - \bar{q})^{-1}$ of [54, (4.6)] is sharpened to the strict bound $-\partial_{xx}u_\gamma < 1/\lambda_\gamma$ (Lemma 2.8). The constants in §2.3 are uniform for κ in a compact interval. The first variation is computed at every order parameter in the class $\mathcal{U}_r$ of order parameters equal to one on $[r, 1]$, in every direction that stays in $\mathcal{U}_r$ (Lemma 2.9), and not only at a minimizer. Finally, the Lipschitz estimates of §2.2 cover activations in class (B), which are concave and may tend to $-\infty$ quadratically.

Throughout this chapter, G denotes a standard Gaussian variable and G' an independent copy of it, and ϕ and Φ denote the standard Gaussian density and distribution function. All functions are Borel measurable, and $\|\cdot\|_{L^1}$ is the norm of $L^1([0, 1])$.

2.1 Order parameters and the Cole–Hopf recursion

Recall from §1.1 that an *order parameter* is a nondecreasing, right-continuous function $\gamma : [0, 1] \rightarrow [0, 1]$ that is identically one on $[\bar{q}, 1]$ for some $\bar{q} < 1$, and that $\mathcal{U}$ is the set of order parameters. For $\gamma \in \mathcal{U}$ we write

$$Q_\gamma = \inf\{q \in [0, 1] : \gamma(q) = 1\} < 1, \quad \lambda_\gamma(q) = \int_q^1 \gamma(t) dt.$$

Thus γ is the distribution function of a probability measure on $[0, Q_\gamma]$. More generally, we identify every probability measure on $[0, 1]$ with its distribution function, a nondecreasing right-continuous function $\gamma : [0, 1] \rightarrow [0, 1]$ with $\gamma(1) = 1$.

The Cole–Hopf step. For $m \geq 0$, $s \geq 0$, and a function $f : \mathbb{R} \rightarrow [-\infty, \infty)$ that is bounded above, define

$$\mathcal{T}_{m,s}f(x) = \begin{cases} \frac{1}{m} \log \mathbb{E} e^{mf(x + \sqrt{s}G)}, & m > 0, \\ \mathbb{E} f(x + \sqrt{s}G), & m = 0, \end{cases} \quad (2.1)$$

with the conventions $e^{-\infty} = 0$ and $\log 0 = -\infty$. The expectations exist because f is bounded above, $\mathcal{T}_{m,s}f$ takes values in $[-\infty, \sup f]$, and $\mathcal{T}_{m,0}f = f$. For functions f, g that are bounded above, $a \in \mathbb{R}$, $m \geq 0$, and $s, s' \geq 0$, the following hold.

(T1) If $f \leq g + a$ on $\mathbb{R}$, then $\mathcal{T}_{m,s}f \leq \mathcal{T}_{m,s}g + a$ on $\mathbb{R}$.

(T2) $\mathcal{T}_{m,s}\mathcal{T}_{m,s'}f = \mathcal{T}_{m,s+s'}f$.

(T3) $t\mathcal{T}_{m,s}f = \mathcal{T}_{m/t,s}(tf)$ for every $t > 0$.

(T4) $\mathcal{T}_{m,s}f \geq \mathcal{T}_{0,s}f$.

If $f \equiv -\infty$, then $\mathcal{T}_{m,s}f \equiv -\infty$ for all m, s , and (T2)–(T4) for f are immediate; below we assume $f \not\equiv -\infty$, so that $\sup f$ is finite. Property (T1) holds because $y \mapsto e^{my}$ is nondecreasing and $\mathcal{T}_{m,s}(g+a) = \mathcal{T}_{m,s}g + a$. For (T2), write $x + \sqrt{s+s'}G$ as $x + \sqrt{s}G + \sqrt{s'}G'$. If $m > 0$, Tonelli's theorem applied to $e^{mf} \geq 0$ gives

$$\mathbb{E} \exp\{m \mathcal{T}_{m,s'}f(x + \sqrt{s}G)\} = \mathbb{E} e^{mf(x + \sqrt{s}G + \sqrt{s'}G')} = \mathbb{E} e^{mf(x + \sqrt{s+s'}G)},$$

and if $m = 0$ the same argument applies to $\sup f - f \geq 0$. Property (T3) follows from (2.1), and (T4) is Jensen's inequality $\mathbb{E}e^{mf} \geq e^{m\mathbb{E}f}$ for $m > 0$ and an equality for $m = 0$.

The operator $\mathcal{T}_{m,s}$ solves the Parisi equation over an interval on which the order parameter equals a constant m ; this is the Cole–Hopf transformation. Indeed, if u solves (1.3) with $\gamma = m > 0$, then $h = e^{mu}$ satisfies

$$\partial_q h + \frac{1}{2} \partial_{xx} h = mh \left(\partial_q u + \frac{1}{2} \partial_{xx} u + \frac{m}{2} (\partial_x u)^2 \right) = 0,$$

the backward heat equation, whose solutions are Gaussian averages of their later values. When $m = 0$ the equation is itself the backward heat equation. The recursion below turns this computation into a definition.

The step recursion. Let $T \in [0, 1]$. A *step function* on $[0, T)$ is a function $\gamma = \sum_{j=0}^n m_j \mathbf{1}_{[t_j, t_{j+1})}$ with $0 = t_0 < t_1 < \dots < t_{n+1} = T$ and $m_j \geq 0$. For such γ and a function $F : \mathbb{R} \rightarrow [-\infty, \infty)$ that is bounded above, define $u_\gamma(\cdot, \cdot; F)$ on $[0, T] \times \mathbb{R}$ by the backward recursion

$$u_\gamma(T, \cdot; F) = F, \quad u_\gamma(q, \cdot; F) = \mathcal{T}_{m_j, t_{j+1}-q} u_\gamma(t_{j+1}, \cdot; F) \quad (q \in [t_j, t_{j+1})). \quad (2.2)$$

We call F the *terminal datum* and T the *terminal time*; the terminal time is always clear from the context. By (T2), two consecutive steps with the same value compose into one step, so inserting a breakpoint does not change u_γ . Since two representations of γ have a common refinement, $u_\gamma(\cdot, \cdot; F)$ depends only on γ and F . For $T = 0$ there are no steps and $u_\gamma(0, \cdot; F) = F$. The recursion goes back to Parisi [105]; see also [68], [128, Chapter 14], and [98, Chapter 3].

A *step order parameter* is an order parameter of the form

$$\gamma(q) = m_j \quad \text{for } q \in [t_j, t_{j+1}), \quad 0 \leq j \leq n, \quad 0 = t_0 < t_1 < \dots < t_n < t_{n+1} = 1, \quad (2.3)$$

with $0 \leq m_0 \leq \dots \leq m_n = 1$ and $\gamma(1) = 1$. For such γ and an activation U that is bounded above, $u_\gamma(\cdot, \cdot; U)$ is given by (2.2) with $T = 1$ and $F = U$. The same definition applies to every step function γ that is the distribution function of a probability measure on $[0, 1]$; its value m_n on the last interval $[t_n, 1)$ need not be one.

The properties of $\mathcal{T}_{m,s}$ pass to the recursion. By (T1), for functions F, F' bounded above and $a \in \mathbb{R}$,

$$F \leq F' + a \text{ on } \mathbb{R} \implies u_\gamma(q, x; F) \leq u_\gamma(q, x; F') + a \text{ on } [0, T] \times \mathbb{R}. \quad (2.4)$$

In particular $F \mapsto u_\gamma(q, x; F)$ is nondecreasing and commutes with the addition of constants. For real-valued F, F' whose recursions are finite on $[0, T] \times \mathbb{R}$ (in particular, when both terminal data are bounded below by a quadratic),

$$\sup_{x \in \mathbb{R}} |u_\gamma(q, x; F) - u_\gamma(q, x; F')| \leq \sup_{x \in \mathbb{R}} |F - F'|. \quad (2.5)$$

By (T4), (T1) and the case $m = 0$ of (T2), $u_\gamma(q, x; F) \geq \mathbb{E}F(x + \sqrt{T - q} G)$. Hence, if $F(y) \geq -C(1 + y^2)$ for all y , then

$$u_\gamma(q, x; F) \geq \mathbb{E}F(x + \sqrt{T - q} G) \geq -C(1 + x^2 + T - q) \geq -C(2 + x^2). \quad (2.6)$$

If $\gamma = 1$ on $[\bar{q}, T)$ for some $\bar{q} \in [0, T)$, then (T2) gives

$$u_\gamma(\bar{q}, x; F) = \mathcal{T}_{1, T - \bar{q}} F(x) = \log \mathbb{E} e^{F(x + \sqrt{T - \bar{q}} G)}. \quad (2.7)$$

We next record the regularity of the recursion. Suppose that F is real-valued and continuous, and that $-C(1 + y^2) \leq F(y) \leq C$ for a constant C . Then $u = u_\gamma(\cdot, \cdot; F)$ is real-valued and continuous on $[0, T] \times \mathbb{R}$, and on each open interval (t_j, t_{j+1}) it is C^∞ and solves

$$\partial_q u + \frac{1}{2} \partial_{xx} u + \frac{m_j}{2} (\partial_x u)^2 = 0. \quad (2.8)$$

By (2.6) and induction over the steps, it suffices to prove the following for a continuous function f with $-C'(1 + y^2) \leq f(y) \leq C'$, a number $m \geq 0$, and $g(s, x) = \mathcal{T}_{m,s} f(x)$: the function g is C^∞ on $(0, \infty) \times \mathbb{R}$ with $\partial_s g = \frac{1}{2} \partial_{xx} g + \frac{m}{2} (\partial_x g)^2$, and $g(s, x) \rightarrow f(x_0)$ as $(s, x) \rightarrow (0, x_0)$. For $m > 0$ and $s > 0$,

$$e^{mg(s,x)} = \int_{\mathbb{R}} \frac{e^{-(y-x)^2/(2s)}}{\sqrt{2\pi s}} e^{mf(y)} dy, \quad 0 < e^{mf} \leq e^{mC'}.$$

Differentiation under the integral shows that the right side is positive, C^∞ in (s, x) , and solves $\partial_s h = \frac{1}{2} \partial_{xx} h$; reading the Cole–Hopf computation backwards gives the equation for g . For $m = 0$ the same argument applies to g itself, since $|f(y)| \leq C'(1 + y^2)$. For the limit, fix $\rho > 0$. For $m > 0$,

$$|\mathbb{E} e^{mf(x + \sqrt{s} G)} - e^{mf(x)}| \leq \sup_{|y-x| \leq \rho} |e^{mf(y)} - e^{mf(x)}| + 2e^{mC'} \mathbb{P}\{\sqrt{s}|G| > \rho\},$$

and for $m = 0$ the last term is replaced by $\mathbb{E}[|f(x + \sqrt{s} G) - f(x)|; \sqrt{s}|G| > \rho]$, which is at most $C''(1 + x^2) \mathbb{P}\{\sqrt{s}|G| > \rho\}^{1/2}$ for $s \leq 1$, with C'' depending only on C' , by the Cauchy–Schwarz inequality. Letting $s \downarrow 0$ and then $\rho \downarrow 0$, the continuity of f shows that $\mathcal{T}_{m,s} f \rightarrow f$ locally uniformly as $s \downarrow 0$ (for $m > 0$ because e^{mf} is continuous and positive), and hence $g(s, x) \rightarrow f(x_0)$.

General order parameters. For $\gamma \in \mathcal{U}$ choose $\bar{q} \in [Q_\gamma, 1)$. There are step order parameters γ_n with $\gamma_n = 1$ on $[\bar{q}, 1]$ and $\|\gamma_n - \gamma\|_{L^1} \leq 1/n$. Indeed, let $0 = t_0 < \dots < t_\ell = \bar{q}$ have mesh at most $1/n$, and put $\gamma_n(q) = \gamma(t_j)$ on $[t_j, t_{j+1})$ for $j < \ell$ and $\gamma_n = 1$ on $[\bar{q}, 1]$. Since γ is nondecreasing,

$$\int_0^1 |\gamma - \gamma_n| dq \leq \sum_{j < \ell} (t_{j+1} - t_j) (\gamma(t_{j+1}) - \gamma(t_j)) \leq \frac{1}{n}.$$

Further breakpoints may be added to the partition without affecting this bound. The same construction with $\bar{q} = 1$ (and $\gamma_n(1) = 1$) approximates the distribution function of any probability measure on $[0, 1]$ by nondecreasing step functions with values in $[0, 1]$. For general γ , and for the terminal data treated below (activations in class (A) or class (B), and the hard wall), the solution u_γ is defined as the limit of u_{γ_n} . That the limit exists and does not depend on the approximating sequence follows from the Lipschitz estimates in γ : Lemma 2.3 for class (A), Lemma 2.4 for class (B), and Lemma 2.7 for the hard wall. The properties (2.4), (2.5) and (2.6) pass to these limits. For mixed p -spin models, Auffinger and W.-K. Chen define the solution for general order parameters in the same way [8, Section 2].

The entropy and the cutoff. The definitions (1.4) and (1.5) involve a number $\bar{q} \in [Q_\gamma, 1)$, and neither $\text{CS}(\gamma)$ nor $u_\gamma(\cdot, \cdot; \kappa)$ depends on this choice. We prove this here for the entropy and for step order parameters; the extension to general $\gamma \in \mathcal{U}$ is given after Lemma 2.7. For CS, let $Q_\gamma \leq \bar{q} < \bar{q}' < 1$. Since $\lambda_\gamma(q) = 1 - q$ on $[Q_\gamma, 1]$,

$$\frac{1}{2} \int_{\bar{q}}^{\bar{q}'} \frac{dq}{\lambda_\gamma(q)} + \frac{1}{2} \log(1 - \bar{q}') = \frac{1}{2} \log \frac{1 - \bar{q}}{1 - \bar{q}'} + \frac{1}{2} \log(1 - \bar{q}') = \frac{1}{2} \log(1 - \bar{q}).$$

Equivalently, since the integrand below vanishes on $[Q_\gamma, 1]$,

$$\text{CS}(\gamma) = \frac{1}{2} \int_0^1 \left(\frac{1}{\lambda_\gamma(q)} - \frac{1}{1 - q} \right) dq. \quad (2.9)$$

For the hard wall, put

$$f_{\bar{q}}(x) = \log \Phi \left(\frac{x - \kappa}{\sqrt{1 - \bar{q}}} \right) = \log \mathbb{P} \{ x + \sqrt{1 - \bar{q}} G \geq \kappa \},$$

the terminal datum (1.5) at time $\bar{q}$; its dependence on κ is suppressed. Let γ be a step order parameter and $Q_\gamma \leq \bar{q} < \bar{q}' < 1$. Since $\gamma = 1$ on $[\bar{q}, \bar{q}']$, (2.7) with terminal time $\bar{q}'$ gives

$$\mathcal{T}_{1, \bar{q}' - \bar{q}} f_{\bar{q}'}(x) = \log \mathbb{P} \{ x + \sqrt{\bar{q}' - \bar{q}} G' + \sqrt{1 - \bar{q}'} G \geq \kappa \} = f_{\bar{q}}(x).$$

Thus the recursion from $f_{\bar{q}'}$ at time $\bar{q}'$ produces $f_{\bar{q}}$ at time $\bar{q}$, and the two definitions of $u_\gamma(q, x; \kappa)$ agree on $[0, \bar{q}]$. The same computation with the hard constraint $H_\kappa = \log \mathbf{1}_{[\kappa, \infty)}$

of §1.1 as terminal datum at time one gives $\mathcal{T}_{1,1-\bar{q}}H_\kappa = f_{\bar{q}}$, so for every step order parameter γ and $\bar{q} \in [Q_\gamma, 1)$,

$$u_\gamma(q, x; H_\kappa) = u_\gamma(q, x; \kappa) \quad (q \in [0, \bar{q}]), \quad \text{hence} \quad \mathcal{P}_\kappa(\gamma) = \alpha u_\gamma(0, 0; H_\kappa) + \text{CS}(\gamma). \quad (2.10)$$

Taking $\bar{q} = q$ in the definition, we obtain for every step order parameter

$$u_\gamma(q, x; \kappa) = \log \Phi\left(\frac{x - \kappa}{\sqrt{1 - q}}\right), \quad Q_\gamma \leq q < 1. \quad (2.11)$$

Classes with a common cutoff. For $0 \leq r < 1$ let

$$\mathcal{U}_r = \{\gamma \in \mathcal{U} : \gamma = 1 \text{ on } [r, 1]\} = \{\gamma \in \mathcal{U} : Q_\gamma \leq r\}.$$

The entropy is Lipschitz on each $\mathcal{U}_r$. For $\gamma_1, \gamma_2 \in \mathcal{U}_r$ use the common cutoff $\bar{q} = r$. On $[0, r]$ we have $\lambda_{\gamma_i} \geq \lambda_{\gamma_i}(r) = 1 - r$, and $|\lambda_{\gamma_1}(q) - \lambda_{\gamma_2}(q)| \leq \|\gamma_1 - \gamma_2\|_{L^1}$. Hence

$$|\text{CS}(\gamma_1) - \text{CS}(\gamma_2)| \leq \frac{1}{2} \int_0^r \frac{|\lambda_{\gamma_1} - \lambda_{\gamma_2}|}{\lambda_{\gamma_1} \lambda_{\gamma_2}} dq \leq \frac{r}{2(1-r)^2} \|\gamma_1 - \gamma_2\|_{L^1} \quad (\gamma_1, \gamma_2 \in \mathcal{U}_r). \quad (2.12)$$

Each $\mathcal{U}_r$ is a compact metric space for the L^1 distance. The L^1 distance is a metric on $\mathcal{U}_r$, because two right-continuous functions that agree almost everywhere agree on $[0, 1)$. By Helly's selection theorem, a sequence in $\mathcal{U}_r$ has a subsequence converging at every point of $[0, 1]$ to a nondecreasing function g with values in $[0, 1]$ and $g = 1$ on $[r, 1]$. The right-continuous version of g lies in $\mathcal{U}_r$ and differs from g at countably many points only, so the subsequence converges to it in L^1 by dominated convergence.

2.2 Continuity in the order parameter for smooth activations

We now show that $\gamma \mapsto u_\gamma(0, 0; U)$ is Lipschitz for the L^1 distance when U belongs to class (A) or class (B). Three arguments need this. The upper bound of Chapter 3 is proved for step order parameters and is transferred to all of $\mathcal{U}$ within a class $\mathcal{U}_r$. The lower bound of Chapter 4 compares two overlap laws in the Wasserstein distance, and these laws need not equal one near $q = 1$. The truncation argument of Chapter 5 needs an explicit Lipschitz constant. We therefore allow distribution functions γ, ν of arbitrary probability measures on $[0, 1]$, for which

$$W_1(\gamma, \nu) = \int_0^1 |\gamma(q) - \nu(q)| dq.$$

We use one pathwise bound for diffusions whose drift has linear growth. Let B be a standard Brownian motion on $[0, 1]$ and $B^* = \sup_{t \leq 1} |B_t|$, so that $\mathbb{E}(B^*)^2 \leq 4$ by Doob's inequality. If $0 \leq q \leq T \leq 1$ and $dY_s = \beta(s, Y_s) ds + dB_s$ on $[q, T]$ with $Y_q = x$

and $|\beta(s, y)| \leq a + K'|y|$, then $|Y_s| \leq |x| + 2B^* + a + K' \int_q^s |Y_t| dt$, since $T - q \leq 1$, and Gronwall's inequality gives

$$\sup_{s \in [q, T]} |Y_s| \leq (|x| + 2B^* + a)e^{K'}. \quad (2.13)$$

The next lemma computes the first two derivatives of one Cole–Hopf step. It is the only place where the form of $\mathcal{T}_{m,s}$ enters the continuity estimates.

Lemma 2.1. *Let $f \in C^2(\mathbb{R})$ be bounded above with f'' bounded, let $m \geq 0$ and $s > 0$, and put $g = \mathcal{T}_{m,s}f$. For $x \in \mathbb{R}$ let P_x be the law of $Y = x + \sqrt{s}G$ tilted by $e^{mf(Y)}$, that is, $dP_x / d\text{Law}(Y) = e^{mf(Y)} / \mathbb{E}e^{mf(Y)}$. Then $g \in C^2(\mathbb{R})$, $g \leq \sup f$, and*

$$g'(x) = \mathbb{E}_{P_x} f'(Y), \quad g''(x) = \mathbb{E}_{P_x} f''(Y) + m \text{Var}_{P_x}(f'(Y)). \quad (2.14)$$

Consequently the following hold.

- (a) If $|f'| \leq L$ on $\mathbb{R}$, then $|g'| \leq L$ and $|g''| \leq \sup |f''| + mL^2$.
- (b) If f is concave and $f'' \geq -K$ on $\mathbb{R}$, then g is concave and $g'' \geq -K$.

Proof. Since f' has at most linear growth, f'' is bounded, and $e^{mf} \leq e^{m \sup f}$, the function e^{mf} and its first two derivatives $mf'e^{mf}$ and $(mf'' + m^2 f'^2)e^{mf}$, evaluated at $x + \sqrt{s}G$, are dominated by $C(1 + |G|)^2$ for x in a compact set, where C depends on the compact set. Hence we may differentiate $\mathbb{E}e^{mf(x + \sqrt{s}G)}$ twice under the expectation, and dividing by $\mathbb{E}e^{mf(Y)}$ gives (2.14) for $m > 0$. For $m = 0$ we have $P_x = \text{Law}(Y)$, and the same argument applied to $\mathbb{E}f(x + \sqrt{s}G)$ gives $g' = \mathbb{E}f'(Y)$ and $g'' = \mathbb{E}f''(Y)$. Part (a) follows from (2.14) and $\text{Var}_{P_x} f'(Y) \leq L^2$. In part (b), $g'' \geq -K$ by (2.14). If $m = 0$, then g is an average of concave functions. If $m > 0$, put $h(x, y) = \exp\{mf(y) - (y - x)^2/(2s)\}$. Its logarithm is concave in (x, y) , and $\int_{\mathbb{R}} h(x, y) dy = \sqrt{2\pi s} e^{mg(x)}$. By Prékopa's theorem [108, Theorem 6], the marginal of a log-concave function on $\mathbb{R}^2$ is log-concave, so mg is concave. $\square$

For the hard-wall terminal datum, the formula for g'' in (2.14) appears in [54, (4.17)]. Part II, Lemma A.1 proves the same formulas for concave nondecreasing f ; here part (a) also covers data that are neither concave nor monotone, as in class (A). As in Part II, Lemma A.1, part (b) gives concavity one step at a time. Hence Lemma 2.2(i) holds for all step functions with values in $[0, 1]$, whereas Part II, Lemma A.2 and [54, proof of (4.6)] assume the coefficient nondecreasing.

The next lemma compares the solutions for two step coefficients with the same terminal datum, at a general terminal time T . Part II, Lemma A.2 proves the identity (2.15) for a class of concave nondecreasing terminal data that contains the hard-wall datum.

Lemma 2.2. Let $T \in (0, 1]$ and let $F \in C^2(\mathbb{R})$ be bounded above with F'' bounded. Let $\gamma, \nu : [0, T] \rightarrow [0, 1]$ be step functions, let $u_\gamma = u_\gamma(\cdot, \cdot; F)$ and $u_\nu = u_\nu(\cdot, \cdot; F)$ be given by (2.2) with terminal datum F at time T , and put $w = u_\gamma - u_\nu$. Assume one of the following.

(A) $|F'| \leq L$ on $\mathbb{R}$.

(B) F is concave and $F'' \geq -K$ on $\mathbb{R}$ for a constant $K \geq 0$. In this case put $A = \sup F - F(0) + |F'(0)| + K$.

Then the following hold for all $(q, x) \in [0, T] \times \mathbb{R}$.

(i) In case (A), $u_\gamma(q, \cdot) \in C^2(\mathbb{R})$ and $|\partial_x u_\gamma| \leq L$. In case (B), $u_\gamma(q, \cdot) \in C^2(\mathbb{R})$ and

$$\begin{aligned} -K \leq \partial_{xx} u_\gamma \leq 0, \quad F(0) - |F'(0)||x| - \frac{K}{2}(1+x^2) \leq u_\gamma(q, x) \leq \sup F, \\ |\partial_x u_\gamma(q, x)| \leq A + K|x|. \end{aligned}$$

The same holds for u_ν .

(ii) Let B be a standard Brownian motion and let Y solve $dY_s = \beta(s, Y_s) ds + dB_s$ on $[q, T]$ with $Y_q = x$, where $\beta = \frac{\gamma}{2}(\partial_x u_\gamma + \partial_x u_\nu)$. This equation has a unique strong solution, and

$$w(q, x) = \frac{1}{2} \int_q^T (\gamma - \nu)(s) \mathbb{E}[(\partial_x u_\nu)^2(s, Y_s)] ds. \quad (2.15)$$

(iii) In case (A), $|w(q, x)| \leq \frac{L^2}{2} \int_q^T |\gamma - \nu| ds$. In case (B),

$$|w(q, x)| \leq C_{A,K}(1+x^2) \int_q^T |\gamma - \nu| ds, \quad C_{A,K} = A^2 + 3K^2 e^{2K}(16 + A^2).$$

Proof. Refining the partitions, we may assume that γ and ν are constant on the intervals $[t_j, t_{j+1})$ of a common partition $0 = t_0 < \dots < t_{n+1} = T$. We divide the proof into four steps.

STEP 1. Proof of (i). For $q \in [t_j, t_{j+1})$, $u_\gamma(q, \cdot)$ is $\mathcal{T}_{\gamma(t_j), t_{j+1}-q}$ applied to $u_\gamma(t_{j+1}, \cdot)$. Induction over the intervals, starting from the last, and Lemma 2.1 give the following: in case (A), $u_\gamma(q, \cdot) \in C^2$, $|\partial_x u_\gamma| \leq L$, and $|\partial_{xx} u_\gamma| \leq \sup |F''| + (n+1)L^2$ on $[0, T] \times \mathbb{R}$; in case (B), $u_\gamma(q, \cdot) \in C^2$ is concave with $\partial_{xx} u_\gamma \geq -K$. In case (B), $u_\gamma \leq \sup F$, and (2.6) together with $F(y) \geq F(0) + F'(0)y - \frac{K}{2}y^2$ gives

$$u_\gamma(q, x) \geq \mathbb{E}F(x + \sqrt{T-q} G) \geq F(0) + F'(0)x - \frac{K}{2}(x^2 + T - q),$$

which is the asserted lower bound, since $T - q \leq 1$. For $x \in \{-1, 0, 1\}$ the right side is at least $F(0) - |F'(0)| - K$. By concavity,

$$u_\gamma(q, 1) - u_\gamma(q, 0) \leq \partial_x u_\gamma(q, 0) \leq u_\gamma(q, 0) - u_\gamma(q, -1),$$

so $|\partial_x u_\gamma(q, 0)| \leq \sup F - F(0) + |F'(0)| + K = A$, and $|\partial_{xx} u_\gamma| \leq K$ gives $|\partial_x u_\gamma(q, x)| \leq A + K|x|$.

STEP 2. *The equation for w .* By §2.1, u_γ and u_ν are continuous on $[0, T] \times \mathbb{R}$, and on each open interval (t_j, t_{j+1}) they are smooth and solve (2.8) with the constant coefficients γ and ν . Subtracting the two equations and using

$$\frac{\gamma}{2}(\partial_x u_\gamma)^2 - \frac{\nu}{2}(\partial_x u_\nu)^2 = \beta \partial_x w + \frac{\gamma - \nu}{2}(\partial_x u_\nu)^2,$$

we obtain on each $(t_j, t_{j+1}) \times \mathbb{R}$

$$\partial_q w + \frac{1}{2} \partial_{xx} w + \beta \partial_x w = -\frac{\gamma - \nu}{2}(\partial_x u_\nu)^2, \quad w(T, \cdot) = 0. \quad (2.16)$$

By Step 1, β is Lipschitz in y uniformly in s , and $|\beta(s, y)| \leq a + K'|y|$ with $(a, K') = (L, 0)$ in case (A) and $(a, K') = (A, K)$ in case (B). It is Borel, since $\partial_x u_\gamma$ and $\partial_x u_\nu$ are pointwise limits of difference quotients of continuous functions. Hence the equation for Y has a unique strong solution [79, Theorems 5.2.5 and 5.2.9], and (2.13) gives $\mathbb{E} \sup_{s \in [q, T]} Y_s^2 < \infty$. Moreover $|\partial_x w(s, y)| \leq 2(a + K'|y|)$ by (i), and $|w(s, y)| \leq C(1 + y^2)$ for a constant C , because u_γ and u_ν lie between the lower bound in (2.6) and $\sup F$ (in case (A), $F(y) \geq F(0) - L|y|$).

STEP 3. *Itô's formula and the passage through the breakpoints.* Fix j with $t_{j+1} > q$, put $t'_j = \max\{q, t_j\}$, and let $t'_j < s_1 < s_2 < t_{j+1}$. On $[s_1, s_2] \times \mathbb{R}$ the function w is smooth, so Itô's formula applies to $w(s, Y_s)$ on $[s_1, s_2]$. The stochastic integral $\int_{s_1}^{s_2} \partial_x w(s, Y_s) dB_s$ has mean zero, because $\mathbb{E} \int_{s_1}^{s_2} (\partial_x w)^2(s, Y_s) ds \leq 4\mathbb{E}(a + K' \sup_s |Y_s|)^2 < \infty$. Taking expectations and using (2.16),

$$\mathbb{E} w(s_2, Y_{s_2}) - \mathbb{E} w(s_1, Y_{s_1}) = -\frac{1}{2} \int_{s_1}^{s_2} (\gamma - \nu)(s) \mathbb{E} (\partial_x u_\nu)^2(s, Y_s) ds.$$

Now let $s_1 \downarrow t'_j$ and $s_2 \uparrow t_{j+1}$. The function w is continuous on $[0, T] \times \mathbb{R}$ and Y has continuous paths. Since $|w| \leq C(1 + y^2)$ and $(\partial_x u_\nu)^2(s, y) \leq (a + K'|y|)^2$ are dominated along Y by $C'(1 + \sup_s Y_s^2)$, which is integrable, dominated convergence gives the same identity with $s_1 = t'_j$ and $s_2 = t_{j+1}$. Summing over j , and using $Y_q = x$ and $w(T, \cdot) = 0$, gives (2.15).

STEP 4. *Proof of (iii).* In case (A), $(\partial_x u_\nu)^2 \leq L^2$ in (2.15). In case (B), $(\partial_x u_\nu)^2(s, y) \leq 2A^2 + 2K^2 y^2$ by (i). By (2.13) with $(a, K') = (A, K)$ and $\mathbb{E}(B^*)^2 \leq 4$,

$$\mathbb{E} Y_s^2 \leq e^{2K} \mathbb{E} (|x| + 2B^* + A)^2 \leq 3e^{2K} (x^2 + 16 + A^2),$$

so

$$\mathbb{E} (\partial_x u_\nu)^2(s, Y_s) \leq 2A^2 + 6K^2 e^{2K} (x^2 + 16 + A^2) \leq 2C_{A,K} (1 + x^2),$$

and (2.15) gives the bound. $\square$

The identity (2.15) is the Feynman–Kac formula of Jagannath and Tobasco [76, Lemma 14], and part (iii) is the analogue of Guerra’s bound (see the beginning of this chapter). Case (B) covers concave terminal data whose derivative grows linearly, such as the hard-wall datum and the activations in class (B).

For activations in class (A), the Lipschitz estimate holds in the uniform norm and at every time; this is the form used in the upper bound.

Lemma 2.3. *Let U belong to class (A) and put $L = \|U'\|_\infty$. Then for all $q \in [0, 1]$ and all distribution functions γ_1, γ_2 of probability measures on $[0, 1]$, in particular for all $\gamma_1, \gamma_2 \in \mathcal{U}$,*

$$\sup_{x \in \mathbb{R}} |u_{\gamma_1}(q, x; U) - u_{\gamma_2}(q, x; U)| \leq \frac{L^2}{2} \int_q^1 |\gamma_1 - \gamma_2| \, ds. \quad (2.17)$$

If moreover $\gamma_1, \gamma_2 \in \mathcal{U}_{\bar{q}}$ for some $\bar{q} < 1$, then

$$|\mathcal{P}_U(\gamma_1) - \mathcal{P}_U(\gamma_2)| \leq \left(\frac{\alpha L^2}{2} + \frac{\bar{q}}{2(1 - \bar{q})^2} \right) \|\gamma_1 - \gamma_2\|_{L^1}. \quad (2.18)$$

Proof. For step functions, (2.17) is Lemma 2.2(iii) in case (A), with $T = 1$ and $F = U$. For general γ_1, γ_2 , let $\gamma_{i,n}$ be step approximations as in §2.1. By (2.17) for step functions, $(u_{\gamma_{i,n}}(q, \cdot; U))_n$ is Cauchy in the uniform norm, and two approximating sequences of γ_i have the same limit, since they can be interleaved. This proves that u_{γ_i} is well defined, and (2.17) passes to the limit. Finally, (2.18) is α times (2.17) at $(q, x) = (0, 0)$ plus (2.12) with $r = \bar{q}$. $\square$

The next lemma gives the Lipschitz estimate for both classes at the origin, with an explicit constant.

Lemma 2.4. *Let U belong to class (A) or class (B). If U belongs to class (A), put $C_U^A = L^2/2$ with $L = \|U'\|_\infty$. If U belongs to class (B), put $C_U^B = C_{A,K}$, the constant of Lemma 2.2(iii) with*

$$K = \|U''\|_\infty, \quad A = \sup U - U(0) + |U'(0)| + K.$$

Let C_U be the smallest of the numbers C_U^A, C_U^B that are defined. Then, for any two probability measures γ, ν on $[0, 1]$,

$$|u_\gamma(0, 0; U) - u_\nu(0, 0; U)| \leq C_U W_1(\gamma, \nu). \quad (2.19)$$

Consequently, for $r \in [0, 1]$ and $\gamma, \nu \in \mathcal{U}_r$,

$$|\mathcal{P}_U(\gamma) - \mathcal{P}_U(\nu)| \leq \left(\alpha C_U + \frac{r}{2(1 - r)^2} \right) \|\gamma - \nu\|_{L^1}. \quad (2.20)$$

Proof. If U belongs to class (A), (2.19) with the constant C_U^A is (2.17) at $(q, x) = (0, 0)$. Let U belong to class (B). Then $F = U$ satisfies case (B) of Lemma 2.2 with $T = 1$ and the constants K and A above. For step functions γ, ν with values in $[0, 1]$, Lemma 2.2(iii) gives

$$|u_\gamma(q, x; U) - u_\nu(q, x; U)| \leq C_{A,K}(1 + x^2) \int_q^1 |\gamma - \nu| ds \quad \text{for all } (q, x) \in [0, 1] \times \mathbb{R}.$$

For general γ , approximate it by step functions as in §2.1. By this estimate the values at (q, x) form a Cauchy sequence, locally uniformly in (q, x) , and the limit does not depend on the approximating sequence. This defines $u_\gamma(q, x; U)$ (for U in both classes it agrees with the limit in Lemma 2.3), and the estimate passes to the limit. At $(q, x) = (0, 0)$ it gives (2.19) with the constant C_U^B . Finally, (2.20) follows from (2.19), the identity $W_1(\gamma, \nu) = \|\gamma - \nu\|_{L^1}$ and (2.12). $\square$

2.3 The hard-wall functional at a finite cutoff

We now study $\mathcal{P}_\kappa(\gamma)$ as a deterministic functional of the order parameter. The top Q_γ of the support is not continuous for the L^1 distance, since adding a small mass near one moves it. We therefore work on the classes $\mathcal{U}_r$, $0 \leq r < 1$, and evaluate every $\gamma \in \mathcal{U}_r$ with the common terminal time r . All constants below depend on r and on a compact interval I containing κ , but not on the number of steps of a step order parameter. None of the results concerns the limit $r \rightarrow 1$. Part II, §§2.1–2.5 and Appendix A, which do not use Part I, prove the results of this section other than the martingale identity (2.28), in a stronger form with constants that are uniform for $|\kappa| \leq K_0$, and we derive them from there. Part II defines the hard-wall solution for step order parameters by the same recursion, with the terminal datum f_r defined below, and for general order parameters by the same approximation within $\mathcal{U}_r$ (Part II, Definition 2.2(iii)), so the two definitions agree.

Gaussian quantities. We use the functions

$$R(z) = \frac{\phi(z)}{\Phi(z)}, \quad \bar{R}(z) = z + R(z), \quad V(z) = R(z)\bar{R}(z) \quad (z \in \mathbb{R}).$$

Lemma 2.5. *The following statements hold.*

(i) For $t > 0$,

$$\frac{t\phi(t)}{1+t^2} < \Phi(-t) < \frac{\phi(t)}{t}, \quad \text{equivalently} \quad t < R(-t) < t + \frac{1}{t}.$$

(ii) $R' = -V$ and $\bar{R}' = 1 - V$.

(iii) For $z \in \mathbb{R}$ let T_z have the law of $z + G$ conditioned on the event $\{z + G \geq 0\}$, that is, the law on $[0, \infty)$ with density proportional to $e^{zt-t^2/2}$. Then $\mathbb{E}T_z = \bar{R}(z)$ and $\text{Var}(T_z) = 1 - V(z)$. Consequently $\bar{R} > 0$ and $0 < V < 1$.

Proof. The functions R , $\bar{R}$, V , and the variable T_z are those of Part II, §2.1. Part (i) is Part II, Lemma 2.1(b), and parts (ii) and (iii) are contained in Part II, Lemma 2.1(a). $\square$

Part (i) consists of Gordon's bounds for the Mills ratio [65], and the bound $V < 1$ in part (iii) is due to Sampford [113]. El Alaoui and Sellke collect these facts, in terms of the inverse Mills ratio, in [54, Lemma 17].

The terminal datum. Fix $r \in [0, 1)$. For $\gamma \in \mathcal{U}_r$, the hard-wall solution $u_\gamma(\cdot, \cdot; \kappa)$ on $[0, r] \times \mathbb{R}$ is given by the recursion (2.2) with terminal time r and terminal datum

$$f_r(x) = \log \Phi\left(\frac{x - \kappa}{\sqrt{1 - r}}\right),$$

by (1.5) with $\bar{q} = r$; for step order parameters this is the definition, and for general $\gamma \in \mathcal{U}_r$ it is completed in Lemma 2.7. Using the same terminal time for all of $\mathcal{U}_r$ avoids the discontinuity of $\gamma \mapsto Q_\gamma$. With $z = (x - \kappa)/\sqrt{1 - r}$, Lemma 2.5 gives

$$f_r(x) \leq 0, \quad f'_r(x) = \frac{R(z)}{\sqrt{1 - r}} > 0, \quad 0 < -f''_r(x) = \frac{V(z)}{1 - r} < \frac{1}{1 - r}. \quad (2.21)$$

Every derivative of f_r has polynomial growth, uniformly for κ in a compact interval; this gives the case $r = 0$ of Lemma 2.6(v). Indeed, $R' = -zR - R^2$, so every derivative of R is a polynomial in z and R , and $0 < R(z) \leq R(0) + |z|$ because $-1 < R' < 0$ by Lemma 2.5(ii),(iii). Throughout the rest of this section, I is a compact interval, $\kappa \in I$, $K = (1 - r)^{-1}$, and constants depend only on r and I unless stated otherwise.

Notation. The following abbreviations are used in this section and in Chapter 6. We fix $\kappa \in \mathbb{R}$ and $\gamma \in \mathcal{U}$, and write

$$\begin{aligned} Q &= Q_\gamma, & \delta &= 1 - Q, & \lambda &= \lambda_\gamma, & u &= u_\gamma(\cdot, \cdot; \kappa), \\ b &= \partial_x u, & c &= -\partial_{xx} u, & v &= \lambda c. \end{aligned}$$

Thus $\lambda(q) \geq \lambda(Q) = \delta > 0$ for $q \in [0, Q]$, and u is defined on $[0, 1) \times \mathbb{R}$ by (2.11). We write

$$\mathcal{L} = \partial_q + \frac{1}{2} \partial_{xx} + \gamma b \partial_x$$

for the generator of the diffusion

$$dX_q = \gamma(q) b(q, X_q) dq + dB_q, \quad X_0 = 0, \quad (2.22)$$

where B is a standard Brownian motion. For $\gamma \in \mathcal{U}_r$ we consider X on $[0, r]$, and we write $X^{s,x}$ for the solution of the equation in (2.22) on $[s, r]$ with $X_s^{s,x} = x$; it exists and is unique by Lemma 2.6(iii) for step order parameters and by Lemma 2.7 in general. On $[0, Q]$, X is the Parisi diffusion (1.8) of §1.1, and the processes defined with two admissible values

of r agree on their common interval. It is also the diffusion of Part II, (2.14). When two order parameters $\gamma, \tilde{\gamma}$ are compared, we indicate the order parameter as a subscript, as in $u_{\tilde{\gamma}}, b_{\gamma}$, or $X^{\tilde{\gamma}}$. With this notation, the hard-wall functional (1.6) with cutoff $\bar{q} = Q$ reads

$$\mathcal{P}_\kappa(\gamma) = \alpha u(0,0) + \frac{1}{2} \int_0^Q \frac{dq}{\lambda(q)} + \frac{1}{2} \log \delta = \alpha u(0,0) + \frac{1}{2} \int_0^1 \left(\frac{1}{\lambda(q)} - \frac{1}{1-q} \right) dq, \quad (2.23)$$

the second form by (2.9), and $\mathcal{P}_*(\kappa) = \inf_{\gamma \in \mathcal{U}} \mathcal{P}_\kappa(\gamma)$.

A priori bounds and continuity. The next lemma collects the bounds on the solution and the diffusion that justify Itô's formula and the Feynman–Kac formula in the rest of this part.

Lemma 2.6. *Fix $r \in [0, 1)$ and a compact interval I . There is a constant C , and for each $p \geq 1$ and $k \geq 1$ there are constants C_p and C'_k , depending only on r, I , and p or k , such that for every $\kappa \in I$ and every step order parameter $\gamma \in \mathcal{U}_r$ the following hold on $[0, r] \times \mathbb{R}$.*

- (i) $-C(1 + x^2) \leq u(q, x) \leq 0$.
- (ii) *On each open interval where γ is constant, u is smooth and*

$$\mathcal{L}b = 0, \quad \mathcal{L}c = \gamma c^2.$$

- (iii) $0 \leq c \leq (1 - r)^{-1}$ and $|b(q, x)| \leq C(1 + |x|)$. *In particular the drift γb in (2.22) is Lipschitz in x with constant $(1 - r)^{-1}$ and has linear growth, so the equation in (2.22) has a unique strong solution $X^{s,x}$ from every starting point $(s, x) \in [0, r] \times \mathbb{R}$.*
- (iv) $\mathbb{E} \sup_{t \in [s, r]} |X_t^{s,x}|^p \leq C_p(1 + |x|)^p$; *in particular $\mathbb{E} \sup_{q \leq r} |X_q|^p \leq C_p$.*
- (v) $|\partial_x^k u(q, x)| \leq C'_k(1 + |x|)^{C'_k}$, *and $\partial_x^k u$ is continuous on $[0, r] \times \mathbb{R}$.*

Proof. For $r = 0$ the statements concern $u = f_0$ only and follow from (2.21) and the polynomial growth of the derivatives of f_0 noted after it; let $r > 0$. Every statement is contained in Part II, Chapter 2, applied with $K_0 = \max_{\kappa \in I} |\kappa|$; the constants there depend only on r, K_0 , and p or k . Part (i) and the bound on b in (iii) are Part II, Lemma 2.4(b). By Part II, Lemma 2.4(c), $0 < c < 1/\lambda \leq (1 - r)^{-1}$ on $[0, r] \times \mathbb{R}$, since $\lambda \geq \lambda(r) = 1 - r$ there. Hence $|\partial_x(\gamma b)| = \gamma c \leq (1 - r)^{-1}$, and the drift γb has linear growth. Strong existence and uniqueness for (2.22) are stated in Part II after (2.14). Part (ii) is Part II, Lemma 2.4(a) and Lemma 2.9(a). Part (iv) is Part II, Lemma 2.6(a), since $1 + |x|^p \leq (1 + |x|)^p$ for $p \geq 1$. Part (v) is Part II, Lemma 2.4(a),(b): the bound there is $C_k(1 + |x|^{p_k}) \leq 2C_k(1 + |x|)^{p_k}$, so one may take $C'_k = \max\{2C_k, p_k\}$. $\square$

Parts (i) and (iii) correspond to the bounds [54, (4.4)–(4.6)], and part (v) is a weaker form of [54, (4.7)], with polynomial growth in place of boundedness.

The next lemma is the hard-wall counterpart of Lemma 2.4. It extends the solution, its derivatives, and the diffusion to general order parameters, and it is the source of the

compactness and continuity used in Chapter 6. For mixed p -spin models, the convergence of all spatial derivatives of the solution as the order parameter converges is [8, Proposition 1].

Lemma 2.7. *Fix $r \in [0, 1)$ and a compact interval I . There is a constant C depending only on r and I , and for each $p \geq 1$ a constant C_p depending only on r , I , and p , such that for all $\kappa \in I$ and all step order parameters $\gamma, \tilde{\gamma} \in \mathcal{U}_r$, with $\varepsilon = \|\tilde{\gamma} - \gamma\|_{L^1}$,*

$$|u_{\tilde{\gamma}}(q, x) - u_{\gamma}(q, x)| \leq C\varepsilon(1 + x^2), \quad |b_{\tilde{\gamma}}(q, x) - b_{\gamma}(q, x)| \leq C\varepsilon(1 + |x|) \quad (2.24)$$

on $[0, r] \times \mathbb{R}$. For every $k \geq 2$ and $A_0 > 0$,

$$\sup \left\{ |\partial_x^k u_{\tilde{\gamma}}(q, x) - \partial_x^k u_{\gamma}(q, x)| : q \in [0, r], |x| \leq A_0 \right\} \longrightarrow 0 \quad \text{as } \varepsilon \rightarrow 0,$$

uniformly in $\kappa \in I$ and in the step order parameters. When $X^{\tilde{\gamma}}$ and X^{γ} start from the same point $(s, x) \in [0, r] \times \mathbb{R}$ and are driven by the same Brownian motion,

$$\mathbb{E} \sup_{t \in [s, r]} |X_t^{\tilde{\gamma}} - X_t^{\gamma}|^p \leq C_p \varepsilon^p (1 + |x|)^p. \quad (2.25)$$

Consequently the solution u , its spatial derivatives, and the diffusion X extend from step order parameters to every $\gamma \in \mathcal{U}_r$ by L^1 approximation within $\mathcal{U}_r$, and the bounds of Lemma 2.6(i), (iii)–(v) and of this lemma hold for all $\gamma, \tilde{\gamma} \in \mathcal{U}_r$. If $\gamma \in \mathcal{U}_r$ equals a constant m on an interval $[a, a'] \subset [0, r]$, then $u(q, \cdot) = \mathcal{T}_{m, a' - q} u(a', \cdot)$ for $q \in [a, a']$, and u is smooth and satisfies $\mathcal{L}b = 0$ and $\mathcal{L}c = \gamma c^2$ on $(a, a') \times \mathbb{R}$. Finally,

$$|\mathcal{P}_{\kappa}(\tilde{\gamma}) - \mathcal{P}_{\kappa}(\gamma)| \leq \left(\alpha C + \frac{r}{2(1-r)^2} \right) \|\tilde{\gamma} - \gamma\|_{L^1} \quad (\gamma, \tilde{\gamma} \in \mathcal{U}_r, \kappa \in I). \quad (2.26)$$

In particular $\mathcal{P}_{\kappa}$ is continuous on the compact set $\mathcal{U}_r$, uniformly for $\kappa \in I$, and attains its minimum over $\mathcal{U}_r$.

Proof. For $r = 0$ the class $\mathcal{U}_0$ has one element; let $r > 0$. We apply the results of Part II, Chapter 2 with $K_0 = \max_{\kappa \in I} |\kappa|$; they hold for all $\gamma, \tilde{\gamma} \in \mathcal{U}_r$, and their constants depend only on r , K_0 , and p or k . The solution for general $\gamma \in \mathcal{U}_r$ is the limit of the solutions for step approximations in $\mathcal{U}_r$ (Part II, Definition 2.2(iii)). By Part II, Lemma 2.4(f), for every $k \geq 0$,

$$|\partial_x^k u_{\tilde{\gamma}}(q, x) - \partial_x^k u_{\gamma}(q, x)| \leq C_k \varepsilon (1 + |x|^{p_k}) \quad \text{on } [0, r] \times \mathbb{R},$$

with $p_0 = 2$ and $p_1 = 1$. The cases $k = 0, 1$ give (2.24), the cases $k \geq 2$ give the convergence of the higher derivatives, and the derivatives of u_{γ} are the limits of those of its step approximations. The bound (2.25) is Part II, Lemma 2.6(b); in particular the diffusions of step approximations converge to X^{γ} . Part II, Lemma 2.4(a)–(c) and Lemma 2.6(a) hold for general $\gamma \in \mathcal{U}_r$, so the proof of Lemma 2.6 gives its parts (i) and

(iii)–(v) for all $\gamma \in \mathcal{U}_r$. If $\gamma = m$ on $[a, a']$, then $u(q, \cdot) = \mathcal{T}_{m, a' - q} u(a', \cdot)$ for $q \in [a, a']$ by Part II, Lemma 2.5(a), and u is smooth and satisfies $\mathcal{L}b = 0$ and $\mathcal{L}c = \gamma c^2$ on $(a, a') \times \mathbb{R}$ by Part II, Lemma 2.4(a) and Lemma 2.9(a). Finally, the first bound in (2.24) at $(0, 0)$ controls the energy term of $\mathcal{P}_\kappa$, and (2.12) controls the entropy term; this proves (2.26). Compactness of $\mathcal{U}_r$ was shown in §2.1. $\square$

We record a consequence of Lemma 2.7. If $\gamma \in \mathcal{U}_r \subset \mathcal{U}_{r'}$ with $r < r' < 1$, then the step approximations of γ in $\mathcal{U}_r$ also lie in $\mathcal{U}_{r'}$, and for them the solutions computed with the cutoffs r and r' agree on $[0, r]$ by the discussion before (2.10). Passing to the limit, $u_\gamma(\cdot, \cdot; \kappa)$ and $\mathcal{P}_\kappa(\gamma)$ do not depend on the cutoff for any $\gamma \in \mathcal{U}$, and (2.11) holds for every $\gamma \in \mathcal{U}$. This is also Part II, Lemma 2.3.

We next record the signs of b , c , and $1 - v$, which sharpen Lemma 2.6(iii) to strict inequalities, and the martingale structure along the diffusion. They are the starting point of the analysis in Chapter 6.

Lemma 2.8. *Let $r \in [0, 1)$, $\kappa \in \mathbb{R}$, and $\gamma \in \mathcal{U}_r$. Then, on $[0, r] \times \mathbb{R}$,*

$$b > 0, \quad 0 < v < 1, \quad \text{so that} \quad 0 < c < \frac{1}{\lambda} \leq \frac{1}{1 - r}. \quad (2.27)$$

Moreover, for $q \in [0, r]$,

$$b(q, X_q) = b(0, 0) - \int_0^q c(s, X_s) dB_s; \quad (2.28)$$

in particular $b(q, X_q)$ is a square-integrable martingale on $[0, r]$.

Proof. The signs (2.27) are Part II, Lemma 2.4(b),(c), applied with $K_0 = |\kappa|$; the last inequality holds because $\lambda \geq \lambda(r) = 1 - r$ on $[0, r]$. It remains to prove (2.28). For $r = 0$ there is nothing to prove; let $r > 0$ and take $I = \{\kappa\}$.

Step order parameters. Let γ be a step order parameter and fix $(s, x) \in [0, r] \times \mathbb{R}$. By Itô's formula on each open interval of constancy and $\mathcal{L}b = 0$, $db(t, X_t^{s,x}) = \partial_x b(t, X_t^{s,x}) dB_t = -c(t, X_t^{s,x}) dB_t$ there. Since b is continuous on $[0, r] \times \mathbb{R}$ with linear growth and $0 \leq c \leq K$ by Lemma 2.6(iii),(v), the identity extends to the endpoints of the intervals as in Step 3 of the proof of Lemma 2.2, and

$$b(t, X_t^{s,x}) = b(s, x) - \int_s^t c(t', X_{t'}^{s,x}) dB_{t'} \quad (t \in [s, r]).$$

With $(s, x) = (0, 0)$ this is (2.28), and since $0 \leq c \leq K$ the stochastic integral is a square-integrable martingale on $[0, r]$.

General order parameters. Let $\gamma \in \mathcal{U}_r$, let $\gamma_n \in \mathcal{U}_r$ be step approximations, and write $X^n = X^{\gamma_n}$, $b_n = b_{\gamma_n}$, and $c_n = c_{\gamma_n}$. By Lemma 2.7, $c_n \rightarrow c$ uniformly on compact

subsets of $[0, r] \times \mathbb{R}$, and $X^n \rightarrow X$ in every moment, uniformly on $[0, r]$. By (2.24) and the Lipschitz bound $|\partial_x b| \leq K$,

$$|b_n(q, X_q^n) - b(q, X_q)| \leq C\varepsilon_n(1 + |X_q^n|) + K|X_q^n - X_q|,$$

with $\varepsilon_n = \|\gamma_n - \gamma\|_{L^1}$, so $b_n(q, X_q^n) \rightarrow b(q, X_q)$ in L^2 . Moreover $c_n(s, X_s^n) \rightarrow c(s, X_s)$ in probability for each s , by the locally uniform convergence of c_n , the continuity of c , and the convergence of X^n ; since $0 \leq c_n, c \leq K$, dominated convergence and Itô's isometry give $\int_0^q c_n(s, X_s^n) dB_s \rightarrow \int_0^q c(s, X_s) dB_s$ in L^2 . The identity (2.28) for γ_n therefore passes to the limit, almost surely for each fixed $q \in [0, r]$. Both sides of (2.28) have continuous paths, since b and X are continuous and the stochastic integral has a continuous version, so the identity holds simultaneously for all rational q on one event of probability one, and then for all $q \in [0, r]$ by continuity. $\square$

The martingale identity (2.28) is the hard-wall form of [7, Lemma 2]. Part II, Lemma 2.9(b) states the martingale property of $b(q, X_q)$ without the stochastic integral.

The first variation. For $\gamma \in \mathcal{U}_r$ and $q \in [0, r]$ set

$$D(q) = \alpha \mathbb{E} b(q, X_q)^2 - \int_0^q \lambda(t)^{-2} dt, \quad S(q) = \alpha \lambda(q)^2 \mathbb{E} c(q, X_q)^2. \quad (2.29)$$

The first term of D comes from the energy term of $\mathcal{P}_\kappa$ and the second from the entropy term, as the proof of Part II, Proposition 2.11(a) shows. Since $v = \lambda c$, these are the functions D and S of Part II, (2.22). The function D and the first variation below are the basic tools of Chapter 6.

Lemma 2.9. *Let $r \in [0, 1)$, $\kappa \in \mathbb{R}$, and $\gamma, \tilde{\gamma} \in \mathcal{U}_r$. The function D is continuously differentiable on $[0, r]$ with*

$$D' = \frac{S - 1}{\lambda^2}.$$

Moreover, with $\gamma_\theta = \gamma + \theta(\tilde{\gamma} - \gamma) \in \mathcal{U}_r$ for $\theta \in [0, 1]$,

$$\left. \frac{d}{d\theta} \mathcal{P}_\kappa(\gamma_\theta) \right|_{\theta=0+} = \frac{1}{2} \int_0^r (\tilde{\gamma} - \gamma)(q) D(q) dq. \quad (2.30)$$

Proof. The formula for D' is among the generator identities of Part II, Lemma 2.9(c)(i), and (2.30) is the first equality in Part II, (2.28) (Proposition 2.11(a)); both hold for all $\gamma, \tilde{\gamma} \in \mathcal{U}_r$ and $\kappa \in \mathbb{R}$. $\square$

Formula (2.30) is the hard-wall analogue of the directional derivative of the Parisi functional of mixed p -spin models computed by W.-K. Chen [38, Theorem 2]; see also [55, Proposition 6.8]. The first variation of El Alaoui and Sellke [54, (4.15)] involves the same function D ; the beginning of this chapter compares the two. The formula for D'

is the analogue of [8, Proposition 3], where Auffinger and W.-K. Chen differentiate the corresponding function for mixed p -spin models. Talagrand computed the derivative of the Parisi functional when the order parameter is transported by a map close to the identity [126].

The upper bound

This chapter proves that $\mathcal{P}_U(\gamma)$ bounds the free energy from above for every order parameter γ and every activation U in class (A) (Theorem 3.1), and deduces upper bounds for the feasible volume and for the maximal margin. Lemma 3.2 is a change of measure on a Ruelle probability cascade with marks, which turns Gibbs averages over the leaves into expectations along a Markov chain, and Corollary 3.3 deduces that conditional overlaps are nondecreasing in the branching level. Lemma 3.4 computes a Gaussian cascade exactly. Lemma 3.5 gives the derivatives of the interpolating free energy. Lemma 3.6 proves concentration of the free energy and computes its derivatives in the perturbation parameters, and Lemmas 3.7 and 3.8 use it to control the conditional covariance of the two overlaps. Lemma 3.9 bounds the spherical term by the Crisanti–Sommers entropy. Proposition 3.10 combines these facts through a maximum principle in the interpolation time. Corollary 3.11 passes to the hard wall, with a bounded source term, and Lemma 3.13 and Corollary 3.14, with Lemma 3.12, bound the maximal margin.

Step 1 of §1.3 outlines the argument and introduces the notation used below. The proof follows that of [92, Theorem 4.1], a step in Mourrat’s proof that the Hamilton–Jacobi limit bounds the free energy of bipartite models from above [92, Theorem 1.1]. As there, the atoms of the order parameter have equal mass, the derivative in t is a sum of products of conditional overlap means and a conditional covariance (see [92, (2.23)]), and Mourrat’s perturbation enforces the Ghirlanda–Guerra identities only at contact points (§§3.3 and 3.4). Four things differ. The row side carries a non-Gaussian activation U , with no sign condition on U' , in place of Mourrat’s bilinear Gaussian energy. Only the spin side has free cascade levels, so the viscosity argument becomes a maximum principle in t for the maximum over p (§3.7). The monotonicity of the response overlaps ζ_l follows from a change of measure on a marked cascade; its Gaussian analogue is the monotonicity of the normalized derivatives of the enriched free energy in the cascade parameters [92, Lemma 2.4] (see also [128, Proposition 14.3.2]). The bound closes on the Parisi functional $\mathcal{P}_U(\gamma_h)$ instead of the solution of a Hamilton–Jacobi equation.

Throughout this chapter, $b_a = g_a/\sqrt{N}$ for the rows g_a of $\mathbf{G}$, $\alpha_N = M/N \rightarrow \alpha$, and $\alpha' = \sup_N \alpha_N < \infty$. For U in class (A) we write $U_{\max} = \sup U$ and $L = \|U'\|_\infty$.

Theorem 3.1. *Let U belong to class (A). For every $\gamma \in \mathcal{U}$,*

$$\limsup_{N \rightarrow \infty} F_N(U) \leq \mathcal{P}_U(\gamma). \quad (3.1)$$

Moreover, $\mathbb{P}\{f_N(U) > \mathcal{P}_U(\gamma) + \varepsilon\} \rightarrow 0$ for every $\varepsilon > 0$.

Chapter 5 combines Theorem 3.1 with the lower bound of Chapter 4.

3.1 A change of measure on a marked cascade

The interpolation places Gaussian fields on the nodes of a Ruelle probability cascade. Its free energy is an iterated logarithmic moment over the levels of the tree, and to differentiate it we need to express averages over the cascade weights as expectations along a Markov chain on the marks. The following change-of-measure identity does this. It also gives the law of the branching level of two leaves sampled from the weights, and it shows that conditional overlaps are nondecreasing in the level.

The cascade. Fix $k \geq 0$ and $0 < \theta_1 < \dots < \theta_k < 1$, and set $\theta_0 = 0$, $\theta_{k+1} = 1$, and $w_l = \theta_{l+1} - \theta_l$ for $0 \leq l \leq k$. The nodes of depth l are the strings $\beta \in \mathbb{N}_*^l$, with the empty string as the root, and the leaves are the nodes of depth k . For a leaf β we write $\beta|l$ for its ancestor at depth l , and for two leaves we put $\beta \wedge \beta' = \max\{l : \beta|l = \beta'|l\}$. At every node of depth $l-1$, independently, place a Poisson point process on $(0, \infty)$ with intensity measure $\theta_l x^{-1-\theta_l} dx$. Its points, in decreasing order, are attached to the edges from that node to its children; π_β denotes the point attached to the edge into the node β . For a leaf β put

$$\Pi_\beta = \prod_{l=1}^k \pi_{\beta|l}, \quad \Pi = \sum_{\beta} \Pi_\beta, \quad v_\beta = \frac{\Pi_\beta}{\Pi}.$$

The weights (v_β) are the Ruelle probability cascade [112]. For $k = 0$ the root is the only leaf, and $\Pi = v = 1$.

Independently of the cascade, attach to each node β of depth $l \geq 1$ a random *mark* e_β with law P_l on a standard Borel space; marks at different nodes are independent. The root mark e_0 is fixed for now; below it contains the disorder shared by all leaves. For a leaf β write $e_{[\beta]} = (e_0, e_{\beta|1}, \dots, e_{\beta|k})$ for the marks along its path. Let $X_k(e_0, \dots, e_k)$ be a measurable function of a path of marks, and define backwards, for $l = k, \dots, 1$,

$$X_{l-1}(e_0, \dots, e_{l-1}) = \frac{1}{\theta_l} \log \int e^{\theta_l X_l} dP_l(e_l), \quad K_l(de_l | e_0, \dots, e_{l-1}) = e^{\theta_l(X_l - X_{l-1})} P_l(de_l). \quad (3.2)$$

We assume that every X_l is finite on every path. Then each K_l is a probability kernel. We write $\mathbb{E}_K$ for expectation over the Markov chain $(e_0, e_1, \dots, e_k)$ started from e_0 with the kernels $K_1, \dots, K_k$, and, for $0 \leq l \leq k$, $\mathbb{E}_{K,l}$ for expectation over two paths (e, e') that coincide up to depth l , follow the chain up to depth l , and then continue independently with the kernels $K_{l+1}, \dots, K_k$.

Lemma 3.2. *The quantities Π and $W = \sum_{\beta} v_{\beta} e^{X_k(e_{[\beta]})}$ are positive and finite almost surely. Let $\omega_{\beta} = v_{\beta} e^{X_k(e_{[\beta]})} / W$ be the associated Gibbs weights on the leaves. Then, for every measurable φ that is nonnegative, or integrable under the law on the right side,*

$$\mathbb{E} \sum_{\beta} \omega_{\beta} \varphi(e_{[\beta]}) = \mathbb{E}_{\mathbf{K}} \varphi(e), \quad (3.3)$$

$$\mathbb{E} \sum_{\beta, \beta'} \omega_{\beta} \omega_{\beta'} \mathbf{1}_{\{\beta \wedge \beta' = l\}} \varphi(e_{[\beta]}, e_{[\beta']}) = w_l \mathbb{E}_{\mathbf{K}, l} \varphi(e, e'), \quad 0 \leq l \leq k. \quad (3.4)$$

Moreover, conditionally on every value of the root mark e_0 ,

$$\log W = X_0(e_0) + \log \Pi' - \log \Pi, \quad \Pi' \stackrel{d}{=} \Pi, \quad \mathbb{E} |\log \Pi|^2 < \infty. \quad (3.5)$$

In words, (3.3) says that the path of marks of a leaf sampled from the Gibbs weights is distributed as the chain with the kernels K_l , and (3.4) says that two independently sampled leaves branch at depth l with probability w_l , whatever the function X_k , and that their paths then share the first l marks and continue independently.

These are known properties of Ruelle probability cascades, and they rest on the invariance property of Bolthausen and Sznitman [19]; see [98, Chapter 2]. With marks and the kernels K_l , (3.3) and (3.4) correspond to [103, Theorems 6 and 7], and (3.5) to [103, (3.12)–(3.13)]; Panchenko and Talagrand use them to write Guerra's interpolation on a cascade. See also [4]. Steps 1 and 3 of the proof follow [103, Lemmas 2 and 3], and Step 4 follows Mourrat's derivation of the law of $\beta \wedge \beta'$ by Gaussian integration by parts [93], which adapts the derivation of [98, (2.82)].

Proof. For $k = 0$ we have $W = e^{X_0(e_0)}$, $\omega = 1$, and $w_0 = 1$, and all assertions hold with $\Pi = \Pi' = 1$. Let $k \geq 1$. We divide the proof into four steps.

STEP 1. A Poisson computation. Let (x_n) be a Poisson process with intensity $\theta x^{-1-\theta} dx$, let (e_n) be independent marks with law P , and let f be measurable with $c = \int e^{\theta f} dP < \infty$. By the mapping and marking theorems for Poisson processes [81, Chapters 2 and 5], the pairs $(x_n e^{f(e_n)}, e_n)$ form a Poisson process with intensity

$$c \theta y^{-1-\theta} dy \otimes c^{-1} e^{\theta f(e)} P(de).$$

Indeed, for a test function χ the substitution $y = x e^{f(e)}$ gives

$$\iint \chi(x e^{f(e)}, e) \theta x^{-1-\theta} dx P(de) = \iint \chi(y, e) \theta y^{-1-\theta} e^{\theta f(e)} dy P(de).$$

After the first coordinate is multiplied by $c^{-1/\theta}$, the first factor becomes $\theta y^{-1-\theta} dy$ again. Since the intensity is a product, after ranking the first coordinates the marks are independent with law $c^{-1} e^{\theta f} P$ and independent of the ranked points. If each mark also carries

an independent subtree and f depends only on the first component of the mark, the conditional law of the subtree given the first component is unchanged.

STEP 2. *The totals.* The sum S_θ of the points of one process satisfies $\mathbb{E}e^{-uS_\theta} = \exp\{-\Gamma(1-\theta)u^\theta\}$ for $u \geq 0$, so it is positive and finite almost surely. The identities

$$\begin{aligned} x^{-a} &= \frac{1}{\Gamma(a)} \int_0^\infty u^{a-1} e^{-ux} du \quad (a > 0), \\ x^a &= \frac{a}{\Gamma(1-a)} \int_0^\infty (1 - e^{-ux}) u^{-a-1} du \quad (0 < a < 1), \end{aligned}$$

applied to $x = S_\theta$ and integrated, show that S_θ has negative moments of all orders and positive moments of all orders $a < \theta$; the second integral converges at $u = 0$ because $1 - e^{-\Gamma(1-\theta)u^\theta}$ is of order u^θ . In particular $\mathbb{E}|\log S_\theta|^2 < \infty$. Next, Π is a deterministic multiple of a copy of S_{θ_1} . If S_n denotes the total of the subtree below the n -th child of the root, then $\Pi = \sum_n x_n S_n$, and by Step 1 with $f = \log S_n$ the points $x_n S_n$ form a Poisson process with intensity $\mathbb{E}[S^{\theta_1}] \theta_1 x^{-1-\theta_1} dx$. By induction on k , the subtree total S is a multiple of a copy of S_{θ_2} , so $\mathbb{E}S^{\theta_1} < \infty$ because $\theta_1 < \theta_2$. This proves the assertions about Π .

STEP 3. *The change of measure.* For a node β of depth $l \geq 1$ with parent β^- , replace the edge weight by

$$\tilde{\pi}_\beta = \pi_\beta e^{X_l(e_{[\beta]}) - X_{l-1}(e_{[\beta^-]})},$$

where $e_{[\beta]}$ denotes the marks along the path from the root to β . Given the ancestors of β , the multiplier depends only on the mark e_β , and $\int e^{\theta_l(X_l - X_{l-1})} dP_l = 1$ by (3.2). Apply Step 1 at the root with $c = 1$, carrying the independent subtree of each child as part of its mark, and then apply it recursively below each child. By induction on the depth, after ranking at each node the modified weights $\tilde{\pi}$ form a cascade with the same law as the original one and independent of the marks, while the mark of a child of depth l has conditional law K_l given its ancestors. The relabeling by ranks preserves the ancestry. The multipliers telescope along a path:

$$\Pi_\beta e^{X_k(e_{[\beta]})} = e^{X_0(e_0)} \prod_{l=1}^k \tilde{\pi}_{\beta|l}.$$

Summing over β gives $W = e^{X_0(e_0)} \Pi' / \Pi$ with $\Pi' = \sum_\beta \prod_l \tilde{\pi}_{\beta|l}$, which has the law of Π . This is (3.5), including the positivity and finiteness of W . Moreover, $\omega_\beta = \prod_l \tilde{\pi}_{\beta|l} / \Pi'$ are the weights of the relabeled cascade, which is independent of the marks. This proves (3.3). It also proves (3.4) with the coefficient $\bar{w}_l = \mathbb{E} \sum_{\beta \wedge \beta' = l} v_\beta v_{\beta'}$ in place of w_l , because, given the relabeled cascade, the marks along the paths to two leaves with $\beta \wedge \beta' = l$ share the first l marks and then continue independently.

STEP 4. The branching probabilities. The coefficient $\bar{w}_l$ does not depend on the marks, so we may compute it with a convenient choice. Fix $1 \leq l \leq k$ and $s > 0$, let the mark of each node ν of depth l be a standard Gaussian variable G_ν (the other marks play no role), and take $X_k(e_{[\beta]}) = sG_{\beta|l}$. Then $X_m = sG_{\beta|l}$ for $m \geq l$ and $X_{l-1} = \theta_l s^2/2$, so K_l is the law $N(\theta_l s, 1)$, and (3.3) gives

$$\mathbb{E} \sum_{|\nu|=l} \omega_\nu G_\nu = \theta_l s, \quad \omega_\nu = \sum_{\beta|l=\nu} \omega_\beta.$$

The sum over nodes is absolutely integrable, by (3.3) applied to the path function $|G_{\beta|l}|$. Conditionally on all other variables, $\omega_\nu = Ae^{sG_\nu}/(Ae^{sG_\nu} + B)$ with A, B independent of G_ν , so Gaussian integration by parts gives $\mathbb{E}[G_\nu \omega_\nu] = s \mathbb{E}[\omega_\nu(1 - \omega_\nu)]$. Summing over ν and using $\sum_\nu \omega_\nu = 1$ gives $\mathbb{E} \sum_\nu \omega_\nu^2 = 1 - \theta_l$. By (3.4) with the coefficients $\bar{w}_m$ and $\varphi = 1$, the left side equals $\sum_{m \geq l} \bar{w}_m$. Hence $\sum_{m \geq l} \bar{w}_m = 1 - \theta_l$ for $1 \leq l \leq k$, and also for $l = 0$ since $\sum_m \bar{w}_m = 1$. Taking differences gives $\bar{w}_l = \theta_{l+1} - \theta_l = w_l$. $\square$

We write $\langle \cdot \rangle$ for averages under the Gibbs weights and their independent replicas, that is, leaves $\beta^1, \beta^2, \dots$ sampled independently from (ω_β) . Conditional averages given $\beta^1 \wedge \beta^2 = l$ are taken with respect to the probability $\mathbb{E} \langle \cdot \mathbf{1}_{\{\beta^1 \wedge \beta^2 = l\}} \rangle / w_l$.

Corollary 3.3. *Let f be a bounded function of a path of marks with values in a Euclidean space, and let $F_l = \mathbb{E}_K[f \mid e_0, \dots, e_l]$. Then*

$$\mathbb{E} \langle f(e_{[\beta^1]}) \cdot f(e_{[\beta^2]}) \mid \beta^1 \wedge \beta^2 = l \rangle = \mathbb{E}_K \|F_l\|^2,$$

and the right side is nonnegative and nondecreasing in l . Moreover,

$$\mathbb{E}[\log W \mid e_0] = X_0(e_0), \quad \text{Var}(\log W \mid e_0) \leq 4 \text{Var}(\log \Pi). \quad (3.6)$$

Proof. In (3.4), condition on the common part of the two paths. Their continuations are independent, so the conditional expectation of $f(e) \cdot f(e')$ is $\|F_l\|^2$. The sequence (F_l) is a martingale for the chain, so $\mathbb{E}_K \|F_l\|^2$ is nondecreasing by conditional Jensen. For (3.6), take conditional expectations in (3.5), using $\Pi' \stackrel{d}{=} \Pi$ given e_0 . For the variance, center both logarithms at $\mathbb{E} \log \Pi$ and use $(x - y)^2 \leq 2x^2 + 2y^2$; independence of Π and Π' is not needed. $\square$

In Lemma 3.5(iv) we apply the monotonicity in l to the response overlap of the activation.

3.2 Gaussian cascades

Both bounds on the free energy use one explicit computation: a Gaussian field on a cascade, integrated against a Gaussian weight at each leaf. In this chapter it bounds

the spherical term (Lemma 3.9); in Chapter 4 it evaluates the cavity term. The useful observation is that the field along a sampled path, divided by its precision, is a Gaussian martingale. Its variances give every overlap of the added coordinates, and the normalization of their squared norm identifies the value with a Crisanti–Sommers entropy.

Let $\mathcal{K} = \{p \in \mathbb{R}^{k+1} : 0 \leq p_0 \leq \dots \leq p_k\}$, and put $\Delta p_0 = p_0$ and $\Delta p_l = p_l - p_{l-1}$ for $1 \leq l \leq k$. For $x \in \mathcal{K}$ with $x_k < 1$ let

$$\gamma_x(q) = \sum_{l=0}^k w_l \mathbf{1}_{\{q \geq x_l\}}, \quad \lambda_{\gamma_x}(s) = \int_s^1 \gamma_x(q) dq = 1 - \sum_{l=0}^k w_l \max(x_l, s);$$

$\gamma_x \in \mathcal{U}$ is the order parameter with atoms of mass w_l at x_l .

Fix $n \geq 1$ and $p \in \mathcal{K}$. Let (y_ν) be independent standard Gaussian vectors in $\mathbb{R}^n$ indexed by all nodes ν of the tree, including the root, independent of the cascade, and put

$$Y_\beta = \sum_{l=0}^k \sqrt{\Delta p_l} y_{\beta|l}, \quad Y_{\beta,l} = \sum_{j=0}^l \sqrt{\Delta p_j} y_{\beta|j} \quad (0 \leq l \leq k),$$

for a leaf β . The coordinates of Y are independent, and each has covariance $p_{\beta \wedge \beta'}$ between the leaves β and β' . We call Y the *Gaussian cascade field with levels p* . Summation by parts gives

$$\sum_{l=1}^k \theta_l \Delta p_l = p_k - \sum_{l=0}^k w_l p_l. \quad (3.7)$$

For $b > \sum_{l=1}^k \theta_l \Delta p_l$ define, for $0 \leq l \leq k$,

$$d_l = b - \sum_{j=l+1}^k \theta_j \Delta p_j, \quad m_l = \frac{p_0}{d_0^2} + \sum_{j=1}^l \frac{\Delta p_j}{d_{j-1} d_j}, \quad \mathcal{G}(b, p) = \frac{p_0}{2d_0} + \frac{1}{2} \sum_{j=1}^k \frac{1}{\theta_j} \log \frac{d_j}{d_{j-1}}, \quad (3.8)$$

and put $\varrho = m_k + 1/b$. Since $d_l - d_{l-1} = \theta_l \Delta p_l \geq 0$ and $d_0 = b - p_k + \sum_l w_l p_l > 0$, we have $0 < d_0 \leq \dots \leq d_k = b$ and $0 \leq m_0 \leq \dots \leq m_k$. We call d_l the *precision* at depth l . The top precision b is a scalar and is unrelated to the rows b_a . Let ν_n be the standard Gaussian measure on $\mathbb{R}^n$, and let $\mathcal{M}$ be the random probability measure on pairs (β, ε) of a leaf and a vector of $\mathbb{R}^n$ given by

$$\mathcal{M}(\beta, d\varepsilon) \propto v_\beta \nu_n(d\varepsilon) \exp \left\{ \varepsilon \cdot Y_\beta - \frac{b-1}{2} \|\varepsilon\|^2 \right\}.$$

We write $\mathbb{E}_{\mathcal{M}}$ and $\mathbb{P}_{\mathcal{M}}$ for expectation and probability over the cascade, the field Y , and independent samples $(\beta^1, \varepsilon^1), (\beta^2, \varepsilon^2), \dots$ from $\mathcal{M}$, and we put $J = \beta^1 \wedge \beta^2$. Conditional expectations given $J = l$ are taken with respect to $\mathbb{E}_{\mathcal{M}}[\cdot \mathbf{1}_{\{J=l\}}] / \mathbb{P}_{\mathcal{M}}(J = l)$.

Lemma 3.4. *In this setting the following hold.*

(i) The normalization of $\mathcal{M}$ is finite, and

$$\frac{1}{n} \mathbb{E} \log \sum_{\beta} v_{\beta} e^{\|Y_{\beta}\|^2/(2b)} = \mathcal{G}(b, p), \quad (3.9)$$

$$\frac{1}{n} \mathbb{E} \log \sum_{\beta} v_{\beta} \int \nu_n(d\varepsilon) e^{\varepsilon \cdot Y_{\beta} - (b-1)\|\varepsilon\|^2/2} = \mathcal{G}(b, p) - \frac{1}{2} \log b. \quad (3.10)$$

(ii) Under $\mathbb{P}_{\mathcal{M}}$, the process $M_l = Y_{\beta^1, l}/d_l$, $0 \leq l \leq k$, is a centered Gaussian martingale with independent increments and

$$M_0 \sim N(0, m_0 I_n), \quad M_l - M_{l-1} \sim N(0, (m_l - m_{l-1}) I_n) \quad (1 \leq l \leq k), \quad (3.11)$$

and conditionally on β^1 and Y the vector ε^1 has the law $N(M_k, b^{-1} I_n)$. In particular,

$$\text{Law}_{\mathcal{M}}(\varepsilon^1) = N(0, \varrho I_n), \quad (3.12)$$

$$\mathbb{P}_{\mathcal{M}}(J = l) = w_l, \quad \mathbb{E}_{\mathcal{M}} \left[\frac{\varepsilon^1 \cdot \varepsilon^2}{n} \mid J = l \right] = m_l \quad (0 \leq l \leq k). \quad (3.13)$$

(iii) The levels satisfy the duality identity

$$\varrho p_k - \sum_{l=0}^k w_l m_l p_l = b\varrho - 1. \quad (3.14)$$

(iv) The order parameter $\gamma_{m/\varrho}$, with atoms of mass w_l at m_l/ϱ , belongs to $\mathcal{U}$, its largest atom is $m_k/\varrho = 1 - 1/(\varrho b)$, and

$$\mathcal{G}(b, p) - \frac{1}{2} \log b = \text{CS}(\gamma_{m/\varrho}) + \frac{1}{2} \log \varrho. \quad (3.15)$$

If $\varrho = 1$, then moreover, for $0 \leq l \leq k$,

$$\lambda_{\gamma_m}(m_l) = \frac{1}{d_l}, \quad p_l = \int_0^{m_l} \frac{ds}{\lambda_{\gamma_m}(s)^2}, \quad \text{CS}(\gamma_m) = \mathcal{G}(b, p) - \frac{1}{2} \log b, \quad (3.16)$$

$$p_k - \sum_{l=0}^k w_l m_l p_l = b - 1, \quad (3.17)$$

and $d_l \geq 1$, $1 \leq b \leq 1 + p_k$, and $m_k \leq 1 - (1 + p_k)^{-1}$.

(v) If Y^c is a real Gaussian cascade field with levels $c \in \mathcal{K}$ (the case $n = 1$), then

$$\mathbb{E} \log \sum_{\beta} v_{\beta} e^{Y_{\beta}^c} = \frac{1}{2} \left(c_k - \sum_{l=0}^k w_l c_l \right). \quad (3.18)$$

By (3.4), the branching probabilities w_l do not depend on the function integrated at the leaves. In particular, they remain w_l when the ε -integral in the definition of $\mathcal{M}$ is restricted to a ball. The restricted measure, used in §4.5, is defined for every b , including the case $d_0 \leq 0$ in which $\mathcal{M}$ is not defined. In (3.15), ϱ is the mean of $\|\varepsilon^1\|^2/n$ under $\mathbb{P}_{\mathcal{M}}$, by (3.12); when it equals one, the value (3.10) is exactly the Crisanti–Sommers entropy of the law of the conditional overlaps m_J .

The value $\mathcal{G}(b, p)$, together with the term $(b - 1 - \log b)/2$, is the spherical part of the expression minimized over b in the Parisi formula for spherical models of Talagrand [125] and W.-K. Chen [37]. Parts (iii) and (iv) identify it with a Crisanti–Sommers entropy, in line with Talagrand’s theorem that the spherical Parisi formula agrees with the Crisanti–Sommers formula [44, 125]. Part (v) is the linear case computed by Arguin [4]; see also [100, (22)].

Proof. We divide the proof into five steps. Steps 1, 2, and 5 apply Lemma 3.2 to the Gaussian increments; Steps 3 and 4 are algebraic.

STEP 1. The recursion. We use the marks $e_0 = \sqrt{p_0} y_\emptyset$ at the root and $e_v = \sqrt{\Delta p_l} y_v$ at a node v of depth $l \geq 1$, with laws $P_l = N(0, \Delta p_l I_n)$, and write $x_l = e_0 + e_1 + \dots + e_l$ for the partial sums along a path, so that $Y_{\beta, l} = x_l$ along the path of β . Since $\int v_n(d\varepsilon) e^{\varepsilon \cdot x - (b-1)\|\varepsilon\|^2/2} = b^{-n/2} e^{\|x\|^2/(2b)}$ for $b > 0$, the leaf function of (3.10) is $X_k = \|x_k\|^2/(2b) - \frac{n}{2} \log b$, and that of (3.9) is $\|x_k\|^2/(2b)$. For one coordinate, $\theta \in (0, 1)$, $\Delta \geq 0$, $d > \theta\Delta$, and a standard Gaussian variable G , completing the square gives

$$\frac{1}{\theta} \log \mathbb{E} \exp \left(\frac{\theta(x + \sqrt{\Delta} G)^2}{2d} \right) = \frac{x^2}{2(d - \theta\Delta)} + \frac{1}{2\theta} \log \frac{d}{d - \theta\Delta}.$$

Since $d_l - \theta_l \Delta p_l = d_{l-1} > 0$, backward induction in (3.2) shows that $X_l(e_0, \dots, e_l) = \|x_l\|^2/(2d_l) + a_l$ with $a_{l-1} = a_l + \frac{n}{2\theta_l} \log(d_l/d_{l-1})$, where $a_k = -\frac{n}{2} \log b$ for (3.10) and $a_k = 0$ for (3.9). All values are finite, so Lemma 3.2 applies. By (3.5), n times the left side of (3.10) equals $\mathbb{E} X_0(e_0) = np_0/(2d_0) + a_0 = n[\mathcal{G}(b, p) - \frac{1}{2} \log b]$, and the same computation with $a_k = 0$ gives (3.9). This proves (i).

STEP 2. The law of a sampled path. Under $\mathcal{M}$ the leaf β^1 is sampled from the Gibbs weights of Lemma 3.2 with the leaf function X_k of (3.10), and, given the leaf and Y , the vector ε^1 has density proportional to $e^{\varepsilon \cdot Y_{\beta^1} - b\|\varepsilon\|^2/2}$ with respect to Lebesgue measure, that is, the law $N(Y_{\beta^1}/b, b^{-1}I_n)$. By (3.2), the kernel $K_l(d\varepsilon_l \mid e_0, \dots, e_{l-1})$ has density proportional to $\exp\{\theta_l \|x_{l-1} + e_l\|^2/(2d_l) - \|e_l\|^2/(2\Delta p_l)\}$. The coefficient of $-\|e_l\|^2/2$ in the exponent is $1/\Delta p_l - \theta_l/d_l = d_{l-1}/(d_l \Delta p_l)$, so K_l is the Gaussian law with mean $\theta_l \Delta p_l x_{l-1}/d_{l-1}$ and covariance $(\Delta p_l d_l/d_{l-1}) I_n$ (the point mass at 0 if $\Delta p_l = 0$). Since $d_{l-1} + \theta_l \Delta p_l = d_l$, under the chain

$$x_l = \frac{d_l}{d_{l-1}} x_{l-1} + \left(\frac{\Delta p_l d_l}{d_{l-1}} \right)^{1/2} \xi_l, \quad \text{hence} \quad \frac{x_l}{d_l} = \frac{x_{l-1}}{d_{l-1}} + \left(\frac{\Delta p_l}{d_{l-1} d_l} \right)^{1/2} \xi_l,$$

with independent standard Gaussian vectors ξ_l that are independent of the root mark. The root mark keeps its law $N(0, p_0 I_n)$, so $x_0/d_0 \sim N(0, (p_0/d_0^2)I_n)$. By (3.3), averaged over the root mark, the process $M_l = x_l/d_l$ along the path of β^1 has this law, which is (3.11) by the definition of m_l . Since $Y_{\beta^1}/b = M_k$, the vector ε^1 is $M_k + b^{-1/2}\xi'$ with an independent standard Gaussian vector ξ' , which gives (3.12). By (3.4), $J = l$ with probability w_l , and the two paths then share $x_0, \dots, x_l$ and continue independently with the same kernels. Given x_l , the vectors M_k^1 and M_k^2 of the two paths are therefore independent, each with conditional mean $M_l = x_l/d_l$, and $\mathbb{E}_{\mathcal{M}}[\varepsilon^1 \cdot \varepsilon^2 \mid J = l] = \mathbb{E}\|M_l\|^2 = nm_l$. All functions involved are polynomials in Gaussian vectors, so the integrability required in Lemma 3.2 holds. This proves (ii).

STEP 3. The duality identity. Let $\lambda_\varrho(s) = \varrho - \sum_l w_l \max(m_l, s)$ for $s \in [0, m_k]$. This function is nonincreasing, with $\lambda'_\varrho(s) = -\sum_{l:m_l < s} w_l$ for $s \notin \{m_l\}$; it is constant on $[0, m_0]$ and affine with slope $-\theta_j$ on $[m_{j-1}, m_j]$. Moreover $\lambda_\varrho(m_k) = \varrho - m_k = 1/b = 1/d_k$, and

$$\lambda_\varrho(m_{j-1}) - \lambda_\varrho(m_j) = \theta_j(m_j - m_{j-1}) = \frac{\theta_j \Delta p_j}{d_{j-1} d_j} = \frac{1}{d_{j-1}} - \frac{1}{d_j},$$

since $d_j - d_{j-1} = \theta_j \Delta p_j$. By downward induction, $\lambda_\varrho(m_j) = 1/d_j$ for all j ; in particular $\lambda_\varrho \geq 1/b > 0$ on $[0, m_k]$. On $[0, m_0]$ we have $\lambda_\varrho = 1/d_0$, and on $[m_{j-1}, m_j]$ we have $(1/\lambda_\varrho)' = \theta_j/\lambda_\varrho^2$ and $(\log \lambda_\varrho)' = -\theta_j/\lambda_\varrho$. Hence the constant piece contributes $m_0 d_0^2 = p_0$ and $m_0 d_0 = p_0/d_0$, and the j -th affine piece contributes $(d_j - d_{j-1})/\theta_j = \Delta p_j$ and $\theta_j^{-1} \log(d_j/d_{j-1})$, to the two integrals

$$\int_0^{m_l} \frac{ds}{\lambda_\varrho(s)^2} = p_l \quad (0 \leq l \leq k), \quad \int_0^{m_k} \frac{ds}{\lambda_\varrho(s)} = 2\mathcal{G}(b, p). \quad (3.19)$$

(If $m_{j-1} = m_j$, then $\Delta p_j = 0$ and $d_j = d_{j-1}$, and both contributions vanish.) Insert the first identity of (3.19) into the left side of (3.14) and use Fubini's theorem:

$$\varrho p_k - \sum_l w_l m_l p_l = \int_0^{m_k} \frac{\varrho - \sum_{l:m_l > s} w_l m_l}{\lambda_\varrho(s)^2} ds.$$

The numerator equals $\lambda_\varrho(s) + s \sum_{l:m_l \leq s} w_l = \lambda_\varrho(s) - s \lambda'_\varrho(s)$ for almost every s , so the integrand is $(s/\lambda_\varrho(s))'$, and the integral is $m_k/\lambda_\varrho(m_k) = b m_k = b\varrho - 1$. This proves (3.14) and (iii).

STEP 4. The entropy. Suppose first that $\varrho = 1$. Then $m_k = 1 - 1/b < 1$, $\gamma_m \in \mathcal{U}$, and $\lambda_\varrho = \lambda_{\gamma_m}$ on $[0, m_k]$. Step 3 gives the first two identities in (3.16), and (3.14) becomes (3.17). By (3.19) and the definition of CS with the cutoff $m_k \in [Q_{\gamma_m}, 1)$,

$$\text{CS}(\gamma_m) = \frac{1}{2} \int_0^{m_k} \frac{ds}{\lambda_{\gamma_m}(s)} + \frac{1}{2} \log(1 - m_k) = \mathcal{G}(b, p) - \frac{1}{2} \log b.$$

Moreover $1/d_l = \lambda_{\gamma_m}(m_l) \leq \lambda_{\gamma_m}(0) \leq 1$, so $b = d_k \geq d_l \geq 1$; since $m_l, p_l \geq 0$, (3.17) gives $b - 1 \leq p_k$, and then $m_k = 1 - 1/b \leq 1 - (1 + p_k)^{-1}$. For general $\varrho > 0$, multiply the levels and the top precision by ϱ . The parameters $(\varrho b, \varrho p)$ satisfy the standing condition, and by (3.8) they have precisions ϱd_l , conditional levels m_l/ϱ , and the same value $\mathcal{G}(\varrho b, \varrho p) = \mathcal{G}(b, p)$, because p_0/d_0 and the ratios d_j/d_{j-1} do not change. Their top level satisfies $m_k/\varrho + 1/(\varrho b) = 1$. The case $\varrho = 1$, applied to them, shows that $\gamma_{m/\varrho} \in \mathcal{U}$ with largest atom $m_k/\varrho = 1 - 1/(\varrho b)$, and

$$\text{CS}(\gamma_{m/\varrho}) = \mathcal{G}(\varrho b, \varrho p) - \frac{1}{2} \log(\varrho b) = \mathcal{G}(b, p) - \frac{1}{2} \log b - \frac{1}{2} \log \varrho.$$

This proves (iv).

STEP 5. *The linear field.* Apply Lemma 3.2 with the marks of Step 1 for the levels c (with $n = 1$) and the leaf function $X_k = x_k$. For $\theta \in (0, 1)$ and $\Delta \geq 0$, $\theta^{-1} \log \mathbb{E} e^{\theta(x + \sqrt{\Delta} G)} = x + \theta\Delta/2$, so $X_l = x_l + \frac{1}{2} \sum_{j>l} \theta_j \Delta c_j$ and $\mathbb{E} X_0(e_0) = \frac{1}{2} \sum_{j=1}^k \theta_j \Delta c_j$. By (3.5) and (3.7) for c , this is (3.18). For $k = 0$ both sides of (3.18) vanish. $\square$

3.3 The interpolation and its derivatives

We now define the interpolating free energy and compute its derivatives. Throughout this section the activation U is in class (A). From now on we fix an integer $K \geq 2$ and use the cascade of §3.1 with

$$k = K - 1, \quad \theta_l = \frac{l}{K} \quad (0 \leq l \leq K), \quad \text{so that } w_l = \frac{1}{K} \quad (0 \leq l \leq k).$$

We keep writing w_l , which makes the role of each factor visible. We also fix the levels of the row field,

$$h = (h_0, \dots, h_k) \in \mathcal{K}, \quad h_k < 1,$$

so that γ_h is the order parameter with atoms of mass w_l at h_l . The spin field has levels $p \in \mathcal{K}$. At every node β of the cascade, including the root, take independent standard Gaussian vectors $y_\beta \in \mathbb{R}^N$ and $z_\beta \in \mathbb{R}^M$, independent of $\mathbf{G}$, and set, for a leaf β ,

$$Y_\beta = \sum_{l=0}^k \sqrt{\Delta p_l} y_{\beta|l}, \quad \Xi_\beta = \sum_{l=0}^k \sqrt{\Delta h_l} z_{\beta|l},$$

where $\Delta h_0 = h_0$ and $\Delta h_l = h_l - h_{l-1}$. Thus Y and Ξ are the Gaussian cascade fields with levels p and h in dimensions N and M . The remaining variance $1 - h_k$ of the row field is carried by an auxiliary standard Gaussian vector $\xi \in \mathbb{R}^M$ inside each leaf, so that each coordinate of $\Xi_\beta + \sqrt{1 - h_k} \xi$ has variance one, as does $b_a \cdot \sigma$ for fixed $\sigma \in S_N$. This equality of variances is what makes the diagonal terms cancel when the activation is differentiated in t . Without the cascade, interpolations of this form were used for the perceptron by Talagrand [127, Chapter 2].

The perturbation. To obtain approximate Ghirlanda–Guerra identities we add a small Gaussian field whose covariance is a polynomial in the spin overlap and in the branching level. Fix an enumeration $(\vartheta_n)_{n \geq 1}$ of $\mathbb{Q} \cap [0, 1]$ and an integer $j_+ \geq 1$, and let $\mathcal{I} = \{1, \dots, j_+\}^3$. For $j = (j_1, j_2, j_3) \in \mathcal{I}$ put

$$c_j(R, q) = (\vartheta_{j_1} R + \vartheta_{j_2} q)^{j_3}, \quad \text{so that} \quad |c_j(R, q)| \leq 2^{j_3} \quad (|R| \leq 1, 0 \leq q \leq 1).$$

We use centered Gaussian fields $H^j(\sigma, \beta)$, $j \in \mathcal{I}$, indexed by $\sigma \in S_N$ and leaves β , independent of each other and of all previous variables, with covariance

$$\mathbb{E} H^j(\sigma, \beta) H^j(\sigma', \beta') = N c_j(R, J/k), \quad R = \frac{\sigma \cdot \sigma'}{N}, \quad J = \beta \wedge \beta'. \quad (3.20)$$

This perturbation, together with the prefactor $s_N = N^{-1/16}$ used below, is the one of Mourrat [92, (4.4)–(4.5)]. It has the form of Panchenko’s perturbation for multi-species models [100, (24)–(28)] (see [98] for one species), with the branching level J/k in place of the overlap of a second species. Panchenko averages over random coefficients in the perturbation; here, as in [92], the perturbation coefficients are chosen at a contact point (§3.4). Mourrat proves that such a field exists in [92, Lemma A.2]; the construction below places its Gaussian tensors on the nodes of the cascade, so that they become part of the marks. For integrability we use the following explicit construction. For $m \geq 1$ let $a_{m,0} = 0$ and $a_{m,l} = (l/k)^m - ((l-1)/k)^m$ for $1 \leq l \leq k$, and let $a_{0,0} = 1$ and $a_{0,l} = 0$ for $l \geq 1$. Thus $a_{m,l} \geq 0$ and $\sum_{l \leq J} a_{m,l} = (J/k)^m$ for all $m \geq 0$ and $0 \leq J \leq k$. Attach to each node β independent standard Gaussian tensors $g_\beta^{j,i} \in (\mathbb{R}^N)^{\otimes i}$, $0 \leq i \leq j_3$, and set

$$H^j(\sigma, \beta) = \sum_{i=0}^{j_3} \sum_{l=0}^k c_{j,i,l} N^{-i/2} \langle g_{\beta|l}^{j,i}, \sigma^{\otimes i} \rangle, \quad c_{j,i,l}^2 = N \binom{j_3}{i} \vartheta_{j_1}^i \vartheta_{j_2}^{j_3-i} a_{j_3-i,l}.$$

Only the nodes $\beta|l$ with $l \leq \beta \wedge \beta'$ are shared by two leaves, so the covariance is

$$\sum_{i=0}^{j_3} \binom{j_3}{i} \vartheta_{j_1}^i \vartheta_{j_2}^{j_3-i} N^{1-i} (\sigma \cdot \sigma')^i \sum_{l \leq J} a_{j_3-i,l} = N \sum_{i=0}^{j_3} \binom{j_3}{i} (\vartheta_{j_1} R)^i (\vartheta_{j_2} J/k)^{j_3-i},$$

which is (3.20) by the binomial theorem. Since $a_{m,0} = 0$ unless $m = 0$, only the tensor of order $i = j_3$ has a root component, with $c_{j,j_3,0}^2 = N \vartheta_{j_1}^{j_3} \leq N$.

The interpolating free energy. Let $s_N = N^{-1/16}$ and $x = (x_j)_{j \in \mathcal{I}} \in \mathbb{R}^{\mathcal{I}}$. A configuration is a triple (σ, ξ, β) with $\sigma \in S_N$, $\xi \in \mathbb{R}^M$, and β a leaf, with prior $\mu_N(d\sigma) N(0, I_M)(d\xi) v_\beta$. For $t \in [0, 1]$, $p \in \mathcal{K}$, and $x \in \mathbb{R}^{\mathcal{I}}$ the Hamiltonian is

$$\begin{aligned} \mathcal{H}(\sigma, \xi, \beta) &= \sum_{a=1}^M U(S_a) + Y_\beta \cdot \sigma + s_N \sum_{j \in \mathcal{I}} x_j H^j(\sigma, \beta), \\ S_a &= \sqrt{t} b_a \cdot \sigma + \sqrt{1-t} (\Xi_{\beta,a} + \sqrt{1-h_k} \xi_a). \end{aligned}$$

Let $W = \sum_{\beta} v_{\beta} \iint e^{\mathcal{H}(\sigma, \xi, \beta)} \mu_N(d\sigma) N(0, I_M)(d\xi)$ be the partition function, and put

$$u_N(t, p, x) = \frac{1}{N} \mathbb{E} \log W - \frac{p_k}{2}, \quad \psi_{\sigma, N}(p) = \frac{1}{N} \mathbb{E} \log \sum_{\beta} v_{\beta} \int e^{Y_{\beta} \cdot \sigma} \mu_N(d\sigma) - \frac{p_k}{2}.$$

The function $\psi_{\sigma, N}$ is the spherical part of u_N at $t = 0$ (see (3.25)), and the subtraction of $p_k/2$ normalizes it: $\psi_{\sigma, N}(0) = 0$. This is a marked cascade in the sense of §3.1. The root mark consists of $\mathbf{G}$, $y_{\emptyset}$, $z_{\emptyset}$, and the root tensors; the mark of a node β of depth $l \geq 1$ consists of y_{β} , z_{β} , and the tensors $g_{\beta}^{j,i}$; and $X_k = \log \iint e^{\mathcal{H}} d\mu_N dN(0, I_M)$ is the logarithm of the *leaf integral*. From now on $\langle \cdot \rangle$ denotes the Gibbs measure on configurations, proportional to $v_{\beta} e^{\mathcal{H}}$ times the product prior, together with its independent replicas $(\sigma^l, \xi^l, \beta^l)$. We write

$$R_{12} = \frac{\sigma^1 \cdot \sigma^2}{N}, \quad J = J_{12} = \beta^1 \wedge \beta^2, \quad X_{12} = \frac{1}{M} \sum_{a=1}^M U'(S_a^1) U'(S_a^2),$$

where S_a^l is S_a evaluated at the l -th replica, and

$$r_l = \mathbb{E} \langle R_{12} \mid J = l \rangle, \quad \zeta_l = \alpha_N \mathbb{E} \langle X_{12} \mid J = l \rangle.$$

By Lemma 3.2, $\mathbb{E} \langle \mathbf{1}_{\{J=l\}} \rangle = w_l$ whatever the parameters, and conditional averages given $J = l$ are as in §3.1. For random variables A, B of two replicas we write $\text{Cov}(A, B \mid J = l) = \mathbb{E} \langle AB \mid J = l \rangle - \mathbb{E} \langle A \mid J = l \rangle \mathbb{E} \langle B \mid J = l \rangle$.

The next lemma collects what the comparison uses: continuity, Guerra's formula for the derivative in t , the right derivatives in p in the directions $\mathbf{1}_{\geq m}$, and the monotonicity of ζ_l . For $0 \leq m \leq k$ let $\mathbf{1}_{\geq m} \in \mathcal{K}$ be the vector with coordinates $\mathbf{1}_{\{l \geq m\}}$; moving p in this direction increases Δp_m and leaves the other increments unchanged.

Lemma 3.5. *The following hold.*

- (i) *The function u_N is continuous on $[0, 1] \times \mathcal{K} \times \mathbb{R}^{\mathcal{I}}$, and so is $\mathbb{E} \langle f(R_{12}, J_{12}) \rangle$ for every bounded measurable function f on $[-1, 1] \times \{0, \dots, k\}$.*
- (ii) *For $0 < t < 1$, u_N is differentiable in t , and*

$$\partial_t u_N = -\frac{\alpha_N}{2} \mathbb{E} \langle X_{12} (R_{12} - h_J) \rangle = -\frac{1}{2} \sum_{l=0}^k w_l \left[(r_l - h_l) \zeta_l + \text{Cov}(R_{12}, \alpha_N X_{12} \mid J = l) \right]. \quad (3.21)$$

- (iii) *For every (t, p, x) and $0 \leq m \leq k$, the function $\epsilon \mapsto u_N(t, p + \epsilon \mathbf{1}_{\geq m}, x)$ on $[0, \infty)$ has right derivative at $\epsilon = 0$ equal to*

$$-\frac{1}{2} \mathbb{E} \langle R_{12} \mathbf{1}_{\{J \geq m\}} \rangle = -\frac{1}{2} \sum_{l=m}^k w_l r_l. \quad (3.22)$$

(iv) The conditional response overlaps are ordered:

$$0 \leq \zeta_0 \leq \zeta_1 \leq \cdots \leq \zeta_k \leq \alpha_N L^2. \quad (3.23)$$

Proof. Given its leaf β , a replica (σ, ξ) has the *leaf measure*, proportional to $e^{\mathcal{H}(\sigma, \xi, \beta)}$ times the prior $\mu_N \otimes N(0, I_M)$. It depends only on the path of marks $e_{[\beta]}$, and we write $\langle \cdot \rangle_\beta$ for averages under it. Given the leaves, replicas are independent. We divide the proof into six steps.

STEP 1. Integrability. For a node β of depth m let $\Theta_m(e_\beta)$ be the sum of the Euclidean norms of y_β, z_β , and all tensors $g_\beta^{j,i}$; at the root add the Frobenius norm $\|\mathbf{G}\|_F$. On any compact set of parameters there is C , depending also on N , such that

$$|X_l(e_0, \dots, e_l)| \leq C \left(1 + \sum_{m=0}^l \Theta_m(e_m) \right), \quad 0 \leq l \leq k. \quad (3.24)$$

For $l = k$, the upper bound follows from $U \leq U_{\max}$, $|Y_\beta \cdot \sigma| \leq \sqrt{N}|Y_\beta|$, and $|\langle g, \sigma^{\otimes i} \rangle| \leq N^{i/2} \|g\|$, which hold uniformly in (σ, ξ) . For the lower bound, apply Jensen's inequality to the leaf integral and use $U(y) \geq U(0) - L|y|$, $\int |b_a \cdot \sigma| d\mu_N \leq |b_a|$ (since $\int \sigma \sigma^\top d\mu_N = I_N$), $\int |\xi_a| dN(0, I_M) \leq 1$, $\int Y_\beta \cdot \sigma d\mu_N = 0$, and the bound on the tensors. Backward induction in (3.2) gives (3.24) at every level: Jensen's inequality gives the lower bound, and the upper bound uses that each Θ_m has finite exponential moments of all orders. In particular the hypotheses of Lemma 3.2 hold, and each kernel density $e^{\theta_l(X_l - X_{l-1})}$ is at most $\exp\{C(1 + \sum_{m \leq l} \Theta_m)\}$.

STEP 2. Continuity. Fix the marks. The leaf integral is continuous in (t, p, x) by dominated convergence, since its integrand is continuous in the parameters and bounded, uniformly in (σ, ξ) and on compact parameter sets, by $\exp\{C(1 + \sum_m \Theta_m)\}$. Backward induction in (3.2), with dominated convergence justified by (3.24), shows that every X_l and every kernel density is continuous in the parameters. Since $\mathbb{E} \log W = \mathbb{E} X_0(e_0)$ by (3.5) and $|X_0| \leq C(1 + \Theta_0)$, u_N is continuous. For the second claim in (i), let $\varphi(e, e')$ be the average of $f(\sigma \cdot \sigma' / N, l)$ under the product of the leaf measures of two paths e, e' . It is bounded by $\sup |f|$ and continuous in the parameters. By (3.4), applied for a fixed root mark, $\mathbb{E} \langle f(R_{12}, J) \mathbf{1}_{\{J=l\}} \rangle = w_l \mathbb{E} \mathbb{E}_{K,l} \varphi$, and $\mathbb{E}_{K,l}$ integrates φ against a product of kernel densities, each continuous and bounded by $\exp\{C(1 + \sum_m \Theta_m)\}$. Dominated convergence proves (i).

STEP 3. Differentiation under the recursion. Let ∂ denote the derivative in one of the parameters $t \in (0, 1)$ and Δp_m , the latter in the range $\Delta p_m > 0$. Locally uniformly in the parameters, $|\partial \mathcal{H}| \leq C(1 + |\xi| + \sum_m \Theta_m)$. Hence $\partial X_k = \langle \partial \mathcal{H} \rangle_\beta$, and, bounding $\langle |\xi| \rangle_\beta \leq e^{\sup \mathcal{H} - X_k} \int |\xi| dN(0, I_M)$ with (3.24), we get $|\partial X_k| \leq \exp\{C(1 + \sum_m \Theta_m)\}$. Differentiating (3.2) gives $\partial X_{l-1} = \int \partial X_l dK_l$, and the same kind of bound propagates to

every level, which justifies each differentiation under the integral. Hence $\partial X_0 = \mathbb{E}_K \partial X_k$, and by (3.5) and (3.3)

$$\partial \mathbb{E} \log W = \mathbb{E} \mathbb{E}_K \partial X_k = \mathbb{E} \langle \partial \mathcal{H} \rangle.$$

STEP 4. *Gaussian integration by parts.* Each Gaussian coordinate at a node of depth m enters the leaf integrals of all leaves below that node, and only those. Fix one such coordinate and let it vary in a compact set. Since $|U'| \leq L$ and $|\sigma| = \sqrt{N}$, the derivative of each descendant Hamiltonian in this coordinate is bounded, uniformly over the spin and auxiliary variables. Thus the perturbed descendant leaf integrals are bounded above and below by common finite multiples of their original values, their first derivatives satisfy the same domination, and so do their second derivatives, by the bound on U'' . The series over leaves defining the Gibbs averages therefore converge locally uniformly together with their derivatives, and by (3.3) and (3.24) the sum over nodes of the expected absolute contributions is finite. This justifies Gaussian integration by parts node by node. Summing over nodes and coordinates gives, for every row a and every $0 \leq m \leq k$,

$$\begin{aligned} \mathbb{E} \langle y_{\beta|m} \cdot \sigma \rangle &= \sqrt{\Delta p_m} N (1 - \mathbb{E} \langle R_{12} \mathbf{1}_{\{J \geq m\}} \rangle), \\ \mathbb{E} \langle z_{\beta|m,a} U'(S_a) \rangle &= \sqrt{(1-t) \Delta h_m} (\mathbb{E} \langle U''(S_a) + U'(S_a)^2 \rangle - \mathbb{E} \langle U'(S_a^1) U'(S_a^2) \mathbf{1}_{\{J \geq m\}} \rangle), \\ \mathbb{E} \langle \xi_a U'(S_a) \rangle &= \sqrt{(1-t)(1-h_k)} \mathbb{E} \langle U''(S_a) + U'(S_a)^2 \rangle, \\ \sum_{i=1}^N \mathbb{E} \langle g_{ai} U'(S_a) \sigma_i \rangle &= \sqrt{tN} (\mathbb{E} \langle U''(S_a) + U'(S_a)^2 \rangle - \mathbb{E} \langle U'(S_a^1) U'(S_a^2) R_{12} \rangle). \end{aligned}$$

Here β is the leaf of the sampled configuration. In the first line, the derivative of $\langle \sigma_i \mathbf{1}_{\{\beta|m=\nu\}} \rangle$ in the coordinate $y_{\nu,i}$ produces the one-replica term $\sqrt{\Delta p_m} \langle \sigma_i^2 \mathbf{1}_{\{\beta|m=\nu\}} \rangle$ and the two-replica term $-\sqrt{\Delta p_m} \langle \sigma_i^1 \sigma_i^2 \mathbf{1}_{\{\beta^1|m=\beta^2|m=\nu\}} \rangle$; summing over i and over the nodes ν of depth m gives the formula, because $|\sigma|^2 = N$ and $\mathbf{1}_{\{\beta^1|m=\beta^2|m\}} = \mathbf{1}_{\{J \geq m\}}$. The second and fourth lines are obtained in the same way; there the factor $U'' + (U')^2$ comes from differentiating $U'(S_a) e^{U(S_a)}$. The variable ξ is integrated inside the leaf measure, so in the third line only the one-replica term appears.

STEP 5. *The derivative in t .* For $0 < t < 1$, $\partial_t \mathcal{H} = \sum_a U'(S_a) \partial_t S_a$ with

$$\partial_t S_a = \frac{b_a \cdot \sigma}{2\sqrt{t}} - \frac{\Xi_{\beta,a} + \sqrt{1-h_k} \xi_a}{2\sqrt{1-t}}, \quad b_a \cdot \sigma = N^{-1/2} \sum_i g_{ai} \sigma_i, \quad \Xi_{\beta,a} = \sum_m \sqrt{\Delta h_m} z_{\beta|m,a}.$$

The last three identities of Step 4, together with $\sum_m \Delta h_m = h_k$ and $\sum_m \Delta h_m \mathbf{1}_{\{J \geq m\}} = h_J$, give for each row a

$$\begin{aligned} \mathbb{E} \langle U'(S_a) \partial_t S_a \rangle &= \frac{1}{2} \left(\mathbb{E} \langle U'' + (U')^2 \rangle - \mathbb{E} \langle U'(S_a^1) U'(S_a^2) R_{12} \rangle \right) \\ &\quad - \frac{1}{2} \left(\mathbb{E} \langle U'' + (U')^2 \rangle - \mathbb{E} \langle U'(S_a^1) U'(S_a^2) h_J \rangle \right), \end{aligned}$$

with $U'' + (U')^2$ evaluated at S_a . The one-replica terms cancel because $b_a \cdot \sigma$ and $\Xi_{\beta,a} + \sqrt{1 - h_k} \xi_a$ both have variance one. Summing over a and dividing by N gives the first identity in (3.21), by Step 3; the term $-p_k/2$ of u_N does not depend on t . For the second, split the average according to the value of J and use $\mathbb{E}\langle R_{12} \alpha_N X_{12} \mid J = l \rangle = r_l \zeta_l + \text{Cov}(R_{12}, \alpha_N X_{12} \mid J = l)$. This proves (ii).

STEP 6. *The derivatives in p and the monotonicity.* Moving p along $\mathbf{1}_{\geq m}$ increases Δp_m and p_k at unit rate and leaves the other increments fixed. If $\Delta p_m > 0$, then $\partial_{\Delta p_m} \mathcal{H} = y_{\beta|m} \cdot \sigma / (2\sqrt{\Delta p_m})$, and the first identity of Step 4 and Step 3 show that $\frac{1}{N} \mathbb{E} \log W$ has derivative $\frac{1}{2}(1 - \mathbb{E}\langle R_{12} \mathbf{1}_{\{J \geq m\}} \rangle)$; subtracting $\frac{1}{2}$ for the term $-p_k/2$ gives (3.22). If $\Delta p_m = 0$, this formula holds at $p + \epsilon \mathbf{1}_{\geq m}$ for every $\epsilon > 0$, and by (i), applied with $f(R, l) = R \mathbf{1}_{\{l \geq m\}}$, both u_N and the right side of (3.22) are continuous as $\epsilon \downarrow 0$. The mean value theorem then gives the right derivative at $\epsilon = 0$. Finally, $\mathbb{E}\langle R_{12} \mathbf{1}_{\{J \geq m\}} \rangle = \sum_{l \geq m} w_l r_l$ by the definition of r_l . This proves (iii).

For (iv), fix the root mark and apply Corollary 3.3 to the path function with coordinates $\langle U'(S_a) \rangle_{\beta} / \sqrt{M}$, $a \leq M$, whose norm is at most L . Since replicas are independent given their leaves, $\langle U'(S_a^1) U'(S_a^2) \rangle$ given the two leaves equals $\langle U'(S_a) \rangle_{\beta^1} \langle U'(S_a) \rangle_{\beta^2}$, so for fixed root mark $\mathbb{E}\langle X_{12} \mid J = l \rangle$ is the quantity $\mathbb{E}_K \|F_l\|^2$ of that corollary. It is nonnegative, nondecreasing in l , and at most L^2 . These properties survive averaging over the root mark, because w_l does not depend on it. This proves (3.23). $\square$

The endpoints. At $t = 1$, $p = 0$, and $x = 0$ the Hamiltonian is $\sum_a U(b_a \cdot \sigma)$, which depends neither on ξ nor on the leaf, so $W = Z_N(U)$. At $t = 0$ and $x = 0$ the leaf integral is the product of $\int e^{Y_{\beta} \cdot \sigma} \mu_N(d\sigma)$, which depends only on the vectors y , and of

$$\prod_{a=1}^M \int \exp \{ U(\Xi_{\beta,a} + \sqrt{1 - h_k} \xi_a) \} N(0, 1)(d\xi_a),$$

which depends only on the vectors z ; these are independent components of the marks. For independent F, F' we have $\theta^{-1} \log \mathbb{E} e^{\theta(F+F')} = \theta^{-1} \log \mathbb{E} e^{\theta F} + \theta^{-1} \log \mathbb{E} e^{\theta F'}$, so the recursion (3.2) factorizes into a spin part and a row part, and by (3.5) so does $\mathbb{E} \log W$. The spin part is $N\psi_{\sigma,N}(p) + Np_k/2$. The row part factorizes in the same way over the rows, whose coordinates $z_{\cdot,a}$ of the node vectors are independent. For one row it is the recursion (2.2) for the order parameter γ_h , evaluated at $(0, 0)$: the auxiliary variable gives the step $\mathcal{T}_{1,1-h_k}$ over $[h_k, 1]$; the level $l \geq 1$ gives the step $\mathcal{T}_{\theta_l, \Delta h_l}$ over $[h_{l-1}, h_l]$, where $\gamma_h = \theta_l$; and the root gives the heat step $\mathcal{T}_{0,h_0}$ over $[0, h_0]$, where $\gamma_h = 0$. Levels with $\Delta h_l = 0$ give empty intervals and the identity step $\mathcal{T}_{\theta_l, 0}$, so they can be omitted. Hence

$$u_N(0, p, 0) = \psi_{\sigma,N}(p) + \alpha_N u_{\gamma_h}(0, 0; U), \quad u_N(1, 0, 0) = F_N(U). \quad (3.25)$$

The identification of the row part with the recursion (2.2) is the representation of the Parisi functional through Ruelle probability cascades (see [98, Chapter 2] and [4]).

3.4 Concentration and identities at a contact

The comparison argument evaluates the interpolation only at maximum points of a deterministic function of the parameters. At such points we need the Ghirlanda–Guerra identities for the perturbed Gibbs measure to hold approximately. They follow from concentration of the derivatives of $\log W$ in the perturbation parameters x . Since the points are deterministic, concentration at each fixed parameter suffices.

This section follows Steps 1–3 of the proof of [92, Theorem 4.1]. There a quadratic penalty around the point $(1, \dots, 1)$ [92, (4.18)–(4.19)] keeps the perturbation coefficients at the contact point close to one; the bound on the Hessian in x at a contact point and the concentration of the free energy [92, Proposition 4.2] control the fluctuations of the perturbation Hamiltonians; and Gaussian integration by parts gives approximate Ghirlanda–Guerra identities [92, (4.28)]. Lemmas 3.6 and 3.7 carry out these steps for the present interpolation, and the penalty appears in the proof of Proposition 3.10.

Lemma 3.6. *Fix K , h , and $\mathcal{I}$, and a compact subset of $[0, 1] \times \mathcal{K} \times \mathbb{R}^{\mathcal{I}}$. There is a constant C such that*

$$\text{Var} \left(\frac{1}{N} \log W \right) \leq \frac{C}{N}$$

on this set. Moreover, almost surely $x \mapsto \log W$ is convex and C^2 on $\mathbb{R}^{\mathcal{I}}$, with

$$\partial_{x_j} \log W = s_N \langle H^j \rangle, \quad (3.26)$$

$$\partial_{x_j}^2 \log W = s_N^2 \text{Var}_{\langle \cdot \rangle} (H^j). \quad (3.27)$$

The function u_N is convex and differentiable in x , with $N \partial_{x_j} u_N = s_N \mathbb{E} \langle H^j \rangle$, $|\partial_{x_j} u_N| \leq 2^{j_+ + 1} N^{-1/8} |x_j|$, and, with $C_{j_+} = 2^{j_+}$,

$$|u_N(t, p, x) - u_N(t, p, 0)| \leq C_{j_+} N^{-1/8} |x|^2 \quad ((t, p, x) \in [0, 1] \times \mathcal{K} \times \mathbb{R}^{\mathcal{I}}). \quad (3.28)$$

Proof. We divide the proof into three steps.

STEP 1. The variance. By the law of total variance, $\text{Var}(\log W) = \mathbb{E} \text{Var}(\log W \mid e_0) + \text{Var}(\mathbb{E}[\log W \mid e_0])$. By (3.6) the first term is at most $4 \text{Var}(\log \Pi)$, a constant depending only on the cascade, and $\mathbb{E}[\log W \mid e_0] = X_0(e_0)$ is a function of the root Gaussian variables $\mathbf{G}$, $y_\emptyset$, $z_\emptyset$, and the root tensors. As in Steps 2 and 3 of the proof of Lemma 3.5, applied to these variables (the derivatives of $\mathcal{H}$ in them are bounded), X_0 is continuously differentiable in them and its gradient is the $\mathbb{E}_K$ -average of the gradient of the leaf logarithm, so by Jensen’s inequality its squared norm is at most the $\mathbb{E}_K$ -average of the squared norm of the leaf gradient. The derivative of the leaf logarithm in g_{ai} is $\sqrt{t/N} \langle U'(S_a) \sigma_i \rangle_\beta$; by Jensen’s inequality and $|\sigma|^2 = N$ the sum of its squares over a, i is at most $(t/N) MNL^2 = Nt\alpha_N L^2$. The derivative in $y_\emptyset$ is $\sqrt{p_0} \langle \sigma \rangle_\beta$, with squared norm at most Np_0 . The derivative in $z_{\emptyset, a}$ is $\sqrt{(1-t)h_0} \langle U'(S_a) \rangle_\beta$, with squared norm

summed over a at most $N\alpha_N h_0 L^2$. The root tensor of H^j has coefficient $c_{j,j_3,0}^2 \leq N$, and the derivative of the leaf logarithm in it has squared norm at most $s_N^2 x_j^2 c_{j,j_3,0}^2 N^{-j_3} |\sigma|^{2j_3} \leq s_N^2 N x_j^2$. Altogether

$$|\nabla X_0|^2 \leq N(t\alpha_N L^2 + p_0 + \alpha_N h_0 L^2 + N^{-1/8} |x|^2),$$

and the Gaussian Poincaré inequality [21, Theorem 3.20] gives $\text{Var}(X_0) \leq \mathbb{E}|\nabla X_0|^2 \leq CN$ on the compact set. Hence $\text{Var}(N^{-1} \log W) \leq N^{-2}(4 \text{Var} \log \Pi + CN) \leq C/N$.

STEP 2. *The derivatives in x .* Let $\bar{\Theta}_\beta = \sum_{j \in \mathcal{I}} \sum_{i,l} c_{j,i,l} \|g_{\beta|l}^{j,i}\|$, so that $\sum_j \sup_\sigma |H^j(\sigma, \beta)| \leq \bar{\Theta}_\beta$, and let $A_\beta(x)$ be the leaf integral. For each $x_0 \in \mathbb{Q}^{\mathcal{I}}$ and each integer $n \geq 1$,

$$\sum_\beta v_\beta A_\beta(x_0) (1 + \bar{\Theta}_\beta + \bar{\Theta}_\beta^2) e^{s_N n \bar{\Theta}_\beta} < \infty \quad \text{almost surely.}$$

Indeed, the sum equals $W(x_0)$ times a Gibbs average whose expectation, by (3.3), is an $\mathbb{E}_K$ -expectation of $(1 + \bar{\Theta} + \bar{\Theta}^2) e^{s_N n \bar{\Theta}}$, which is finite by the kernel density bound of Step 1 in the proof of Lemma 3.5. On the intersection of these countably many events, since $A_\beta(x) \leq A_\beta(x_0) e^{s_N |x - x_0|_1 \bar{\Theta}_\beta}$, the series defining $W(x)$ and the series of its first two derivatives converge locally uniformly in x . Termwise differentiation gives (3.26) and (3.27). More generally, the Hessian of $\log W$ in x is s_N^2 times the covariance matrix of $(H^j)_{j \in \mathcal{I}}$ under $\langle \cdot \rangle$, which is positive semidefinite, so $\log W$ is convex in x . For a convex function the difference quotients $(\log W(x) - \log W(x - \epsilon \mathbf{e}_j))/\epsilon$ and $(\log W(x + \epsilon \mathbf{e}_j) - \log W(x))/\epsilon$, where $\mathbf{e}_j$ is the j -th coordinate vector of $\mathbb{R}^{\mathcal{I}}$, bound $\partial_{x_j} \log W$ from below and above, are integrable, and converge monotonically as $\epsilon \downarrow 0$. Hence $\partial_{x_j} \mathbb{E} \log W = \mathbb{E} \partial_{x_j} \log W$, and u_N is convex in x . A convex function on $\mathbb{R}^{\mathcal{I}}$ whose partial derivatives exist is differentiable, so u_N is differentiable in x .

STEP 3. *Replica integration by parts.* For a bounded measurable function f of the overlap arrays $(R_{ll'}, J_{ll'})_{l,l' \leq n}$ of n replicas, Gaussian integration by parts in the tensors gives

$$\mathbb{E} \langle f H^j(1) \rangle = s_N x_j N \left(\sum_{l=1}^n \mathbb{E} \langle f c_j(R_{1l}, J_{1l}/k) \rangle - n \mathbb{E} \langle f c_j(R_{1,n+1}, J_{1,n+1}/k) \rangle \right), \quad (3.29)$$

where $H^j(1) = H^j(\sigma^1, \beta^1)$. To justify the exchange of the sum over nodes with the expectation, differentiate with respect to one tensor $g_v^{j,i}$ at one node v of depth m . Each replica l contributes the indicator that β^l descends from v , and contracting the two tensors gives $N^i R_{1l}^i$. Summing the indicators over the nodes of depth m gives $\mathbf{1}_{\{J_{1l} \geq m\}}$, and summing the squared coefficients over i and m gives the covariance (3.20). The expected absolute sum over nodes is bounded by a constant times $\mathbb{E} \langle \|g_{\beta|m}^{j,i}\| \rangle$, which is finite by (3.3) and (3.24). Taking $n = 1$ and $f = 1$ in (3.29), and using $|c_j| \leq 2^{j_3}$, gives $|\mathbb{E} \langle H^j \rangle| \leq 2^{j_3+1} s_N |x_j| N$, hence $|\partial_{x_j} u_N| = s_N N^{-1} |\mathbb{E} \langle H^j \rangle| \leq 2^{j_3+1} N^{-1/8} |x_j|$. Integrating along the segment from 0 to x gives (3.28), since $j_3 \leq j_+$. $\square$

For two families of overlap arrays $A^{ll'}, B^{ll'}$, a test function f of the arrays of the first n replicas, and a function c of two real variables, the Ghirlanda–Guerra discrepancy is

$$\mathfrak{D}_n(f, c) = \mathbb{E}\langle f c(A^{1,n+1}, B^{1,n+1}) \rangle - \frac{1}{n} \mathbb{E}\langle f \rangle \mathbb{E}\langle c(A^{12}, B^{12}) \rangle - \frac{1}{n} \sum_{l=2}^n \mathbb{E}\langle f c(A^{1l}, B^{1l}) \rangle.$$

The Ghirlanda–Guerra identities [64] state that $\mathfrak{D}_n(f, c) = 0$. For $c = c_j$ this is the form used in [92, (4.28) and (5.6)], and it is the analogue, for two overlap arrays, of the multi-species discrepancy in [100, (31)].

Lemma 3.7. *Fix a compact set of parameters $(t, p) \in [0, 1] \times \mathcal{K}$. There is a constant C with the following property. Suppose that (t, p) lies in this set, that $|x - \mathbf{1}| \leq 1/2$, where $\mathbf{1} \in \mathbb{R}^{\mathcal{I}}$ has all coordinates equal to one, and that for every $y \in \mathbb{R}^{\mathcal{I}}$*

$$0 \leq u_N(x + y) - u_N(x) - \nabla_x u_N(x) \cdot y \leq |y|^2, \quad (3.30)$$

where $u_N(x)$ abbreviates $u_N(t, p, x)$. Then for $A = R$ and $B = J/k$, every $j \in \mathcal{I}$, every $n \geq 1$, and every measurable function f of the overlap arrays of n replicas with $|f| \leq 1$,

$$|\mathfrak{D}_n(f, c_j)| \leq CN^{-1/8}.$$

Proof. We show that H^j has small fluctuations, both under the Gibbs measure and over the disorder, at the scale relevant to (3.29). We divide the proof into three steps.

STEP 1. Thermal fluctuation. Adding (3.30) at $y = \pm \epsilon \mathbf{e}_j$ gives $u_N(x + \epsilon \mathbf{e}_j) + u_N(x - \epsilon \mathbf{e}_j) - 2u_N(x) \leq 2\epsilon^2$, and the expectation of the corresponding second difference of $\log W$ is N times the left side. This second difference is nonnegative by convexity, and, divided by ϵ^2 , it converges to $s_N^2 \text{Var}_{\langle \cdot \rangle}(H^j)$ by (3.27). Fatou's lemma gives $\mathbb{E} \text{Var}_{\langle \cdot \rangle}(H^j) \leq 2Ns_N^{-2} = 2N^{9/8}$.

STEP 2. Disorder fluctuation. Put $\hat{u}_N = N^{-1} \log W - p_k/2$, so that $u_N = \mathbb{E} \hat{u}_N$, and fix $\epsilon \in (0, 1]$. Let $\Delta = \max |\hat{u}_N(z) - u_N(z)|$ over the three points $z \in \{x, x + \epsilon \mathbf{e}_j, x - \epsilon \mathbf{e}_j\}$. By convexity of $\hat{u}_N$ in x_j and the upper bound in (3.30),

$$\partial_j \hat{u}_N(x) \leq \frac{\hat{u}_N(x + \epsilon \mathbf{e}_j) - \hat{u}_N(x)}{\epsilon} \leq \frac{u_N(x + \epsilon \mathbf{e}_j) - u_N(x) + 2\Delta}{\epsilon} \leq \partial_j u_N(x) + \epsilon + \frac{2\Delta}{\epsilon},$$

and in the same way, with the secant through $x - \epsilon \mathbf{e}_j$, $\partial_j \hat{u}_N(x) \geq \partial_j u_N(x) - \epsilon - 2\Delta/\epsilon$. The three points are deterministic and lie in a fixed compact set, so Lemma 3.6 gives $\mathbb{E} \Delta^2 \leq C/N$. With $\epsilon = N^{-1/4}$,

$$\mathbb{E} (\partial_j \hat{u}_N(x) - \partial_j u_N(x))^2 \leq 2\epsilon^2 + \frac{8\mathbb{E} \Delta^2}{\epsilon^2} \leq CN^{-1/2}.$$

By (3.26), $\partial_j \hat{u}_N = N^{-1} s_N \langle H^j \rangle$ and $\partial_j u_N = N^{-1} s_N \mathbb{E} \langle H^j \rangle$, so $\mathbb{E} (\langle H^j \rangle - \mathbb{E} \langle H^j \rangle)^2 \leq N^2 s_N^{-2} CN^{-1/2} = CN^{13/8}$. Adding the thermal fluctuation, $\mathbb{E} \langle (H^j - \mathbb{E} \langle H^j \rangle)^2 \rangle \leq CN^{13/8}$.

STEP 3. *Conclusion.* Subtract from (3.29) for n replicas the same identity with $n = 1$ and $f = 1$, multiplied by $\mathbb{E}\langle f \rangle$. The terms with $l = 1$ cancel, because $c_j(R_{11}, J_{11}/k) = c_j(1, 1)$ is a constant, and what remains is

$$\mathbb{E}\langle f(H^j(1) - \mathbb{E}\langle H^j \rangle) \rangle = -s_N x_j N n \mathfrak{D}_n(f, c_j).$$

By the Cauchy–Schwarz inequality the left side is at most $CN^{13/16}$. Dividing by $ns_N x_j N \geq \frac{1}{2}N^{15/16}$, which uses $x_j \geq 1/2$, gives $|\mathfrak{D}_n(f, c_j)| \leq CN^{-1/8}$. $\square$

3.5 Synchronization

We use Panchenko’s synchronization of overlaps, in the quantitative form proved by Mourrat. Panchenko proved that, under the Ghirlanda–Guerra identities, the overlap of each species in a multi-species model is a nondecreasing Lipschitz function of the total overlap [100, Theorem 4]. The proof rests on his theorem that the Ghirlanda–Guerra identities imply ultrametricity [97]. He proved the analogue for vector spins in [101]. Mourrat’s synchronization theorem [92, Theorem 5.3] is a variant for two overlap arrays under approximate identities. The following form of it bounds the conditional variance of one overlap given the other, once approximate Ghirlanda–Guerra identities hold for a sufficiently rich finite family of test functions. Chapter 4 uses the same synchronization in the lower bound.

Lemma 3.8. *Fix an enumeration (ϑ_n) of $\mathbb{Q} \cap [0, 1]$. For every $\varepsilon > 0$ there is an integer $n_\varepsilon \geq 1$ with the following property. Let a random probability measure be supported on the product of the unit balls of two Hilbert spaces, and let $A^{l'}$ and $B^{l'}$ be the two inner-product arrays of independent samples from it. Suppose that*

$$|\mathfrak{D}_n(f, c_j)| \leq \frac{1}{n_\varepsilon} \quad \text{for all } n \leq n_\varepsilon \text{ and } j \in \{1, \dots, n_\varepsilon\}^3,$$

and every continuous test function f of the arrays of n samples with $|f| \leq 1$, where $c_j(u, v) = (\vartheta_{j_1} u + \vartheta_{j_2} v)^{j_3}$. If the averaged law of B^{12} is $K^{-1} \sum_{l=1}^K \delta_{q_l}$, where $-1 = q_0 < q_1 < \dots < q_K \leq 1$, then

$$\mathbb{E}\langle (A^{12} - \mathbb{E}\langle A^{12} \mid B^{12} \rangle)^2 \rangle \leq \frac{12}{K} + \varepsilon K^2 \max_{0 \leq l < K} (q_{l+1} - q_l)^{-1}. \quad (3.31)$$

Proof. This is [92, Proposition 5.5 and (5.13)], stated there with a number $\delta > 0$ in place of $1/n_\varepsilon$ and with the conditions required for $n, j_1, j_2, j_3 \leq \lfloor \delta^{-1} \rfloor$; its hypothesis on the test functions and on the law of B^{12} is the one above, and $\mathbb{E}\langle A^{12} \mid B^{12} \rangle$ is the conditional expectation under the averaged measure. Our form follows by taking $n_\varepsilon = \lceil \delta^{-1} \rceil$, because then $1/n_\varepsilon \leq \delta$ and $\lfloor \delta^{-1} \rfloor \leq n_\varepsilon$. $\square$

We apply Lemma 3.8 with the following choices. Recall that $K = k + 1$ and $w_l = 1/K$ for $0 \leq l \leq k$. The maps $\sigma \mapsto \sigma/\sqrt{N}$ and $\beta \mapsto k^{-1/2} \sum_{l=1}^k \mathbf{u}_{\beta|l}$, where $(\mathbf{u}_v)$ is an orthonormal family indexed by the nodes of depth at least one, represent R and J/k as inner products of unit vectors, since $k^{-1} \#\{1 \leq l \leq k : \beta|l = \beta'|l\} = (\beta \wedge \beta')/k$. Up to the factor $k^{-1/2}$, this is the embedding of [92, (A.7)]. The image of the Gibbs measure under the product of these maps is a random probability measure on the product of two unit balls. By Lemma 3.2, J/k has averaged law exactly uniform on $\{0, 1/k, \dots, 1\}$, for every value of the parameters, including the perturbation. In the notation of Lemma 3.8 this is $q_l = (l-1)/k$ for $1 \leq l \leq K$, so the gaps $q_{l+1} - q_l$ are 1 (for $l = 0$) and $1/k$, and the right side of (3.31) is at most $12/K + \varepsilon K^3$.

We choose the parameters in the following order. Given K , let $\varepsilon = K^{-4}$, let $n_K = n_\varepsilon$ be given by Lemma 3.8, and take $j_+ = n_K$, so that the functions c_j , $j \in \mathcal{I}$, are exactly those of Lemma 3.8. At any point satisfying the hypotheses of Lemma 3.7, and for N so large that $CN^{-1/8} \leq 1/n_K$, the hypotheses of Lemma 3.8 hold. The left side of (3.31) equals $\sum_l w_l \text{Var}(R_{12} \mid J = l)$, since $\mathbb{E}\langle R_{12} \mid J \rangle = r_J$, and $12/K + K^{-4}K^3 = 13/K$. Hence, with

$$\mathcal{E}_N = \frac{1}{2} \sum_{l=0}^k w_l \text{Cov}(R_{12}, \alpha_N X_{12} \mid J = l) \quad \text{and} \quad \eta_N = \frac{\alpha_N L^2}{2} \sqrt{\frac{13}{K}}, \quad (3.32)$$

we have $\sum_l w_l \text{Var}(R_{12} \mid J = l) \leq 13/K$ and $|\mathcal{E}_N| \leq \eta_N$ at such a point. The second bound follows from the first: since $|X_{12}| \leq L^2$, the conditional Cauchy–Schwarz inequality gives $|\text{Cov}(R_{12}, \alpha_N X_{12} \mid J = l)| \leq \alpha_N L^2 \text{Var}(R_{12} \mid J = l)^{1/2}$, and the Cauchy–Schwarz inequality over l , with $\sum_l w_l = 1$, gives $\sum_l w_l \text{Var}(R_{12} \mid J = l)^{1/2} \leq (13/K)^{1/2}$. By (3.21),

$$\partial_t u_N = -\frac{1}{2} \sum_{l=0}^k w_l (r_l - h_l) \zeta_l - \mathcal{E}_N. \quad (3.33)$$

The tolerance $\varepsilon = K^{-4}$ and the bound $13/K$ are those of Mourrat [92, (4.31)]. There the conditional variances of both overlaps enter. Here only the spin overlap is synchronized with J , and the response overlap enters through the bound $|X_{12}| \leq L^2$, which gives the square root in η_N .

3.6 The spherical term

At $t = 0$ the spins interact only with the cascade field Y , through the term $\psi_{\sigma,N}(p)$ of (3.25). This is the only geometric term in the comparison. It enters through the maximum over p of $\psi_{\sigma,N}(p) + \frac{1}{2} \sum_l w_l h_l p_l$, and we show that this maximum is at most $\text{CS}(\gamma_h) + o_N(1)$. The idea is to replace the uniform measure on the sphere by a Gaussian measure with a well-chosen variance $1/b$, for which the cascade recursion is computed exactly by Lemma 3.4. This replacement, at the cost $(b-1-\log b)/2$, is the large deviation argument behind

the Parisi formula for spherical models [125]; see also [37, 93]. Here only an upper bound is needed, and it holds for each N with an error of order $\log N/N$.

Lemma 3.9. *The following hold.*

- (i) For every N and every $p \in \mathcal{K}$, $\psi_{\sigma,N}(p) \leq \sqrt{p_k} - \frac{1}{2} \sum_l w_l p_l$.
- (ii) For $h \in \mathcal{K}$ with $h_k < 1$,

$$\bar{\psi}_N := \sup_{p \in \mathcal{K}} \left\{ \psi_{\sigma,N}(p) + \frac{1}{2} \sum_{l=0}^k w_l h_l p_l \right\} \leq \text{CS}(\gamma_h) + o_N(1), \quad (3.34)$$

where $o_N(1)$ depends only on K and h .

Proof. Part (ii) rests on part (i).

Proof of (i). The leaf logarithm $\log \int e^{Y_\beta \cdot \sigma} d\mu_N$ is $\sqrt{N\Delta p_l}$ -Lipschitz in the Gaussian vector $y_{\beta|l}$, since its gradient is $\sqrt{\Delta p_l} \langle \sigma \rangle_\beta$. Differentiating (3.2) shows that each X_{l-1} keeps these Lipschitz bounds in the earlier vectors, since its gradient is a K_l -average. Gaussian concentration in the form $\mathbb{E} e^{\theta(F - \mathbb{E}F)} \leq e^{\theta^2 \ell^2/2}$ for an ℓ -Lipschitz function F of a standard Gaussian vector [21, Theorem 5.5], applied at each level $l = k, \dots, 1$, shows that X_{l-1} exceeds the plain average of X_l over $y_{\beta|l}$ by at most $N\theta_l \Delta p_l/2$. The root is averaged without a logarithm. Therefore, by (3.5),

$$\frac{1}{N} \mathbb{E} \log \sum_\beta v_\beta \int e^{Y_\beta \cdot \sigma} d\mu_N \leq \frac{1}{N} \mathbb{E} \log \int e^{Y \cdot \sigma} d\mu_N + \frac{1}{2} \sum_{l=1}^k \theta_l \Delta p_l \leq \sqrt{p_k} + \frac{1}{2} \sum_{l=1}^k \theta_l \Delta p_l,$$

where $Y \sim N(0, p_k I_N)$, and we used $Y \cdot \sigma \leq \sqrt{N}|Y|$ and $\mathbb{E}|Y| \leq \sqrt{N p_k}$. By (3.7) this proves (i).

Proof of (ii). We divide the proof into three steps.

STEP 1. Localization. By (i), since $h_l \leq h_k$ and $p_l \geq 0$,

$$\psi_{\sigma,N}(p) + \frac{1}{2} \sum_l w_l h_l p_l \leq \sqrt{p_k} - \frac{1}{2} \sum_l w_l (1 - h_l) p_l \leq \sqrt{p_k} - \frac{1}{2} w_k (1 - h_k) p_k.$$

The value at $p = 0$ is zero, and the right side is negative when $p_k > \bar{p} := 4/[w_k(1 - h_k)]^2$. So the supremum may be restricted to the set $\{p \in \mathcal{K} : p_k \leq \bar{p}\}$, and we prove the bound uniformly over this set. Fix such a p .

STEP 2. Choice of the variance. In the notation of (3.8) with $n = N$, as b increases every d_l increases, so $m_k + 1/b$ decreases continuously. It tends to zero as $b \rightarrow \infty$, and to infinity as $b \downarrow \sum_{j \geq 1} \theta_j \Delta p_j$: if $p_0 > 0$, the term p_0/d_0^2 diverges; if $p_0 = 0$ and $p \neq 0$, then for the first index j_* with $\Delta p_{j_*} > 0$ we have $d_{j_*-1} = b - \sum_{j \geq 1} \theta_j \Delta p_j$, and the j_* -th term of m_k diverges; if $p = 0$, the term $1/b$ diverges. We choose b as the unique solution of $m_k + 1/b = 1$,

so that Lemma 3.4(iv) with $\varrho = 1$ applies. In particular $b \geq 1$. Also $b < 2(\bar{p} + 1)$: since $d_l \geq d_0 \geq b - p_k \geq b - \bar{p}$, we have $m_k \leq p_k/(b - \bar{p})^2 \leq \bar{p}/(b - \bar{p})^2$ for $b > \bar{p}$, and at $b = 2(\bar{p} + 1)$ this plus $1/b$ is at most $\frac{1}{8} + \frac{1}{2} < 1$, because $(\bar{p} + 2)^2 \geq 8\bar{p}$.

STEP 3. *Comparison with the sphere.* For fixed Y , the average of $e^{\rho Y \cdot \sigma}$ over $\sigma \in S_N$ is nondecreasing in $\rho \geq 0$: by the symmetry $\sigma \mapsto -\sigma$ it equals the average of $\cosh(\rho Y \cdot \sigma)$, whose derivative in ρ is nonnegative. If $\varepsilon \sim N(0, b^{-1}I_N)$, then $\varepsilon = |\varepsilon|N^{-1/2}\sigma$ with σ uniform on S_N and independent of $|\varepsilon|$, and $b|\varepsilon|^2$ has the chi-square law χ_N^2 with N degrees of freedom. Restricting $\mathbb{E}e^{Y \cdot \varepsilon} = e^{|Y|^2/(2b)}$ to $\{|\varepsilon|^2 \geq N\}$ and using the monotonicity gives, pathwise,

$$\log \sum_{\beta} v_{\beta} \int e^{Y_{\beta} \cdot \sigma} d\mu_N \leq \log \sum_{\beta} v_{\beta} e^{|Y_{\beta}|^2/(2b)} - \log \mathbb{P}\{\chi_N^2 \geq bN\}.$$

For $N \geq 2$ and $b \geq 1$,

$$-\frac{1}{N} \log \mathbb{P}\{\chi_N^2 \geq bN\} \leq \frac{b-1-\log b}{2} + \frac{\frac{1}{2} \log(\pi N) + \log b + 1 + \frac{1}{6N}}{N}. \quad (3.35)$$

Indeed, the chi-square density $x^{N/2-1}e^{-x/2}/(2^{N/2}\Gamma(N/2))$ is nonincreasing on $[N-2, \infty)$, so $\mathbb{P}\{\chi_N^2 \geq bN\}$ is at least twice its value at $bN+2$. Then (3.35) follows from $(bN+2)^{N/2-1} \geq (bN)^{N/2-1}$ and Stirling's bound $\Gamma(z) \leq \sqrt{2\pi}z^{z-1/2}e^{-z+1/(12z)}$ with $z = N/2$, after collecting terms. Take expectations, divide by N , and use (3.9) (with $n = N$), (3.16) and (3.17) in the form $\frac{b-1}{2} - \frac{p_k}{2} = -\frac{1}{2} \sum_l w_l m_l p_l$. Since $1 \leq b < 2(\bar{p} + 1)$, this gives

$$\psi_{\sigma, N}(p) + \frac{1}{2} \sum_l w_l h_l p_l \leq \text{CS}(\gamma_m) + \frac{1}{2} \sum_l w_l p_l (h_l - m_l) + C_{\bar{p}} \frac{\log N}{N}.$$

It remains to show that $\text{CS}(\gamma_h) \geq \text{CS}(\gamma_m) + \frac{1}{2} \sum_l w_l p_l (h_l - m_l)$. The map $x \mapsto \text{CS}(\gamma_x)$ is convex on the convex set $\{x \in \mathcal{K} : x_k < 1\}$. Indeed, compute the entropies at two points x, x' with a common cutoff $q' \in [\max(x_k, x'_k), 1)$: for each $s < q'$, $\lambda_{\gamma_x}(s) = 1 - \sum_l w_l \max(x_l, s)$ is positive and concave in x , so its reciprocal is convex. Hence $f(\epsilon) = \text{CS}(\gamma_{m+\epsilon(h-m)})$ is convex on $[0, 1]$, and $\text{CS}(\gamma_h) \geq \text{CS}(\gamma_m) + f'(0+)$; this one-sided derivative is all we need, also when m lies on the boundary of $\mathcal{K}$. With the common cutoff $q' = \max(h_k, m_k)$, all the functions λ involved are at least $1 - q' > 0$ on $[0, q']$, and $|\max(m_l + \epsilon(h_l - m_l), s) - \max(m_l, s)| \leq \epsilon|h_l - m_l|$. For almost every s , the right derivative of $\max(m_l + \epsilon(h_l - m_l), s)$ at $\epsilon = 0$ is $(h_l - m_l)\mathbf{1}_{\{m_l > s\}}$. Dominated convergence and (3.16) give

$$f'(0+) = \sum_l \frac{w_l}{2} (h_l - m_l) \int_0^{m_l} \frac{ds}{\lambda_{\gamma_m}(s)^2} = \frac{1}{2} \sum_l w_l p_l (h_l - m_l).$$

This bounds the left side of the previous display by $\text{CS}(\gamma_h) + C_{\bar{p}} \log N/N$, uniformly over $p \in \mathcal{K}$ with $p_k \leq \bar{p}$, and proves (ii). $\square$

3.7 The comparison

We now show that, up to the synchronization error, the free energy at $t = 1$ is at most the maximum over p of the interpolation at $t = 0$. The reason is the following. At a point where p maximizes $u_N(t, \cdot, x) + \frac{1}{2} \sum_l w_l h_l p_l$, the product term in (3.33) has a sign, so $\partial_t u_N$ is at most the synchronization error. A maximum principle in t turns this into a bound on the growth of the maximum over p . Maximizing also over x , with a quadratic penalty, selects a perturbation at which Lemma 3.7 applies.

This is a finite- N form of Step 5 of the proof of [92, Theorem 4.1], where the first-order conditions [92, (4.39)–(4.40)] at a contact point and the monotonicity of [92, Lemma 2.4] give the sign [92, (4.32)] needed for the supersolution property. Here the maximum principle in t for the maximum over p takes the place of viscosity solutions and of the comparison principle [92, Proposition 3.2]. The use of first-order conditions in the levels to control the derivative in t also parallels the stationarization of Stojnic's fully lifted interpolation for bilinearly indexed random processes [120, 121]. There the levels on both sides are chosen at stationary points along the interpolation path, and the derivative in t vanishes when the overlaps concentrate.

Proposition 3.10. *Let U belong to class (A), let $K \geq 2$ and $h \in \mathcal{K}$ with $h_k < 1$, let n_K and $j_+ = n_K$ be as in §3.5, and let η_N and $\bar{\psi}_N$ be given by (3.32) and (3.34). There is N_0 , depending only on U, K, h , and α' , such that for $N \geq N_0$*

$$F_N(U) \leq \bar{\psi}_N + \alpha_N u_{\gamma_h}(0, 0; U) + \eta_N + 3C_{j_+} |\mathcal{I}| N^{-1/8}. \quad (3.36)$$

Consequently,

$$\limsup_{N \rightarrow \infty} F_N(U) \leq \mathcal{P}_U(\gamma_h) + \frac{\alpha L^2}{2} \sqrt{\frac{13}{K}}. \quad (3.37)$$

Proof. Fix $t_0 \in (0, 1)$ and $\varsigma \in (0, 1]$, and consider on $[0, t_0] \times \mathcal{K} \times \mathbb{R}^{\mathcal{I}}$ the function

$$\Psi(t, p, x) = u_N(t, p, x) + \frac{1}{2} \sum_l w_l h_l p_l - |x - \mathbf{1}|^2 - (\eta_N + \varsigma)t.$$

We divide the proof into four steps.

STEP 1. Localization. We show that Ψ attains its maximum and that all maximum points lie in a compact set that does not depend on N, t_0 , or ς . We take $N \geq N_0$, with N_0 fixed at the end of this step; in particular

$$(a) \quad 2C_{j_+} N^{-1/8} \leq \min\left\{\frac{1}{2}, |\mathcal{I}|^{-1}\right\},$$

where C_{j_+} is the constant of (3.28). Then $C_{j_+} N^{-1/8} |x|^2 \leq 2C_{j_+} N^{-1/8} (|x - \mathbf{1}|^2 + |\mathcal{I}|) \leq \frac{1}{2} |x - \mathbf{1}|^2 + 1$. Since $U \leq U_{\max}$, the leaf integral is at most $e^{MU_{\max}} \int e^{Y_{\beta} \cdot \sigma} d\mu_N$ when $x = 0$, so $u_N(t, p, 0) \leq \alpha_N U_{\max} + \psi_{\sigma, N}(p)$. With (3.28), Lemma 3.9(i) and $h_l \leq h_k$,

$$\Psi(t, p, x) \leq \alpha' U_{\max}^+ + 1 + \sqrt{p_k} - \frac{1}{2} w_k (1 - h_k) p_k - \frac{1}{2} |x - \mathbf{1}|^2.$$

On the other hand, by (3.28), (3.25), $\psi_{\sigma,N}(0) = 0$, and the bound $u_{\gamma_h}(0,0;U) \geq \mathbb{E}U(G) \geq U(0) - L$ from (2.6),

$$\Psi(0,0,1) \geq u_N(0,0,0) - C_{j_+} N^{-1/8} |\mathcal{I}| \geq -\alpha'(|U(0)| + L) - 1.$$

Since $w_k(1 - h_k) > 0$ and $0 \leq p_l \leq p_k$ on $\mathcal{K}$, we have $\Psi < \Psi(0,0,1)$ outside $[0, t_0] \times \mathcal{K}_0$, for a compact set $\mathcal{K}_0 \subset \mathcal{K} \times \mathbb{R}^{\mathcal{I}}$ that depends only on $U, K, h, \mathcal{I}$, and α' . As Ψ is continuous by Lemma 3.5(i), it attains its maximum, and every maximum point lies in $[0, t_0] \times \mathcal{K}_0$. Let N_0 be the least integer such that, for all $N \geq N_0$, condition (a) holds and also

- (b) $2^{j_+} N^{-1/8} |x| \leq \frac{1}{2}$ for all $(p, x) \in \mathcal{K}_0$, and
- (c) $C N^{-1/8} \leq 1/n_K$, where C is the constant of Lemma 3.7 for the compact set $[0, 1] \times \{p : (p, x) \in \mathcal{K}_0 \text{ for some } x\}$.

Since $\mathcal{K}_0, \mathcal{I}$, and n_K are determined by U, K, h , and α' , so is N_0 ; it does not depend on t_0 or ς .

STEP 2. Conditions at a maximum point. Let (t_*, p_*, x_*) be a maximum point. First, x_* maximizes $x \mapsto u_N(t_*, p_*, x) - |x - \mathbf{1}|^2$, so $\nabla_x u_N = 2(x_* - \mathbf{1})$ there. The derivative bound of Lemma 3.6 gives $|x_* - \mathbf{1}| \leq 2^{j_+} N^{-1/8} |x_*|$, so $|x_* - \mathbf{1}| \leq \frac{1}{2}$ by (b). Maximality in x gives $u_N(x_* + y) - u_N(x_*) \leq |x_* + y - \mathbf{1}|^2 - |x_* - \mathbf{1}|^2 = \nabla_x u_N(x_*) \cdot y + |y|^2$ for all $y \in \mathbb{R}^{\mathcal{I}}$, and convexity of u_N in x gives the lower bound in (3.30). Thus Lemma 3.7 applies, and by (c) and §3.5, $|\mathcal{E}_N| \leq \eta_N$ at (t_*, p_*, x_*) .

Second, p_* maximizes $p \mapsto u_N(t_*, p, x_*) + \frac{1}{2} \sum_l w_l h_l p_l$ over $\mathcal{K}$. Since $p_* + \epsilon \mathbf{1}_{\geq m} \in \mathcal{K}$ for $\epsilon \geq 0$, the right derivative in this direction is nonpositive, and by (3.22)

$$A_m := \frac{1}{2} \sum_{l=m}^k w_l (h_l - r_l) \leq 0 \quad (0 \leq m \leq k). \quad (3.38)$$

Third, t_* maximizes $t \mapsto \Psi(t, p_*, x_*)$ over $[0, t_0]$.

STEP 3. Every maximum point has $t_* = 0$. Suppose that $t_* > 0$. Since $t_* \leq t_0 < 1$, u_N is differentiable in t at (t_*, p_*, x_*) by Lemma 3.5(ii), and maximality in t gives $\partial_t u_N(t_*, p_*, x_*) \geq \eta_N + \varsigma$. We now show the opposite inequality. Put $\zeta_{-1} = 0$, so that $\zeta_l = \sum_{m \leq l} (\zeta_m - \zeta_{m-1})$. Summation by parts, (3.23) and (3.38) give

$$\sum_l w_l (r_l - h_l) \zeta_l = \sum_{m=0}^k (\zeta_m - \zeta_{m-1}) \sum_{l=m}^k w_l (r_l - h_l) = -2 \sum_{m=0}^k (\zeta_m - \zeta_{m-1}) A_m \geq 0.$$

By (3.33) and Step 2, $\partial_t u_N = -\frac{1}{2} \sum_l w_l (r_l - h_l) \zeta_l - \mathcal{E}_N \leq |\mathcal{E}_N| \leq \eta_N$, a contradiction. Hence $t_* = 0$.

STEP 4. *Conclusion.* Compare Ψ at $(t_0, 0, \mathbf{1})$ with its value at a maximum point, which has $t_* = 0$. By (3.28) and (3.25),

$$\begin{aligned} u_N(t_0, 0, \mathbf{1}) - (\eta_N + \varsigma)t_0 &\leq \Psi(0, p_*, x_*) \\ &\leq \psi_{\sigma, N}(p_*) + \frac{1}{2} \sum_l w_l h_l p_{*, l} + \alpha_N u_{\gamma_h}(0, 0; U) \\ &\quad + C_{j_+} N^{-1/8} |x_*|^2 - |x_* - \mathbf{1}|^2. \end{aligned}$$

By (a), the last two terms are at most $2C_{j_+} N^{-1/8} |\mathcal{I}|$, and the first two are at most $\bar{\psi}_N$. Letting $t_0 \uparrow 1$, by continuity of u_N , and then $\varsigma \downarrow 0$,

$$u_N(1, 0, \mathbf{1}) \leq \bar{\psi}_N + \alpha_N u_{\gamma_h}(0, 0; U) + \eta_N + 2C_{j_+} N^{-1/8} |\mathcal{I}|.$$

Finally, $F_N(U) = u_N(1, 0, 0) \leq u_N(1, 0, \mathbf{1}) + C_{j_+} N^{-1/8} |\mathcal{I}|$ by (3.25) and (3.28). This proves (3.36). By Lemma 3.9(ii), $\bar{\psi}_N \leq \text{CS}(\gamma_h) + o_N(1)$, and $\alpha_N \rightarrow \alpha$, so (3.36) gives (3.37). $\square$

Proof of Theorem 3.1. Let $\gamma \in \mathcal{U}$ and $K \geq 2$. Approximate γ by the order parameter $\gamma_K = \lfloor K\gamma \rfloor / K$, which has K atoms of mass $1/K$ at the quantiles $h_l = \inf\{q : \gamma(q) \geq (l+1)/K\}$, $0 \leq l \leq k = K-1$. Indeed, by right-continuity $h_l \leq q$ if and only if $\gamma(q) \geq (l+1)/K$, so $\gamma_h(q) = K^{-1} \#\{l : h_l \leq q\} = \lfloor K\gamma(q) \rfloor / K$. Thus $h \in \mathcal{K}$, $h_k = Q_\gamma < 1$, $\gamma - 1/K < \gamma_K \leq \gamma$, so that $\|\gamma_K - \gamma\|_{L^1} \leq 1/K$, and $\gamma_K = \gamma = 1$ on $[Q_\gamma, 1]$. By (3.37) for this h , and by (2.18) with $\bar{q} = Q_\gamma$, since $\gamma, \gamma_K \in \mathcal{U}_{Q_\gamma}$,

$$\limsup_{N \rightarrow \infty} F_N(U) \leq \mathcal{P}_U(\gamma) + \frac{1}{K} \left(\frac{\alpha L^2}{2} + \frac{Q_\gamma}{2(1-Q_\gamma)^2} \right) + \frac{\alpha L^2}{2} \sqrt{\frac{13}{K}}.$$

Letting $K \rightarrow \infty$ proves (3.1).

If $L = 0$, the activation is constant and $f_N(U) = F_N(U)$, so the probability bound follows from (3.1). Let $L > 0$. The map $\mathbf{G} \mapsto f_N(U)$ is Lipschitz: its derivative in g_{ai} is $N^{-3/2} \langle U'(b_a \cdot \sigma) \sigma_i \rangle$, where now $\langle \cdot \rangle$ is the Gibbs measure of $Z_N(U)$, and by Jensen's inequality and $|\sigma|^2 = N$ the sum of the squares is at most $\alpha_N L^2 / N$. Gaussian concentration [21, Theorem 5.6] gives $\mathbb{P}\{f_N(U) - F_N(U) > \varepsilon/2\} \leq \exp(-\varepsilon^2 N / (8\alpha_N L^2))$, and $F_N(U) \leq \mathcal{P}_U(\gamma) + \varepsilon/2$ for large N by (3.1).

The parameters and limits are chosen, with U and γ fixed, in the following order: first K (and with it h), then the synchronization tolerance $\varepsilon = K^{-4}$, the integer n_K , and the perturbation family $\mathcal{I}$; then, for each $N \geq N_0$, the auxiliary $t_0 \uparrow 1$ and $\varsigma \downarrow 0$ inside the proof of Proposition 3.10; then $N \rightarrow \infty$; and finally $K \rightarrow \infty$. $\square$

3.8 Hard constraints

We now pass to the hard wall. Recall that $V_N(\kappa) = Z_N(H_\kappa)$ with $H_\kappa = \log \mathbf{1}_{[\kappa, \infty)}$ and $\log 0 = -\infty$. Chapter 9 also needs the feasible volume weighted by a bounded function of the fields, so we allow a source term from the start.

Terminal data with a hard wall. Let γ be a step order parameter as in (2.3), and let $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ be bounded and measurable. The terminal datum $H_\kappa + \varphi$ is bounded above by $\|\varphi\|_\infty$, so the recursion (2.2) with $T = 1$ defines $u_\gamma(\cdot, \cdot; H_\kappa + \varphi)$, independently of the representation of γ , and we put

$$\mathcal{P}_{H_\kappa + \varphi}(\gamma) = \alpha u_\gamma(0, 0; H_\kappa + \varphi) + \text{CS}(\gamma).$$

The function $u_\gamma(q, x; H_\kappa)$ is finite for $q < 1$: for $q \leq t_n$, the last breakpoint of γ , it equals $u_\gamma(q, x; \kappa)$ by (2.10), and for $t_n \leq q < 1$ it equals $\log \mathbb{P}\{x + \sqrt{1-q} G \geq \kappa\}$ by (2.7). By (2.4), applied in both directions, for bounded measurable φ, φ' the functions $u_\gamma(\cdot, \cdot; H_\kappa + \varphi)$ are finite on $[0, 1) \times \mathbb{R}$, and

$$|u_\gamma(q, x; H_\kappa + \varphi) - u_\gamma(q, x; H_\kappa + \varphi')| \leq \sup_{y \in \mathbb{R}} |\varphi(y) - \varphi'(y)|, \quad (q, x) \in [0, 1) \times \mathbb{R}. \quad (3.39)$$

By (2.10), $\mathcal{P}_{H_\kappa}(\gamma) = \mathcal{P}_\kappa(\gamma)$.

The wall approximation. We approximate the hard wall from above by activations in class (A) that decrease to it. Let

$$\Omega(y) = y^+ - \arctan(y^+), \quad U_\ell(y; \kappa) = -\ell \Omega(\kappa - y) \quad (\ell \geq 1). \quad (3.40)$$

The function Ω vanishes on $(-\infty, 0]$, is positive on $(0, \infty)$, and is C^2 with $\Omega'(y) = (y^+)^2 / (1 + (y^+)^2) \in [0, 1]$ and $\Omega''(y) = 2y^+ / (1 + (y^+)^2)^2 \in [0, 1]$. Hence $U_\ell(\cdot; \kappa)$ belongs to class (A), with $\sup U_\ell = 0$ and $|U'_\ell|, |U''_\ell| \leq \ell$. Moreover, $e^{U_\ell} = 1$ on $[\kappa, \infty)$, and $e^{U_\ell(y)}$ decreases to zero as $\ell \rightarrow \infty$ for $y < \kappa$. Thus $e^{U_\ell} \downarrow \mathbf{1}_{[\kappa, \infty)}$ pointwise.

Corollary 3.11. *Let $\kappa \in \mathbb{R}$ and $\varepsilon > 0$.*

(i) *For every $\gamma \in \mathcal{U}$,*

$$\mathbb{P}\left\{\frac{1}{N} \log V_N(\kappa) > \mathcal{P}_\kappa(\gamma) + \varepsilon\right\} \longrightarrow 0.$$

(ii) *Let γ be a step order parameter and let $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ be bounded and Lipschitz. Then*

$$\mathbb{P}\left\{\frac{1}{N} \log \int_{S_N(\kappa)} \exp\left\{\sum_{a=1}^M \varphi(b_a \cdot \sigma)\right\} \mu_N(d\sigma) > \mathcal{P}_{H_\kappa + \varphi}(\gamma) + \varepsilon\right\} \longrightarrow 0.$$

Part (ii) with $\varphi = 0$ is part (i) for step order parameters. Chapter 9 uses part (ii) with φ a real multiple of a bounded Lipschitz test function translated by κ , and Chapter 8 uses part (i).

Proof. We prove (ii) first, since (i) is deduced from it.

Proof of (ii). Let φ be L_φ -Lipschitz. The activation must be C^2 , so we first smooth φ . Let χ be a smooth probability density supported in $[-1, 1]$, and for $\tau \in (0, 1]$ let $\varphi_\tau = \varphi * \chi_\tau$

with $\chi_\tau(y) = \tau^{-1}\chi(y/\tau)$. Then φ_τ is smooth, $\|\varphi_\tau\|_\infty \leq \|\varphi\|_\infty$, and $|\varphi_\tau - \varphi| \leq L_\varphi\tau$. Moreover $\varphi'_\tau = \varphi' * \chi_\tau$ and $\varphi''_\tau = \varphi' * \chi'_\tau$, where φ' is the almost everywhere derivative, so $|\varphi'_\tau| \leq L_\varphi$ and $|\varphi''_\tau| \leq L_\varphi\|\chi'\|_{L^1}/\tau$. Hence $\tilde{U}_\ell = U_\ell(\cdot; \kappa) + \varphi_\tau$, with U_ℓ from (3.40), belongs to class (A) for every ℓ .

Since $U_\ell(\cdot; \kappa) = 0$ on $[\kappa, \infty)$ and the integrand of $Z_N(\tilde{U}_\ell)$ is positive,

$$\int_{S_N(\kappa)} e^{\sum_a \varphi(b_a \cdot \sigma)} \mu_N(d\sigma) \leq e^{ML_\varphi\tau} \int_{S_N(\kappa)} e^{\sum_a \varphi_\tau(b_a \cdot \sigma)} \mu_N(d\sigma) \leq e^{ML_\varphi\tau} Z_N(\tilde{U}_\ell).$$

So the normalized logarithm in (ii) is at most $f_N(\tilde{U}_\ell) + \alpha' L_\varphi\tau$.

We show that $u_\gamma(0, 0; \tilde{U}_\ell)$ decreases to $u_\gamma(0, 0; H_\kappa + \varphi_\tau)$ as $\ell \rightarrow \infty$. The terminal data satisfy $e^{\tilde{U}_\ell} \downarrow \mathbf{1}_{[\kappa, \infty)} e^{\varphi_\tau} = e^{H_\kappa + \varphi_\tau}$ pointwise, with values in $(0, e^{\|\varphi\|_\infty}]$. The recursion (2.2) applies the same operators $\mathcal{T}_{m,s}$ to both terminal data. Each is nondecreasing, so by induction over the steps the functions obtained from $\tilde{U}_\ell$ decrease in ℓ ; they are bounded above by $\|\varphi\|_\infty$, and their limits are the functions obtained from $H_\kappa + \varphi_\tau$, which are finite by (3.39). Indeed, at a step with $m > 0$ the functions $e^{mf_\ell} \in (0, e^{m\|\varphi\|_\infty}]$ decrease to e^{mf} , and at a step with $m = 0$ the functions $\|\varphi\|_\infty - f_\ell \geq 0$ increase to $\|\varphi\|_\infty - f$; bounded and monotone convergence apply. Hence $\mathcal{P}_{\tilde{U}_\ell}(\gamma) \downarrow \mathcal{P}_{H_\kappa + \varphi_\tau}(\gamma)$, and $\mathcal{P}_{H_\kappa + \varphi_\tau}(\gamma) \leq \mathcal{P}_{H_\kappa + \varphi}(\gamma) + \alpha L_\varphi\tau$ by (3.39).

Now choose τ with $(\alpha' + \alpha)L_\varphi\tau \leq \varepsilon/3$, and then ℓ with $\mathcal{P}_{\tilde{U}_\ell}(\gamma) \leq \mathcal{P}_{H_\kappa + \varphi_\tau}(\gamma) + \varepsilon/3$. Then $\mathcal{P}_{\tilde{U}_\ell}(\gamma) + \alpha' L_\varphi\tau \leq \mathcal{P}_{H_\kappa + \varphi}(\gamma) + 2\varepsilon/3$, and Theorem 3.1, applied to $\tilde{U}_\ell$ and γ , gives

$$\mathbb{P}\left\{\frac{1}{N} \log \int_{S_N(\kappa)} e^{\sum_a \varphi(b_a \cdot \sigma)} \mu_N(d\sigma) > \mathcal{P}_{H_\kappa + \varphi}(\gamma) + \varepsilon\right\} \leq \mathbb{P}\left\{f_N(\tilde{U}_\ell) > \mathcal{P}_{\tilde{U}_\ell}(\gamma) + \frac{\varepsilon}{3}\right\} \rightarrow 0.$$

Proof of (i). Let $\gamma \in \mathcal{U}$ and $r = Q_\gamma < 1$. The step order parameters $\gamma_n \in \mathcal{U}_r$ of §2.1 converge to γ in $L^1([0, 1])$, so $u_{\gamma_n}(0, 0; \kappa) \rightarrow u_\gamma(0, 0; \kappa)$ by the definition of u_γ through Lemma 2.7, and $\text{CS}(\gamma_n) \rightarrow \text{CS}(\gamma)$ by (2.12). Hence there is a step order parameter $\gamma' \in \mathcal{U}_r$ with $\mathcal{P}_\kappa(\gamma') \leq \mathcal{P}_\kappa(\gamma) + \varepsilon/2$. Part (ii) with $\varphi = 0$, the order parameter γ' , and $\varepsilon/2$ gives the claim, since $\mathcal{P}_{H_\kappa}(\gamma') = \mathcal{P}_\kappa(\gamma')$. $\square$

3.9 Volume from one feasible point

A bound on the volume does not by itself bound the largest feasible margin, because a feasible set may have zero spherical volume. Lemma 3.13 shows that one feasible point at margin κ produces exponentially small, but not superexponentially small, volume at margin $\kappa - \eta$, on a set where the fields $b_a \cdot \sigma$ change little. The set it constructs is used again in Chapter 9. The proof uses Šidák's inequality for a Gaussian vector whose covariance may be singular; Chapter 8 uses the same extension.

Lemma 3.12. *Let $(\xi_1, \dots, \xi_M)$ be a centered Gaussian vector with an arbitrary, possibly singular, covariance matrix, and let $c_1, \dots, c_M \geq 0$. Then*

$$\mathbb{P}\{|\xi_a| \leq c_a \text{ for all } a \leq M\} \geq \prod_{a=1}^M \mathbb{P}\{|\xi_a| \leq c_a\}.$$

Proof. For a nonsingular covariance matrix this is Šidák's inequality [116, Corollary 1], proved independently by Khatri [80]. If $c_a = 0$ and $\text{Var } \xi_a > 0$ for some a , the right side vanishes, and a coordinate with $c_a = 0$ and $\xi_a = 0$ contributes the factor one to both sides and may be removed. So we may assume that all $c_a > 0$. Let $\xi' \sim N(0, I_M)$ be independent of ξ , and for $\epsilon > 0$ put $\xi^\epsilon = \xi + \epsilon \xi'$, whose covariance matrix is nonsingular. Then $\mathbb{P}\{|\xi_a^\epsilon| \leq c_a \forall a\} \geq \prod_a \mathbb{P}\{|\xi_a^\epsilon| \leq c_a\}$. Let $\epsilon \downarrow 0$ along a sequence. Each factor converges to $\mathbb{P}\{|\xi_a| \leq c_a\}$: if $\text{Var } \xi_a > 0$, because ξ_a^ϵ and ξ_a are centered Gaussian with variances converging and a continuous limiting law; if $\xi_a = 0$, because $\mathbb{P}\{|\epsilon \xi'_a| \leq c_a\} \rightarrow 1$. Since $\xi^\epsilon \rightarrow \xi$ pointwise and the intervals are closed, the indicator of $\{|\xi_a^\epsilon| \leq c_a \forall a\}$ has lim sup at most the indicator of $\{|\xi_a| \leq c_a \forall a\}$. Fatou's lemma, applied to one minus the indicators, proves the lemma. $\square$

Lemma 3.13. *Fix $\alpha' < \infty$. There are constants $K_{\text{cap}} \geq 1$, $C_{\text{cap}} < \infty$, and N_0 , depending only on α' , with the following property. Let $N \geq N_0$, $M \leq \alpha'N$, and let $b_1, \dots, b_M \in \mathbb{R}^N$ be deterministic vectors with $|b_a|^2 \leq 2$. Let $\kappa \in \mathbb{R}$ and $\eta \in (0, 1]$, suppose that some $\sigma_0 \in S_N$ satisfies $b_a \cdot \sigma_0 \geq \kappa$ for all a , and put*

$$s_0 = \min \left\{ \frac{\eta}{2K_{\text{cap}}}, \sqrt{\frac{\eta}{2|\kappa| + 1}} \right\}.$$

There is a Borel set $T \subset \{\sigma \in S_N : b_a \cdot \sigma \geq \kappa - \eta \text{ for all } a\}$ with

$$\begin{aligned} \mu_N(T) &\geq \exp \left\{ -N(\log(2/s_0) + C_{\text{cap}}) \right\}, \\ |b_a \cdot \sigma - b_a \cdot \sigma_0| &\leq s_0^2 |b_a \cdot \sigma_0| + s_0 K_{\text{cap}} \quad (\sigma \in T, a \leq M). \end{aligned} \tag{3.41}$$

If κ ranges over a compact interval I , then $\log(2/s_0) \leq \log(1/\eta) + C_I$ with C_I depending only on α' and I .

Proof. We divide the proof into four steps.

STEP 1. Constants. Let $c_0 = \frac{1}{2}(\log 2 - \frac{1}{2}) > 0$. For $X \sim \chi_d^2$ and $0 < y < 1$, Markov's inequality applied to e^{-uX} with $u = (1-y)/(2y)$ gives the Chernoff bound $\mathbb{P}\{X \leq yd\} \leq e^{uyd}(1+2u)^{-d/2} = \exp\{-\frac{d}{2}(y-1-\log y)\}$; with $y = \frac{1}{2}$ this is $\mathbb{P}\{X \leq d/2\} \leq e^{-c_0 d}$. Let $\pi(z) = \mathbb{P}\{|\sqrt{2}G| \leq z/2\}$, and choose $K_{\text{cap}} \geq 1$ with $\alpha' \log(1/\pi(K_{\text{cap}})) \leq c_0/4$, which is possible because $\pi(z) \rightarrow 1$ as $z \rightarrow \infty$. Put $C_{\text{cap}} = c_0/4 + \log 2$, and let $N_0 \geq 3$ be such that $e^{-c_0(N-1)} \leq \frac{1}{2}e^{-c_0 N/4}$ for $N \geq N_0$.

STEP 2. Coordinates. Let P be the orthogonal projection onto $\sigma_0^\perp$. Every $\sigma \in S_N$ with $|\sigma \cdot \sigma_0| < N$ can be written uniquely as

$$\sigma = \rho \sigma_0 + s\sqrt{N}v, \quad \rho = \frac{\sigma \cdot \sigma_0}{N} \in (-1, 1), \quad s = \sqrt{1 - \rho^2}, \quad v \in \sigma_0^\perp, \quad |v| = 1.$$

Under μ_N , ρ and v are independent, v is uniform on the unit sphere of $\sigma_0^\perp$, and ρ has density $c'_N(1 - \rho^2)^{(N-3)/2}$ on $(-1, 1)$, with $c'_N = \Gamma(N/2)/(\sqrt{\pi}\Gamma((N-1)/2))$. This constant is nondecreasing in N by log-convexity of Γ , and equals $1/2$ at $N = 3$. Let

$$T = \left\{ \sigma : \rho \in [\sqrt{1 - s_0^2}, \sqrt{1 - s_0^2/4}], \sqrt{N}|b_a \cdot v| \leq K_{\text{cap}} \text{ for all } a \leq M \right\}.$$

It is a Borel set, and on it $s \in [s_0/2, s_0]$ and $\rho \geq 0$. Also $s_0 \leq \eta/(2K_{\text{cap}}) \leq 1/2$.

STEP 3. Fields. For $\sigma \in T$, $b_a \cdot \sigma - b_a \cdot \sigma_0 = (\rho - 1)b_a \cdot \sigma_0 + s\sqrt{N}b_a \cdot v$, and $0 \leq 1 - \rho \leq s^2 \leq s_0^2$ because $\rho = \sqrt{1 - s^2} \geq 1 - s^2$. This gives the second bound in (3.41). For feasibility, $b_a \cdot \sigma_0 \geq \kappa$ and $\rho \in [1 - s^2, 1]$ imply $\rho b_a \cdot \sigma_0 \geq \kappa - s^2|\kappa|$: if $\kappa > 0$, then $\rho b_a \cdot \sigma_0 \geq (1 - s^2)\kappa$; if $b_a \cdot \sigma_0 \geq 0 \geq \kappa$, the left side is nonnegative; and if $\kappa \leq b_a \cdot \sigma_0 < 0$, then $\rho b_a \cdot \sigma_0 \geq b_a \cdot \sigma_0$. Therefore

$$b_a \cdot \sigma \geq \kappa - s_0^2|\kappa| - s_0K_{\text{cap}} \geq \kappa - \frac{\eta}{2} - \frac{\eta}{2},$$

using $s_0^2 \leq \eta/(2|\kappa| + 1)$ and $s_0 \leq \eta/(2K_{\text{cap}})$. Thus every $\sigma \in T$ is feasible at margin $\kappa - \eta$.

STEP 4. Volume. On the interval of ρ defining T we have $1 - \rho^2 \geq s_0^2/4$, and the interval has length $\frac{3}{4}s_0^2/(\sqrt{1 - s_0^2/4} + \sqrt{1 - s_0^2}) \geq \frac{3}{8}s_0^2$. Hence its probability is at least $\frac{1}{2}(s_0/2)^{N-3} \cdot \frac{3}{8}s_0^2 = \frac{3}{2s_0}(s_0/2)^N \geq (s_0/2)^N$. For the direction, represent $v = P\xi/|P\xi|$ with $\xi \sim N(0, I_N)$, so that $b_a \cdot v = (Pb_a) \cdot \xi/|P\xi|$. On the event $\{|P\xi|^2 \geq N/4\} \cap \{|(Pb_a) \cdot \xi| \leq K_{\text{cap}}/2 \text{ for all } a\}$ we have $\sqrt{N}|b_a \cdot v| \leq K_{\text{cap}}$ for all a . The variables $(Pb_a) \cdot \xi$ are centered and jointly Gaussian with variances $|Pb_a|^2 \leq 2$, and $\mathbb{P}\{|N(0, v)| \leq K_{\text{cap}}/2\}$ is nonincreasing in v . By Lemma 3.12, the choice of K_{cap} , and $M \leq \alpha'N$,

$$\mathbb{P}\{|(Pb_a) \cdot \xi| \leq K_{\text{cap}}/2 \text{ for all } a\} \geq \pi(K_{\text{cap}})^M \geq e^{-c_0N/4}.$$

Since $|P\xi|^2$ has the law χ_{N-1}^2 and $N/4 \leq (N-1)/2$, Step 1 gives $\mathbb{P}\{|P\xi|^2 < N/4\} \leq e^{-c_0(N-1)} \leq \frac{1}{2}e^{-c_0N/4}$. The directional event therefore has probability at least $\frac{1}{2}e^{-c_0N/4}$. By independence of ρ and v ,

$$\mu_N(T) \geq \left(\frac{s_0}{2}\right)^N \frac{1}{2}e^{-c_0N/4} \geq \exp\{-N(\log(2/s_0) + C_{\text{cap}})\}.$$

Finally, for $\kappa \in I$ and $\eta \leq 1$ we have $\sqrt{\eta/(2|\kappa| + 1)} \geq \eta/\sqrt{2|\kappa| + 1}$, so $s_0 \geq c_I\eta$ with $c_I = \min\{1/(2K_{\text{cap}}), (2\sup_{\kappa \in I}|\kappa| + 1)^{-1/2}\}$, and $\log(2/s_0) \leq \log(1/\eta) + \log(2/c_I)$. $\square$

Corollary 3.14. Let $\alpha > 0$. For every real $\kappa > \kappa_c(\alpha)$, $\mathbb{P}\{\kappa_N \geq \kappa\} \rightarrow 0$.

Proof. Choose $\eta \in (0, 1]$ with $\kappa - \eta > \kappa_c$, and let N_0 and $C_V = \log(2/s_0) + C_{\text{cap}}$ be given by Lemma 3.13 for $\alpha' = \sup_N \alpha_N$, the margin κ , and this η . Since $\kappa - \eta > \kappa_c$, the definition of κ_c gives $\mathcal{P}_*(\kappa - \eta) = -\infty$, so there is $\gamma \in \mathcal{U}$ with $\mathcal{P}_{\kappa-\eta}(\gamma) < -C_V - 1$. By Corollary 3.11(i) with $\varepsilon = 1$,

$$\mathbb{P}\left\{\frac{1}{N} \log V_N(\kappa - \eta) \geq -C_V\right\} \leq \mathbb{P}\left\{\frac{1}{N} \log V_N(\kappa - \eta) > \mathcal{P}_{\kappa-\eta}(\gamma) + 1\right\} \longrightarrow 0.$$

On the other hand, suppose that $N \geq N_0$, $\kappa_N \geq \kappa$, and $\max_a |b_a|^2 \leq 2$. A maximizer σ_0 of $\min_a b_a \cdot \sigma$ over the compact set S_N satisfies $b_a \cdot \sigma_0 \geq \kappa$ for all a , so Lemma 3.13 gives $V_N(\kappa - \eta) \geq \mu_N(T) \geq e^{-NC_V}$. Finally, $N|b_a|^2$ has the law χ_N^2 , and the Chernoff bound $\mathbb{P}\{\chi_N^2 \geq 2N\} \leq e^{-N(1-\log 2)/2}$ (the upper-tail analogue of Step 1 in the proof of Lemma 3.13) gives

$$\mathbb{P}\left\{\max_a |b_a|^2 > 2\right\} \leq \alpha' N e^{-N(1-\log 2)/2} \longrightarrow 0.$$

Hence $\mathbb{P}\{\kappa_N \geq \kappa\} \leq \mathbb{P}\{N^{-1} \log V_N(\kappa - \eta) \geq -C_V\} + \mathbb{P}\{\max_a |b_a|^2 > 2\} \rightarrow 0. \quad \square$

The lower bound

In this chapter we prove the lower bound $\liminf_N F_N(U) \geq \mathcal{P}_*(U)$ for smooth activations U (Theorem 4.1). Together with the upper bound of Chapter 3, it gives the Parisi formula in Chapter 5. The argument is outlined in Step 2 of §1.3.

§4.1 introduces the spherical coordinates, the Poisson row count, and the three systems that are compared, and §4.2 proves the overlap representation theorem (Theorem 4.6) from the external results behind it. §4.3 defines the perturbation and states the perturbation, continuity, and precision lemmas, which are proved in §4.8. §4.4 reduces the increment to the row term and the coordinate term. §4.5 evaluates them on cascades with Lemma 3.4 and removes the cutoff on $\|\varepsilon\|$ there (Lemma 4.10). §4.6 proves the rotation estimate, and §4.7 proves Theorem 4.1.

The two cavity steps, adding constraints and adding coordinates, follow the cavity method for the perceptron in the replica-symmetric regime [115], [127, Chapters 2 and 3], [128, Chapter 8]. As in [37, 99], the trial order parameter comes from the limiting overlap law of the system. Compared with W.-K. Chen's spherical form [37] of the Aizenman–Sims–Starr scheme and with its multi-species version by Bates and Sohn [16], two points differ. First, in both works the cavity functional is identified with the spherical Parisi functional, as the number of cavity coordinates grows, through Talagrand's computation [125]; here the coordinate term is an explicit Gaussian integral on cascades, and the rotation invariance of the patterns identifies it with the Crisanti–Sommers entropy. Second, the other species in the Ghirlanda–Guerra identities is not a group of coordinates, as in [16, 100], but the response vector.

Activation class. Throughout this chapter, U satisfies, for some $U_{\max} < \infty$ and $L < \infty$ (enlarged, if necessary, so that $L \geq 1 + |U(0)|$),

$$U \in C^4(\mathbb{R}), \quad U \leq U_{\max}, \quad |U'(x)| \leq L(1 + |x|), \quad |U^{(j)}(x)| \leq L \quad (2 \leq j \leq 4), \quad (4.1)$$

and in addition either U' is bounded or U is concave. In the first case U belongs to class (A), and in the second to class (B). The derivatives of orders three and four are used only in the perturbation of §4.3, in the error estimates of the cavity interpolation, and in Lemma 4.9. They control the first and second derivatives of $xU'(x)$ and $U'(x)^2$. In Chapter 5 this extra smoothness is removed by mollification. By (4.1) and Taylor's

formula,

$$-C_0(1+x^2) \leq U(x) \leq U_{\max}, \quad C_0 = 2L. \quad (4.2)$$

Conventions. Throughout this chapter, C and c denote positive constants that may depend on α , $U_{\max}$, and L , but not on N ; their values may change from one occurrence to the next. A dependence on the cavity dimension n or on the cutoff Λ is indicated by a subscript, as in $C_{n,\Lambda}$ or $O_{n,\Lambda}(\cdot)$. The cavity dimension n is fixed until the last step of the proof. We write ν_n for the standard Gaussian measure on $\mathbb{R}^n$ and G for a standard Gaussian variable. The letter a is a row index, and $\mathbf{G}, \mathbf{G}', \mathbf{G}''$ denote Gaussian matrices with rows g_a, g'_a, g''_a . The letter b , with or without subscripts, denotes a precision parameter of a Gaussian integral, as in Lemma 3.4. The rows $b_a = g_a / \sqrt{N}$ of Chapter 1 and the derivative $b = \partial_x u$ of §2.3 are not used in this chapter.

The goal of the chapter is the following theorem.

Theorem 4.1. *For every $\alpha > 0$ and every activation U satisfying (4.1), with either bounded U' or concave U ,*

$$\liminf_{N \rightarrow \infty} F_N(U) \geq \mathcal{P}_*(U) = \inf_{\gamma \in \mathcal{U}} \{ \alpha u_\gamma(0, 0; U) + \text{CS}(\gamma) \}.$$

4.1 Spherical coordinates, the row count, and the three systems

We first separate the added coordinates without changing the surface measure. Throughout this chapter

$$N' = N + n.$$

For $\sigma \in S_N$ and $\varepsilon \in \mathbb{R}^n$ with $\|\varepsilon\|^2 < N'$, define

$$\chi(\varepsilon) = \sqrt{\frac{N' - \|\varepsilon\|^2}{N}}, \quad \rho(\sigma, \varepsilon) = (\chi(\varepsilon)\sigma, \varepsilon) \in S_{N'}.$$

The uniform probability measure $\mu_{N'}$ on $S_{N'}$ is the image of $\mu_N(d\sigma) \otimes f_{N,n}(\varepsilon) d\varepsilon$ under $(\sigma, \varepsilon) \mapsto \rho(\sigma, \varepsilon)$, where

$$f_{N,n}(\varepsilon) = \frac{\Gamma(N'/2)}{\Gamma(N/2)(N'/2)^{n/2}} \frac{1}{(2\pi)^{n/2}} \left(1 - \frac{\|\varepsilon\|^2}{N'}\right)^{N/2-1} \mathbf{1}_{\{\|\varepsilon\|^2 < N'\}}. \quad (4.3)$$

Indeed, write a uniform point of $S_{N'}$ as $\sqrt{N'} \eta / \|\eta\|$ for a standard Gaussian vector $\eta = (\eta_1, \eta_2) \in \mathbb{R}^N \times \mathbb{R}^n$. Its last n coordinates are $\varepsilon = \sqrt{N'} \eta_2 / \|\eta\|$, and its first N coordinates are $\chi(\varepsilon)\sigma$ with $\sigma = \sqrt{N} \eta_1 / \|\eta_1\|$. The vector σ is uniform on S_N and independent of $(\|\eta_1\|, \eta_2)$, hence of ε . The variable $\|\varepsilon\|^2 / N' = \|\eta_2\|^2 / \|\eta\|^2$ has the beta distribution with parameters $n/2$ and $N/2$, and the direction of ε is uniform and independent of its norm. Writing the beta density in polar coordinates on $\mathbb{R}^n$ gives (4.3). For $\Lambda \geq 1$ let

$$\mathbb{B}_\Lambda = \{\varepsilon \in \mathbb{R}^n : \|\varepsilon\|^2 \leq \Lambda n\}.$$

Stirling's formula, $\Gamma(N'/2)/\Gamma(N/2) = (N'/2)^{n/2}(1 + O_n(N^{-1}))$, and the expansion $(N/2 - 1)\log(1 - \|\varepsilon\|^2/N') = -\|\varepsilon\|^2/2 + O_{n,\Lambda}(N^{-1})$ give, uniformly on $\mathbb{B}_\Lambda$,

$$f_{N,n}(\varepsilon) = \frac{e^{-\|\varepsilon\|^2/2}}{(2\pi)^{n/2}}(1 + O_{n,\Lambda}(N^{-1})), \quad \chi(\varepsilon) = 1 + \frac{n - \|\varepsilon\|^2}{2N} + O_{n,\Lambda}(N^{-2}). \quad (4.4)$$

W.-K. Chen [37] separates the cavity coordinates of spherical models by a decomposition of the uniform measure of the same kind. The expansion of $f_{N,n}$ reflects the classical fact that a fixed number of coordinates of a uniform point on a high-dimensional sphere are asymptotically independent standard Gaussian variables; see [47].

Gaussian matrices. We use the following standard bound on the operator norm of a Gaussian matrix [130, Corollary 5.35], here and in Chapters 5 and 8.

Lemma 4.2. *Let $\mathbf{A}$ be an $M \times N$ matrix with independent standard Gaussian entries. For every $t \geq 0$,*

$$\mathbb{P}\{\|\mathbf{A}\|_{\text{op}} > \sqrt{M} + \sqrt{N} + t\} \leq 2e^{-t^2/2}.$$

Consequently $\mathbb{E}\|\mathbf{A}\|_{\text{op}}^2 \leq 4(M + N) + 8$, and $\mathbb{E}\exp(c\|\mathbf{A}\|_{\text{op}}^2/N) \leq 3\exp(4c(M + N)/N)$ for $0 < c \leq N/8$.

Proof. The tail bound is the upper half of [130, Corollary 5.35]. Put $T = (\|\mathbf{A}\|_{\text{op}} - \sqrt{M} - \sqrt{N})_+$, so that $\mathbb{P}(T > t) \leq 2e^{-t^2/2}$ and $\|\mathbf{A}\|_{\text{op}}^2 \leq 4(M + N) + 2T^2$. Then $\mathbb{E}T^2 = \int_0^\infty 2t\mathbb{P}(T > t) dt \leq 4$, which gives the second bound. For $0 < s \leq 1/4$,

$$\mathbb{E}e^{sT^2} = 1 + \int_0^\infty 2st e^{st^2} \mathbb{P}(T > t) dt \leq 1 + \frac{2s}{1/2 - s} \leq 3,$$

and the choice $s = 2c/N$ gives the third bound. $\square$

Poissonization. We replace the number M_N of rows by a Poisson variable $M \sim \text{Poi}(\alpha N)$, and when passing to dimension N' we add independently $\pi \sim \text{Poi}(\alpha n)$ rows, so that the N' -dimensional system has $M + \pi \sim \text{Poi}(\alpha N')$ rows; from now on the letter π denotes this count. This changes the free energy by $o(1)$. To check this, let $h_N(j)$ be the expected log partition function in dimension N with j rows. Adding a fresh row g changes h_N by $\mathbb{E}\log\langle e^{U(g \cdot \sigma / \sqrt{N})} \rangle$, where $\langle \cdot \rangle$ is the Gibbs measure before the row is added. By Jensen's inequality and $U \leq U_{\max}$, this lies between $\mathbb{E}U(G)$ and $U_{\max}$, because $g \cdot \sigma / \sqrt{N}$ is standard Gaussian for every fixed σ and g is independent of the Gibbs measure. By (4.2), $|\mathbb{E}U(G)| < \infty$. Hence h_N is Lipschitz with a constant independent of N , and

$$|\mathbb{E}h_N(M) - h_N(M_N)| \leq C\mathbb{E}|M - M_N| \leq C(\sqrt{\alpha N} + |\alpha N - M_N|) = o(N).$$

Poisson numbers of constraints are standard in the study of diluted spin glasses [57, 102].

Conditioning on the number of rows. The concentration estimates below are proved at a fixed number of rows, so we carry out the cavity computation conditionally on the value of M , for values in the window

$$W_N = \{j \in \mathbb{N} : |j - \alpha N| \leq N^{2/3}\}.$$

The Poisson tail bound $\mathbb{P}(|M - \alpha N| > t) \leq 2 \exp(-t^2 / (2(\alpha N + t)))$ gives $\mathbb{P}(M \notin W_N) \leq 2e^{-cN^{1/3}}$, and for large N every $j \in W_N$ satisfies $j \leq 2\alpha N$. Step 2 of §4.7 reduces the proof to one value of M in W_N . From §4.3 on, M denotes a fixed integer in W_N , expectations are conditional on this value, and all bounds and all terms $o(1)$ are uniform in $M \in W_N$; the exceptions are Lemma 4.7(i) and the quantities Ξ_N of Step 1 of §4.7, which concern the Poissonized systems. The number $\pi \sim \text{Poi}(\alpha n)$ of rows added in dimension N' remains Poisson throughout.

The three systems. We compare three systems. Each has M rows, and the N' -dimensional system has π further rows.

- (a) The N -dimensional system has configuration space S_N , rows $g_a^{(N)} \in \mathbb{R}^N$ for $a \leq M$, and Hamiltonian $\sum_{a \leq M} U(g_a^{(N)} \cdot \sigma / \sqrt{N})$.
- (b) The N' -dimensional system has configuration space $S_{N'}$ and rows $(g_a, g'_a) \in \mathbb{R}^N \times \mathbb{R}^n$ for $a \leq M$, together with π further rows $\tilde{g}_i \in \mathbb{R}^{N'}$. We write $\langle \cdot \rangle_{N'}$ for its Gibbs measure with the first M rows only.
- (c) The *reference system* has configuration space S_N , the rows $g_a \in \mathbb{R}^N$ for $a \leq M$ (the first N coordinates of the rows of the N' -dimensional system), and the fields

$$S_a(\sigma) = \frac{g_a \cdot \sigma}{\sqrt{N'}}.$$

All rows are independent standard Gaussian vectors. We write $\mathbf{G}, \mathbf{G}'$ for the $M \times N$ and $M \times n$ matrices with rows g_a and g'_a . In the coordinates $\rho = \rho(\sigma, \varepsilon)$ the fields of the N' -dimensional system are

$$\frac{(g_a, g'_a) \cdot \rho}{\sqrt{N'}} = \chi(\varepsilon) S_a(\sigma) + w_a(\varepsilon), \quad w_a(\varepsilon) = \frac{g'_a \cdot \varepsilon}{\sqrt{N'}}. \quad (4.5)$$

The fields of the N -dimensional system can be realized from the same rows. Let $g''_a \in \mathbb{R}^N$ be independent standard Gaussian vectors, forming the matrix $\mathbf{G}''$, and put

$$\eta_a(\sigma) = c_N g''_a \cdot \sigma, \quad c_N^2 = \frac{1}{N} - \frac{1}{N'} = \frac{n}{NN'}. \quad (4.6)$$

As processes indexed by $\sigma \in S_N$, the centered Gaussian families $(S_a + \eta_a)_{a \leq M}$ and $(g_a^{(N)} \cdot \sigma / \sqrt{N})_{a \leq M}$ have the same covariance $\sigma \cdot \sigma' / N$, hence the same law. Both the N' -dimensional and the N -dimensional partition functions are therefore integrals against the reference system with modified fields.

Observables of the reference system. The following quantities determine the coordinate term:

$$\begin{aligned} A_N(\sigma) &= \frac{1}{N} \sum_{a \leq M} S_a(\sigma) U'(S_a(\sigma)), & B_N(\sigma) &= \frac{1}{N'} \sum_{a \leq M} U''(S_a(\sigma)), \\ X_N(\sigma, \sigma') &= \frac{1}{N'} \sum_{a \leq M} U'(S_a(\sigma)) U'(S_a(\sigma')), & R_N(\sigma, \sigma') &= \frac{\sigma \cdot \sigma'}{N}, \\ b_N(\sigma) &= 1 + A_N(\sigma) - B_N(\sigma). \end{aligned} \quad (4.7)$$

The spin overlap is R_N . We call X_N the *response overlap*; it is a Gram matrix, hence positive semidefinite, and its diagonal is $X_N(\sigma, \sigma) = N'^{-1} \sum_a U'(S_a(\sigma))^2$. The terms A_N and B_N arise below from the radial factor $\chi(\varepsilon)$ and from the second-order term of a Gaussian integration by parts, respectively, and $b_N - 1$ will be the coefficient of $-(\|\varepsilon\|^2 - n)/2$ in the exponent of the Gaussian integral over ε .

Let Ω_N be the event on which the operator norms of $\mathbf{G}$, $\mathbf{G}''$, and $\mathbf{G}'$ are at most $C_1 \sqrt{N}$. Since $M \leq 2\alpha N$, Lemma 4.2 with $t = \sqrt{N}$ shows that $\mathbb{P}(\Omega_N^c) \leq 6e^{-N/2}$ for $N \geq n$, with the constant $C_1 = \sqrt{2\alpha} + 4$. Since $\sum_a S_a(\sigma)^2 = \|\mathbf{G}\sigma\|^2/N' \leq \|\mathbf{G}\|_{\text{op}}^2$, the growth bounds in (4.1) give, for every $\sigma \in S_N$,

$$\sum_{a \leq M} S_a(\sigma)^2 \leq \|\mathbf{G}\|_{\text{op}}^2, \quad |b_N(\sigma)| + X_N(\sigma, \sigma) + |B_N(\sigma)| \leq C \left(1 + \frac{\|\mathbf{G}\|_{\text{op}}^2}{N}\right). \quad (4.8)$$

On Ω_N the right sides are at most CN and C . By Lemma 4.2, for every $c > 0$ and $N \geq N_0(c)$,

$$\mathbb{E} \exp \left(c \sup_{\sigma \in S_N} (|b_N(\sigma)| + X_N(\sigma, \sigma)) \right) \leq C_c, \quad (4.9)$$

uniformly in $M \leq 2\alpha N$. In particular, all polynomial moments of these quantities and of $\|\mathbf{G}\|_{\text{op}}^2/N$ are bounded.

4.2 The overlap representation theorem

The cavity computation requires a description of the limiting joint law of the spin overlap R_N and the response overlap X_N . We use a general result about random probability measures on Hilbert spaces, which does not involve the perceptron Hamiltonian. In this section we state it, recall the three external results on which it rests, and derive it from them.

Setting. For each $N \geq 1$, let $\mathfrak{G}_N$ be a random probability measure on the product of the unit balls of two separable Hilbert spaces, which may depend on N . For independent samples $(x_N^\ell, y_N^\ell)_{\ell \geq 1}$ from $\mathfrak{G}_N$ put

$$R_{\ell\ell'}^{1,N} = x_N^\ell \cdot x_N^{\ell'}, \quad R_{\ell\ell'}^{2,N} = y_N^\ell \cdot y_N^{\ell'} \quad (\ell, \ell' \geq 1).$$

We write $\mathbb{E}\langle \cdot \rangle_N$ for expectation over the random measure and its samples. The arrays take values in $[-1, 1]$, they are positive semidefinite, and they are invariant in law under finite permutations of the sample labels. Suppose that the pair of arrays converges in finite-dimensional distribution to a pair (R^1, R^2) , and write $\mathbb{E}\langle \cdot \rangle$ for expectations of the limiting arrays. The limiting arrays have the same three properties. For $r \geq 1$, rational $v_1, v_2 \geq 0$ and an integer $m \geq 1$ put $Q(x, y) = (v_1x + v_2y)^m$. For a bounded measurable function f of the arrays of the first r samples, the Ghirlanda–Guerra discrepancy of §3.4, written with r replicas since n is the cavity dimension here, is

$$\begin{aligned} \mathfrak{D}_r^N(f, Q) &= \mathbb{E}\langle fQ(R_{1,r+1}^{1,N}, R_{1,r+1}^{2,N}) \rangle_N - \frac{1}{r} \mathbb{E}\langle f \rangle_N \mathbb{E}\langle Q(R_{12}^{1,N}, R_{12}^{2,N}) \rangle_N \\ &\quad - \frac{1}{r} \sum_{\ell=2}^r \mathbb{E}\langle fQ(R_{1\ell}^{1,N}, R_{1\ell}^{2,N}) \rangle_N. \end{aligned} \quad (4.10)$$

Here f may depend on the diagonal entries of the arrays. For a single random measure $\mathfrak{G}$ we drop the index N and write $\mathfrak{D}_r(f, Q)$.

Ruelle probability cascade arrays with paired levels. We use the Ruelle probability cascade of §3.1: a depth $k \geq 0$, parameters $0 = \theta_0 < \theta_1 < \dots < \theta_k < \theta_{k+1} = 1$, masses $w_l = \theta_{l+1} - \theta_l$ for $0 \leq l \leq k$, weights (v_β) indexed by the leaves β of the infinitely branching tree of depth k , and the depth $\beta \wedge \beta'$ of the last common ancestor of two leaves. If β^1 and β^2 are sampled independently from the weights (v_β) , then the averaged probability that $\beta^1 \wedge \beta^2 = l$ is w_l , by (3.4) with the leaf function zero. Given levels $0 \leq q_0 \leq \dots \leq q_k$ and $0 \leq p_0 \leq \dots \leq p_k$, the *Ruelle probability cascade array with paired levels* (q_l, p_l) is the pair of arrays with off-diagonal entries $(q_{\beta^\ell \wedge \beta^{\ell'}}, p_{\beta^\ell \wedge \beta^{\ell'}})$, $\ell \neq \ell'$, where $\beta^1, \beta^2, \dots$ are independent samples from (v_β) . The law of its pair of $(1, 2)$ entries is $\sum_{l=0}^k w_l \delta_{(q_l, p_l)}$. A single sequence of levels $0 \leq s_0 < \dots < s_k$ gives the *Ruelle probability cascade array with levels* (s_l) , whose overlap law is $\sum_l w_l \delta_{s_l}$; every probability measure with finite support in $[0, \infty)$ is of this form.

External results. We use three results from the literature. The first is the Dovbysh–Sudakov representation [51] of positive semidefinite exchangeable arrays, in the form of [96, Proposition 1]. A random array $(R_{\ell\ell'})_{\ell, \ell' \geq 1}$ is called a *Gram–de Finetti array* if it is symmetric, every finite principal submatrix is positive semidefinite, and its law is invariant under finite permutations of the indices.

Lemma 4.3. *Let $(R_{\ell\ell'})$ be a Gram–de Finetti array with $R_{\ell\ell} \leq 1$ almost surely. There is a random probability measure $\mathfrak{G}$ on the unit ball of a separable Hilbert space such that, for independent samples $\sigma^1, \sigma^2, \dots$ from $\mathfrak{G}$, the off-diagonal array $(R_{\ell\ell'})_{\ell \neq \ell'}$ has the law of $(\sigma^\ell \cdot \sigma^{\ell'})_{\ell \neq \ell'}$.*

Proof. By [96, Proposition 1], the law of the array is a mixture over a random parameter ω of the laws of $(h_\ell \cdot h_{\ell'} + a_\ell \delta_{\ell\ell'})_{\ell, \ell'}$, where (h_ℓ, a_ℓ) are independent samples from a

probability measure η_ω on $\ell^2 \times [0, \infty)$. Since $\|h_\ell\|^2 \leq \|h_\ell\|^2 + a_\ell = R_{\ell\ell} \leq 1$ almost surely, η_ω is concentrated on pairs with $\|h\| \leq 1$ for almost every ω . The law of the first coordinate under η_ω is the required random measure. $\square$

The second result, from [49, Sections 5.6–5.7], characterizes the scalar arrays that satisfy the Ghirlanda–Guerra identities. Let $\mathfrak{G}$ be a random probability measure on the unit ball of a separable Hilbert space, with overlap array $R = (\sigma^\ell \cdot \sigma^{\ell'})_{\ell \neq \ell'}$ and overlap law $\zeta(A) = \mathbb{E}\langle \mathbf{1}_{\{R_{12} \in A\}} \rangle$. We say that $\mathfrak{G}$ satisfies the *Ghirlanda–Guerra identities* [64] if for every $r \geq 1$, every bounded measurable function f of $(R_{\ell\ell'})_{\ell \neq \ell' \leq r}$, and every bounded measurable $\psi : \mathbb{R} \rightarrow \mathbb{R}$,

$$\mathbb{E}\langle f\psi(R_{1,r+1}) \rangle = \frac{1}{r}\mathbb{E}\langle f \rangle \mathbb{E}\langle \psi(R_{12}) \rangle + \frac{1}{r} \sum_{\ell=2}^r \mathbb{E}\langle f\psi(R_{1\ell}) \rangle. \quad (4.11)$$

Part (a) of the following lemma is Panchenko’s characterization of such measures, which rests on his proof of ultrametricity [97] (see also [98, Theorem 2.14]); the positivity of the overlap is Talagrand’s positivity principle (see [98, Theorem 2.16]). The identities for Ruelle probability cascades used in part (b), [49, Theorem 5.28], are due to Bovier and Kurkova [24].

Lemma 4.4. *Let $\mathfrak{G}$ satisfy the Ghirlanda–Guerra identities, with overlap array R and overlap law ζ .*

- (a) *The measure ζ is supported in $[0, 1]$, and the law of R is determined by ζ .*
- (b) *If ζ_h , $h \geq 1$, are probability measures with finite support in $[0, 1]$ that converge weakly to ζ , then the Ruelle probability cascade arrays whose overlap laws are ζ_h converge in finite-dimensional distribution to R .*

Proof. Part (a) is [49, Theorem 5.29], together with $|R_{12}| \leq 1$. For (b), [49, Corollary 5.32] shows that the cascade arrays converge in finite-dimensional distribution (the Poisson–Dirichlet cascades of [49] are the Ruelle probability cascades of §3.1). The limit has overlap law ζ , and it satisfies the identities (4.11): each cascade array satisfies them by [49, Theorem 5.28], so the limit satisfies them for continuous f and ψ , and hence for bounded measurable f and ψ , because for fixed ψ (respectively f) both sides are integrals against finite signed measures. The cascade array with levels $s_0 < \dots < s_k$ in $[0, 1]$ is the off-diagonal part of the Gram array of independent samples from the random measure $\sum_\beta v_\beta \delta_{h_\beta}$, where $h_\beta = \sum_{l=0}^k (s_l - s_{l-1})^{1/2} e_{\beta|l}$ with $s_{-1} = 0$ and an orthonormal family (e_ν) indexed by the nodes; its diagonal entries equal $s_k \leq 1$. Along a subsequence these Gram arrays converge in finite-dimensional distribution, and the limit is a Gram–de Finetti array with diagonal entries at most one whose off-diagonal part is the limit of the cascade arrays. By Lemma 4.3 this off-diagonal part is the overlap array of a random measure on a unit ball, and this measure satisfies the Ghirlanda–Guerra identities, since these are statements about the law of the off-diagonal array. By (a), the limit has the law of R . $\square$

The third result is Mourrat's quantitative synchronization theorem [92, Theorem 5.3], a finitary variant of Panchenko's synchronization theorem; see §3.5. For a real random variable Y let $F_Y^{-1}(v) = \inf\{s \in \mathbb{R} : \mathbb{P}(Y \leq s) \geq v\}$ for $v \in [0, 1]$, and let V be uniform on $[0, 1]$. The pair $(F_{Y_1}^{-1}(V), F_{Y_2}^{-1}(V))$ is the *monotone coupling* of the laws of Y_1 and Y_2 .

Lemma 4.5. *Fix an enumeration $(\vartheta_i)_{i \geq 1}$ of $\mathbb{Q} \cap [0, 1]$, and put $c_j(x, y) = (\vartheta_{j_1}x + \vartheta_{j_2}y)^{j_3}$ for $j = (j_1, j_2, j_3)$, as in Lemma 3.8.*

- (a) *For every $\epsilon > 0$ there is $\delta > 0$ with the following property. Let $\mathfrak{G}$ be a random probability measure on the product of the unit balls of two Hilbert spaces, with arrays R^1, R^2 as above. Suppose that $|\mathfrak{D}_r(f, c_j)| \leq \delta$ for all $r, j_1, j_2, j_3 \leq \lfloor \delta^{-1} \rfloor$ and all continuous functions f of the arrays of r samples with $\|f\|_\infty \leq 1$, where $\mathfrak{D}_r$ is (4.10) for $\mathfrak{G}$. Then, for every $\varphi \in C^\infty(\mathbb{R}^2)$,*

$$\left| \mathbb{E} \langle \varphi(R_{12}^1, R_{12}^2) \rangle - \mathbb{E} \varphi(F_1^{-1}(V), F_2^{-1}(V)) \right| \leq \epsilon (\|\varphi\|_\infty + \|\nabla \varphi\|_\infty),$$

where F_a^{-1} is the function F_Y^{-1} for the law of R_{12}^a under $\mathbb{E} \langle \cdot \rangle$.

- (b) *If $Y_{1,N} \rightarrow Y_1$ and $Y_{2,N} \rightarrow Y_2$ in law, then the monotone couplings converge in law: $(F_{Y_{1,N}}^{-1}(V), F_{Y_{2,N}}^{-1}(V)) \rightarrow (F_{Y_1}^{-1}(V), F_{Y_2}^{-1}(V))$.*

Proof. Part (a) is [92, Theorem 5.3], whose hypothesis is formulated with the same defects, for continuous test functions of the arrays of r samples including their diagonal entries. Part (b) is [92, Lemma 5.4]. $\square$

Theorem 4.6. *In the setting above, assume that for every $r \geq 1$, every integer $m \geq 1$, and all rational $v_1, v_2 \geq 0$,*

$$\sup_{\|f\|_\infty \leq 1} |\mathfrak{D}_r^N(f, Q)| \longrightarrow 0 \quad (N \rightarrow \infty), \quad (4.12)$$

where the supremum is over bounded measurable functions of the arrays of the first r samples. Then the limiting pair satisfies the Ghirlanda–Guerra identities

$$\mathbb{E} \langle f Q(R_{1,r+1}^1, R_{1,r+1}^2) \rangle = \frac{1}{r} \mathbb{E} \langle f \rangle \mathbb{E} \langle Q(R_{12}^1, R_{12}^2) \rangle + \frac{1}{r} \sum_{\ell=2}^r \mathbb{E} \langle f Q(R_{1\ell}^1, R_{1\ell}^2) \rangle \quad (4.13)$$

for every bounded measurable f of the arrays of its first r samples. Moreover:

- (i) *The off-diagonal entries of R^1 and R^2 are nonnegative almost surely.*
- (ii) *The off-diagonal pair of arrays is a limit in finite-dimensional distribution of Ruelle probability cascade arrays with paired nondecreasing levels $(q_l, p_l) \in [0, 1]^2$.*

Part (ii) is the form, for a pair of Gram arrays, of Panchenko's synchronization combined with his approximation by Ruelle probability cascades [100]. Step 1 of the proof follows the proof of [100, Theorem 3] in Section 3 there, Step 2 the positivity argument of [100, Section 4], and Step 4 the approximation by cascades in [100, Section 5].

In Step 3 we use Mourrat's quantitative form, Lemma 4.5(a). It applies to the Gram pairs at finite N , before passing to the limit, so no representing measure of the limiting arrays needs to preserve their diagonal values.

Proof. We divide the proof into four steps.

STEP 1. The identities in the limit. For continuous f , finite-dimensional convergence and (4.12) give (4.13), since all arrays take values in $[-1, 1]$. Both sides of (4.13) are integrals of f against finite signed measures on the arrays of r samples, and they agree on continuous functions, so they agree for bounded measurable f . Choosing $(v_1, v_2) = (1, 0)$, $(0, 1)$ and $(1/2, 1/2)$ gives (4.11) with $\psi(x) = x^m$, $m \geq 1$, for the scalar arrays R^1 , R^2 , and $S/2$, where $S = R^1 + R^2$, and for test functions f of their off-diagonal entries. It also holds for constant ψ , since the sum of its three coefficients is one. By Weierstrass approximation it holds for continuous ψ on $[-1, 1]$, and, again by equality of finite measures, for bounded measurable ψ .

STEP 2. Positivity. Each of the arrays R^1 , R^2 , and $S/2$ is a Gram-de Finetti array with diagonal entries at most one, as a limit of such arrays. By Lemma 4.3, its off-diagonal part is the overlap array of a random measure on a unit ball, and this measure satisfies the Ghirlanda-Guerra identities by Step 1. Lemma 4.4(a) gives (i), and it shows that the law of the off-diagonal array $S/2$ is determined by the law of $S_{12}/2$.

STEP 3. Monotone coupling. Let $\epsilon > 0$, and let $\delta > 0$ be given by Lemma 4.5(a). The hypothesis of Lemma 4.5(a) involves finitely many tuples (r, j_1, j_2, j_3) , and by (4.12) it holds for $\mathfrak{G}_N$ for all sufficiently large N . Hence, for $\varphi \in C^\infty(\mathbb{R}^2)$,

$$\left| \mathbb{E} \langle \varphi(R_{12}^{1,N}, R_{12}^{2,N}) \rangle_N - \mathbb{E} \varphi(F_{1,N}^{-1}(V), F_{2,N}^{-1}(V)) \right| \leq \epsilon (\|\varphi\|_\infty + \|\nabla \varphi\|_\infty),$$

where $F_{a,N}^{-1}$ is defined from the law of $R_{12}^{a,N}$. As $N \rightarrow \infty$, the first term converges to $\mathbb{E} \langle \varphi(R_{12}^1, R_{12}^2) \rangle$, and by Lemma 4.5(b) the second converges to $\mathbb{E} \varphi(F_1^{-1}(V), F_2^{-1}(V))$. Since ϵ is arbitrary, the law of (R_{12}^1, R_{12}^2) is the monotone coupling of its marginals. The maps F_1^{-1} and F_2^{-1} are nondecreasing, so the image of $(0, 1)$ under $v \mapsto (F_1^{-1}(v), F_2^{-1}(v))$ is totally ordered for the coordinatewise order, and so is its closure. Thus the support C of the law of (R_{12}^1, R_{12}^2) , a compact subset of $[0, 1]^2$ by (i), is totally ordered. Two points (x, y) and (x', y') of C with $x + y \leq x' + y'$ therefore satisfy

$$0 \leq x' - x \leq (x' + y') - (x + y), \quad 0 \leq y' - y \leq (x' + y') - (x + y).$$

In particular the sum determines both coordinates: on the compact set $\{x + y : (x, y) \in C\}$, which is the support of the law of S_{12} , there are nondecreasing 1-Lipschitz functions Q_1, Q_2 with $(x, y) = (Q_1(x + y), Q_2(x + y))$ on C . We extend them to nondecreasing 1-Lipschitz functions on $\mathbb{R}$, by linear interpolation across the gaps of this compact set

and by constants outside its hull. Almost surely $R_{12}^1 = Q_1(S_{12})$ and $R_{12}^2 = Q_2(S_{12})$. By exchangeability and a countable intersection, the same holds for every off-diagonal pair of indices.

STEP 4. Approximation by cascades. Let F^{-1} be the function F_Y^{-1} for the law ζ_S of $S_{12}/2$, and let ζ_h be the law of $F^{-1}(\lceil hV \rceil / h)$. It has finite support contained in the support of ζ_S , and it converges weakly to ζ_S as $h \rightarrow \infty$, since $0 \leq \lceil hV \rceil / h - V < 1/h$ almost surely and F^{-1} is continuous at almost every point of $(0, 1)$. By Steps 1 and 2 and Lemma 4.4(b), the cascade arrays with overlap laws ζ_h , with levels $s_0^h < \dots < s_k^h$ in $[0, 1]$, converge in finite-dimensional distribution to $(S_{\ell\ell'}/2)_{\ell \neq \ell'}$. Apply the maps $s \mapsto (Q_1(2s), Q_2(2s))$ to their entries. The resulting arrays are the cascade arrays with the paired levels $(q_l, p_l) = (Q_1(2s_l^h), Q_2(2s_l^h))$, which are nondecreasing in l and lie in $[0, 1]^2$, because $2s_l^h$ lies in the support of the law of S_{12} . By continuity of Q_1 and Q_2 they converge in finite-dimensional distribution to $(Q_1(S_{\ell\ell'}), Q_2(S_{\ell\ell'}))_{\ell \neq \ell'} = (R_{\ell\ell'}^1, R_{\ell\ell'}^2)_{\ell \neq \ell'}$. This proves (ii). $\square$

4.3 Perturbations, continuity, and the precision identity

The cavity computation needs three properties of the Gibbs measures, which we obtain by adding a small perturbation to each Hamiltonian. First, the Ghirlanda–Guerra identities (4.13) for the spin and response overlaps must hold in the limit, so that Theorem 4.6 applies. Second, the per-configuration observables b_N and $X_N(\sigma, \sigma)$ must concentrate, so that they can be replaced by constants in the limit. Third, the perturbations must be compatible with adding coordinates and rows. In this section we define the perturbation and state its properties (Lemma 4.7), a continuity lemma for the functionals that appear in the cavity computation (Lemma 4.8), and an identity for the mean precision (Lemma 4.9). We then explain how these lemmas are combined with Theorem 4.6. The three lemmas are proved in §4.8. Every assertion in this section concerns a fixed number n of added coordinates and a fixed cutoff Λ for their squared norm; removing the cutoff is a separate step.

The perturbation. Consider a d -dimensional system, $d \in \{N, N'\}$, with configurations $\rho \in S_d$ and rows $g_a \in \mathbb{R}^d$ indexed by a finite set $\mathcal{A}$ (in the N' -dimensional system, $\mathcal{A} = \{1, \dots, M\}$ for the Gibbs measure $\langle \cdot \rangle_{N'}$, and $\mathcal{A}$ also contains the π added rows when these are present), and write $x_a(\rho) = g_a \cdot \rho / \sqrt{d}$. The perturbation has two parts.

Scalar part. Put

$$\tau_1(x) = xU'(x) - U''(x), \quad \tau_2(x) = U'(x)^2, \quad \xi_d = d^{-1/4},$$

and add $\xi_d \sum_{a \in \mathcal{A}} (u_1 \tau_1 + u_2 \tau_2)(x_a(\rho))$ to the Hamiltonian, where $u_1, u_2 \in [1, 2]$ are parameters. Equivalently, the activation U is replaced by

$$U^\xi = U + \xi(u_1 \tau_1 + u_2 \tau_2), \quad \xi = \xi_d. \quad (4.14)$$

By (4.1), $|\tau_i(x)| \leq C_\tau(1+x^2)$ for a constant C_τ depending only on L , and the first and second derivatives of τ_i grow at most linearly. The derivative of the Hamiltonian in u_i is $\xi_d \sum_a \tau_i(x_a(\rho))$. For the reference fields these sums are exactly

$$\sum_{a \leq M} \tau_2(S_a) = N' X_N(\sigma, \sigma), \quad \sum_{a \leq M} \tau_1(S_a) = N A_N - N' B_N = N(b_N - 1) - n B_N, \quad (4.15)$$

so concentration of these derivatives gives concentration of $X_N(\sigma, \sigma)$ and b_N .

Gaussian part. We first bound the response vectors without changing them on a typical event. Fix a smooth function $\psi_0 : [0, \infty) \rightarrow (0, 1]$ with $\psi_0 = 1$ on $[0, 3/4]$, $\psi_0(t) \leq 1/t$ for $t \geq 1$, and $\psi_0(t) = 1/t$ for $t \geq 2$, and define $\omega(y) = \psi_0(\|y\|)y$ on every Euclidean space. Then ω equals the identity on the ball of radius $3/4$, takes values in the closed unit ball, and its first and second derivatives are bounded by a constant that does not depend on the dimension. Put

$$\Psi_d(\rho) = \frac{(U'(x_a(\rho)))_{a \in \mathcal{A}}}{\sqrt{C_X d}},$$

where C_X is a constant depending only on α and L . We choose C_X so large that, on the event Ω_N , the vector $\Psi_d(\rho)$ with $\mathcal{A} = \{1, \dots, M\}$ has norm at most $1/\sqrt{2}$ for every $\rho \in S_d$ in each of the three systems. We also require the same bound in the N' -dimensional system when $\mathcal{A}$ contains the π added rows as well, on the event

$$\Omega'_N = \Omega_N \cap \{\pi \leq N, \|\tilde{\mathbf{G}}\|_{\text{op}} \leq C_1 \sqrt{N}\},$$

where $\tilde{\mathbf{G}}$ is the $\pi \times N'$ matrix with rows $\tilde{g}_i$. This is possible because $\|\Psi_d(\rho)\|^2 \leq 2L^2(\#\mathcal{A} + \sum_a x_a(\rho)^2)/(C_X d)$ by (4.1), on Ω_N the fields of the three systems ((a)–(c) of §4.1, with the fields $S_a + \eta_a$ for the N -dimensional system) satisfy $\sum_{a \leq M} x_a(\rho)^2 \leq CN$, and the fields of the added rows satisfy $\sum_{i \leq \pi} (\tilde{g}_i \cdot \rho)^2 / N' \leq \|\tilde{\mathbf{G}}\|_{\text{op}}^2$. By Lemma 4.2 and the Poisson tail of π , $\mathbb{P}(\Omega'_N) \leq C_n e^{-N/2}$. Set

$$\hat{\Psi}_d = \omega \circ \Psi_d, \quad \hat{X}_d(\rho, \rho') = \hat{\Psi}_d(\rho) \cdot \hat{\Psi}_d(\rho'), \quad R(\rho, \rho') = \frac{\rho \cdot \rho'}{d}.$$

Since $1/\sqrt{2} < 3/4$, we have $\hat{\Psi}_d = \Psi_d$ on Ω_N when $\mathcal{A} = \{1, \dots, M\}$. Hence, on Ω_N , $\hat{X}_d$ is the response overlap $d^{-1} \sum_a U'(x_a(\rho))U'(x_a(\rho'))$ of the d -dimensional system divided by C_X .

Enumerate the pairs $v = (v_1, v_2)$ of nonnegative rationals with $v_1 + v_2 \leq 1$ as $v^{(1)}, v^{(2)}, \dots$, write $j(v)$ for the index of v , and put $c_{v,m} = 2^{-j(v)-m}$ for integers $m \geq 1$. Let $h_{v,m}$ be independent (conditionally on the rows) centered Gaussian processes on S_d with covariance

$$Q_{v,m}(\rho, \rho') = (v_1 R(\rho, \rho') + v_2 \hat{X}_d(\rho, \rho'))^m.$$

These processes exist. Indeed, by the binomial theorem

$$Q_{v,m} = \sum_{i=0}^m \binom{m}{i} v_1^i v_2^{m-i} R^i \hat{X}_d^{m-i},$$

and $R(\rho, \rho')^i \hat{X}_d(\rho, \rho')^{m-i}$ is the inner product of the tensors $(\rho/\sqrt{d})^{\otimes i} \otimes \hat{\Psi}_d(\rho)^{\otimes(m-i)}$ and $(\rho'/\sqrt{d})^{\otimes i} \otimes \hat{\Psi}_d(\rho')^{\otimes(m-i)}$. Hence

$$\begin{aligned} h_{v,m}(\rho) &= \langle T_{v,m}, F_{v,m}(\rho) \rangle, \\ F_{v,m}(\rho) &= \left(\binom{m}{i}^{1/2} v_1^{i/2} v_2^{(m-i)/2} (\rho/\sqrt{d})^{\otimes i} \otimes \hat{\Psi}_d(\rho)^{\otimes(m-i)} \right)_{0 \leq i \leq m}, \end{aligned} \quad (4.16)$$

where $T_{v,m} = (T_{v,m,i})_{0 \leq i \leq m}$ is a family of independent standard Gaussian tensors of the matching shapes, has covariance $F_{v,m}(\rho) \cdot F_{v,m}(\rho') = Q_{v,m}(\rho, \rho')$. Every diagonal covariance satisfies $\|F_{v,m}(\rho)\|^2 = Q_{v,m}(\rho, \rho) \leq 1$, since $R(\rho, \rho) = 1$, $\hat{X}_d(\rho, \rho) \leq 1$ and $v_1 + v_2 \leq 1$. With $s_d = d^{1/3}$ and parameters $u_{v,m} \in [1, 2]$ we add

$$s_d \sum_{v,m} c_{v,m} u_{v,m} h_{v,m}(\rho)$$

to the Hamiltonian. The series converges almost surely, uniformly in ρ (see (4.41) below). The Gaussian part is modeled on Panchenko's perturbation for the multi-species Sherrington–Kirkpatrick model [100, (24)–(28)], which has weights of the same form 2^{-j-m} , parameters uniform on $[1, 2]$, and a strength N^γ with $1/4 < \gamma < 1/2$ (here $s_d = d^{1/3}$). In [100] the covariance is a power of a weighted sum of the overlaps within the species. Here it is a power of a weighted sum of the spin overlap and the overlap of the response vectors, which the cutoff ϖ keeps in the unit ball.

Parameters and the three systems. The parameters u_1, u_2 and $u_{v,m}$ are independent and uniform on $[1, 2]$, and they take the same values in the N -dimensional, the N' -dimensional, and the reference system. The first two systems use the construction above with $d = N$ and $d = N'$. The reference system uses the scalar part $\xi_{N'} \sum_{a \leq M} (u_1 \tau_1 + u_2 \tau_2)(S_a(\sigma))$, and as its Gaussian part it uses, by definition, the Gaussian part of the N' -dimensional system with the rows $a \leq M$, evaluated at $\rho(\sigma, 0) = (\sqrt{N'/N} \sigma, 0)$. This common restriction is what makes the comparison in Lemma 4.7(iv) possible. We have $R(\rho(\sigma, 0), \rho(\sigma', 0)) = R_N(\sigma, \sigma')$ and $x_a(\rho(\sigma, 0)) = \sqrt{N'/N} S_a(\sigma)$. Thus the response vector of the reference perturbation is built from $U'(\sqrt{N'/N} S_a)$ rather than $U'(S_a)$. Since $|U'(\sqrt{N'/N} S) - U'(S)| \leq L|S| n/N$, the Cauchy–Schwarz inequality shows that on Ω_N

$$\left| \hat{X}_{N'}(\rho(\sigma, 0), \rho(\sigma', 0)) - \frac{X_N(\sigma, \sigma')}{C_X} \right| \leq \frac{Cn}{N} \quad (\sigma, \sigma' \in S_N). \quad (4.17)$$

Finally, the reference system is tilted by $\exp(nB_N(\sigma)/2)$. We write $\langle \cdot \rangle_{\text{ref}}$ for the Gibbs measure of the perturbed and tilted reference system. The tilt will cancel between the two cavity terms in §4.4.

Invariance. The Gaussian tensors are invariant in law under orthogonal transformations of their spin slots, and $\Psi_d(O\rho)$ computed with the rows g_a equals $\Psi_d(\rho)$ computed with the rows $O^\top g_a$. Since the rows are standard Gaussian vectors, the perturbed N' -dimensional system is invariant in law under a simultaneous orthogonal transformation of the configurations and of the rows. In particular, for every orthogonal O of $\mathbb{R}^{N'}$ and every bounded measurable F ,

$$\mathbb{E}\langle F(O\rho^1, \dots, O\rho^r) \rangle_{N'} = \mathbb{E}\langle F(\rho^1, \dots, \rho^r) \rangle_{N'}. \quad (4.18)$$

Properties of the perturbation. For a system with Gibbs measure $\langle \cdot \rangle$ and a Gaussian component with covariance $Q = Q_{v,m}$, write $Q_{\ell\ell'} = Q(\rho^\ell, \rho^{\ell'})$. For a bounded measurable function f of the arrays $(R(\rho^\ell, \rho^{\ell'}), \hat{X}_d(\rho^\ell, \rho^{\ell'}))_{\ell, \ell' \leq r}$ of r replicas, the defect in the identity (4.13) is $|\mathfrak{D}_r(f, Q)|$, where $\mathfrak{D}_r$ is the discrepancy (4.10) of the pair $(R, \hat{X}_d)$ with $(v_1, v_2) = v$; its terms $Q(R_{\ell\ell'}^1, R_{\ell\ell'}^2)$ are then the values $Q_{\ell\ell'}$. For the reference system the arrays are those of the points $\rho(\sigma^\ell, 0)$.

Lemma 4.7. *Fix n and Λ . In the following statements $o(1)$ denotes a quantity that tends to zero as $N \rightarrow \infty$, uniformly in the perturbation parameters and in $M \in W_N$, except where a statement specifies otherwise: in (ii) the defects tend to zero after averaging over the parameters, and in (iii) uniformly on the sets $\mathcal{G}_N(M)$.*

- (i) *For the N -dimensional system with $M \sim \text{Poi}(\alpha N)$ rows, the perturbation changes the expected log partition function by $o(N)$.*
- (ii) *For the reference system let $A_i(\sigma) = N'^{-1} \sum_{a \leq M} \tau_i(S_a(\sigma))$, $i = 1, 2$. After averaging over the perturbation parameters, $\mathbb{E}\langle |A_i - \mathbb{E}\langle A_i \rangle_{\text{ref}}| \rangle_{\text{ref}} \rightarrow 0$. By (4.15), the observables b_N and $X_N(\sigma, \sigma)$ therefore concentrate under $\langle \cdot \rangle_{\text{ref}}$ about their expectations, and so does the diagonal entry $\hat{X}_{N'}(\rho(\sigma, 0), \rho(\sigma, 0))$ of its perturbation.*
- (iii) *There are sets $\mathcal{G}_N = \mathcal{G}_N(M)$ of perturbation parameters with $\inf_{M \in W_N} \mathbb{P}(\mathcal{G}_N(M)) \rightarrow 1$ such that*

$$\sup_{M \in W_N} \sup_{\mathcal{G}_N(M)} \sup_{\|f\|_\infty \leq 1} |\mathfrak{D}_r(f, Q_{v,m})| \rightarrow 0$$

for the reference system and every fixed r and (v, m) , and for the N' -dimensional system with the Gibbs measure $\langle \cdot \rangle_{N'}$ and every fixed r and (v, m) with $v_2 = 0$, and such that the concentration defects of (ii), without the average over the parameters, also tend to zero uniformly on $\mathcal{G}_N(M)$ and in $M \in W_N$. The sets $\mathcal{G}_N(M)$ do not depend on Λ .

- (iv) **Comparison with the reference perturbation.** (a) *In the partition function of the N' -dimensional system with M rows, restricted to $\varepsilon \in \mathbb{B}_\Lambda$, replacing the Gaussian part of the perturbation at $\rho(\sigma, \varepsilon)$ by its value at $\rho(\sigma, 0)$ changes the expected logarithm by $o(1)$, and changes the Gibbs expectations of bounded measurable functions of finitely many spin overlaps and cavity coordinates by $o(1)$ times their supremum norm.* (b) *In the representation of the N -dimensional system through the fields $S_a + \eta_a$ of (4.6), replacing its Gaussian perturbation*

by that of the reference system, and the strengths s_N, ξ_N by $s_{N'}, \xi_{N'}$, changes the expected log partition function by $o(1)$.

- (v) Rows. Adding the π further rows to the N' -dimensional system, including the corresponding change of its perturbation, increases the expected log partition function by at least

$$\mathbb{E} \log \left\langle \exp \left\{ \sum_{i \leq \pi} U(y_i(\rho)) \right\} \right\rangle_{N'} - o(1),$$

where $y_i(\rho) = \tilde{g}_i \cdot \rho / \sqrt{N'}$ are the fields of the added rows. Moreover, for $N \geq N_0(n)$ the difference between the expected log partition functions of the N' -dimensional system with $M + \pi$ rows and of the N -dimensional system with M rows is at least $-C_n$, uniformly in the perturbation parameters and in $M \in W_N$, and for every N and every number M of rows it is at most $C_n(N + M)$ in absolute value.

The second lemma provides continuity of the quantities in the cavity computation under convergence of overlap arrays. We apply it both to the Gibbs measures of §4.4 and to finite Ruelle probability cascades, so we state it for a random probability measure $\mathfrak{G}$ on a measurable space E . We write $\langle \cdot \rangle$ for the average over independent samples $\rho^1, \rho^2, \dots$ from $\mathfrak{G}$ (replicas) and $\mathbb{E} \langle \cdot \rangle$ for the expectation over $\mathfrak{G}$ and the replicas. Let R and X be random positive semidefinite kernels on E with $R(\rho, \rho) = 1$, let b be a random function on E , and let $z : E \rightarrow \mathbb{R}^n$ and $y, y_1, y_2, \dots : E \rightarrow \mathbb{R}$ be random fields, all defined on one probability space with $\mathfrak{G}$ and jointly measurable. For r replicas write $\mathbf{A}_r = (R(\rho^\ell, \rho^{\ell'}), X(\rho^\ell, \rho^{\ell'}), b(\rho^\ell))_{\ell, \ell' \leq r}$ for their arrays. For such an array $\mathbf{a}$ let $N_{\mathbf{a}}$ be the centered Gaussian law of vectors $(z^\ell, y^\ell, y_1^\ell, y_2^\ell, \dots)_{\ell \leq r}$ under which the families $z, y, y_1, y_2, \dots$ are independent and

$$\mathbb{E} z_i^\ell z_j^{\ell'} = \delta_{ij} X_{\ell\ell'}, \quad \mathbb{E} y^\ell y^{\ell'} = n R_{\ell\ell'} X_{\ell\ell'}, \quad \mathbb{E} y_i^\ell y_i^{\ell'} = R_{\ell\ell'};$$

these matrices are positive semidefinite by the Schur product theorem. We assume:

- (G) For every r and every bounded measurable function Φ of $\mathbf{A}_r$ and of the values at $\rho^1, \dots, \rho^r$ of z, y , and finitely many y_i , $\mathbb{E} \langle \Phi \rangle = \mathbb{E} \langle \bar{\Phi}(\mathbf{A}_r) \rangle$, where $\bar{\Phi}(\mathbf{a})$ is the integral of $\Phi(\mathbf{a}, \cdot)$ against $N_{\mathbf{a}}$.

This holds, for instance, if $E = S_d$ and, conditionally on $(\mathfrak{G}, R, X, b)$, the fields are independent centered Gaussian processes with covariances $\delta_{ij} X$, nRX , and R , and are independent of the replicas. This is the situation in §4.4. The *numerator*, or cavity partition function on $\mathbb{B}_\Lambda$, and the *denominator* are

$$\mathcal{N}_\Lambda = \left\langle \int_{\mathbb{B}_\Lambda} v_n(d\varepsilon) e^{\varepsilon \cdot z(\rho) - (\|\varepsilon\|^2 - n)(b(\rho) - 1)/2} \right\rangle, \quad \mathcal{D} = \langle e^{y(\rho)} \rangle,$$

and the *cavity measure* is the random probability measure on $E \times \mathbb{B}_\Lambda$ with density proportional to the integrand of $\mathcal{N}_\Lambda$ with respect to $\mathfrak{G} \otimes v_n$.

Lemma 4.8. *Consider a sequence of such random measures satisfying (G) for which, for every $c > 0$, the quantities*

$$\mathbb{E}\langle \exp(cX(\rho, \rho) + c|b(\rho)|) \rangle$$

are bounded along the sequence from some index on, and suppose that the finite-dimensional distributions of the arrays converge. Then the following quantities converge, and their limits depend only on the limiting law of the arrays:

- (i) $\mathbb{E} \log \mathcal{N}_\Lambda$ and $\mathbb{E} \log \mathcal{D}$;
- (ii) *expectations under the cavity measure of bounded continuous functions of the arrays and the cavity coordinates of finitely many replicas;*
- (iii) *the expected logarithmic increment $\mathbb{E} \log \langle \exp\{\sum_{i \leq \pi} U(y_i(\rho))\} \rangle$, for $\pi \sim \text{Poi}(\alpha n)$ independent of the other variables, whenever U is continuous and satisfies (4.2).*

In particular, if $\mathbb{E}\langle |b - \bar{b}_N| \rangle \rightarrow 0$ for constants $\bar{b}_N \rightarrow b_\infty$, then b may be replaced by the constant b_∞ in these limits.

Lemma 4.8 is the analogue for the perceptron of the continuity of the cavity functionals in the law of the overlap array, proved for mixed p -spin models in [99, Lemma 3] and for spherical models in [37]. The difference is that the functionals here also depend on the response overlap and on the precision b , and that part (iii) treats the row term, whose activation may be unbounded below.

The third lemma is an identity for the mean of b_N under the reference system, obtained by Gaussian integration by parts in the rows g_a ; it turns $S_a U'(S_a)$ into $U''(S_a) + U'(S_a)^2$ minus a two-replica term. In the limit it shows that the smallest precision of the Gaussian integral over the added coordinates is at least one (Step 3 of §4.7).

Lemma 4.9. *Fix n . For two replicas σ^1, σ^2 of the reference system, uniformly in the perturbation parameters and in $M \in W_N$,*

$$\mathbb{E}\langle b_N \rangle_{\text{ref}} = 1 + \mathbb{E}\langle X_N(\sigma, \sigma) \rangle_{\text{ref}} - \mathbb{E}\langle R_N(\sigma^1, \sigma^2) X_N(\sigma^1, \sigma^2) \rangle_{\text{ref}} + o(1).$$

Finite cascades as random measures. Lemma 4.8 is also applied to finite Ruelle probability cascades carrying prescribed diagonal values. Take a cascade as in §4.2, with depth k , weights (v_β) , and paired levels $0 \leq q_0 \leq \dots \leq q_k \leq 1$ and $0 \leq p_0 \leq \dots \leq p_k$, and let $\bar{p} \geq p_k$ and $\bar{b} \in \mathbb{R}$. Independently of (v_β) , let Y^p , Y^c , and $Y^{q,1}, Y^{q,2}, \dots$ be independent Gaussian cascade fields (§3.2), with values in $\mathbb{R}^n$, $\mathbb{R}$, and $\mathbb{R}$, and with levels p , $c = (nq_l p_l)_{l \leq k}$, and q , respectively; the levels c_l are nonnegative and nondecreasing. The *cascade realization* with diagonal values $(1, \bar{p}, \bar{b})$ is the random measure $\mathfrak{G} = \sum_\beta v_\beta \delta_\beta \otimes \nu_n \otimes \nu_1 \otimes \nu_1^{\otimes \mathbb{N}}$ on $E = \mathbb{N}_*^k \times \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}^{\mathbb{N}}$, with $b \equiv \bar{b}$, with kernels R and X equal to $q_{\beta \wedge \beta'}$ and $p_{\beta \wedge \beta'}$ at distinct points with leaves β, β' and to 1 and $\bar{p}$ at equal points, and with the fields

$$z(\rho) = Y_\beta^p + \sqrt{\bar{p} - p_k} x, \quad y(\rho) = Y_\beta^c + \sqrt{n(\bar{p} - q_k p_k)} x', \quad y_i(\rho) = Y_\beta^{q,i} + \sqrt{1 - q_k} x_i''$$

at $\rho = (\beta, x, x', x'')$. The within-state coordinates x, x', x'' carry the nonnegative differences between the diagonal values and the top levels. Two independent samples from $\mathfrak{G}$ are distinct points almost surely, so the arrays of the replicas are cascade arrays with paired levels (q_l, p_l) off the diagonal, with diagonal values 1, $\bar{p}$, and $\bar{b}$. Given the cascade weights and the leaves of the replicas, the values of the fields at the replicas are centered Gaussian with the covariances in (G), because the cascade fields are independent of the weights and the within-state coordinates are independent standard Gaussian. This gives (G), and the moment condition of Lemma 4.8 holds because $X(\rho, \rho) = \bar{p}$ and $b = \bar{b}$ are constant.

How the lemmas are combined with the overlap theorem. Fix $M \in W_N$ and parameters in the set $\mathcal{G}_N(M)$ of Lemma 4.7(iii) (in §4.7, $M = M_N$ depends on N), and pass to a subsequence along which the finite-dimensional distributions of the arrays (R_N, X_N, b_N) under $\mathbb{E}\langle \cdot \rangle_{\text{ref}}$ converge; they are tight by (4.8) and (4.9). The reference system satisfies (G) with $\mathfrak{G} = \langle \cdot \rangle_{\text{ref}}$, $R = R_N$, $X = X_N$, and $b = b_N$, for fields that are, conditionally on the disorder, independent Gaussian processes with the covariances in (G); for instance

$$\begin{aligned} z_i(\sigma) &= \sum_{a \leq M} \frac{U'(S_a(\sigma))}{\sqrt{N'}} g_{ai}, & y(\sigma) &= \sqrt{n} \sum_{a \leq M} \sum_{i \leq N} \frac{\sigma_i U'(S_a(\sigma))}{\sqrt{NN'}} g'_{ai}, \\ y_j(\sigma) &= \sum_{i \leq N} \frac{\sigma_i}{\sqrt{N}} g''_{ji}, \end{aligned}$$

with independent standard Gaussian variables $g_{ai}, g'_{ai}, g''_{ji}$, independent of the reference system. By (4.9), it satisfies the moment condition of Lemma 4.8. We apply Theorem 4.6 to the Gram pairs generated by $\sigma/\sqrt{N}$ and $\hat{\Psi}_{N'}(\rho(\sigma, 0))$. Their limiting arrays have $R^1 = R$ and $R^2 = X/C_X$, where (R, X) is the limit of (R_N, X_N) , for three reasons.

- (a) Both feature vectors lie in unit balls, and on Ω_N the response Gram array differs from X_N/C_X by $O(n/N)$, by (4.17); moreover $\mathbb{P}(\Omega_N^c) \rightarrow 0$.
- (b) The diagonals concentrate by Lemma 4.7(ii)–(iii), so the limiting diagonal values are 1 and p_∞/C_X , where

$$p_\infty = \lim \mathbb{E}\langle X_N(\sigma, \sigma) \rangle_{\text{ref}}$$

along a further subsequence.

- (c) Lemma 4.7(iii) gives vanishing discrepancies $\mathfrak{D}_r^N(f, Q)$ of (4.10), uniformly over bounded tests, for each fixed number of replicas and each $Q = Q_{v,m}$, that is, for $v_1 + v_2 \leq 1$. Homogeneity of degree m extends the conclusion to every nonnegative rational pair (v_1, v_2) , so (4.12) holds.

By Theorem 4.6(ii), after multiplying the second levels of its approximating cascades by C_X , the off-diagonal limit of (R_N, X_N) is a limit in finite-dimensional distribution of Ruelle probability cascade arrays with paired nondecreasing levels (q_l, p_l) . The observable b_N

concentrates about $\mathbb{E}\langle b_N \rangle_{\text{ref}}$, which converges along a further subsequence to a constant b_∞ ; by the last assertion of Lemma 4.8, b_N may be replaced by b_∞ in all limits.

In the limit, $0 \leq R_{12} \leq 1$ and $0 \leq X_{12} \leq p_\infty$ almost surely. Nonnegativity is Theorem 4.6(i), and $R_N \leq 1$. For the last bound, the Cauchy–Schwarz inequality gives $X_N(\sigma^1, \sigma^2) \leq (X_N(\sigma^1, \sigma^1) + X_N(\sigma^2, \sigma^2))/2$, so $\mathbb{E}\langle (X_N(\sigma^1, \sigma^2) - p_\infty)_+ \rangle_{\text{ref}} \leq \mathbb{E}\langle |X_N(\sigma, \sigma) - p_\infty| \rangle_{\text{ref}} \rightarrow 0$, and the left side converges to the corresponding expectation in the limit by uniform integrability. Projecting the levels of the approximating cascades onto $[0, 1] \times [0, p_\infty]$, which is monotone, continuous, and the identity on the values of the limiting arrays, we may assume $0 \leq q_l \leq 1$ and $0 \leq p_l \leq p_\infty$. The cascade realizations of these cascades with diagonal values $(1, p_\infty, b_\infty)$ then have arrays converging to the limiting arrays of (R_N, X_N, b_N) . We apply Lemma 4.8 to the sequence that alternates the reference systems along our subsequence (with b_N replaced by b_∞) and these cascade realizations, whose arrays converge to the same limit. We conclude that the limits along our subsequence of the quantities in Lemma 4.8(i)–(ii) for the reference system are the limits of their values on the cascade realizations.

For the spin overlaps of the N' -dimensional system we apply Theorem 4.6 to the Gram pairs generated by $\rho/\sqrt{N'}$ and the zero vector. Then every Q in its hypothesis is $(v_1 x)^m = v_1^m x^m$, so the hypothesis reduces to the defects with $v = (1, 0)$ of Lemma 4.7(iii). By parts (i)–(ii) of the theorem, the limiting spin overlap array is a limit in finite-dimensional distribution of Ruelle probability cascade arrays with nondecreasing levels in $[0, 1]$, and the overlap laws of these cascades converge to the limiting overlap law $\tilde{\zeta}$ of the N' -dimensional system.

4.4 The two cavity terms

We now compute the increment of the expected log partition function from dimension N to dimension N' at fixed perturbation parameters and a fixed number $M \in W_N$ of rows. Write $\log Z_{N'}^{M+\pi}$, $\log Z_{N'}^M$, and $\log Z_N^M$ for the log partition functions of the perturbed N' -dimensional system with all $M + \pi$ rows, of the same system with its first M rows only, and of the perturbed N -dimensional system. The increment splits as

$$\mathbb{E} \log Z_{N'}^{M+\pi} - \mathbb{E} \log Z_N^M = \underbrace{(\mathbb{E} \log Z_{N'}^{M+\pi} - \mathbb{E} \log Z_{N'}^M)}_{\text{row term}} + \underbrace{(\mathbb{E} \log Z_{N'}^M - \mathbb{E} \log Z_N^M)}_{\text{coordinate term}}. \quad (4.19)$$

The row term. By Lemma 4.7(v), the row term is at least $\mathbb{E} \log \langle \exp\{\sum_{i \leq \pi} U(y_i(\rho))\} \rangle_{N'} - o(1)$, where the fields y_i of the added rows are centered Gaussian with covariance $\rho \cdot \rho' / N'$ and are independent of $\langle \cdot \rangle_{N'}$. Thus $\langle \cdot \rangle_{N'}$, with $R(\rho, \rho') = \rho \cdot \rho' / N'$, $X \equiv 0$, $b \equiv 0$, and the fields y_i of the added rows, satisfies (G) and the moment condition of Lemma 4.8. By Lemma 4.8(iii), which applies by (4.2), the limit of this quantity along a subsequence along which the spin overlap array converges depends only on the limiting law of that

array. For parameters in $\mathcal{G}_N(M)$ this array is a limit of Ruelle probability cascade arrays with nondecreasing levels in $[0, 1]$ whose overlap laws ζ_h converge to the limiting overlap law $\tilde{\zeta}$ (§4.3). Consider the h -th of these cascades, with levels $0 \leq q_0 \leq \dots \leq q_k \leq 1$ (we drop the index h from its levels and masses), and its cascade realization with diagonal values $(1, 0, 0)$ and $p \equiv 0$. Given π , the fields $y_i(\rho) = Y_\beta^{q,i} + \sqrt{1 - q_k} x_i''$ of different rows are independent, so

$$\left\langle \exp \left\{ \sum_{i \leq \pi} U(y_i(\rho)) \right\} \right\rangle = \sum_{\beta} v_{\beta} \prod_{i \leq \pi} \mathbb{E} [e^{U(Y_\beta^{q,i} + \sqrt{1 - q_k} G)} \mid Y^{q,i}].$$

Apply Lemma 3.2 with the marks $(e_v^{(i)})_{i \leq \pi}$, where $e_v^{(i)}$ are the Gaussian increments of $Y^{q,i}$ at the node v (with variance q_0 at the root and $q_l - q_{l-1}$ at depth l), and with the leaf function $\sum_{i \leq \pi} \mathcal{T}_{1,1-q_k} U(Y_\beta^{q,i})$. The leaf function is bounded above and at least $-C \sum_i (1 + (Y_\beta^{q,i})^2)$ by (4.2) and (2.6), so all quantities in (3.2) are finite. Since the marks of different rows are independent and the leaf function is a sum over the rows, the recursion (3.2) factorizes over the rows. For one row, its steps are $\mathcal{T}_{\theta_l, q_l - q_{l-1}}$ at depth l , and the average over the root mark is $\mathcal{T}_{0, q_0}$. By (3.5) and the recursion (2.2), the expected logarithm of the display equals

$$\pi [\mathcal{T}_{0, q_0} \mathcal{T}_{\theta_1, q_1 - q_0} \cdots \mathcal{T}_{\theta_k, q_k - q_{k-1}} \mathcal{T}_{1, 1 - q_k} U](0) = \pi u_{\zeta_h}(0, 0; U),$$

where $\zeta_h = \sum_l w_l \delta_{q_l}$ is the overlap law of the cascade, a probability measure on $[0, 1]$ whose distribution function equals 0 on $[0, q_0)$ and θ_{l+1} on $[q_l, q_{l+1})$ for $0 \leq l \leq k$, with $q_{k+1} = 1$ (some of these intervals may be empty); its steps are exactly the operators in the display. Averaging over $\pi \sim \text{Poi}(\alpha n)$ gives $\alpha n u_{\zeta_h}(0, 0; U)$. As $h \rightarrow \infty$, these values converge, by Lemma 4.8(iii) applied to the sequence that alternates the N' -dimensional systems and the cascade realizations, to the limit of the lower bound for the row term. Moreover $u_{\zeta_h}(0, 0; U) \rightarrow u_{\tilde{\zeta}}(0, 0; U)$ by Lemma 2.4, since the laws ζ_h converge weakly on $[0, 1]$ and hence in the Wasserstein distance. Thus, along the subsequences used below,

$$\liminf_N (\text{row term}) \geq \alpha n u_{\tilde{\zeta}}(0, 0; U). \quad (4.20)$$

In §4.7 we show that $\tilde{\zeta}$ equals the limiting overlap law ζ of the reference system.

Reduction of the coordinate term to the reference system. By (4.3) and (4.5), $Z_{N'}^M$ is an integral over (σ, ε) against $\mu_N(d\sigma) f_{N,n}(\varepsilon) d\varepsilon$. Restricting ε to the ball $\mathbb{B}_\Lambda$ decreases it, so the coordinate term is bounded below by the same expression with $Z_{N'}^M$ replaced by its restriction to $\mathbb{B}_\Lambda$. By Lemma 4.7(iv)(a) we may replace the Gaussian part of the perturbation at $\rho(\sigma, \varepsilon)$ by its value at $\rho(\sigma, 0)$, at a cost $o(1)$. After this replacement the Gaussian part depends only on σ . We add and subtract the activation terms $\sum_a U^{\tilde{\zeta}}(S_a)$ at the reference fields, with $\tilde{\zeta} = \zeta_{N'}$, which include the scalar part, and factor out the

partition function of the tilted reference system. Similarly, by Lemma 4.7(iv)(b) the N -dimensional system may be written, at a cost $o(1)$, with the fields $S_a + \eta_a$ of (4.6), the reference perturbation, and the strength $\xi_{N'}$. Since the reference partition function is common to both terms, it cancels, and the coordinate term is at least

$$\begin{aligned} \mathbb{E} \log \left\langle \int_{\mathbb{B}_\Lambda} f_{N,n}(\varepsilon) e^{\sum_a [U^\xi(\chi(\varepsilon)S_a + w_a(\varepsilon)) - U^\xi(S_a)]} d\varepsilon e^{-nB_N/2} \right\rangle_{\text{ref}} \\ - \mathbb{E} \log \left\langle e^{\sum_a [U^\xi(S_a + \eta_a) - U^\xi(S_a)]} e^{-nB_N/2} \right\rangle_{\text{ref}} - o(1). \end{aligned} \quad (4.21)$$

The factors $e^{-nB_N/2}$ undo the tilt of $\langle \cdot \rangle_{\text{ref}}$; they appear in both terms. Here $S_a = S_a(\sigma)$, $B_N = B_N(\sigma)$, and $\langle \cdot \rangle_{\text{ref}}$ acts on σ .

Gaussian interpolation for the numerator. We replace the activation increment in the first term of (4.21) by a Gaussian field. Let $z(\sigma) = (z_1(\sigma), \dots, z_n(\sigma))$ be, conditionally on the disorder, a centered Gaussian process with $\mathbb{E} z_i(\sigma) z_j(\sigma') = \delta_{ij} X_N(\sigma, \sigma')$, independent of everything else (for instance the process z of §4.3). For $0 \leq t \leq 1$ put $S_a^t = \chi(\varepsilon)S_a + \sqrt{t} w_a(\varepsilon)$ and

$$\Delta_t = \sum_a [U^\xi(S_a^t) - U^\xi(S_a)] + \sqrt{1-t} \varepsilon \cdot z(\sigma) + (1-t) \frac{\|\varepsilon\|^2}{2} B_N(\sigma),$$

and let $\varphi(t)$ be the first term of (4.21) with the exponent replaced by Δ_t ; thus $\varphi(1)$ is that term. Write $\langle \cdot \rangle_t$ for the associated Gibbs measure on $S_N \times \mathbb{B}_\Lambda$. We first discuss the terms produced by the unperturbed activation U and then the terms that carry a factor ξ or ξ^2 .

The field $w_a(\varepsilon)$ is centered Gaussian with $\mathbb{E} w_a(\varepsilon) w_a(\varepsilon') = \varepsilon \cdot \varepsilon' / N'$, independently over a , and it is independent of the reference system. Gaussian integration by parts in the fresh columns g'_a produces the diagonal terms $\|\varepsilon\|^2 (U'' + U'^2)(S_a^t)$ (the U'' part from differentiating U' , the U'^2 part from differentiating the Gibbs density of the same replica) and a two-replica term. Integration by parts in z produces the U'^2 part and the two-replica term at the reference fields S_a , with the opposite sign, and the derivative of the explicit term $(1-t)\|\varepsilon\|^2 B_N/2$ supplies the U'' part at the reference fields, also with the opposite sign. The result, for the activation U , is

$$\begin{aligned} \varphi'(t) &= \frac{1}{2N'} \sum_a \mathbb{E} \langle \|\varepsilon\|^2 [(U'' + U'^2)(S_a^t) - (U'' + U'^2)(S_a)] \rangle_t \\ &\quad - \frac{1}{2N'} \sum_a \mathbb{E} \langle (\varepsilon^1 \cdot \varepsilon^2) [U'(S_a^{t,1}) U'(S_a^{t,2}) - U'(S_a^1) U'(S_a^2)] \rangle_t, \end{aligned} \quad (4.22)$$

where the superscripts 1,2 refer to two replicas $(\sigma^1, \varepsilon^1)$, $(\sigma^2, \varepsilon^2)$. The first line collects the diagonal contributions and the second the two-replica covariance contributions.

We bound (4.22) on Ω_N . Uniformly on $S_N \times \mathbb{B}_\Lambda$,

$$\sum_a S_a^2 \leq CN, \quad \sum_a |S_a^t - S_a|^2 \leq C_{n,\Lambda}.$$

The second bound follows from $S_a^t - S_a = (\chi(\varepsilon) - 1)S_a + \sqrt{t}w_a(\varepsilon)$, from $\chi(\varepsilon) - 1 = O_{n,\Lambda}(N^{-1})$, and from $\sum_a w_a(\varepsilon)^2 \leq \|\mathbf{G}'\|_{\text{op}}^2 \|\varepsilon\|^2 / N' \leq C\|\varepsilon\|^2$. Since $|U'|$ grows at most linearly and U'', U''' are bounded, the function $\omega_U = U'' + U'^2$ satisfies $|\omega_U(x) - \omega_U(x')| \leq C(1 + |x| + |x - x'|)|x - x'|$, and the Cauchy–Schwarz inequality gives

$$\begin{aligned} \frac{1}{N'} \sum_a |\omega_U(S_a^t) - \omega_U(S_a)| &\leq \frac{C}{N'} \left(\sum_a (1 + |S_a| + |S_a^t - S_a|)^2 \right)^{1/2} \left(\sum_a |S_a^t - S_a|^2 \right)^{1/2} \\ &\leq C_{n,\Lambda} N^{-1/2}. \end{aligned}$$

The product difference in the second line obeys the same estimate, after writing it as $[U'(S_a^{t,1}) - U'(S_a^1)]U'(S_a^{t,2}) + U'(S_a^1)[U'(S_a^{t,2}) - U'(S_a^2)]$. Since $\|\varepsilon\|^2 \leq \Lambda n$ and $|\varepsilon^1 \cdot \varepsilon^2| \leq \Lambda n$ on $\mathbb{B}_\Lambda$, the right side of (4.22) is $O_{n,\Lambda}(N^{-1/2})$ on Ω_N . On Ω_N^c the integrand is bounded by a polynomial in the operator norms, whose moments are bounded, so its contribution is at most $C_{n,\Lambda} \mathbb{P}(\Omega_N^c)^{1/2}$.

The terms carrying ξ or ξ^2 arise because the interpolation uses the activation U^ξ of (4.14) while B_N and X_N are defined with U . In (4.22) they replace $\omega_U(S_a^t)$ by $\omega_{U^\xi}(S_a^t)$ and $U'(S_a^{t,1})U'(S_a^{t,2})$ by $(U^\xi)'(S_a^{t,1})(U^\xi)'(S_a^{t,2})$. By (4.1) the first and second derivatives of τ_1 and τ_2 grow at most linearly, and the first derivatives of U as well; hence $|\omega_{U^\xi}(x) - \omega_U(x)| \leq C(\xi + \xi^2)(1 + x^2)$ and $|(U^\xi)'(x)(U^\xi)'(x') - U'(x)U'(x')| \leq C(\xi + \xi^2)(1 + x^2 + x'^2)$. Since $N'^{-1} \sum_a (1 + (S_a^t)^2) \leq C$ on Ω_N , these terms contribute $O_{n,\Lambda}(\xi_{N'} + \xi_{N'}^2) = o(1)$ on Ω_N ; this is sufficient, although it is weaker than the rate $N^{-1/2}$ available for the activation U .

We conclude that $\varphi(1) - \varphi(0) = o(1)$ for fixed n and Λ . The same holds for Gibbs expectations: differentiating $\mathbb{E}\langle F \rangle_t$ for a bounded measurable function F of the spin overlaps and cavity coordinates of r replicas, which does not depend on the fresh Gaussian variables, inserts at most $r + 2$ replicas and the same covariance differences, so $|\partial_t \mathbb{E}\langle F \rangle_t| \leq C_{r,n,\Lambda} \|F\|_\infty (N^{-1/2} + \xi_{N'} + \xi_{N'}^2)$ on Ω_N , up to a contribution $C_{r,n,\Lambda} \|F\|_\infty \mathbb{P}(\Omega_N^c)^{1/2}$.

The numerator at $t = 0$. By Taylor's formula and (4.4), uniformly on $\mathbb{B}_\Lambda$ and on Ω_N ,

$$\begin{aligned} \sum_a [U(\chi(\varepsilon)S_a) - U(S_a)] &= (\chi(\varepsilon) - 1) \sum_a S_a U'(S_a) + O\left((\chi(\varepsilon) - 1)^2 \sum_a S_a^2\right) \\ &= \frac{n - \|\varepsilon\|^2}{2} A_N + O_{n,\Lambda}(N^{-1}), \end{aligned}$$

and the corresponding terms carrying ξ are $O_{n,\Lambda}(\xi_{N'})$. Adding $\|\varepsilon\|^2 B_N / 2$ and using $b_N - 1 = A_N - B_N$, the exponent Δ_0 equals

$$\varepsilon \cdot z(\sigma) - \frac{\|\varepsilon\|^2 - n}{2} (b_N - 1) + \frac{n}{2} B_N + o(1).$$

The term $nB_N/2$ cancels the factor $e^{-nB_N/2}$ in (4.21), and $f_{N,n}(\varepsilon) d\varepsilon$ may be replaced by $\nu_n(d\varepsilon)$ by (4.4). This explains the definitions of A_N , B_N , and b_N .

Gaussian interpolation for the denominator. The second term of (4.21) is treated by the same interpolation. The field $\eta_a(\sigma) = c_N g_a'' \cdot \sigma$ has covariance $c_N^2 \sigma \cdot \sigma' = (n/N') R_N(\sigma, \sigma')$, and on Ω_N we have $\sum_a \eta_a^2 \leq c_N^2 \|\mathbf{G}''\|_{\text{op}}^2 N \leq C_n$. Let $y(\sigma)$ be, conditionally on the disorder, a centered Gaussian process with covariance $n R_N(\sigma, \sigma') X_N(\sigma, \sigma')$, and interpolate with

$$\sum_a [U^\xi(S_a + \sqrt{t} \eta_a) - U^\xi(S_a)] + \sqrt{1-t} y(\sigma) + (1-t) \frac{n}{2} B_N(\sigma).$$

Integration by parts gives (4.22) with $\|\varepsilon\|^2$ replaced by n and $\varepsilon^1 \cdot \varepsilon^2$ replaced by $n R_N(\sigma^1, \sigma^2)$, and the same estimates show that the two endpoints differ by $o(1)$. At $t = 0$ the activation increment is replaced by $y(\sigma) + n B_N(\sigma)/2$, and $n B_N/2$ again cancels the factor $e^{-n B_N/2}$.

The coordinate term. Combining the preceding paragraphs, for fixed n and Λ the coordinate term is at least

$$\mathbb{E} \log \left\langle \int_{\mathbb{B}_\Lambda} \nu_n(d\varepsilon) e^{\varepsilon \cdot z(\sigma) - (\|\varepsilon\|^2 - n)(b_N(\sigma) - 1)/2} \right\rangle_{\text{ref}} - \mathbb{E} \log \langle e^{y(\sigma)} \rangle_{\text{ref}} - o(1). \quad (4.23)$$

The same comparison holds for Gibbs expectations of bounded continuous functions of finitely many replicas, including their cavity coordinates: under the Gibbs measure of the N' -dimensional system with M rows, conditioned on $\varepsilon \in \mathbb{B}_\Lambda$ for each replica, such expectations differ by $o(1)$ from those under the cavity measure of the first term of (4.23), where the spin overlaps are $R_N(\sigma^\ell, \sigma^{\ell'})$. Indeed, the chain of comparisons above consists of Lemma 4.7(iv)(a), the interpolation, whose effect on Gibbs expectations was bounded above, and the replacement of an exponent and a density by quantities that differ from them by $o(1)$ uniformly on $\Omega_N \times \mathbb{B}_\Lambda$. Moreover, on $\mathbb{B}_\Lambda$ the spin overlap $\rho^\ell \cdot \rho^{\ell'} / N' = [\chi(\varepsilon^\ell) \chi(\varepsilon^{\ell'}) n R_N(\sigma^\ell, \sigma^{\ell'}) + \varepsilon^\ell \cdot \varepsilon^{\ell'}] / N'$ of the N' -dimensional system differs from $R_N(\sigma^\ell, \sigma^{\ell'})$ by $O_{n,\Lambda}(N^{-1})$, and a continuous function is uniformly continuous on the compact set of overlaps in $[-1, 1]$ and cavity coordinates in $\mathbb{B}_\Lambda$.

For parameters in $\mathcal{G}_N(M)$ and along the subsequences of §4.3, Lemma 4.8 allows us to replace b_N by its limit b_∞ . Thus, for fixed n and Λ , the coordinate term is bounded below, up to $o(1)$, by

$$\mathbb{E} \log \left\langle \int_{\mathbb{B}_\Lambda} \nu_n(d\varepsilon) e^{\varepsilon \cdot z(\sigma) - (\|\varepsilon\|^2 - n)(b_\infty - 1)/2} \right\rangle_{\text{ref}} - \mathbb{E} \log \langle e^{y(\sigma)} \rangle_{\text{ref}}, \quad (4.24)$$

where, conditionally on the reference system, z and y are centered Gaussian with

$$\mathbb{E} z_i(\sigma) z_j(\sigma') = \delta_{ij} X_N(\sigma, \sigma'), \quad \mathbb{E} y(\sigma) y(\sigma') = n R_N(\sigma, \sigma') X_N(\sigma, \sigma').$$

The first term of (4.24) is the expected logarithm of the numerator $\mathcal{N}_\Lambda$ of Lemma 4.8 for the reference system with $b = b_\infty$, and the second is minus that of the denominator $\mathcal{D}$; we call the two expected logarithms the numerator and the denominator. By Lemma 4.8(i), both converge along the subsequence, and by §4.3 their limits are the limits of their values on cascade realizations. We compute these values next.

4.5 The cavity quantities on cascades

We evaluate the numerator and the denominator of (4.24) on a cascade realization, using Lemma 3.4, and we show that the cutoff Λ can be removed when the smallest precision is bounded below. Throughout this section we use the notation of §3.2: a Ruelle probability cascade of depth k with parameters θ_l and masses w_l , the branching level $J = \beta^1 \wedge \beta^2$ of two replicas, and, for levels $p \in \mathcal{K}$ and a precision b , the precisions d_l , the mean levels m_l , the function $\mathcal{G}(b, p)$, the mean squared norm $\varrho = m_k + 1/b$, and the tilted measure $\mathcal{M}$ defined in (3.8) and below it.

Fix a cascade with paired levels $0 \leq q_0 \leq \dots \leq q_k \leq 1$ and $0 \leq p_0 \leq \dots \leq p_k$, and its cascade realization with diagonal values $(1, \bar{p}, \bar{b})$ (§4.3). The overlap law of the cascade is $\sum_l w_l \delta_{q_l}$, and the law of the response overlap is $\sum_l w_l \delta_{p_l}$. On the realization, the response field z is the cascade field Y_β^p plus the within-state variable $\sqrt{\bar{p} - p_k} x$. Averaging x inside the Gibbs average multiplies the integrand of the numerator by $e^{(\bar{p} - p_k)\|\varepsilon\|^2/2}$, which changes the coefficient of $-\|\varepsilon\|^2/2$ from $\bar{b} - 1$ to $b - 1$, where

$$b = \bar{b} - (\bar{p} - p_k). \quad (4.25)$$

By (3.7), the smallest precision $d_0 = b - \sum_{l \geq 1} \theta_l \Delta p_l$ of (3.8) is expressed through the mean of the response overlap:

$$d_0 = b - p_k + \sum_{l=0}^k w_l p_l. \quad (4.26)$$

When $d_0 > 0$, all precisions are positive and Lemma 3.4 applies.

The cavity quantities on a cascade realization. Averaging the within-state coordinates inside the Gibbs average gives the following values. The numerator of (4.24), with b_∞ replaced by $\bar{b}$, equals $\Theta_\Lambda + n(\bar{b} - 1)/2$, where

$$\Theta_\Lambda = \mathbb{E} \log \sum_{\beta} v_{\beta} \int_{\mathbb{B}_\Lambda} v_n(\mathrm{d}\varepsilon) e^{\varepsilon \cdot Y_\beta^p - (b-1)\|\varepsilon\|^2/2},$$

and we write Θ for the same quantity with ε integrated over $\mathbb{R}^n$. When $d_0 > 0$, (3.10) gives $\Theta = n[\mathcal{G}(b, p) - \frac{1}{2} \log b]$. The denominator equals

$$\mathbb{E} \log \sum_{\beta} v_{\beta} e^{Y_\beta^c + n(\bar{p} - q_k p_k)/2} = \frac{n}{2} \left(\bar{p} - \sum_{l=0}^k w_l q_l p_l \right),$$

by (3.18) with the levels $c_l = n q_l p_l$; here $\bar{p} \geq p_k \geq q_k p_k$. Hence, when $d_0 > 0$, the unrestricted numerator minus the denominator, divided by n , is

$$\mathcal{G}(b, p) - \frac{1}{2} \log b + \frac{1}{2} \left(\bar{b} - 1 - \bar{p} + \sum_{l=0}^k w_l q_l p_l \right). \quad (4.27)$$

The last term vanishes in the limit by Lemma 4.9 (Step 3 of §4.7).

The cutoff is removed with the following lemma. It concerns the unrestricted tilted measure $\mathcal{M}$ of §3.2 with the field $Y = Y^p$, defined when $d_0 > 0$, and the random probability measure $\mathcal{M}_\Lambda$ on pairs (β, ε) with density proportional to $v_\beta v_n(d\varepsilon) e^{\varepsilon \cdot Y_\beta - (b-1)\|\varepsilon\|^2/2} \mathbf{1}_{\mathbb{B}_\Lambda}(\varepsilon)$, which is defined for every b ; when $d_0 > 0$, $\mathcal{M}_\Lambda$ is $\mathcal{M}$ conditioned on $\{\varepsilon \in \mathbb{B}_\Lambda\}$.

Lemma 4.10. *Fix the cavity dimension n and constants $P_0, B_0, c_0 > 0$, and consider finite cascades with*

$$0 \leq p_k \leq P_0, \quad d_0 \geq c_0, \quad b \leq B_0.$$

There are constants $C, c' > 0$, depending only on n, P_0, B_0, c_0 , such that for all $\Lambda \geq 1$

$$0 \leq \Theta - \Theta_\Lambda \leq Ce^{-c'\Lambda}, \quad \mathbb{P}_{\mathcal{M}}(\varepsilon^1 \notin \mathbb{B}_\Lambda) \leq Ce^{-c'\Lambda}. \quad (4.28)$$

Moreover $\varrho = m_k + 1/b \leq P_0/c_0^2 + 1/c_0$.

Proof. Every denominator $d_{l-1}d_l$ and d_0^2 in (3.8) is at least c_0^2 , and the increments of p sum to $p_k \leq P_0$, so $m_k \leq P_0/c_0^2$; also $1/b \leq 1/d_0 \leq 1/c_0$. By (3.12), one replica ε^1 has the law $N(0, \varrho I_n)$ under $\mathbb{P}_{\mathcal{M}}$, and the Gaussian tail bound for $\|\varepsilon^1\|^2$ gives the second estimate in (4.28), for Λ larger than a constant; enlarging C covers the remaining $\Lambda \geq 1$.

A small averaged probability of $\{\varepsilon \notin \mathbb{B}_\Lambda\}$ alone would not bound $\Theta - \Theta_\Lambda$, so we use convexity. Define the leaf functions

$$\begin{aligned} f(y) &= \log \int v_n(d\varepsilon) e^{\varepsilon \cdot y - (b-1)\|\varepsilon\|^2/2}, \\ f_\Lambda(y) &= \log \int_{\mathbb{B}_\Lambda} v_n(d\varepsilon) e^{\varepsilon \cdot y - (b-1)\|\varepsilon\|^2/2} \quad (y \in \mathbb{R}^n). \end{aligned}$$

Then $f_\Lambda \leq f$, so $\Theta_\Lambda \leq \Theta$. The map $\varphi \mapsto \mathbb{E} \log \sum_\beta v_\beta e^{\varphi(Y_\beta)}$ is convex on leaf functions, and for a bounded direction ψ its derivative at f is $\mathbb{E}_{\mathcal{M}} \psi(Y_{\beta^1})$, by dominated convergence. Applying convexity with the bounded direction $\psi_t = \max\{f_\Lambda, f - t\} - f \in [-t, 0]$ gives

$$\Theta - \mathbb{E} \log \sum_\beta v_\beta e^{\max\{f_\Lambda, f-t\}(Y_\beta)} \leq \mathbb{E}_{\mathcal{M}} [f(Y_{\beta^1}) - f_\Lambda(Y_{\beta^1})],$$

and letting $t \rightarrow \infty$, by monotone convergence, $\Theta - \Theta_\Lambda \leq \mathbb{E}_{\mathcal{M}} [f(Y_{\beta^1}) - f_\Lambda(Y_{\beta^1})]$. Now $\exp(f_\Lambda - f)(y)$ is the probability that a vector with law $N(y/b, b^{-1}I_n)$ lies in $\mathbb{B}_\Lambda$, and by Lemma 3.4(ii), $Y_{\beta^1}/b = M_k \sim N(0, m_k I_n)$ under $\mathbb{P}_{\mathcal{M}}$. Hence

$$\Theta - \Theta_\Lambda \leq \mathbb{E} \left[-\log \mathbb{P}\{M_k + b^{-1/2}G' \in \mathbb{B}_\Lambda \mid M_k\} \right],$$

where G' is a standard Gaussian vector in $\mathbb{R}^n$ independent of $M_k \sim N(0, m_k I_n)$.

We split the last expectation according to the size of M_k . On $\{\|M_k\| \leq \sqrt{\Lambda n}/2\}$, the conditional probability of leaving $\mathbb{B}_\Lambda$ is at most $\mathbb{P}(\|G'\|^2 > b\Lambda n/4) \leq \mathbb{P}(\|G'\|^2 > c_0\Lambda n/4) \leq Ce^{-c'\Lambda}$, and for Λ larger than a constant the negative logarithm is at most twice this bound, since $-\log(1-x) \leq 2x$ for $0 \leq x \leq 1/2$. For all M_k and $\Lambda \geq 1$, integrating the Gaussian density of $M_k + b^{-1/2}G'$ over the fixed ball $\mathbb{B}_1$ and using $\|x - M_k\|^2 \leq 2\|x\|^2 + 2\|M_k\|^2$ gives

$$\mathbb{P}\{M_k + b^{-1/2}G' \in \mathbb{B}_\Lambda \mid M_k\} \geq c_2 \exp[-C_2(1 + \|M_k\|^2)],$$

with c_2, C_2 depending only on n, c_0, B_0 , since $c_0 \leq b \leq B_0$. The contribution of $\{\|M_k\| > \sqrt{\Lambda n}/2\}$ is therefore bounded by a Gaussian tail expectation of $\log(1/c_2) + C_2(1 + \|M_k\|^2)$, which is also at most $Ce^{-c'\Lambda}$ uniformly for $m_k \leq P_0/c_0^2$. Enlarging C covers the remaining $\Lambda \geq 1$. $\square$

The measure $\mathcal{M}_\Lambda$ is the law of the leaves and cavity coordinates under the cavity measure of the cascade realization on $\mathbb{B}_\Lambda$, after the within-state coordinates are integrated out. For r replicas, $\mathcal{M}_\Lambda$ is $\mathcal{M}$ conditioned on the event that $\varepsilon^\ell \in \mathbb{B}_\Lambda$ for $\ell \leq r$, realization by realization. If a probability measure gives mass $1 - \delta$ to an event, then conditioning each of r replicas on this event changes the expectation of a bounded function F of the r replicas by at most $2r\|F\|_\infty\delta$, because the conditioning event for the r replicas has mass at least $1 - r\delta$. Since this bound is linear in δ , it survives averaging over the disorder:

$$|\mathbb{E}_{\mathcal{M}_\Lambda} F - \mathbb{E}_{\mathcal{M}} F| \leq 2r\|F\|_\infty \mathbb{P}_{\mathcal{M}}(\varepsilon^1 \notin \mathbb{B}_\Lambda). \quad (4.29)$$

4.6 Rotation invariance

The two identities that determine the entropy come from the following lemma. It uses only the invariance (4.18) of the N' -dimensional system under simultaneous rotations of the configurations and the rows. In the coordinates $\rho = \rho(\sigma, \varepsilon)$, the quantity R^{cav} below is the overlap $\varepsilon^1 \cdot \varepsilon^2/n$ of the added coordinates.

Lemma 4.11. *Suppose that the disorder-averaged law of a random Gibbs measure on $S_{N'}$ is invariant under simultaneous orthogonal transformations of its replicas. For two replicas ρ^1, ρ^2 define*

$$R = \frac{\rho^1 \cdot \rho^2}{N'}, \quad R^{\text{cav}} = \frac{1}{n} \sum_{i=N+1}^{N'} \rho_i^1 \rho_i^2.$$

Then

$$\mathbb{E}\langle (R^{\text{cav}} - R)^2 \rangle \leq \frac{3N'}{n(N' - 1)}. \quad (4.30)$$

Moreover, the averaged one-replica law is uniform on $S_{N'}$, and the law of its last n coordinates converges to ν_n as $N \rightarrow \infty$.

Proof. By the invariance, we may replace two given replicas by their images under an independent Haar-distributed rotation without changing the averaged law. Condition on two unit vectors $\rho^1/\sqrt{N'}$ and $\rho^2/\sqrt{N'}$ with inner product R . Their images can be written as x and $Rx + \sqrt{1-R^2}x^\perp$, where x is uniform on the unit sphere of $\mathbb{R}^{N'}$ and, given x , $x^\perp$ is uniform on the unit sphere of the orthogonal complement of x . For the n coordinate indices $i \in \{N+1, \dots, N'\}$ put $A = \sum_i x_i^2$ and $B = \sum_i x_i x_i^\perp$, so that $R^{\text{cav}} = (N'/n)(RA + \sqrt{1-R^2}B)$. The variable A has the beta distribution with parameters $n/2$ and $N/2$, so

$$\mathbb{E}A = \frac{n}{N'}, \quad \text{Var}(A) = \frac{2nN}{N'^2(N'+2)} \leq \frac{2n}{N'(N'-1)}.$$

Given x , $\mathbb{E}[x_i^\perp x_j^\perp \mid x] = (\delta_{ij} - x_i x_j)/(N' - 1)$, hence $\mathbb{E}[B \mid x] = 0$ and $\mathbb{E}[B^2 \mid x] = (A - A^2)/(N' - 1)$. Therefore

$$\mathbb{E}B^2 \leq \frac{n}{N'(N'-1)}, \quad \mathbb{E}[(A - n/N')B] = 0.$$

Consequently

$$\mathbb{E}[(RA + \sqrt{1-R^2}B - Rn/N')^2] = R^2 \text{Var}(A) + (1-R^2)\mathbb{E}B^2 \leq \frac{3n}{N'(N'-1)},$$

and multiplying by $(N'/n)^2$ gives (4.30). The averaged one-replica law is a rotation-invariant probability measure on $S_{N'}$, hence uniform. By (4.3) its last n coordinates have the density $f_{N,n}$, which converges pointwise to the standard Gaussian density; by Scheffé's lemma the laws converge to ν_n in total variation. $\square$

4.7 Proof of the lower bound

The order of the limits matters, because the Gaussian integral over all of $\mathbb{R}^n$ in the numerator is not a bounded function of the overlaps. For fixed n we first let $N \rightarrow \infty$ along a subsequence at a fixed cutoff Λ , then let the index h of the approximating cascades tend to infinity, then let $\Lambda \rightarrow \infty$; the cavity dimension n is sent to infinity last.

Proof of Theorem 4.1. We divide the proof into eight steps.

STEP 1. Increments. Let Ξ_N be the expected log partition function of the perturbed N -dimensional system with $M \sim \text{Poi}(\alpha N)$ rows, averaged also over the perturbation parameters. By the Poissonization estimate of §4.1 and Lemma 4.7(i),

$$\frac{\Xi_N}{N} = F_N(U) + o(1). \quad (4.31)$$

For every fixed n ,

$$\liminf_{N \rightarrow \infty} \frac{\Xi_N}{N} \geq \frac{1}{n} \liminf_{N \rightarrow \infty} (\Xi_{N+n} - \Xi_N). \quad (4.32)$$

To verify this, let ℓ' be smaller than n times the right side. Then $\Xi_{N+n} - \Xi_N \geq \ell'$ for all $N \geq N_0$. Every $N \geq N_0 + n$ can be written as $N = N_1 + jn$ with $N_0 \leq N_1 < N_0 + n$ and $j \geq 1$, and telescoping along the residue class of N_1 gives $\Xi_N \geq \min_{N_0 \leq N_1 < N_0 + n} \Xi_{N_1} + j\ell'$. Since $j/N \rightarrow 1/n$, the left side of (4.32) is at least ℓ'/n . This is the inequality on which the Aizenman–Sims–Starr scheme rests; see [100, (53)].

Let $\ell_n = \liminf_N (\Xi_{N+n} - \Xi_N)$. Then $\ell_n < \infty$, since Ξ_N/N is bounded above by $\alpha U_{\max} + o(1)$ and (4.32) holds; and $\ell_n \geq -C_n$ by Step 2.

STEP 2. Choice of parameters and row count. For fixed perturbation parameters and a number M of rows, let $\Delta_N(M)$ be the increment on the left of (4.19), computed conditionally on the number of rows; thus $\Xi_{N+n} - \Xi_N$ is the average of $\Delta_N(M)$ over the parameters and over $M \sim \text{Poi}(\alpha N)$, because $M + \pi \sim \text{Poi}(\alpha N')$. By Lemma 4.7(v), $\Delta_N(M) \geq -C_n$ uniformly in the parameters and in $M \in W_N$, and $|\Delta_N(M)| \leq C_n(N + M)$ for every M . Since $\mathbb{P}(M \notin W_N) \leq 2e^{-cN^{1/3}}$ and $\mathbb{E}(N + M)^2 \leq CN^2$, the Cauchy–Schwarz inequality shows that the event $\{M \notin W_N\}$ contributes $o(1)$ to the average; in particular $\ell_n \geq -C_n$. Choose a subsequence along which $\Xi_{N+n} - \Xi_N \rightarrow \ell_n$. Let $\mathcal{G}'_N$ be the event that $M \in W_N$ and that the parameters lie in the set $\mathcal{G}_N(M)$ of Lemma 4.7(iii). Its probability is at least $\mathbb{P}(M \in W_N) \inf_{M \in W_N} \mathbb{P}(\mathcal{G}_N(M)) \rightarrow 1$. Since $\Delta_N \geq -C_n$ on $\{M \in W_N\}$, the average of $\Delta_N(M)$ over $\mathcal{G}'_N$ is at most

$$\frac{\Xi_{N+n} - \Xi_N + C_n \mathbb{P}(\mathcal{G}'_N)^c + o(1)}{\mathbb{P}(\mathcal{G}'_N)} = \ell_n + o(1)$$

along the subsequence. Hence we can choose $M_N \in W_N$ and parameters in $\mathcal{G}_N(M_N)$ whose increment satisfies $\Delta_N(M_N) \leq \ell_n + o(1)$. We fix these parameters and this row count from now on.

STEP 3. One limit of the overlap arrays, and positive precision. Along a further subsequence, the finite-dimensional distributions of the arrays (R_N, X_N, b_N) under $\mathbb{E}\langle \cdot \rangle_{\text{ref}}$ converge, $\mathbb{E}\langle b_N \rangle_{\text{ref}} \rightarrow b_\infty$, $\mathbb{E}\langle X_N(\sigma, \sigma) \rangle_{\text{ref}} \rightarrow p_\infty$, and the spin overlap array of the N' -dimensional system under $\mathbb{E}\langle \cdot \rangle_{N'}$ converges. Write $\mathbb{L}$ for the limiting law of the arrays of the reference system, with diagonal values $(1, p_\infty, b_\infty)$, and $\mathbb{E}_{\mathbb{L}}$ for expectations under it; let ζ be the law of R_{12} under $\mathbb{L}$, and $\tilde{\zeta}$ the limiting overlap law of the N' -dimensional system. These limits do not depend on Λ . The error terms $o(1)$ of §4.4 depend on Λ , but for every fixed Λ they tend to zero along the whole sequence.

By §4.3, there are finite cascades, indexed by $h \geq 1$, with depths k_h and paired levels $0 \leq q_l^h \leq 1$ and $0 \leq p_l^h \leq p_\infty$, whose cascade realizations with diagonal values $(1, p_\infty, b_\infty)$ have arrays converging to $\mathbb{L}$. We attach the index h to the quantities

of §4.5 for the h -th cascade, and write $\mathbb{E}_{L_h}$ for expectations under its arrays; the law of (R_{12}, X_{12}) under it is $\sum_l w_l^h \delta_{(q_l^h, p_l^h)}$, and its precision (4.25) is $b_h = b_\infty - p_\infty + p_{k_h}^h$. The product $R_N(\sigma^1, \sigma^2) X_N(\sigma^1, \sigma^2)$ has bounded second moments by (4.8) and (4.9), so $\mathbb{E}\langle R_N(\sigma^1, \sigma^2) X_N(\sigma^1, \sigma^2) \rangle_{\text{ref}} \rightarrow \mathbb{E}_L[R_{12} X_{12}]$, and Lemma 4.9 gives

$$b_\infty = 1 + p_\infty - \mathbb{E}_L[R_{12} X_{12}]. \quad (4.33)$$

This has two consequences. First, by (4.26) the smallest precision of the h -th cascade is $d_0^h = b_\infty - p_\infty + \mathbb{E}_{L_h}[X_{12}]$, and as $h \rightarrow \infty$

$$d_0^h \longrightarrow b_\infty - p_\infty + \mathbb{E}_L[X_{12}] = 1 + \mathbb{E}_L[X_{12}(1 - R_{12})] \geq 1,$$

because $X_{12} \geq 0$ and $R_{12} \leq 1$ under L (§4.3); the convergence holds because X_{12} takes values in $[0, p_\infty]$ under all these laws. Also $b_h = 1 + p_{k_h}^h - \mathbb{E}_L[R_{12} X_{12}] \leq 1 + p_\infty$, since $R_{12} X_{12} \geq 0$ under L . So there is h_0 such that, for $h \geq h_0$, the tilted measure $\mathcal{M}^h$ of §3.2 is defined, Lemma 3.4 applies, and Lemma 4.10 applies with $c_0 = 1/2$ and $P_0 = B_0 = 1 + p_\infty$, with constants C, c' independent of h . Second, by (4.33) the last term of (4.27) tends to zero:

$$b_\infty - 1 - p_\infty + \sum_l w_l^h q_l^h p_l^h = \mathbb{E}_{L_h}[R_{12} X_{12}] - \mathbb{E}_L[R_{12} X_{12}] \longrightarrow 0. \quad (4.34)$$

STEP 4. The increment and the cutoff. Fix $\Lambda \geq 1$. By the choice of parameters in Step 2, (4.19), (4.20) and (4.24), ℓ_n is at least $\alpha n u_{\tilde{\zeta}}(0, 0; U)$ plus the limit of the numerator minus the limit of the denominator. By §4.3, these limits are the limits as $h \rightarrow \infty$ of the values $\Theta_\Lambda^h + n(b_\infty - 1)/2$ and $\frac{n}{2}(p_\infty - \sum_l w_l^h q_l^h p_l^h)$ on the cascade realizations (§4.5). By (4.34) their difference is $\Theta_\Lambda^h + o_h(1)$, so $\lim_h \Theta_\Lambda^h$ exists and $\ell_n \geq \alpha n u_{\tilde{\zeta}}(0, 0; U) + \lim_h \Theta_\Lambda^h$. For $h \geq h_0$, (4.28) and (3.10) give $\Theta_\Lambda^h \geq n[\mathcal{G}(b_h, p^h) - \frac{1}{2} \log b_h] - C e^{-c' \Lambda}$. Letting $\Lambda \rightarrow \infty$,

$$\frac{\ell_n}{n} \geq \alpha u_{\tilde{\zeta}}(0, 0; U) + \limsup_{h \rightarrow \infty} \left[\mathcal{G}(b_h, p^h) - \frac{1}{2} \log b_h \right]. \quad (4.35)$$

STEP 5. The rotation identities on the cascades. Let $q_h = m_{k_h}^h + 1/b_h$, the mean of $\|\varepsilon^1\|^2/n$ under $\mathcal{M}^h$ by (3.12). We show that $\tilde{\zeta} = \zeta$ and

$$q_h \rightarrow 1, \quad \limsup_{h \rightarrow \infty} \mathbb{E}_{\mathcal{M}^h}[(\varepsilon^1 \cdot \varepsilon^2/n - q_j^h)^2] \leq \frac{3}{n}. \quad (4.36)$$

All three follow from the fact that, for every bounded continuous function F of the spin overlaps and the cavity coordinates of r replicas,

$$\lim_{h \rightarrow \infty} \mathbb{E}_{\mathcal{M}^h} F = \lim_{N' \rightarrow \infty} \mathbb{E}\langle F \rangle_{N'}, \quad (4.37)$$

where both limits exist, the second along our subsequence. Here the spin overlap of two distinct replicas is $\rho^\ell \cdot \rho^{\ell'}/N'$ for the N' -dimensional system and $q_{\beta^\ell \wedge \beta^{\ell'}}^h$ for $\mathcal{M}^h$, and the cavity coordinates are the last n coordinates of ρ^ℓ and the vectors ε^ℓ , respectively.

To prove (4.37), fix Λ . We condition the Gibbs measure $\langle \cdot \rangle_{N'}$ on $\{\varepsilon \in \mathbb{B}_\Lambda\}$ for each replica. As for (4.29), this changes $\mathbb{E}\langle F \rangle_{N'}$ by at most $2r\|F\|_\infty \mathbb{E}\langle \mathbf{1}_{\{\varepsilon \notin \mathbb{B}_\Lambda\}} \rangle_{N'}$, and by Lemma 4.11, whose hypothesis holds by (4.18), this expected missing mass tends to $\nu_n(\mathbb{B}_\Lambda^c)$. By the comparison statement after (4.23), the conditioned expectation differs by $o(1)$ from the expectation of F under the cavity measure of the reference system, computed with the spin overlaps R_N . By the last assertion of Lemma 4.8, by Lemma 4.8(ii) and by §4.3, the latter converges to the limit as $h \rightarrow \infty$ of the expectation of F under the cavity measures of the cascade realizations on $\mathbb{B}_\Lambda$, which is $\lim_h \mathbb{E}_{\mathcal{M}_\Lambda^h} F$ (§4.5). For $h \geq h_0$, (4.29) and (4.28) give $|\mathbb{E}_{\mathcal{M}_\Lambda^h} F - \mathbb{E}_{\mathcal{M}^h} F| \leq 2r\|F\|_\infty Ce^{-c'\Lambda}$. Hence the upper and lower limits of $\mathbb{E}_{\mathcal{M}^h} F$ and of $\mathbb{E}\langle F \rangle_{N'}$ all lie within $2r\|F\|_\infty(\nu_n(\mathbb{B}_\Lambda^c) + Ce^{-c'\Lambda})$ of $\lim_h \mathbb{E}_{\mathcal{M}_\Lambda^h} F$. Since Λ is arbitrary, the four limits coincide, which proves (4.37). We apply it three times.

- (a) $F = f(R_{12})$ for a continuous function f on $[-1, 1]$. The right side of (4.37) is $\int f d\tilde{\zeta}$. Under $\mathcal{M}^h$ the level J has law (w_l^h) by (3.13), so the left side is $\lim_h \mathbb{E}_{L_h} f(R_{12}) = \int f d\tilde{\zeta}$. Hence $\tilde{\zeta} = \zeta$.
- (b) $F = \min\{\|\varepsilon\|^2/n, T\}$ for $T > 0$. By Lemma 4.11, the averaged one-replica law of the last n coordinates of the N' -dimensional system converges to ν_n , so the right side is $\mathbb{E} \min\{\|G'\|^2/n, T\}$ for a standard Gaussian vector G' . For $h \geq h_0$, $\varepsilon^1 \sim N(0, \varrho_h I_n)$ under $\mathcal{M}^h$ with $\varrho_h \leq \bar{\varrho} = P_0/c_0^2 + 1/c_0$ (Lemma 4.10), and $\mathbb{E}[(\|G'\|^2/n)^2] = 1 + 2/n \leq 3$. So removing the truncation changes the two sides by at most $3\bar{\varrho}^2/T$ and $3/T$. Letting $T \rightarrow \infty$ gives $\varrho_h \rightarrow 1$.
- (c) $F = \min\{(\varepsilon^1 \cdot \varepsilon^2/n - R_{12})^2, T\}$. In the N' -dimensional system $\varepsilon^1 \cdot \varepsilon^2/n$ is the quantity R^{cav} of Lemma 4.11, so $\mathbb{E}\langle F \rangle_{N'} \leq 3N'/(n(N' - 1))$, and the left side of (4.37) is at most $3/n$. For $h \geq h_0$ the fourth moment of $\varepsilon^1 \cdot \varepsilon^2/n - q_J^h$ under $\mathcal{M}^h$ is bounded uniformly in h , since $0 \leq q_J^h \leq 1$ and $\varepsilon^1, \varepsilon^2$ have the law $N(0, \varrho_h I_n)$ with $\varrho_h \leq \bar{\varrho}$. So removing the truncation costs $O(1/T)$ uniformly in h . This gives (4.36).

In particular, the row term in (4.35) is $\alpha u_\zeta(0, 0; U)$.

STEP 6. *The entropy.* For $h \geq h_0$, Lemma 3.4(iv) gives

$$\mathcal{G}(b_h, p^h) - \frac{1}{2} \log b_h = \text{CS}(\gamma_h) + \frac{1}{2} \log \varrho_h, \quad \gamma_h = \gamma_{m^h/\varrho_h} \in \mathcal{U}, \quad (4.38)$$

where γ_h is the order parameter with atoms of mass w_l^h at m_l^h/ϱ_h . The trial order parameters γ_h are close to ζ . The law $\zeta_h = \sum_l w_l^h \delta_{q_l^h}$ of R_{12} under the h -th cascade array converges weakly to ζ , hence in the Wasserstein distance W_1 on $[0, 1]$. Under $\mathbb{P}_{\mathcal{M}^h}$ the level J has law (w_l^h) , so $(q_J^h, m_J^h/\varrho_h)$ is a coupling of ζ_h and γ_h , and since $0 \leq m_J^h \leq m_{k_h}^h \leq \varrho_h$,

$$W_1(\zeta_h, \gamma_h) \leq \mathbb{E}|q_J^h - m_J^h| + |\varrho_h - 1| \leq (\mathbb{E}(q_J^h - m_J^h)^2)^{1/2} + |\varrho_h - 1|,$$

with expectations under $\mathbb{P}_{\mathcal{M}^h}$. By (3.13), m_J^h is the conditional mean of $\varepsilon^1 \cdot \varepsilon^2/n$ given J , and q_J^h is a function of J . So conditional Jensen's inequality and (4.36) give $\limsup_h \mathbb{E}(q_J^h - m_J^h)^2 \leq 3/n$. Hence

$$\limsup_{h \rightarrow \infty} W_1(\zeta, \gamma_h) \leq \sqrt{3/n}. \quad (4.39)$$

STEP 7. *The lower bound for fixed n .* Let $\delta > 0$. By (4.35), $\tilde{\zeta} = \zeta$, (4.38), $q_h \rightarrow 1$, and (4.39), there is h such that $\gamma = \gamma_h \in \mathcal{U}$ satisfies

$$\alpha u_\zeta(0, 0; U) + \text{CS}(\gamma) \leq \frac{\ell_n}{n} + \delta, \quad W_1(\zeta, \gamma) \leq \sqrt{3/n} + \delta.$$

Lemma 2.4 gives a constant C_U with $|u_\zeta(0, 0; U) - u_\gamma(0, 0; U)| \leq C_U W_1(\zeta, \gamma)$, so

$$\mathcal{P}_U(\gamma) = \alpha u_\gamma(0, 0; U) + \text{CS}(\gamma) \leq \frac{\ell_n}{n} + \alpha C_U \sqrt{3/n} + (1 + \alpha C_U) \delta.$$

Since $\gamma \in \mathcal{U}$, the left side is at least $\mathcal{P}_*(U)$. Letting $\delta \rightarrow 0$ and using (4.31) and (4.32),

$$\liminf_{N \rightarrow \infty} F_N(U) \geq \frac{\ell_n}{n} \geq \mathcal{P}_*(U) - \alpha C_U \sqrt{3/n}. \quad (4.40)$$

STEP 8. *The limit $n \rightarrow \infty$.* The constant C_U does not depend on n . Letting $n \rightarrow \infty$ in (4.40) proves Theorem 4.1. The perturbation and the Poissonization were already removed in (4.31). Neither a minimizer of $\mathcal{P}_U$ nor the upper bound of Chapter 3 was used. $\square$

4.8 Proofs of the perturbation, continuity, and precision lemmas

We keep the notation of §4.3. Throughout, $d \in \{N, N'\}$, $\mathcal{A}$ is the set of rows of the system under consideration, $\mathbf{G}_d$ is its matrix of rows, and constants may depend on U , α , n , and Λ , but not on N . Since $M \in W_N$, the number of rows satisfies $\#\mathcal{A} \leq 2\alpha N$, except in part (i), where M is Poisson, and in part (v), where the π added rows are included. All bounds on the log partition function hold uniformly for parameters $u_i, u_{v,m}$ in $[1/2, 5/2]$, which allows difference quotients at the endpoints of $[1, 2]$. We use the Gaussian Poincaré inequality [21, Theorem 3.20]: for a standard Gaussian vector $\mathbf{g}$ in a Euclidean space and a continuously differentiable function Φ with $\mathbb{E}\|\nabla\Phi(\mathbf{g})\|^2 < \infty$, $\text{Var}\Phi(\mathbf{g}) \leq \mathbb{E}\|\nabla\Phi(\mathbf{g})\|^2$.

We first record an elementary estimate for replacing an activation by a smaller one on fresh rows. It is used in part (v) of Lemma 4.7 and in part (iii) of Lemma 4.8.

Lemma 4.12. *Let $\langle \cdot \rangle$ be a random probability measure, and let $V_1 \leq V_2$ and $Y \geq 0$ be random functions on its space such that $V_2 \leq A$ for a constant A , $V_1 \geq -Y$, and $\mathbb{E}[\langle Y \rangle^2] < \infty$. Then, for every $T > 0$,*

$$0 \leq \mathbb{E} \log \langle e^{V_2} \rangle - \mathbb{E} \log \langle e^{V_1} \rangle \leq e^T \mathbb{E} \langle e^{V_2} - e^{V_1} \rangle + \frac{\mathbb{E}[(A_+ + \langle Y \rangle)^2]}{T}.$$

Proof. Put $Z_i = \langle e^{V_i} \rangle$ and $\Delta = \log Z_2 - \log Z_1 \geq 0$. By Jensen's inequality $\log Z_1 \geq \langle V_1 \rangle \geq -\langle Y \rangle$, and $\log Z_2 \leq A$; so both logarithms are integrable and $\Delta \leq A_+ + \langle Y \rangle$. Since $\log x \leq x - 1$, also $\Delta \leq (Z_2 - Z_1)/Z_1$. On $\{Z_1 \geq e^{-T}\}$ this gives $\Delta \leq e^T(Z_2 - Z_1)$. On $\{Z_1 < e^{-T}\}$ we have $\langle Y \rangle \geq -\log Z_1 > T$, so $\Delta \leq A_+ + \langle Y \rangle \leq (A_+ + \langle Y \rangle)^2/T$. Taking expectations proves the claim. $\square$

In both applications V_1 and V_2 are sums of activations over π fresh rows y_i , and each $y_i(\rho)$ is standard Gaussian for one replica, so that $\mathbb{E}((1 + y_i^2)^2) = 6$. Recall the constant C_0 of (4.2).

Proof of Lemma 4.7. We divide the proof into eight steps.

STEP 1. Tensor estimates. We first record a bound that controls both the countable sum of Gaussian components and the derivatives of the Gaussian perturbation in the rows. If T is a standard Gaussian multilinear form with m slots, each of dimension at most $d + \#\mathcal{A}$, its injective norm $\|T\|_{\text{inj}}$ (the supremum of $|T(x_1, \dots, x_m)|$ over unit vectors) satisfies

$$\mathbb{E}\|T\|_{\text{inj}}^2 \leq Cm(d + \#\mathcal{A}) \log(12m). \quad (4.41)$$

Indeed, a $1/(4m)$ -net of the unit ball of $\mathbb{R}^D$ can be chosen with at most $(12m)^D$ points, so the product net over the m slots has logarithmic cardinality at most $m(d + \#\mathcal{A}) \log(12m)$. Changing one argument at a time from a point of the unit balls to a nearest net point changes T by at most $\|T\|_{\text{inj}}/(4m)$ per slot, so $\|T\|_{\text{inj}}$ is at most $4/3$ times the maximum over the net. The values of T at the net points are centered Gaussian with variance at most one, and $\mathbb{E} \max_{i \leq K} g_i^2 \leq 4 \log K + 2$ for K such variables g_i , by $\mathbb{E} e^{g_i^2/4} \leq \sqrt{2}$. This proves (4.41). Differentiating a tensor contraction in its response argument $y = \hat{\Psi}_d(\rho)$ frees one slot and sums over at most m positions; the other arguments are contracted with vectors of norm at most one, and the coefficients in (4.16) are at most one because $\sum_i \binom{m}{i} v_1^i v_2^{m-i} \leq 1$. Hence $\sup_\rho \|\partial_y h_{v,m}(\rho)\| \leq m \sum_{i \leq m} \|T_{v,m,i}\|_{\text{inj}}$, and by (4.41), with no attempt to optimize the power of m ,

$$\mathbb{E} \sup_\rho \|\partial_y h_{v,m}(\rho)\|^2 \leq C(m+1)^5 \log(12(m+1))(d + \#\mathcal{A}). \quad (4.42)$$

These polynomial factors are summable against $c_{v,m} = 2^{-j(v)-m}$. In the same way $\sup_\rho |h_{v,m}(\rho)| \leq \sum_{i \leq m} \|T_{v,m,i}\|_{\text{inj}}$, so $(\mathbb{E} \sup_\rho |h_{v,m}(\rho)|^2)^{1/2} \leq C(m+1)^{3/2} \log(12(m+1))^{1/2} (d + \#\mathcal{A})^{1/2}$ by the Cauchy-Schwarz inequality and (4.41). Hence $\sum_{v,m} c_{v,m} (\mathbb{E} \sup_\rho |h_{v,m}|^2)^{1/2} < \infty$. By Tonelli's theorem the series defining the Gaussian part converges absolutely almost surely, uniformly in ρ , and by Minkowski's inequality the supremum over ρ of its tail after finitely many components tends to zero in L^2 . The same summable majorants, with the polynomial factors of (4.42), apply to the

derivatives of the finite truncations used below. Every statement below is first proved with finitely many components and then passes to the limit.

STEP 2. *Part (i).* Since $|\tau_i(x)| \leq C_\tau(1 + x^2)$, the scalar part changes $\log Z$ by at most $C\tilde{\xi}_d \sum_a (1 + x_a(\rho)^2) \leq C\tilde{\xi}_d (\#\mathcal{A} + \|\mathbf{G}_d\|_{\text{op}}^2)$, whose expectation is at most $C\tilde{\xi}_d N$ by Lemma 4.2 (averaged over M). For the Gaussian part, Jensen's inequality, conditionally on everything except the tensors, shows that its expected effect lies between 0 and $s_d^2 \sup_\rho \sum_{v,m} c_{v,m}^2 u_{v,m}^2 Q_{v,m}(\rho, \rho)/2 \leq Cs_d^2$: the lower bound because the Gibbs average of the Gaussian part under the unperturbed measure has mean zero, the upper bound because $\mathbb{E}e^h = e^{\text{Var}(h)/2}$ for a centered Gaussian variable h . The total expected cost is therefore $O(d^{3/4} + d^{2/3}) = o(d)$.

STEP 3. *Concentration of the free energy.* For the reference and the N' -dimensional system with $\mathcal{A} = \{1, \dots, M\}$, we show that

$$\text{Var}(\log Z) \leq Cd, \quad v_d := \sup \mathbb{E}|\log Z - \mathbb{E} \log Z| \leq C\sqrt{d}, \quad (4.43)$$

where the supremum is over the parameters. Apply the Gaussian Poincaré inequality, with finitely many tensors, in the rows and the tensor entries. Its right side is the expected squared gradient, which has three parts. The gradient in the rows through the activation contributes at most $C(M + \|\mathbf{G}_d\|_{\text{op}}^2)$, because $\sum_i \rho_i^2 = d$ and the first derivative of U^ξ grows at most linearly. The gradient in the tensor entries contributes at most Cs_d^2 , because each $h_{v,m}$ is linear in its tensor with coefficient vector of norm at most one. For the gradient in the rows through the response vector, write $y = \hat{\Psi}_d(\rho)$; then

$$\partial_{g_{ai}} h_{v,m} = [D\omega(\Psi_d)^\top \partial_y h_{v,m}]_a U''(x_a(\rho)) \frac{\rho_i}{d\sqrt{C_X}}.$$

Summing the squares over a and i and using $\|\rho\|^2 = d$, (4.42) bounds the expected contribution by $Cs_d^2(1 + M/d)$. For the reference system, the tilt $nB_N/2$ adds a row gradient of squared norm at most $C_n M/d^2$. Since $\mathbb{E}\|\mathbf{G}_d\|_{\text{op}}^2 \leq C(M + d)$ by Lemma 4.2 and $M \leq 2\alpha N$, the variance is at most $C(M + d) + Cs_d^2(1 + M/d) \leq Cd$. The bound passes to infinitely many components by Step 1, which proves (4.43).

STEP 4. *Part (ii).* Steps 4 and 5 follow the proof of the Ghirlanda–Guerra identities in [98, Theorem 3.2]; see also [100, Theorem 2]. Consider a scalar parameter $u = u_i$. In a d -dimensional system it enters the Hamiltonian as $d\tilde{\xi}_d u A(\rho)$ with $A = d^{-1} \sum_a \tau_i(x_a(\rho))$; for the reference system $d = N'$, $x_a = S_a$, and $A = A_i$. The function $u \mapsto \log Z$ is convex, with first derivative $d\tilde{\xi}_d \langle A \rangle$ and second derivative $(d\tilde{\xi}_d)^2 \langle (A - \langle A \rangle)^2 \rangle$. Integrating the second derivative over $[1, 2]$ and using that $\mathbb{E}\langle |A| \rangle$ is bounded by (4.8),

$$\int_1^2 \mathbb{E} \langle (A - \langle A \rangle)^2 \rangle du \leq \frac{C}{d\tilde{\xi}_d}.$$

For the fluctuation of $\langle A \rangle$ with the disorder we use the following consequence of convexity: if f, g are convex and differentiable, then for $t > 0$

$$|f'(u) - g'(u)| \leq g'(u+t) - g'(u-t) + \frac{|f-g|(u+t) + |f-g|(u-t) + 2|f-g|(u)}{t}.$$

(Bound $f'(u)$ above by $(f(u+t) - f(u))/t$ and below by $(f(u) - f(u-t))/t$, compare with g , and use that g' is nondecreasing.) Apply it to $f = \log Z$ and $g = \mathbb{E} \log Z$ as functions of u and integrate over $u \in [1, 2]$. For $0 < t \leq 1/2$, the boundary integrals give

$$\int_1^2 (g'(u+t) - g'(u-t)) du = \int_{2-t}^{2+t} g'(u) du - \int_{1-t}^{1+t} g'(u) du \leq 4t \sup |g'| \leq Ctd\xi_d,$$

where the supremum is over $[1/2, 5/2]$. Together with (4.43) this gives $\int_1^2 \mathbb{E} |\langle A \rangle - \mathbb{E} \langle A \rangle| du \leq Ct + 4v_d/(td\xi_d)$, and the choice $t = (v_d/(d\xi_d))^{1/2}$, which is at most $1/2$ for large N , gives

$$\int_1^2 \mathbb{E} |\langle A \rangle - \mathbb{E} \langle A \rangle| du \leq C\sqrt{v_d/(d\xi_d)} = O(d^{-1/8}).$$

Together with the thermal bound, $\int_1^2 \mathbb{E} \langle |A - \mathbb{E} \langle A \rangle| \rangle du \rightarrow 0$, uniformly in the other parameters and in $M \in W_N$. For the reference system, (4.15) gives $A_2 = X_N(\sigma, \sigma)$ and $A_1 = (N/N')(b_N - 1) - (n/N')B_N$, and $(n/N')B_N = O(n/N)$ by (4.8), so $X_N(\sigma, \sigma)$ and b_N concentrate. By (4.17), the diagonal entry of the reference perturbation differs from $X_N(\sigma, \sigma)/C_X$ by $O(n/N)$ on Ω_N , and both are bounded by polynomial moments off Ω_N ; so it concentrates as well.

STEP 5. Part (iii). Fix one Gaussian component, write $\lambda = s_d c_{v,m} u_{v,m}$, $h = h_{v,m}$ and $Q = Q_{v,m}$. Gaussian integration by parts in the entries of its tensor, conditionally on the rows (so that the dependence of Q on the rows produces no extra derivative), gives $\mathbb{E} \langle fh(\rho^1) \rangle = \lambda \mathbb{E} \langle f(\sum_{\ell \leq r} Q_{1\ell} - rQ_{1,r+1}) \rangle$ for a test function f of the arrays of r replicas, which does not depend on the tensors, and $\mathbb{E} \langle h \rangle = \lambda(\mathbb{E} \langle Q_{11} \rangle - \mathbb{E} \langle Q_{12} \rangle)$. Subtracting $\mathbb{E} \langle f \rangle$ times the second identity from the first and rearranging, we obtain for $\|f\|_\infty \leq 1$

$$r|\mathfrak{D}_r(f, Q)| \leq \mathbb{E} \langle |Q_{11} - \mathbb{E} \langle Q_{11} \rangle| \rangle + \lambda^{-1} \mathbb{E} \langle |h - \mathbb{E} \langle h \rangle| \rangle.$$

The function $u_{v,m} \mapsto \log Z$ is convex, with derivative $s_d c_{v,m} \langle h \rangle$ and second derivative $(s_d c_{v,m})^2 \langle (h - \langle h \rangle)^2 \rangle$. As in Step 4, since $|\mathbb{E} \langle h \rangle| \leq 2\lambda$, the thermal part of the second term, integrated over $u_{v,m} \in [1, 2]$, is at most $C/(s_d c_{v,m})$, and the difference quotient inequality bounds the disorder part by $C\sqrt{v_d}/(s_d c_{v,m})$. For each fixed component the resulting bound $C(1 + \sqrt{v_d})/(s_d c_{v,m})$ tends to zero, since $d^{1/4}/s_d = d^{-1/12}$. For the reference system, the diagonal term tends to zero after averaging over the parameters, because $Q_{11} = (v_1 + v_2 \tilde{X}_{N'}(\rho(\sigma, 0), \rho(\sigma, 0)))^m$ is an m -Lipschitz function of this diagonal entry on $[0, 1]$, and the entry concentrates by Step 4. For the N' -dimensional system we only need

$v_2 = 0$; then $Q_{11} = v_1^m$ is constant and the diagonal term vanishes. The bound does not depend on f , so the supremum over $\|f\|_\infty \leq 1$ may be taken before averaging over the parameters.

We now construct the sets $\mathcal{G}_N(M)$. As in [99, Lemma 1], we combine the defects in a weighted sum; there the parameters are also chosen so that the cavity increment is controlled, as in Step 2 of §4.7. Enumerate the countably many defects under consideration: the suprema over $\|f\|_\infty \leq 1$ of the defects $|\mathfrak{D}_r(f, Q_{v,m})|$ of the two systems for all r and (v, m) (with $v_2 = 0$ for the N' -dimensional system), and the concentration defects $\mathbb{E}\langle |A_i - \mathbb{E}\langle A_i \rangle| \rangle$ at fixed parameters. Each defect is a measurable function of the parameters. For fixed disorder the Hamiltonian is continuous in the parameters, because the series defining the Gaussian part converges uniformly by Step 1, so Gibbs averages of jointly measurable functions are jointly measurable in the disorder and the parameters. By Tonelli's theorem $\mathbb{E}\langle g \rangle$ is then measurable in the parameters for every nonnegative jointly measurable g . This covers the concentration defects, since $|A_i - \mathbb{E}\langle A_i \rangle|$ is jointly measurable in the disorder, the configuration, and the parameters. For fixed parameters, $f \mapsto \mathfrak{D}_r(f, Q_{v,m})$ is the integral of f against a finite signed measure on the compact space of arrays of r replicas, so the supremum over $\|f\|_\infty \leq 1$ is the total variation norm of this measure. It equals the supremum of $\mathfrak{D}_r(\mathbf{1}_A - \mathbf{1}_{A^c}, Q_{v,m})$ over the sets A of a countable algebra that generates the Borel sets of this space, and is therefore measurable. Each defect d_i is bounded by a constant $\bar{c}_i$ independent of N and M (the Ghirlanda–Guerra defects by 2, the concentration defects by (4.8)), and its average over the parameters tends to zero uniformly in $M \in W_N$, by the bounds above, which depend on M only through $M \leq 2\alpha N$. Choose weights $w_i > 0$ with $\sum_i w_i \bar{c}_i < \infty$, and let $\epsilon_N(M)$ be the average over the parameters of $\sum_i w_i d_i$. By dominated convergence, applied to the weighted sum of the suprema over $M \in W_N$ of the averaged defects, $\bar{\epsilon}_N = \sup_{M \in W_N} \epsilon_N(M) \rightarrow 0$. By Markov's inequality, the set $\mathcal{G}_N(M)$ where $\sum_i w_i d_i \leq \bar{\epsilon}_N^{1/2}$ has probability at least $1 - \bar{\epsilon}_N^{1/2}$, and on $\mathcal{G}_N(M)$ each fixed defect d_i is at most $\bar{\epsilon}_N^{1/2}/w_i$. None of these defects involves Λ , so neither does $\mathcal{G}_N(M)$. This proves (iii).

STEP 6. *Part (iv)(a).* We first retain a finite set $\mathcal{B}$ of components $\beta = (v, m)$; no constant below depends on $\mathcal{B}$. Write $s = s_{N'}$, $\lambda_\beta = c_{v,m} u_{v,m}$, and $h_\beta = h_{v,m}$, and let C_ω be a common bound for the operator norm of $D\omega(z)$ and the norm of the bilinear map $D^2\omega(z)$, uniform in z and in the dimension. Use the path $\rho^t = \rho(\sigma, t\varepsilon)$, $0 \leq t \leq 1$, on $S_{N'}$. Keep the activation part of the Hamiltonian at $\rho(\sigma, \varepsilon)$ fixed, and evaluate the Gaussian part $P_t(\sigma, \varepsilon) = s \sum_\beta \lambda_\beta h_\beta(\rho^t)$ at ρ^t . Let Z_t be the partition function restricted to $\varepsilon \in \mathbb{B}_\Lambda$ and $\langle \cdot \rangle_t$ its Gibbs measure, and assume $N \geq 2\Lambda n$. For a replica $(\sigma^p, \varepsilon^p)$ put $\rho_p = \rho(\sigma^p, t\varepsilon^p)$, $r_p = \rho_p / \sqrt{N'}$, $\zeta^p = \varepsilon^p / \sqrt{N'}$, and

$$x_{pa} = x_a(\rho_p), \quad \bar{x}_{pa} = x_a(\rho(\sigma^p, \varepsilon^p)), \quad \Psi_p = \Psi_{N'}(\rho_p), \quad y_p = \omega(\Psi_p), \quad J_p = D\omega(\Psi_p).$$

The fields $\bar{x}_{pa}$ enter the activation part, and $\sum_j |\zeta_j^p \zeta_j^q| \leq \Lambda n / N'$. A dot denotes ∂_t .

The derivative. By (4.16), $F_\beta(\rho) \cdot F_\beta(\rho')$ is the polynomial $(v_1 r \cdot r' + v_2 y \cdot y')^m$ in the spin arguments $r = \rho/\sqrt{N'}$, r' and the response arguments y, y' . Differentiating in the first pair of arguments,

$$D_{pq}^\beta := \partial_t F_\beta(\rho_p) \cdot F_\beta(\rho_q) = m K_{pq}^{m-1} (v_1 \dot{r}_p \cdot r_q + v_2 \dot{y}_p \cdot y_q), \quad K_{pq} = v_1 r_p \cdot r_q + v_2 y_p \cdot y_q, \quad (4.44)$$

and $|K_{pq}| \leq 1$. Since $\partial_t \log Z_t = \langle \partial_t P_t \rangle_t$, Gaussian integration by parts in the tensor entries, conditionally on the rows, gives

$$\frac{d}{dt} \mathbb{E} \log Z_t = s^2 \sum_\beta \lambda_\beta^2 \mathbb{E} \langle D_{11}^\beta - D_{12}^\beta \rangle_t.$$

Let f be a bounded measurable function of the configurations of r replicas. The derivative of $\mathbb{E} \langle f \rangle_t$ is $\mathbb{E} \langle f \sum_{p \leq r+1} a_p \partial_t P_t(\sigma^p, \varepsilon^p) \rangle_t$, with $a_p = 1$ for $p \leq r$ and $a_{r+1} = -r$. Integration by parts in the tensor entries, now with $r+1$ replicas in the Gibbs average, gives

$$\frac{d}{dt} \mathbb{E} \langle f \rangle_t = s^2 \sum_\beta \lambda_\beta^2 \sum_{p \leq r+1} \sum_{q \leq r+2} a_p b_q \mathbb{E} \langle f D_{pq}^\beta \rangle_t,$$

with $b_q = 1$ for $q \leq r+1$ and $b_{r+2} = -(r+1)$, so that $\sum_p |a_p| = 2r$ and $\sum_q |b_q| = 2(r+1)$. With finitely many components, $\partial_t P_t$ and all derivatives in the rows used below are bounded by polynomials in $\|[\mathbf{G}, \mathbf{G}']\|_{\text{op}}$, where $[\mathbf{G}, \mathbf{G}']$ is the $M \times N'$ matrix with rows (g_a, g'_a) , and the injective norms of the tensors, which have moments of all orders. This justifies the differentiation under $\mathbb{E}$ and the integrations by parts. Both derivatives are therefore controlled by the following bound: if f is a bounded measurable function of the configurations of k replicas, which depends neither on t nor on the disorder, and $p, q \leq k$, then

$$|\mathbb{E} \langle f D_{pq}^\beta \rangle_t| \leq C_{n,\Lambda} k \|f\|_\infty \left(\frac{m^2}{N} + \frac{ms}{N^{3/2}} \right). \quad (4.45)$$

Spin and radial parts. We have $\dot{\rho}_p = (\dot{\chi}_p \sigma^p, \varepsilon^p)$ with $\chi_p = \chi(t\varepsilon^p)$ and $\dot{\chi}_p = -t \|\varepsilon^p\|^2 / (N\chi_p)$. Since $N \geq 2\Lambda n$, we have $1/2 \leq \chi_p^2 \leq N'/N \leq 2$, so $|\dot{\chi}_p| \leq 2\Lambda n/N$ and $\chi_p \leq \sqrt{2}$. Since $\|\rho_p\|^2 = N'$, the diagonal term $\dot{r}_p \cdot r_p$ vanishes, and for $p \neq q$

$$|\dot{r}_p \cdot r_q| = \frac{|\dot{\chi}_p \chi_q \sigma^p \cdot \sigma^q + t \varepsilon^p \cdot \varepsilon^q|}{N'} \leq \frac{4\Lambda n}{N}.$$

So the spin part of D_{pq}^β is at most $4\Lambda n m / N$. For the response part, $\dot{y}_p = J_p \dot{\Psi}_p$ and $\dot{\Psi}_{pa} = U''(x_{pa}) \dot{x}_{pa} / \sqrt{C_X N'}$, where by (4.5)

$$\dot{x}_{pa} = \dot{\chi}_p S_a(\sigma^p) + \sum_{j \leq n} g'_{aj} \zeta_j^p.$$

Put $Y^{pq} = J_p^\top y_q \in \mathbb{R}^M$, so that $\|Y^{pq}\| \leq C_\omega$ and $\dot{y}_p \cdot y_q = \sum_a Y_a^{pq} U''(x_{pa}) \dot{x}_{pa} / \sqrt{C_X N'}$. Since $\sum_a S_a(\sigma)^2 \leq \|\mathbf{G}\|_{\text{op}}^2$, the radial part of $m v_2 K_{pq}^{m-1} \dot{y}_p \cdot y_q$ is at most

$2\Lambda n m L C_\omega \|\mathbf{G}\|_{\text{op}} / (N\sqrt{C_X N'})$, and $\mathbb{E}\|\mathbf{G}\|_{\text{op}} \leq C\sqrt{N}$ by Lemma 4.2. Both parts contribute at most $C_{n,\Lambda} m \|f\|_\infty / N$ to (4.45).

Integration by parts in the fresh columns. The remaining part of D_{pq}^β is $\sum_{a,j} g'_{aj} A_{aj}^{pq}$, where

$$A_{aj}^{pq} = \frac{mv_2}{\sqrt{C_X N'}} \zeta_j^p K_{pq}^{m-1} U''(x_{pa}) Y_a^{pq}.$$

Write $\partial_{aj} = \partial/\partial g'_{aj}$ and let $\mathcal{H}_t(\sigma, \varepsilon) = \sum_a U^\xi(x_a(\rho(\sigma, \varepsilon))) + P_t(\sigma, \varepsilon)$ be the Hamiltonian. The function f does not depend on g'_{aj} . Gaussian integration by parts in g'_{aj} gives

$$\mathbb{E} \left\langle f \sum_{a,j} g'_{aj} A_{aj}^{pq} \right\rangle_t = \sum_{a,j} \mathbb{E} \langle f \partial_{aj} A_{aj}^{pq} \rangle_t + \sum_{\ell \leq k+1} \eta_\ell \sum_{a,j} \mathbb{E} \langle f A_{aj}^{pq} \partial_{aj} \mathcal{H}_t(\sigma^\ell, \varepsilon^\ell) \rangle_t, \quad (4.46)$$

where $\eta_\ell = 1$ for $\ell \leq k$ and $\eta_{k+1} = -k$. The terms $\ell \leq k$ come from the Gibbs densities of the replicas of $f A^{pq}$ (the same replica when $\ell \in \{p, q\}$, another replica otherwise), and the term $\ell = k+1$ comes from the normalization; $\sum_\ell |\eta_\ell| = 2k$. Only the a -th field depends on g'_{aj} , and

$$\partial_{aj} x_{pa} = t \zeta_j^p, \quad \partial_{aj} \bar{x}_{pa} = \zeta_j^p, \quad \partial_{aj} \Psi_p = \frac{t \zeta_j^p U''(x_{pa})}{\sqrt{C_X N'}} e_a,$$

where e_a is the a -th unit vector of $\mathbb{R}^M$. In the following bounds we use $|U''|, |U'''| \leq L$, $|K_{pq}| \leq 1$, $t, v_2 \leq 1$, and $\sum_j |\zeta_j^p \zeta_j^q| \leq \Lambda n / N'$.

Derivatives of the coefficient. The coefficient A_{aj}^{pq} depends on g'_{aj} through four factors.

- (a) Through $U''(x_{pa})$ the derivative is $U'''(x_{pa}) t \zeta_j^p$. This is the only term with the ℓ^1 norm of Y^{pq} , and $\sum_a |Y_a^{pq}| \leq \sqrt{M} \|Y^{pq}\|$. Its sum over a, j is at most $m L \Lambda n C_\omega \sqrt{M} / (\sqrt{C_X N'}^{3/2})$, which is $O_{n,\Lambda}(m/N)$ since $M \leq 2\alpha N$.
- (b) Through the cutoff Jacobian J_p in $Y_a^{pq} = \sum_b (J_p)_{ba} y_{qb}$, the derivative of Y_a^{pq} is $t \zeta_j^p U''(x_{pa}) D^2 \omega(\Psi_p)[e_a, e_a] \cdot y_q / \sqrt{C_X N'}$. Each term $|D^2 \omega(\Psi_p)[e_a, e_a] \cdot y_q|$ is at most C_ω , so the sum over a of this Hessian trace costs a factor M . The sum over a, j is at most $m L^2 \Lambda n M C_\omega / (C_X N'^2) = O_{n,\Lambda}(m/N)$.
- (c) Through the other response vector y_q , whose derivative is $t \zeta_j^q U''(x_{qa}) J_q e_a / \sqrt{C_X N'}$, the derivative of Y_a^{pq} is $t \zeta_j^q U''(x_{qa}) (J_p^\top J_q)_{aa} / \sqrt{C_X N'}$. Since $\sum_a |(J_p^\top J_q)_{aa}| \leq \|J_p\|_F \|J_q\|_F \leq M C_\omega^2$, the sum over a, j is at most $m L^2 \Lambda n M C_\omega^2 / (C_X N'^2) = O_{n,\Lambda}(m/N)$. When $p = q$, (b) and (c) are the two derivatives of $J_p^\top y_p$.
- (d) Through the remaining tensor slots, that is, through K_{pq}^{m-1} , with

$$\partial_{aj} K_{pq} = \frac{v_2 t}{\sqrt{C_X N'}} (\zeta_j^p U''(x_{pa}) Y_a^{pq} + \zeta_j^q U''(x_{qa}) Y_a^{qp}),$$

since the spin part of K_{pq} does not depend on the rows. The two row vectors in each product are paired before norms are taken: $\sum_a |Y_a^{pq}|(|Y_a^{pq}| + |Y_a^{qp}|) \leq 2C_\omega^2$, with no factor M . The sum over a, j is at most $2m(m-1)L^2\Lambda nC_\omega^2/(C_X N'^2) = O_{n,\Lambda}(m^2/N^2)$.

The first sum on the right of (4.46) is therefore at most $C_{n,\Lambda}\|f\|_\infty m^2/N$.

The activation part of the Gibbs density. The activation part of $\partial_{aj}\mathcal{H}_t(\sigma^\ell, \varepsilon^\ell)$ is $(U^\xi)'(\bar{x}_{\ell a})\zeta_j^\ell$, and

$$\sum_{a,j} A_{aj}^{pq} (U^\xi)'(\bar{x}_{\ell a})\zeta_j^\ell = \frac{mv_2 K_{pq}^{m-1} \zeta^p \cdot \zeta^\ell}{\sqrt{C_X N'}} \sum_a U''(x_{pa}) Y_a^{pq} (U^\xi)'(\bar{x}_{\ell a}).$$

By (4.1) and the growth of the derivatives of τ_1 and τ_2 , $|(U^\xi)'(x)| \leq C(1+|x|)$. Since $\sum_a \bar{x}_{\ell a}^2 \leq \|[\mathbf{G}, \mathbf{G}']\|_{\text{op}}^2$ for every configuration, the Cauchy–Schwarz inequality bounds this sum by $CmL\Lambda nC_\omega(\sqrt{M} + \|[\mathbf{G}, \mathbf{G}']\|_{\text{op}})/(\sqrt{C_X} N'^{3/2})$. By Lemma 4.2, $\mathbb{E}\|[\mathbf{G}, \mathbf{G}']\|_{\text{op}} \leq C\sqrt{N}$, so each replica ℓ contributes $C_{n,\Lambda}\|f\|_\infty m/N$.

The Gaussian part of the Gibbs density. Let $\Gamma_\ell = \sum_{\beta'} \lambda_{\beta'} \partial_y h_{\beta'}$, evaluated at ρ_ℓ , be the response gradient of the Gaussian part. Then $\partial_{aj} P_t(\sigma^\ell, \varepsilon^\ell) = s\Gamma_\ell \cdot \partial_{aj} y_\ell = st\zeta_j^\ell U''(x_{\ell a})(J_\ell^\top \Gamma_\ell)_a / \sqrt{C_X N'}$, and

$$\sum_{a,j} A_{aj}^{pq} \partial_{aj} P_t(\sigma^\ell, \varepsilon^\ell) = \frac{stmv_2 K_{pq}^{m-1} \zeta^p \cdot \zeta^\ell}{C_X N'} \sum_a U''(x_{pa}) U''(x_{\ell a}) Y_a^{pq} (J_\ell^\top \Gamma_\ell)_a.$$

We bound Γ_ℓ by a pathwise envelope. By Step 1, $\|\partial_y h_{v,m}\| \leq m \sum_{i \leq m} \|T_{v,m,i}\|_{\text{inj}}$ at all arguments in the unit balls. Hence $\|\Gamma_\ell\| \leq W$ for every configuration, where $W = \sum_{v,m} \lambda_{v,m} m \sum_{i \leq m} \|T_{v,m,i}\|_{\text{inj}}$ does not depend on the Gibbs measure. By (4.41) with $d = N'$ and $\#\mathcal{A} = M$,

$$\mathbb{E}W \leq B_* \sqrt{N' + M}, \quad B_* = \frac{5}{2} C^{1/2} \sum_{v,m} c_{v,m} m(m+1) (m \log(12m))^{1/2} < \infty.$$

The Cauchy–Schwarz inequality in a bounds the last sum by $L^2 C_\omega^2 W$, so each replica ℓ contributes at most $smL^2\Lambda nC_\omega^2\|f\|_\infty \mathbb{E}W/(C_X N'^2) \leq C_{n,\Lambda}\|f\|_\infty ms/N^{3/2}$. Bounding the coordinates of $J_\ell^\top \Gamma_\ell$ one at a time would lose a factor $\sqrt{M}$.

Summing over ℓ with the weights η_ℓ and adding the previous parts proves (4.45).

Summation over the components. Since $u_{v,m} \leq 5/2$ and $c_{v,m} = 2^{-j(v)-m}$ with $j(v) \geq 1$,

$$\sum_{v,m} \lambda_{v,m}^2 m^2 \leq \frac{25}{4} \sum_{j \geq 1} 4^{-j} \sum_{m \geq 1} m^2 4^{-m} = \frac{125}{81}.$$

By (4.45) with $k = 2$ and $m \leq m^2$, the derivative of $\mathbb{E} \log Z_t$ is at most $C_{n,\Lambda}(s^2/N + s^3/N^{3/2})$. Integrating over $t \in [0, 1]$, the complete error is bounded by

$$C_{n,\Lambda}(s_{N'}^2/N + s_{N'}^3/N^{3/2}) = o(1). \quad (4.47)$$

For a bounded measurable function f of the spin overlaps and cavity coordinates of r replicas, (4.45) applies with $k = r + 2$, and the integration by parts in the fresh columns introduces the replica $r + 3$. So $|\partial_t \mathbb{E} \langle f \rangle_t| \leq 4r(r+1)(r+2)C_{n,\Lambda} \|f\|_\infty (s^2/N + s^3/N^{3/2})$.

Infinitely many components and rare events. The bounds above do not depend on $\mathcal{B}$. Let $\mathcal{B}_K$ be the set of the first K components and $\delta_K = s \sum_{\beta \notin \mathcal{B}_K} \lambda_\beta \sup_\rho |h_\beta(\rho)|$. By Step 1, $\delta_K \rightarrow 0$ almost surely and in L^1 . Truncating the Gaussian part to $\mathcal{B}_K$ changes $\log Z_t$ by at most δ_K , and it changes $\langle f \rangle_t$ by at most $\|f\|_\infty \min(2, e^{2r\delta_K} - 1)$, since the Gibbs density of r replicas changes by a factor in $[e^{-2r\delta_K}, e^{2r\delta_K}]$. By dominated convergence, the bounds pass to the full series at $t = 0$ and $t = 1$. No restriction to Ω_N was used: the bounds on ϖ and its derivatives hold everywhere, and $\|\mathbf{G}\|_{\text{op}}$, $\|[\mathbf{G}, \mathbf{G}']\|_{\text{op}}$, and W enter only through their unconditional expectations.

STEP 7. *Part (iv)(b).* In the representation through the fields $x_a = S_a + \eta_a$, the N -dimensional system has the activation $U^{\tilde{N}}(x_a)$ and the Gaussian part $s_N \sum_\beta \lambda_\beta h_\beta$ with the response vector $\varpi(\Psi^1)$, where

$$\Psi^t(\sigma) = \frac{(U'(S_a(\sigma) + t\eta_a(\sigma)))_{a \leq M}}{\sqrt{C_X N}} \quad (0 \leq t \leq 1).$$

Conditionally on the rows, each Gaussian part considered below consists of independent centered Gaussian processes on S_N with covariances $(v_1 R_N(\sigma, \sigma') + v_2 y(\sigma) \cdot y(\sigma'))^m$ for some response vector y with values in the unit ball. The expected log partition function depends on the Gaussian part only through these covariances. We retain finitely many components and pass to the limit as in Step 6, and we make four replacements.

The response vector. Replace $\varpi(\Psi^1)$ by $\varpi(\Psi^t)$, keeping the activation at $S_a + \eta_a$, and let t run from 1 to 0. Define y_p , J_p , K_{pq} , and Y^{pq} as in Step 6, with $r_p = \sigma^p / \sqrt{N}$, $\Psi_p = \Psi^t(\sigma^p)$, $x_{pa} = S_a(\sigma^p) + t\eta_a(\sigma^p)$, and $\zeta^p = c_N \sigma^p \in \mathbb{R}^N$. Then $\sum_i |\zeta_i^p \zeta_i^q| \leq c_N^2 N = n/N'$. The spin argument $\sigma / \sqrt{N}$ does not move, so (4.44) has no spin or radial part, and $\dot{x}_{pa} = \eta_a(\sigma^p) = \sum_i g''_{ai} \zeta_i^p$. Hence $D_{pq}^\beta = \sum_{a,i} g''_{ai} A_{ai}^{pq}$ with

$$A_{ai}^{pq} = \frac{mv_2}{\sqrt{C_X N}} \zeta_i^p K_{pq}^{m-1} U''(x_{pa}) Y_a^{pq},$$

the coefficient of Step 6 with $\varepsilon^p / \sqrt{N'}$ replaced by $c_N \sigma^p$ and N' replaced by N in the normalization of the response vector. Integrate by parts in the entries g''_{ai} of $\mathbf{G}''$. Identity (4.46) holds with $\partial_{g''_{ai}} x_{pa} = t \zeta_i^p$, and the derivative of the activation field $\bar{x}_{\ell a} = S_a(\sigma^\ell) + \eta_a(\sigma^\ell)$ is ζ_i^ℓ . The bounds (a)–(d) on the coefficient and the bounds on the activation and Gaussian parts of the Gibbs density hold as in Step 6, with three changes: $\Lambda n/N'$ becomes n/N' ; $\|[\mathbf{G}, \mathbf{G}']\|_{\text{op}}$ becomes $\|\mathbf{G}\|_{\text{op}} + c_N \sqrt{N} \|\mathbf{G}''\|_{\text{op}}$, which bounds $\|(\bar{x}_{\ell a})_a\|$ and has expectation at most $C\sqrt{N}$; and W is formed with the tensors of dimension $N + M$, so

$\mathbb{E}W \leq B_*\sqrt{N+M}$. Thus (4.45) holds with C_n and $s = s_N$, and the summation of Step 6 bounds the change of $\mathbb{E} \log Z$ by $C_n(s_N^2/N + s_N^3/N^{3/2})$.

Normalization and scaling. At $t = 0$ the response vector is $y^0 = \varpi(\Psi^0)$, while that of the reference perturbation is $y^{\text{ref}} = \varpi(\Psi^{\text{ref}})$ with $\Psi^{\text{ref}}(\sigma) = (U'(\sqrt{N'/N} S_a(\sigma)))_a / \sqrt{C_X N'}$ (see §4.3). The two differ in the normalization (N instead of N') and in the factor $\sqrt{N'/N}$ in the argument of U' . Since $1/\sqrt{N} - 1/\sqrt{N'} \leq n/(2N^{3/2})$, $|U'(\sqrt{N'/N} S) - U'(S)| \leq L|S|n/(2N)$, and $\|(U'(S_a))_a\| \leq L(\sqrt{M} + \|\mathbf{G}\|_{\text{op}})$, and since ϖ is C_ϖ -Lipschitz,

$$\sup_{\sigma} \|y^0(\sigma) - y^{\text{ref}}(\sigma)\| \leq \delta_N, \quad \delta_N = \frac{C_\varpi L n (\sqrt{M} + \|\mathbf{G}\|_{\text{op}})}{\sqrt{C_X} N^{3/2}}, \quad \mathbb{E} \delta_N \leq \frac{Cn}{N}.$$

All response vectors lie in the unit ball, and $|k^m - k'^m| \leq m|k - k'|$ for $k, k' \in [-1, 1]$, so the two covariances of each component differ by at most $2m\delta_N$. Let h_β^0 and h_β^{ref} be conditionally independent Gaussian processes with these covariances, and interpolate with $\sqrt{1-t}h_\beta^0 + \sqrt{t}h_\beta^{\text{ref}}$. Gaussian integration by parts gives the derivative $(s_N^2/2) \sum_\beta \lambda_\beta^2 \mathbb{E} \langle \Delta Q_{11}^\beta - \Delta Q_{12}^\beta \rangle$, where ΔQ^β is the difference of the two covariances. Its absolute value is at most $2s_N^2 \mathbb{E} \delta_N \sum_\beta \lambda_\beta^2 m = O_n(s_N^2/N)$.

The strengths. The derivative of the expected log partition function in s is $\sum_\beta \lambda_\beta \mathbb{E} \langle h_\beta \rangle = s \sum_\beta \lambda_\beta^2 \mathbb{E} \langle Q_{11}^\beta - Q_{12}^\beta \rangle$, by integration by parts, hence $O(s)$. Since $|s_{N'} - s_N| \leq nN^{-2/3}$, changing s_N to $s_{N'}$ costs $O(nN^{-1/3})$. The derivative in ξ is $\mathbb{E} \langle \sum_a (u_1 \tau_1 + u_2 \tau_2)(x_a) \rangle$, which is at most $C(M + \mathbb{E} \|\mathbf{G}\|_{\text{op}}^2 + c_N^2 N \mathbb{E} \|\mathbf{G}''\|_{\text{op}}^2) = O(N)$. Since $|\xi_{N'} - \xi_N| \leq nN^{-5/4}$, changing ξ_N to $\xi_{N'}$ costs $O(nN^{-1/4})$. As in Step 6, no restriction to Ω_N is needed.

STEP 8. *Part (v).* When rows are added, the response vector gains coordinates, and we realize the tensors of the system with M rows as the restrictions of those of the system with $M + \pi$ rows to the old response coordinates. On the event Ω'_N of §4.3, which depends only on the rows and on π , the map ϖ is the identity for the old and the new response vectors, by the choice of C_X . There the Gaussian part of the perturbation changes by a centered Gaussian field that is linear in the tensor entries involving a new coordinate, and these entries are independent of the old ones. Since the logarithm of a Gibbs average of e^h is at least the Gibbs average of h , averaging over these entries shows that this change has a nonnegative effect on the expected log partition function. On the complement, the change of the log partition function is at most $s_{N'} \sum_{v,m} c_{v,m} u_{v,m}$ times the suprema of the old and new components in absolute value, whose second moment is at most CN^2 by Step 1, while $\mathbb{P}(\Omega'_N)^c \leq C_n e^{-N/2}$; by the Cauchy–Schwarz inequality the complement contributes $o(1)$. The added rows carry the activation $U^{\xi_{N'}}$. Since $|\tau_i(x)| \leq C_\tau(1+x^2)$, $u_1 \leq 5/2$, $u_2 \geq 0$, and $\tau_2 \geq 0$, we have $U^{\xi_{N'}} \geq U - c(1+x^2)$ with $c = 4C_\tau \xi_{N'}$, and $c \leq 1$ for N large; the expected logarithmic increment is monotone in the activation. The fields y_i of the added rows

are standard Gaussian for each configuration and independent of $\langle \cdot \rangle_{N'}$. Conditionally on π , apply Lemma 4.12 to $\langle \cdot \rangle_{N'}$ with $V_2 = \sum_{i \leq \pi} U(y_i)$, $V_1 = \sum_{i \leq \pi} [U(y_i) - c(1 + y_i^2)]$, $A = \pi U_{\max}$, and $Y = (C_0 + 1) \sum_{i \leq \pi} (1 + y_i^2)$. Then $\mathbb{E}[\langle Y \rangle^2] \leq \mathbb{E}\langle Y^2 \rangle \leq 6(C_0 + 1)^2 \pi^2$, and $e^{V_2} - e^{V_1} \leq e^{\pi U_{\max}^+ c} \sum_{i \leq \pi} (1 + y_i^2)$ has expectation at most $2\pi c e^{\pi U_{\max}^+}$. Averaging over $\pi \sim \text{Poi}(\alpha n)$, whose exponential moments are finite, we obtain for every $T > 0$

$$0 \leq \mathbb{E} \log \left\langle e^{\sum_{i \leq \pi} U(y_i)} \right\rangle_{N'} - \mathbb{E} \log \left\langle e^{\sum_{i \leq \pi} [U(y_i) - c(1 + y_i^2)]} \right\rangle_{N'} \leq C_n (c e^T + T^{-1}),$$

uniformly in the Gibbs measure. Choosing $T = T_N \rightarrow \infty$ so slowly that $\xi_{N'} e^{T_N} \rightarrow 0$, we see that replacing $U^{\xi_{N'}}$ by U on the added rows costs $o(1)$. This proves the first assertion of (v).

For the uniform lower bound, take $\Lambda = 2$ in the coordinate term. After (4.47), Lemma 4.7(iv)(b) and the interpolation of §4.4, whose errors are uniform in the parameters and in $M \in W_N$, the coordinate term is at least the difference (4.23) minus $o(1)$. Jensen's inequality, for the Gibbs average and for the normalized restriction of ν_n to the centered ball $\mathbb{B}_2$, bounds the logarithm of the numerator below by $\log \nu_n(\mathbb{B}_2) - C_n \langle 1 + |b_N| \rangle$, whose expectation is at least $-C_n$ by (4.8). Jensen's inequality in the field y bounds the expected logarithm of the denominator above by $\mathbb{E} \log \langle e^{n X_N(\sigma, \sigma)/2} \rangle \leq n \mathbb{E} \sup_{\sigma} X_N(\sigma, \sigma)/2 \leq C_n$. By the first assertion of (v) and Jensen's inequality for the fresh rows, the row term is at least $\alpha n \mathbb{E} U(G) - o(1) \geq -C_n$. Combining these bounds proves the lower bound in the last assertion of (v) for $N \geq N_0(n)$. For the upper bound on the absolute value, for any number of rows $|\mathbb{E} \log Z| \leq C(d + \#\mathcal{A}) + C s_d^2$ in each system: without its Gaussian part, the Hamiltonian is at most $C(\#\mathcal{A} + \|\mathbf{G}_d\|_{\text{op}}^2) + C_n \#\mathcal{A}/d$ in absolute value, and Jensen's inequality in the tensors bounds the effect of the Gaussian part between 0 and $C s_d^2$. $\square$

Proof of Lemma 4.9. We fix the perturbation parameters and the number $M \leq 2\alpha N$ of rows. The Gibbs measure $\langle \cdot \rangle_{\text{ref}}$ has density proportional to e^H with respect to μ_N , where

$$H(\sigma) = \sum_{a \leq M} U^{\xi}(S_a(\sigma)) + P(\sigma) + \frac{n}{2N'} \sum_{a \leq M} U''(S_a(\sigma)), \quad \xi = \xi_{N'},$$

$P(\sigma)$ is the Gaussian part of the perturbation of the N' -dimensional system evaluated at $\rho(\sigma, 0)$, and the last sum is the tilt $nB_N/2$. The row g_a enters H through $S_a = g_a \cdot \sigma / \sqrt{N'}$ and, in P , through the response coordinate $U'(x_a) / \sqrt{C_X N'}$ with $x_a = g_a \cdot \sigma / \sqrt{N}$. Hence $\partial_{g_{ai}} H(\sigma) = (\sigma_i / \sqrt{N'}) \phi_a(\sigma)$ with

$$\begin{aligned} \phi_a &= (U^{\xi})'(S_a) + \frac{n}{2N'} U'''(S_a) + \psi_a, \\ \psi_a &= \sqrt{\frac{N'}{N}} \frac{U''(x_a)}{\sqrt{C_X N'}} s_{N'} \sum_{v,m} c_{v,m} u_{v,m} [D\omega(\Psi_{N'})^{\top} \partial_y h_{v,m}]_{a'}, \end{aligned}$$

all evaluated at σ and $\rho(\sigma, 0)$. Since $S_a U'(S_a) = N'^{-1/2} \sum_i g_{ai} \sigma_i U'(S_a)$, Gaussian integration by parts in g_{ai} , conditionally on all other variables, gives

$$\mathbb{E} \langle S_a U'(S_a) \rangle_{\text{ref}} = \frac{N}{N'} \mathbb{E} \left[\langle U''(S_a) \rangle_{\text{ref}} + \langle U'(S_a) \phi_a \rangle_{\text{ref}} - \langle R_N(\sigma^1, \sigma^2) U'(S_a(\sigma^1)) \phi_a(\sigma^2) \rangle_{\text{ref}} \right].$$

The first term comes from differentiating $U'(S_a)$, and the other two from differentiating the Gibbs density; we used $\sum_i \sigma_i^2 = N$ and $\sum_i \sigma_i^1 \sigma_i^2 = N R_N(\sigma^1, \sigma^2)$.

We justify the integration by parts. Write $h_{v,m}(\rho) = H_{v,m}(\rho, \hat{\Psi}_{N'}(\rho))$, where $H_{v,m}(\rho, y) = \langle T_{v,m}, F_{v,m}(\rho, y) \rangle$ and $F_{v,m}(\rho, y)$ is (4.16) with $\hat{\Psi}_d(\rho)$ replaced by y . For $\rho \in S_{N'}$ and $\|y\| \leq 1$, the derivative $\partial_y H_{v,m}(\rho, y)$ has norm at most $\mathcal{I}_{v,m} = m \sum_i \|T_{v,m,i}\|_{\text{inj}}$, as in Step 1 of the previous proof. By (4.41) and Minkowski's inequality, $\mathcal{I} = \sum_{v,m} c_{v,m} \mathcal{I}_{v,m}$ is square integrable, and it does not depend on the rows. Since $\hat{\Psi}_{N'}$ takes values in the unit ball, $D\omega$ is bounded, $|U''| \leq L$, and $|\sigma_i| \leq \sqrt{N}$, the derivative in g_{ai} of the term of index (v, m) of P is bounded by $C_N S_{N'} c_{v,m} \mathcal{I}_{v,m}$, uniformly in σ and g_{ai} . By the Weierstrass M-test, P is continuously differentiable in g_{ai} , jointly continuous with its derivative in (σ, g_{ai}) , and its derivative is the termwise one; the other two parts of H are C^1 in g_{ai} since $U \in C^4$. Differentiating under the integral over the compact sphere S_N shows that $F(g_{ai}) = \langle \sigma_i U'(S_a) \rangle_{\text{ref}}$, with all other variables fixed, is C^1 with the derivative used above. By $|\sigma_i| \leq \sqrt{N}$, $|S_a| \leq \|g_a\|$, and the growth bounds in (4.1), $|F| + |F'| \leq C_N (1 + \|g_a\|)^2 (1 + \mathcal{I})$, which is integrable because $\mathcal{I}$ is independent of the rows. Hence Stein's lemma $\mathbb{E}[gF(g)] = \mathbb{E}[F'(g)]$ for a standard Gaussian g , valid for absolutely continuous F with $\mathbb{E}|F'(g)| < \infty$, applies conditionally on the other variables, and Fubini's theorem gives the display.

Summing over $a \leq M$, dividing by N , and writing $\phi_a = U'(S_a) + e_a$, the definitions in (4.7) give

$$\begin{aligned} \mathbb{E} \langle b_N \rangle_{\text{ref}} - 1 - \mathbb{E} \langle X_N(\sigma, \sigma) \rangle_{\text{ref}} + \mathbb{E} \langle R_N(\sigma^1, \sigma^2) X_N(\sigma^1, \sigma^2) \rangle_{\text{ref}} \\ = \frac{1}{N'} \sum_{a \leq M} \mathbb{E} \left[\langle U'(S_a) e_a \rangle_{\text{ref}} - \langle R_N(\sigma^1, \sigma^2) U'(S_a(\sigma^1)) e_a(\sigma^2) \rangle_{\text{ref}} \right]. \end{aligned}$$

Since $|R_N| \leq 1$, the Cauchy–Schwarz inequality over a bounds the right side by

$$\frac{2}{N'} \mathbb{E} \left[\sup_{\sigma} \|U'(S(\sigma))\| \sup_{\sigma} \|e(\sigma)\| \right],$$

with Euclidean norms over $a \leq M$. We bound the factors separately, using $\sum_a S_a^2 \leq \|\mathbf{G}\|_{\text{op}}^2$ and $\mathbb{E} \|\mathbf{G}\|_{\text{op}}^2 \leq C(M + N)$ (Lemma 4.2).

(a) Since $|U'(x)| \leq L(1 + |x|)$, $\sup_{\sigma} \|U'(S(\sigma))\|^2 \leq 2L^2(M + \|\mathbf{G}\|_{\text{op}}^2)$.

(b) The scalar perturbation contributes $\xi(u_1 \tau'_1 + u_2 \tau'_2)(S_a)$ to e_a . Since $\tau'_1 = U' + xU'' - U'''$ and $\tau'_2 = 2U'U''$ grow at most linearly, its norm is at most $C\xi(M + \|\mathbf{G}\|_{\text{op}}^2)^{1/2}$.

- (c) The tilt contributes $\frac{n}{2N'}U'''(S_a)$, of norm at most $nL\sqrt{M}/(2N')$.
- (d) The Gaussian part contributes ψ_a . Since $D\omega$ is bounded, $u_{v,m} \leq 5/2$ and $|U''| \leq L$, we have $\|\psi\| \leq Cs_{N'}N^{-1/2}\sum_{v,m}c_{v,m}\sup_{\rho}\|\partial_y h_{v,m}(\rho)\|$. By (4.42) and the Cauchy–Schwarz inequality with the summable weights $c_{v,m}$, $\mathbb{E}\sup_{\sigma}\|\psi(\sigma)\|^2 \leq Cs_{N'}^2(N' + M)/N$.

Since $M \leq 2\alpha N$, the right side is at most $CN^{-1}(CN)^{1/2}(\zeta_{N'}^2N + n^2/N + s_{N'}^2)^{1/2} \leq C_n(\zeta_{N'} + N^{-1} + s_{N'}N^{-1/2})$, uniformly in the parameters and in $M \in W_N$. This tends to zero, since $\zeta_{N'} = N'^{-1/4}$ and $s_{N'} = N'^{1/3}$. $\square$

Proof of Lemma 4.8. We divide the proof into four steps.

STEP 1. Reduction to the arrays. Let S be the closed set of arrays in which the R - and X -blocks are positive semidefinite; all arrays of the random measures and their limits lie in S . By (G), for every r and every bounded measurable function Φ of the arrays of r replicas and of the values of the fields at these replicas, $\mathbb{E}\langle\Phi\rangle = \mathbb{E}\langle\bar{\Phi}(\mathbf{A}_r)\rangle$. If Φ is continuous, then $\bar{\Phi}$ is continuous on S , since centered Gaussian laws depend continuously on their covariance matrices. Hence $\mathbb{E}\langle\Phi\rangle$ converges under convergence of the finite-dimensional distributions of the arrays, and its limit depends only on the limiting law.

STEP 2. Moment bounds. Write $W(\rho) = \int_{\mathbb{B}_\Lambda} \nu_n(d\varepsilon)e^{\varepsilon \cdot z(\rho) - (\|\varepsilon\|^2 - n)(b(\rho) - 1)/2}$, so that $\mathcal{N}_\Lambda = \langle W \rangle$. On the fixed ball, $W \leq \exp\{\sqrt{\Lambda n}\|z(\rho)\| + C_{n,\Lambda}(1 + |b(\rho)|)\}$, and Jensen's inequality for the normalized restriction of ν_n to the centered ball gives $W \geq \nu_n(\mathbb{B}_\Lambda) \exp\{-C_{n,\Lambda}(1 + |b(\rho)|)\}$. Under N_a , $z(\rho^1)$ has the law $N(0, X_{11}I_n)$, and $\mathbb{E}e^{c\|z\|} \leq 2^n e^{nc^2 X_{11}/2}$ for such a vector. By (G), applied to truncations and by monotone convergence, and by the moment hypothesis, $\mathbb{E}\langle W^q \rangle$ and $\mathbb{E}\langle W^{-q} \rangle$ are bounded along the sequence for every fixed $q \geq 1$. By Jensen's inequality, $\langle W \rangle^{\pm q} \leq \langle W^{\pm q} \rangle$, so $\mathbb{E}[\mathcal{N}_\Lambda^q + \mathcal{N}_\Lambda^{-q}]$ is bounded as well. The same holds for $\mathcal{D} = \langle e^y \rangle$, since $y(\rho^1)$ has the law $N(0, nX_{11})$ under N_a .

STEP 3. Parts (i) and (ii). Let $T \geq 1$, put $W_T = \min\{W, T\}$ and $\mathcal{N}_T = \langle W_T \rangle$. Then $0 \leq \log \mathcal{N}_\Lambda - \log \mathcal{N}_T \leq (\mathcal{N}_\Lambda - \mathcal{N}_T)/\mathcal{N}_T$, with $\mathcal{N}_\Lambda - \mathcal{N}_T \leq \langle W^2 \rangle/T$ and $\mathcal{N}_T \geq \langle \min\{W, 1\} \rangle$, whose negative moments are bounded by Step 2. By the Cauchy–Schwarz inequality, $\mathbb{E}|\log \mathcal{N}_\Lambda - \log \mathcal{N}_T| \leq C/T$ uniformly along the sequence. Next let $0 < \delta < 1$, and let P be a polynomial with $|P(x) - \log \max\{x, \delta\}| \leq \delta$ on $[0, T]$. Since $\mathcal{N}_T \in (0, T]$ and $\mathbb{P}(\mathcal{N}_T < \delta) \leq \delta^2 \mathbb{E}\mathcal{N}_T^{-2}$, the moment bounds of Step 2 give

$$\mathbb{E}|\log \mathcal{N}_T - P(\mathcal{N}_T)| \leq \delta + \mathbb{E}[(|\log \mathcal{N}_T| + |\log \delta|) \mathbf{1}_{\{\mathcal{N}_T < \delta\}}] \leq \delta + C\delta |\log \delta|,$$

uniformly along the sequence. Finally, $\mathbb{E}[\mathcal{N}_T^j] = \mathbb{E}\langle \prod_{\ell \leq j} W_T(\rho^\ell) \rangle$ is the expectation of a bounded continuous function of the arrays of j replicas and of the values of z at them, so it converges by Step 1, with a limit that depends only on the limiting law. Letting $N \rightarrow \infty$, then $\delta \rightarrow 0$ and $T \rightarrow \infty$, proves the assertion of (i) for $\mathcal{N}_\Lambda$; the denominator $\mathcal{D}$ is treated

in the same way. For (ii), the expectation of a bounded continuous function F of the arrays and the cavity coordinates of r replicas under the cavity measure is $\mathbb{E}[\langle \Phi_F \rangle / \mathcal{N}_\Lambda^r]$, where

$$\Phi_F = \int_{\mathbb{B}_\Lambda^r} F \prod_{\ell \leq r} e^{\varepsilon^\ell \cdot z(\rho^\ell) - (\|\varepsilon^\ell\|^2 - n)(b(\rho^\ell) - 1)/2} \nu_n(d\varepsilon^\ell)$$

satisfies $|\Phi_F| \leq \|F\|_\infty \prod_\ell W(\rho^\ell)$. Let Φ_F^T be defined in the same way with each factor multiplied by $W_T(\rho^\ell)/W(\rho^\ell)$; it is a bounded continuous function of the arrays and of the values of z at the r replicas, and $|\Phi_F^T| \leq \|F\|_\infty \prod_\ell W_T(\rho^\ell)$. Replacing one factor at a time gives $|\langle \Phi_F \rangle - \langle \Phi_F^T \rangle| \leq r\|F\|_\infty \mathcal{N}_\Lambda^{r-1}(\mathcal{N}_\Lambda - \mathcal{N}_T)$, and $|\mathcal{N}_\Lambda^{-r} - \mathcal{N}_T^{-r}| \langle |\Phi_F^T| \rangle \leq r\|F\|_\infty(\mathcal{N}_\Lambda - \mathcal{N}_T)/\mathcal{N}_T$. As above, both errors are at most C/T in expectation, uniformly along the sequence. Finally, $|\langle \Phi_F^T \rangle|/\mathcal{N}_T^r \leq \|F\|_\infty$, and replacing $x \mapsto x^{-r}$ by a polynomial that approximates $\max\{x, \delta\}^{-r}$ within δ on $[0, T]$ costs at most $\|F\|_\infty(T^r\delta + 3\mathbb{P}(\mathcal{N}_T < \delta))$ in expectation. The expectations $\mathbb{E}[\langle \Phi_F^T \rangle \mathcal{N}_T^j] = \mathbb{E}[\langle \Phi_F^T(\rho^1, \dots, \rho^r) \prod_{\ell=r+1}^{r+j} W_T(\rho^\ell) \rangle]$ converge by Step 1, and letting $N \rightarrow \infty$, then $\delta \rightarrow 0$ and $T \rightarrow \infty$ proves (ii).

STEP 4. *Part (iii).* A quadratic lower bound on U suffices; negative exponential moments of U are not needed. Fix π and $T > 0$, and put $U_T = \max\{U, -2T\}$. Apply Lemma 4.12 with $V_1 = \sum_{i \leq \pi} U(y_i)$, $V_2 = \sum_{i \leq \pi} U_T(y_i)$, $A = \pi U_{\max}^+$, and $Y = C_0 \sum_{i \leq \pi} (1 + y_i^2)$; by (G) and $R(\rho, \rho) = 1$, each $y_i(\rho)$ is standard Gaussian for one replica, so $\mathbb{E}[\langle Y \rangle^2] \leq \mathbb{E}\langle Y^2 \rangle \leq 6C_0^2\pi^2$. Since $0 \leq e^{U_T} - e^U \leq e^{-2T}$ and all activations are at most $U_{\max}^+$, writing $e^{V_2} - e^{V_1}$ as a telescoping sum over the rows gives $e^{V_2} - e^{V_1} \leq \pi e^{(\pi-1)U_{\max}^+} e^{-2T}$. Hence

$$0 \leq \mathbb{E} \log \langle e^{V_2} \rangle - \mathbb{E} \log \langle e^{V_1} \rangle \leq \pi e^{(\pi-1)U_{\max}^+} e^{-T} + \frac{C\pi^2}{T}$$

uniformly over the random measures. For fixed T and π the weight e^{V_2} lies between $e^{-2\pi T}$ and $e^{\pi U_{\max}^+}$ and is a bounded continuous function of the values of $y_1, \dots, y_\pi$, so the argument of Step 3, with $\log$ approximated by polynomials on this compact interval, applies to it. Letting $T \rightarrow \infty$ proves (iii) for a fixed number of rows. For a Poisson number of rows, independent of the other variables, the bounds above are integrable in π , and $|\mathbb{E} \log \langle e^{V_1} \rangle| \leq C(1 + \pi)$ given π , so dominated convergence applies.

Finally, if $\mathbb{E} \langle |b - \bar{b}_N| \rangle \rightarrow 0$ and $\bar{b}_N \rightarrow b_\infty$, then the arrays with b and with b replaced by b_∞ have the same limiting finite-dimensional distributions, and the random measures with the constant b_∞ satisfy the hypotheses of the lemma. Since the limits depend only on the limiting law, b may be replaced by b_∞ . $\square$

Proof of the Parisi formula

In this chapter we prove Theorem 1.1: for U in class (A) or class (B), $F_N(U) \rightarrow \mathcal{P}_*(U)$ and $f_N(U) \rightarrow \mathcal{P}_*(U)$ in probability, the infimum defining $\mathcal{P}_*(U)$ is attained, and every minimizer γ satisfies

$$Q_\gamma \leq r_U, \quad r_U = \frac{2\alpha C_U}{1 + 2\alpha C_U}, \quad (5.1)$$

where C_U is the constant of Lemma 2.4. In class (A), $C_U \leq C_U^A = \|U'\|_\infty^2/2$, and since $x \mapsto x/(1+x)$ is increasing, $r_U \leq \alpha\|U'\|_\infty^2/(1 + \alpha\|U'\|_\infty^2)$, as stated in Theorem 1.1. We combine the upper bound of Chapter 3 (Theorem 3.1, which applies to every activation of class (A)) with the lower bound of Chapter 4 (Theorem 4.1). The lower bound was proved under the additional smoothness (4.1); Lemma 5.1 removes it by mollification. The upper bound requires a bounded first derivative, which may fail in class (B); there we approximate U from above by the activations $-T + \log(1 + e^{U+T})$, which belong to class (A). Convergence in probability follows from the Gaussian Poincaré inequality and the operator-norm bound of Lemma 4.2. Finally, the truncation lemma (Lemma 5.2) reduces the infimum to the compact class $\mathcal{U}_{r_U}$, on which $\mathcal{P}_U$ is continuous, and places every minimizer in $\mathcal{U}_{r_U}$.

5.1 Mollification

Fix an even smooth probability density ϑ on $\mathbb{R}$ supported in $[-1, 1]$, and for $0 < \epsilon \leq 1$ put $\vartheta_\epsilon(x) = \epsilon^{-1}\vartheta(x/\epsilon)$ and $U_\epsilon = U * \vartheta_\epsilon$.

Lemma 5.1. *Let U belong to class (A) or class (B), and put $\delta_\epsilon = \epsilon^2\|U''\|_\infty/2$.*

- (a) U_ϵ satisfies (4.1) and belongs to the same class as U .
- (b) $|U_\epsilon(x) - U(x)| \leq \delta_\epsilon$ for all $x \in \mathbb{R}$.
- (c) For every N , pointwise in the disorder, $|f_N(U_\epsilon) - f_N(U)| \leq \alpha_N \delta_\epsilon$, where $\alpha_N = M_N/N$.
For every $\gamma \in \mathcal{U}$, $|\mathcal{P}_{U_\epsilon}(\gamma) - \mathcal{P}_U(\gamma)| \leq \alpha \delta_\epsilon$; consequently $|\mathcal{P}_*(U_\epsilon) - \mathcal{P}_*(U)| \leq \alpha \delta_\epsilon$.

Proof. We prove the three parts in order.

Proof of (a). The function U_ϵ is smooth, and for $2 \leq j \leq 4$ its derivative $U_\epsilon^{(j)} = U'' * \vartheta_\epsilon^{(j-2)}$ is bounded by $\|U''\|_\infty \|\vartheta_\epsilon^{(j-2)}\|_{L^1}$; these bounds depend on ϵ , which is allowed in (4.1). Since ϑ_ϵ is a probability density supported in $[-1, 1]$, we have $\sup U_\epsilon \leq \sup U$, $\|U_\epsilon''\|_\infty \leq$

$\|U''\|_\infty, \|U'_\epsilon\|_\infty \leq \|U'\|_\infty$, and $|U'_\epsilon(x)| \leq \sup_{|y| \leq 1} |U'(x-y)| \leq |U'(0)| + \|U''\|_\infty(1+|x|)$. Convolution with a nonnegative kernel preserves concavity. Hence U_ϵ belongs to the class of U and satisfies (4.1), with L chosen large in terms of $|U_\epsilon(0)|, |U'(0)|, \|U''\|_\infty$, and ϵ .

Proof of (b). By Taylor's formula, $U(x-y) = U(x) - U'(x)y + \frac{1}{2}U''(\bar{x})y^2$ for some $\bar{x}$ between x and $x-y$. Integrating against $\vartheta_\epsilon(y) dy$, the linear term vanishes because ϑ_ϵ is even, and the quadratic term is at most $\frac{1}{2}\|U''\|_\infty \int y^2 \vartheta_\epsilon(y) dy \leq \delta_\epsilon$.

Proof of (c). By (b), $e^{-M_N \delta_\epsilon} \leq Z_N(U_\epsilon)/Z_N(U) \leq e^{M_N \delta_\epsilon}$, which gives the first assertion. For the second, (2.5), which holds for every $\gamma \in \mathcal{U}$ since U and U_ϵ belong to the same class by (a), and (b) give $|u_\gamma(0,0;U_\epsilon) - u_\gamma(0,0;U)| \leq \delta_\epsilon$. The entropy term does not depend on the activation, so $|\mathcal{P}_{U_\epsilon}(\gamma) - \mathcal{P}_U(\gamma)| \leq \alpha \delta_\epsilon$, and taking the infimum over γ gives the last assertion. $\square$

5.2 Truncation near $q = 1$

The functional $\mathcal{P}_U$ is continuous on each compact class $\mathcal{U}_r$, $r < 1$, by (2.20), but $\mathcal{U}$ is not compact: mass of an order parameter can escape to $q = 1$. When the mass above the level r_U of (5.1) is moved down to r_U , the energy term rises, measured against $(1-\gamma(s)) ds$ on $[r_U, 1]$, at rate at most αC_U , and the entropy falls at rate at least $s/(2(1-s))$, which exceeds αC_U for $s > r_U$. The next lemma is the analogue for $\mathcal{P}_U$ of a fact about the spherical mixed p -spin model, whose Parisi formula was proved in [37, 125]: its Parisi measures are supported in an interval $[0, q_1]$ with $q_1 < 1$ depending only on the model (see the discussion in [125] and [8, Section 5]).

Lemma 5.2. *Let U belong to class (A) or class (B), and let r_U be as in (5.1). For $\gamma \in \mathcal{U}$ let $\hat{\gamma} = \gamma$ on $[0, r_U)$ and $\hat{\gamma} = 1$ on $[r_U, 1]$. Then $\hat{\gamma} \in \mathcal{U}_{r_U}$ and*

$$\mathcal{P}_U(\hat{\gamma}) - \mathcal{P}_U(\gamma) \leq - \int_{r_U}^1 (1-\gamma(s)) \left(\frac{s}{2(1-s)} - \alpha C_U \right) ds \leq 0. \quad (5.2)$$

The inequality $\mathcal{P}_U(\hat{\gamma}) \leq \mathcal{P}_U(\gamma)$ is strict if $Q_\gamma > r_U$. Consequently $\mathcal{P}_*(U) = \inf_{\gamma \in \mathcal{U}_{r_U}} \mathcal{P}_U(\gamma)$, and every minimizer γ of $\mathcal{P}_U$ satisfies $Q_\gamma \leq r_U$.

Proof. Since $\gamma \leq 1$, the function $\hat{\gamma}$ is nondecreasing and right-continuous, and it equals one on $[r_U, 1]$; so $\hat{\gamma} \in \mathcal{U}_{r_U}$. Moreover $\hat{\gamma} - \gamma = (1-\gamma)\mathbf{1}_{[r_U, 1]} \geq 0$, and $1-\gamma(s) = 0$ for $s \geq Q_\gamma$. We divide the rest of the proof into three steps.

STEP 1. *The energy term.* As distribution functions of probability measures on $[0, 1]$, γ and $\hat{\gamma}$ satisfy $W_1(\gamma, \hat{\gamma}) = \int_0^1 (\hat{\gamma} - \gamma) ds = \int_{r_U}^1 (1-\gamma(s)) ds$, so (2.19) gives

$$\alpha u_{\hat{\gamma}}(0,0;U) - \alpha u_\gamma(0,0;U) \leq \alpha C_U \int_{r_U}^1 (1-\gamma(s)) ds.$$

STEP 2. *The entropy term.* Integrating $\gamma \leq \hat{\gamma} \leq 1$ over $[q, 1]$ gives $0 < \lambda_\gamma(q) \leq \lambda_{\hat{\gamma}}(q) \leq 1 - q$ for $q < 1$, and

$$\lambda_{\hat{\gamma}}(q) - \lambda_\gamma(q) = \int_{\max\{q, r_U\}}^1 (1 - \gamma(s)) \, ds.$$

The integrands in (2.9) for γ and for $\hat{\gamma}$ take values in $[0, (1 - Q_\gamma)^{-1}]$, because $\lambda_{\hat{\gamma}} \geq \lambda_\gamma \geq 1 - Q_\gamma$ on $[0, Q_\gamma]$, and they vanish on $[Q_\gamma, 1]$. Subtracting the two instances of (2.9) and using $\lambda_\gamma \lambda_{\hat{\gamma}} \leq (1 - q)^2$,

$$\begin{aligned} \text{CS}(\gamma) - \text{CS}(\hat{\gamma}) &= \frac{1}{2} \int_0^1 \frac{\lambda_{\hat{\gamma}}(q) - \lambda_\gamma(q)}{\lambda_\gamma(q) \lambda_{\hat{\gamma}}(q)} \, dq \geq \frac{1}{2} \int_0^1 \frac{\lambda_{\hat{\gamma}}(q) - \lambda_\gamma(q)}{(1 - q)^2} \, dq \\ &= \frac{1}{2} \int_{r_U}^1 (1 - \gamma(s)) \frac{s}{1 - s} \, ds. \end{aligned}$$

The last equality is Tonelli's theorem: the function $(1 - \gamma(s))(1 - q)^{-2}$ is nonnegative on the region $\{(q, s) : 0 \leq q \leq s, r_U \leq s < 1\}$, and $\int_0^s (1 - q)^{-2} \, dq = s/(1 - s)$.

STEP 3. *Conclusion.* Adding the two estimates gives the first inequality in (5.2). The function $s \mapsto s/(1 - s)$ is increasing on $[0, 1)$ and equals $2\alpha C_U$ at $s = r_U$, so the integrand in (5.2) is nonnegative, and it is positive at every $s \in (r_U, Q_\gamma)$, where $\gamma(s) < 1$. If $Q_\gamma > r_U$, this interval has positive length and the inequality is strict. Since $\mathcal{U}_{r_U} \subset \mathcal{U}$, we obtain $\mathcal{P}_*(U) = \inf_{\mathcal{U}_{r_U}} \mathcal{P}_U$. If γ is a minimizer with $Q_\gamma > r_U$, then $\mathcal{P}_U(\hat{\gamma}) < \mathcal{P}_U(\gamma) = \mathcal{P}_*(U)$, which is impossible. Therefore every minimizer satisfies $Q_\gamma \leq r_U$. $\square$

5.3 The two bounds and attainment of the infimum

Proof of Theorem 1.1. Let U belong to class (A) or class (B). We divide the proof into four steps: the lower bound, the upper bound, convergence in probability, and attainment together with (5.1).

STEP 1. *The lower bound.* Let $0 < \epsilon \leq 1$. By Lemma 5.1(a), U_ϵ satisfies (4.1) and belongs to the class of U , so U'_ϵ is bounded or U_ϵ is concave; hence Theorem 4.1 gives $\liminf_N F_N(U_\epsilon) \geq \mathcal{P}_*(U_\epsilon)$. By Lemma 5.1(c) and $\alpha_N \rightarrow \alpha$,

$$\liminf_{N \rightarrow \infty} F_N(U) \geq \liminf_{N \rightarrow \infty} F_N(U_\epsilon) - \alpha \delta_\epsilon \geq \mathcal{P}_*(U_\epsilon) - \alpha \delta_\epsilon \geq \mathcal{P}_*(U) - 2\alpha \delta_\epsilon.$$

Letting $\epsilon \rightarrow 0$ gives $\liminf_N F_N(U) \geq \mathcal{P}_*(U)$. The limits are taken in the order: first ϵ is fixed and $N \rightarrow \infty$ (inside Theorem 4.1, the cavity dimension n is sent to infinity after N), then $\epsilon \rightarrow 0$.

STEP 2. *The upper bound.* If U belongs to class (A), Theorem 3.1 gives $\limsup_N F_N(U) \leq \mathcal{P}_U(\gamma)$ for every $\gamma \in \mathcal{U}$, hence $\limsup_N F_N(U) \leq \mathcal{P}_*(U)$.

Let now U belong to class (B), and put $B = \sup U$ and $K = \|U''\|_\infty$. If $K = 0$, then U is affine and bounded above, hence constant, and it belongs to class (A). Assume $K > 0$. We first show that $|U'|$ is controlled by the distance of U from its supremum:

$$U'(x)^2 \leq 2K(B - U(x)), \quad x \in \mathbb{R}. \quad (5.3)$$

Indeed, Taylor's formula gives $B \geq U(x + y) \geq U(x) + U'(x)y - Ky^2/2$ for every $y \in \mathbb{R}$, and the choice $y = U'(x)/K$ gives $B \geq U(x) + U'(x)^2/(2K)$.

For $T > 0$ let $s(v) = e^v/(1 + e^v)$ and

$$U_T(x) = -T + \log(1 + e^{U(x)+T}).$$

Then $U_T \geq U$, since $\log(1 + e^v) \geq v$; U_T decreases as T increases, since $\partial_T U_T = s(U + T) - 1 < 0$; and $U_T \downarrow U$ pointwise, since $U_T - U = \log(1 + e^{-U-T})$. Moreover U_T belongs to class (A). It is C^2 with $\sup U_T \leq -T + \log(1 + e^{B+T})$, and its derivatives are

$$U'_T = s(U + T) U', \quad U''_T = s(U + T) U'' + s'(U + T) (U')^2.$$

By (5.3), with $v = U(x) + T \leq B + T$,

$$|U'_T(x)|^2 \leq 2K s(v)^2 (B + T - v), \quad |U''_T(x)| \leq K + 2K s'(v) (B + T - v).$$

Since $s(v) \leq e^v$ and $0 \leq s'(v) \leq e^v$, both right sides are bounded uniformly in $v \leq B + T$. Hence U'_T and U''_T are bounded.

Fix a step order parameter γ and $T > 0$. Since $U \leq U_T$, we have $Z_N(U) \leq Z_N(U_T)$, and Theorem 3.1 applied to U_T gives

$$\limsup_{N \rightarrow \infty} F_N(U) \leq \limsup_{N \rightarrow \infty} F_N(U_T) \leq \mathcal{P}_{U_T}(\gamma).$$

We now let $T \rightarrow \infty$. The quantity $u_\gamma(0, 0; U_T)$ is computed by the recursion (2.2) started from the terminal datum U_T , and we show by induction over the steps that the intermediate functions decrease pointwise to those computed from U ; they decrease in T by (2.4). At a step with positive parameter m , the functions e^{mh_T} decrease to e^{mh} and are bounded by $e^{m \sup h_{T_0}}$ for $T \geq T_0$, so monotone convergence applies. At a step with parameter zero, the functions h_T are bounded above by $\sup h_{T_0}$ and below by h , and $h(x) \geq -C(2 + x^2)$ by (2.6), since $U(x) \geq U(0) + U'(0)x - Kx^2/2 \geq -C(1 + x^2)$. Dominated convergence therefore applies. Hence $u_\gamma(0, 0; U_T) \downarrow u_\gamma(0, 0; U)$ and $\mathcal{P}_{U_T}(\gamma) \downarrow \mathcal{P}_U(\gamma)$, so

$$\limsup_{N \rightarrow \infty} F_N(U) \leq \mathcal{P}_U(\gamma)$$

for every step order parameter γ . Finally let $\gamma \in \mathcal{U}$ be arbitrary, and fix $\bar{q} \in [Q_\gamma, 1)$. The step order parameters γ_j of §2.1 with $\gamma_j = 1$ on $[\bar{q}, 1]$ and $\|\gamma_j - \gamma\|_{L^1} \rightarrow 0$ satisfy

$\mathcal{P}_U(\gamma_j) \rightarrow \mathcal{P}_U(\gamma)$ by (2.20) with $r = \bar{q}$. Therefore $\limsup_N F_N(U) \leq \mathcal{P}_U(\gamma)$ for every $\gamma \in \mathcal{U}$, that is, $\limsup_N F_N(U) \leq \mathcal{P}_*(U)$. The order of limits is: first the step order parameter γ and T are fixed and $N \rightarrow \infty$; then $T \rightarrow \infty$; then γ is approximated.

Combining Steps 1 and 2, $\lim_N F_N(U) = \mathcal{P}_*(U)$ in both classes.

STEP 3. Convergence in probability. We apply the Gaussian Poincaré inequality (§4.8) to $f_N(U)$ as a function of the entries g_{ai} of $\mathbf{G}$; it is continuously differentiable, with $\partial_{g_{ai}} f_N(U) = N^{-3/2} \langle U'(g_a \cdot \sigma / \sqrt{N}) \sigma_i \rangle$, where $\langle \cdot \rangle$ is the Gibbs measure with density proportional to $\exp\{\sum_a U(g_a \cdot \sigma / \sqrt{N})\}$. By the Cauchy–Schwarz inequality and $\|\sigma\|^2 = N$, $\sum_i \langle U'(g_a \cdot \sigma / \sqrt{N}) \sigma_i \rangle^2 \leq N \langle U'(g_a \cdot \sigma / \sqrt{N})^2 \rangle$, so

$$\text{Var}(f_N(U)) \leq \frac{1}{N^2} \mathbb{E} \left\langle \sum_{a \leq M_N} U'(g_a \cdot \sigma / \sqrt{N})^2 \right\rangle.$$

In class (A) the right side is at most $\alpha_N \|U'\|_\infty^2 / N$. In class (B), $|U'(x)| \leq |U'(0)| + \|U''\|_\infty |x| \leq C(1 + |x|)$ and $\sum_a (g_a \cdot \sigma)^2 / N \leq \|\mathbf{G}\|_{\text{op}}^2$, and Lemma 4.2 gives $\mathbb{E} \|\mathbf{G}\|_{\text{op}}^2 \leq 4(M_N + N) + 8$. Hence the right side is at most $C(M_N + \mathbb{E} \|\mathbf{G}\|_{\text{op}}^2) / N^2 = O(1/N)$. In both cases $f_N(U) - F_N(U) \rightarrow 0$ in probability by Chebyshev's inequality, and Steps 1 and 2 give $f_N(U) \rightarrow \mathcal{P}_*(U)$ in probability.

STEP 4. Attainment and the location of minimizers. By Lemma 5.2, $\mathcal{P}_*(U)$ is the infimum of $\mathcal{P}_U$ over $\mathcal{U}_{r_U}$. The class $\mathcal{U}_{r_U}$ is a compact metric space for the L^1 distance (§2.1), and $\mathcal{P}_U$ is Lipschitz on it by (2.20). A continuous real function on a compact metric space attains its infimum, so $\mathcal{P}_U(\gamma_*) = \mathcal{P}_*(U)$ for some $\gamma_* \in \mathcal{U}_{r_U}$. By Lemma 5.2, every minimizer satisfies (5.1). This proves Theorem 1.1. $\square$

The variational problem at hard margins

This chapter studies the deterministic minimization problem $\mathcal{P}_*(\kappa) = \inf_{\gamma \in \mathcal{U}} \mathcal{P}_\kappa(\gamma)$ of (1.6) and proves Theorem 1.3; Step 3 of §1.3 outlines the argument. No random constraints appear; the only random objects are the Parisi diffusion and auxiliary Gaussian variables. §6.1 proves the value identity and the existence of minimizers, §6.2 proves the slope bound, and §6.3 proves weak top stability, the replica symmetry of constrained minimizers at very negative margins, and the finiteness of κ_c . §6.4 proves the divergence of $\mathcal{P}_*$ at κ_c and deduces the replica symmetry of minimizers at very negative margins. §6.5 shows that a very negative value pins the top of every minimizer close to one, and deduces the collapse of minimizers at κ_c . Chapter 7 uses the definition of κ_c and the Gaussian bounds of Lemma 6.1; Chapter 8 uses the results below to identify the hard-wall volume and the largest margin, and Chapter 9 uses the collapse of minimizers to identify the gap law. The value identity (Lemma 6.2), the existence of minimizers (Theorem 6.3), the slope bound (Lemma 6.4), weak top stability (Lemma 6.5), the finiteness of κ_c (Proposition 6.7) and the divergence at κ_c (Theorem 6.8) are also used in Part II, whose §§2.6 and 2.7 state them and refer to the proofs given here (Part II, Lemmas 2.13–2.15, Theorem 2.16, Proposition 2.17 and Theorem 2.18).

The stationarity conditions in Lemma 6.2 and (6.9) are the analogues for $\mathcal{P}_\kappa$ of the first-order and obstacle conditions for Parisi measures of mixed p -spin models, due to Talagrand [126], to Jagannath and Tobasco [77], and to W.-K. Chen [38, Theorem 2 and Proposition 1], and of their spherical versions, which appear in the work of Jagannath and Tobasco [78] and of Auffinger and W.-K. Chen [8, Proposition 4]. Jagannath and Tobasco characterized Parisi measures as the measures carried by the minimum points of an antiderivative of the first variation [77, Proposition 1.1]; there, convexity makes the condition also sufficient. For the spherical model, whose Crisanti–Sommers functional is the closer analogue of $\mathcal{P}_\kappa$, the corresponding characterization is due to Talagrand [125, Proposition 2.1]. For $\mathcal{P}_\kappa$ itself, El Alaoui and Sellke obtained $D = 0$ and $S = 1$ at a minimizer whose support is an interval, in directions supported on that interval [54, Proposition 16]; see the opening of Chapter 2. Here the conditions on D (Lemma 6.2 and (6.9)) hold for every minimizer over $\mathcal{U}_r$, with no assumption on its support and including the case $Q = r$ of an active cutoff; the condition $S(Q) = 1$ appears only at an accumulation point of the support (proof of Lemma 6.5). The value identity holds for every order parameter, and so does the slope bound, with a constant depending only

on κ . These two results, weak top stability, the pinning of the top, and the collapse of minimizers at κ_c do not appear in these works.

Conventions. Throughout this chapter, $\alpha > 0$ is fixed, and we use the notation of §2.3. For a margin $\kappa \in \mathbb{R}$ and $\gamma \in \mathcal{U}$ we write

$$\begin{aligned} Q &= Q_\gamma, & \delta &= 1 - Q, & \lambda &= \lambda_\gamma, & u &= u_\gamma(\cdot, \cdot; \kappa), \\ b &= \partial_x u, & c &= -\partial_{xx} u, & v &= \lambda c, \end{aligned}$$

where u is computed with the cutoff Q , so that $u(Q, x) = \log \Phi((x - \kappa)/\sqrt{\delta})$ by (2.11). Thus $\lambda(q) = \delta + \int_q^Q \gamma$ for $q \leq Q$, and in particular $\lambda \geq \delta$ on $[0, Q]$. The generator is $\mathcal{L} = \partial_q + \frac{1}{2}\partial_{xx} + \gamma b \partial_x$, the diffusion X solves $dX_q = \gamma(q)b(q, X_q) dq + dB_q$ with $X_0 = 0$, and D and S are the functions in (2.29). For $r < 1$, $\mathcal{U}_r$ is the class of $\gamma \in \mathcal{U}$ with $\gamma = 1$ on $[r, 1]$; for $\gamma \in \mathcal{U}_r$ the functions X , D , and S are defined on $[0, r]$. We write μ_γ for the probability measure on $[0, 1]$ with distribution function γ ; its support lies in $[0, Q]$. When the order parameter or the margin varies, we write $u_{\gamma, \kappa}$, $b_{\gamma, \kappa}$, and so on; thus $b_{\gamma, \kappa}(0, 0) = \partial_x u_\gamma(0, 0; \kappa)$. As in §2.3, $R = \phi/\Phi$, $\bar{R}(z) = z + R(z)$, and $V = R\bar{R}$; we also use

$$h(z) = \log \Phi(z) + \frac{1}{2}R(z)^2, \quad m(k) = \mathbb{E}(k - G)_+^2,$$

where G denotes a standard Gaussian variable independent of B . Finally, the *normalized top position* is

$$Z = \frac{X_Q - \kappa}{\sqrt{\delta}}, \quad \text{so that} \quad u(Q, X_Q) = \log \Phi(Z), \quad b(Q, X_Q) = \frac{R(Z)}{\sqrt{\delta}}.$$

We use the following facts from §2.3. For each $r < 1$ the class $\mathcal{U}_r$ is compact in $L^1([0, 1])$, and $\mathcal{P}_\kappa$ is continuous on $\mathcal{U}_r$, uniformly for κ in compact intervals (Lemma 2.7); in particular $\mathcal{P}_\kappa$ attains its minimum over $\mathcal{U}_r$. By (2.27), $b > 0$ and $0 < v < 1$, so $0 < c < 1/\lambda \leq 1/\delta$, and by (2.28), $b(q, X_q)$ is a square-integrable martingale. The function D is continuously differentiable with $D' = (S - 1)/\lambda^2$ (Lemma 2.9), and the first variation of $\mathcal{P}_\kappa$ on $\mathcal{U}_r$ is (2.30). The martingale property gives

$$\sqrt{\delta} b(0, 0) = \sqrt{\delta} \mathbb{E} b(Q, X_Q) = \mathbb{E} R(Z). \quad (6.1)$$

Step approximation with a fixed top. Some identities below are first proved for step order parameters, for which u is smooth on every interval of constancy of γ and Itô's formula applies there; polynomial growth of u and its derivatives (Lemma 2.6) and the moment bounds for X make every stochastic integral below a true martingale. Let $\gamma \in \mathcal{U}$ with $Q > 0$ and, for $n \geq 1$, put

$$\gamma_n(q) = \gamma(jQ/n) \quad \text{for } q \in [jQ/n, (j+1)Q/n), \quad 0 \leq j < n, \quad \gamma_n = 1 \text{ on } [Q, 1]. \quad (6.2)$$

Then $\gamma_n \in \mathcal{U}_Q$ is a step order parameter with $\gamma_n \leq \gamma$. Its top is exactly Q , because $\gamma < 1$ on $[0, Q)$, and $\|\gamma_n - \gamma\|_{L^1} \leq Q/n$, because γ is nondecreasing with values in $[0, 1]$. By Lemma 2.7 with $r = Q$, the functions u_{γ_n} , b_{γ_n} , and c_{γ_n} converge to u , b , and c locally uniformly on $[0, Q] \times \mathbb{R}$ with bounds of polynomial growth that are uniform in n , the diffusions X^{γ_n} converge to X in every moment, uniformly on $[0, Q]$, and $\mathcal{P}_\kappa(\gamma_n) \rightarrow \mathcal{P}_\kappa(\gamma)$. Since b and c have spatial derivatives of polynomial growth and $0 < c_{\gamma_n}, c < 1/\delta$, it follows that $\mathbb{E}b_{\gamma_n}(q, X_q^{\gamma_n})^2 \rightarrow \mathbb{E}b(q, X_q)^2$ and $\mathbb{E}c_{\gamma_n}(q, X_q^{\gamma_n})^2 \rightarrow \mathbb{E}c(q, X_q)^2$ uniformly in $q \in [0, Q]$. Indeed, for $R > 0$ split each expectation according to whether $|X_q^{\gamma_n}| \vee |X_q| \leq R$. On this event the integrands differ by at most a constant times the sum of $\sup_{[0, Q] \times [-R, R]} (|b_{\gamma_n} - b| + |c_{\gamma_n} - c|)$ and $C_R |X_q^{\gamma_n} - X_q|$, because the functions and their spatial derivatives are bounded on $[0, Q] \times [-R, R]$ uniformly in n . On the complement the integrands are bounded by $C(1 + |X_q^{\gamma_n}| + |X_q|)^p$, whose contribution is at most $C_p R^{-1}$ uniformly in q and n by the uniform moment bounds. Letting $n \rightarrow \infty$ and then $R \rightarrow \infty$ gives the uniform convergence. Since also $\lambda_{\gamma_n} \rightarrow \lambda$ uniformly with $\lambda_{\gamma_n} \geq \delta$, the functions D of γ_n converge to D uniformly on $[0, Q]$. Since the top is fixed, the normalized top positions Z_n of γ_n converge to Z in L^2 ; as $0 < R(z) \leq R(0) + |z| < 1 + |z|$ (the derivative of R lies in $(-1, 0)$ by Lemma 2.5), every function g with $|g'| \leq C(1 + |z|)$, such as $\log \Phi$, R^2 , and h , satisfies $\mathbb{E}g(Z_n) \rightarrow \mathbb{E}g(Z)$.

6.1 A value identity and existence of minimizers

We first show that a finite value bounds the top of every minimizer away from one. The mechanism is an identity (Lemma 6.2) that rewrites the value along the Parisi diffusion. It involves the function h , whose sign is determined in the next lemma. The lemma also collects the other Gaussian facts used in this chapter and in Chapter 7. The bound $V < 1$ in part (i) is Sampford's inequality, as noted after Lemma 2.5.

Lemma 6.1. *The following statements hold.*

- (i) $R' = -V$ and $0 < V < 1$. The function $\bar{R}$ is positive and strictly increasing, R is strictly decreasing, and $R(z) > z_- = \max\{-z, 0\}$.
- (ii) V is strictly decreasing.
- (iii) $V(z)^2 \leq 2R(z)$ for every $z \in \mathbb{R}$.
- (iv) $h' = R(1 - V) > 0$ and $h < 0$.
- (v) With $C_0 = -\log \Phi(-1)$, for every $t \in \mathbb{R}$,

$$-\frac{1}{2}t_+^2 - \log(1 + t_+) - C_0 \leq \log \Phi(-t) \leq -\frac{1}{2}t_+^2. \quad (6.3)$$

- (vi) For every $t \in \mathbb{R}$ and $\theta > 0$, with $\log_+ y = \max\{\log y, 0\}$,

$$\log \Phi(-t) \geq -\frac{1}{2}(t + \theta)_+^2 - \log_+ \frac{1}{\theta} - C_0. \quad (6.4)$$

Proof. The functions R , $\bar{R}$, V , and h are those of Part II, §2.1.

Proof of (i)–(iv). These are contained in Part II, Lemma 2.1. Its part (a) gives $R' = -V$, $0 < V < 1$, $\bar{R} > 0$, $\bar{R}' = 1 - V > 0$, and $R(z) > z_-$, which is (i), since $R' = -V < 0$. Assertion (ii) is the inequality $V' < 0$ in (2.2), and assertions (iii) and (iv) are parts (e) and (d) of that lemma.

Proof of (v). The upper bound holds for $t \leq 0$ because $\Phi \leq 1$, and for $t > 0$ it is the Chernoff bound $\mathbb{P}(G \geq t) \leq e^{-t^2} \mathbb{E}e^{tG} = e^{-t^2/2}$. For the lower bound, if $t \leq 1$ then $\log \Phi(-t) \geq \log \Phi(-1) = -C_0$. If $t \geq 1$, the lower Mills bound of Lemma 2.5(i) and $2t^2 \geq 1 + t^2$ give

$$\Phi(-t) > \frac{t\phi(t)}{1+t^2} \geq \frac{\phi(t)}{2t} \geq \frac{e^{-t^2/2}}{2\sqrt{2\pi}(1+t)},$$

and $\log(2\sqrt{2\pi}) < 1.62 < 1.84 < C_0$.

Proof of (vi). If $t \leq 0$, the right side is at most $-C_0$, and (v) gives $\log \Phi(-t) \geq -C_0$. Let $t > 0$. Since $\frac{1}{2}(t+\theta)^2 \geq \frac{1}{2}t^2 + \theta t$, by (v) it suffices to show $\log(1+t) \leq \theta t + \log_+(1/\theta)$. If $\theta \geq 1$, then $\log(1+t) \leq t \leq \theta t$. If $\theta < 1$, the concave function $t \mapsto \log(1+t) - \theta t$ is maximal at $t = 1/\theta - 1$, where it equals $\log(1/\theta) - 1 + \theta < \log(1/\theta)$. $\square$

The value identity below involves the first-variation function D . At a minimizer over $\mathcal{U}_r$, stationarity says that the antiderivative of $-D/2$ is minimal on the support of μ_γ ; this is the source of the sign information, as in the characterizations of Parisi measures recalled at the start of the chapter. In the proof, (6.7) is the control representation of Auffinger and W.-K. Chen [7, Theorem 3] and of Jagannath and Tobasco [76, Lemma 18], evaluated at its optimal control $b(q, X_q)$.

Lemma 6.2. *For every $\kappa \in \mathbb{R}$ and $\gamma \in \mathcal{U}$,*

$$\mathcal{P}_\kappa(\gamma) = \frac{1}{2} \log \delta + \alpha \mathbb{E}h(Z) - \frac{\delta}{2} D(Q) - \frac{1}{2} \int_0^Q \gamma(q) D(q) \, dq. \quad (6.5)$$

If $r \in [0, 1)$ and γ minimizes $\mathcal{P}_\kappa$ over $\mathcal{U}_r$, then $\int_0^Q \gamma D \, dq = 0$ and $D(Q) \geq 0$, with $D(Q) = 0$ if $Q < r$; consequently

$$\mathcal{P}_\kappa(\gamma) < \frac{1}{2} \log \delta. \quad (6.6)$$

If γ minimizes $\mathcal{P}_\kappa$ over $\mathcal{U}$, then $D(Q) = 0$ and $\mathcal{P}_\kappa(\gamma) = \frac{1}{2} \log \delta + \alpha \mathbb{E}h(Z)$.

Proof. We first prove (6.5). If $Q = 0$, then $\gamma \equiv 1$, $\delta = 1$, $Z = -\kappa$, $u(0, 0) = \log \Phi(-\kappa)$, $\text{CS}(\gamma) = 0$, and $D(0) = \alpha b(0, 0)^2 = \alpha R(-\kappa)^2$, and both sides of (6.5) equal $\alpha \log \Phi(-\kappa)$. Let $Q > 0$. Both sides of (6.5) converge along the approximations (6.2), so we may assume

that γ is a step order parameter. By Itô's formula on each interval of constancy of γ , the equation (1.3) and the continuity of u ,

$$\begin{aligned} du(q, X_q) &= \left(\partial_q u + \frac{1}{2} \partial_{xx} u + \gamma b^2 \right) (q, X_q) dq + b(q, X_q) dB_q \\ &= \frac{1}{2} \gamma(q) b(q, X_q)^2 dq + b(q, X_q) dB_q. \end{aligned}$$

The stochastic integral has mean zero because $b(q, X_q)$ is square integrable. Since $u(Q, X_Q) = \log \Phi(Z)$,

$$u(0, 0) = \mathbb{E} \log \Phi(Z) - \frac{1}{2} \int_0^Q \gamma(q) \mathbb{E} b(q, X_q)^2 dq. \quad (6.7)$$

By (2.29), $\alpha \mathbb{E} b(q, X_q)^2 = D(q) + \int_0^q \lambda(t)^{-2} dt$. Multiplying by $\gamma(q)$, integrating over $[0, Q]$, exchanging the order of integration, and using $\int_t^Q \gamma = \lambda(t) - \delta$, we obtain

$$\alpha \int_0^Q \gamma(q) \mathbb{E} b(q, X_q)^2 dq = \int_0^Q \gamma D dq + \int_0^Q \frac{dt}{\lambda(t)} - \delta \int_0^Q \frac{dt}{\lambda(t)^2}.$$

At the terminal time $b(Q, X_Q) = R(Z) / \sqrt{\delta}$, so the definition of $D(Q)$ gives $\delta \int_0^Q \lambda^{-2} dt = \alpha \mathbb{E} R(Z)^2 - \delta D(Q)$. Substituting these two identities into (6.7) and adding $\text{CS}(\gamma) = \frac{1}{2} \int_0^Q \lambda^{-1} dq + \frac{1}{2} \log \delta$, the integral $\frac{1}{2} \int_0^Q \lambda^{-1}$ cancels and we obtain

$$\mathcal{P}_\kappa(\gamma) = \frac{1}{2} \log \delta + \alpha \mathbb{E} \left[\log \Phi(Z) + \frac{1}{2} R(Z)^2 \right] - \frac{\delta}{2} D(Q) - \frac{1}{2} \int_0^Q \gamma D dq,$$

which is (6.5).

Now let γ minimize over $\mathcal{U}_r$, and write $\mu = \mu_\gamma$. If $r = 0$, then $\gamma \equiv 1$, $Q = 0$, and $D(0) = \alpha b(0, 0)^2 > 0$, and the stationarity assertions hold; so let $r > 0$. Put

$$\mathcal{H}(t) = \mathcal{H}_\gamma(t) = -\frac{1}{2} \int_0^t D(q) dq \quad (t \in [0, r]),$$

so that $\mathcal{H}' = -D/2$. For $t \in [0, r]$ the order parameter $\mathbf{1}_{[t, 1]}$ belongs to $\mathcal{U}_r$, and so does $\gamma + \theta(\mathbf{1}_{[t, 1]} - \gamma)$ for $\theta \in [0, 1]$. Minimality and (2.30) give $\int_0^r (\mathbf{1}_{[t, 1]} - \gamma) D dq \geq 0$. Here $\int_0^r \mathbf{1}_{[t, 1]} D dq = \int_t^r D dq = 2(\mathcal{H}(t) - \mathcal{H}(r))$, while, since $\gamma(q) = \mu([0, q])$, Fubini's theorem gives $\int_0^r \gamma D dq = \int (\int_s^r D(q) dq) \mu(ds) = 2 \int (\mathcal{H}(s) - \mathcal{H}(r)) \mu(ds)$. Hence

$$\mathcal{H}(t) \geq \int \mathcal{H} d\mu \quad (0 \leq t \leq r). \quad (6.8)$$

Since μ is a probability measure on $[0, Q] \subset [0, r]$, (6.8) forces $\int \mathcal{H} d\mu = \min_{[0, r]} \mathcal{H}$, so $\mathcal{H}$ equals its minimum μ -almost everywhere and, being continuous, on $\text{supp } \mu$. The point

Q belongs to $\text{supp } \mu$: if $Q > 0$, then $\gamma(q) < 1$ for $q < Q$, so $\mu((q, Q]) > 0$ for every $q < Q$; if $Q = 0$, then μ is the point mass at 0. Therefore

$$\int_0^Q \gamma D \, dq = \int \left(\int_s^Q D(q) \, dq \right) \mu(ds) = 2 \int (\mathcal{H}(s) - \mathcal{H}(Q)) \mu(ds) = 0.$$

Next, $D(0) = \alpha b(0,0)^2 > 0$ because $X_0 = 0$ and $b > 0$. Since $r > 0$, $\mathcal{H}'(0) = -D(0)/2 < 0$, so 0 is not a minimum point of $\mathcal{H}$ on $[0, r]$, and $0 \notin \text{supp } \mu$; in particular $Q > 0$. Since Q is a minimum point of $\mathcal{H}$ on $[0, r]$, $\mathcal{H}(q) \geq \mathcal{H}(Q)$ for $q < Q$, and the left derivative satisfies $-D(Q)/2 = \mathcal{H}'(Q) \leq 0$; thus $D(Q) \geq 0$. If moreover $Q < r$, then Q is an interior minimum point of $\mathcal{H}$, so $D(Q) = -2\mathcal{H}'(Q) = 0$. This proves the stationarity assertions for every $r \in [0, 1)$. With $h < 0$ from Lemma 6.1(iv), they and (6.5) give $\mathcal{P}_\kappa(\gamma) \leq \frac{1}{2} \log \delta + \alpha \mathbb{E}h(Z) < \frac{1}{2} \log \delta$.

Finally, a minimizer γ over $\mathcal{U}$ belongs to $\mathcal{U}_r$ for every $r \in (Q, 1)$ and minimizes over $\mathcal{U}_r \subset \mathcal{U}$. Since $Q < r$, the preceding paragraph gives $D(Q) = 0$, and (6.5) with $\int_0^Q \gamma D \, dq = 0$ gives the last assertion. $\square$

We record for later use two consequences obtained in the proof. If γ minimizes $\mathcal{P}_\kappa$ over $\mathcal{U}_r$ with $r > 0$, and $\mathcal{H}$ is as in the proof (with $\mathcal{H}' = -D/2$), then

$$\text{supp } \mu_\gamma \subset \underset{[0, r]}{\text{argmin}} \mathcal{H}, \quad 0 \notin \text{supp } \mu_\gamma. \quad (6.9)$$

For comparison, the order parameter $\gamma \equiv 1$ has $Q = 0$, $\delta = 1$, $u(0, 0) = \log \Phi(-\kappa)$ and $\text{CS}(\gamma) = 0$. Hence

$$\mathcal{P}_*(\kappa) \leq \mathcal{P}_\kappa(1) = \alpha \log \Phi(-\kappa) < 0 \quad (\kappa \in \mathbb{R}). \quad (6.10)$$

For $r \in [0, 1)$ we write

$$\mathcal{P}_*^r(\kappa) = \min_{\gamma \in \mathcal{U}_r} \mathcal{P}_\kappa(\gamma) \quad (6.11)$$

for the constrained minimum. Since the classes $\mathcal{U}_r$ increase with r and their union is $\mathcal{U}$, $\mathcal{P}_*^r(\kappa)$ is nonincreasing in r and decreases to $\mathcal{P}_*(\kappa)$ as $r \uparrow 1$. Also $\mathcal{P}_*^0(\kappa) = \alpha \log \Phi(-\kappa)$, since $\mathcal{U}_0 = \{1\}$.

Theorem 6.3. *Let $\kappa \in \mathbb{R}$ with $\mathcal{P}_*(\kappa) > -\infty$. Then the infimum defining $\mathcal{P}_*(\kappa)$ is attained. Every minimizer γ satisfies*

$$1 - Q_\gamma > e^{2\mathcal{P}_*(\kappa)},$$

that is, every minimizer lies in $\mathcal{U}_r$ with $r = 1 - e^{2\mathcal{P}_*(\kappa)} \in (0, 1)$, and the set of minimizers is compact in $L^1([0, 1])$.

Proof. By (6.10), $-\infty < \mathcal{P}_*(\kappa) < 0$, so $r_* = 1 - e^{2\mathcal{P}_*(\kappa)}$ lies in $(0, 1)$. For $r \in [0, 1)$ let γ_r minimize $\mathcal{P}_\kappa$ over $\mathcal{U}_r$, with top Q_r and $\delta_r = 1 - Q_r$. By (6.6),

$$\frac{1}{2} \log \delta_r > \mathcal{P}_\kappa(\gamma_r) \geq \mathcal{P}_*(\kappa),$$

so $Q_r < r_*$, and every γ_r belongs to the single compact class $\mathcal{U}_{r_*}$. Hence $\mathcal{P}_*^r(\kappa) \geq \mathcal{P}_*^{r_*}(\kappa)$ for every $r < 1$, and letting $r \uparrow 1$ gives $\mathcal{P}_*(\kappa) = \mathcal{P}_*^{r_*}(\kappa)$, which is attained.

Every minimizer γ over $\mathcal{U}$ belongs to $\mathcal{U}_r$ for $r \in (Q_\gamma, 1)$ and minimizes over this $\mathcal{U}_r$, so $\frac{1}{2} \log(1 - Q_\gamma) > \mathcal{P}_\kappa(\gamma) = \mathcal{P}_*(\kappa)$ by Lemma 6.2, which is the asserted bound; in particular $\gamma \in \mathcal{U}_{r_*}$. The set of minimizers is $\{\gamma \in \mathcal{U}_{r_*} : \mathcal{P}_\kappa(\gamma) = \mathcal{P}_*(\kappa)\}$, which is closed in the compact class $\mathcal{U}_{r_*}$ by continuity, hence compact. $\square$

The theorem concerns minimizers, and (6.6) gives the same bound for constrained minimizers. Neither says that the tops of an arbitrary minimizing sequence stay away from one: adding a vanishing mass near one changes the top without changing the L^1 limit. Like Lemma 5.2, the theorem is the analogue for $\mathcal{P}_\kappa$ of the fact about spherical Parisi measures recalled before that lemma: they are supported in an interval $[0, q_1]$ with $q_1 < 1$. Here the bound on the top comes from the value identity, and it depends on $\mathcal{P}_*(\kappa)$.

6.2 The slope in the margin

To move values between margins we need a bound on how fast $\mathcal{P}_\kappa(\gamma)$ changes with κ . The following bound holds for every order parameter, which is what makes it usable for near-minimizers and for constrained minimizers alike.

Lemma 6.4. *Let $C_\kappa = \frac{1}{2}(\kappa + \sqrt{\kappa^2 + 4})$. For every $\gamma \in \mathcal{U}$ and $\kappa \in \mathbb{R}$,*

$$\mathbb{E}(\kappa - B_Q)_+ \leq \lambda(0) b(0, 0) < C_\kappa. \quad (6.12)$$

The map $\kappa \mapsto \mathcal{P}_\kappa(\gamma)$ is twice continuously differentiable, with $\partial_\kappa \mathcal{P}_\kappa(\gamma) = -\alpha b(0, 0) < 0$ and $\partial_\kappa^2 \mathcal{P}_\kappa(\gamma) = -\alpha c(0, 0) < 0$; in particular it is strictly decreasing and strictly concave. Consequently $\mathcal{P}_*$, as a function with values in $[-\infty, 0)$, is nonincreasing and concave, and $\kappa \mapsto b_{\gamma, \kappa}(0, 0)$ is increasing.

Proof. The upper bound comes from a maximum principle for an explicit combination of b and c . Put $\tau = 1 - q$ and

$$F = \tau c - \lambda b^2 - (x - \kappa)b$$

on $[0, Q] \times \mathbb{R}$. At $q = Q$ we have $\tau = \lambda = \delta$ and, with $z = (x - \kappa)/\sqrt{\delta}$, $b(Q, x) = R(z)/\sqrt{\delta}$ and $c(Q, x) = -R'(z)/\delta = V(z)/\delta$. Thus

$$F(Q, x) = V(z) - R(z)^2 - zR(z) = V(z) - R(z)\bar{R}(z) = 0.$$

If $Q = 0$ this gives $F(0, 0) = 0$. Let $Q > 0$, and first let γ be a step order parameter. For smooth f, g we have $\mathcal{L}(fg) = f\mathcal{L}g + g\mathcal{L}f + \partial_x f \partial_x g$. On each interval of constancy of γ ,

Lemma 2.6(ii) gives $\mathcal{L}b = 0$ and $\mathcal{L}c = \gamma c^2$; together with $\partial_x b = -c$, $\mathcal{L}\tau = -1$, $\lambda' = -\gamma$, and $\mathcal{L}(x - \kappa) = \gamma b$ we obtain

$$\mathcal{L}(\tau c) = -c + \gamma \tau c^2, \quad \mathcal{L}(\lambda b^2) = -\gamma b^2 + \lambda c^2, \quad \mathcal{L}((x - \kappa)b) = \gamma b^2 - c.$$

For instance, $\mathcal{L}(b^2) = 2b\mathcal{L}b + (\partial_x b)^2 = c^2$, and $\mathcal{L}((x - \kappa)b) = (x - \kappa)\mathcal{L}b + b\mathcal{L}(x - \kappa) + \partial_x b$. Hence

$$\mathcal{L}F = (\gamma\tau - \lambda)c^2 \leq 0,$$

because γ is nondecreasing and therefore $\lambda(q) = \int_q^1 \gamma \geq \gamma(q)(1 - q)$. Itô's formula on each interval of constancy gives $F(0, 0) = \mathbb{E}F(Q, X_Q) - \mathbb{E} \int_0^Q \mathcal{L}F(q, X_q) dq \geq 0$. For general γ the inequality $F(0, 0) \geq 0$ follows from the step approximations (6.2), since $b(0, 0)$, $c(0, 0)$ and $\lambda(0)$ converge along them. At $(0, 0)$ this inequality reads $c(0, 0) \geq \lambda(0)b(0, 0)^2 - \kappa b(0, 0)$, and $\lambda(0)c(0, 0) = v(0, 0) < 1$. Multiplying by $\lambda(0) > 0$, the number $y = \lambda(0)b(0, 0) > 0$ satisfies $y^2 - \kappa y < 1$. Hence y lies below the positive root C_κ of $y^2 - \kappa y - 1$.

For the lower bound write $X_Q = B_Q + \Delta$ with $\Delta = \int_0^Q \gamma(q)b(q, X_q) dq \geq 0$. Since $b(q, X_q)$ is a martingale, $\mathbb{E}\Delta = b(0, 0) \int_0^Q \gamma = b(0, 0)(\lambda(0) - \delta)$. Since $(\kappa - X_Q)_+ = \sqrt{\delta} Z_-$ and $R(z) \geq z_-$ (Lemma 6.1(i)), (6.1) gives $\mathbb{E}(\kappa - X_Q)_+ \leq \sqrt{\delta} \mathbb{E}R(Z) = \delta b(0, 0)$. Finally $(\kappa - B_Q)_+ \leq (\kappa - X_Q)_+ + \Delta$, and adding the two expectations gives $\mathbb{E}(\kappa - B_Q)_+ \leq \lambda(0)b(0, 0)$.

The terminal datum depends on κ only through $x - \kappa$, and the recursion (2.2) commutes with translations in x , so $u_{\gamma, \kappa}(q, x) = u_{\gamma, 0}(q, x - \kappa)$. By Lemma 2.6(v), which holds for every $\gamma \in \mathcal{U}_r$ by Lemma 2.7, $u_{\gamma, 0}(0, \cdot)$ has continuous derivatives of all orders. Therefore $\kappa \mapsto u_{\gamma, \kappa}(0, 0)$ is twice continuously differentiable with $\partial_\kappa u_{\gamma, \kappa}(0, 0) = -b(0, 0)$ and $\partial_\kappa^2 u_{\gamma, \kappa}(0, 0) = -c(0, 0)$, and $\partial_\kappa b_{\gamma, \kappa}(0, 0) = c(0, 0) > 0$. Since $\text{CS}(\gamma)$ does not depend on κ , $\partial_\kappa \mathcal{P}_\kappa(\gamma) = -\alpha b(0, 0) < 0$ and $\partial_\kappa^2 \mathcal{P}_\kappa(\gamma) = -\alpha c(0, 0) < 0$. An infimum of nonincreasing concave functions is nonincreasing and concave. $\square$

We use the lemma mostly through the tangent-line inequality: by concavity, for every $\gamma \in \mathcal{U}$ and all margins k, k' ,

$$\mathcal{P}_{k'}(\gamma) \leq \mathcal{P}_k(\gamma) - \alpha b_{\gamma, k}(0, 0) (k' - k). \quad (6.13)$$

Recall from (1.6) that $\kappa_c = \sup\{\kappa : \mathcal{P}_*(\kappa) > -\infty\}$. Since $\mathcal{P}_*$ is nonincreasing, $\mathcal{P}_*(\kappa) > -\infty$ for $\kappa < \kappa_c$ and $\mathcal{P}_*(\kappa) = -\infty$ for $\kappa > \kappa_c$. Proposition 6.7 shows that κ_c is finite, and Theorem 6.8 shows that $\mathcal{P}_*(\kappa_c) = -\infty$.

6.3 Weak top stability and finiteness of the critical margin

To show that $\kappa_c > -\infty$ we must find margins with $\mathcal{P}_*(\kappa) > -\infty$. For an order parameter with a single support point the value is explicit, so it suffices to show that, at very negative

margins, constrained minimizers have a single support point. The next lemma gives a quantitative obstruction to a second support point, namely a lower bound on $S(Q)$, and a lower bound on the slope in terms of $S(Q)$. Together with the slope bound, these two bounds give $\lambda(0) \leq C\sqrt{\delta}$ for a minimizer; this is used again in Proposition 6.11. The proof follows the endpoint analysis of Auffinger and W.-K. Chen [8, Proposition 3 and proof of Theorem 4]: when support points accumulate at Q , Rolle's theorem gives $S(Q) = 1$. The differences are that both terms in (6.16) are nonnegative, because $c > 0$, while the corresponding cubic term in [8, Proposition 3] enters with a negative sign, and that an isolated top is handled by the quantitative inequality (6.18).

Lemma 6.5. *Let $\kappa \in \mathbb{R}$ and $r \in [0, 1)$, and let γ minimize $\mathcal{P}_\kappa$ over $\mathcal{U}_r$. If μ_γ has more than one support point, then $\alpha > 1$ and*

$$S(Q) \geq s_\alpha := \min \left\{ \frac{1}{2}, \frac{1}{16\alpha} \right\}.$$

Moreover, every $\gamma \in \mathcal{U}$ satisfies, at every margin κ ,

$$\sqrt{\delta} b(0, 0) \geq \frac{S(Q)}{2\alpha}. \quad (6.14)$$

Consequently, every minimizer γ of $\mathcal{P}_\kappa$ over $\mathcal{U}_r$ satisfies

$$\lambda(0) \leq \max \left\{ 1, \frac{2\alpha C_\kappa}{s_\alpha} \right\} \sqrt{\delta}. \quad (6.15)$$

Proof. We first show that S/λ^2 is nondecreasing on $[0, r]$ and that $S < \alpha$, for every $\gamma \in \mathcal{U}_r$. Let first γ be a step order parameter. On each interval of constancy of γ , $\mathcal{L}c = \gamma c^2$ gives $\mathcal{L}(c^2) = 2c \mathcal{L}c + (\partial_x c)^2 = 2\gamma c^3 + (\partial_x c)^2$, so Itô's formula gives

$$\frac{d}{dq} \mathbb{E} c(q, X_q)^2 = \mathbb{E} [(\partial_x c)^2 + 2\gamma c^3](q, X_q) \geq 0 \quad (6.16)$$

there, because $c > 0$. The function $q \mapsto \mathbb{E} c(q, X_q)^2$ is continuous, because c is continuous and bounded by $(1-r)^{-1}$, so (6.16) shows that it is nondecreasing. For general $\gamma \in \mathcal{U}_r$ take step order parameters $\gamma_n \in \mathcal{U}_r$ with $\gamma_n \rightarrow \gamma$ in L^1 . By Lemma 2.7, $c_{\gamma_n} \rightarrow c$ locally uniformly on $[0, r] \times \mathbb{R}$ and $X_q^{\gamma_n} \rightarrow X_q$ in L^2 ; since $0 < c_{\gamma_n} < (1-r)^{-1}$, dominated convergence gives $\mathbb{E} c_{\gamma_n}(q, X_q^{\gamma_n})^2 \rightarrow \mathbb{E} c(q, X_q)^2$ for each q , and a pointwise limit of nondecreasing functions is nondecreasing. Since $S(q)/\lambda(q)^2 = \alpha \mathbb{E} c(q, X_q)^2$, the ratio S/λ^2 is nondecreasing. Also $S = \alpha \mathbb{E} v(q, X_q)^2 < \alpha$ because $0 < v < 1$. Recall that D is continuously differentiable with $D' = (S-1)/\lambda^2$; in particular S is continuous.

Now let γ minimize over $\mathcal{U}_r$, and suppose that $\mu = \mu_\gamma$ has more than one support point. Then $\gamma \neq 1$, so $r > 0$, and we use (6.9) with $\mathcal{H}$ as in the proof of Lemma 6.2. The support of μ contains Q and a second point, which lies in $(0, Q)$ because $0 \notin \text{supp } \mu$. Every

$p \in \text{supp } \mu \cap (0, Q)$ is an interior minimum point of $\mathcal{H}$ on $[0, r]$, so $D(p) = -2\mathcal{H}'(p) = 0$. By Lemma 6.2, $D(Q) \geq 0$; this includes the case $Q = r$ of an active cutoff. One of two cases occurs.

Suppose first that there are $p_n \in \text{supp } \mu$ with $0 < p_n < p_{n+1} < Q$ and $p_n \uparrow Q$. By Rolle's theorem there are $\theta_n \in (p_n, p_{n+1})$ with $D'(\theta_n) = 0$, that is, $S(\theta_n) = 1$. Continuity of S gives $S(Q) = 1 \geq s_\alpha$, and $1 = S(\theta_n) < \alpha$.

Otherwise the closed set $\text{supp } \mu \cap [0, Q)$ is nonempty and contained in $[0, Q - \varepsilon]$ for some $\varepsilon > 0$, so $p = \max(\text{supp } \mu \cap [0, Q))$ exists, $0 < p < Q$, and $D(p) = 0$. Since $\mu((p, Q)) = 0$, γ is constant on $[p, Q)$, equal to $\gamma(p) = \mu([0, p])$, and $\gamma(p) > 0$ because every neighborhood of $p \in \text{supp } \mu$ has positive mass while $\mu((p, Q)) = 0$. Then

$$0 \leq D(Q) - D(p) = \int_p^Q \frac{S(q) - 1}{\lambda(q)^2} dq. \quad (6.17)$$

If $\alpha \leq 1$, the integrand is negative because $S < \alpha$, which is impossible; so $\alpha > 1$. Put

$$\sigma = S(Q), \quad \lambda_0 = \lambda(Q) = \delta, \quad \lambda_1 = \lambda(p) = \lambda_0 + \gamma(p)(Q - p) > \lambda_0.$$

On $[p, Q]$ the function λ is affine with $\lambda' = -\gamma(p)$. We change variables to $\ell = \lambda(q)$, so that $d\ell = -\gamma(p) dq$, and bound $S(q)/\lambda(q)^2$ in two ways: by $S(Q)/\lambda(Q)^2 = \sigma/\lambda_0^2$, since S/λ^2 is nondecreasing, and by α/ℓ^2 , since $S < \alpha$. Multiplying (6.17) by $\gamma(p)$ we obtain

$$\frac{1}{\lambda_0} - \frac{1}{\lambda_1} \leq \int_{\lambda_0}^{\lambda_1} \min\left(\frac{\sigma}{\lambda^2}, \frac{\alpha}{\ell^2}\right) d\ell. \quad (6.18)$$

If $\lambda_1 \leq 2\lambda_0$, the left side of (6.18) is $(\lambda_1 - \lambda_0)/(\lambda_0\lambda_1) \geq (\lambda_1 - \lambda_0)/(2\lambda_0^2)$ and the right side is at most $\sigma(\lambda_1 - \lambda_0)/\lambda_0^2$, so $\sigma \geq \frac{1}{2}$. If $\lambda_1 > 2\lambda_0$, the left side is at least $1/(2\lambda_0)$, and the right side is at most the integral over $(0, \infty)$. Since $\sigma > 0$ (because $c > 0$), the two terms of the minimum agree at $\ell_* = \lambda_0\sqrt{\alpha/\sigma}$, so this integral equals $\sigma\ell_*/\lambda_0^2 + \alpha/\ell_* = 2\sqrt{\alpha\sigma}/\lambda_0$. Hence $\frac{1}{2} \leq 2\sqrt{\alpha\sigma}$, that is, $\sigma \geq 1/(16\alpha)$. Both cases give $S(Q) \geq s_\alpha$.

For (6.14), let $\gamma \in \mathcal{U}$. Since $\lambda(Q) = \delta$ and $c(Q, x) = V((x - \kappa)/\sqrt{\delta})/\delta$,

$$S(Q) = \alpha \mathbb{E} V(Z)^2. \quad (6.19)$$

Lemma 6.1(iii) and (6.1) give $S(Q) \leq 2\alpha \mathbb{E} R(Z) = 2\alpha\sqrt{\delta} b(0, 0)$.

Finally, let γ minimize over $\mathcal{U}_r$. If μ_γ has a single support point, then $\gamma = \mathbf{1}_{[Q, 1]}$ and $\lambda(0) = \delta \leq \sqrt{\delta}$. Otherwise the first assertion and (6.14) give $\sqrt{\delta} b(0, 0) \geq s_\alpha/(2\alpha)$, and Lemma 6.4 gives $\lambda(0) < C_\kappa/b(0, 0) \leq 2\alpha C_\kappa\sqrt{\delta}/s_\alpha$. This proves (6.15). $\square$

In the physics literature, the conditions $D = 0$ and $S = 1$ appear as the variational equation for the order parameter and its derivative in [58], where $S = 1$ on a continuous part of the support is interpreted as marginal stability. For the replica-symmetric order

parameter $\mathbf{1}_{[Q,1]}$ with $D(Q) = 0$, the condition $S(Q) = 1$, with $S(Q)$ as in (6.19), defines their de Almeida–Thouless line, the analogue of the instability found in [45] for the Sherrington–Kirkpatrick model. The replica-symmetric phase at negative margins is the region below this line in their phase diagram.

For $q \in [0, 1)$ the order parameter $\mathbf{1}_{[q,1]}$, which is the distribution function of the point mass at q , has top q and $\lambda = 1 - q$ on $[0, q]$. On $[0, q)$ it vanishes, so u solves the heat equation there and $X = B$. Its value is therefore

$$\mathcal{P}_\kappa^{\text{RS}}(q) := \mathcal{P}_\kappa(\mathbf{1}_{[q,1]}) = \alpha \mathbb{E} \log \Phi\left(\frac{\sqrt{q}G - \kappa}{\sqrt{1-q}}\right) + \frac{q}{2(1-q)} + \frac{1}{2} \log(1-q). \quad (6.20)$$

These are the replica-symmetric order parameters. The next lemma shows that at very negative margins every constrained minimizer is of this form, with values bounded below. Jagannath and Tobasco used first-order conditions in a similar way to prove replica symmetry in parts of the phase diagram of mixed p -spin models [77, Theorem 1.9]. Their argument checks a sufficient condition for optimality, which relies on convexity; here we instead exclude a second support point for every constrained minimizer.

Lemma 6.6. *There is $\kappa_0 = \kappa_0(\alpha) \leq -1$ such that for every $\kappa \leq \kappa_0$ and every $r \in [0, 1)$, every minimizer of $\mathcal{P}_\kappa$ over $\mathcal{U}_r$ equals $\mathbf{1}_{[q,1]}$ for some $q \in [0, r]$, and*

$$\mathcal{P}_\kappa^{\text{RS}}(q) \geq -\frac{1}{4} - \alpha C_0 \quad (0 \leq q < 1). \quad (6.21)$$

Proof. For $z \geq 0$ we have $V(z) \leq 2\phi(z)(z+1)$ by Part II, Lemma 2.1(e), so $V(z) \rightarrow 0$ as $z \rightarrow \infty$. Since V is decreasing (Lemma 6.1(ii)), we can choose $z_0 \geq 1$ with $\alpha V(z)^2 \leq \alpha V(z_0)^2 < s_\alpha/2$ for all $z \geq z_0$. Since Φ and m are increasing, we can then choose $\kappa_0 \leq -z_0$ with $\alpha\Phi(\kappa_0 + z_0) < s_\alpha/2$ and $\alpha m(\kappa_0 + 1) \leq \frac{1}{2}$. Let $\kappa \leq \kappa_0$; then these three inequalities hold with κ in place of κ_0 .

Let $\gamma \in \mathcal{U}$ with $Q > 0$. Since $\gamma \geq 0$ and $b > 0$, $X_Q = B_Q + \int_0^Q \gamma b \, dq \geq B_Q$. Using (6.19), the bound $V < 1$ on $\{Z < z_0\}$, and $\alpha V^2 < s_\alpha/2$ on $\{Z \geq z_0\}$, we get

$$S(Q) \leq \alpha \mathbb{P}(Z < z_0) + \frac{s_\alpha}{2} \leq \alpha \mathbb{P}(B_Q < \kappa + z_0\sqrt{\delta}) + \frac{s_\alpha}{2} \leq \alpha \Phi(\kappa + z_0) + \frac{s_\alpha}{2} < s_\alpha.$$

Here $\mathbb{P}(B_Q < \kappa + z_0\sqrt{\delta}) \leq \mathbb{P}(B_Q < \kappa + z_0) = \Phi((\kappa + z_0)/\sqrt{Q})$ because $z_0\sqrt{\delta} \leq z_0$, and $(\kappa + z_0)/\sqrt{Q} \leq \kappa + z_0$ because $\kappa + z_0 \leq 0$ and $\sqrt{Q} \leq 1$. An order parameter with $Q = 0$ is the point mass at 0. Hence, by Lemma 6.5, for every $r \in [0, 1)$ each minimizer of $\mathcal{P}_\kappa$ over $\mathcal{U}_r$ has a single support point, that is, it equals $\mathbf{1}_{[q,1]}$ for some $q \in [0, r]$.

To bound (6.20) below, apply (6.4) with $t = (\kappa - \sqrt{q}G)/\sqrt{1-q}$ and $\theta = 1/\sqrt{1-q} \geq 1$:

$$\log \Phi\left(\frac{\sqrt{q}G - \kappa}{\sqrt{1-q}}\right) \geq -\frac{(\kappa + 1 - \sqrt{q}G)_+^2}{2(1-q)} - C_0.$$

Since $\kappa + 1 \leq 0$ and $\sqrt{q} \leq 1$, for $q > 0$ we have $\mathbb{E}(\kappa + 1 - \sqrt{q} G)_+^2 = q m((\kappa + 1)/\sqrt{q}) \leq q m(\kappa + 1)$, and this also holds for $q = 0$. With $\alpha m(\kappa + 1) \leq \frac{1}{2}$ and $\delta = 1 - q$ we obtain

$$\mathcal{P}_\kappa^{\text{RS}}(q) \geq \frac{q}{4\delta} + \frac{1}{2} \log \delta - \alpha C_0 = \frac{1}{4} \left(\frac{1}{\delta} - 2 \log \frac{1}{\delta} \right) - \frac{1}{4} - \alpha C_0 \geq -\frac{1}{4} - \alpha C_0,$$

because $y \geq 2 \log y$ for all $y > 0$. □

Proposition 6.7. *For every $\alpha > 0$,*

$$-\infty < \kappa_c(\alpha) \leq \kappa_{\text{RS}}(\alpha) < \infty,$$

where $\kappa_{\text{RS}}(\alpha)$ is the unique solution of $\alpha m(\kappa_{\text{RS}}) = 1$. In particular $\kappa_c(\alpha) < 0$ if $\alpha > 2$.

Proof. The function m is continuous and strictly increasing, since $m'(k) = 2\mathbb{E}(k - G)_+ > 0$, and it tends to 0 as $k \rightarrow -\infty$ and to ∞ as $k \rightarrow \infty$. Hence κ_{RS} exists, is unique, and is finite.

The upper bound. Let $\alpha m(\kappa) > 1$. The upper bound in (6.3) gives

$$\mathcal{P}_\kappa^{\text{RS}}(q) \leq \frac{q - \alpha \mathbb{E}(\kappa - \sqrt{q} G)_+^2}{2(1 - q)} + \frac{1}{2} \log(1 - q).$$

As $q \uparrow 1$ the numerator tends to $1 - \alpha m(\kappa) < 0$ by dominated convergence, so the right side tends to $-\infty$. Hence $\mathcal{P}_*(\kappa) = -\infty$ whenever $\alpha m(\kappa) > 1$, and $\kappa_c \leq \kappa_{\text{RS}}$.

The lower bound. Let κ_0 be as in Lemma 6.6. By that lemma and (6.21), $\mathcal{P}_*^r(\kappa_0) \geq \inf_{q \in [0,1]} \mathcal{P}_{\kappa_0}^{\text{RS}}(q) \geq -\frac{1}{4} - \alpha C_0$ for every $r < 1$. Letting $r \uparrow 1$ gives $\mathcal{P}_*(\kappa_0) > -\infty$, hence $\kappa_c \geq \kappa_0 > -\infty$.

Finally $m(0) = \mathbb{E}G_-^2 = \frac{1}{2}$, so $\alpha m(0) > 1$ if $\alpha > 2$. Since m is increasing and $\alpha m(\kappa_{\text{RS}}) = 1$, this gives $\kappa_c \leq \kappa_{\text{RS}} < 0$. □

El Alaoui and Sellke observed the bound $\kappa_c \leq \kappa_{\text{RS}}$, stated in terms of the constraint density, in the remark after [54, Conjecture 1]; their argument is the one above, with $\gamma = \mathbf{1}_{[q,1]}$ and $q \rightarrow 1$. For $\alpha > 2$, combined with Theorem 1.2(ii), the bound $\kappa_c \leq \kappa_{\text{RS}}$ recovers the replica-symmetric upper bound on the capacity, which Stojnic proved for every margin [118, Section 2.3].

6.4 Divergence at the critical margin

Since κ_c is finite, we can approach it from both sides. Above κ_c the value is $-\infty$, but the constrained minima $\mathcal{P}_*^r(k)$ of (6.11) take every sufficiently negative finite value. The value inequality (6.6) and the slope bound transport such a value down to $\kappa_c - \eta$.

Theorem 6.8. *For every $\alpha > 0$ the function $\mathcal{P}_*$ is finite, nonincreasing, concave, and continuous on $(-\infty, \kappa_c)$, and $\mathcal{P}_*(\kappa) = -\infty$ for $\kappa \geq \kappa_c$. More precisely, with $C_* = C_{\kappa_c}$, for all sufficiently small $\eta > 0$,*

$$\mathcal{P}_*(\kappa_c - \eta) \leq \frac{1}{2} \log \eta + \frac{1}{2} \log(2\alpha C_*) + \frac{1}{2}. \quad (6.22)$$

Proof. Let $\kappa < \kappa_c$. By the definition of κ_c as a supremum there is $\kappa' > \kappa$ with $\mathcal{P}_*(\kappa') > -\infty$, and monotonicity (Lemma 6.4) gives $\mathcal{P}_*(\kappa) \geq \mathcal{P}_*(\kappa') > -\infty$; also $\mathcal{P}_*(\kappa) < 0$ by (6.10). Thus $\mathcal{P}_*$ is finite on the open interval $(-\infty, \kappa_c)$, which is nonempty by Proposition 6.7, and it is nonincreasing and concave there by Lemma 6.4; a finite concave function on an open interval is continuous. For $\kappa > \kappa_c$, $\mathcal{P}_*(\kappa) = -\infty$ by the definition of κ_c .

Fix $k > \kappa_c$. We first show that $r \mapsto \mathcal{P}_*^r(k)$ is continuous on $[0, 1]$. Let $r < r' \leq r'' < 1$ and let γ' minimize $\mathcal{P}_k$ over $\mathcal{U}_{r'}$. The order parameter $\max\{\gamma', \mathbf{1}_{[r,1]}\}$ belongs to $\mathcal{U}_r$ and is at L^1 distance at most $r' - r$ from γ' . If ω is a modulus of continuity of $\mathcal{P}_k$ on the compact class $\mathcal{U}_{r''}$, then $\mathcal{P}_*^{r'}(k) \leq \mathcal{P}_*^r(k) \leq \mathcal{P}_*^{r''}(k) + \omega(r' - r)$. This proves continuity on $[0, r'']$ for every $r'' < 1$.

Moreover $\mathcal{P}_*^0(k) = \alpha \log \Phi(-k) < 0$, and $\mathcal{P}_*^r(k) \downarrow \mathcal{P}_*(k) = -\infty$ as $r \uparrow 1$. By the intermediate value theorem, every $w < \alpha \log \Phi(-k)$ equals $\mathcal{P}_*^r(k)$ for some $r \in [0, 1]$. Let γ minimize $\mathcal{P}_k$ over this $\mathcal{U}_r$, so that $\mathcal{P}_k(\gamma) = w$. By (6.6), $1 - Q_\gamma > e^{2w}$, and hence $\lambda_\gamma(0) \geq 1 - Q_\gamma > e^{2w}$. For $k' < k$, the tangent-line bound (6.13) and the slope bound (6.12) give

$$\mathcal{P}_*(k') \leq \mathcal{P}_{k'}(\gamma) \leq w + \alpha(k - k') b_{\gamma,k}(0, 0) \leq w + \alpha(k - k') C_k e^{-2w}.$$

Now let $\eta > 0$ and $w < \alpha \log \Phi(-\kappa_c)$. The function $k \mapsto \alpha \log \Phi(-k)$ is continuous, so $w < \alpha \log \Phi(-k)$ for all $k > \kappa_c$ sufficiently close to κ_c , and for such k the bound holds with $k' = \kappa_c - \eta$. Letting $k \downarrow \kappa_c$ and using the continuity of $k \mapsto C_k$, we obtain

$$\mathcal{P}_*(\kappa_c - \eta) \leq w + \alpha \eta C_* e^{-2w}.$$

The choice $w = \frac{1}{2} \log(2\alpha C_* \eta)$ makes the last term equal to $\frac{1}{2}$, and it is admissible when $\frac{1}{2} \log(2\alpha C_* \eta) < \alpha \log \Phi(-\kappa_c)$, that is, for all sufficiently small η . This proves (6.22). Since $\mathcal{P}_*$ is nonincreasing, $\mathcal{P}_*(\kappa_c) \leq \mathcal{P}_*(\kappa_c - \eta)$ for every such η , so $\mathcal{P}_*(\kappa_c) = -\infty$. $\square$

We return to the deep margins of Lemma 6.6. There the variational problem is replica symmetric: the infimum is a minimum over the one-parameter family (6.20).

Corollary 6.9. *Let $\kappa_0 = \kappa_0(\alpha)$ be as in Lemma 6.6, and let $\kappa \leq \kappa_0$. Then $\kappa < \kappa_c$, and every minimizer of $\mathcal{P}_\kappa$ over $\mathcal{U}$ equals $\mathbf{1}_{[q,1]}$ for some $q \in (0, 1)$. In particular*

$$\mathcal{P}_*(\kappa) = \min_{q \in [0,1]} \mathcal{P}_\kappa^{\text{RS}}(q) > -\infty.$$

If $\alpha \leq 1$, the same holds for every $\kappa < \kappa_c$.

Proof. By Lemma 6.6 and the proof of Proposition 6.7, $\mathcal{P}_*(\kappa) \geq \mathcal{P}_*(\kappa_0) > -\infty$, so minimizers exist by Theorem 6.3, and $\kappa < \kappa_c$ because $\mathcal{P}_*(\kappa_c) = -\infty$ by Theorem 6.8. A minimizer γ minimizes $\mathcal{P}_\kappa$ over $\mathcal{U}_r$ for every $r \in (Q, 1)$, so $\gamma = \mathbf{1}_{[Q, 1]}$ by Lemma 6.6, and $Q > 0$ by (6.9). Hence $\mathcal{P}_*(\kappa) = \mathcal{P}_\kappa^{\text{RS}}(Q) \geq \inf_q \mathcal{P}_\kappa^{\text{RS}}(q) \geq \mathcal{P}_*(\kappa)$, and the infimum is attained at $q = Q$. If $\alpha \leq 1$ and $\kappa < \kappa_c$, minimizers exist by Theorem 6.3, and Lemma 6.5 shows directly that each of them has a single support point. $\square$

6.5 Pinning of the top and collapse of minimizers

By (6.6), a finite value keeps the top of a minimizer away from one. For a minimizer over all of $\mathcal{U}$ the value identity is exact, and it also gives the converse: a very negative value forces the top close to one. The reason is that the terminal term $\alpha \mathbb{E}h(Z)$ is only logarithmically negative, since $h(z) \geq -\log(1 + z_-) - C_0$, and the slope bound controls the negative part of Z .

Lemma 6.10. *Let $\kappa \in \mathbb{R}$ with $\mathcal{P}_*(\kappa) > -\infty$, let γ minimize $\mathcal{P}_\kappa$ over $\mathcal{U}$, and put $\delta = 1 - Q_\gamma$. Then*

$$\frac{1+\alpha}{2} \log \delta - \alpha(C_0 + \log(1 + C_\kappa)) \leq \mathcal{P}_*(\kappa) < \frac{1}{2} \log \delta, \quad (6.23)$$

with $C_0 = -\log \Phi(-1)$ as in Lemma 6.1 and C_κ as in Lemma 6.4.

Proof. The upper bound in (6.23) is (6.6). For the lower bound, the last assertion of Lemma 6.2 gives $\mathcal{P}_*(\kappa) = \mathcal{P}_\kappa(\gamma) = \frac{1}{2} \log \delta + \alpha \mathbb{E}h(Z)$. We first show that

$$h(z) \geq -\log(1 + z_-) - C_0 \quad (z \in \mathbb{R}).$$

For $z \geq 0$ this holds because $h(z) \geq \log \Phi(z) \geq \log \frac{1}{2} > -C_0$. For $z = -t < 0$, (6.3) gives $\log \Phi(-t) \geq -\frac{1}{2}t^2 - \log(1 + t) - C_0$, and $R(-t) > t$ by Lemma 6.1(i), so $\frac{1}{2}R(z)^2 \geq \frac{1}{2}t^2$. Since $x \mapsto -\log(1 + x)$ is convex on $[0, \infty)$, Jensen's inequality gives $\mathbb{E}h(Z) \geq -\log(1 + \mathbb{E}Z_-) - C_0$. By Lemma 6.1(i), (6.1), Lemma 6.4 and $\lambda(0) \geq \delta$,

$$\mathbb{E}Z_- \leq \mathbb{E}R(Z) = \sqrt{\delta} b(0, 0) < \frac{\sqrt{\delta} C_\kappa}{\lambda(0)} \leq \frac{C_\kappa}{\sqrt{\delta}}.$$

Since $\delta \leq 1$, $1 + C_\kappa / \sqrt{\delta} \leq (1 + C_\kappa) / \sqrt{\delta}$. Hence

$$\mathcal{P}_*(\kappa) \geq \frac{1}{2} \log \delta - \alpha \log \frac{1 + C_\kappa}{\sqrt{\delta}} - \alpha C_0 = \frac{1+\alpha}{2} \log \delta - \alpha(C_0 + \log(1 + C_\kappa)).$$

$\square$

Combined with the upper bound of Theorem 6.8 at the margin $\kappa_c - \eta$, the lemma gives the collapse of minimizers used in Chapter 9.

Proposition 6.11. *Let $\alpha > 0$. There are constants $C, \eta_0 > 0$, depending only on α , such that for $0 < \eta \leq \eta_0$ every minimizer γ of $\mathcal{P}_{\kappa_c - \eta}$ over $\mathcal{U}$ satisfies*

$$1 - Q_\gamma \leq C\eta^{1/(1+\alpha)}, \quad \lambda_\gamma(0) \leq C\eta^{1/(2+2\alpha)}. \quad (6.24)$$

Proof. Put $C_* = C_{\kappa_c}$. Since C_κ is increasing in κ , $C_\kappa \leq C_*$ for $\kappa \leq \kappa_c$. By Theorem 6.8 there is $\eta_0 \in (0, 1]$ such that $\mathcal{P}_*(\kappa_c - \eta) \leq \frac{1}{2} \log \eta + C_1$ for $0 < \eta \leq \eta_0$, with $C_1 = \frac{1}{2} \log(2\alpha C_*) + \frac{1}{2}$. Fix such η , put $k = \kappa_c - \eta$, and let γ be a minimizer of $\mathcal{P}_k$ over $\mathcal{U}$, which exists by Theorem 6.3; write $\delta = 1 - Q_\gamma$. By (6.23) with $C_k \leq C_*$,

$$\frac{1+\alpha}{2} \log \delta - C_2 \leq \mathcal{P}_*(k) \leq \frac{1}{2} \log \eta + C_1, \quad C_2 = \alpha(C_0 + \log(1 + C_*)),$$

so $\delta \leq C_3 \eta^{1/(1+\alpha)}$ with $C_3 = e^{2(C_1+C_2)/(1+\alpha)}$. The minimizer γ minimizes $\mathcal{P}_k$ over $\mathcal{U}_r$ for every $r \in (Q_\gamma, 1)$, so (6.15) applies: with $C_4 = \max\{1, 2\alpha C_*/s_\alpha\}$, $\lambda_\gamma(0) \leq C_4 \sqrt{\delta} \leq C_4 C_3^{1/2} \eta^{1/(2+2\alpha)}$. $\square$

The proof uses the exact value of $\mathcal{P}_*$ at a minimizer, so it says nothing about the tops of near-minimizers. For $\alpha > 2$, the bound $1 - Q_\gamma \leq C\eta^{1/(1+\alpha)}$ for every γ with $\mathcal{P}_{\kappa_c - \eta}(\gamma) \leq \mathcal{P}_*(\kappa_c - \eta) + 1$, and a lower bound $\mathcal{P}_*(\kappa_c - \eta) \geq \alpha \log \eta - C$ of the same logarithmic order as (6.22), are proved in Part II through a comparison with the zero-temperature functional (Part II, Proposition 2.24); in this part only Corollary 1.7 uses the lower bound.

Theorem 6.8 and Proposition 6.11 are the bounds behind the comparison with the physical picture of jamming in Chapter 1; the mass $\lambda_\gamma(0)$ measures the order parameter below the top.

6.6 Proof of the theorem on the variational problem

We collect the results of this chapter; recall that $Q_\gamma = 1 - \delta$ is the top of the support of μ_γ and that $\lambda_\gamma(0) = \int_0^1 \gamma$.

Proof of Theorem 1.3. Part (i) is Theorem 6.3, part (ii) is Proposition 6.7, and part (iii) is Theorem 6.8, with $C_* = C_{\kappa_c}$. In part (iv), the pinning inequality is Lemma 6.10, and the bounds on minimizers at the margin $\kappa_c - \eta$ are (6.24) in Proposition 6.11. Part (v) is Corollary 6.9. This proves the theorem. $\square$

The critical margin near $\alpha = 2$

This chapter proves Theorem 1.4: for constraint densities slightly above two, the critical margin lies below Gardner's prediction κ_{RS} by at least a constant times the cube of the distance to two. Throughout the chapter, $\alpha = 2 + e$ with $e > 0$ small.

Step 4 of §1.3 outlines the argument. It uses only the definition (1.6) of κ_c , the Cole–Hopf recursion of §2.1 for one explicit order parameter, the tail bound (6.3) and one-dimensional Gaussian calculus, part of which is taken from Part II (Lemmas 2.1 and 3.1). Lemma 7.1 records the behavior of κ_{RS} near $\alpha = 2$. Lemma 7.2 bounds the value of an explicit order parameter with two support points near one by an explicit Gaussian integral J_ℓ , and deduces that one negative value of J_ℓ bounds κ_c from above. Lemma 7.3 expands J_ℓ , and the proof of Theorem 1.4 combines the three lemmas.

In the notation of Part II, $J_\ell(k) = J^k(\ell^{-1}\mathbf{1}_{[1-\ell,1]})$ is the value of the zero-temperature functional J^k (Part II, (2.41)) at the coefficient ℓ^{-1} carried by the block $[1 - \ell, 1]$. Part II proves the bound of Theorem 1.4 with this coefficient (Part II, Theorem 3.13), and transfers it to κ_c through the characterization of κ_c as the threshold of J^k (Part II, Theorem 2.21). The same characterization and a dual certificate for J^k give the matching bound $\kappa_{\text{RS}} - \kappa_c \leq Ce^3$, so the order e^3 is sharp (Corollary 1.6). Lemma 7.3 is a two-sided form of the bound in Step 3 of the proof of Part II, Theorem 3.13. Here Lemma 7.2 bounds $\mathcal{P}_\kappa$ by J_ℓ directly, so we need only the explicit form of J_ℓ and no comparison of $\mathcal{P}_\kappa$ with J^k .

Stojnic proved upper bounds on the capacity through a partially lifted version of his random duality theory, and numerical evaluation indicates that they lie below Gardner's prediction [117]; he later proposed fully lifted versions, evaluated numerically at finite levels of lifting [119]. The upper bound of Montanari, Zhong, and Zhou, discussed after Corollary 1.5, is expressed through the Gaussian exponential moment $\mathbb{E} \exp(-t(k - G)_+^2)$ [90, Definition 2], and $e^{\mathcal{E}(z)}$ in (7.4) below is this moment at $t = \frac{1}{2}$ and $k = -z$. The quadratic penalty that replaces the hard-wall datum in Lemma 7.2 is the jamming-limit boundary condition of Franz and Parisi [58, (15)], whose scaling regime near jamming also places the relevant overlaps close to one. What differs here is the trial order parameter, whose second atom sits at the bottom of a thin layer of relative width $\ell \asymp e^2$ below the top, and the analysis of the gain of this layer to order e^3 near $\alpha = 2$, where κ_{RS} is close to zero.

Besides $m(k) = \mathbb{E}(k - G)_+^2$ we use $m_1(k) = \mathbb{E}(k - G)_+$ and $\phi_0 = \phi(0) = (2\pi)^{-1/2}$. By

Part II, Lemma 2.1(g),

$$m(k) = (1 + k^2)\Phi(k) + k\phi(k), \quad m_1(k) = k\Phi(k) + \phi(k), \quad m' = 2m_1, \quad m'_1 = \Phi, \quad (7.1)$$

and $\Phi - m = -k m_1$. In particular $m_1 > 0$, m is increasing and convex, and m_1 is increasing. For $\alpha = 2 + e$, Gardner's prediction κ_{RS} is defined by $\alpha m(\kappa_{\text{RS}}) = 1$ (Proposition 6.7).

Lemma 7.1. *There is a numerical constant $e_1 \in (0, 1]$ such that for $0 < e \leq e_1$ and $\alpha = 2 + e$,*

$$-\frac{1}{2} < \kappa_{\text{RS}} < 0, \quad \alpha\Phi(\kappa_{\text{RS}}) - 1 = \alpha|\kappa_{\text{RS}}| m_1(\kappa_{\text{RS}}) \geq c_\Phi e, \quad c_\Phi = \frac{m_1(-1/2)}{4\phi_0} > 0. \quad (7.2)$$

Proof. Let e_g be the numerical constant of Part II, Lemma 3.1, and put $e_1 = \min\{e_g, 1\}$. For $0 < e \leq e_1$, part (a) there gives $-\frac{1}{2} < \kappa_{\text{RS}} < 0$, and part (b) gives the identity in (7.2) and the lower bound $c_\Phi e$, where $c_\Phi > 0$ because $m_1 > 0$. $\square$

Part II, Lemma 3.1(a), also gives the expansion of κ_{RS} quoted in the introduction: for $0 < e \leq e_1$,

$$\kappa_{\text{RS}}(2 + e) = -\frac{e}{8\phi_0} + O(e^2) = -\frac{\sqrt{2\pi}}{8} e + O(e^2). \quad (7.3)$$

We now introduce the explicit Gaussian integral. For $z \in \mathbb{R}$ put

$$\mathcal{E}(z) = \log \mathbb{E} \exp \left(-\frac{1}{2}(z + G)_-^2 \right), \quad (7.4)$$

and for $0 < \ell \leq \frac{1}{2}$ and $k \in \mathbb{R}$ put

$$J_\ell(k) = \alpha \ell \mathbb{E} \mathcal{E}(Y_\ell(k)) + \frac{1 - \ell}{4} + \frac{\ell \log 2}{2}, \quad Y_\ell(k) = \frac{\sqrt{1 - \ell} G - k}{\sqrt{\ell}}. \quad (7.5)$$

Since $e^{-(z+G)_-^2/2} \geq \mathbf{1}_{\{z+G \geq 0\}}$ and $(z + G)_-^2 \leq (z + G)^2$, we have $\mathcal{E}(z) \geq \log \Phi(z)$ and $\mathcal{E}(z) \geq \log \mathbb{E} e^{-(z+G)^2/2}$; the Gaussian integral

$$\mathbb{E} e^{-(z+G)^2/2} = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-(g+z/2)^2 - z^2/4} dg = \frac{e^{-z^2/4}}{\sqrt{2}} \quad (7.6)$$

follows from $g^2 + (z + g)^2 = 2(g + z/2)^2 + z^2/2$. The function $\mathcal{E}$ is nondecreasing, because $(z + G)_-$ is nonincreasing in z . Hence $0 \geq \mathcal{E}(z) \geq \mathcal{E}(\min\{z, 0\}) \geq -\frac{1}{4}z_-^2 - \frac{1}{2} \log 2$ for all z , and $J_\ell(k)$ is finite. Since $Y_\ell(k)$ is decreasing in k , the map $k \mapsto J_\ell(k)$ is nonincreasing.

The next lemma is the only place where the Parisi functional enters. It evaluates the functional at an order parameter with atoms at two levels near one, and bounds its energy term by replacing the hard-wall datum with a quadratic penalty.

Lemma 7.2. Let $\kappa \in \mathbb{R}$, $0 < \ell \leq \frac{1}{2}$, and $\delta \in (0, 1)$ with $\delta < (1 - \delta)\ell$. Put

$$Q = 1 - \delta, \quad p = Q(1 - \ell), \quad a = \frac{\delta}{Q\ell} \in (0, 1), \quad \gamma_\delta = a\mathbf{1}_{[p, Q]} + \mathbf{1}_{[Q, 1]}.$$

Then $\gamma_\delta \in \mathcal{U}$, $Q_{\gamma_\delta} = Q$, and

$$\mathcal{P}_\kappa(\gamma_\delta) \leq \frac{Q}{\delta} J_\ell\left(\frac{\kappa}{\sqrt{Q}}\right) + \frac{1}{2} \log \delta. \quad (7.7)$$

Consequently, if $J_\ell(k) < 0$ for some $k \in \mathbb{R}$ and $\ell \in (0, \frac{1}{2}]$, then $\kappa_c \leq k$.

Proof. The function γ_δ is nondecreasing, right-continuous, equal to one on $[Q, 1]$ and at most $a < 1$ on $[0, Q]$, so $\gamma_\delta \in \mathcal{U}$ and $Q_{\gamma_\delta} = Q$. It is a step order parameter, so $u = u_{\gamma_\delta}(\cdot, \cdot; \kappa)$ is given on $[0, Q]$ by the recursion (2.2) with terminal time Q and terminal datum $f(x) = \log \Phi((x - \kappa)/\sqrt{\delta})$:

$$u(p, \cdot) = \mathcal{T}_{a, Q\ell} f, \quad u(0, \cdot) = \mathcal{T}_{0, p} u(p, \cdot),$$

because $Q - p = Q\ell$. By the upper bound in (6.3) with $t = (\kappa - x)/\sqrt{\delta}$,

$$\delta f(x) \leq g(x) := -\frac{1}{2}(\kappa - x)_+^2 \quad (x \in \mathbb{R}).$$

By the homogeneity (T3) with $t = \delta$ and the monotonicity (T1), and since $a/\delta = 1/(Q\ell)$,

$$\begin{aligned} \delta u(p, x) &= \mathcal{T}_{1/(Q\ell), Q\ell}(\delta f)(x) \leq \mathcal{T}_{1/(Q\ell), Q\ell} g(x) \\ &= Q\ell \log \mathbb{E} \exp\left(-\frac{(\kappa - x - \sqrt{Q\ell} G)_+^2}{2Q\ell}\right) = Q\ell \mathcal{E}\left(\frac{x - \kappa}{\sqrt{Q\ell}}\right), \end{aligned}$$

where in the last step we used $(\kappa - x - \sqrt{Q\ell} G)/\sqrt{Q\ell} = -(z + G)$ with $z = (x - \kappa)/\sqrt{Q\ell}$, $-(z + G)_+ = (z + G)_-$, and the definition (7.4). Applying the monotone operator $\mathcal{T}_{0, p}$ and using $p = Q(1 - \ell)$,

$$\delta u(0, 0) \leq Q\ell \mathbb{E} \mathcal{E}\left(\frac{\sqrt{p} G - \kappa}{\sqrt{Q\ell}}\right) = Q\ell \mathbb{E} \mathcal{E}\left(Y_\ell(\kappa/\sqrt{Q})\right).$$

For the entropy term $\text{CS}(\gamma_\delta)$ we use the cutoff Q in (1.4). On $[p, Q]$, $\lambda_{\gamma_\delta}(q) = \delta + a(Q - q)$, and on $[0, p]$, $\lambda_{\gamma_\delta}(q) = \delta + aQ\ell = 2\delta$. Hence

$$\begin{aligned} \text{CS}(\gamma_\delta) &= \frac{1}{2} \int_0^p \frac{dq}{2\delta} + \frac{1}{2} \int_p^Q \frac{dq}{\delta + a(Q - q)} + \frac{1}{2} \log \delta = \frac{p}{4\delta} + \frac{1}{2a} \log \frac{\delta + aQ\ell}{\delta} + \frac{1}{2} \log \delta \\ &= \frac{Q}{\delta} \left(\frac{1 - \ell}{4} + \frac{\ell \log 2}{2} \right) + \frac{1}{2} \log \delta. \end{aligned}$$

Adding $au(0, 0)$ gives (7.7).

For the last assertion, let $J_\ell(k) < 0$ and $\kappa' > k$. As $\delta \downarrow 0$ we have $Q \rightarrow 1$ and $\kappa'/\sqrt{Q} \rightarrow \kappa' > k$, so for all sufficiently small δ the hypothesis $\delta < (1 - \delta)\ell$ holds and $\kappa'/\sqrt{Q} \geq k$. For such δ , (7.7) and the monotonicity of J_ℓ give

$$\mathcal{P}_*(\kappa') \leq \mathcal{P}_{\kappa'}(\gamma_\delta) \leq \frac{Q}{\delta} J_\ell(k) + \frac{1}{2} \log \delta,$$

and the right side tends to $-\infty$ as $\delta \downarrow 0$. Hence $\mathcal{P}_*(\kappa') = -\infty$ for every $\kappa' > k$, and (1.6) gives $\kappa_c \leq k$. $\square$

The next lemma expands J_ℓ in the width ℓ . The first term of the expansion is the value obtained from the replica-symmetric order parameters, and the second is the gain produced by the thin layer.

Lemma 7.3. *There is a numerical constant C_J such that for $0 < \alpha \leq 3$, $0 < \ell \leq \frac{1}{2}$, and $|k| \leq 1$,*

$$\left| J_\ell(k) - \frac{1 - \alpha m(k)}{4} + \frac{\ell}{2} d (\alpha \Phi(k) - 1) \right| \leq C_J \ell^{3/2}, \quad d = \log 2 - \frac{1}{2}. \quad (7.8)$$

Moreover $d > 0$.

Proof. We divide the proof into three steps. First we isolate an integrable remainder of $\mathcal{E}$, then we rewrite J_ℓ exactly with it, and finally we expand the explicit terms.

STEP 1. *An integrable remainder.* Put

$$\mathcal{R}(z) = \mathcal{E}(z) + \frac{1}{4} z_-^2 + \frac{\log 2}{2} \mathbf{1}_{\{z < 0\}} \quad (z \in \mathbb{R}).$$

Let $z < 0$ and write $A(z) = e^{-z^2/4}/\sqrt{2}$, so that $\mathcal{R}(z) = \log(e^{\mathcal{E}(z)}/A(z))$ by (7.6). Since $(z + G)_-^2 \leq (z + G)^2$, with equality on $\{z + G \leq 0\}$,

$$0 \leq e^{\mathcal{E}(z)} - A(z) = \mathbb{E} \left[1 - e^{-(z+G)^2/2}; z + G > 0 \right] \leq \mathbb{P}(z + G > 0) = \Phi(z).$$

Hence $0 \leq \mathcal{R}(z) \leq \log(1 + \Phi(z)/A(z)) \leq \sqrt{2} e^{z^2/4} \Phi(z)$. For $-1 \leq z < 0$ the right side is at most $\sqrt{2} e^{1/4}$, and for $z = -t \leq -1$ the Mills bound of Lemma 2.5(i) gives $\sqrt{2} e^{t^2/4} \Phi(-t) \leq \sqrt{2} e^{t^2/4} \phi(t)/t \leq \sqrt{2} \phi_0 e^{-t^2/4}$. Now let $z \geq 0$. Then $\Phi(z) \leq e^{\mathcal{E}(z)} \leq 1$, so $0 \leq -\mathcal{R}(z) = -\mathcal{E}(z) \leq -\log \Phi(z) = -\log(1 - \Phi(-z)) \leq 2\Phi(-z)$, because $\Phi(-z) \leq \frac{1}{2}$ and $-\log(1 - y) \leq 2y$ for $0 \leq y \leq \frac{1}{2}$. Since $\int_0^\infty \Phi(-z) dz = \mathbb{E}G_+ = \phi_0$, these bounds give $\|\mathcal{R}\|_{L^1(\mathbb{R})} < \infty$.

STEP 2. *An exact formula.* Write $Y = Y_\ell(k)$ and, as in Part II, put $m(k; t) = \mathbb{E}(k - \sqrt{t} G)_+^2$ for $t > 0$, so that $m(k; 1) = m(k)$. Then $\mathbb{E}Y_-^2 = m(k; 1 - \ell)/\ell$ and $\mathbb{P}(Y < 0) = \Phi(k/\sqrt{1 - \ell})$, so the definitions (7.5) of J_ℓ and of $\mathcal{R}$ give

$$J_\ell(k) = \frac{1 - \ell - \alpha m(k; 1 - \ell)}{4} + \frac{\ell \log 2}{2} \left(1 - \alpha \Phi(k/\sqrt{1 - \ell}) \right) + \alpha \ell \mathbb{E} \mathcal{R}(Y). \quad (7.9)$$

The variable Y is Gaussian with variance $(1 - \ell)/\ell \geq 1/(2\ell)$, so its density is at most $\phi_0\sqrt{2\ell}$, and

$$|\alpha\ell \mathbb{E}\mathcal{R}(Y)| \leq 3\sqrt{2}\phi_0\|\mathcal{R}\|_{L^1}\ell^{3/2}.$$

STEP 3. *Expansion of the explicit terms.* Gaussian integration by parts, $\frac{d}{dt}\mathbb{E}F(\sqrt{t}G) = \frac{1}{2}\mathbb{E}F''(\sqrt{t}G)$ applied to $F(y) = (k - y)_+^2$, gives $\partial_t m(k; t) = \Phi(k/\sqrt{t})$. For $t \in [\frac{1}{2}, 1]$,

$$\left| \frac{d}{dt}\Phi(k/\sqrt{t}) \right| = \frac{|k|\phi(k/\sqrt{t})}{2t^{3/2}} \leq \sqrt{2}\phi_0|k| \leq 1,$$

so $|\Phi(k/\sqrt{t}) - \Phi(k)| \leq 1 - t$. Integrating over $t \in [1 - \ell, 1]$ gives $|m(k; 1 - \ell) - m(k) + \ell\Phi(k)| \leq \ell^2/2$, and also $|\Phi(k/\sqrt{1 - \ell}) - \Phi(k)| \leq \ell$. Inserting these bounds into (7.9),

$$J_\ell(k) = \frac{1 - \alpha m(k)}{4} + \frac{\ell}{4}(\alpha\Phi(k) - 1) - \frac{\ell \log 2}{2}(\alpha\Phi(k) - 1) + \varepsilon,$$

where $|\varepsilon| \leq \frac{\alpha\ell^2}{8} + \frac{\alpha\ell^2 \log 2}{2} + 3\sqrt{2}\phi_0\|\mathcal{R}\|_{L^1}\ell^{3/2}$. The two middle terms combine to $-\frac{\ell}{2}d(\alpha\Phi(k) - 1)$, and $\ell^2 \leq \ell^{3/2}$, so (7.8) holds with $C_J = \frac{3}{8} + \frac{3}{2}\log 2 + 3\sqrt{2}\phi_0\|\mathcal{R}\|_{L^1}$. Finally $\log 2 = \int_1^2 t^{-1} dt > \frac{1}{2}$, so $d > 0$. $\square$

Proof of Theorem 1.4. Let c_Φ , C_J , and d be as in Lemmas 7.1 and 7.3, and put

$$t = \left(\frac{d c_\Phi}{4C_J} \right)^2, \quad c = \frac{t d c_\Phi}{8}.$$

Let $e_0 \leq e_1$ be so small that $te_0^2 \leq \frac{1}{2}$, $ce_0^3 \leq \frac{1}{2}$, and $3tde_0^2 \leq 1$. Fix $0 < e < e_0$, put $\alpha = 2 + e \leq 3$, $\ell = te^2$, $s = ce^3$, and $k = \kappa_{RS} - s$. By Lemma 7.1, $-1 < k < 0$, so Lemma 7.3 applies and gives

$$J_\ell(k) \leq \frac{1 - \alpha m(k)}{4} - \frac{\ell d}{2}(\alpha\Phi(k) - 1) + C_J\ell^{3/2}.$$

We bound the two explicit terms at $k = \kappa_{RS} - s$ by their values at κ_{RS} . Since m is convex with $m' = 2m_1$, and $m_1(\kappa_{RS}) \leq m_1(0) = \phi_0$ because $\kappa_{RS} < 0$,

$$1 - \alpha m(k) = \alpha(m(\kappa_{RS}) - m(k)) \leq \alpha m'(\kappa_{RS})s \leq 2\alpha\phi_0 s.$$

Since $\Phi' \leq \phi_0$, (7.2) gives $\alpha\Phi(k) - 1 \geq \alpha\Phi(\kappa_{RS}) - 1 - \alpha\phi_0 s \geq c_\Phi e - \alpha\phi_0 s$. Using $\alpha\phi_0 \leq 3\phi_0 < \frac{3}{2}$, we obtain

$$J_\ell(k) \leq \frac{3}{4}s - \frac{\ell d}{2}c_\Phi e + \frac{3}{4}d\ell s + C_J\ell^{3/2} = e^3 \left(\frac{3}{4}c - \frac{tdc_\Phi}{2} + C_J t^{3/2} \right) + \frac{3}{4}d t c e^5.$$

By the choice of t , $C_J t^{3/2} = \frac{1}{4}tdc_\Phi$, and $\frac{3}{4}c = \frac{3}{32}tdc_\Phi$. By the choice of e_0 , $\frac{3}{4}d t c e^5 \leq \frac{1}{4}c e^3 = \frac{1}{32}tdc_\Phi e^3$. Hence

$$J_\ell(k) \leq \left(\frac{3}{32} - \frac{16}{32} + \frac{8}{32} + \frac{1}{32} \right) t d c_\Phi e^3 = -\frac{1}{8} t d c_\Phi e^3 < 0.$$

Since $\ell \leq \frac{1}{2}$, Lemma 7.2 gives $\kappa_c \leq k = \kappa_{RS} - ce^3$. $\square$

The constant c is not optimized, and the proof does not identify the limit of $(\kappa_{\text{RS}} - \kappa_c)/e^3$.

From soft free energy to hard feasible volume

This chapter proves the lower bounds in Theorem 1.2(i),(ii) and completes the proof of Theorem 1.2; Step 5 of §1.3 outlines the argument. The matching upper bounds are Corollaries 3.11 and 3.14; the second rests on the volume estimate from one feasible point in Lemma 3.13. §8.1 shows that a configuration whose total squared violation $\sum_a (k_0 - b_a \cdot \sigma)_+^2$ at a level k_0 is at most tN can be displaced by a vector of norm at most $4\sqrt{tN}$ to a point that satisfies every constraint at a slightly smaller margin (Lemma 8.2). This rests on the conditioning of the pattern matrix on sparse sets of rows (Lemma 8.1). §8.2 transfers volume by a Gaussian perturbation tangent to the sphere (Lemma 8.4). §8.3 combines the two with the Parisi formula for a smooth concave penalty (Theorem 8.5), and §8.4 proves Theorem 1.2.

For $\kappa \geq 0$ the feasibility problem is convex: a nonzero point of the ball that satisfies the constraints can be moved out along its ray to the sphere, since $b_a \cdot (t\sigma) \geq b_a \cdot \sigma \geq \kappa$ for $t \geq 1$, so the feasible set on the sphere is nonempty exactly when its convex relaxation in the ball contains a nonzero point. The passage from smooth concave penalties to the hard constraints in [115] and [128, Chapter 8] uses this convexity. For negative margins we replace convexity by a repair of small violations. For the present model, a rounding step of this kind appears in [54, Lemma 11]. There El Alaoui and Sellke project a point with small total violation onto the feasible polyhedron, and they bound the distance by convex duality and the smallest positive singular value of the pattern matrix. Montanari, Zhong, and Zhou use this rounding step in the proofs of [90, Theorems 4.1 and 5.1]. The repair of Lemma 8.2 also rests on convex duality. Our argument differs in three respects. The pattern matrix enters through its operator norm and its conditioning on sparse sets of rows, and not through its smallest positive singular value. The repaired point satisfies the constraints at a margin lowered by an additive slack 2τ , which allows the rows with small violation to decrease. Finally, the conclusion is a lower bound on the feasible volume rather than the existence of one feasible point.

Throughout this chapter, and in Chapter 9, $A = N^{-1/2}\mathbf{G}$ denotes the $M \times N$ matrix with rows $b_a = g_a / \sqrt{N}$, so that $(A\sigma)_a = b_a \cdot \sigma$. We write $\|\cdot\|$ for the Euclidean norm, $\|A\|_{\text{op}}$ for the operator norm, $\mathbf{1} \in \mathbb{R}^M$ for the vector with all entries equal to one, and $y_+ = \max\{y, 0\}$, applied entrywise to vectors. Inequalities between vectors are entrywise. Thus

$$S_N(\kappa) = \{\sigma \in S_N : A\sigma \geq \kappa \mathbf{1}\}, \quad V_N(\kappa) = \mu_N(S_N(\kappa)).$$

Since $M/N \rightarrow \alpha$, we may and do assume that N is so large that $\alpha N/2 \leq M \leq (\alpha + 1)N$.

8.1 Repairing a small violation vector

Let $u \in \mathbb{R}^N$ and $k_0 \in \mathbb{R}$, and let $w = (k_0 \mathbf{1} - Au)_+$ be the vector of violations of the constraints at margin k_0 . We look for a displacement d with $A(u + d) \geq (k_0 - 2\tau)\mathbf{1}$, where $\tau > 0$ is small. If $\|w\|^2 \leq tN$, then at most tN/τ^2 rows have violation larger than τ . The displacement must raise these rows by almost their violation, while it may lower every other row by up to τ . A displacement with this property and with norm comparable to $\|w\|$ exists when $A^\top$ is well conditioned on vectors supported on few rows. For $1 \leq m \leq \min\{M, N\}$ we therefore define the smallest singular value of $A^\top$ over vectors supported on at most m rows,

$$s_m(A) = \inf\{\|A^\top y\| : y \in \mathbb{R}^M, \|y\| = 1, |\text{supp } y| \leq m\}.$$

Lemma 8.1 shows that the Gaussian matrix A has this property, together with a bound on its operator norm, with overwhelming probability, and Lemma 8.2 then constructs the displacement. A lower bound on $s_m(A)$ is the lower half of the restricted isometry property [28] for the $N \times M$ matrix $A^\top$, and we prove it by the net argument of [12]; see also [130, Theorem 5.65].

The proof of Lemma 8.1 uses the operator-norm bound of Lemma 4.2, nets, and one chi-square tail bound. For each $\varepsilon \in (0, 1]$, the unit sphere of $\mathbb{R}^n$ contains an ε -net with at most $(1 + 2/\varepsilon)^n$ points (volumetric bound). If X has the chi-square law with n degrees of freedom, then $\mathbb{E}e^{-X/6} = (3/4)^{n/2}$, and Markov's inequality applied to $e^{-X/6}$ gives

$$\mathbb{P}\{X \leq 3n/4\} \leq e^{n/8} \left(\frac{3}{4}\right)^{n/2} = e^{-n(\frac{1}{2} \log \frac{4}{3} - \frac{1}{8})} \leq e^{-n/64}, \quad (8.1)$$

since $\frac{1}{2} \log \frac{4}{3} - \frac{1}{8} > 0.018 > \frac{1}{64}$.

Lemma 8.1. *There are constants $s \in (0, \min\{\alpha, 1\}/4)$, $K \geq 1$, and $c > 0$, depending only on α , such that for all sufficiently large N the event*

$$E_N = \left\{ \|A\|_{\text{op}} \leq K, \quad s_{\lfloor 2sN \rfloor}(A) \geq \frac{1}{2} \right\}$$

has probability at least $1 - e^{-cN}$.

Proof. We divide the proof into two steps: we first treat the operator norm and then the sparse singular values.

STEP 1. The operator norm. Put $K = \sqrt{\alpha + 1} + 2$. Then $K \geq 1$, and K depends only on α . Lemma 4.2, applied to the matrix $\mathbf{G} = \sqrt{N}A$ with $t = \sqrt{N}$, gives $\mathbb{P}\{\|A\|_{\text{op}} > \sqrt{M/N} + 2\} \leq 2e^{-N/2}$. Since $M \leq (\alpha + 1)N$, it follows that $\mathbb{P}\{\|A\|_{\text{op}} > K\} \leq 2e^{-N/2}$.

STEP 2. Sparse singular values. Let $s \in (0, \min\{\alpha, 1\}/4)$, to be chosen below, and put $m = \lfloor 2sN \rfloor$. For large N we have $1 \leq m \leq 2sN \leq \min\{M, N\}$, because $M \geq \alpha N/2$. Fix a set S of m rows, let A_S be the $m \times N$ submatrix of A formed by the rows in S , and fix a unit vector $v \in \mathbb{R}^S$. The vector $A_S^\top v = N^{-1/2} \sum_{a \in S} v_a g_a$ has the law $N(0, N^{-1}I_N)$, so $N\|A_S^\top v\|^2$ has the chi-square law with N degrees of freedom, and (8.1) gives $\mathbb{P}\{\|A_S^\top v\|^2 \leq 3/4\} \leq e^{-N/64}$. Let $\mathcal{N}_S$ be a $1/(4K)$ -net of the unit sphere of $\mathbb{R}^S$ with $|\mathcal{N}_S| \leq (1 + 8K)^m$. Suppose that $\|A\|_{\text{op}} \leq K$ and that $\|A_S^\top v_0\|^2 > 3/4$ for every $v_0 \in \mathcal{N}_S$. For a unit vector $v \in \mathbb{R}^S$ choose $v_0 \in \mathcal{N}_S$ with $\|v - v_0\| \leq 1/(4K)$. Then

$$\|A_S^\top v\| \geq \|A_S^\top v_0\| - \|A\|_{\text{op}}\|v - v_0\| \geq \frac{\sqrt{3}}{2} - \frac{1}{4} > \frac{1}{2}.$$

Every $y \in \mathbb{R}^M$ with $|\text{supp } y| \leq m$ is supported in some set S of exactly m rows, and then $\|A^\top y\| = \|A_S^\top y_S\|$. Hence, if $\|A\|_{\text{op}} \leq K$ and $s_m(A) < 1/2$, then $\|A_S^\top v_0\|^2 \leq 3/4$ for some set S of m rows and some $v_0 \in \mathcal{N}_S$. There are at most $\binom{M}{m} \leq (eM/m)^m$ sets S , so a union bound over S and over $v_0 \in \mathcal{N}_S$ gives

$$\mathbb{P}\left\{\|A\|_{\text{op}} \leq K, s_m(A) < \frac{1}{2}\right\} \leq \exp\left\{m \log \frac{e(1+8K)M}{m} - \frac{N}{64}\right\}.$$

The map $y \mapsto y \log(e(1+8K)M/y)$ is increasing on $(0, (1+8K)M]$, and $m \leq 2sN \leq M \leq (\alpha+1)N$. Hence the exponent is at most $N\{2s \log(e(1+8K)(\alpha+1)/(2s)) - 1/64\}$. Since $x \log(C/x) \rightarrow 0$ as $x \downarrow 0$, we may choose s , depending only on α and K , so small that $2s \log(e(1+8K)(\alpha+1)/(2s)) \leq 1/128$. The probability above is then at most $e^{-N/128}$.

Adding the two failure probabilities gives $\mathbb{P}(E_N^c) \leq 2e^{-N/2} + e^{-N/128} \leq e^{-N/256}$ for large N . This proves the lemma with $c = 1/256$. $\square$

The next lemma is an error bound of Hoffman type [70] for the relaxed system $Ax \geq (k_0 - 2\tau)\mathbf{1}$: if the violation vector $(k_0\mathbf{1} - Au)_+$ has norm at most $\sqrt{tN}$, then the distance from u to the solution set is at most $4\sqrt{tN}$. The constant 4 does not depend on N , at the price of the slack 2τ and of a smallness condition on t .

Lemma 8.2. *Let s and K be as in Lemma 8.1, suppose that the event E_N occurs, and let $N \geq 1/s$. Let $k_0 \in \mathbb{R}$, $\tau > 0$, and $t > 0$ with*

$$t \leq \frac{\tau^2 s}{32K^2}.$$

If $u \in \mathbb{R}^N$ satisfies $\|(k_0\mathbf{1} - Au)_+\|^2 \leq tN$, then there is $d \in \mathbb{R}^N$ with

$$A(u + d) \geq (k_0 - 2\tau)\mathbf{1}, \quad \|d\| \leq 4\sqrt{tN}.$$

Proof. Put $w = (k_0\mathbf{1} - Au)_+$ and $S = \{a : w_a > \tau\}$. By the definition of w we have $Au \geq k_0\mathbf{1} - w$. It therefore suffices to find $d \in \mathbb{R}^N$ with

$$Ad \geq w - 2\tau\mathbf{1}, \quad \|d\| \leq 4\|w\|, \tag{8.2}$$

because then $A(u + d) \geq (k_0 - 2\tau)\mathbf{1}$ and $\|d\| \leq 4\sqrt{tN}$. The target $w - 2\tau\mathbf{1}$ asks that each row be raised by its violation w_a up to an error 2τ . For the rows $a \notin S$ this allows a decrease, since $w_a \leq \tau$ there. We divide the proof into two steps.

STEP 1. *Counting.* Since $\tau^2|S| \leq \|w\|^2 \leq tN$ and $t \leq \tau^2 s / (32K^2) \leq \tau^2 s / 32$, we have $|S| \leq sN/32$. Put $p = \lfloor sN \rfloor$ and $m = \lfloor 2sN \rfloor$. As $sN \geq 1$, we have $p \geq 1$ and $p \geq sN/2$. The integer $|S| + p$ is at most $(33/32)sN \leq 2sN$, hence at most m . Moreover $m \leq \min\{M, N\}$, because $s < \min\{\alpha, 1\}/4$ and $M \geq \alpha N/2$. On E_N , by the assumption on t ,

$$s_m(A) - \frac{\|A\|_{\text{op}}\|w\|}{\tau\sqrt{p}} \geq \frac{1}{2} - \frac{K\sqrt{tN}}{\tau\sqrt{sN/2}} = \frac{1}{2} - \frac{\sqrt{2}K\sqrt{t}}{\tau\sqrt{s}} \geq \frac{1}{4}. \quad (8.3)$$

STEP 2. *Separation.* Put $R_0 = 4\|w\|$ and $v = w - 2\tau\mathbf{1}$. The required d exists if and only if v belongs to

$$\mathcal{K} = \{Ad - z : d \in \mathbb{R}^N, \|d\| \leq R_0, z \in \mathbb{R}^M, z \geq 0\},$$

since $v = Ad - z$ with $z \geq 0$ means $Ad \geq v$. The set $\mathcal{K}$ is convex. It is closed, being the sum of the compact set $\{Ad : \|d\| \leq R_0\}$ and the closed cone $\{-z : z \geq 0\}$. Its support function $h_{\mathcal{K}}(y) = \sup_{x \in \mathcal{K}} y \cdot x$ equals $R_0\|A^\top y\|$ if $y \geq 0$: the supremum of $y \cdot Ad$ over $\|d\| \leq R_0$ is $R_0\|A^\top y\|$, and $\sup_{z \geq 0} (-y \cdot z) = 0$. If some $y_a < 0$, then $h_{\mathcal{K}}(y) = +\infty$. By the separating hyperplane theorem, a point lies in the closed convex set $\mathcal{K}$ if and only if $y \cdot v \leq h_{\mathcal{K}}(y)$ for every $y \in \mathbb{R}^M$. It therefore suffices to prove

$$y \cdot v \leq R_0\|A^\top y\| \quad \text{for every } y \in \mathbb{R}^M \text{ with } y \geq 0. \quad (8.4)$$

Fix $y \geq 0$, and write y_S and y_{S^c} for its restrictions to S and to S^c . For $a \in S$ we have $v_a \leq w_a$, and for $a \notin S$ we have $v_a = w_a - 2\tau \leq -\tau$. Hence

$$y \cdot v \leq y_S \cdot w_S - \tau\|y_{S^c}\|_1.$$

The right side of (8.4) is nonnegative, so we may assume $y \cdot v > 0$. Then

$$\tau\|y_{S^c}\|_1 < y_S \cdot w_S \leq \|y_S\|\|w\|. \quad (8.5)$$

Let $T \subseteq S^c$ be the set of indices of the p largest entries of y_{S^c} (all of S^c if $|S^c| \leq p$), and write $y = y' + y''$, where y' is the restriction of y to $S \cup T$ and y'' its restriction to $S^c \setminus T$. Each entry of y'' is at most the p -th largest entry of y_{S^c} , and hence at most $\|y_{S^c}\|_1/p$. Therefore

$$\|y''\|^2 \leq \frac{\|y_{S^c}\|_1}{p} \|y''\|_1 \leq \frac{\|y_{S^c}\|_1^2}{p}.$$

This splitting of a vector by the sizes of its entries is standard in compressed sensing; see [56]. The vector y' has at most $|S| + p \leq m$ nonzero entries and $\|y'\| \geq \|y_S\|$. By the definition of $s_m(A)$, by (8.5) and by (8.3),

$$\begin{aligned} \|A^\top y\| &\geq \|A^\top y'\| - \|A^\top y''\| \geq s_m(A)\|y_S\| - \|A\|_{\text{op}} \frac{\|y_{S^c}\|_1}{\sqrt{p}} \\ &\geq \left(s_m(A) - \frac{\|A\|_{\text{op}}\|w\|}{\tau\sqrt{p}} \right) \|y_S\| \geq \frac{1}{4} \|y_S\|. \end{aligned}$$

Combining this with (8.5) gives $y \cdot v \leq \|y_S\|\|w\| \leq 4\|w\|\|A^\top y\| = R_0\|A^\top y\|$. This proves (8.4), and hence (8.2). $\square$

8.2 Transferring volume by a tangent Gaussian perturbation

Let $B \subseteq S_N$, and suppose that every $\sigma \in B$ can be displaced by a vector d of norm at most $r\sqrt{N}$ to a point with $A(\sigma + d) \geq k\mathbf{1}$. We want to show that the volume of a slightly relaxed feasible set is not much smaller than $\mu_N(B)$. The displaced points $\sigma + d$ may form a set of small volume, so we perturb σ itself. For $\sigma \in S_N$ let $P_\sigma = I_N - \sigma\sigma^\top/N$ be the orthogonal projection onto $\sigma^\perp = \{x \in \mathbb{R}^N : x \cdot \sigma = 0\}$. Fix $\rho > 0$, let ξ be a standard Gaussian vector in $\mathbb{R}^N$, and put

$$Y(\sigma, \xi) = \sqrt{N} \frac{\sigma + \rho P_\sigma \xi}{\|\sigma + \rho P_\sigma \xi\|} \in S_N.$$

Since $P_\sigma \xi \perp \sigma$, we have $\|\sigma + \rho P_\sigma \xi\|^2 = N + \rho^2 \|P_\sigma \xi\|^2 \geq N$. Thus Y is a continuous map from $S_N \times \mathbb{R}^N$ to S_N , and $Y(\sigma, \xi)$ is obtained from $\sigma + \rho P_\sigma \xi$ by multiplication with a number in $(0, 1]$. In particular, normalization can only raise a negative margin. Two properties of Y are used. If σ is uniform on S_N and independent of ξ , then so is $Y(\sigma, \xi)$. For each fixed σ that admits a small repair, the point $Y(\sigma, \xi)$ is feasible at a slightly smaller margin with probability at least $e^{-N\varepsilon}$ for a small $\varepsilon > 0$. The second property rests on two Gaussian inequalities. The first compares a symmetric event with its translate. It is a standard lower bound for the Gaussian measure of a shifted symmetric set; see [84, 85].

Lemma 8.3. *Let ξ be a standard Gaussian vector in $\mathbb{R}^N$, let $\mathcal{R} \subseteq \mathbb{R}^N$ be a Borel set with $\mathcal{R} = -\mathcal{R}$, and let $y \in \mathbb{R}^N$. Then*

$$\mathbb{P}\{\xi \in y + \mathcal{R}\} \geq e^{-\|y\|^2/2} \mathbb{P}\{\xi \in \mathcal{R}\}.$$

Proof. The standard Gaussian density φ_N on $\mathbb{R}^N$ satisfies $\varphi_N(x + y) = \varphi_N(x)e^{-x \cdot y - \|y\|^2/2}$. Hence

$$\begin{aligned} \mathbb{P}\{\xi \in y + \mathcal{R}\} &= \int_{\mathcal{R}} \varphi_N(x + y) dx = e^{-\|y\|^2/2} \mathbb{E}[e^{-\xi \cdot y} \mathbf{1}_{\mathcal{R}}(\xi)] \\ &= e^{-\|y\|^2/2} \mathbb{E}[\cosh(\xi \cdot y) \mathbf{1}_{\mathcal{R}}(\xi)], \end{aligned}$$

where the last equality holds because ξ and $-\xi$ have the same law and $\mathcal{R} = -\mathcal{R}$. Since $\cosh \geq 1$, this proves the lemma. $\square$

The second is Šidák's inequality in the form of Lemma 3.12. In the next lemma the matrix A is fixed, and probabilities refer to ξ and to the auxiliary uniform point in (b).

Lemma 8.4. *Let $K > 0$ be such that $\max_a \|b_a\| \leq K$. Let $k < 0$, $0 < r \leq 1/2$, $\eta > 0$, and $\rho > 0$, suppose that $\exp(-\eta^2/(2K^2\rho^2)) \leq 1/2$, and put*

$$k' = \frac{k}{1-r} - \eta.$$

(a) *If $\sigma \in S_N$ and $d \in \mathbb{R}^N$ satisfy $\|d\| \leq r\sqrt{N}$ and $A(\sigma + d) \geq k\mathbf{1}$, then*

$$\mathbb{P}\{Y(\sigma, \xi) \in S_N(k')\} \geq \exp\left\{-\frac{2r^2N}{\rho^2} - 2Me^{-\eta^2/(2K^2\rho^2)}\right\}. \quad (8.6)$$

(b) *If $B \subseteq S_N$ is a Borel set such that every $\sigma \in B$ admits a vector d as in (a), then*

$$V_N(k') \geq \mu_N(B) \exp\left\{-\frac{2r^2N}{\rho^2} - 2Me^{-\eta^2/(2K^2\rho^2)}\right\}. \quad (8.7)$$

Proof. We divide the proof into four steps. The first three prove (a): we move the repair displacement into the tangent space, compare the perturbed point with its translate, and bound the probability of the untranslated event. The fourth proves (b) by integration.

STEP 1. A tangent displacement. The Gaussian vector $\rho P_\sigma \xi$ lies in $\sigma^\perp$, so we first replace d by a vector in $\sigma^\perp$. Write $d = c\sigma + P_\sigma d$ with $c = d \cdot \sigma / N$. Then $|c| \leq \|d\|/\sqrt{N} \leq r \leq 1/2$, and $d' = P_\sigma d / (1+c) \in \sigma^\perp$ satisfies

$$\sigma + d' = \frac{(1+c)\sigma + d - c\sigma}{1+c} = \frac{\sigma + d}{1+c}.$$

Since $k < 0$ and $0 < 1/(1+c) \leq 1/(1-r)$,

$$A(\sigma + d') = \frac{A(\sigma + d)}{1+c} \geq \frac{k}{1+c} \mathbf{1} \geq \frac{k}{1-r} \mathbf{1}, \quad \|d'\| \leq \frac{\|d\|}{1-r} \leq 2r\sqrt{N}. \quad (8.8)$$

STEP 2. Shifting the Gaussian vector. Let

$$\mathcal{R} = \{x \in \mathbb{R}^N : |b_a \cdot P_\sigma x| \leq \eta/\rho \text{ for all } a \leq M\},$$

a closed set with $\mathcal{R} = -\mathcal{R}$, and put $y = d'/\rho$. If $\xi \in y + \mathcal{R}$, then, since $P_\sigma y = y$,

$$\sigma + \rho P_\sigma \xi = \sigma + d' + \rho P_\sigma (\xi - y), \quad A(\sigma + \rho P_\sigma \xi) \geq \frac{k}{1-r} \mathbf{1} - \eta \mathbf{1} = k' \mathbf{1},$$

by (8.8). Since $k' < 0$ and $Y(\sigma, \xi)$ is obtained from $\sigma + \rho P_\sigma \xi$ by multiplication with a number in $(0, 1]$, also $AY(\sigma, \xi) \geq k' \mathbf{1}$, that is, $Y(\sigma, \xi) \in S_N(k')$. By Lemma 8.3 and $\|y\|^2 = \|d'\|^2 / \rho^2 \leq 4r^2 N / \rho^2$,

$$\mathbb{P}\{Y(\sigma, \xi) \in S_N(k')\} \geq \mathbb{P}\{\xi \in y + \mathcal{R}\} \geq e^{-2r^2 N / \rho^2} \mathbb{P}\{\xi \in \mathcal{R}\}. \quad (8.9)$$

STEP 3. The untranslated event. The variables $\zeta_a = b_a \cdot P_\sigma \xi = (P_\sigma b_a) \cdot \xi$, $a \leq M$, are centered and jointly Gaussian, with $\text{Var } \zeta_a = \|P_\sigma b_a\|^2 \leq K^2$, and $\{\xi \in \mathcal{R}\} = \{|\zeta_a| \leq \eta / \rho \text{ for all } a\}$. By Lemma 3.12, the probability of this event is at least $\prod_a \mathbb{P}\{|\zeta_a| \leq \eta / \rho\}$. For $z \geq 0$ the Gaussian tail bound $2\mathbb{P}\{G > z\} \leq e^{-z^2/2}$, together with the fact that $\mathbb{P}\{|N(0, v)| > z\}$ is nondecreasing in v , gives $\mathbb{P}\{|\zeta_a| > \eta / \rho\} \leq e^{-\eta^2 / (2K^2 \rho^2)} \leq 1/2$. Using $\log(1 - x) \geq -2x$ for $0 \leq x \leq 1/2$, we obtain

$$\mathbb{P}\{\xi \in \mathcal{R}\} \geq (1 - e^{-\eta^2 / (2K^2 \rho^2)})^M \geq \exp\{-2Me^{-\eta^2 / (2K^2 \rho^2)}\}.$$

Together with (8.9) this proves (8.6).

STEP 4. Integration. Let Σ be a random point with law μ_N , independent of ξ . We first show that $Y(\Sigma, \xi)$ also has law μ_N . For an orthogonal matrix O we have $P_{O\sigma} = OP_\sigma O^\top$, and hence $Y(O\sigma, O\xi) = OY(\sigma, \xi)$. The pair $(O\Sigma, O\xi)$ has the same law as (Σ, ξ) , so $OY(\Sigma, \xi) = Y(O\Sigma, O\xi)$ has the same law as $Y(\Sigma, \xi)$. The law ν of $Y(\Sigma, \xi)$ is therefore a Borel probability measure on S_N that is invariant under all orthogonal transformations, and such a measure equals μ_N . Indeed, let H be the Haar probability measure on the compact group $O(N)$, which is invariant under multiplication from the left and from the right. Let f be a bounded Borel function on S_N . For $\sigma, \sigma' \in S_N$ choose $O_0 \in O(N)$ with $\sigma' = O_0 \sigma$. The right invariance of H gives $\int f(O\sigma') H(dO) = \int f(OO_0\sigma) H(dO) = \int f(O\sigma) H(dO)$, so this integral is a constant c_f that does not depend on σ . By the invariance of ν and Fubini's theorem, $\int f d\nu = \int \int f(O\sigma) H(dO) \nu(d\sigma) = c_f$, and the same computation gives $\int f d\mu_N = c_f$. Hence $\nu = \mu_N$.

The map Y is continuous and $S_N(k')$ is closed, so by independence and Fubini's theorem $\sigma \mapsto \mathbb{P}\{Y(\sigma, \xi) \in S_N(k')\}$ is Borel and

$$\begin{aligned} V_N(k') &= \mathbb{P}\{Y(\Sigma, \xi) \in S_N(k')\} \geq \mathbb{P}\{\Sigma \in B, Y(\Sigma, \xi) \in S_N(k')\} \\ &= \int_B \mathbb{P}\{Y(\sigma, \xi) \in S_N(k')\} \mu_N(d\sigma). \end{aligned}$$

By (a), the integrand is at least the right side of (8.6) for every $\sigma \in B$. This proves (8.7). $\square$

No measurable choice of the repair displacement is needed in (b): the lower bound (8.6) holds at every point of B .

8.3 The hard-volume lower bound

We now define the penalties to which the Parisi formula is applied. Fix $\chi \in C^\infty(\mathbb{R})$ with values in $[0, 1]$, $\chi = 0$ on $(-\infty, 0]$ and $\chi = 1$ on $[1, \infty)$. For $\omega > 0$, $\beta > 0$, and $\ell \in \mathbb{R}$ put

$$h_\omega(y) = \int_0^y \int_0^{y'} 2\chi(v/\omega) dv dy', \quad U_{\beta,\ell}(x) = -\beta h_\omega(\ell - x).$$

The dependence of $U_{\beta,\ell}$ on ω and χ is suppressed. Since $h''_\omega(y) = 2\chi(y/\omega)$ vanishes for $y \leq 0$, lies in $[0, 2]$ and equals 2 for $y \geq \omega$, and since $h_\omega(0) = h'_\omega(0) = 0$, the function h_ω vanishes on $(-\infty, 0]$ and is nondecreasing. Moreover $h'_\omega(y) \geq 2(y - \omega)$ for $y \geq \omega$, so that $h_\omega(y) \geq (y - \omega)_+^2$ for all y . Consequently:

- (P1) $U_{\beta,\ell} = 0$ on $[\ell, \infty)$ and $U_{\beta,\ell}(x) \leq -\beta(\ell - \omega - x)_+^2$ for $x \in \mathbb{R}$. In particular $H_\ell \leq U_{\beta,\ell} \leq 0$, where $H_\ell(x) = \log \mathbf{1}_{\{x \geq \ell\}}$.
- (P2) $U_{\beta,\ell}$ is smooth and concave, with $\sup U_{\beta,\ell} = 0$ and $-2\beta \leq U''_{\beta,\ell} \leq 0$. Thus $U_{\beta,\ell}$ belongs to class (B).

The following theorem gives the lower bounds in Theorem 1.2(i),(ii). The restriction to negative margins enters only through the transfer of Lemma 8.4.

Theorem 8.5. *Let $\alpha > 0$ and $\kappa < \min\{\kappa_c(\alpha), 0\}$. Then $\mathcal{P}_*(\kappa)$ is finite, and for every $\varepsilon > 0$,*

$$\mathbb{P}\left\{\frac{1}{N} \log V_N(\kappa) \geq \mathcal{P}_*(\kappa) - \varepsilon\right\} \longrightarrow 1. \quad (8.10)$$

Consequently $\mathbb{P}\{\kappa_N \geq \kappa\} \rightarrow 1$.

Proof. We may assume $\varepsilon \leq 1$. We divide the proof into five steps: regularity of $\mathcal{P}_*$, a lower bound for the free energy of one penalty, the choice of parameters, the passage from the free energy to the volume of approximately feasible configurations, and the repair and transfer of Lemmas 8.2 and 8.4.

STEP 1. Regularity of $\mathcal{P}_*$. By Theorem 6.8, the function $\mathcal{P}_*$ is finite, nonincreasing, and continuous on $(-\infty, \kappa_c)$. In particular $\mathcal{P}_*(\kappa)$ is finite and $\mathcal{P}_*$ is continuous at κ .

STEP 2. The penalty free energy. Let $\ell \in \mathbb{R}$ and $\beta, \omega > 0$, and write $U = U_{\beta,\ell}$. We show that for every $\varepsilon' > 0$,

$$\mathbb{P}\left\{\frac{1}{N} \log Z_N(U) \geq \mathcal{P}_*(\ell) - \varepsilon'\right\} \longrightarrow 1. \quad (8.11)$$

By (P2) and Theorem 1.1, $N^{-1} \log Z_N(U) = f_N(U) \rightarrow \mathcal{P}_*(U)$ in probability, so it suffices to prove $\mathcal{P}_*(U) \geq \mathcal{P}_*(\ell)$. Let $\gamma \in \mathcal{U}$ and $r \in [Q_\gamma, 1)$. First let $\gamma \in \mathcal{U}_r$ be a step order parameter. By (P1), $H_\ell \leq U$, so (2.4) and (2.10) give $u_\gamma(0, 0; U) \geq u_\gamma(0, 0; H_\ell) = u_\gamma(0, 0; \ell)$. The entropy term $\text{CS}(\gamma)$ does not depend on the activation, so $\mathcal{P}_U(\gamma) \geq \mathcal{P}_\ell(\gamma) \geq \mathcal{P}_*(\ell)$. For general $\gamma \in \mathcal{U}$, choose step order parameters $\gamma_n \in \mathcal{U}_r$ with $\gamma_n \rightarrow \gamma$ in $L^1([0, 1])$ (§2.1).

Then $\mathcal{P}_U(\gamma_n) \geq \mathcal{P}_*(\ell)$ for every n . Since U belongs to class (B), (2.20) in Lemma 2.4 gives $\mathcal{P}_U(\gamma_n) \rightarrow \mathcal{P}_U(\gamma)$. Hence $\mathcal{P}_U(\gamma) \geq \mathcal{P}_*(\ell)$ for every $\gamma \in \mathcal{U}$, that is, $\mathcal{P}_*(U) \geq \mathcal{P}_*(\ell)$, and (8.11) follows.

STEP 3. Choice of parameters. Let s and K be the constants of Lemma 8.1. They depend only on α . We choose the remaining parameters in the following order, and each choice depends only on the earlier ones and on $\alpha, \kappa, \varepsilon$.

- (i) The level ℓ . By Step 1, choose $\ell \in (\kappa, \min\{\kappa_c, 0\})$ with $\mathcal{P}_*(\ell) \geq \mathcal{P}_*(\kappa) - \varepsilon/4$. Then $\mathcal{P}_*(\ell)$ is finite. Put $\vartheta = (\ell - \kappa)/5$. The margin decreases from ℓ to κ in four stages: the penalty has transition width $\omega = \vartheta$, the repair of Lemma 8.2 with $\tau = \vartheta$ loses 2ϑ , and the transfer of Lemma 8.4 with $\eta = \vartheta$ loses at most ϑ through the factor $1/(1-r)$ and ϑ through η .
- (ii) The perturbation size $\rho > 0$, chosen so that $2(\alpha + 1) \exp(-\vartheta^2/(2K^2\rho^2)) \leq \varepsilon/8$. Since $M \leq (\alpha + 1)N$ and $\varepsilon \leq 1$, this gives $2M \exp(-\vartheta^2/(2K^2\rho^2)) \leq N\varepsilon/8$ and $\exp(-\vartheta^2/(2K^2\rho^2)) \leq 1/16$.
- (iii) The violation level $t > 0$, chosen so small that

$$t \leq \frac{\vartheta^2 s}{32K^2}, \quad \frac{32t}{\rho^2} \leq \frac{\varepsilon}{8}, \quad r := 4\sqrt{t} \leq \min\left\{\frac{1}{2}, \frac{\vartheta}{2|\kappa|}\right\}.$$

The first condition is the hypothesis of Lemma 8.2 with $\tau = \vartheta$. The second bounds the shift cost $2r^2/\rho^2 = 32t/\rho^2$ in Lemma 8.4. The third is the condition $r \leq 1/2$ of that lemma and bounds the margin lost in the factor $1/(1-r)$.

- (iv) The penalty strength $\beta = (|\mathcal{P}_*(\ell)| + 1)/t$, so that configurations with total squared violation above tN carry weight at most $e^{-N(|\mathcal{P}_*(\ell)| + 1)}$ in $Z_N(U)$.

Write $U = U_{\beta, \ell}$ with $\omega = \vartheta$, and put $k_0 = \ell - \vartheta$ and $k = \ell - 3\vartheta < 0$.

STEP 4. Approximately feasible configurations. Let

$$B = \{\sigma \in S_N : \|(k_0 \mathbf{1} - A\sigma)_+\|^2 \leq tN\},$$

a closed subset of S_N . By (P1), $\exp\{\sum_a U(b_a \cdot \sigma)\} \leq \exp\{-\beta\|(k_0 \mathbf{1} - A\sigma)_+\|^2\}$, which is at most 1 on B and at most $e^{-\beta tN}$ off B . Hence

$$Z_N(U) \leq \mu_N(B) + e^{-\beta tN} \leq \mu_N(B) + e^{N(\mathcal{P}_*(\ell) - 1)},$$

because $-\beta t = -|\mathcal{P}_*(\ell)| - 1 \leq \mathcal{P}_*(\ell) - 1$. On the event in (8.11) with $\varepsilon' = \varepsilon/8$ we have $Z_N(U) \geq e^{N(\mathcal{P}_*(\ell) - \varepsilon/8)}$. Since $\mathcal{P}_*(\ell) - 1 \leq \mathcal{P}_*(\ell) - \varepsilon/8 - 7/8$,

$$\mu_N(B) \geq e^{N(\mathcal{P}_*(\ell) - \varepsilon/8)} (1 - e^{-7N/8}) \geq e^{N(\mathcal{P}_*(\ell) - \varepsilon/4)}$$

as soon as $N \geq 8 \log 2/\varepsilon$, since then $1 - e^{-7N/8} \geq 1/2 \geq e^{-N\varepsilon/8}$.

STEP 5. Repair and transfer. By Lemma 8.1 and Step 4, with probability tending to one the event E_N occurs and $\mu_N(B) \geq e^{N(\mathcal{P}_*(\ell) - \varepsilon/4)}$. We work on this event, with $N \geq \max\{1/s, 8 \log 2/\varepsilon\}$.

Let $\sigma \in B$. Lemma 8.2, applied with $u = \sigma$, $k_0 = \ell - \vartheta$, $\tau = \vartheta$, and t (condition (iii)), gives $d \in \mathbb{R}^N$ with $A(\sigma + d) \geq (k_0 - 2\vartheta)\mathbf{1} = k\mathbf{1}$ and $\|d\| \leq 4\sqrt{tN} = r\sqrt{N}$. On E_N we have $\max_a \|b_a\| \leq \|A\|_{\text{op}} \leq K$, because $b_a = A^\top e_a$ for the a -th unit vector e_a of $\mathbb{R}^M$. Condition (ii) gives $\exp(-\vartheta^2/(2K^2\rho^2)) \leq 1/2$, so Lemma 8.4(b) applies with this K , the margin $k < 0$, $r \leq 1/2$, $\eta = \vartheta$, and ρ . By conditions (ii) and (iii), it gives

$$V_N(k') \geq \mu_N(B) \exp\left\{-\frac{32tN}{\rho^2} - 2Me^{-\vartheta^2/(2K^2\rho^2)}\right\} \geq \mu_N(B)e^{-N\varepsilon/4}, \quad k' = \frac{k}{1-r} - \vartheta.$$

The margin k' is at least κ . Indeed, since $\kappa < k < 0$ and $r \leq 1/2$,

$$\frac{k}{1-r} = k - \frac{r|k|}{1-r} \geq k - 2r|k| \geq k - 2r|\kappa| \geq k - \vartheta,$$

by condition (iii), and so $k' \geq \ell - 3\vartheta - \vartheta - \vartheta = \kappa$. Hence $S_N(k') \subseteq S_N(\kappa)$ and

$$\frac{1}{N} \log V_N(\kappa) \geq \frac{1}{N} \log \mu_N(B) - \frac{\varepsilon}{4} \geq \mathcal{P}_*(\ell) - \frac{\varepsilon}{2} \geq \mathcal{P}_*(\kappa) - \frac{3\varepsilon}{4}.$$

This proves (8.10).

On the same event $V_N(\kappa) \geq e^{N(\mathcal{P}_*(\kappa) - \varepsilon)} > 0$, because $\mathcal{P}_*(\kappa)$ is finite. Then $S_N(\kappa)$ is nonempty, that is, some $\sigma \in S_N$ has $\min_a b_a \cdot \sigma \geq \kappa$, and $\kappa_N \geq \kappa$. Hence $\mathbb{P}\{\kappa_N \geq \kappa\} \rightarrow 1$. $\square$

8.4 Proof of the capacity theorem

Proof of Theorem 1.2. Let $\alpha > 2$. By Proposition 6.7, $-\infty < \kappa_c(\alpha) \leq \kappa_{\text{RS}}(\alpha) < \infty$ and, since $\alpha > 2$, $\kappa_c(\alpha) < 0$.

Proof of (i). Fix $\kappa < \kappa_c$ and $\varepsilon > 0$. Since $\kappa < \kappa_c < 0$, Theorem 8.5 shows that $\mathcal{P}_*(\kappa)$ is finite and that

$$\mathbb{P}\left\{\frac{1}{N} \log V_N(\kappa) \geq \mathcal{P}_*(\kappa) - \varepsilon\right\} \rightarrow 1.$$

For the upper bound choose $\gamma \in \mathcal{U}$ with $\mathcal{P}_\kappa(\gamma) \leq \mathcal{P}_*(\kappa) + \varepsilon/2$. Corollary 3.11(i), applied with this γ , the margin κ , and $\varepsilon/2$, gives $\mathbb{P}\{N^{-1} \log V_N(\kappa) > \mathcal{P}_*(\kappa) + \varepsilon\} \rightarrow 0$. The two bounds prove (i).

Proof of (ii). Let $\varepsilon > 0$. By Theorem 8.5 at the margin $\kappa_c - \varepsilon < 0$, $\mathbb{P}\{\kappa_N < \kappa_c - \varepsilon\} \rightarrow 0$. By Corollary 3.14, $\mathbb{P}\{\kappa_N \geq \kappa_c + \varepsilon\} \rightarrow 0$. Hence $\kappa_N \rightarrow \kappa_c$ in probability.

Proof of (iii). By Theorem 6.8, $\mathcal{P}_*(\kappa) = -\infty$ for every $\kappa \geq \kappa_c$. Let $\kappa \geq \kappa_c$ and $C < \infty$. Choose $\gamma \in \mathcal{U}$ with $\mathcal{P}_\kappa(\gamma) < -C - 1$, and apply Corollary 3.11(i) with this γ , the margin κ , and $\varepsilon = 1$:

$$\mathbb{P}\left\{\frac{1}{N} \log V_N(\kappa) > -C\right\} \leq \mathbb{P}\left\{\frac{1}{N} \log V_N(\kappa) > \mathcal{P}_\kappa(\gamma) + 1\right\} \longrightarrow 0.$$

This includes the critical margin $\kappa = \kappa_c$. □

The assumption $\alpha > 2$ enters the proof of part (i) only through the inequality $\kappa_c(\alpha) < 0$. Theorem 8.5 and Corollary 3.11 hold for every $\alpha > 0$, so the proof of part (i) shows that $N^{-1} \log V_N(\kappa) \rightarrow \mathcal{P}_*(\kappa)$ in probability for every $\alpha > 0$ and every $\kappa < \min\{\kappa_c(\alpha), 0\}$. Such κ exist because $\kappa_c(\alpha) > -\infty$ by Proposition 6.7.

Gap statistics

This chapter proves Theorems 1.8 and 1.9; Step 6 of §1.3 outlines the argument. Both theorems concern averages of a test function ψ over the gaps $b_a \cdot \sigma - k$ of the constraints, and both rest on the deviation inequality of Proposition 9.8. It compares such averages over every set $T \subseteq S_N(k)$ of positive volume with $\alpha \mathbb{E} \psi(X_1^{\gamma, k} - k)$, up to an error that involves a source strength $\beta > 0$, the volume term $\mathcal{P}_k(\gamma) - N^{-1} \log \mu_N(T)$, and the mass $\lambda_\gamma(0)$.

§9.1 counts the contacts at optimal configurations. §9.2 writes the Parisi equation for a step order parameter as a finite recursion, introduces the Markov chain that describes the linearization of this recursion, and identifies that chain with the Parisi diffusion. §9.3 proves the quadratic source estimate, and §9.4 combines it with Jensen's inequality and the upper bound of Chapter 3 to prove Proposition 9.8. For Theorem 1.9, at a fixed margin $\kappa < \kappa_c$, T consists of the configurations in $S_N(\kappa)$ whose gap averages exceed, or fall below, $\alpha \mathbb{E} \psi(X_1^{\gamma, \kappa} - \kappa)$ by more than ε , for a minimizer γ of $\mathcal{P}_\kappa$ (§9.5). For Theorem 1.8, T is the set of Lemma 3.13 near one configuration σ_0 of nearly maximal margin (§9.6), and §9.7 combines this with the collapse of minimizers (Proposition 6.11) and with the contacts of Proposition 9.2.

Notation. Throughout this chapter, $\alpha_N = M/N \rightarrow \alpha$ and $\bar{\alpha} = \sup_N \alpha_N < \infty$, and $A = N^{-1/2} \mathbf{G}$ is the matrix with rows b_a , as in Chapter 8. The letter k denotes a real margin. A *test function* is a bounded Lipschitz function $\psi : [0, \infty) \rightarrow \mathbb{R}$; we extend it to $\mathbb{R}$ by $\psi(x) = \psi(0)$ for $x < 0$, which preserves its bound and its Lipschitz constant. We call ψ L -Lipschitz if $|\psi(x) - \psi(y)| \leq L|x - y|$ for all x, y . For $\sigma \in S_N$ and $k, \beta \in \mathbb{R}$ put

$$\Psi_N(\sigma, k) = \frac{1}{N} \sum_{a=1}^M \psi(b_a \cdot \sigma - k), \quad f_N(k, \beta) = \frac{1}{N} \log \int_{S_N(k)} e^{N\beta \Psi_N(\sigma, k)} \mu_N(d\sigma),$$

so that $f_N(k, 0) = N^{-1} \log V_N(k)$, and $f_N(k, \beta) = f_N(H_k + \beta \psi(\cdot - k))$ in the notation of (1.1). The test function is suppressed from the notation. For a Borel set $T \subseteq S_N$ with $\mu_N(T) > 0$, $\langle \cdot \rangle_T$ denotes expectation under the probability measure $\mu_N(\cdot \cap T) / \mu_N(T)$, and $\langle \cdot \rangle_{N, k} = \langle \cdot \rangle_{S_N(k)}$ when $V_N(k) > 0$. Thus $\int \psi d\nu_{N, \sigma} = \Psi_N(\sigma, m_N(\sigma))$, and the average in Theorem 1.9 is $\langle \Psi_N(\sigma, \kappa) \rangle_{N, \kappa}$. Several statements below involve a supremum over a random set of configurations, such as the set of maximizers of m_N . A statement that such an event has probability tending to one, has probability at least $1 - e^{-cN}$, or occurs almost

surely means that it contains an event, defined without reference to the random set, with the same property. In particular no measurable selection of maximizers is needed.

Let $C_{\text{op}} = K$ be the constant of Lemma 8.1, which depends only on α . By that lemma, $\|A\|_{\text{op}} \leq C_{\text{op}}$ with probability tending to one. Each $N\|b_a\|^2$ has the chi-square law with N degrees of freedom, and Markov's inequality applied to $e^{N\|b_a\|^2/4}$, whose mean is $2^{N/2}$, gives $\mathbb{P}\{N\|b_a\|^2 \geq 2N\} \leq e^{-N/2} 2^{N/2} = e^{-N(1-\log 2)/2}$. A union bound over $a \leq M$ gives $\mathbb{P}\{\max_a \|b_a\|^2 > 2\} \leq \bar{\alpha} N e^{-N(1-\log 2)/2} \rightarrow 0$. Hence the event

$$\Omega_N = \left\{ \|A\|_{\text{op}} \leq C_{\text{op}}, \max_{a \leq M} \|b_a\|^2 \leq 2 \right\}$$

has probability tending to one. On Ω_N the Cauchy–Schwarz inequality gives, for every $\sigma \in S_N$,

$$\frac{1}{N} \sum_{a \leq M} |b_a \cdot \sigma| \leq \frac{\sqrt{M}}{N} \|A\sigma\| \leq \sqrt{\alpha_N} C_{\text{op}}. \quad (9.1)$$

9.1 Contacts at optimal configurations

We first count the contacts at maximizers of m_N . The argument uses only the Gaussian matrix, and not the variational problem. Its probabilistic input is Wendel's theorem on random half-spaces, which we restate for Gaussian vectors.

Lemma 9.1. *Let $M, N \geq 1$ and let $g_1, \dots, g_M$ be independent standard Gaussian vectors in $\mathbb{R}^N$. Then*

$$\mathbb{P}\{\text{there is } y \in \mathbb{R}^N \text{ with } g_a \cdot y > 0 \text{ for all } a \leq M\} = 2^{1-M} \sum_{j=0}^{N-1} \binom{M-1}{j}. \quad (9.2)$$

Proof. This is [131, (1)]. Wendel proved it for random vectors $x_1, \dots, x_M$ in $\mathbb{R}^N$ whose joint law is invariant under the reflection of each x_a through the origin and such that, with probability one, every N of them are linearly independent. The probability computed there is that of the event that the inner products of all x_a with some vector y are positive. Both hypotheses hold here. The g_a are independent and have symmetric laws. For every set of N rows, the determinant of the corresponding $N \times N$ matrix is a nonzero polynomial in its N^2 independent Gaussian entries, and therefore vanishes with probability zero. See also [43, Theorem 1 and (20)]. $\square$

The right side of (9.2) is the probability that a binomial variable with parameters $M-1$ and $1/2$ is at most $N-1$. For $M/N \rightarrow \alpha > 2$ it tends to zero exponentially fast, and this is the only place where $\alpha > 2$ enters the count of contacts.

To describe the maximizers we use the following objects. For a set S of N rows such that $\{g_a : a \in S\}$ is linearly independent, let $y_S \in \mathbb{R}^N$ be the unique solution of $g_a \cdot y_S = -1$ for $a \in S$, and put

$$u_S = \frac{y_S}{\|y_S\|}, \quad k_S = -\frac{1}{\|y_S\|}. \quad (9.3)$$

Recall that a contact of σ is a row a with $b_a \cdot \sigma = m_N(\sigma)$.

Proposition 9.2. *Let $\alpha > 2$.*

(i) *For every $c < (\alpha - 2)^2 / (2\alpha)$ and all sufficiently large N ,*

$$\mathbb{P}\{\kappa_N \geq 0\} \leq e^{-cN}. \quad (9.4)$$

(ii) *Almost surely the following holds. If $\kappa_N < 0$, then every maximizer σ_* of m_N has exactly N contacts, the set S of its contacts has linearly independent rows $\{g_a : a \in S\}$, and, with the notation (9.3),*

$$\sigma_* = \sqrt{N} u_S, \quad \kappa_N = k_S. \quad (9.5)$$

(iii) *Let $0 < t < \sqrt{2\pi}$ and $\varepsilon > \alpha \log 3 / \log(\sqrt{2\pi}/t)$. With probability tending to one, every maximizer σ_* of m_N satisfies*

$$\#\{a \leq M : 0 < b_a \cdot \sigma_* - \kappa_N \leq t\} < \varepsilon N. \quad (9.6)$$

Proof. Maximizers of m_N exist because m_N is continuous on the compact set S_N . If $M < N$, some $\sigma \in S_N$ is orthogonal to all rows, so $\kappa_N \geq 0$ and (ii) is vacuous. Since $\alpha > 2$, we have $M \geq N$ for large N , and we assume $M \geq N$ from now on. Let $\mathcal{G}$ be the event on which every N of the rows g_a are linearly independent and, for every set S of N rows and every $a \notin S$, $g_a \cdot y_S \neq -1$. We claim that $\mathbb{P}(\mathcal{G}) = 1$. The first requirement holds almost surely by the determinant argument in the proof of Lemma 9.1. For the second, fix S and $a \notin S$. Conditionally on the rows in S , the vector g_a is standard Gaussian, so $g_a \cdot y_S$ is Gaussian with variance $\|y_S\|^2 > 0$ and equals -1 with probability zero. A union over the finitely many pairs (S, a) proves the claim. We divide the rest of the proof into four steps.

STEP 1. *Proof of (i).* Suppose that $\mathcal{G}$ holds and $\kappa_N \geq 0$, and let σ be a maximizer of m_N , so that $g_a \cdot \sigma \geq 0$ for all a . Let $J = \{a : g_a \cdot \sigma = 0\}$. The rows g_a , $a \in J$, lie in the $(N-1)$ -dimensional space $\sigma^\perp$. On $\mathcal{G}$ every N rows are linearly independent, so $|J| \leq N-1$ and the rows in J are linearly independent. Hence there is $v \in \mathbb{R}^N$ with $g_a \cdot v = 1$ for all $a \in J$, and for small $\tau > 0$ the vector $y = \sigma + \tau v$ satisfies $g_a \cdot y > 0$ for all a . By Lemma 9.1,

$$\mathbb{P}\{\kappa_N \geq 0\} \leq \mathbb{P}\{Y_{M-1} \leq N-1\},$$

where Y_n denotes a binomial variable with parameters n and $1/2$. For $0 < h < 1/2$ and $s > 0$, Markov's inequality and $\cosh x \leq e^{x^2/2}$ give

$$\mathbb{P}\{Y_n \leq n/2 - nh\} \leq e^{s(n/2 - nh)} \mathbb{E}e^{-sY_n} = e^{-snh} \cosh(s/2)^n \leq e^{-snh + ns^2/8},$$

and the choice $s = 4h$ gives $\mathbb{P}\{Y_n \leq n/2 - nh\} \leq e^{-2nh^2}$. Apply this with $n = M - 1$ and $h_N = \frac{1}{2} - \frac{N-1}{M-1}$, which is positive for large N because $(N-1)/(M-1) \rightarrow 1/\alpha < 1/2$. Since

$$\frac{2(M-1)h_N^2}{N} \longrightarrow 2\alpha \left(\frac{1}{2} - \frac{1}{\alpha} \right)^2 = \frac{(\alpha - 2)^2}{2\alpha},$$

this proves (9.4).

STEP 2. Maximizers as farthest points of a polytope. Suppose that $\kappa_N < 0$. Consider the polyhedron

$$K_N = \{y \in \mathbb{R}^N : g_a \cdot y \geq -1 \text{ for all } a \leq M\},$$

which is closed, convex, and contains 0. It is bounded. Indeed, if $y_n \in K_N$ and $\|y_n\| \rightarrow \infty$, then a limit point d of $y_n/\|y_n\|$ satisfies $\|d\| = 1$ and $g_a \cdot d \geq 0$ for all a , and $\sigma = \sqrt{N}d$ would have $m_N(\sigma) \geq 0 > \kappa_N$. For $\sigma \in S_N$ we have $m_N(\sigma) \leq \kappa_N < 0$, and $y = \sigma/(\sqrt{N}|m_N(\sigma)|)$ lies in K_N with $\|y\| = 1/|m_N(\sigma)|$. Conversely, if $y \in K_N \setminus \{0\}$, then $\sigma = \sqrt{N}y/\|y\|$ satisfies $m_N(\sigma) \geq -1/\|y\|$, so that $\|y\| \leq 1/|m_N(\sigma)| \leq 1/|\kappa_N|$. Hence $\max_{K_N} \|y\| = 1/|\kappa_N|$, the farthest points of K_N from the origin are exactly the points $y_* = \sigma_*/(\sqrt{N}|\kappa_N|)$ for maximizers σ_* , and $b_a \cdot \sigma_* = \kappa_N$ if and only if $g_a \cdot y_* = -1$. This description of the largest negative margin through the largest norm of a point of a polytope appears in [90, Section 2.2].

A farthest point y_* is an extreme point of K_N : if $y_* = \frac{1}{2}(y_1 + y_2)$ with $y_1 \neq y_2$ in K_N , strict convexity gives $\|y_*\|^2 < \frac{1}{2}(\|y_1\|^2 + \|y_2\|^2) \leq \max_{K_N} \|y\|^2$. At an extreme point the active rows $\{g_a : g_a \cdot y_* = -1\}$ span $\mathbb{R}^N$. Otherwise some $v \neq 0$ is orthogonal to all of them, the inactive rows satisfy $g_a \cdot y_* > -1$, and $y_* \pm \tau v \in K_N$ for small $\tau > 0$. Hence every maximizer has at least N contacts, and they contain N linearly independent rows. The relation between extreme points and active constraints is standard in linear programming; see [18, Theorem 2.3].

STEP 3. Proof of (ii). Suppose that $\mathcal{G}$ holds and $\kappa_N < 0$, let σ_* be a maximizer, and let y_* be as in Step 2. Choose a set S of N linearly independent contacts. Then $y_* = y_S$. If σ_* had a further contact $a \notin S$, then $g_a \cdot y_S = -1$, which is excluded on $\mathcal{G}$. Hence S is the set of contacts of σ_* , and $\|y_S\| = \|y_*\| = 1/|\kappa_N|$ gives (9.5).

STEP 4. Proof of (iii). Put $j_N = \lceil \varepsilon N \rceil$. If $j_N > M - N$, then on $\mathcal{G} \cap \{\kappa_N < 0\}$ the asserted bound is automatic, since there are only $M - N$ noncontacts. We may therefore assume $j_N \leq M - N$ in the counting argument below. Suppose that $\mathcal{G}$ holds, $\kappa_N < 0$, and some maximizer σ_* violates (9.6). Let S be its set of contacts. By (9.5), $b_a \cdot \sigma_* - \kappa_N = g_a \cdot u_S - k_S$ for every a . Hence there are a set S of N linearly independent rows and a set J of j_N rows

outside S such that $g_a \cdot u_S \in (k_S, k_S + t]$ for all $a \in J$. Fix S and J . Conditionally on the rows in S , the unit vector u_S and the number k_S are fixed, and the variables $g_a \cdot u_S$, $a \notin S$, are independent standard Gaussian variables. Since the standard Gaussian density is at most $1/\sqrt{2\pi}$, the conditional probability of the event for (S, J) is at most $(t/\sqrt{2\pi})^{j_N}$. The number of pairs (S, J) is the multinomial coefficient $M!/(N! j_N! (M - N - j_N)!)$, which is at most 3^M . By a union bound, by $t/\sqrt{2\pi} < 1$, and by (i),

$$\begin{aligned} \mathbb{P}\{\text{some maximizer violates (9.6)}\} &\leq \mathbb{P}\{\kappa_N \geq 0\} + 3^M \left(\frac{t}{\sqrt{2\pi}}\right)^{\varepsilon N} \\ &= \mathbb{P}\{\kappa_N \geq 0\} + \exp\left\{-N\left(\varepsilon \log \frac{\sqrt{2\pi}}{t} - \alpha_N \log 3\right)\right\}. \end{aligned}$$

Both terms tend to zero, the second because $\alpha_N \rightarrow \alpha$ and $\varepsilon \log(\sqrt{2\pi}/t) > \alpha \log 3$. $\square$

Parts (i) and (ii) show that, with probability at least $1 - e^{-cN}$, every optimal configuration is isostatic in the sense recalled in Chapter 1: its N contacts match the $N - 1$ degrees of freedom on the sphere together with the optimized margin. In the replica computation of Franz, Parisi, Sevelev, Urbani, and Zamponi, isostaticity at jamming for $\kappa < 0$ is derived from the scaling of the full replica symmetry breaking solution [59, (100)]. Part (iii) is elementary but not sharp. In the proof of Theorem 1.8 it gives the bound (1.13) on the limiting gap law.

9.2 The finite recursion and the Parisi chain

Gap averages enter the Parisi functional through its terminal datum. For a step order parameter the Parisi equation is an explicit recursion, which accepts a hard wall with a source as terminal datum. We set up this recursion, define the Markov chain $\mathcal{Q}$ that describes its linearization, and identify $\mathcal{Q}$ with the Parisi diffusion.

Let $\gamma \in \mathcal{U}$ be a step order parameter. We write

$$\begin{aligned} 0 = t_0 < t_1 < \cdots < t_n < t_{n+1} = 1, \quad 0 \leq m_0 \leq m_1 \leq \cdots \leq m_n = 1, \\ \gamma = m_j \text{ on } [t_j, t_{j+1}), \end{aligned}$$

and $\Delta_j = t_{j+1} - t_j$. Then $Q_\gamma \leq t_n$ and

$$\lambda_\gamma(t_j) = \sum_{i=j}^n m_i \Delta_i \quad (0 \leq j \leq n), \quad \lambda_\gamma(0) = \sum_{j=0}^n m_j \Delta_j. \quad (9.7)$$

The representation need not be minimal: we allow $t_n > Q_\gamma$ and consecutive steps with equal values, so that further breakpoints may always be inserted. For a measurable

terminal datum $F : \mathbb{R} \rightarrow [-\infty, \infty)$ that is bounded above we define $u_\gamma(\cdot, \cdot; F)$ at the breakpoints by $u_\gamma(1, x; F) = F(x)$ and, for $j = n, n-1, \dots, 0$,

$$u_\gamma(t_j, x; F) = \begin{cases} \frac{1}{m_j} \log \mathbb{E} \exp\{m_j u_\gamma(t_{j+1}, x + \sqrt{\Delta_j} G; F)\}, & m_j > 0, \\ \mathbb{E} u_\gamma(t_{j+1}, x + \sqrt{\Delta_j} G; F), & m_j = 0, \end{cases} \quad (9.8)$$

where G is a standard Gaussian variable. This is the recursion (2.2) with terminal time one, evaluated at the breakpoints. Every function in the recursion is bounded above by $\sup F$. By (T2), inserting breakpoints does not change $u_\gamma(t_j, \cdot; F)$ at the original breakpoints, and $u_\gamma(0, 0; F)$ does not depend on the representation (§2.1).

The hard wall. For $F = H_k = \log \mathbf{1}_{[k, \infty)}$, (2.10) with $\bar{q} = t_n$ gives

$$u_j^0 := u_\gamma(t_j, \cdot; H_k) = u_\gamma(t_j, \cdot; k), \quad j \leq n,$$

the hard-wall solution computed with the cutoff t_n ; in particular $u_n^0(x) = \log \Phi((x - k)/\sqrt{1 - t_n})$. We put $u_{n+1}^0 = H_k$. By Lemma 2.6(i), (iii), and (v), applied with $r = t_n$ and $\kappa = k$, each u_j^0 with $j \leq n$ is smooth and concave on $\mathbb{R}$ (since $-\partial_{xx} u_j^0 \geq 0$ by (iii)), and $-C(1 + x^2) \leq u_j^0(x) \leq 0$.

The source. For a test function ψ and $\beta \in \mathbb{R}$ we use the terminal datum $H_k + \beta\psi(\cdot - k)$ and write $u_j^\beta = u_\gamma(t_j, \cdot; H_k + \beta\psi(\cdot - k))$. The source $\beta\psi(\cdot - k)$ is bounded by $|\beta|\|\psi\|_\infty$, so (2.4), applied in both directions with $a = |\beta|\|\psi\|_\infty$, gives $|u_j^\beta - u_j^0| \leq |\beta|\|\psi\|_\infty$ for $j \leq n$. In particular every u_j^β with $j \leq n$ is finite. For a terminal datum F as above we put

$$\mathcal{P}_F(\gamma) = \alpha u_\gamma(0, 0; F) + \text{CS}(\gamma), \quad \mathcal{P}_{k, \beta\psi}(\gamma) = \mathcal{P}_{H_k + \beta\psi(\cdot - k)}(\gamma).$$

Thus $\mathcal{P}_U(\gamma)$ and $\mathcal{P}_{k, 0}(\gamma) = \mathcal{P}_k(\gamma)$ are the functionals of §1.1.

The chain. Fix $k \in \mathbb{R}$ and let $p_\Delta(x, z) = (2\pi\Delta)^{-1/2} e^{-(z-x)^2/(2\Delta)}$ be the Gaussian kernel. Let $\mathcal{Q}$ be the law of the Markov chain $X_{t_0}, X_{t_1}, \dots, X_{t_n}, X_1$ with $X_{t_0} = 0$ and transitions

$$\mathcal{Q}_j(dz | x) = p_{\Delta_j}(x, z) \exp\{m_j u_{j+1}^0(z) - m_j u_j^0(x)\} dz, \quad j = 0, \dots, n, \quad (9.9)$$

from $X_{t_j} = x$ to $X_{t_{j+1}}$. We call $\mathcal{Q}$ the *Parisi chain*. By (9.8), $e^{m_j u_j^0(x)} = \int p_{\Delta_j}(x, z) e^{m_j u_{j+1}^0(z)} dz$, so each $\mathcal{Q}_j(\cdot | x)$ is a probability measure. If $m_j = 0$, it is the Gaussian law $N(x, \Delta_j)$. The measure $\mathcal{Q}_j(\cdot | x)$ is carried by I_j , where $I_j = \mathbb{R}$ for $j < n$ and $I_n = [k, \infty)$, and $\mathcal{Q}_n(\cdot | x)$ is $N(x, \Delta_n)$ conditioned on $[k, \infty)$.

The next lemma identifies $\mathcal{Q}$ with the Parisi diffusion of §1.1. The reason is that on an interval where $\gamma = m > 0$, the Parisi diffusion is the Doob transform [50, 110] of Brownian motion by the space-time harmonic function e^{mu_γ} . Part (a) also shows that the law of $X_1^{\gamma, k}$ may be computed with any cutoff $r \geq Q_\gamma$, which is used for general order parameters.

Lemma 9.3. Let $\gamma \in \mathcal{U}$, $k \in \mathbb{R}$, and $r \in [Q_\gamma, 1)$, and let $u = u_\gamma(\cdot, \cdot; k)$ be computed with the cutoff r . Let X solve $dX_q = \gamma(q) \partial_x u(q, X_q) dq + dB_q$ on $[0, r]$ with $X_0 = 0$, for a standard Brownian motion B , and let X_1 be distributed, given $(X_q)_{q \leq r}$, as $X_r + \sqrt{1-r} G$ conditioned on $\{X_r + \sqrt{1-r} G \geq k\}$.

- (a) $((X_q)_{q \leq Q_\gamma}, X_1)$ has the law of the Parisi diffusion $X^{\gamma, k}$.
- (b) If γ is a step order parameter represented as above with $t_n = r$, then $(X_{t_0}, \dots, X_{t_n}, X_1)$ has the law $\mathcal{Q}$. In particular, for every representation of γ , X_1 has the same law under $\mathcal{Q}$ as $X_1^{\gamma, k}$.

Proof. The drift $\gamma(q) \partial_x u(q, x)$ is Lipschitz in x , uniformly in q , and has linear growth. This is Lemma 2.6(iii), which extends to all of $\mathcal{U}_r$ by Lemma 2.7. Hence the equation has a unique strong solution from every starting point. In particular solutions are unique in law, and X is a Markov process. We use the following fact.

Doob transform. Let $[a, b] \subseteq [0, r]$ and $m > 0$ be such that $\gamma = m$ on $[a, b]$, and suppose that $w = e^{mu}$ satisfies $w(q, x) = \int p_{b-q}(x, z) w(b, z) dz$ on $[a, b] \times \mathbb{R}$. Then, given $X_a = x$, the variable X_b has density $p_{b-a}(x, z) w(b, z) / w(a, x)$. Indeed, let B be a Brownian motion started from x at time a . Since $u \leq 0$ (Lemma 2.6(i)), we have $0 < w \leq 1$. By the representation of w and the Markov property of B , the process $\Theta_q = w(q, B_q) / w(a, x)$ is a bounded positive martingale on $[a, b]$ with $\Theta_a = 1$. The function w is smooth on $[a, b] \times \mathbb{R}$ and solves $\partial_q w + \frac{1}{2} \partial_{xx} w = 0$ there, and Itô's formula gives $d\Theta_q = \Theta_q \partial_x \log w(q, B_q) dB_q$ on $[a, b]$. By Girsanov's theorem [79, Theorem 3.5.1], under the probability measure $\Theta_{b'} \cdot \mathbb{P}$, for every $b' < b$, the process $B_q - \int_a^q \partial_x \log w(q', B_{q'}) dq'$ is a Brownian motion on $[a, b']$. Since $\partial_x \log w = m \partial_x u = \gamma \partial_x u$ on $[a, b]$, under $\Theta_b \cdot \mathbb{P}$ the process B solves the equation for X on $[a, b']$ started from x , for every $b' < b$, and by continuity of paths the laws of $(B_q)_{q \leq b}$ under $\Theta_b \cdot \mathbb{P}$ and of $(X_q)_{q \leq b}$ given $X_a = x$ coincide. Hence X_b given $X_a = x$ has the law of B_b under $\Theta_b \cdot \mathbb{P}$, whose density is $p_{b-a}(x, z) w(b, z) / w(a, x)$.

Proof of (a). On $[0, Q_\gamma]$ the function u does not depend on r , so the law of $(X_q)_{q \leq Q_\gamma}$ is that of the Parisi diffusion. If $Q_\gamma = r$ there is nothing more to prove. Otherwise, on $[Q_\gamma, r]$ we have $\gamma = 1$ and, by (2.11), $e^{u(q, x)} = h_k(q, x) := \mathbb{P}\{x + \sqrt{1-q} G \geq k\}$. This function satisfies $h_k(q, x) = \int p_{r-q}(x, z) h_k(r, z) dz$, because a sum of independent centered Gaussian variables is Gaussian. The Doob transform with $m = 1$ shows that, given $X_{Q_\gamma} = x$, the variable X_r has density $p_{r-Q_\gamma}(x, y) h_k(r, y) / h_k(Q_\gamma, x)$. Given $X_r = y$, the variable X_1 has density $p_{1-r}(y, z) \mathbf{1}_{\{z \geq k\}} / h_k(r, y)$. Multiplying, integrating over y , and using $\int p_{r-Q_\gamma}(x, y) p_{1-r}(y, z) dy = p_{1-Q_\gamma}(x, z)$, we find that X_1 given $X_{Q_\gamma} = x$ has density $p_{1-Q_\gamma}(x, z) \mathbf{1}_{\{z \geq k\}} / h_k(Q_\gamma, x)$, that is, the law of $x + \sqrt{1-Q_\gamma} G$ conditioned on $\{x + \sqrt{1-Q_\gamma} G \geq k\}$. By the Markov property this conditional law does not depend on $(X_q)_{q < Q_\gamma}$. This is the definition in §1.1.

Proof of (b). Let $j < n$. If $m_j > 0$, then $w = e^{m_j u}$ satisfies the hypothesis of the Doob transform on $[t_j, t_{j+1}]$ by the recursion (2.2), so the transition from X_{t_j} to $X_{t_{j+1}}$ is $\mathcal{Q}_j$. If

$m_j = 0$, the drift vanishes on $[t_j, t_{j+1})$ and the transition is $N(x, \Delta_j) = \mathcal{Q}_j(\cdot | x)$. The last transition is $\mathcal{Q}_n$ by construction. By the Markov property, $(X_{t_0}, \dots, X_{t_n}, X_1)$ has the law $\mathcal{Q}$. The last assertion follows from (a). $\square$

Accordingly, for a step order parameter γ we write $\mathbb{E}_\gamma$ for expectation under the Parisi chain $\mathcal{Q}$. By Lemma 9.3(b), expectations of functions of X_1 under $\mathcal{Q}$ agree with those under the Parisi diffusion $X^{\gamma,k}$. A general order parameter is handled by approximation with step order parameters that share its cutoff: every $\gamma \in \mathcal{U}_r$ is the $L^1([0, 1])$ -limit of step order parameters in $\mathcal{U}_r$ (§2.1).

Lemma 9.4. *Fix $r \in [0, 1)$ and $k \in \mathbb{R}$. If $\gamma_i, \gamma \in \mathcal{U}_r$ and $\gamma_i \rightarrow \gamma$ in $L^1([0, 1])$, then*

$$\mathcal{P}_k(\gamma_i) \rightarrow \mathcal{P}_k(\gamma), \quad \lambda_{\gamma_i}(0) \rightarrow \lambda_\gamma(0), \quad \mathbb{E}\varphi(X_1^{\gamma_i,k}) \rightarrow \mathbb{E}\varphi(X_1^{\gamma,k})$$

for every bounded continuous $\varphi : \mathbb{R} \rightarrow \mathbb{R}$.

Proof. The first limit is the continuity of $\mathcal{P}_k$ on $\mathcal{U}_r$ in Lemma 2.7, and $|\lambda_{\gamma_i}(0) - \lambda_\gamma(0)| \leq \|\gamma_i - \gamma\|_{L^1}$. For the third, let X^{γ_i} and X^γ be the diffusions of Lemma 9.3 with cutoff r , driven by the same Brownian motion. By Lemma 9.3(a),

$$\mathbb{E}\varphi(X_1^{\gamma,k}) = \mathbb{E}\pi_\varphi(X_r^\gamma), \quad \mathbb{E}\varphi(X_1^{\gamma_i,k}) = \mathbb{E}\pi_\varphi(X_r^{\gamma_i}), \quad \pi_\varphi(x) = \frac{\int_k^\infty \varphi(z) p_{1-r}(x, z) dz}{\Phi((x - k)/\sqrt{1 - r})}.$$

The numerator of π_φ is continuous in x by dominated convergence, and the denominator is continuous and positive, so π_φ is bounded and continuous. By Lemma 2.7, $\mathbb{E}|X_r^{\gamma_i} - X_r^\gamma|^2 \rightarrow 0$. Hence $X_r^{\gamma_i} \rightarrow X_r^\gamma$ in probability, and $\mathbb{E}\pi_\varphi(X_r^{\gamma_i}) \rightarrow \mathbb{E}\pi_\varphi(X_r^\gamma)$ by bounded convergence. $\square$

9.3 A quadratic source estimate

For configurations of nearly maximal margin the source strength will be large compared with the errors we must control, so we need the dependence of $\mathcal{P}_{k,\beta\psi}(\gamma)$ on β beyond first order, with a constant proportional to $\lambda_\gamma(0)$. Write $w_j = u_j^\beta - u_j^0$. By (9.8) and (9.9), $w_j(x)$ equals $m_j^{-1} \log \int e^{m_j w_{j+1}} d\mathcal{Q}_j(\cdot | x)$ when $m_j > 0$: an iterated logarithmic moment along the Parisi chain. If each transition were the Gaussian law $N(x, \Delta_j)$, the logarithmic moment of $m_j\beta$ times an L -Lipschitz function would exceed $m_j\beta$ times its mean by at most $m_j^2\beta^2 L^2 \Delta_j/2$. After division by m_j , summing $m_j\beta^2 L^2 \Delta_j/2$ over j would give the bound $\frac{1}{2}\beta^2 L^2 \lambda_\gamma(0)$ by (9.7). The transitions of the Parisi chain are Gaussian laws tilted by the log-concave weights $e^{m_j u_{j+1}^0}$. The next lemma shows that they share two properties of Gaussian laws, and the proof of Lemma 9.6 shows that these properties suffice. The variance bound in (a), from which (b) follows, is a case of the Brascamp–Lieb inequality

[25], and the sub-Gaussian bound in (a) also follows from Caffarelli's contraction theorem [27] and Gaussian concentration. In one dimension a direct proof is short.

Lemma 9.5. *Let $\gamma \in \mathcal{U}$ be a step order parameter, $k \in \mathbb{R}$, $0 \leq j \leq n$, and $x \in \mathbb{R}$, and let W have law $\mathcal{Q}_j(\cdot \mid x)$.*

(a) $\mathbb{E}e^{t(W-\mathbb{E}W)} \leq e^{t^2\Delta_j/2}$ for every $t \in \mathbb{R}$. Consequently $\text{Var } W \leq \Delta_j$ and $\mathbb{E}(W - \mathbb{E}W)^4 \leq 16\Delta_j^2$.

(b) If $f : I_j \rightarrow \mathbb{R}$ is bounded and L -Lipschitz, then $x \mapsto \int f d\mathcal{Q}_j(\cdot \mid x)$ is L -Lipschitz on $\mathbb{R}$.

Proof. Write $m = m_j$ and $\Delta = \Delta_j$. If $m = 0$, then $j < n$ (since $m_n = 1$), so $I_j = \mathbb{R}$ and W has the law $N(x, \Delta)$. Then (a) holds with equality in the first claim and $\mathbb{E}(W - \mathbb{E}W)^4 = 3\Delta^2$, and (b) holds because $\int f d\mathcal{Q}_j(\cdot \mid x) = \mathbb{E}f(x + \sqrt{\Delta}G)$. Let $m > 0$.

Proof of (a). Completing the square gives $e^{tz}p_\Delta(x, z) = e^{tx+t^2\Delta/2}p_\Delta(x + t\Delta, z)$. Since $e^{mu_j^0(y)} = \int p_\Delta(y, z)e^{mu_{j+1}^0(z)} dz$ for every $y \in \mathbb{R}$, by (9.8), the definition (9.9) gives

$$\mathbb{E}e^{tW} = \exp\left\{tx + \frac{1}{2}t^2\Delta + mu_j^0(x + t\Delta) - mu_j^0(x)\right\}, \quad t \in \mathbb{R}. \quad (9.10)$$

In particular W has exponential moments of all orders, and differentiating (9.10) at $t = 0$ gives $\mathbb{E}W = x + m\Delta \partial_x u_j^0(x)$. Since u_j^0 is concave, $u_j^0(x + t\Delta) - u_j^0(x) \leq t\Delta \partial_x u_j^0(x)$, and therefore $\mathbb{E}e^{t(W-\mathbb{E}W)} \leq e^{t^2\Delta/2}$. Put $Y = W - \mathbb{E}W$. Applying this bound at t and at $-t$ gives $\mathbb{E}[e^{tY} + e^{-tY} - 2] \leq 2(e^{t^2\Delta/2} - 1)$. Since $e^y + e^{-y} - 2 = 2\sum_{i \geq 1} y^{2i}/(2i)!$ is at least y^2 and at least $y^4/12$, we obtain $t^2\mathbb{E}Y^2 \leq 2(e^{t^2\Delta/2} - 1)$, which gives $\text{Var } W \leq \Delta$ as $t \rightarrow 0$, and, with $t^2 = 1/\Delta$, $\mathbb{E}Y^4 \leq 24(e^{1/2} - 1)\Delta^2 \leq 16\Delta^2$.

Proof of (b). By (9.9), $\mathcal{Q}_j(dz \mid x)$ is proportional to $e^{xz/\Delta}p(z) dz$ on I_j , with $p(z) = e^{-z^2/(2\Delta)+mu_{j+1}^0(z)} \leq e^{-z^2/(2\Delta)}$. The Gaussian bound on p justifies differentiation under the integral, which gives

$$\frac{d}{dx} \int f d\mathcal{Q}_j(\cdot \mid x) = \frac{1}{\Delta} \text{Cov}(f(W), W).$$

With W' an independent copy of W , $\text{Cov}(f(W), W) = \frac{1}{2}\mathbb{E}[(f(W) - f(W'))(W - W')]$, which is at most $\frac{L}{2}\mathbb{E}(W - W')^2 = L \text{Var } W \leq L\Delta$ in absolute value, by (a). $\square$

Lemma 9.6. *Let $\gamma \in \mathcal{U}$ be a step order parameter, $k \in \mathbb{R}$, and let ψ be an L -Lipschitz test function. Put $g_\gamma = \alpha\mathbb{E}\psi(X_1^{\gamma,k} - k)$. For every $\beta \in \mathbb{R}$,*

$$\mathcal{P}_{k,\beta\psi}(\gamma) \leq \mathcal{P}_k(\gamma) + \beta g_\gamma + \frac{\alpha L^2 \beta^2}{2} \lambda_\gamma(0). \quad (9.11)$$

Proof. Fix $\beta \in \mathbb{R}$. Since $\text{CS}(\gamma)$ does not depend on β , it suffices to show that

$$D_\beta = u_0^\beta(0) - u_0^0(0) - \beta\mathbb{E}_\gamma\psi(X_1 - k)$$

satisfies $D_\beta \leq \frac{1}{2}L^2\beta^2\lambda_\gamma(0)$. Neither D_β nor $\lambda_\gamma(0)$ depends on the representation of γ : the values $u_0^\beta(0)$ and $u_0^0(0)$ do not, as shown after (9.8), and the law of X_1 under $\mathcal{Q}$ does not, by Lemma 9.3(b). Fix $\theta \in (0, 1]$ and, by inserting breakpoints, choose a representation with $\Delta_j \leq \theta$ for all j . All objects below refer to this representation. We divide the rest of the proof into three steps.

STEP 1. Conditional expectations. Put $\psi_{n+1}(z) = \psi(z - k)$ on I_n and $\psi_j(x) = \int \psi_{j+1} d\mathcal{Q}_j(\cdot | x)$ for $j = n, \dots, 0$. Then $\psi_j(x)$ is the conditional expectation of $\psi(X_1 - k)$ given $X_{t_j} = x$ under $\mathcal{Q}$, and $\psi_0(0) = \mathbb{E}_\gamma \psi(X_1 - k)$. Each ψ_j is bounded by $\|\psi\|_\infty$, and by Lemma 9.5(b) and backward induction each ψ_j is L -Lipschitz.

STEP 2. The recursion for the difference. Put $w_j = u_j^\beta - u_j^0$ on $\mathbb{R}$ for $j \leq n$, and $w_{n+1} = \beta\psi_{n+1}$ on I_n . Then $|w_j| \leq |\beta|\|\psi\|_\infty$ for all j . Let $j \leq n$. On I_j we have $u_{j+1}^\beta = u_{j+1}^0 + w_{j+1}$, and for $j = n$ both $e^{u_{n+1}^\beta}$ and $e^{u_{n+1}^0}$ vanish outside I_n . Dividing (9.8) for u_j^β by the same formula for u_j^0 and using (9.9) gives

$$\begin{aligned} w_j(x) &= \frac{1}{m_j} \log \int e^{m_j w_{j+1}} d\mathcal{Q}_j(\cdot | x) \quad (m_j > 0), \\ w_j(x) &= \int w_{j+1} d\mathcal{Q}_j(\cdot | x) \quad (m_j = 0). \end{aligned} \tag{9.12}$$

STEP 3. Induction. Put

$$\varepsilon_\theta = \frac{\beta^2 L^2}{2} + \frac{8}{3} |\beta|^3 L^3 e^{2|\beta|\|\psi\|_\infty} \theta^{1/2},$$

and $\lambda_\gamma(t_{n+1}) = \lambda_\gamma(1) = 0$. We show by backward induction that

$$w_j \leq \beta\psi_j + \varepsilon_\theta \lambda_\gamma(t_j) \quad (j = n+1, n, \dots, 0). \tag{9.13}$$

For $j = n+1$ this holds with equality. Assume it for $j+1$, and recall from (9.7) that $\lambda_\gamma(t_j) = \lambda_\gamma(t_{j+1}) + m_j \Delta_j$. If $m_j = 0$, integrating against $\mathcal{Q}_j(\cdot | x)$ gives $w_j \leq \beta\psi_j + \varepsilon_\theta \lambda_\gamma(t_{j+1}) = \beta\psi_j + \varepsilon_\theta \lambda_\gamma(t_j)$. Let $m_j > 0$, let W have law $\mathcal{Q}_j(\cdot | x)$, and put $Y = \psi_{j+1}(W) - \psi_j(x)$, so that $\mathbb{E}Y = 0$ and $|Y| \leq 2\|\psi\|_\infty$. By (9.12) and the induction hypothesis,

$$w_j(x) \leq \varepsilon_\theta \lambda_\gamma(t_{j+1}) + \beta\psi_j(x) + \frac{1}{m_j} \log \mathbb{E} e^{m_j \beta Y}.$$

We bound the last term by a second-order expansion. With $a = m_j \beta$, the inequalities $\log(1+y) \leq y$ and $e^y \leq 1 + y + \frac{1}{2}y^2 + \frac{1}{6}|y|^3 e^{|y|}$ give

$$\log \mathbb{E} e^{aY} \leq \frac{a^2}{2} \mathbb{E} Y^2 + \frac{|a|^3}{6} e^{2|a|\|\psi\|_\infty} \mathbb{E} |Y|^3.$$

Let W' be an independent copy of W . The second moment is $\mathbb{E} Y^2 = \frac{1}{2} \mathbb{E} (\psi_{j+1}(W) - \psi_{j+1}(W'))^2 \leq L^2 \text{Var } W \leq L^2 \Delta_j$, by Lemma 9.5(a). For the third moment, Y is the

conditional expectation of $\psi_{j+1}(W) - \psi_{j+1}(W')$ given W . By Jensen's inequality, the inequality $(y - y')^4 \leq 8(y^4 + y'^4)$, and Lemma 9.5(a),

$$\mathbb{E}Y^4 \leq L^4 \mathbb{E}(W - W')^4 \leq 16L^4 \mathbb{E}(W - \mathbb{E}W)^4 \leq 256L^4 \Delta_j^2,$$

and hence $\mathbb{E}|Y|^3 \leq (\mathbb{E}Y^2 \mathbb{E}Y^4)^{1/2} \leq 16L^3 \Delta_j^{3/2}$. Since $0 < m_j \leq 1$, $|a| \leq |\beta|$, and $\Delta_j^{1/2} \leq \theta^{1/2}$,

$$\frac{1}{m_j} \log \mathbb{E} e^{m_j \beta Y} \leq m_j \Delta_j \left(\frac{\beta^2 L^2}{2} + \frac{8}{3} |\beta|^3 L^3 e^{2|\beta| \|\psi\|_\infty} \theta^{1/2} \right) = \varepsilon_\theta m_j \Delta_j,$$

which completes the induction. At $j = 0$ and $x = 0$, (9.13) gives $D_\beta = w_0(0) - \beta \psi_0(0) \leq \varepsilon_\theta \lambda_\gamma(0)$. Since D_β does not depend on θ , letting $\theta \rightarrow 0$ gives $D_\beta \leq \frac{1}{2} \beta^2 L^2 \lambda_\gamma(0)$. Multiplying by α proves (9.11). $\square$

9.4 Gap averages on sets of positive volume

We now combine the quadratic source estimate with the upper bound of Chapter 3 for the feasible volume weighted by a source. The latter is Corollary 3.11(ii), which we restate in the present notation.

Lemma 9.7. *Let $k, \beta \in \mathbb{R}$, let $\gamma \in \mathcal{U}$ be a step order parameter, and let ψ be a test function. For every $\varepsilon > 0$,*

$$\mathbb{P}\{f_N(k, \beta) \leq \mathcal{P}_{k, \beta \psi}(\gamma) + \varepsilon\} \rightarrow 1.$$

Proof. The function $\beta \psi(\cdot - k)$ is bounded and Lipschitz on $\mathbb{R}$, and $\sum_a \beta \psi(b_a \cdot \sigma - k) = N \beta \Psi_N(\sigma, k)$. Hence $f_N(k, \beta)$ is the normalized logarithm of the integral in Corollary 3.11(ii) with the margin k and the source $\beta \psi(\cdot - k)$. The value $\mathcal{P}_{H_k + \beta \psi(\cdot - k)}(\gamma)$ there is defined by the recursion (2.2) with the terminal datum $H_k + \beta \psi(\cdot - k)$. Evaluated at the breakpoints this is (9.8), so it equals $\mathcal{P}_{k, \beta \psi}(\gamma)$, and it does not depend on the representation of γ . Corollary 3.11(ii) is therefore the claim. $\square$

We now state the deviation inequality that both gap theorems use. It holds for every order parameter; the step approximation of Lemma 9.4 is performed inside the proof.

Proposition 9.8. *Let $k \in \mathbb{R}$, $\gamma \in \mathcal{U}$, $\beta > 0$, and $\varepsilon > 0$, let ψ be an L -Lipschitz test function, and put $g = \alpha \mathbb{E} \psi(X_1^{\gamma, k} - k)$. With probability tending to one, every Borel set $T \subseteq S_N(k)$ with $\mu_N(T) > 0$ satisfies*

$$|\langle \Psi_N(\sigma, k) \rangle_T - g| \leq \frac{\mathcal{P}_k(\gamma) - N^{-1} \log \mu_N(T)}{\beta} + \frac{\alpha L^2 \beta}{2} \lambda_\gamma(0) + \varepsilon. \quad (9.14)$$

Proof. We divide the proof into two steps: we first treat step order parameters and then approximate.

STEP 1. *Step order parameters.* Let γ be a step order parameter. For every Borel $T \subseteq S_N(k)$ with $\mu_N(T) > 0$, restricting the integral defining $f_N(k, \beta)$ to T and applying Jensen's inequality to $\langle e^{N\beta\Psi_N(\sigma, k)} \rangle_T$ gives

$$f_N(k, \beta) \geq \frac{1}{N} \log \mu_N(T) + \beta \langle \Psi_N(\sigma, k) \rangle_T.$$

By Lemma 9.3(b), $g = \alpha \mathbb{E}_\gamma \psi(X_1 - k)$. By Lemma 9.7 with $\beta\varepsilon$ in place of ε , and by Lemma 9.6, with probability tending to one,

$$f_N(k, \beta) \leq \mathcal{P}_{k, \beta\psi}(\gamma) + \beta\varepsilon \leq \mathcal{P}_k(\gamma) + \beta g + \frac{\alpha L^2 \beta^2}{2} \lambda_\gamma(0) + \beta\varepsilon.$$

On this event, which does not depend on T , the two inequalities give

$$\langle \Psi_N(\sigma, k) \rangle_T - g \leq \frac{\mathcal{P}_k(\gamma) - N^{-1} \log \mu_N(T)}{\beta} + \frac{\alpha L^2 \beta}{2} \lambda_\gamma(0) + \varepsilon$$

for every such T . The test function $-\psi$ is also L -Lipschitz, and replacing ψ by $-\psi$ replaces Ψ_N by $-\Psi_N$ and g by $-g$. The same argument therefore bounds $g - \langle \Psi_N(\sigma, k) \rangle_T$ on a second event whose probability tends to one. The intersection of the two events has probability tending to one.

STEP 2. *General order parameters.* Choose $r \in [Q_\gamma, 1)$ and step order parameters $\gamma_i \in \mathcal{U}_r$ with $\gamma_i \rightarrow \gamma$ in $L^1([0, 1])$. By Lemma 9.4, applied with the bounded continuous function $\varphi = \psi(\cdot - k)$, we have $\mathcal{P}_k(\gamma_i) \rightarrow \mathcal{P}_k(\gamma)$, $\lambda_{\gamma_i}(0) \rightarrow \lambda_\gamma(0)$, and $g_i = \alpha \mathbb{E} \psi(X_1^{\gamma_i, k} - k) \rightarrow g$. Fix i such that

$$\frac{|\mathcal{P}_k(\gamma_i) - \mathcal{P}_k(\gamma)|}{\beta} + \frac{\alpha L^2 \beta}{2} |\lambda_{\gamma_i}(0) - \lambda_\gamma(0)| + |g_i - g| \leq \frac{\varepsilon}{2}.$$

Step 1 applied to γ_i with $\varepsilon/2$ in place of ε proves the proposition for γ . $\square$

El Alaoui and Sellke described the empirical law of the fields along the iterates of their algorithm through the Parisi diffusion [54, Section 3]. Empirical laws of the projections $b_a \cdot \sigma$ along data-dependent directions σ were also studied by Montanari and Zhou [89]; see also [73]. Montanari and Zhou characterized the laws attainable by a class of efficient algorithms through a stochastic control problem. They passed between these laws and the largest averages of test functions by duality, and obtained a variational principle that extends the Parisi formula. In Proposition 9.8 the test function enters the terminal datum of the Parisi functional as a source, and the bound applies to every set of positive volume in the feasible set.

9.5 Gaps of typical feasible configurations

We apply Proposition 9.8 at a fixed margin $\kappa < \kappa_c$, with a minimizer γ of $\mathcal{P}_\kappa$, to the sets of configurations whose gap average deviates from $\alpha \mathbb{E}\psi(X_1^{\gamma, \kappa} - \kappa)$. The hard-volume lower bound of Theorem 1.2(i) controls the volume term.

Proof of Theorem 1.9. Let $\alpha > 2$ and $\kappa < \kappa_c(\alpha)$. We divide the proof into three steps.

STEP 1. Preliminaries. By Theorem 6.8, $\mathcal{P}_*(\kappa) > -\infty$. Hence minimizers of $\mathcal{P}_\kappa$ exist (Theorem 6.3). By Theorem 1.2(i), $V_N(\kappa) > 0$, so that $\langle \cdot \rangle_{N, \kappa}$ is well defined, with probability tending to one. Let γ be a minimizer of $\mathcal{P}_\kappa$, so that $\mathcal{P}_\kappa(\gamma) = \mathcal{P}_*(\kappa)$, and put $\lambda = \lambda_\gamma(0)$. Then $\lambda \geq 1 - Q_\gamma > 0$. Let ψ be an L -Lipschitz test function with $L > 0$, put $g = \alpha \mathbb{E}\psi(X_1^{\gamma, \kappa} - \kappa)$, and fix $\varepsilon > 0$.

STEP 2. Deviation sets. Put $\beta = \varepsilon / (2\alpha L^2 \lambda)$, so that $\frac{1}{2}\alpha L^2 \beta \lambda = \varepsilon/4$. Let $\mathcal{E}_N$ be the event on which the conclusion of Proposition 9.8 holds for the margin κ , the minimizer γ , this β , and $\varepsilon/8$ in place of ε , and on which

$$\frac{1}{N} \log V_N(\kappa) \geq \mathcal{P}_*(\kappa) - \frac{\beta \varepsilon}{8}.$$

By Proposition 9.8 and Theorem 1.2(i), $\mathbb{P}(\mathcal{E}_N) \rightarrow 1$. The function $\sigma \mapsto \Psi_N(\sigma, \kappa)$ is continuous, so the sets

$$T_\pm = \{\sigma \in S_N(\kappa) : \pm(\Psi_N(\sigma, \kappa) - g) > \varepsilon\}$$

are Borel. Suppose that $\mathcal{E}_N$ holds and $\mu_N(T_+) > 0$. Every $\sigma \in T_+$ has $\Psi_N(\sigma, \kappa) - g > \varepsilon$, and $\mu_N(T_+) > 0$, so $\langle \Psi_N(\sigma, \kappa) \rangle_{T_+} - g > \varepsilon$. By (9.14) with $T = T_+$,

$$\varepsilon < \frac{\mathcal{P}_*(\kappa) - N^{-1} \log \mu_N(T_+)}{\beta} + \frac{\varepsilon}{4} + \frac{\varepsilon}{8}, \quad \text{that is,} \quad \frac{1}{N} \log \mu_N(T_+) < \mathcal{P}_*(\kappa) - \frac{5\beta \varepsilon}{8}.$$

Together with the lower bound on $V_N(\kappa)$ in the definition of $\mathcal{E}_N$, this gives $\mu_N(T_+) < e^{-N\beta\varepsilon/2} V_N(\kappa)$. The same bound holds for T_- , and it also holds if $\mu_N(T_\pm) = 0$. Since $\beta\varepsilon/2 = \varepsilon^2 / (4\alpha L^2 \lambda)$, on $\mathcal{E}_N$

$$\mu_N(\{\sigma \in S_N(\kappa) : |\Psi_N(\sigma, \kappa) - g| > \varepsilon\}) \leq 2 \exp\left\{-\frac{\varepsilon^2 N}{4\alpha L^2 \lambda}\right\} V_N(\kappa) \leq \exp\left\{-\frac{\varepsilon^2 N}{8\alpha L^2 \lambda}\right\} V_N(\kappa)$$

as soon as $\varepsilon^2 N \geq 8\alpha L^2 \lambda \log 2$. This proves (1.14).

STEP 3. Gibbs averages and the law of X_1 . On $\mathcal{E}_N$, since $|\Psi_N(\sigma, \kappa)| \leq \bar{\alpha} \|\psi\|_\infty$ and $|g| \leq \alpha \|\psi\|_\infty$, Step 2 gives, for large N ,

$$|\langle \Psi_N(\sigma, \kappa) \rangle_{N, \kappa} - g| \leq \varepsilon + (\bar{\alpha} + \alpha) \|\psi\|_\infty \exp\left\{-\frac{\varepsilon^2 N}{8\alpha L^2 \lambda}\right\} \leq 2\varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, $\langle \Psi_N(\sigma, \kappa) \rangle_{N, \kappa} \rightarrow g$ in probability. The averages $\langle \Psi_N(\sigma, \kappa) \rangle_{N, \kappa}$ do not depend on γ , and a limit in probability is unique. Hence $\mathbb{E}\psi(X_1^{\gamma, \kappa} - \kappa)$ is the same for all minimizers γ , for every bounded Lipschitz ψ . The variable $X_1^{\gamma, \kappa} - \kappa$ takes values in $[0, \infty)$, and bounded Lipschitz functions determine a finite Borel measure on $[0, \infty)$: for a closed set F the functions $(1 - j \operatorname{dist}(\cdot, F))_+$, $j \geq 1$, are bounded and Lipschitz and decrease to $\mathbf{1}_F$, and a finite Borel measure is determined by its values on closed sets. Hence all minimizers induce the same law of $X_1^{\gamma, \kappa}$. $\square$

In the replica computation of Franz, Parisi, Sevelev, Urbani, and Zamponi, the gap distribution is obtained as the derivative of the free energy with respect to a potential acting on the gaps, and it is expressed through the solution of the Fokker–Planck equation of the Parisi diffusion [59, (37)]. The proof of Proposition 9.8 perturbs the hard wall by a source and compares free energies in the same way. Theorem 1.9 is a rigorous counterpart of this description for the uniform measure on $S_N(\kappa)$ at a fixed margin $\kappa < \kappa_c$.

9.6 Configurations near the optimal margin

We turn to configurations of nearly maximal margin. To apply Proposition 9.8 to one feasible configuration σ_0 , we need a set of positive volume near σ_0 on which the fields $b_a \cdot \sigma$, and hence the gap averages, change little. Lemma 3.13 provides such a set: a subset of a small cap around σ_0 that is feasible at a slightly smaller margin and has volume at least $e^{-N(\log(1/\eta) + C)}$. This volume cost, of order $N \log(1/\eta)$ on the logarithmic scale, is what must be compared with the collapse of $\lambda_\gamma(0)$ near κ_c . We record the consequence for gap averages.

Lemma 9.9. *Let I be a compact interval. There are constants N_0 and $C_1 \geq 0$, depending only on $\bar{\alpha}$ and I , with the following property. Let $N \geq N_0$, suppose that Ω_N holds, and let $k \in I$, $\sigma_0 \in S_N(k)$, and $0 < \eta \leq 1$. There is a Borel set $T \subseteq S_N(k - \eta)$ with $-N^{-1} \log \mu_N(T) \leq \log(1/\eta) + C_1$ such that, for every L -Lipschitz test function ψ and every $\sigma \in T$,*

$$|\Psi_N(\sigma, k - \eta) - \Psi_N(\sigma_0, k)| \leq C_\psi \eta, \quad C_\psi = L(\sqrt{\bar{\alpha}} C_{\text{op}} + 2\bar{\alpha}). \quad (9.15)$$

Proof. Let $K_{\text{cap}} \geq 1$, C_{cap} , and N_0 be the constants of Lemma 3.13 for $\alpha' = \bar{\alpha}$, let C_I be the constant there for the interval I , and put $C_1 = \max\{C_{\text{cap}} + C_I, 0\}$. On Ω_N the rows satisfy $\|b_a\|^2 \leq 2$, and $M \leq \bar{\alpha}N$, so Lemma 3.13 applies to the margin k , the point σ_0 , and this η . Since $k \in I$, it gives a Borel set $T \subseteq S_N(k - \eta)$ with $\mu_N(T) \geq \exp\{-N(\log(2/s_0) + C_{\text{cap}})\} \geq \exp\{-N(\log(1/\eta) + C_1)\}$ and

$$|b_a \cdot \sigma - b_a \cdot \sigma_0| \leq s_0^2 |b_a \cdot \sigma_0| + s_0 K_{\text{cap}}, \quad \sigma \in T, \quad a \leq M,$$

where $s_0 \leq \eta/(2K_{\text{cap}})$ by the definition of s_0 there. As $K_{\text{cap}} \geq 1$ and $\eta \leq 1$, we have $s_0 K_{\text{cap}} \leq \eta/2$ and $s_0^2 \leq \eta^2/4 \leq \eta/4$. Since ψ is L -Lipschitz, for $\sigma \in T$ and $a \leq M$,

$$|\psi(b_a \cdot \sigma - k + \eta) - \psi(b_a \cdot \sigma_0 - k)| \leq L(|b_a \cdot \sigma - b_a \cdot \sigma_0| + \eta) \leq L\left(\frac{\eta}{4}|b_a \cdot \sigma_0| + \frac{3\eta}{2}\right).$$

Averaging over a , and using (9.1) for σ_0 and $\alpha_N \leq \bar{\alpha}$, gives $|\Psi_N(\sigma, k - \eta) - \Psi_N(\sigma_0, k)| \leq L\eta(\frac{1}{4}\sqrt{\bar{\alpha}}C_{\text{op}} + \frac{3}{2}\bar{\alpha}) \leq C_\psi\eta$. $\square$

We now combine the cap with Proposition 9.8. The following proposition separates the probabilistic inputs, which are the convergence of κ_N and the bounds already proved, from the deterministic input, which concerns only the variational problem: along some sequence of margins $k_n \uparrow \kappa_c$ there are order parameters whose mass $\lambda_\gamma(0)$ is small compared with the volume cost $\log(1/(\kappa_c - k_n))$ of the cap.

Proposition 9.10. *Assume that $\kappa_N \rightarrow \kappa_c$ in probability, and let ψ be a test function. Suppose that there are margins $k_n < \kappa_c$ with $k_n \rightarrow \kappa_c$ and order parameters $\gamma_n \in \mathcal{U}$ such that*

$$\lambda_{\gamma_n}(0) \left(1 + [\mathcal{P}_{k_n}(\gamma_n)]_+ + \log \frac{1}{\kappa_c - k_n}\right) \rightarrow 0. \quad (9.16)$$

Then $g_n = \alpha \mathbb{E} \psi(X_1^{\gamma_n, k_n} - k_n)$ converges to a limit $g(\psi)$, and for every $\varepsilon > 0$ there is $\rho > 0$ such that

$$\mathbb{P} \left\{ \sup_{\sigma \in S_N: m_N(\sigma) \geq \kappa_c - \rho} |\Psi_N(\sigma, m_N(\sigma)) - g(\psi)| \leq \varepsilon \right\} \rightarrow 1. \quad (9.17)$$

In particular, since $m_N(\sigma_*) = \kappa_N$ for every maximizer σ_* of m_N and $\mathbb{P}\{\kappa_N \geq \kappa_c - \rho\} \rightarrow 1$, the conclusion applies simultaneously to all maximizers.

Proof. Let ψ be L -Lipschitz with $L > 0$ (a Lipschitz function is L -Lipschitz for every large L), and write $\lambda_n = \lambda_{\gamma_n}(0) > 0$. Discarding finitely many n , we assume $\kappa_c - k_n \leq 1$, and we put $\eta_n = (\kappa_c - k_n)/2 \leq \frac{1}{2}$, so that $k_n + \eta_n = \kappa_c - \eta_n$. Let N_0 and C_1 be the constants of Lemma 9.9 for the interval $I = [\kappa_c - 1, \kappa_c]$, let C_ψ be as in (9.15), and put

$$\ell_n = 1 + \left[\mathcal{P}_{k_n}(\gamma_n) + \log \frac{1}{\eta_n} + C_1 \right]_+, \quad \omega_n = L(2\alpha\lambda_n\ell_n)^{1/2} + (C_\psi + 2\bar{\alpha}L)\eta_n.$$

We divide the proof into four steps.

STEP 1. *Trapping at the margin $\kappa_c - \eta_n$.* Fix n and $\varepsilon > 0$, and put $\beta_n = (2\ell_n/(\alpha L^2 \lambda_n))^{1/2}$, so that

$$\frac{\ell_n}{\beta_n} + \frac{\alpha L^2 \beta_n}{2} \lambda_n = L(2\alpha\lambda_n\ell_n)^{1/2}.$$

Consider the event on which Ω_N holds and the conclusion of Proposition 9.8 holds for the margin k_n , the order parameter γ_n , the source strength β_n , and this ε . Its probability tends

to one. On this event, for $N \geq N_0$, let $\sigma_0 \in S_N(\kappa_c - \eta_n)$, and let T be the set of Lemma 9.9 for the margin $k = \kappa_c - \eta_n \in I$ and $\eta = \eta_n$. Then $T \subseteq S_N(k - \eta_n) = S_N(k_n)$ and $-N^{-1} \log \mu_N(T) \leq \log(1/\eta_n) + C_1$, so that $\mathcal{P}_{k_n}(\gamma_n) - N^{-1} \log \mu_N(T) \leq \ell_n$. Averaging (9.15) over T gives $|\langle \Psi_N(\sigma, k_n) \rangle_T - \Psi_N(\sigma_0, \kappa_c - \eta_n)| \leq C_\psi \eta_n$. Therefore (9.14) gives

$$\sup_{\sigma_0 \in S_N(\kappa_c - \eta_n)} |\Psi_N(\sigma_0, \kappa_c - \eta_n) - g_n| \leq L(2\alpha\lambda_n\ell_n)^{1/2} + C_\psi\eta_n + \varepsilon.$$

STEP 2. Changing the reference margin. Since ψ is L -Lipschitz, $|\Psi_N(\sigma, r) - \Psi_N(\sigma, r')| \leq \alpha_N L|r - r'| \leq \bar{\alpha}L|r - r'|$. With probability tending to one, $\kappa_N \leq \kappa_c + \eta_n$. Let σ satisfy $m_N(\sigma) \geq \kappa_c - \eta_n$. Then $\sigma \in S_N(\kappa_c - \eta_n)$ and $0 \leq m_N(\sigma) - (\kappa_c - \eta_n) \leq \kappa_N - \kappa_c + \eta_n \leq 2\eta_n$. Together with Step 1 this shows that, with probability tending to one,

$$\sup_{\sigma \in S_N: m_N(\sigma) \geq \kappa_c - \eta_n} |\Psi_N(\sigma, m_N(\sigma)) - g_n| \leq \omega_n + \varepsilon. \quad (9.18)$$

STEP 3. The errors vanish. Since $\log(1/\eta_n) = \log 2 + \log(1/(\kappa_c - k_n))$, $\log(1/(\kappa_c - k_n)) \geq 0$, and $C_1 \geq 0$,

$$\ell_n \leq (2 + \log 2 + C_1) \left(1 + [\mathcal{P}_{k_n}(\gamma_n)]_+ + \log \frac{1}{\kappa_c - k_n} \right).$$

Hence $\lambda_n \ell_n \rightarrow 0$ by (9.16). Since also $\eta_n \rightarrow 0$, we have $\omega_n \rightarrow 0$.

STEP 4. The centers converge. Fix n and n' . With probability tending to one, $\kappa_N \geq \kappa_c - \min\{\eta_n, \eta_{n'}\}$, and then every maximizer σ_* of m_N lies in both sets in (9.18), for n and for n' . For large N the three events intersect, which gives $|g_n - g_{n'}| \leq \omega_n + \omega_{n'} + 2\varepsilon$. Since ε is arbitrary, $|g_n - g_{n'}| \leq \omega_n + \omega_{n'}$, and $\omega_n \rightarrow 0$. Thus (g_n) is a Cauchy sequence. Let $g(\psi)$ be its limit. Then $|g_n - g(\psi)| \leq \omega_n$.

Given $\varepsilon > 0$, choose n with $2\omega_n \leq \varepsilon/2$ and put $\rho = \eta_n$. By (9.18) with $\varepsilon/2$ in place of ε , with probability tending to one, $|\Psi_N(\sigma, m_N(\sigma)) - g(\psi)| \leq \omega_n + \varepsilon/2 + |g_n - g(\psi)| \leq \varepsilon$ for all σ with $m_N(\sigma) \geq \kappa_c - \rho$. This proves (9.17). $\square$

9.7 Gaps of nearly optimal configurations

We first identify the limits of gap averages of near-maximizers, one test function at a time, by applying Proposition 9.10 to minimizers. The collapse of minimizers, Proposition 6.11, supplies its hypothesis. The measure ν_α is then obtained by tightness, which we derive from the operator norm bound (9.1) at maximizers. Finally, Proposition 9.2 gives the contacts and the small positive gaps at maximizers.

Proof of Theorem 1.8. Let $\alpha > 2$. By Theorem 1.2, $\kappa_c < 0$ and $\kappa_N \rightarrow \kappa_c$ in probability. We divide the proof into six steps.

STEP 1. *Identification along one sequence.* Fix a test function ψ and margins $k_n < \kappa_c$ with $k_n \rightarrow \kappa_c$. By Theorem 6.8, $\mathcal{P}_*(k_n) > -\infty$, and by Theorem 6.3 there are minimizers γ_n of $\mathcal{P}_{k_n}$; fix any. By (6.10), $\mathcal{P}_{k_n}(\gamma_n) = \mathcal{P}_*(k_n) \leq \alpha \log \Phi(-k_n) < 0$, so that $[\mathcal{P}_{k_n}(\gamma_n)]_+ = 0$. By Proposition 6.11, applied with $\eta = \kappa_c - k_n$, there are C and η_0 , depending only on α , such that $\lambda_{\gamma_n}(0) \leq C(\kappa_c - k_n)^{1/(2+2\alpha)}$ as soon as $\kappa_c - k_n \leq \eta_0$ (the second bound in (6.24)). Since $x^{1/(2+2\alpha)}(1 + \log(1/x)) \rightarrow 0$ as $x \downarrow 0$, the condition (9.16) holds. Proposition 9.10 shows that $\alpha \mathbb{E}\psi(X_1^{\gamma_n, k_n} - k_n)$ converges to a limit $g(\psi)$, and that for every $\varepsilon > 0$ there is $\rho > 0$ with

$$\mathbb{P}\left\{\sup_{\sigma \in S_N: m_N(\sigma) \geq \kappa_c - \rho} |\Psi_N(\sigma, m_N(\sigma)) - g(\psi)| \leq \varepsilon\right\} \rightarrow 1. \quad (9.19)$$

Given two sequences of margins and minimizers, the sequence obtained by interleaving them is again of this form, so the two limits coincide. Thus $g(\psi)$ depends only on ψ and α .

STEP 2. *Mean gap at maximizers.* Maximizers of m_N exist by compactness, and $m_N(\sigma_*) = \kappa_N$. For $\sigma \in S_N$, by (9.1),

$$\int x v_{N,\sigma}(\mathrm{d}x) = \frac{1}{N} \sum_a (b_a \cdot \sigma - m_N(\sigma)) \leq \sqrt{\alpha_N} C_{\text{op}} + \alpha_N |m_N(\sigma)|$$

on Ω_N . Put $C_2 = \sqrt{2\alpha} C_{\text{op}} + 2\alpha(|\kappa_c| + 1)$. With probability tending to one, Ω_N holds and $|\kappa_N| \leq |\kappa_c| + 1$. Then, for N so large that $\alpha_N \leq 2\alpha$, every maximizer satisfies $\int x v_{N,\sigma_*}(\mathrm{d}x) \leq C_2$, and hence

$$v_{N,\sigma_*}((t, \infty)) \leq \frac{C_2}{t} \quad (t > 0). \quad (9.20)$$

STEP 3. *The measure v_α ; parts (i) and (iii).* Let k_n and γ_n be as in Step 1, and let $\pi_n = \alpha \text{Law}(X_1^{\gamma_n, k_n} - k_n)$, a measure on $[0, \infty)$ of mass α . For $t > 0$ let $\chi_t(x) = \min\{1, (x - t)_+\}$, a bounded Lipschitz function with $\mathbf{1}_{[t+1, \infty)} \leq \chi_t \leq \mathbf{1}_{(t, \infty)}$. We first show that $g(\chi_t) \leq C_2/t$. Given $\varepsilon > 0$, let ρ be as in (9.19) for χ_t . With probability tending to one, $\kappa_N \geq \kappa_c - \rho$, the event in (9.19) holds, and the conclusion of Step 2 holds. For large N these events intersect, and for a maximizer σ_* they give

$$g(\chi_t) \leq \int \chi_t \mathrm{d}v_{N,\sigma_*} + \varepsilon \leq v_{N,\sigma_*}((t, \infty)) + \varepsilon \leq \frac{C_2}{t} + \varepsilon.$$

Hence $g(\chi_t) \leq C_2/t$, and

$$\limsup_{n \rightarrow \infty} \pi_n([t+1, \infty)) \leq \lim_{n \rightarrow \infty} \int \chi_t \mathrm{d}\pi_n = g(\chi_t) \leq \frac{C_2}{t}.$$

For every $\varepsilon > 0$, choose t so that $C_2/t < \varepsilon/2$. The tail masses above $t+1$ are then less than ε for all sufficiently large n . Enlarging the compact interval to control the finitely

many remaining measures proves that (π_n) is tight. Apply Prokhorov's theorem [109] to the probability measures π_n/α on the Polish space $[0, \infty)$ and rescale. Every subsequence has a further subsequence converging weakly to a measure π of mass α , and by Step 1, $\int \psi d\pi = g(\psi)$ for every test function ψ . Bounded Lipschitz functions determine a finite measure on $[0, \infty)$, as shown in the proof of Theorem 1.9, so π depends neither on the subsequence nor on the sequence (k_n, γ_n) . We call it ν_α . Then $\nu_\alpha([0, \infty)) = \alpha$ and $\int \psi d\nu_\alpha = g(\psi)$ for every test function ψ . Along every such sequence, every subsequence of (π_n) has a further subsequence converging weakly to ν_α , and hence $\pi_n \rightarrow \nu_\alpha$ weakly. This proves (iii). Since $\int \psi d\nu_{N,\sigma} = \Psi_N(\sigma, m_N(\sigma))$, (9.19) is (i).

STEP 4. Weak convergence at maximizers. Let ψ be a bounded continuous function on $[0, \infty)$, and let $0 < \varepsilon \leq 1$. Choose $t \geq 1$ with $C_2/t \leq \varepsilon$ and $\nu_\alpha((t, \infty)) \leq \varepsilon$, and let $\chi(x) = \min\{1, (t+1-x)_+\}$. The function $\psi\chi$ is continuous and vanishes on $[t+1, \infty)$, hence uniformly continuous. Its piecewise linear interpolation on a sufficiently fine uniform grid of $[0, t+1]$, extended by zero, is a bounded Lipschitz function ψ' with $\sup|\psi\chi - \psi'| \leq \varepsilon$. Since $|\psi(1-\chi)| \leq \|\psi\|_\infty \mathbf{1}_{(t, \infty)}$, every maximizer satisfies

$$\begin{aligned} \left| \int \psi d\nu_{N,\sigma_*} - \int \psi d\nu_\alpha \right| &\leq \left| \int \psi' d\nu_{N,\sigma_*} - \int \psi' d\nu_\alpha \right| + \varepsilon(\alpha_N + \alpha) \\ &\quad + \|\psi\|_\infty (\nu_{N,\sigma_*}((t, \infty)) + \nu_\alpha((t, \infty))). \end{aligned}$$

By part (i), applied to ψ' , and since $\kappa_N \rightarrow \kappa_c$, the first term on the right is at most ε for all maximizers with probability tending to one. By (9.20) and the choice of t , for N so large that $\alpha_N \leq \alpha + 1$ the remaining terms are at most $(2\alpha + 1)\varepsilon + 2\|\psi\|_\infty \varepsilon$ with probability tending to one. Hence, for every bounded continuous ψ , $\sup_{\sigma_*} |\int \psi d\nu_{N,\sigma_*} - \int \psi d\nu_\alpha| \rightarrow 0$ in probability, the supremum being over all maximizers. This is the weak convergence in (ii).

STEP 5. Contacts and the atom at zero. By Proposition 9.2(i) and (ii), for every $c < (\alpha - 2)^2/(2\alpha)$ and all sufficiently large N , with probability at least $1 - e^{-cN}$ every maximizer has exactly N contacts. This is the first assertion of (ii). On this event $\nu_{N,\sigma_*}(\{0\}) = 1$ for every maximizer. Let $\varepsilon > 0$, and choose $t \in (0, \sqrt{2\pi})$ so small that $\alpha \log 3 / \log(\sqrt{2\pi}/t) < \varepsilon$. By Proposition 9.2(iii), with probability tending to one every maximizer also satisfies $\nu_{N,\sigma_*}((0, t]) < \varepsilon$, and hence $\nu_{N,\sigma_*}([0, t]) < 1 + \varepsilon$.

For $h > 0$ let $\psi_h(x) = (1 - x/h)_+$, a test function with $\mathbf{1}_{\{0\}} \leq \psi_h \leq \mathbf{1}_{[0,h]}$. With probability tending to one, $\int \psi_h d\nu_{N,\sigma_*} \geq \nu_{N,\sigma_*}(\{0\}) = 1$ and $\int \psi_h d\nu_{N,\sigma_*} \leq \nu_{N,\sigma_*}([0, t]) < 1 + \varepsilon$. By part (i) at maximizers, $\int \psi_h d\nu_{N,\sigma_*} \rightarrow \int \psi_h d\nu_\alpha$ in probability for every $h > 0$. Hence $\int \psi_h d\nu_\alpha \geq 1$ for every $h > 0$, and letting $h \downarrow 0$ gives $\nu_\alpha(\{0\}) \geq 1$ by dominated convergence. Also $\nu_\alpha(\{0\}) \leq \int \psi_t d\nu_\alpha \leq 1 + \varepsilon$ for every $\varepsilon > 0$. Therefore $\nu_\alpha(\{0\}) = 1$.

STEP 6. Small positive gaps. Let $t > 0$ be a continuity point of ν_α . For $0 < h < t$ let ψ^+ equal one on $[0, t]$, zero on $[t+h, \infty)$ and be linear in between, and let ψ^- equal one on $[0, t-h]$,

zero on $[t, \infty)$ and be linear in between. Then $\int \psi^- d\nu_{N, \sigma_*} \leq \nu_{N, \sigma_*}([0, t]) \leq \int \psi^+ d\nu_{N, \sigma_*}$. The outer integrals converge in probability to $\int \psi^\pm d\nu_\alpha$ by part (i), and both of these tend to $\nu_\alpha([0, t])$ as $h \downarrow 0$ because $\nu_\alpha(\{t\}) = 0$. So $\nu_{N, \sigma_*}([0, t]) \rightarrow \nu_\alpha([0, t])$ in probability, uniformly over maximizers. With probability tending to one $\nu_{N, \sigma_*}(\{0\}) = 1$, and therefore

$$\frac{1}{N} \#\{a \leq M : 0 < b_a \cdot \sigma_* - \kappa_N \leq t\} = \nu_{N, \sigma_*}([0, t]) - 1 \longrightarrow \nu_\alpha([0, t]) - \nu_\alpha(\{0\}) = \nu_\alpha((0, t]) \quad (9.21)$$

in probability. Together with Steps 4 and 5 this proves (ii).

It remains to prove (1.13). For $0 < t < \sqrt{2\pi}$ put $\delta(t) = \alpha \log 3 / \log(\sqrt{2\pi}/t)$, a continuous function of t . Let $t_1 \in (0, \sqrt{2\pi})$ and $\varepsilon > \delta(t_1)$, and let $t' \leq t_1$ be a continuity point of ν_α . By Proposition 9.2(iii) with t_1 in place of t , with probability tending to one every maximizer satisfies

$$\frac{1}{N} \#\{a \leq M : 0 < b_a \cdot \sigma_* - \kappa_N \leq t'\} \leq \frac{1}{N} \#\{a \leq M : 0 < b_a \cdot \sigma_* - \kappa_N \leq t_1\} < \varepsilon.$$

By (9.21) at t' , this gives $\nu_\alpha((0, t']) \leq \varepsilon$. Since this holds for every $\varepsilon > \delta(t_1)$, letting $\varepsilon \downarrow \delta(t_1)$ gives $\nu_\alpha((0, t']) \leq \delta(t_1)$. The continuity points of ν_α are dense, because ν_α has at most countably many atoms, so $\nu_\alpha((0, t_1)) \leq \delta(t_1)$ by continuity of measures along increasing unions. For $0 < t < \sqrt{2\pi}$, letting $t_1 \downarrow t$ gives $\nu_\alpha((0, t]) = \lim_{t_1 \downarrow t} \nu_\alpha((0, t_1)) \leq \delta(t)$, by continuity of δ . This proves (1.13). $\square$

The bound (1.13) tends to zero only logarithmically as $t \downarrow 0$. It comes from the union bound over sets of N rows in Proposition 9.2(iii). The limiting measure ν_α is identified only through the Parisi minimizers in part (iii), and its behavior near zero is not determined by our arguments. Franz and Parisi [58] and Franz, Parisi, Sevelev, Urbani, and Zamponi [59, (75) and (106)] predicted that, at jamming with $\kappa < 0$, the gap distribution consists of an atom carrying the contacts and a density that diverges like $h^{-0.41269\dots}$ as $h \downarrow 0$. This is the exponent predicted for hard spheres in infinite dimension [30]. Theorem 1.8 gives the atom, and the power law remains open.

PART II

Full replica symmetry breaking in the negative spherical perceptron

We study the Parisi variational problem of the negative spherical perceptron, a mean-field model of jamming, at constraint densities $\alpha = 2 + e$ with $e > 0$ small. We prove that replica symmetry is lost exactly at the de Almeida–Thouless margin, through a continuous transition to full replica symmetry breaking: in a window of width of order e^2 ending at this margin the replica-symmetric order parameter is the unique minimizer, and in a window of the same width beyond it every minimizer has a Parisi measure supported on one interval, with density of order e^{-2} and atoms at both endpoints. In the second window the no-overlap-gap hypothesis of El Alaoui and Sellke holds. The de Almeida–Thouless margin lies $\theta_d e^2(1 + o(1))$ below Gardner’s prediction κ_{RS} for the capacity, for an explicit $\theta_d > 0$, and κ_{RS} exceeds the critical margin of the functional by exact order e^3 . With the identification of the capacity in Part I, Gardner’s formula therefore overestimates the capacity by exact order $(\alpha - 2)^3$. The proof rests on a compactness theorem for minimizers and on a universal flat limit. In the compactness theorem, a balance law implied by the optimality conditions bounds the depth of the support. A jamming identity for the functional and a bound on the jump at the top of the support give the remaining bounds.

Introduction

Let $N \geq 1$, and let μ_N be the uniform probability measure on the sphere $S_N = \{\sigma \in \mathbb{R}^N : \|\sigma\|^2 = N\}$. Fix a constraint density $\alpha > 0$ and a margin $\kappa \in \mathbb{R}$, let $M = M_N$ be integers with $M/N \rightarrow \alpha$, and let $g_1, \dots, g_M$ be independent standard Gaussian vectors in $\mathbb{R}^N$. The spherical perceptron asks for points of the sphere that satisfy the M random linear constraints

$$\frac{g_a \cdot \sigma}{\sqrt{N}} \geq \kappa, \quad a = 1, \dots, M. \quad (1.1)$$

Let $S_N(\kappa)$ be the set of $\sigma \in S_N$ that satisfy (1.1), let $V_N(\kappa) = \mu_N(S_N(\kappa))$ be its volume, and let $\kappa_N = \max_{\sigma \in S_N} \min_a g_a \cdot \sigma / \sqrt{N}$ be the largest achievable margin. The basic questions are the exponential rate $N^{-1} \log V_N(\kappa)$, called the free energy, and the limit of κ_N , called the capacity. For $\kappa \geq 0$ each constraint confines σ to a spherical cap, and the feasible set is an intersection of geodesically convex sets. For $\kappa < 0$ each constraint removes a spherical cap, and the feasible set need be neither convex nor connected. Franz and Parisi proposed this negative spherical perceptron as the simplest mean-field model of jamming [58]: its capacity is a satisfiability (SAT–UNSAT) transition, whose critical behavior was argued to be that of the jamming of hard spheres [59].

Gardner and Derrida computed the free energy by the replica method, under the replica-symmetric ansatz [62, 63]. Their formula predicts the capacity $\kappa_{\text{RS}}(\alpha)$, the solution of

$$\alpha \mathbb{E}(\kappa_{\text{RS}} - G)_+^2 = 1,$$

where G is a standard Gaussian variable; $\kappa_{\text{RS}}(\alpha)$ is negative exactly when $\alpha > 2$. For $\kappa \geq 0$ the replica-symmetric formula was proved by Shcherbina and Tirozzi [115] and by Talagrand [128, Chapter 8]; for this and other rigorous work on perceptrons see Part I, Chapter 1. For $\kappa < 0$ the physical picture is richer [58, 59]. At small constraint densities the replica-symmetric solution is predicted to be correct. As the constraints become more demanding, it loses stability on a de Almeida–Thouless line [45], inside the satisfiable phase. For margins between $\kappa_{1\text{RSB}} \approx -2.05$ and zero [58], this instability is predicted to be a continuous transition to full replica symmetry breaking in the sense of Parisi [104, 105]. In this phase the order parameter, a probability measure that describes the overlaps of typical solutions, has an interval in its support. The jamming line then lies inside the broken phase, and Gardner’s formula for the capacity should be an upper bound that is not exact. Annesi, Malatesta, and Zamponi computed the satisfiability

threshold numerically from the full replica-symmetry-breaking equations, and located the region of the phase diagram in which the overlap distribution of typical solutions has connected support [3].

Rigorous results for negative margins have been partial. Stojnic proved, using Gordon's comparison inequality [66], that Gardner's prediction κ_{RS} bounds the capacity from above for every margin [118]. The upper bound of Montanari, Zhong, and Zhou [90, Theorem 4.2] shows that Gardner's formula overestimates the capacity when the margin is sufficiently negative, that is, when α is large. Near $\alpha = 2$, where $\kappa_{\text{RS}}(\alpha)$ is close to zero, these bounds do not quantify the separation.

El Alaoui and Sellke stated the Parisi variational problem for this model in the form studied below, and conjectured, following [59], that it gives the free energy and the capacity [54, Conjecture 1]. In their terms, the model has no overlap gap if some minimizer has a Parisi measure supported on an interval on which its distribution function is strictly increasing [54, Definition 2]. Under this hypothesis they constructed an efficient algorithm that, given oracle access to such a minimizer, produces vectors of norm close to $\sqrt{QN}$ with a small vector of constraint violations, where Q is the top of the support of that minimizer [54, Theorem 1]; they expect these vectors to approximate barycenters of pure states. They left open whether the hypothesis holds for any parameters. In Part I we prove a Parisi formula for this model (Part I, Theorem 1.1), and we deduce that, for $\alpha > 2$, the free energy converges to the infimum of the Parisi functional below its critical margin and κ_N converges to this critical margin (Part I, Theorem 1.2). This part studies the variational problem itself.

The structure of Parisi minimizers has been determined rigorously for spin glasses of Sherrington–Kirkpatrick (SK) type. Following Guerra's interpolation bound [68], Talagrand proved the Parisi formula for the SK model and mixed even p -spin models [124], and Panchenko extended it to general mixtures [98, 99]. For spherical models with even interactions, Talagrand proved the Crisanti–Sommers formula [125] (see [44] for the formula), and W.-K. Chen proved it for general mixtures [37]. Auffinger and W.-K. Chen proved that the Parisi functional is strictly convex, so that its minimizer is unique [7], and Jagannath and Tobasco gave another proof [76, Theorem 2]. Auffinger and W.-K. Chen also established structural properties of Parisi measures [8]; see also [126]. Jagannath and Tobasco recast the Parisi functional through stochastic dynamic programming, which gives the well-posedness of the Parisi equation and its control representation [76], and they derived first-variation criteria for its minimizer [77]. W.-K. Chen developed variational representations and optimality conditions [38]. By strict convexity, an order parameter that satisfies these first-order conditions is the Parisi measure. Toninelli proved that the replica-symmetric formula fails below the de Almeida–Thouless line of the SK model [129]. Jagannath and Tobasco proved that every replica-symmetric Parisi measure of a mixed p -spin model satisfies the de Almeida–Thouless

condition, and that replica symmetry holds up to the de Almeida–Thouless line in parts of the phase diagram [77, Theorems 1.8, 1.9, and 1.13]. W.-K. Chen proved that the de Almeida–Thouless line is the exact transition line for the SK model with a centered Gaussian external field [39], and Lopatto proved that replica symmetry holds up to the de Almeida–Thouless line for the SK model with a uniform external field $h > 0$ [86].

The support of the Parisi measure has since been determined in several regimes. Auffinger, W.-K. Chen, and Zeng proved that the zero-temperature Parisi measure of mixed p -spin models has infinitely many points in its support [11]. Zhou proved full replica symmetry breaking for the SK model at zero external field and inverse temperatures slightly above the critical value (one in the usual SK normalization) [134]. There the Parisi measure is supported on an interval, with a smooth density below its endpoint and an atom at the endpoint. A recent preprint extends this description to every inverse temperature above one [87], and a recent preprint by H.-B. Chen shows that at zero temperature the SK order parameter has full support on $[0, 1)$ [32]. Another recent preprint constructs mixed even p -spin Ising models whose Parisi measures have exactly k steps of replica symmetry breaking, for every k , together with transitions from r to $r + 1$ steps [5]. These structural results rely on the strict convexity of the Parisi functional and on the resulting characterization of the unique Parisi measure by its first-order conditions.

The perceptron differs from these models in three ways. First, the terminal datum of its Parisi equation, (1.2) below, is the logarithm of a Gaussian probability, and it becomes singular as the top of the support of the Parisi measure approaches one. This hard wall must be kept at a distance from every minimizer. Second, the functional is not known to be convex. Its first-order conditions are necessary for minimality but are not known to be sufficient, and minimizers are not known to be unique. A candidate that satisfies the first-order conditions therefore cannot be certified as a minimizer, and every minimizer has to be described. Third, the transition studied here takes place in a singular limit with several scales. For $\alpha = 2 + e$ the relevant margins lie within $O(e^2)$ of κ_{RS} , the relevant overlaps lie within $O(e^2)$ of one, and the Parisi measure has a density of order e^{-2} . The scaling limit is itself a nontrivial variational problem, and the order parameter must be controlled both near the top of its support and far below it.

To our knowledge, full replica symmetry breaking has not been established for any perceptron model. Our contribution is to determine the minimizers near the de Almeida–Thouless line for constraint densities slightly above two. Let $\alpha = 2 + e$ with $e > 0$ small. We prove that the replica-symmetric order parameter is the unique minimizer at the de Almeida–Thouless margin κ_d and in a window of width of order e^2 below it. In a window of the same width above κ_d , every minimizer has a Parisi measure supported on a single interval $[q_0, Q]$, with a density of order e^{-2} on (q_0, Q) and atoms at both endpoints (Theorem 1.2). Thus replica symmetry is lost exactly at the de Almeida–Thouless line, through a continuous transition that passes directly to full replica symmetry breaking,

as predicted in [58, 59]. We also locate the three thresholds. The critical margin κ_c of the functional satisfies $\kappa_d < \kappa_c < \kappa_{RS}$, the distance $\kappa_{RS} - \kappa_d$ is $\theta_d e^2 + O(e^3 \log(1/e))$ for an explicit $\theta_d > 0$, and $\kappa_{RS} - \kappa_c$ is of exact order e^3 (Theorem 1.1). Combined with the capacity theorem of Part I, the last statement shows that Gardner's formula overestimates the capacity by an amount of exact order $(\alpha - 2)^3$ (Part I, Corollary 1.6).

The main step is a compactness theorem (Theorem 1.3). In variables adapted to the scale e^2 , it places the order parameter of every minimizer, at all margins $\kappa_{RS} - \theta e^2$ with θ in a fixed compact subset of $(0, \infty)$, in one compact class. The obstruction is a thin tail of the support at large depth, which costs little in the value of the functional. We exclude it with a balance law, obtained by integrating the first-order conditions against an integrating factor. On the compact class the functional has a universal limit, the flat functional, whose unique minimizer at the crossing is replica symmetric (Theorem 1.4). A local analysis of the first-order conditions near the crossing then gives Theorem 1.2.

1.1 The Parisi variational problem

An *order parameter* is a right-continuous nondecreasing function $\gamma: [0, 1] \rightarrow [0, 1]$ with $\gamma = 1$ on a neighborhood of one. We write $\mathcal{U}$ for the set of order parameters and μ_γ for the probability measure on $[0, 1]$ with distribution function γ , called the *Parisi measure* of γ . For $\gamma \in \mathcal{U}$ put

$$Q = \inf\{q : \gamma(q) = 1\} < 1, \quad \delta = 1 - Q,$$

$$\lambda_\gamma(q) = \int_q^1 \gamma(t) dt, \quad q_0 = \min \text{supp } \mu_\gamma.$$

Thus $\text{supp } \mu_\gamma \subset [q_0, Q]$ and $q_0, Q \in \text{supp } \mu_\gamma$. We call $\lambda_\gamma(0)$ the *mass* of γ , and we call γ *replica symmetric* if μ_γ is a point mass, that is, if $\gamma = \mathbf{1}_{[q, 1]}$ for some $q \in [0, 1]$.

Let ϕ and Φ be the standard Gaussian density and distribution function, and put $\phi_0 = \phi(0) = (2\pi)^{-1/2}$. Fix $\alpha > 0$ and $\kappa \in \mathbb{R}$. For $\gamma \in \mathcal{U}$ let $u = u_\gamma(\cdot, \cdot; \kappa)$ solve the Parisi equation

$$\partial_q u = -\frac{1}{2}(\partial_{xx} u + \gamma(q)(\partial_x u)^2) \quad \text{on } [0, Q) \times \mathbb{R}, \quad u(Q, x) = \log \Phi\left(\frac{x - \kappa}{\sqrt{\delta}}\right). \quad (1.2)$$

On an interval where γ is constant the equation is solved by a Cole–Hopf transformation [42, 71], and for general γ the solution is the limit of the solutions for step order parameters (Chapter 2). This construction goes back to Parisi [105], the equation (1.2) to Duplantier [52], and the passage to general γ uses Lipschitz continuity in the order parameter, due to Guerra [67]; see also [76, Lemma 14 and Remark 15]. For this terminal datum, El Alaoui and Sellke constructed the solution in the same way and proved its regularity [54, Section 4]. The terminal datum is $\log \mathbb{P}(x + \sqrt{\delta} G \geq \kappa)$. It is the hard-wall

form of the constraint (1.1), smoothed by the Gaussian fluctuations of the overlaps above Q . The *Parisi functional* and its infimum are

$$\mathcal{P}_\kappa(\gamma) = \alpha u_\gamma(0, 0; \kappa) + \frac{1}{2} \int_0^Q \frac{dq}{\lambda_\gamma(q)} + \frac{1}{2} \log \delta, \quad \mathcal{P}_*(\kappa) = \inf_{\gamma \in \mathcal{U}} \mathcal{P}_\kappa(\gamma). \quad (1.3)$$

The last two terms of $\mathcal{P}_\kappa$ form the Crisanti–Sommers entropy of spherical spin glasses [37, 44, 125]. For the replica-symmetric order parameters the functional is explicit:

$$\mathcal{P}_\kappa^{\text{RS}}(q) := \mathcal{P}_\kappa(\mathbf{1}_{[q,1]}) = \alpha \mathbb{E} \log \Phi\left(\frac{\sqrt{q}G - \kappa}{\sqrt{1-q}}\right) + \frac{q}{2(1-q)} + \frac{1}{2} \log(1-q) \quad (0 \leq q < 1). \quad (1.4)$$

This is Gardner’s replica-symmetric functional [62, 63], and its minimum over q is Gardner’s formula for the free energy. The *critical margin* of the functional is

$$\kappa_c(\alpha) = \sup\{\kappa \in \mathbb{R} : \mathcal{P}_*(\kappa) > -\infty\}. \quad (1.5)$$

By Proposition 2.17 and Theorems 2.16 and 2.18, $\kappa_c(\alpha)$ is finite, $\kappa_c(\alpha) \leq \kappa_{\text{RS}}(\alpha)$, and $\kappa_c(\alpha) < 0$ for $\alpha > 2$. For $\kappa < \kappa_c$ the infimum $\mathcal{P}_*(\kappa)$ is attained, and every minimizer satisfies $1 - Q > e^{2\mathcal{P}_*(\kappa)}$. Moreover $\mathcal{P}_*(\kappa) \rightarrow -\infty$ as $\kappa \uparrow \kappa_c$.

Every theorem of this part is a deterministic statement about $\mathcal{P}_\kappa$ and its minimizers. The results of §§2.6 and 2.7 are proved in Part I, Chapter 6, and are only stated here. No other proof in this part uses Part I. The random model enters only through the consequences of Part I recalled after Theorems 1.1 and 1.2. Unlike the Parisi functionals of SK-type models, $\mathcal{P}_\kappa$ is not known to be convex, and its minimizers are not known to be unique in general. Our theorems therefore describe every minimizer. By a minimizer we always mean a global minimizer of $\mathcal{P}_\kappa$ over $\mathcal{U}$.

The first-order conditions for minimizers are expressed through the *Parisi diffusion*

$$dX_q = \gamma(q) \partial_x u(q, X_q) dq + dB_q, \quad X_0 = 0,$$

where B is a standard Brownian motion, and through the functions

$$D(q) = \alpha \mathbb{E} [\partial_x u(q, X_q)^2] - \int_0^q \frac{dt}{\lambda_\gamma(t)^2}, \quad \mathcal{H}_\gamma(t) = -\frac{1}{2} \int_0^t D(q) dq. \quad (1.6)$$

The right derivative of $\vartheta \mapsto \mathcal{P}_\kappa(\gamma + \vartheta(\tilde{\gamma} - \gamma))$ at $\vartheta = 0$ equals $\int \mathcal{H}_\gamma d\mu_{\tilde{\gamma}} - \int \mathcal{H}_\gamma d\mu_\gamma$. Consequently every minimizer γ satisfies the obstacle condition $\text{supp } \mu_\gamma \subset \text{argmin}_{[0,1]} \mathcal{H}_\gamma$ (Proposition 2.11); in particular $D = 0$ on $\text{supp } \mu_\gamma$. These conditions are necessary but, in the absence of convexity, not known to be sufficient.

For a replica-symmetric order parameter $\mathbf{1}_{[q,1]}$ the diffusion is a Brownian motion up to the level q , and everything is explicit. Let $R = \phi/\Phi$, and let

$$V(z) = R(z)(z + R(z)) = -\frac{d^2}{dz^2} \log \Phi(z),$$

which takes values in $(0, 1)$. The derivative of $\mathcal{P}_\kappa^{\text{RS}}$ at q is $-\frac{1}{2}D(q)$, where D is the function (1.6) of $\mathbf{1}_{[q,1]}$, and one half of the curvature of the obstacle potential at the atom is

$$\Lambda_R = \frac{1}{2}\mathcal{H}_{\mathbf{1}_{[q,1]}}''(q) = \frac{1 - \alpha \text{EV}(Z)^2}{4(1 - q)^2}, \quad Z = \frac{\sqrt{q}G - \kappa}{\sqrt{1 - q}} \quad (1.7)$$

(Lemma 3.3). This is the replicon eigenvalue of the replica-symmetric solution. The equation $\alpha \text{EV}(Z)^2 = 1$ is the de Almeida–Thouless condition of the physics analysis of this model [59, (52)]; for perceptrons, the stability of the replica-symmetric solution was studied by Gardner and Derrida [63] and by Bouten [23]. If q is a critical point of $\mathcal{P}_\kappa^{\text{RS}}$ and $\Lambda_R < 0$, then moving a small amount of mass from the atom at q to a nearby level lowers the functional, so $\mathbf{1}_{[q,1]}$ is not a minimizer.

1.2 Main results

Three margins organize our results: Gardner’s prediction κ_{RS} , the critical margin κ_c , and the de Almeida–Thouless margin κ_d , which we now define. For $0 < \alpha < \pi^2/4$ and $\kappa < \kappa_{\text{RS}}(\alpha)$, the function $\mathcal{P}_\kappa^{\text{RS}}$ has exactly one critical point $q^*(\kappa)$ in $(0, 1)$, which is its unique minimizer on $[0, 1)$ (Lemma 3.5 and Lemma 3.3(c)). Thus $\mathbf{1}_{[q^*(\kappa), 1]}$ is the only replica-symmetric order parameter whose atom satisfies $D = 0$, and $\mathcal{P}_\kappa^{\text{RS}}(q^*(\kappa)) = \min_q \mathcal{P}_\kappa^{\text{RS}}(q)$ is Gardner’s formula for the free energy. Let $\Lambda_R(\kappa)$ denote the replicon eigenvalue (1.7) at $q = q^*(\kappa)$. For $\alpha = 2 + e$ with $e > 0$ sufficiently small, there is exactly one margin $\kappa_d(\alpha) \in (\kappa_{\text{RS}} - c_2, \kappa_{\text{RS}})$ with $\Lambda_R(\kappa_d) = 0$, where $c_2 > 0$ is a numerical constant, and

$$\Lambda_R(\kappa) < 0 \quad (\kappa_d < \kappa < \kappa_{\text{RS}}), \quad \Lambda_R(\kappa) > 0 \quad (\kappa_{\text{RS}} - c_2 \leq \kappa < \kappa_d)$$

(Proposition 3.8). We call $\kappa_d(\alpha)$ the *de Almeida–Thouless margin*, and put $q_d = q^*(\kappa_d)$ and $\gamma_d = \mathbf{1}_{[q_d, 1]}$. On $(\kappa_d, \kappa_{\text{RS}})$ the replica-symmetric solution is unstable. Finally, put

$$c_0 = \int_{\mathbb{R}} V(z)(1 - V(z)) \, dz \in (0, \infty), \quad \theta_d = \frac{(2\pi)^{3/2}}{128 c_0^2}.$$

Our first result is the ordering of the three margins near $\alpha = 2$.

Theorem 1.1. *There are numerical constants $e_0, c > 0$ and $C < \infty$ such that for every $0 < e < e_0$ and $\alpha = 2 + e$,*

$$\kappa_{\text{RS}}(\alpha) - \kappa_d(\alpha) = \theta_d e^2 + O(e^3 \log(1/e)), \quad c e^3 \leq \kappa_{\text{RS}}(\alpha) - \kappa_c(\alpha) \leq C e^3. \quad (1.8)$$

In particular $\kappa_d(\alpha) < \kappa_c(\alpha) < \kappa_{\text{RS}}(\alpha)$, and $\kappa_c(\alpha) - \kappa_d(\alpha) = \theta_d e^2 + O(e^3 \log(1/e))$.

Since $\kappa_{\text{RS}}(2 + e) = -e/(8\phi_0) + O(e^2)$ (Lemma 3.1), the three margins are of order e , and they lie within $O(e^2)$ of each other. On the interval (κ_d, κ_c) , whose length is

asymptotic to $\theta_d e^2$, the replica-symmetric solution is unstable while $\mathcal{P}_*$ is finite and attained; there no minimizer is replica symmetric (Corollary 3.14). Gardner's formula overestimates the critical margin of the functional by an amount of exact order e^3 : it is not exact for any small $e > 0$, yet its error is one order smaller than the length of the unstable interval. The expansion of $\kappa_{\text{RS}} - \kappa_d$ is part of Proposition 3.8. The two cubic bounds use a zero-temperature characterization of κ_c , as the threshold at which a functional J^k with the quadratic terminal penalty $-\frac{1}{2}(k-x)_+^2$ first takes a negative value (Theorem 2.21). The upper bound on $\kappa_{\text{RS}} - \kappa_c$ comes from an explicit dual certificate for J^k : a single function $g \geq k - x$ with $\alpha \mathbb{E}g(G)^2 \leq 1$ and $\alpha \mathbb{E}g'(G)^2 \leq 1$ forces $J^k \geq 0$ (Theorems 3.10 and 3.11). The lower bound comes from an explicit coefficient, carried by a layer of width of order e^2 just below the top, at which J^k is negative for $k = \kappa_{\text{RS}} - ce^3$ (Theorem 3.13). Part I, Theorem 1.4, proves the same lower bound by evaluating $\mathcal{P}_\kappa$ directly at an order parameter with two support points near one, whose top layer plays the role of this coefficient. The constants c and C are not matched, and we do not identify the limit of $(\kappa_{\text{RS}} - \kappa_c)/e^3$.

Theorem 1.1 concerns the functional. In Part I we prove that, for every $\alpha > 2$,

$$\frac{1}{N} \log V_N(\kappa) \longrightarrow \mathcal{P}_*(\kappa) \quad (\kappa < \kappa_c(\alpha)), \quad \kappa_N \longrightarrow \kappa_c(\alpha), \quad (1.9)$$

in probability, and that $\mathbb{P}\{N^{-1} \log V_N(\kappa) > -C\} \rightarrow 0$ for every $\kappa \geq \kappa_c(\alpha)$ and every $C < \infty$ (Part I, Theorem 1.2). In Part I this capacity theorem is combined with Theorem 1.1, Corollary 3.14 and Proposition 2.24 (Part I, Corollaries 1.6 and 1.7). For $\alpha = 2 + e$ with $e > 0$ small enough, the capacity $\kappa_c(\alpha)$ satisfies $ce^3 \leq \kappa_{\text{RS}}(\alpha) - \kappa_c(\alpha) \leq Ce^3$, and Gardner's formula for the free energy fails at every margin in $(\kappa_d, \kappa_{\text{RS}})$. For every $\alpha > 2$, the free energy diverges only logarithmically at the capacity. The limit $N \rightarrow \infty$ is taken at fixed e , and the bounds describe how this limit behaves as $e \downarrow 0$.

Our main theorem describes all minimizers at, below, and above the de Almeida–Thouless margin. Put

$$B_V = \int_{\mathbb{R}} V(z)^2 (1 - V(z)) \, dz, \quad x_f = \frac{1}{2} - \frac{c_0}{4B_V}.$$

By Lemma 8.2, $x_f = \int_{\mathbb{R}} V'(z)^2 \, dz / (2B_V)$ and $0 < x_f < \frac{1}{4}$.

Theorem 1.2. *There are numerical constants $e_0, c > 0$ and $C < \infty$ such that the following holds for every $0 < e < e_0$ and $\alpha = 2 + e$.*

- (i) *For every $\kappa \in [\kappa_d - ce^2, \kappa_d]$, the replica-symmetric order parameter $\mathbf{1}_{[q^*(\kappa), 1]}$ is the unique minimizer of $\mathcal{P}_\kappa$. In particular $\gamma_d = \mathbf{1}_{[q_d, 1]}$ is the unique minimizer of $\mathcal{P}_{\kappa_d}$.*
- (ii) *For every $\kappa \in (\kappa_d, \kappa_d + ce^2)$, the functional $\mathcal{P}_\kappa$ has minimizers, and every minimizer γ has the following properties.*

- (a) *The Parisi measure is supported on one nondegenerate interval: $\text{supp } \mu_\gamma = [q_0, Q]$ with $0 < q_0 < Q < 1$.*
- (b) *The distribution function γ is continuously differentiable on (q_0, Q) , and*

$$\frac{c}{e^2} \leq \gamma'(q) \leq \frac{C}{e^2} \quad (q_0 < q < Q).$$

- (c) *μ_γ has an atom at q_0 of mass larger than $\frac{1}{2}x_f$ and an atom at Q of mass larger than $\frac{1}{4}$. As $e \rightarrow 0$ and $(\kappa - \kappa_d)/e^2 \rightarrow 0$, these masses converge to x_f and $1 - x_f$, uniformly over minimizers.*
- (d) *$Q - q_0 \leq C(\kappa - \kappa_d)$, $|t - q_d| \leq C(\kappa - \kappa_d)$ for every $t \in \text{supp } \mu_\gamma$, and $\|\gamma - \gamma_d\|_{L^1} \leq C(\kappa - \kappa_d)$.*

Part (i) says that the replica-symmetric order parameter is optimal at the de Almeida–Thouless margin and in a window of width ce^2 below it. Part (ii) says that immediately beyond this margin every optimal order parameter has full replica symmetry breaking. The support of its Parisi measure is an interval, so no order parameter with finitely many steps of replica symmetry breaking is a minimizer. By (d), the minimizers converge to γ_d as $\kappa \downarrow \kappa_d$. Thus the transition out of replica symmetry is continuous, and it passes directly to full replica symmetry breaking. All constants are numerical, so the two windows have width ce^2 uniformly in α . Since $\kappa_c - \kappa_d = \theta_d e^2 + O(e^3 \log(1/e))$, the window $(\kappa_d, \kappa_d + ce^2)$ is a fixed fraction of the interval (κ_d, κ_c) on which the replica-symmetric solution is unstable.

By (b), the order parameter increases at rate of order e^{-2} across the interval, which is the scale on which it varies near $\alpha = 2$. The continuous part of μ_γ has mass of order $(Q - q_0)/e^2$ (Corollary 8.12), so near the transition most of the mass sits in the two atoms, and most of it in the top atom. The theorem gives no lower bound on $Q - q_0$, and it does not assert uniqueness of the minimizer above κ_d . On the interval of the support, $D \equiv 0$, and the density of μ_γ is given by an explicit formula in terms of the solution of (1.2) (Theorem 8.13).

Theorem 1.2 verifies the hypothesis of El Alaoui and Sellke, whose functional coincides with (1.3). For every $\kappa \in (\kappa_d, \kappa_d + ce^2)$ and every minimizer γ , part (ii) shows that $q_0 < Q$ and that γ is strictly increasing on $[q_0, Q]$, which is the no-overlap-gap condition of [54, Definition 2]. Since minimizers exist, their algorithm [54, Theorem 1] applies at these margins, with any minimizer as the oracle. In part (i) the hypothesis fails, because the unique minimizer is a point mass. Combined with (1.9), part (i) also shows that $N^{-1} \log V_N(\kappa) \rightarrow \mathcal{P}_\kappa^{\text{RS}}(q^*(\kappa))$ in probability for $\kappa \in [\kappa_d - ce^2, \kappa_d]$. Hence, near $\alpha = 2$, Gardner’s formula for the free energy is exact at the de Almeida–Thouless margin and in a window of width ce^2 below it.

The proofs rest on control of all minimizers at the natural scales near $\alpha = 2$, uniformly in a window of margins. For $\theta > 0$ put

$$\kappa_e(\theta) = \kappa_{\text{RS}}(2 + e) - e^2\theta,$$

and measure overlaps from the top through the rescaled level $s = (1 - q)/e^2$. For $\gamma \in \mathcal{U}$ let

$$\zeta(s) = \gamma(1 - e^2s), \quad d = \frac{1 - Q}{e^2}, \quad s_0 = \frac{1 - q_0}{e^2}, \quad \ell_0 = \frac{\lambda_\gamma(0)}{e^2}$$

be the *rescaled profile*, the *rescaled top gap*, the *rescaled depth*, and the *rescaled mass*. The profile ζ is nonincreasing, equals one on $[0, d]$, vanishes beyond s_0 , and has integral ℓ_0 . A replica-symmetric order parameter has $d = s_0 = \ell_0$. By Theorem 1.1, $\kappa_d = \kappa_e(\theta_d + O(e \log(1/e)))$, and by Theorem 3.11(iii) the margins $\kappa_e(\theta)$, θ in a compact interval $\Theta \subset (0, \infty)$, lie below κ_c by at least $\frac{1}{2}e^2 \min \Theta$ for small e , so that minimizers exist.

Theorem 1.3. *Let $\Theta \subset (0, \infty)$ be a compact interval. There are constants*

$$0 < d_1 \leq D_{\text{top}} < \infty, \quad 0 < \ell_{\min} \leq L < \infty, \quad S_* < \infty, \quad m_* < 1, \quad e_0 > 0,$$

depending only on Θ , such that for every $0 < e < e_0$ and $\theta \in \Theta$, every minimizer γ of $\mathcal{P}_{\kappa_e(\theta)}$ satisfies

$$d_1 < d \leq D_{\text{top}}, \quad \ell_{\min} \leq \ell_0 \leq L, \quad s_0 \leq S_*, \quad \gamma(Q-) \leq m_*.$$

Moreover, every minimizer that is not replica symmetric satisfies $s_0 \leq 1/(4c_w^2)$, where

$$c_w = \frac{3\phi_0}{4} [\phi(3) - 3\Phi(-3)] > 0.$$

In the original variables, the support of every minimizer lies in $[1 - S_*e^2, 1 - d_1e^2]$, its mass lies between $\ell_{\min}e^2$ and Le^2 , and the jump of γ at the top of the support is at least $1 - m_*$. The difficult statement is the bound on the depth s_0 ; the other bounds use it. Comparison with replica-symmetric trial order parameters and the first-order condition at the bottom of the support give only $s_0 \leq C_0e^{-2/3}$ (Theorem 4.11). This leaves room for minimizers with a large atom near the top, above a thin tail of small mass that extends to depths much larger than e^2 . Such a tail costs little in the value of the functional, and indeed near-minimizers can have tails of vanishing mass at arbitrary depth (Remark 5.5). Thus no estimate of the minimal value alone yields Theorem 1.3; the proof uses the first-order conditions of exact minimizers.

Finally, the rescaled problem has a universal limit as $e \rightarrow 0$. A *profile* is a nonincreasing function $\zeta: [0, \infty) \rightarrow [0, 1]$; it is *normalized* if $\int \zeta = 1$. The *normalized profile* of γ is $\tilde{\zeta}(s) = \zeta(\ell_0s) = \gamma(1 - e^2\ell_0s)$, and the replica-symmetric normalized profile is $\mathbf{1}_{[0,1]}$. For normalized profiles ζ with bounded support that equal one on some interval $[0, d_1]$, we define in Definition 7.1 the *flat functional*

$$F_t(\zeta) = A(\zeta) + tI(\zeta), \quad t \in \mathbb{R}.$$

Here $A(\zeta)$ is an explicit entropy term, and $I(\zeta)$ is a renormalized integral over $y \in \mathbb{R}$ of the solution $U = U_\zeta$ of

$$\partial_s U = \frac{1}{2}(\partial_{yy} U + \zeta(s)(\partial_y U)^2), \quad U(s, y) = \log \Phi(y/\sqrt{s}) \quad (0 < s \leq d_1), \quad (1.10)$$

which is the Parisi equation in rescaled variables, with neither e nor the margin. In the limit, the Gaussian law of the bottom of the Parisi diffusion is replaced by Lebesgue measure in y , which is the reason for the name. The term A is maximized by $\mathbf{1}_{[0,1]}$ (Corollary 7.6), so it favors replica symmetry breaking, while I is minimized by $\mathbf{1}_{[0,1]}$ (Theorem 7.13). The parameter at the crossing is $t_d = 1/(4c_0)$.

Theorem 1.4. (i) *Let ζ be a normalized profile with bounded support that equals one on a neighborhood of zero. Then*

$$F_{t_d}(\zeta) - F_{t_d}(\mathbf{1}_{[0,1]}) \geq \frac{9}{1024} \|\zeta - \mathbf{1}_{[0,1]}\|_{L^1}^3, \quad (1.11)$$

and more generally $F_t(\zeta) - F_t(\mathbf{1}_{[0,1]}) \geq \frac{9c_0 t}{256} \|\zeta - \mathbf{1}_{[0,1]}\|_{L^1}^3$ for every $t \geq t_d$. In particular, for every $t \geq t_d$ the replica-symmetric profile $\mathbf{1}_{[0,1]}$ is the unique minimizer of F_t among such profiles.

(ii) *Let $\Theta \subset (0, \infty)$ be a compact interval, and let $0 < d_1 \leq S$. There are $C < \infty$ and $e_0 > 0$, depending only on Θ , d_1 , and S , such that the following holds for $0 < e < e_0$ and $\theta \in \Theta$. There is a number K_e , depending only on e and θ , such that every $\gamma \in \mathcal{U}$ with $d \geq d_1$ and $s_0 \leq S$ satisfies*

$$\left| \mathcal{P}_{\kappa_e(\theta)}(\gamma) - K_e - \Psi_\theta(\ell_0) - e \left\{ \left(\phi_0 - \frac{1}{8\phi_0} \right) \frac{\theta}{\ell_0} + \frac{1}{8} \log \ell_0 + 2\phi_0 + F_{2\phi_0\sqrt{\ell_0}}(\tilde{\zeta}) \right\} \right| \leq Ce^2 \log(1/e), \quad (1.12)$$

where $\Psi_\theta(\ell) = 2\phi_0\theta/\ell + \log \ell$.

The number K_e does not depend on γ ; it is defined at the beginning of §7.2. Thus, to leading order, the value depends on the order parameter only through its mass, via $\Psi_\theta(\ell_0)$, which is minimized at $\ell_0 = 2\phi_0\theta$. The shape of the profile enters at order e through F_t at $t = 2\phi_0\sqrt{\ell_0}$, and at $\theta = \theta_d$ and $\ell_0 = 2\phi_0\theta_d$ this parameter is exactly t_d ; see (7.9). The threshold t_d is sharp: for $t < t_d$ the profile $\mathbf{1}_{[0,1]}$ is not a minimizer of F_t , as shown after Proposition 7.12. This is the flat counterpart of the instability of the replica-symmetric solution beyond κ_d .

Several ingredients of the proofs hold far from $\alpha = 2$ and may be of independent interest. The critical margin $\kappa_c(\alpha)$ equals the threshold $\kappa_J(\alpha)$ of the zero-temperature functional J^k for every $\alpha > 0$ (Theorem 2.21). For every $\alpha > 2$ the infimum diverges

logarithmically at the critical margin, $\alpha \log \eta - C_\alpha \leq \mathcal{P}_*(\kappa_c - \eta) \leq \frac{1}{2} \log \eta + C_\alpha$ for small $\eta > 0$ (Proposition 2.24). Finally, for every $\alpha > 0$, every margin, and every order parameter with $q_0 > 0$, the value $\mathcal{P}_\kappa(\gamma)$ equals the replica-symmetric value $\mathcal{P}_\kappa^{\text{RS}}(1 - \lambda_\gamma(0))$ of the same mass plus an integral over the levels $q \in [q_0, Q]$, with weight $(1 - \gamma(q))/(2\lambda_\gamma(q))$, of an explicit coefficient (Lemma 6.3 and (6.14)). We call this the jamming identity.

Throughout, c and C denote positive numerical constants, and constants with subscripts, such as C_α , depend only on the indicated quantities. Their values may change from one occurrence to the next. Unless stated otherwise, $O(\cdot)$ refers to $e \downarrow 0$ with a numerical implied constant. No constant depends on the order parameter under consideration.

1.3 Ideas of the proof

The proof has two stages. The first is global: it places the rescaled profiles of all minimizers, at all margins $\kappa_e(\theta)$ with θ in a compact interval, in one compact class (Steps 1–3). The second is local: on this class the functional has a universal limit, which forces minimizers near κ_d to be close to γ_d , and the first-order conditions can then be analyzed near γ_d (Steps 4 and 5). Figure 1.1 shows how the steps depend on one another. Throughout we use the function $v = -\lambda_\gamma \partial_{xx} u$, which takes values in $(0, 1)$, and the scalar functions

$$S(q) = \alpha \mathbb{E} v(q, X_q)^2, \quad \chi(q) = \frac{\alpha \mathbb{E} v(q, X_q) - 1}{\lambda_\gamma(q)}, \quad w(q) = \alpha \mathbb{E} [v(1 - v)](q, X_q).$$

The function D of (1.6) is continuously differentiable with $D' = (S - 1)/\lambda_\gamma^2$, and every minimizer satisfies $D = 0$ and $S \leq 1$ on its support (Lemma 2.9 and Proposition 2.11). For a minimizer whose support is an interval, El Alaoui and Sellke proved $D = 0$ and $S = 1$ on the support; see the remark after Proposition 2.11.

Step 1: A priori bounds. Two inputs give the preliminary bounds. The first is the zero-temperature functional J^k . The certified margin $k_* \geq \kappa_{\text{RS}} - Ce^3$ of the dual certificate lies below κ_c . The resulting slack of J^k (Corollary 2.22), together with a universal bound on the mass of minimizers in terms of their top gap (Lemma 2.15), keeps the top away from the wall: $d \geq ce^2$ (Theorem 4.10). The second input is the first-order condition $D(q_0) = 0$ at the bottom of the support. Since the Parisi diffusion is a Brownian motion up to the level q_0 , this condition is an exact Gaussian identity at a single level. In the rescaled space variable $y = (x - \kappa)/e$, the negative half-line controls the mass, and a window of width of order $\sqrt{s_0}$ to the right of zero controls the depth. This gives $\ell_0 \leq L$ and $s_0 \leq C_0 e^{-2/3}$ (Lemma 4.9 and Theorem 4.11). These bounds are proved in Chapters 2–4.

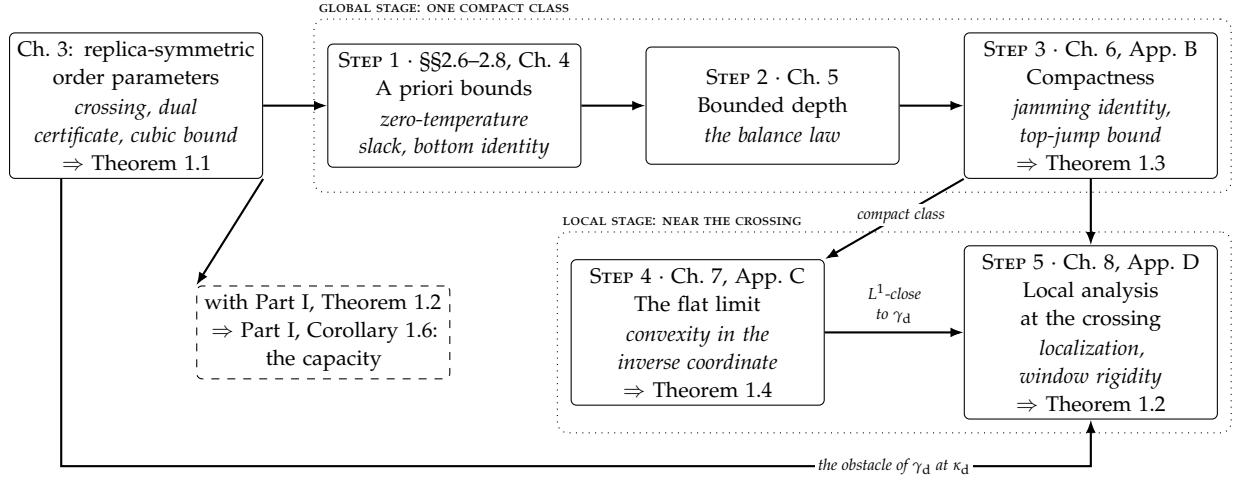


Figure 1.1. Selected dependencies in the two stages of the proof. Each solid box names a step (or, at the left, Chapter 3), the chapters or sections that carry it out and, where there is one, the theorem it proves; an arrow runs from a box to a box that uses it. The dashed box is the consequence for the random model, proved in Part I. The general facts about the Parisi functional in §§2.1–2.5 and Chapter A are used throughout and are not drawn. The results of §§2.6 and 2.7 are proved in Part I, Chapter 6.

Step 2: Bounded depth. To exclude thin tails at large depth we use the balance law

$$(\lambda_\gamma D)' = \chi(0) - \frac{w}{\lambda_\gamma} \quad \text{almost everywhere on } [0, 1), \quad (1.13)$$

valid for every order parameter that satisfies the first-order conditions (Lemma 5.2). It follows from $D' = (S - 1)/\lambda_\gamma^2$, the identity $S - 1 = \lambda_\gamma \chi - w$, and the relation $\chi = \chi(0) + \gamma D$, which holds because D vanishes on the support. The integrating factor λ_γ cancels the unknown term γD exactly. Integrating (1.13) between q_0 and Q , where D vanishes, and passing to rescaled variables gives

$$\int_d^{s_0} \frac{w(s)}{\ell(s)} ds = \tilde{\chi}(s_0 - d), \quad \tilde{\chi} = e^2 \chi(0) = \frac{\alpha \mathbb{E} v(s_0, Y_0) - 1}{\ell_0}, \quad (1.14)$$

where $\ell(s) = \lambda_\gamma(1 - e^2 s)/e^2 \leq \ell_0$ and Y_0 is the rescaled position of the diffusion at the bottom of the support, which is Gaussian. The right side is computed through an exact cancellation. At every level, $v - \mathbf{1}_{\{y < 0\}}$ has integral zero in the rescaled space variable y (Lemma 5.4), so the Gaussian average $\alpha \mathbb{E} v(s_0, Y_0)$ equals $\alpha \mathbb{P}(Y_0 < 0)$ up to $O(e^2(1 + s_0))$, and $\tilde{\chi} \ell_0 = \frac{e}{4} + o(e)$. For the left side, a heat-semigroup lower bound for the rescaled derivative of the solution and a density floor for the diffusion (Lemma 2.7) give $w(s) \geq c_w e \sqrt{s}$ at every level $s \in [d, s_0]$. Hence

$$\frac{2c_w e}{3\ell_0} (s_0^{3/2} - d^{3/2}) \leq \frac{e}{3\ell_0} (s_0 - d),$$

and for a minimizer that is not replica symmetric, $s_0 > d$ and this forces $s_0 \leq 1/(4c_w^2)$. The factors e and ℓ_0 cancel, so no lower bound on the mass or the top gap is needed. This is carried out in Chapter 5.

Step 3: Compactness. It remains to bound the mass and the top gap from below and the mass below the top atom from above. The main tool is the jamming identity (Lemma 6.3). In rescaled variables it reads

$$\mathcal{P}_\kappa(\gamma) = \Pi(\ell_0; q_0) + \int_d^{s_0} \frac{1 - \zeta(t)}{2\ell(t)} \mathcal{J}(t) dt, \quad (1.15)$$

where $\Pi(\ell_0; q_0) = \alpha \mathbb{E} \log \Phi((\sqrt{q_0} G - \kappa)/\sqrt{\lambda_\gamma(0)}) + q_0/(2\lambda_\gamma(0)) + \frac{1}{2} \log \lambda_\gamma(0)$ is the replica-symmetric expression (1.4) with the bottom and the mass decoupled, and $\mathcal{J}$ is an explicit jamming coefficient with values in $(1 - \alpha, 1)$. Near $\alpha = 2$, for profiles of bounded depth,

$$\mathcal{J}(t) \geq c e \min\left\{ \frac{\ell_0}{\sqrt{\ell(t)}}, \frac{1}{2e} \right\} - Ce$$

(Lemma 6.7). Thus the coefficient is at least $-Ce$ at every level, and it is large near the top, where $\ell(t)$ is small. This gives three bounds. A profile of bounded depth and small mass has a value above the replica-symmetric value, so $\ell_0 \geq \ell_{\min}$ (Proposition 6.10). The first-order conditions express the mass $\gamma(Q-)$ below the top atom through a ratio of Gaussian averages of derivatives of u near the top of the support (Lemma 6.11). The tilted-density representation of the law of the diffusion and a Chebyshev association inequality [21] bound this ratio, so that $\gamma(Q-) \leq m_* < 1$ (Theorem 6.13 and Chapter B). This bound is global: it does not assume closeness to a replica-symmetric order parameter. Finally, if the top gap is small and $\gamma(Q-) \leq m_*$, the profile is jammed just below the wall, where $\mathcal{J}$ is large, and (1.15) shows that the replica-symmetric order parameter of the same mass has a smaller value (Proposition 6.14). Hence $d > d_1$. This is carried out in Chapter 6, which completes the proof of Theorem 1.3.

Step 4: The flat limit. On the compact class, Theorem 1.4(ii) holds (Theorem 7.3). To prove Theorem 1.4(i) we parametrize a normalized profile by the inverse $T = \ell^{-1}$ of its primitive ℓ . In this coordinate the entropy term A is affine (Proposition 7.5), and I is convex along every segment from the replica-symmetric profile to a normalized step profile (Corollary 7.11). The convexity comes from an exact formula for the second variation (Proposition 7.10), whose integrand is nonnegative because an associated flat density is monotone (Proposition 7.9). It then suffices that the first variation of F_{t_d} at $\mathbf{1}_{[0,1]}$ is positive in every admissible direction. Along the segment in the inverse coordinate from $\mathbf{1}_{[0,1]}$ toward a normalized step profile with inverse profile T , this first variation equals $\int_{(0,1]} \mathcal{E}(r) d\dot{\sigma}(r)$, where $\dot{\sigma} = (T - \text{id})'$ is nonnegative and nondecreasing. The kernel $\mathcal{E}(r)$ is a positive weighted integral, over $\rho \in (r, 1)$, of the decrease of $\int V^2$ when

Lebesgue measure is transported for the time $\log(1/\rho)$ by the replica-symmetric top diffusion

$$dZ_\zeta = \left(R(Z_\zeta) + \frac{Z_\zeta}{2} \right) d\zeta + dB_\zeta.$$

A synchronous coupling with an Ornstein–Uhlenbeck process gives $\mathcal{E}(r) \geq (1-r)^3/96$ (Proposition 7.12), and convexity turns this cubic contact into the modulus $\frac{9}{1024}$ of (1.11).

Combining the expansion with the modulus gives a stability estimate that is uniform in θ near θ_d (Proposition 7.14): every minimizer at the margin $\kappa_e(\theta)$ satisfies

$$(\ell_0 - 2\phi_0\theta)^2 + e \|\tilde{\zeta} - \mathbf{1}_{[0,1]}\|_{L^1}^3 \leq C(e^2 \log(1/e) + e|\theta - \theta_d|).$$

Consequently, for every $\varepsilon > 0$ there is $\rho > 0$ such that for small e , every minimizer at every margin within ρe^2 of κ_d is within $\varepsilon(1 - q_d)$ of γ_d in L^1 (Corollary 8.8). The flat-limit analysis is carried out in Chapters C and 7; the last consequence is proved in §8.4.

Step 5: Local analysis at the crossing. The last step turns closeness in L^1 into the structure asserted in Theorem 1.2. First, the supports localize (Proposition 8.10): for margins within a small multiple of e^2 of κ_d , every support point of every minimizer lies within a small multiple of e^2 of q_d . The mechanism is the obstacle condition, written through the obstacle potential anchored at a support point and compared with the obstacle potential of γ_d at κ_d . The latter is explicit and has a strict cubic minimum at q_d of depth proportional to e , because the replicon eigenvalue vanishes at the crossing (Proposition 8.4). The comparison uses continuity of S at the natural scale, uniformly over order parameters close to γ_d (Lemma 8.6) and over the margins in the window (Lemma 8.9).

Second, window rigidity (Proposition 8.11) describes every order parameter that satisfies the first-order conditions and has its support in such a window. On a gap of the support the order parameter is constant, and the first-order conditions force a local curvature $\tilde{\Sigma}$ to be nonpositive somewhere in the gap (Lemma 2.12 and Lemma 8.3). In the window, the ratio of $\tilde{\Sigma}$ to the positive function $\mathbb{E}[v^2(1-v)](q, X_q)$ is close to a positive constant of the flat limit (Lemma 8.6), so $\tilde{\Sigma} > 0$ and the support is an interval. On an interval of the support, $D \equiv 0$ and $S \equiv 1$, which give the density of μ_γ explicitly; it is of order e^{-2} . Finally, if $Q > q_0$, comparing the bottom of the interval with the replica-symmetric order parameter of the same mean (Lemma 8.7) gives the width estimate

$$ce \frac{Q - q_0}{e^2} \leq -\tau(\kappa), \quad \tau(\kappa) = 4(1 - q^*(\kappa))^2 \Lambda_R(\kappa),$$

where $\tau(\kappa)$ has the sign of $\kappa_d - \kappa$. At and below κ_d this forces $Q = q_0$, and the minimizer is replica symmetric. Above κ_d it gives $Q - q_0 \leq C(\kappa - \kappa_d)$, while $q_0 = Q$ is excluded because the replicon eigenvalue is negative. This is carried out in Chapters D and 8, and the main theorems are assembled in Chapter 9.

Steps 3 and 5 follow arguments for the SK model. The top-jump bound of Step 3 follows the endpoint strategy of [8, Theorem 4], as in [87, Proposition 9.3]: the first-order conditions at the top of the support relate the mass below the top atom to Gaussian averages of derivatives of u there, and these averages are then bounded. Step 5 follows the near-critical architecture of Zhou [134]: closeness to the replica-symmetric order parameter, then positivity of a local curvature, then exclusion of gaps of the support through the first-order conditions at their endpoints. In [87, Section 9] gaps are excluded through the same endpoint conditions, combined with a sign condition inside the gap. Compared to these works, the local analysis differs in three ways. The functional here is not known to be convex, so the first-order conditions do not identify a minimizer, and every minimizer has to be localized before the local analysis applies. The windows of margins have width ce^2 with numerical constants, uniformly over all minimizers. The Parisi measures have atoms at both endpoints of the support, not only at the top. In every step, the terminal datum is the hard wall, which becomes singular as the top of the support approaches one. The differences in Step 3 are listed in §6.6.

The steps are arranged so that no step uses a later conclusion; the logical order is recorded in §9.1. Chapter A proves the regularity of the Parisi equation used in Chapter 2.

The Parisi functional and its optimality conditions

This chapter collects the deterministic facts about the hard-wall Parisi functional on which the rest of this part is built. §§2.1–2.5 contain general structural facts, valid for every order parameter and every margin. Lemma 2.1 collects the facts about Gaussian functions used throughout this part. We define the functional through a finite-cutoff form of the Parisi equation (Definition 2.2 and Lemma 2.3), prove that its solution is regular and depends Lipschitz continuously on the order parameter (Lemma 2.4), and record the Cole–Hopf description on constant blocks together with a comparison principle (Lemma 2.5). We then introduce the diffusion along which all first-order quantities are evaluated (Lemma 2.6), describe its law (Lemma 2.7), and state the stochastic control representation (Lemma 2.8). Lemma 2.9 gives the generator identities for the derivatives of the solution along the diffusion, and Lemma 2.10 gives a pointwise sandwich inequality. Finally, Proposition 2.11 computes the first variation and derives the obstacle conditions, of Karush–Kuhn–Tucker (KKT) type, satisfied by minimizers, and Lemma 2.12 turns these conditions into a principle for gaps and intervals of the support. The remaining sections, §§2.6–2.9, study the variational problem itself; their contents are described at the beginning of §2.6. Besides the first-order conditions, Step 1 in §1.3 uses three results of these sections: the value inequality (Lemma 2.13), the slack of the zero-temperature functional (Corollary 2.22) and the bound on the mass in terms of the top gap (Lemma 2.15).

Every statement of §§2.1–2.5 holds for general order parameters, not only for step order parameters. The one exception is the control representation of Lemma 2.8, which is stated for step coefficients and is used only through inequalities that extend to general order parameters by continuity.

The functional $\mathcal{P}_\kappa$ defined in (2.11) below is the Parisi functional of the negative spherical perceptron with hard-wall activation [54, 58]; see also Part I, §1.1. Nothing in this part depends on that interpretation.

Many of the facts in §§2.2–2.5 have counterparts in earlier work, which are credited in remarks next to the individual results. What is specific here is the following. The terminal datum is the hard wall $\log \Phi((x - \kappa)/\sqrt{\delta})$, which is increasing and concave in x rather than even, and which is singular as $Q \rightarrow 1$; this is why all estimates are stated in the classes $\mathcal{U}_r$. The first variation is computed at every $\gamma \in \mathcal{U}_r$ in every direction of $\mathcal{U}_r$, and the obstacle conditions are derived for every global minimizer and every KKT point, with no assumption on the support; the remark after Proposition 2.11 compares this with

[54, Proposition 16]. Part I, Chapter 2, derives several of its lemmas on the hard-wall functional from the results of §§2.1–2.5 and Chapter A.

2.1 Gaussian functions and Mills ratios

Throughout, G denotes a standard Gaussian variable, ϕ and Φ denote its density and distribution function, and $\phi_0 = \phi(0) = (2\pi)^{-1/2}$. For $q > 0$ let $\phi_q(x) = q^{-1/2}\phi(x/q^{1/2})$ be the $N(0, q)$ density. We use the functions

$$R(z) = \frac{\phi(z)}{\Phi(z)}, \quad \bar{R}(z) = z + R(z), \quad V(z) = R(z)\bar{R}(z), \quad h(z) = \log \Phi(z) + \frac{1}{2}R(z)^2, \quad (2.1)$$

and

$$m(k) = \mathbb{E}(k - G)_+^2, \quad m_1(k) = \mathbb{E}(k - G)_+, \quad \varrho(z) = R(z)^2 - z_-^2.$$

The function $\log \Phi((x - \kappa)/\sqrt{\delta})$ will be the terminal datum of the Parisi equation. Its first two derivatives in x are $R/\sqrt{\delta}$ and $-V/\delta$, which explains the role of R and V below. For $z \in \mathbb{R}$ let T_z be the random variable $z + G$ conditioned on the event $\{z + G \geq 0\}$; its density on $[0, \infty)$ is proportional to $e^{zt - t^2/2}$.

Lemma 2.1. (a) $R > 0$, $\bar{R} > 0$, $0 < V < 1$, $R' = -V$, and $\bar{R}' = 1 - V \in (0, 1)$. Hence R is decreasing, $\bar{R}$ is increasing, and $z_- < R(z) \leq z_- + R(0)$, where $R(0) = 2\phi_0$. Moreover

$$V' = R(1 - V - \bar{R}^2) < 0, \quad V'' = -(z + 2R)V' - 2V(1 - V), \quad (2.2)$$

so $\bar{R}$ is strictly convex. Finally $\mathbb{E}T_z = \bar{R}(z)$, $\text{Var } T_z = 1 - V(z)$, and $\mathbb{E}T_z^2 = 1 + z\bar{R}(z) > 0$, and

$$1 - V(z) \geq \frac{1}{3}\bar{R}(z)^2 \quad (z \leq 0). \quad (2.3)$$

(b) For $t > 0$,

$$\frac{t\phi(t)}{1+t^2} < \Phi(-t) < \frac{\phi(t)}{t}.$$

Consequently $t < R(-t) < t + t^{-1}$ and $0 < \bar{R}(-t) < t^{-1}$ for $t > 0$, and $\bar{R}(-t) \geq t^{-1} - 3t^{-3}$ and $0 < 1 - V(-t) < 3t^{-2}$ for $t \geq 2$.

(c) For $t \geq 0$, $-\frac{1}{2}t^2 - \log(1+t) - 2 \leq \log \Phi(-t) \leq -\frac{1}{2}t^2$. For $z \geq 0$, $\log \Phi(z) \geq -2\Phi(-z)$.

(d) $h < 0$ and $h' = R(1 - V) > 0$.

(e) $V^2 \leq 2R$. For $z \geq 0$, $R(z) \leq 2\phi(z)$ and $V(z) \leq 2(1+z)\phi(z)$.

(f) The constant $c_0 := \int_{\mathbb{R}} V(1 - V) dz$ lies in $(0, \infty)$, and

$$\int_{\mathbb{R}} (V - \mathbf{1}_{\{z < 0\}}) dz = 0, \quad \int_{\mathbb{R}} (V^2 - \mathbf{1}_{\{z < 0\}}) dz = -c_0.$$

The function ϱ is continuous and positive, $M_\varrho := \sup \varrho < \infty$, and there is a numerical constant C_g with

$$|V(z)^2 - \mathbf{1}_{\{z < 0\}}| + |\varrho(z) - 2 \cdot \mathbf{1}_{\{z < 0\}}| \leq \frac{C_g}{1 + z^2} \quad (z \in \mathbb{R}).$$

(g) $m(k) = (1 + k^2)\Phi(k) + k\phi(k)$, $m_1(k) = k\Phi(k) + \phi(k) = \Phi(k)\bar{R}(k) > 0$, $m' = 2m_1$, $m'_1 = \Phi$, $m'' = 2\Phi$, and $\Phi - m = -k m_1$.

Most of these facts are classical. The two bounds in (b) are Gordon's bounds for the Mills ratio [65], and the bound $V < 1$ in (a) is due to Sampford [113]. The proof of (2.3) uses Khintchine's representation of unimodal distributions (see [46]). Parts of (a)–(e) are used in Part I, Lemma 2.5 and Lemma 6.1; in Lemma 6.1(v) the constant 2 in the lower bound of (c) is replaced by $-\log \Phi(-1) < 2$.

Proof. (a) Integrating against ϕ on $[-z, \infty)$ and using $\phi'(t) = -t\phi(t)$ gives $\mathbb{E}[G \mid G \geq -z] = R(z)$ and $\mathbb{E}[G^2 \mid G \geq -z] = 1 - zR(z)$, so

$$\mathbb{E}T_z = \bar{R}(z), \quad \text{Var}(T_z) = 1 - zR(z) - R(z)^2 = 1 - V(z).$$

Hence $\bar{R} > 0$, $V = R\bar{R} > 0$, and $V < 1$, since T_z is positive and not almost surely constant; also $\mathbb{E}T_z^2 = 1 - V + \bar{R}^2 = 1 + \bar{R}(\bar{R} - R) = 1 + z\bar{R}$. The identities $R' = -zR - R^2 = -V$ and $\bar{R}' = 1 - V$ follow by differentiation. In particular R is decreasing, so $R(z) \leq R(0) = 2\phi_0$ for $z \geq 0$, while $R(z) - R(0) = \int_z^0 V \leq |z|$ for $z < 0$; and $R > -z$ (from $\bar{R} > 0$) gives $R > z_-$. Next

$$V' = R'\bar{R} + R\bar{R}' = R(1 - V - \bar{R}^2) = R\{\text{Var}(T_z) - (\mathbb{E}T_z)^2\},$$

and differentiating once more, with $V(1 - V - \bar{R}^2) = \bar{R}V'$,

$$V'' = -V(1 - V - \bar{R}^2) + R(-V' - 2\bar{R}(1 - V)) = -(\bar{R} + R)V' - 2V(1 - V),$$

which is (2.2) because $\bar{R} + R = z + 2R$. To see $V' < 0$, it suffices that $\mathbb{E}T_z^2 < 2(\mathbb{E}T_z)^2$. The survival function $H(s) = \mathbb{P}(T_z > s)$ has hazard rate $\mathfrak{m}(s - z)$, where $\mathfrak{m}(s) = \phi(s)/\Phi(-s)$. Since $\mathfrak{m}' = \mathfrak{m}(\mathfrak{m} - s)$ and $\mathfrak{m}(s) > s$ (by (b) for $s > 0$, and because $\mathfrak{m} > 0$ for $s \leq 0$), the hazard rate is strictly increasing. Hence $H(s + t) < H(s)H(t)$ for $s, t > 0$, and integrating over $s, t > 0$ gives

$$\frac{1}{2}\mathbb{E}T_z^2 = \int_0^\infty sH(s) ds < \left(\int_0^\infty H\right)^2 = (\mathbb{E}T_z)^2.$$

Then $\bar{R}'' = -V' > 0$. Finally let $z \leq 0$. The density $p(t) = c e^{zt - t^2/2}$ of T_z on $[0, \infty)$ is C^1 , nonincreasing, and tends to zero at infinity. Hence

$$p(t) = \int_t^\infty (-p'(s)) ds = \int_{(0, \infty)} \frac{\mathbf{1}_{\{t < s\}}}{s} \nu(ds), \quad \nu(ds) := s(-p'(s)) ds \geq 0,$$

and integrating in t (Tonelli) shows that ν is a probability measure. So T_z has the law of SU , where $S \sim \nu$ and U is uniform on $[0, 1]$ and independent of S . Therefore $\mathbb{E}T_z^2 = \frac{1}{3}\mathbb{E}S^2 \geq \frac{1}{3}(\mathbb{E}S)^2 = \frac{4}{3}(\mathbb{E}T_z)^2$, and $\text{Var } T_z \geq \frac{1}{3}(\mathbb{E}T_z)^2$ is (2.3).

(b) For $t > 0$ and $j \geq 0$ set $I_j(t) = \int_t^\infty \phi(s)s^{-j} ds$. Since $\phi'(s) = -s\phi(s)$, integration by parts gives the recurrence

$$I_j(t) = \phi(t)t^{-j-1} - (j+1)I_{j+2}(t), \quad 0 < I_j(t) < \phi(t)t^{-j-1}. \quad (2.4)$$

In particular $\Phi(-t) = \phi(t)/t - I_2(t)$ and $0 < I_2(t) < t^{-2}\Phi(-t)$, which proves Gordon's two Mills-ratio bounds and $t < R(-t) < t + t^{-1}$. Iterating (2.4) gives, for every integer $N \geq 1$,

$$\Phi(-t) = \phi(t) \sum_{j=0}^{N-1} (-1)^j (2j-1)!! t^{-2j-1} + (-1)^N (2N-1)!! I_{2N}(t),$$

where $(-1)!! = 1$. Take $N = 3$ and put $\varepsilon = t^{-2} \leq \frac{1}{4}$. Since $I_6 > 0$, this gives $R(-t) > t(1 - \varepsilon + 3\varepsilon^2)^{-1} \geq t(1 + \varepsilon - 3\varepsilon^2)$, because $(1 - \varepsilon + 3\varepsilon^2)(1 + \varepsilon - 3\varepsilon^2) = 1 - (\varepsilon - 3\varepsilon^2)^2 \leq 1$ and both factors are positive. Thus $\bar{R}(-t) > t^{-1} - 3t^{-3}$. For $t \geq 2$ both lower bounds $R(-t) > t$ and $\bar{R}(-t) > t^{-1} - 3t^{-3}$ are positive, so $V(-t) = R(-t)\bar{R}(-t) > 1 - 3t^{-2}$.

(c) The upper bound is the Chernoff bound $\mathbb{P}(G \geq t) \leq e^{-t^2/2}$. For $t \geq 1$, (b) gives $\Phi(-t) \geq \phi(t)/(2t) \geq \phi(t)/(2(1+t))$, and $\log(2\sqrt{2\pi}) < 2$. For $0 \leq t \leq 1$, $\log \Phi(-t) \geq \log \Phi(-1) > -2$. The last claim follows from $\log(1-p) \geq -2p$ for $0 \leq p \leq \frac{1}{2}$.

(d) By (a), $h' = R + RR' = R(1 - V) > 0$. As $z \rightarrow \infty$, $\log \Phi(z) \rightarrow 0$ and $R(z) \rightarrow 0$, so $h < 0$.

(e) For $z \leq 0$, $\bar{R}(z) \leq \bar{R}(0) = 2\phi_0 < 2$, so $V^2 \leq V = R\bar{R} \leq 2R$. For $z \geq 0$, $\Phi(z) \geq \frac{1}{2}$ gives $R \leq 2\phi$, and $R(z) \leq R(0) < 1$ gives $\bar{R}(z) \leq z + 1$; hence $V = R\bar{R} \leq 2(1+z)\phi(z)$. Moreover $V^2/R = R\bar{R}^2 \leq 2\phi(z)(1+z)^2 \leq 8\phi(1) < 2$, because $(1+z)^2\phi(z)$ is maximal at $z = 1$.

(f) Positivity of c_0 follows from $0 < V < 1$. By (b) and (e), $V(1-V) \leq 1-V \leq 3/z^2$ for $z \leq -2$ and $V(1-V) \leq V \leq 2(1+z)\phi(z)$ for $z \geq 0$, so $c_0 < \infty$; the same bounds show that $V - \mathbf{1}_{\{z < 0\}}$ is integrable. Since $V = -R'$,

$$\int_{-T}^T (V - \mathbf{1}_{\{z < 0\}}) dz = R(-T) - R(T) - T = \bar{R}(-T) - R(T) \rightarrow 0 \quad (T \rightarrow \infty),$$

by (b). Since $V^2 - \mathbf{1}_{\{z < 0\}} = (V - \mathbf{1}_{\{z < 0\}}) - V(1-V)$, its integral is $-c_0$. The function ϱ is continuous because R is. For $z \geq 0$, $\varrho = R^2 > 0$; for $z < 0$, $\varrho = \bar{R}(z)(R(z) - z) > 0$. For the pointwise bound, $1 - V^2 \leq 2(1-V) \leq 6/z^2$ for $z \leq -2$, and $V^2 \leq V \leq 2(1+z)\phi(z)$ for $z \geq 0$. For $z = -t \leq -2$ we have $\varrho(-t) = \bar{R}(-t)(R(-t) + t)$, which by (b) lies between $(\frac{1}{t} - \frac{3}{t^3}) \cdot 2t = 2 - \frac{6}{t^2}$ and $\frac{1}{t}(2t + \frac{1}{t}) = 2 + \frac{1}{t^2}$; for $z \geq 0$, $\varrho = R^2 \leq 4\phi^2$. On $[-2, 0]$ both functions are bounded. These estimates also give $M_\varrho < \infty$.

(g) These are direct computations: $m_1(k) = k\Phi(k) - \mathbb{E}[G; G < k] = k\Phi(k) + \phi(k)$; $m' = 2m_1$ by differentiation under the expectation; $m'_1 = \Phi$; the formula for m has derivative $2m_1$ and vanishes at $-\infty$; and $\Phi - m = -k^2\Phi - k\phi = -k m_1$. $\square$

2.2 Order parameters and the finite-cutoff problem

The hard-wall terminal datum at $q = 1$ is singular. Its usual finite-cutoff representation is imposed at the top $Q < 1$ of the support of the order parameter, where the datum is smooth. This top does not depend continuously on the order parameter in L^1 , because adding a small mass near one moves it. We therefore evaluate every order parameter in a class $\mathcal{U}_r$ with a common terminal time $r < 1$, where the terminal datum is smooth. The resulting solution does not depend on r (Lemma 2.3).

Definition 2.2. (i) $\mathcal{U}$ is the set of right-continuous nondecreasing functions $\gamma : [0, 1] \rightarrow [0, 1]$ with $\gamma = 1$ on a neighborhood of 1. For $\gamma \in \mathcal{U}$ put

$$Q = Q_\gamma = \inf\{q : \gamma(q) = 1\} < 1, \quad \delta = 1 - Q, \quad \lambda(q) = \lambda_\gamma(q) = \int_q^1 \gamma(t) dt.$$

Thus $\lambda' = -\gamma$ almost everywhere, $\lambda(q) = 1 - q$ on $[Q, 1]$, $\gamma(q)(1 - q) \leq \lambda(q) \leq 1 - q$, and $\lambda \geq \delta$ on $[0, Q]$. Let μ_γ be the probability measure on $[0, 1]$ with $\mu_\gamma([0, q]) = \gamma(q)$. Then $\text{supp } \mu_\gamma \subset [0, Q]$ and $Q \in \text{supp } \mu_\gamma$; put $q_0 = \min \text{supp } \mu_\gamma$.

- (ii) For $r \in [0, 1)$ let $\mathcal{U}_r = \{\gamma \in \mathcal{U} : \gamma = 1 \text{ on } [r, 1]\}$. Every $\gamma \in \mathcal{U}_r$ has $Q \leq r$, and $\mathcal{U} = \bigcup_{r < 1} \mathcal{U}_r$. A *step order parameter* is a $\gamma \in \mathcal{U}$ that takes finitely many values.
- (iii) Fix $\kappa \in \mathbb{R}$ and $r \in [0, 1)$, and put $f_r(x) = \log \Phi((x - \kappa)/\sqrt{1 - r})$. For $m \geq 0$ set $\mathcal{T}_{m,0}g = g$, and for $s > 0$ define the Gaussian recursion

$$\mathcal{T}_{m,s}g(x) = \frac{1}{m} \log \mathbb{E} \exp(m g(x + \sqrt{s} G)) \quad (m > 0), \quad \mathcal{T}_{0,s}g(x) = \mathbb{E}g(x + \sqrt{s} G). \quad (2.5)$$

If $r = 0$, then $\mathcal{U}_0 = \{\mathbf{1}_{[0,1]}\}$, and we set $u(0, \cdot) = f_0$. If $r > 0$ and $\gamma \in \mathcal{U}_r$ is a step order parameter, choose $0 = t_0 < t_1 < \dots < t_n = r$ such that $\gamma = m_j$ on $[t_{j-1}, t_j)$, and define $u = u_\gamma(\cdot, \cdot; \kappa)$ on $[0, r] \times \mathbb{R}$ by

$$u(r, \cdot) = f_r, \quad u(q, \cdot) = \mathcal{T}_{m_j, t_j - q} u(t_j, \cdot) \quad (q \in [t_{j-1}, t_j)). \quad (2.6)$$

For a general $\gamma \in \mathcal{U}_r$, $u_\gamma(\cdot, \cdot; \kappa)$ is the limit of $u_{\gamma_n}(\cdot, \cdot; \kappa)$ along any sequence of step order parameters $\gamma_n \in \mathcal{U}_r$ with $\gamma_n \rightarrow \gamma$ in $L^1(0, 1)$. The limit exists and does not depend on the sequence, because the Lipschitz bound of Lemma 2.4(f) is first proved for step order parameters (§A.3), and there it shows that (u_{γ_n}) is a Cauchy sequence. We write $u = u_\gamma(\cdot, \cdot; \kappa)$ when γ and κ are fixed, and

$$b = \partial_x u, \quad c = -\partial_{xx} u, \quad v = \lambda c.$$

The right-continuity of γ gives $\gamma(Q) = 1$, while $\gamma < 1$ on $[0, Q)$; hence every neighborhood of Q has positive μ_γ -mass, which is the assertion $Q \in \text{supp } \mu_\gamma$. The inequality $\lambda(q) \geq \gamma(q)(1 - q)$ holds because γ is nondecreasing.

The expectations in (2.5) are finite for every measurable g that is bounded above and satisfies $g(x) \geq -C_0(1 + x^2)$, and $\mathcal{T}_{m,s}$ preserves these two properties, the lower bound because $\mathcal{T}_{m,s}g \geq \mathcal{T}_{0,s}g$ by Jensen's inequality. For such functions $g \leq g'$, and for constants $C \in \mathbb{R}$ and $K > 0$,

$$\mathcal{T}_{m,s}g \leq \mathcal{T}_{m,s}g', \quad \mathcal{T}_{m,s}(g + C) = \mathcal{T}_{m,s}g + C, \quad \mathcal{T}_{m,s}g = K^{-1} \mathcal{T}_{m/K,s}(Kg). \quad (2.7)$$

The first two properties are immediate from (2.5), and the third follows from $\frac{K}{m} \log \mathbb{E} e^{(m/K)Kg} = \frac{K}{m} \log \mathbb{E} e^{mg}$ if $m > 0$, and from linearity if $m = 0$. Consequently the recursion (2.6) is nondecreasing in its terminal datum and commutes with the addition of constants.

On a block where $\gamma = m > 0$, the recursion (2.6) says that e^{mu} is transported by the backward heat equation. This is the Cole–Hopf transform of the Parisi equation

$$u_q = -\frac{1}{2}(u_{xx} + \gamma u_x^2), \quad (2.8)$$

since for constant $\gamma = m$ the function e^{mu} solves $\partial_q e^{mu} = -\frac{1}{2} \partial_{xx} e^{mu}$ exactly when u solves (2.8). Lemma 2.5 below records the consequences used later. The next lemma shows that the cutoff r is only a device.

Lemma 2.3. *Let $\gamma \in \mathcal{U}_r$ and $r \leq r' < 1$. Then $\gamma \in \mathcal{U}_{r'}$, and the solutions computed with the cutoffs r and r' agree on $[0, r] \times \mathbb{R}$. Consequently $u_\gamma(\cdot, \cdot; \kappa)$ is defined on $[0, 1) \times \mathbb{R}$ for every $\gamma \in \mathcal{U}$, independently of the cutoff, and*

$$u_\gamma(q, x; \kappa) = \log \Phi\left(\frac{x - \kappa}{\sqrt{1 - q}}\right) \quad (Q \leq q < 1). \quad (2.9)$$

Proof. Since $\gamma = 1$ on $[r, r']$, the recursion with cutoff r' performs one step with $m = 1$ on $[r, r']$. With G' an independent standard Gaussian,

$$\mathcal{T}_{1,r'-r}f_{r'}(x) = \log \mathbb{P}(\sqrt{1 - r'} G' \leq x - \kappa + \sqrt{r' - r} G) = \log \Phi\left(\frac{x - \kappa}{\sqrt{1 - r}}\right) = f_r(x).$$

Below r the two recursions coincide, which proves the claim for step order parameters. The general case follows by taking limits in Definition 2.2(iii), since step approximations in $\mathcal{U}_r$ also lie in $\mathcal{U}_{r'}$. For $q \in [Q, 1)$ we have $\gamma \in \mathcal{U}_q$, and the cutoff q gives $u(q, \cdot) = f_q$, which is (2.9). $\square$

In particular $u(Q, x) = \log \Phi((x - \kappa)/\sqrt{\delta})$, which is the usual terminal datum at the top of the support. We use the notation u, b, c, v on $[0, 1) \times \mathbb{R}$ without reference to a cutoff. On $[Q, 1)$, with $z = (x - \kappa)/\sqrt{1 - q}$, formula (2.9) and $\lambda(q) = 1 - q$ give

$$b = \frac{R(z)}{\sqrt{1 - q}}, \quad c = \frac{V(z)}{1 - q}, \quad v = V(z). \quad (2.10)$$

For $\alpha > 0$ and $\kappa \in \mathbb{R}$ the *Parisi functional* and its infimum are

$$\begin{aligned} \mathcal{P}_\kappa(\gamma) &= \alpha u_\gamma(0, 0; \kappa) + \text{CS}(\gamma), \quad \text{CS}(\gamma) = \frac{1}{2} \int_0^Q \frac{dq}{\lambda(q)} + \frac{1}{2} \log \delta, \\ \mathcal{P}_*(\kappa) &= \inf_{\gamma \in \mathcal{U}} \mathcal{P}_\kappa(\gamma). \end{aligned} \quad (2.11)$$

We minimize $\mathcal{P}_\kappa$. When the dependence on α matters we write $\mathcal{P}_\kappa(\gamma; \alpha)$ and $\mathcal{P}_*(\kappa; \alpha)$. Since $\lambda(q) = 1 - q$ on $[Q, 1]$, the Crisanti–Sommers entropy $\text{CS}(\gamma)$ satisfies, for every $r \in [Q, 1]$,

$$\text{CS}(\gamma) = \frac{1}{2} \int_0^r \frac{dq}{\lambda(q)} + \frac{1}{2} \log(1 - r) = \frac{1}{2} \int_0^1 \left(\frac{1}{\lambda(q)} - \frac{1}{1 - q} \right) dq \geq 0, \quad (2.12)$$

where the inequality holds because $\lambda(q) \leq 1 - q$. Thus on $\mathcal{U}_r$ the functional involves only the solution on $[0, r]$ and the entropy term $\frac{1}{2} \int_0^r \lambda^{-1} + \frac{1}{2} \log(1 - r)$, in which $\lambda \geq 1 - r$. For $\gamma = \mathbf{1}_{[0,1]}$ one has $Q = 0$, $\text{CS}(\gamma) = 0$, and $\mathcal{P}_\kappa(\gamma) = \alpha \log \Phi(-\kappa) < 0$; hence $\mathcal{P}_*(\kappa) < 0$ for every κ .

2.3 Regularity, constant blocks, and the diffusion

We first record the properties of u proved in Chapter A. All constants in the following lemma depend only on the cutoff r , on a bound $K_0 \geq |\kappa|$, and on the order of differentiation. They do not depend on γ or, for step order parameters, on the number of steps.

Lemma 2.4. Fix $r \in [0, 1)$ and $K_0 \geq 0$. Let $\gamma \in \mathcal{U}_r$, $|\kappa| \leq K_0$, and $u = u_\gamma(\cdot, \cdot; \kappa)$.

(a) Smoothness. For every $k \geq 0$, $\partial_x^k u$ exists and is continuous on $[0, r] \times \mathbb{R}$, and

$$u(q, x) = f_r(x) + \frac{1}{2} \int_q^r (u_{xx} + \gamma u_x^2)(s, x) ds \quad (q \in [0, r], x \in \mathbb{R}).$$

If γ is constant on an interval $[s, t] \subset [0, r]$, then u is C^∞ on $[s, t] \times \mathbb{R}$, its derivatives from this block extend continuously to $[s, t] \times \mathbb{R}$, and (2.8) holds classically for $s < q < t$, with the corresponding one-sided derivatives at the endpoints.

(b) Growth. $-C(1 + x^2) \leq u \leq 0$, $0 < b \leq C(1 + |x|)$, and $|\partial_x^k u(q, x)| \leq C_k(1 + |x|^{p_k})$ for every $k \geq 1$, where p_k depends only on k .

(c) Signs. $0 < c < 1/\lambda$, that is, $0 < v < 1$; and $c_x < 0$, that is, $v_x < 0$. In particular $u(q, \cdot)$ is increasing and strictly concave.

(d) Behavior at $+\infty$. For $x \geq \kappa$, $0 \geq u(q, x) \geq -C \exp\{-(x - \kappa)^2 / (8(1 - q))\}$.

(e) Behavior at $-\infty$. Put

$$\omega(q) = \log \lambda(q) - \frac{1}{2} \log \delta - \frac{1}{2} \int_q^Q \frac{dt}{\lambda(t)} - \frac{1}{2} \log(2\pi) \quad (0 \leq q < 1), \quad (2.13)$$

where for $q > Q$ the integral is oriented, $\int_q^Q = -\int_Q^q$. There are C, x_∞ such that for $q \in [0, r]$ and $x \leq \kappa - x_\infty$,

$$\left| u(q, x) + \frac{(x - \kappa)^2}{2\lambda(q)} + \log(\kappa - x) - \omega(q) \right| \leq \frac{C}{(\kappa - x)^2},$$

$$\left| \lambda(q)b(q, x) - (\kappa - x) - \frac{\lambda(q)}{\kappa - x} \right| \leq \frac{C}{(\kappa - x)^3}.$$

(f) Continuity in γ . For $\gamma, \tilde{\gamma} \in \mathcal{U}_r$ and $k \geq 0$,

$$|\partial_x^k u_{\tilde{\gamma}}(q, x; \kappa) - \partial_x^k u_{\gamma}(q, x; \kappa)| \leq C_k \|\tilde{\gamma} - \gamma\|_{L^1(0,1)} (1 + |x|^{p_k}) \quad (q \in [0, r]),$$

where one may take $p_0 = 2$ and $p_1 = 1$. Moreover $\mathcal{U}_r$ is compact in $L^1(0, 1)$, and $|\mathcal{P}_\kappa(\tilde{\gamma}; \alpha) - \mathcal{P}_\kappa(\gamma; \alpha)| \leq C(1 + \alpha)\|\tilde{\gamma} - \gamma\|_{L^1}$. In particular $\mathcal{P}_\kappa$ is continuous on $\mathcal{U}_r$, uniformly for $|\kappa| \leq K_0$.

(g) Translation. $u_\gamma(q, x; \kappa) = u_\gamma(q, x - \kappa; 0)$.

Every $\gamma \in \mathcal{U}$ belongs to $\mathcal{U}_r$ for all $r \in [Q, 1)$, and by Lemma 2.3 the solution does not depend on r . Hence (a)–(e) describe u on all of $[0, 1) \times \mathbb{R}$, with constants that are uniform on $[0, r]$. The proofs of (a)–(d), (f), and (g) are in §§A.2 and A.3. The proof of (e) uses Lemmas 2.9 and 2.10 below, whose proofs do not use (e), and is given in §A.6. On $[Q, 1)$ all statements can be read off from (2.9). For instance $\omega(q) = \frac{1}{2} \log(1 - q) - \frac{1}{2} \log(2\pi)$ there, which is the constant in the expansion $\log \Phi(z) = -\frac{1}{2}z^2 - \log|z| - \log \sqrt{2\pi} + O(z^{-2})$ at $z = (x - \kappa)/\sqrt{1 - q} \rightarrow -\infty$. Below Q the constant evolves by $\omega' = (1 - 2\gamma)/(2\lambda)$, which is what (2.8) forces on the expansion in (e). The construction of u for general γ , growth bounds, the non-strict form of the bounds on c in (c), and a local form of the Lipschitz bound (f), for x in bounded sets, are in [54, Propositions 12 and 13, Lemma 14]. The proof of (f) rests on a Feynman–Kac identity for the difference of two solutions (Lemma A.2). This identity and the resulting Lipschitz bound appear in [76, Lemma 14], and the Lipschitz bound goes back to Guerra [67] (see [76, Remark 15]).

Next we record the Cole–Hopf description of u on an interval where γ is constant, and the monotonicity of u in the order parameter. Together they are the form in which Jensen’s inequality enters later chapters.

Lemma 2.5. *Let $\gamma \in \mathcal{U}$ and $\kappa \in \mathbb{R}$, and write $u = u_\gamma(\cdot, \cdot; \kappa)$.*

- (a) *Constant blocks. If $0 \leq s < t < 1$ and $\gamma = m$ on $[s, t)$, then $\lambda(q) = \lambda(t) + m(t - q)$ and $u(q, \cdot) = \mathcal{T}_{m, t-q} u(t, \cdot)$ for $q \in [s, t]$. For $m > 0$ this reads $e^{mu(q, x)} = \mathbb{E} e^{mu(t, x + \sqrt{t-q}G)}$, so e^{mu} solves the backward heat equation on $[s, t] \times \mathbb{R}$.*
- (b) *Comparison. If $\tilde{\gamma} \in \mathcal{U}$, $q' \in [0, 1)$ and $\tilde{\gamma} \leq \gamma$ on $[q', 1]$, then $u_{\tilde{\gamma}}(q', \cdot; \kappa) \leq u_\gamma(q', \cdot; \kappa)$. In particular $u_\gamma(q', \cdot; \kappa)$ depends only on the restriction of γ to $[q', 1]$.*
- (c) *Heat lower bound. For $0 \leq q \leq t < 1$ and $x \in \mathbb{R}$, $u(q, x) \geq \mathbb{E} u(t, x + \sqrt{t-q}G)$.*

Proof. (a) For step order parameters this is the definition (2.6), after refining the grid so that s and t are grid points. For general γ it is obtained in §A.3 by passing to the limit along step order parameters equal to m on $[s, t)$. The formula for λ follows from $\lambda' = -m$ on $[s, t)$.

(b) Choose $r \in (q', 1)$ with $\gamma, \tilde{\gamma} \in \mathcal{U}_r$. Lemma A.2, applied with $\xi = \tilde{\gamma}$ and $\tilde{\xi} = \gamma$, gives

$$u_\gamma(q', x; \kappa) - u_{\tilde{\gamma}}(q', x; \kappa) = \frac{1}{2} \mathbb{E} \int_{q'}^r (\gamma - \tilde{\gamma})(t) b_{\tilde{\gamma}}(t, Y_t)^2 dt \geq 0,$$

where Y is the diffusion of that lemma started at (q', x) . If $\tilde{\gamma} = \gamma$ on $[q', 1]$, the inequality holds in both directions.

(c) We may assume $q < t$. Let $\tilde{\gamma} = \gamma \mathbf{1}_{[t, 1]}$. Then $\tilde{\gamma} \in \mathcal{U}$, $\tilde{\gamma} \leq \gamma$, and $\tilde{\gamma} = \gamma$ on $[t, 1]$, so $u_{\tilde{\gamma}}(t, \cdot) = u(t, \cdot)$ by (b). By (b), and by (a) applied to $\tilde{\gamma}$, which vanishes on $[q, t)$,

$$u(q, x) \geq u_{\tilde{\gamma}}(q, x) = \mathcal{T}_{0, t-q} u(t, \cdot)(x) = \mathbb{E} u(t, x + \sqrt{t-q} G).$$

This proves (c). □

We now introduce the diffusion along which all first-order quantities are evaluated. By Lemma 2.4(b),(c), on $[0, r]$ the drift $\gamma(q)b(q, x)$ is Lipschitz in x with constant $(1-r)^{-1}$, since $\partial_x(\gamma b) = -\gamma c$ and $0 < c < 1/\lambda \leq (1-r)^{-1}$, and it has linear growth. Hence for every $\gamma \in \mathcal{U}$ the equation

$$dX_q = \gamma(q) b(q, X_q) dq + dB_q, \quad X_0 = 0, \quad (2.14)$$

driven by a standard Brownian motion B , has a unique strong solution on $[0, 1)$. We call $X = X^{\gamma, \kappa}$ the *Parisi diffusion* [7, Theorem 3 and (17)]; for this functional see [54, (2.8) and (4.13)]. For $(s, x) \in [0, 1) \times \mathbb{R}$ we write $X^{s, x}$ for the solution on $[s, 1)$ started from $X_s^{s, x} = x$. The generator of X acting on smooth functions is

$$\mathcal{L} = \partial_q + \frac{1}{2} \partial_{xx} + \gamma b \partial_x.$$

Since $b > 0$, the drift is nonnegative, and therefore

$$X_q \geq B_q \quad \text{and} \quad X_q^{s, x} \geq x + B_q - B_s \quad (q \geq s). \quad (2.15)$$

For a general γ the function u is not differentiable in q at the discontinuities of γ , so Itô's formula along X is used in the integrated form below. Let $\mathcal{D}$ be the set of functions of the form

$$g(q, x) = \sum_i \pi_i(q, x, \lambda(q), \lambda(q)^{-1}) \prod_j \partial_x^{k_{ij}} u(q, x),$$

finite sums with polynomials π_i and integers $k_{ij} \geq 0$. For $g \in \mathcal{D}$ we define $\mathcal{L}g$ formally: we compute ∂_q of each factor with the rules $\partial_q \partial_x^k u = -\frac{1}{2} \partial_x^{k+2} u - \frac{1}{2} \gamma \partial_x^k (u_x^2)$ (the

x -derivatives of (2.8)) and $\lambda' = -\gamma$, and add $\frac{1}{2}g_{xx} + \gamma b g_x$. Since these rules are affine in γ , we can write $\mathcal{L}g = \mathcal{L}_0g + \gamma \mathcal{L}_1g$ with $\mathcal{L}_0g, \mathcal{L}_1g \in \mathcal{D}$ not involving γ . Where γ is constant, $\mathcal{L}g$ is the classical generator applied to g . The formal time derivative is a derivation on $\mathcal{D}$, and it commutes with ∂_x , because the rule for $\partial_q \partial_x^{k+1}u$ is the x -derivative of the rule for $\partial_q \partial_x^k u$. Hence the formal generator obeys the product rule and the commutation rule

$$\mathcal{L}(fg) = f \mathcal{L}g + g \mathcal{L}f + f_x g_x, \quad \mathcal{L}(g_x) = \partial_x(\mathcal{L}g) + \gamma c g_x \quad (f, g \in \mathcal{D}), \quad (2.16)$$

where the second identity uses $\partial_x(\gamma b g_x) = \gamma b g_{xx} - \gamma c g_x$.

Lemma 2.6. Fix $r \in [0, 1)$, $K_0 \geq 0$, and $p \geq 1$; let $\gamma \in \mathcal{U}_r$ and $|\kappa| \leq K_0$.

- (a) $\mathbb{E} \sup_{s \leq q \leq r} |X_q^{s,x}|^p \leq C(1 + |x|^p)$, with $C = C(r, K_0, p)$.
- (b) If $\tilde{\gamma} \in \mathcal{U}_r$ and $\tilde{X}^{s,x}$ is the diffusion of $\tilde{\gamma}$ driven by the same Brownian motion, then $\mathbb{E} \sup_{s \leq q \leq r} |\tilde{X}_q^{s,x} - X_q^{s,x}|^p \leq C(1 + |x|^p) \|\tilde{\gamma} - \gamma\|_{L^1}^p$.
- (c) Itô's formula along X . For $g \in \mathcal{D}$ and $s \leq t \leq r$,

$$\mathbb{E} g(t, X_t^{s,x}) - g(s, x) = \int_s^t \mathbb{E} [\mathcal{L}_0g + \gamma(q) \mathcal{L}_1g](q, X_q^{s,x}) dq,$$

and $q \mapsto \mathbb{E} h(q, X_q^{s,x})$ is continuous on $[s, r]$ for every $h \in \mathcal{D}$. Consequently $e_g(q) = \mathbb{E} g(q, X_q)$ is Lipschitz on $[0, r]$; its right derivative at q is $\mathbb{E} [\mathcal{L}_0g + \gamma(q) \mathcal{L}_1g](q, X_q)$ and its left derivative is $\mathbb{E} [\mathcal{L}_0g + \gamma(q-) \mathcal{L}_1g](q, X_q)$. If γ is constant on $[s', t']$, then e_g is C^∞ on $[s', t']$ (with one-sided derivatives at the endpoints), and its derivatives are obtained by iterating $\mathcal{L}$ with γ frozen at its value on $[s', t']$.

The proof is in §A.4. In particular, if $\mathcal{L}g = 0$, then $g(q, X_q)$ is a martingale on $[0, 1)$: applying (c) to the diffusion started from (s, X_s) and using the Markov property gives $\mathbb{E}[g(t, X_t) \mid X_s] = g(s, X_s)$ for $s \leq t < 1$.

The first-order quantities of §§2.4 and 2.5 are expectations over the law of X_q . The following structural fact about this law is used for lower bounds on densities in Chapter 5 and for the control of the top of the support in Chapter 6. At $q = 0$ the law of X_0 is a point mass, so the statement concerns $q > 0$ only.

Lemma 2.7. Let $\gamma \in \mathcal{U}$, $\kappa \in \mathbb{R}$, and $q \in (0, 1)$. The law of X_q has a density p_q of the form

$$p_q(x) = e^{\gamma(q)u(q,x)} \phi_q(x) g_q(x), \quad (2.17)$$

where $g_q : \mathbb{R} \rightarrow [0, \infty)$ is nonincreasing. Moreover $g_q \geq 1$ and $e^{\gamma(q)u(q,x)} \geq \Phi((x - \kappa)/\sqrt{1-q})$. Consequently

$$p_q(x) \geq \Phi\left(\frac{x - \kappa}{\sqrt{1-q}}\right) \phi_q(x) \quad (x \in \mathbb{R}). \quad (2.18)$$

Proof. We divide the proof into three steps. First we compute the density when γ is piecewise constant below q , then we approximate, and finally we prove the two lower bounds.

STEP 1. γ piecewise constant on $[0, q]$. Suppose that $\gamma = m_j$ on $[t_{j-1}, t_j]$ for $0 = t_0 < \dots < t_n = q$, with γ arbitrary on $[q, 1]$. On a block $[t_{j-1}, t_j]$ with $m_j > 0$, the function $H = e^{m_j u}$ is positive, bounded by one, and solves $H_q + \frac{1}{2}H_{xx} = 0$ by Lemma 2.5(a). The drift of X on this block is $m_j b = H_x/H$. Thus X is the Doob transform [50, 110] of Brownian motion by H on this block: for $t_{j-1} \leq s < t \leq t_j$,

$$\mathbb{P}(X_t \in dy \mid X_s = x) = \phi_{t-s}(y-x) \frac{H(t, y)}{H(s, x)} dy. \quad (2.19)$$

To verify (2.19), fix a bounded continuous F and put $W(s', x) = \mathbb{E}[(FH)(t, x + B_t - B_{s'})]/H(s', x)$ for $s' \in [s, t]$. The numerator solves the backward heat equation, and dividing by H gives $W_{s'} + \frac{1}{2}W_{xx} + (H_x/H)W_x = 0$. Moreover $|W| \leq \sup |F|$, because $H(s', x) = \mathbb{E}H(t, x + B_t - B_{s'})$, and $W(s', \cdot) \rightarrow F$ locally uniformly as $s' \uparrow t$. Itô's formula shows that $W(s', X_{s'})$ is a bounded martingale on $[s, t]$, whence $\mathbb{E}[F(X_t) \mid X_s = x] = W(s, x)$ by bounded convergence. If $m_j = 0$, (2.19) holds with $H \equiv 1$.

By the Markov property and (2.19), the vector $(X_{t_1}, \dots, X_{t_n})$ has the density

$$\prod_{j=1}^n \phi_{t_j - t_{j-1}}(x_j - x_{j-1}) e^{m_j(u_j - u_{j-1})}, \quad x_0 = 0, \quad u_j = u(t_j, x_j).$$

The product of the Gaussian factors is the density of $(B_{t_1}, \dots, B_{t_n})$. Put $m_{n+1} = \gamma(q)$. Summation by parts gives

$$\sum_{j=1}^n m_j(u_j - u_{j-1}) = \gamma(q) u_n - m_1 u(0, 0) + \sum_{j=1}^n (m_{j+1} - m_j)(-u_j).$$

Given $B_q = x$, the vector $(B_{t_1}, \dots, B_{t_n})$ has the law of $(\beta_1^x, \dots, \beta_n^x)$, where

$$\beta_j^x = \frac{t_j}{q} x + \left(B_{t_j} - \frac{t_j}{q} B_q\right)$$

is the Brownian bridge representation; in particular $\beta_n^x = x$, and the bridge $B_{t_j} - \frac{t_j}{q} B_q$ does not depend on x . Integrating out $x_1, \dots, x_{n-1}$ therefore gives (2.17) with

$$g_q(x) = e^{-m_1 u(0, 0)} \mathbb{E} \exp \left(\sum_{j=1}^n (m_{j+1} - m_j)(-u(t_j, \beta_j^x)) \right).$$

Since $m_{j+1} \geq m_j$, $u(t_j, \cdot)$ is increasing by Lemma 2.4(c), and β_j^x is nondecreasing in x , every summand in the exponent is nonincreasing in x . Hence g_q is nonincreasing.

STEP 2. *General γ .* For partitions $0 = t_0 < \dots < t_n = q$ of $[0, q]$ with mesh tending to zero, let $\gamma_n(t) = \gamma(t_j)$ on $[t_j, t_{j+1})$ and $\gamma_n = \gamma$ on $[q, 1]$. Then $\gamma_n \in \mathcal{U}$ and $\|\gamma_n - \gamma\|_{L^1} \rightarrow 0$. Moreover $\gamma_n(q) = \gamma(q)$ and, by Lemma 2.5(b), $u_{\gamma_n}(q, \cdot) = u(q, \cdot)$. By Step 1 the law of X_q^n (the diffusion of γ_n at time q) has density $\mathfrak{p}g_n$ with $\mathfrak{p} = e^{\gamma(q)u(q, \cdot)}\phi_q$ and g_n nonincreasing, and by Lemma 2.6(b), $X_q^n \rightarrow X_q$ in L^2 . Put $\mathfrak{P}(x) = \int_{-\infty}^x \mathfrak{p} > 0$. Monotonicity gives $g_n(x)\mathfrak{P}(x) \leq \int_{-\infty}^x \mathfrak{p}g_n \leq 1$, so $\sup_n g_n \leq 1/\mathfrak{P}(x_0)$ on $[x_0, \infty)$. By Helly's selection theorem and a diagonal argument, a subsequence converges pointwise on $\mathbb{R}$ to a nonincreasing function $g \geq 0$. For continuous F with compact support in (x_0, ∞) , dominated convergence gives $\mathbb{E}F(X_q^n) = \int F\mathfrak{p}g_n \rightarrow \int F\mathfrak{p}g$, while $\mathbb{E}F(X_q^n) \rightarrow \mathbb{E}F(X_q)$. Hence the law of X_q coincides with $\mathfrak{p}g dx$ on every half-line (x_0, ∞) , and therefore on $\mathbb{R}$.

STEP 3. *The lower bounds.* Since $u \leq 0$ and g_q is nonincreasing, for every x

$$\mathbb{P}(X_q > x) = \int_x^\infty e^{\gamma(q)u} \phi_q g_q \leq g_q(x) \mathbb{P}(B_q > x),$$

while $\mathbb{P}(X_q > x) \geq \mathbb{P}(B_q > x)$ by (2.15). Hence $g_q \geq 1$. For the second bound put $m = \gamma(q)$. For $q \geq Q$ equality holds by (2.9), so let $q < Q$. The order parameter $\tilde{\gamma} = \gamma \mathbf{1}_{[0, q)} + m \mathbf{1}_{[q, Q)} + \mathbf{1}_{[Q, 1]}$ belongs to $\mathcal{U}$, satisfies $\tilde{\gamma} \leq \gamma$ on $[q, 1]$, and equals m on $[q, Q)$ and 1 on $[Q, 1]$. Put $Y = x + \sqrt{Q - q} G$. By Lemma 2.5(b),(a) and (2.9), and since $0 \leq m \leq 1$ and $\Phi \leq 1$,

$$e^{mu(q, x)} \geq e^{mu_{\tilde{\gamma}}(q, x)} = \mathbb{E} \Phi\left(\frac{Y - \kappa}{\sqrt{\delta}}\right)^m \geq \mathbb{E} \Phi\left(\frac{Y - \kappa}{\sqrt{\delta}}\right) = \Phi\left(\frac{x - \kappa}{\sqrt{1 - q}}\right).$$

For $m = 0$ the first three expressions equal one. Combining with $g_q \geq 1$ gives (2.18). $\square$

In the Brownian-bridge representation of [87, Proposition 7.1] for the SK model, the terminal datum is even and the transformed density is even and log-concave. Here the solution is increasing in x , and the conclusion is that g_q is nonincreasing, which gives the density floor (2.18).

Finally, the Parisi equation is the Hamilton–Jacobi–Bellman equation of a stochastic control problem [38, 76]. We use this representation for step coefficients and for a class of terminal data that contains f_r and the quadratic penalty $-\frac{1}{2}(k - x)_+^2$ of the zero-temperature functional. Let $\mathfrak{F}$ be the set of $f \in C^1(\mathbb{R})$ such that $f \leq 0$, f is nondecreasing and concave, f' is Lipschitz, and $f(x) \geq -C_f(1 + x^2)$ for a constant C_f . For $T > 0$, a nondecreasing right-continuous step function $\xi : [0, T) \rightarrow [0, \infty)$ and $f \in \mathfrak{F}$, let $Y = Y_{\xi, f}$ be defined on $[0, T] \times \mathbb{R}$ by (2.6) with r replaced by T , f_r by f , and γ by ξ . A control is a progressively measurable process θ on $[0, T]$ with $\theta = 0$ almost everywhere on $\{\xi = 0\}$ and $\mathbb{E} \int_0^T \theta_q^2 / \xi(q) dq < \infty$, with the convention $0/0 = 0$.

Lemma 2.8. *In this setting, Y and Y_x are continuous on $[0, T] \times \mathbb{R}$, Y is smooth and satisfies $Y_q = -\frac{1}{2}(Y_{xx} + \xi Y_x^2)$ on each open interval of constancy of ξ , and*

$$\begin{aligned} \mathbb{E}f(x + \sqrt{T-q}G) &\leq Y(q, x) \leq 0, & 0 \leq Y_x(q, x) &\leq C(1 + |x|), \\ 0 &\leq -Y_{xx} \leq \text{Lip}(f'). \end{aligned}$$

These bounds hold for $0 \leq q < T$; the first two also hold at $q = T$, while the curvature bound at T is understood almost everywhere in x . For every control θ and $x \in \mathbb{R}$,

$$Y(0, x) \geq \mathbb{E}f\left(x + B_T + \int_0^T \theta_q \, dq\right) - \frac{1}{2}\mathbb{E} \int_0^T \frac{\theta_q^2}{\xi(q)} \, dq, \quad (2.20)$$

with equality for the feedback control $\theta_q^* = \xi(q)Y_x(q, X_q^*)$, where X^* is the unique strong solution of $dX_q^* = \xi(q)Y_x(q, X_q^*) \, dq + dB_q$, $X_0^* = x$. Moreover $Y_x(q, X_q^*)$ is a square-integrable martingale on $[0, T]$ with terminal value $f'(X_T^*)$.

The feedback control θ^* need not be bounded. It is square integrable because Y_x has linear growth and X^* has moments of all orders. The proof is in §A.5. For a step order parameter $\gamma \in \mathcal{U}_r$ with $r > 0$, the lemma applies with $T = r$, $\xi = \gamma$, and $f = f_r$, and then $Y = u_\gamma(\cdot, \cdot; \kappa)$ and $X^* = X$ when $x = 0$. For the SK and mixed p -spin models, Auffinger and W.-K. Chen proved the control representation with the Boué–Dupuis formula [7, Theorem 3], [22], and Jagannath and Tobasco proved it by dynamic programming [76, Lemma 18].

2.4 Generator identities and the sandwich inequality

The first-order optimality conditions are expressed through expectations of derivatives of u along X . We collect the identities for these expectations here. Besides b, c, v we use the function

$$\psi(q, x) = \lambda(q) b(q, x) + x - \kappa, \quad (2.21)$$

which is harmonic for $\mathcal{L}$, and the scalar functions of $q \in [0, 1)$

$$\begin{aligned} D(q) &= \alpha \mathbb{E}b(q, X_q)^2 - \int_0^q \frac{dt}{\lambda(t)^2}, & S(q) &= \alpha \mathbb{E}v(q, X_q)^2, & \chi(q) &= \frac{\alpha \mathbb{E}v(q, X_q) - 1}{\lambda(q)}, \\ w(q) &= \alpha \mathbb{E}[v(1-v)](q, X_q), & \mathcal{N}(q) &= \mathbb{E}v_x(q, X_q)^2, & \tilde{B}(q) &= \mathbb{E}[v^2(1-v)](q, X_q). \end{aligned} \quad (2.22)$$

Thus $S = \alpha \lambda^2 \mathbb{E}c(q, X_q)^2$, and $\mathcal{N} > 0$ and $\tilde{B} > 0$ by Lemma 2.4(c). For $K \in \mathbb{R}$ put

$$\Sigma(q; K) = \mathbb{E}\left[v_{xx}^2 + 4K(3v-1)v_x^2 + 2K^2v^2(1-v)(1-3v)\right](q, X_q), \quad \tilde{x}(q) = \frac{\lambda(q)\mathcal{N}(q)}{2\tilde{B}(q)}. \quad (2.23)$$

Lemma 2.9. Let $\gamma \in \mathcal{U}$ and $\kappa \in \mathbb{R}$.

(a) With $K = \gamma/\lambda$, in the sense of the formal generator of Lemma 2.6 (and classically where γ is constant),

$$\begin{aligned} \mathcal{L}u &= \frac{1}{2}\gamma b^2, & \mathcal{L}b &= 0, & \mathcal{L}c &= \gamma c^2, & \mathcal{L}c_x &= 3\gamma c c_x, & \mathcal{L}\psi &= 0, \\ \mathcal{L}v &= -Kv(1-v), & \mathcal{L}v_x &= K(3v-1)v_x, & \mathcal{L}v_{xx} &= K(3v_x^2 + (4v-1)v_{xx}). \end{aligned}$$

(b) $\psi_x = 1 - v \in (0, 1)$ and $\psi_{xx} = -v_x > 0$, so $\psi(q, \cdot)$ is increasing and strictly convex for every $q \in [0, 1]$. The processes $b(q, X_q)$ and $\psi(q, X_q)$ are martingales on $[0, 1]$: for $0 \leq s \leq t < 1$, $b(s, x) = \mathbb{E}b(t, X_t^{s,x})$ and $\psi(s, x) = \mathbb{E}\psi(t, X_t^{s,x})$. For $s \leq Q$ this gives

$$\psi(s, x) = \mathbb{E}\psi(Q, X_Q^{s,x}) = \sqrt{\delta} \mathbb{E} \bar{R}\left(\frac{X_Q^{s,x} - \kappa}{\sqrt{\delta}}\right) \geq \sqrt{\delta} \bar{R}\left(\frac{\sqrt{\delta}(x - \kappa)}{\lambda(s)}\right) > 0, \quad (2.24)$$

and $b(0, 0) = \delta^{-1/2} \mathbb{E}R(Z)$ with $Z = (X_Q - \kappa)/\sqrt{\delta}$. On $[Q, 1]$, $\psi = \sqrt{1-q} \bar{R}((x - \kappa)/\sqrt{1-q}) > 0$.

(c) The functions $\mathbb{E}b(q, X_q)^2$, S , $\mathbb{E}v(q, X_q)$, χ , w , $\mathcal{N}$, and $\tilde{B}$ are Lipschitz on $[0, r]$ for every $r < 1$, and the following hold.

(i) $\frac{d}{dq} \mathbb{E}b(q, X_q)^2 = \mathbb{E}c(q, X_q)^2$, a continuous function. Hence $D \in C^1([0, 1])$ and $D' = (S - 1)/\lambda^2$.

(ii) $S/\lambda^2 = \alpha \mathbb{E}c(q, X_q)^2$ is nondecreasing on $[0, 1]$.

(iii) $q \mapsto \mathbb{E}v(q, X_q)$ is absolutely continuous with $\frac{d}{dq} \mathbb{E}v(q, X_q) = -\frac{\gamma(q)}{\lambda(q)} \mathbb{E}[v(1-v)](q, X_q)$ for almost every q , and $\chi' = \gamma D'$ almost everywhere.

(iv) At every $q \in [0, 1)$ the functions S , $\mathcal{N}$, and $\tilde{B}$ have right derivatives

$$\begin{aligned} S'(q+) &= \alpha(\mathcal{N}(q) - 2K\tilde{B}(q)), & \mathcal{N}'(q+) &= \mathbb{E}[v_{xx}^2 + 2K(3v-1)v_x^2](q, X_q), \\ \tilde{B}'(q+) &= \mathbb{E}[(1-3v)v_x^2 - Kv^2(1-v)(2-3v)](q, X_q), \end{aligned}$$

with $K = \gamma(q)/\lambda(q)$, and left derivatives given by the same formulas with $K = \gamma(q-)/\lambda(q)$.

(v) $S - 1 = \lambda\chi - w$.

(d) If $\gamma = m$ on $[s, t]$ with $0 \leq s < t < 1$, then $S \in C^\infty([s, t])$ with one-sided derivatives at the endpoints, and with $K = m/\lambda(q)$,

$$S' = \alpha(\mathcal{N} - 2K\tilde{B}) = \frac{2\alpha\tilde{B}}{\lambda}(\tilde{x} - m), \quad S'' = \alpha\Sigma(q; K).$$

Proof. (a) The function u belongs to $\mathcal{D}$, and (2.8) gives $\mathcal{L}u = u_q + \frac{1}{2}u_{xx} + \gamma bu_x = \frac{1}{2}\gamma b^2$. By the commutation rule in (2.16),

$$\mathcal{L}b = \partial_x(\frac{1}{2}\gamma b^2) + \gamma cb = \gamma bb_x + \gamma bc = 0,$$

because $b_x = -c$. Next $\mathcal{L}c = -\mathcal{L}(b_x) = -\partial_x(\mathcal{L}b) - \gamma cb_x = \gamma c^2$, and $\mathcal{L}c_x = \partial_x(\gamma c^2) + \gamma c c_x = 3\gamma c c_x$. Since $\lambda' = -\gamma$, $\mathcal{L}v = -\gamma c + \lambda \gamma c^2 = -Kv(1-v)$. The identities for v_x and v_{xx} follow from (2.16) and $\gamma c = Kv$:

$$\mathcal{L}v_x = -K(1-2v)v_x + Kv v_x = K(3v-1)v_x, \quad \mathcal{L}v_{xx} = K(3v_x^2 + (3v-1)v_{xx}) + Kv v_{xx}.$$

Finally $\mathcal{L}(x - \kappa) = \gamma b$, so $\mathcal{L}\psi = -\gamma b + \lambda \mathcal{L}b + \gamma b = 0$. These are identities between formal expressions affine in γ , that is, they hold separately for the parts $\mathcal{L}_0$ and $\mathcal{L}_1$; they are valid classically where γ is constant.

(b) $\psi_x = \lambda b_x + 1 = 1 - v$ and $\psi_{xx} = -v_x$; the signs come from Lemma 2.4(c). Since b and ψ belong to $\mathcal{D}$ and have zero formal generator, Lemma 2.6(c) applied to the diffusion started at (s, x) gives the martingale identities. At $q = Q$, (2.10) gives $\psi(Q, x) = \sqrt{\delta} \bar{R}((x - \kappa)/\sqrt{\delta})$ and $b(Q, x) = R((x - \kappa)/\sqrt{\delta})/\sqrt{\delta}$, which yields the equality in (2.24) and the formula for $b(0, 0)$. In particular $\psi > 0$, since $\bar{R} > 0$. For the lower bound, let $M_q = (X_q^{s,x} - \kappa)/\lambda(q)$ for $q \in [s, Q]$. Itô's formula and $\lambda' = -\gamma$ give

$$dM_q = \frac{\gamma(q)}{\lambda(q)^2} \psi(q, X_q^{s,x}) dq + \frac{dB_q}{\lambda(q)}. \quad (2.25)$$

The drift is nonnegative because $\psi > 0$, and the martingale part has a deterministic bounded integrand. Hence $\mathbb{E}M_Q \geq M_s = (x - \kappa)/\lambda(s)$. Since $\lambda(Q) = \delta$, we have $(X_Q^{s,x} - \kappa)/\sqrt{\delta} = \sqrt{\delta}M_Q$, and Jensen's inequality for the convex increasing function $\bar{R}$ gives the lower bound. The formula on $[Q, 1)$ is (2.10).

(c) By Lemma 2.6(c) each of the listed functions is Lipschitz on $[0, r]$; for χ also use $\lambda \geq 1 - r$. By (2.16), $\mathcal{L}(b^2) = 2b\mathcal{L}b + b_x^2 = c^2$ contains no factor γ , so $\mathbb{E}b^2$ is C^1 with derivative $\mathbb{E}c^2$, which is continuous by Lemma 2.6(c). As $S = \alpha\lambda^2\mathbb{E}c^2$, this gives $D' = \alpha\mathbb{E}c^2 - \lambda^{-2} = (S - 1)/\lambda^2$, which is (i). Next $\mathcal{L}(c^2) = 2c\mathcal{L}c + c_x^2 = 2\gamma c^3 + c_x^2 \geq 0$, so the Lipschitz function $\mathbb{E}c^2$ has a nonnegative right derivative everywhere and is nondecreasing; this is (ii). By (a), $\mathcal{L}v = -Kv(1-v)$, which gives the derivative of $\mathbb{E}v$ at every continuity point of γ , hence almost everywhere. Then, using $(1/\lambda)' = \gamma/\lambda^2$,

$$\chi' = \frac{\alpha \frac{d}{dq} \mathbb{E}v}{\lambda} + \frac{\gamma(\alpha \mathbb{E}v - 1)}{\lambda^2} = \frac{\gamma}{\lambda^2} (-\alpha \mathbb{E}[v(1-v)] + \alpha \mathbb{E}v - 1) = \frac{\gamma(S - 1)}{\lambda^2} = \gamma D'$$

almost everywhere, which is (iii). For (iv), (a) and (2.16) give

$$\begin{aligned} \mathcal{L}(v^2) &= v_x^2 - 2Kv^2(1-v), & \mathcal{L}(v_x^2) &= v_{xx}^2 + 2K(3v-1)v_x^2, \\ \mathcal{L}(v^3) &= 3v v_x^2 - 3Kv^3(1-v), \end{aligned}$$

and $\tilde{B} = \mathbb{E}v^2 - \mathbb{E}v^3$. Since each of these expressions is affine in γ through K , Lemma 2.6(c) gives the one-sided derivatives. Identity (v) is $\lambda\chi - w = \alpha\mathbb{E}v - 1 - \alpha\mathbb{E}v + \alpha\mathbb{E}v^2$.

(d) On $[s, t]$ we have $\lambda' = -m$, so $K' = K^2$. By Lemma 2.6(c) the functions S , $\mathcal{N}$, and $\tilde{B}$ are C^∞ on $[s, t]$, and their derivatives are given by (c)(iv) with $\gamma = m$. The first formula for

S' is (c)(iv), and the second is its rewriting with (2.23). Differentiating $S'/\alpha = \mathcal{N} - 2K\tilde{B}$ and using $K' = K^2$,

$$\begin{aligned} \frac{S''}{\alpha} &= \mathbb{E}v_{xx}^2 + 2K\mathbb{E}[(3v-1)v_x^2] - 2K^2\tilde{B} \\ &\quad - 2K\mathbb{E}[(1-3v)v_x^2] + 2K^2\mathbb{E}[v^2(1-v)(2-3v)] = \Sigma(q; K), \end{aligned}$$

since $2-3v-1=1-3v$. This proves (d). $\square$

The identities $\mathcal{L}b = 0$ and $\mathcal{L}c = \gamma c^2$ in (a) are the hard-wall forms of [7, Lemma 2], and the one-sided derivatives in (c)(iv) are the analogue of [8, Proposition 3]. The process $X_q + \lambda(q)b(q, X_q)$ is the conditional expectation of the terminal position in [54, (3.19)], and (b) extends the positivity $\psi(q, X_q) \geq 0$ at levels $q \geq Q$ of [54, Lemma 10] to all levels.

The next inequality compares $b\psi$ with v . It is the main pointwise tool for the rescaled analysis of later chapters: it controls v , b , and ψ far from κ in terms of the single parameter $1-q$.

Lemma 2.10. *Let $\gamma \in \mathcal{U}$ and $\kappa \in \mathbb{R}$, and write $\tau = 1-q$.*

(a) *For $(q, x) \in [0, 1) \times \mathbb{R}$,*

$$v \leq b\psi \leq \frac{\tau}{\lambda} v. \quad (2.26)$$

The upper inequality is equivalent to $F := \tau c - \lambda b^2 - (x - \kappa)b = \tau c - b\psi \geq 0$. More precisely, for $q \leq Q$,

$$\begin{aligned} b\psi - v &= \mathbb{E} \int_q^Q [(1-\gamma)c(1-v)](t, X_t^{q,x}) dt, \\ F(q, x) &= \mathbb{E} \int_q^Q [(\lambda - \gamma\tau)c^2](t, X_t^{q,x}) dt, \end{aligned}$$

and both $b\psi - v$ and F vanish on $[Q, 1)$. The function F satisfies $\mathcal{L}F = (\gamma\tau - \lambda)c^2 \leq 0$.

(b) *For $q \in [0, 1)$: $0 < \psi(q, x) \leq \tau/(\kappa - x)$ for $x < \kappa$; $0 < \lambda b(q, x) \leq \tau/(x - \kappa)$ for $x > \kappa$; $\psi(q, \kappa) = \lambda b(q, \kappa) \leq \sqrt{\tau}$, and $\psi(q, \kappa) \geq 2\phi_0\sqrt{\delta}$ if $q \leq Q$;*

$$|v(q, x) - \mathbf{1}_{\{x < \kappa\}}| \leq \min \left\{ 1, \frac{4\tau}{(x - \kappa)^2} \right\};$$

and

$$\int_{-\infty}^{\kappa} (1-v)(q, x) dx = \int_{\kappa}^{\infty} v(q, x) dx = \psi(q, \kappa), \quad \int_{\mathbb{R}} (v(q, x) - \mathbf{1}_{\{x < \kappa\}}) dx = 0.$$

Proof. (a) The product rule (2.16) and Lemma 2.9(a) give $\mathcal{L}(b\psi) = b_x\psi_x = -c(1-v)$, hence

$$\mathcal{L}(v - b\psi) = -\frac{\gamma}{\lambda}v(1-v) + c(1-v) = (1-\gamma)c(1-v),$$

because $\frac{\gamma}{\lambda}v = \gamma c$. Similarly, $\tau' = -1$ and $\lambda' = -\gamma$ give

$$\mathcal{L}(\tau c) = -c + \gamma \tau c^2, \quad \mathcal{L}(\lambda b^2) = -\gamma b^2 + \lambda c^2, \quad \mathcal{L}((x - \kappa)b) = \gamma b^2 - c,$$

so $\mathcal{L}F = (\gamma\tau - \lambda)c^2$, which is nonpositive because $\lambda(q) = \int_q^1 \gamma \geq \gamma(q)\tau$. On $[Q, 1)$, (2.10) and $\lambda = \tau$ give $b\psi = R\bar{R} = V = v$ and $F = V - R\bar{R} = 0$. The functions $v - b\psi$ and F belong to $\mathcal{D}$, so Lemma 2.6(c) on $[q, Q]$, started at (q, x) , gives the two representations. Their integrands are nonnegative, since $\gamma \leq 1$, $c > 0$, $v < 1$, and $\lambda \geq \gamma\tau$. This proves (2.26), the upper bound being $F \geq 0$, because $\tau c = \frac{\tau}{\lambda}v$.

(b) Put $f = \lambda b$, so that $f_x = -v < 0$, $f > 0$, $\psi = f + x - \kappa$ and, by (2.26), $f\psi = \lambda b\psi \leq \tau v < \tau$. For $x < \kappa$, $f = \psi + \kappa - x > \kappa - x$ since $\psi > 0$, so $\psi \leq \tau/f \leq \tau/(\kappa - x)$. For $x > \kappa$, $\psi > x - \kappa$, so $f \leq \tau/(x - \kappa)$. At $x = \kappa$, $\psi = f$ and $\psi^2 \leq \tau$. The lower bound $\psi(q, \kappa) \geq \sqrt{\delta} \bar{R}(0) = 2\phi_0\sqrt{\delta}$ for $q \leq Q$ is (2.24). For $x < \kappa$ let $x' = (x + \kappa)/2$. By convexity of ψ and $\psi > 0$,

$$1 - v(q, x) = \psi_x(q, x) \leq \frac{\psi(q, x')}{x' - x} \leq \frac{2\tau}{\kappa - x} \cdot \frac{2}{\kappa - x}.$$

For $x > \kappa$ let $x' = (x + \kappa)/2$. Since v is decreasing in x ,

$$v(q, x)(x - x') \leq \int_{x'}^x v = f(x') - f(x) \leq f(x') \leq \frac{2\tau}{x - \kappa}.$$

Together with $0 < v < 1$ this proves the bound on $|v - \mathbf{1}_{\{x < \kappa\}}|$. Finally, $\psi_x = 1 - v$ and $\psi(q, x) \rightarrow 0$ as $x \rightarrow -\infty$ give $\int_{-\infty}^{\kappa} (1 - v) = \psi(q, \kappa)$, and $f_x = -v$ and $f(q, x) \rightarrow 0$ as $x \rightarrow \infty$ give $\int_{\kappa}^{\infty} v = f(q, \kappa) = \psi(q, \kappa)$. The difference of the two identities is the last one; all integrals converge absolutely by the bound on $|v - \mathbf{1}_{\{x < \kappa\}}|$. $\square$

2.5 First variation and the obstacle conditions

We now differentiate $\mathcal{P}_{\kappa}$ along segments in $\mathcal{U}_r$. The derivative has the form of a transport cost for the *obstacle potential*

$$\mathcal{H}(t) = \mathcal{H}_{\gamma}(t) = -\frac{1}{2} \int_0^t D(q) \, dq \quad (0 \leq t < 1). \quad (2.27)$$

By Lemma 2.9(c), $\mathcal{H} \in C^2([0, 1))$, $\mathcal{H}' = -D/2$, and $\mathcal{H}'' = (1 - S)/(2\lambda^2)$. Since $D(0) = \alpha b(0, 0)^2 > 0$, $\mathcal{H}$ is strictly decreasing near 0.

Proposition 2.11. *Let $\alpha > 0$ and $\kappa \in \mathbb{R}$.*

(a) *First variation. Let $r \in [0, 1)$ and $\gamma, \tilde{\gamma} \in \mathcal{U}_r$. Then*

$$\left. \frac{d}{d\vartheta} \mathcal{P}_{\kappa}(\gamma + \vartheta(\tilde{\gamma} - \gamma)) \right|_{\vartheta=0+} = \frac{1}{2} \int_0^r (\tilde{\gamma} - \gamma)(q) D(q) \, dq = \int_{[0, r]} \mathcal{H} \, d\mu_{\tilde{\gamma}} - \int_{[0, r]} \mathcal{H} \, d\mu_{\gamma}, \quad (2.28)$$

where D and $\mathcal{H}$ are those of γ .

- (b) Minimizers at a fixed cutoff. Let $r \in (0, 1)$ and let γ minimize $\mathcal{P}_\kappa$ over $\mathcal{U}_r$. Then $\text{supp } \mu_\gamma \subset \text{argmin}_{[0, r]} \mathcal{H}$ and $0 \notin \text{supp } \mu_\gamma$. Every $t \in \text{supp } \mu_\gamma$ with $t < r$ satisfies $D(t) = 0$ and $S(t) \leq 1$. Moreover $D(Q) \geq 0$, with $D(Q) = 0$ and $S(Q) \leq 1$ if $Q < r$, and

$$\int_0^Q \gamma(q) D(q) dq = 0.$$

- (c) Global minimizers and KKT points. We call $\gamma \in \mathcal{U}$ a KKT point of $\mathcal{P}_\kappa$ if $\text{supp } \mu_\gamma \subset \text{argmin}_{[0, r]} \mathcal{H}_\gamma$ for some $r \in (Q, 1)$. If γ minimizes $\mathcal{P}_\kappa$ over $\mathcal{U}$, then for every $r \in (Q, 1)$, γ minimizes $\mathcal{P}_\kappa$ over $\mathcal{U}_r$ and $\text{supp } \mu_\gamma \subset \text{argmin}_{[0, r]} \mathcal{H}$; in particular γ is a KKT point, and

$$\mathcal{H}(t) \geq \mathcal{H}(Q) \quad (0 \leq t < 1), \quad \mathcal{H} = \mathcal{H}(Q) \text{ on } \text{supp } \mu_\gamma,$$

that is, $\text{supp } \mu_\gamma \subset \text{argmin}_{[0, 1]} \mathcal{H}$. If γ is a KKT point (in particular, if γ is a global minimizer), then $0 < q_0 \leq Q < 1$, and $D(t) = 0$ and $S(t) \leq 1$ for every $t \in \text{supp } \mu_\gamma$; in particular $D(q_0) = D(Q) = 0$. Moreover, on $[0, q_0]$ one has $\gamma = 0$, $X_q = B_q$, and $\lambda = \lambda(0)$; S is strictly increasing on $[0, q_0]$, and $D > 0$ on $[0, q_0]$.

- (d) The scalar χ . If γ is a KKT point (in particular, a global minimizer), then $\chi(q) = \chi(0) + \gamma(q)D(q)$ for every $q \in [0, 1]$. In particular χ is constant on $[0, q_0] \cup \text{supp } \mu_\gamma$.

The conditions in (b)–(c) are necessary conditions only. The inequality $S \leq 1$ at support points uses that each support point minimizes $\mathcal{H}$; it does not follow from $D = 0$ on the support alone. For a global minimizer the conditions are obtained in $\mathcal{U}_r$ with $Q < r < 1$, so that every support point is interior to $[0, r]$. Part (d) and Lemma 2.12 below use only the KKT property, so they apply to every KKT point. For this functional, El Alaoui and Sellke computed the first variation at a minimizer whose support is an interval, in directions supported on that interval, and deduced $D = 0$ and $S = 1$ there [54, Proposition 16]. Here (a) holds at every $\gamma \in \mathcal{U}_r$ in every direction of $\mathcal{U}_r$, and (b)–(c) hold for every global minimizer and every KKT point, with no assumption on the support. The first-order and obstacle conditions are the analogues of known conditions for other models. For mixed p -spin models, Jagannath and Tobasco proved the first-variation formula and the obstacle condition [77, Lemma 3.2 and Proposition 1.1], and conditions on the support appear in the work of Talagrand [126], Auffinger and W.-K. Chen [8, Theorem 5], and W.-K. Chen [38, Theorem 2 and Proposition 1]. For spherical models, where the Crisanti–Sommers term enters as here, the obstacle condition appears in the work of Talagrand [125, Proposition 2.1], Jagannath and Tobasco [78], and Auffinger and W.-K. Chen [8, Proposition 4].

Proof. (a) Put $\gamma_\vartheta = \gamma + \vartheta(\tilde{\gamma} - \gamma)$, which lies in the convex set $\mathcal{U}_r$ for $\vartheta \in [0, 1]$. Lemma A.2, applied with $\tilde{\zeta} = \gamma$ and $\tilde{\zeta} = \gamma_\vartheta$ at $(q, x) = (0, 0)$, gives

$$u_{\gamma_\vartheta}(0, 0) - u_\gamma(0, 0) = \frac{\vartheta}{2} \int_0^r (\tilde{\gamma} - \gamma)(q) \mathbb{E} b_\gamma(q, Y_q^\vartheta)^2 dq,$$

where Y^ϑ solves $dY_q^\vartheta = \frac{1}{2}\gamma_\vartheta(q)(b_{\gamma_\vartheta} + b_\gamma)(q, Y_q^\vartheta) dq + dB_q$ with $Y_0^\vartheta = 0$. By Lemma 2.4(b) and (f) with $k = 1$, this drift differs from the drift γb_γ of X by at most $C\vartheta(|\tilde{\gamma} - \gamma|(q) + \|\tilde{\gamma} - \gamma\|_{L^1})(1 + |y|)$, and it is Lipschitz in y with constant $(1 - r)^{-1}$. The Gronwall bound (A.2) and Lemma 2.6(a) give $\mathbb{E} \sup_{q \leq r} |Y_q^\vartheta - X_q|^2 \leq C\vartheta^2$. Since $|\partial_x(b_\gamma^2)| = 2b_\gamma c_\gamma \leq C(1 + |x|)$, it follows that $\mathbb{E} b_\gamma(q, Y_q^\vartheta)^2 \rightarrow \mathbb{E} b_\gamma(q, X_q)^2$ uniformly in $q \in [0, r]$ as $\vartheta \downarrow 0$. Hence the derivative of the energy term $\alpha u_{\gamma_\vartheta}(0, 0)$ at $\vartheta = 0+$ is $\frac{\alpha}{2} \int_0^r (\tilde{\gamma} - \gamma) \mathbb{E} b(q, X_q)^2 dq$. By (2.12) the entropy term is $\text{CS}(\gamma_\vartheta) = \frac{1}{2} \int_0^r \lambda_{\gamma_\vartheta}^{-1} + \frac{1}{2} \log(1 - r)$, with $\lambda_{\gamma_\vartheta} = \lambda_\gamma + \vartheta(\lambda_{\tilde{\gamma}} - \lambda_\gamma) \geq 1 - r$ and $\lambda_{\tilde{\gamma}}(q) - \lambda_\gamma(q) = \int_q^r (\tilde{\gamma} - \gamma)$. Differentiating under the integral and applying Fubini's theorem, its derivative is

$$-\frac{1}{2} \int_0^r \frac{1}{\lambda(q)^2} \int_q^r (\tilde{\gamma} - \gamma)(s) ds dq = -\frac{1}{2} \int_0^r (\tilde{\gamma} - \gamma)(s) \int_0^s \frac{dq}{\lambda(q)^2} ds.$$

Adding the two derivatives gives the first equality in (2.28), by the definition (2.22) of D . For the second, let π be a probability measure on $[0, r]$ with distribution function F_π . By Fubini's theorem,

$$\int_0^r F_\pi(q) D(q) dq = \int_{[0, r]} \int_s^r D(q) dq d\pi(s) = 2 \int_{[0, r]} (\mathcal{H}(s) - \mathcal{H}(r)) d\pi(s).$$

Apply this with $\pi = \mu_{\tilde{\gamma}}$ and $\pi = \mu_\gamma$, whose distribution functions on $[0, r]$ are $\tilde{\gamma}$ and γ , and subtract.

(b) Since γ minimizes over the convex set $\mathcal{U}_r$, the right derivative in (2.28) is nonnegative for every $\tilde{\gamma} \in \mathcal{U}_r$. Taking $\tilde{\gamma} = \mathbf{1}_{[t, 1]}$ with $t \in [0, r]$, so that $\mu_{\tilde{\gamma}}$ is the point mass at t , gives $\mathcal{H}(t) \geq \int \mathcal{H} d\mu_\gamma$. Hence $\min_{[0, r]} \mathcal{H} \geq \int \mathcal{H} d\mu_\gamma \geq \min_{[0, r]} \mathcal{H}$, so $\mathcal{H} = \min_{[0, r]} \mathcal{H}$ μ_γ -almost everywhere, and by continuity on $\text{supp } \mu_\gamma$. Because $\mathcal{H}'(0) < 0$ and $r > 0$, the point 0 is not a minimum point, so $0 \notin \text{supp } \mu_\gamma$. A support point $t \in (0, r)$ is an interior minimum point of the C^2 function $\mathcal{H}$, so $\mathcal{H}'(t) = 0$ and $\mathcal{H}''(t) \geq 0$, that is, $D(t) = 0$ and $S(t) \leq 1$. Since $Q \in \text{supp } \mu_\gamma$, this applies to Q if $Q < r$. If $Q = r$, minimality at the right endpoint gives $\mathcal{H}'(r) \leq 0$, that is, $D(Q) \geq 0$. Finally, by Fubini's theorem and the constancy of $\mathcal{H}$ on $\text{supp } \mu_\gamma$,

$$\int_0^Q \gamma(q) D(q) dq = \int_{[0, Q]} \int_s^Q D(q) dq d\mu_\gamma(s) = 2 \int (\mathcal{H}(s) - \mathcal{H}(Q)) d\mu_\gamma(s) = 0.$$

(c) If γ is a global minimizer and $r \in (Q, 1)$, then $\gamma \in \mathcal{U}_r \subset \mathcal{U}$, so γ minimizes over $\mathcal{U}_r$, and (b) gives $\text{supp } \mu_\gamma \subset \arg\min_{[0, r]} \mathcal{H}$ and $\mathcal{H}(Q) = \min_{[0, r]} \mathcal{H}$. As $r \in (Q, 1)$ is arbitrary, $\mathcal{H} \geq \mathcal{H}(Q)$ on $[0, 1)$, and $\mathcal{H} = \mathcal{H}(Q)$ on $\text{supp } \mu_\gamma$. Now let γ be a KKT point, with $r \in (Q, 1)$ as in the definition. Since $r > 0$ and $\mathcal{H}'(0) < 0$, the point 0 does not minimize $\mathcal{H}$ on $[0, r]$, so $0 \notin \text{supp } \mu_\gamma$ and $q_0 > 0$. Every support point t lies in $(0, Q] \subset (0, r)$ and is therefore an interior minimum point of $\mathcal{H}$ on $[0, r]$; hence $D(t) = 0$ and $S(t) \leq 1$, as in (b). For $q < q_0$, $\gamma(q) = \mu_\gamma([0, q]) = 0$, so the drift in (2.14) vanishes, $X_q = B_q$, and

$\lambda(q) = \lambda(0)$. By Lemma 2.9(c)(iv), $S'(q+) = \alpha\mathcal{N}(q) > 0$ for $q \in [0, q_0)$. A continuous function with positive right derivative is strictly increasing, so S is strictly increasing on $[0, q_0]$. Hence $S < S(q_0) \leq 1$ on $[0, q_0)$, so $D' < 0$ there, and $D > D(q_0) = 0$ on $[0, q_0)$.

(d) By Lemma 2.9(c)(iii), $\chi(q) - \chi(0) = \int_0^q \gamma D' dt$. The integration by parts formula for the right-continuous function γ of bounded variation and the C^1 function D reads

$$\gamma(q)D(q) - \gamma(0)D(0) = \int_{(0,q]} D d\mu_\gamma + \int_0^q \gamma(t)D'(t) dt.$$

Here $\gamma(0) = \mu_\gamma(\{0\}) = 0$ because $0 \notin \text{supp } \mu_\gamma$, and the Stieltjes integral vanishes because $D = 0$ on $\text{supp } \mu_\gamma$ by (c). Hence $\chi(q) - \chi(0) = \gamma(q)D(q)$. The right side vanishes on $[0, q_0)$, where $\gamma = 0$, and on $\text{supp } \mu_\gamma$, where $D = 0$. $\square$

A gap of the support is an interval on which γ is constant. On such an interval the obstacle conditions force a critical point of S with a sign condition on S'' , which by Lemma 2.9(d) is a sign condition on Σ . This is the mechanism by which gaps are excluded in Chapter 8. For the SK model, Zhou excluded gaps of the support through the first-order conditions at their endpoints [134], and Lopatto added a sign condition inside the gap [87, Section 9]; Lemma 2.12 combines the two.

Lemma 2.12. *Let γ be a KKT point of $\mathcal{P}_\kappa$ in the sense of Proposition 2.11(c), with $r \in (Q, 1)$ as in the definition; for instance, let γ be a global minimizer.*

- (a) *Let $p < p'$ be points of $\text{supp } \mu_\gamma$ with $(p, p') \cap \text{supp } \mu_\gamma = \emptyset$. Then $\gamma = m := \gamma(p)$ on $[p, p']$, with $0 < m < 1$. Moreover $D(p) = D(p') = 0$, $S(p) \leq 1$, $S(p') \leq 1$, $\mathcal{H} \geq \mathcal{H}(p) = \mathcal{H}(p')$ on $[p, p']$, $\lambda(q) = \lambda(p') + m(p' - q)$ on $[p, p']$, and $S \in C^\infty([p, p'])$. There is $x^* \in (p, p')$ with*

$$S(x^*) \geq 1, \quad S'(x^*) = 0, \quad S''(x^*) \leq 0;$$

equivalently, $\tilde{x}(x^*) = m$ and $\Sigma(x^*; m/\lambda(x^*)) \leq 0$. If D does not vanish identically on $[p, p']$, then x^* can be taken to be a maximum point of S on $[p, p']$, and $S(x^*) > 1$. If $D \equiv 0$ on $[p, p']$, then $S \equiv 1$, $\tilde{x} \equiv m$, and $\Sigma(\cdot; m/\lambda) \equiv 0$ on $[p, p']$.

- (b) *If $[c_1, c_2] \subset \text{supp } \mu_\gamma$ with $c_1 < c_2$, then $D = 0$ and $S = 1$ on $[c_1, c_2]$, $\gamma = \tilde{x}$ on $[c_1, c_2)$, and $\gamma(q-) = \tilde{x}(q)$ for $q \in (c_1, c_2]$. In particular γ is continuous on (c_1, c_2) .*

Proof. (a) Since $\mu_\gamma((p, p')) = 0$, γ is constant on $[p, p']$. As $p \in \text{supp } \mu_\gamma$ and $\mu_\gamma((p, p')) = 0$, every interval $(p - \varepsilon, p]$ has positive mass, so $m = \gamma(p) > 0$. As $p' \in \text{supp } \mu_\gamma$, $\mu_\gamma([p', 1]) > 0$ and $m = 1 - \mu_\gamma([p', 1]) < 1$. The statements about D and S at p, p' are Proposition 2.11(c). Since $p, p' \in \text{supp } \mu_\gamma \subset \text{argmin}_{[0,r]} \mathcal{H}$ and $[p, p'] \subset [0, Q] \subset [0, r]$, we have $\mathcal{H} \geq \mathcal{H}(p) = \mathcal{H}(p')$ on $[p, p']$. The formula for λ follows from $\lambda' = -m$. Smoothness of S is Lemma 2.9(d) with $[s, t] = [p, p']$. Put $\mathcal{G} = \mathcal{H} - \mathcal{H}(p) \geq 0$ on $[p, p']$, so $\mathcal{G}(p) = \mathcal{G}(p') = 0$ and $\mathcal{G}' = -D/2$. If $\mathcal{G} \equiv 0$, then $D \equiv 0$ and $D' = (S - 1)/\lambda^2 \equiv 0$, so

$S \equiv 1$ and $S' \equiv S'' \equiv 0$ on $[p, p']$; Lemma 2.9(d) and $\tilde{B} > 0$ give $\tilde{x} \equiv m$ and $\Sigma(\cdot; m/\lambda) \equiv 0$. Otherwise $\mathcal{G}$ has a positive maximum at some $s_M \in (p, p')$. The mean value theorem gives $t_1 \in (p, s_M)$ and $t_2 \in (s_M, p')$ with $\mathcal{G}'(t_1) > 0 > \mathcal{G}'(t_2)$, that is, $D(t_1) < 0 < D(t_2)$, and then $x_2 \in (t_1, t_2)$ with $D'(x_2) > 0$, that is, $S(x_2) > 1$. Since $S(p), S(p') \leq 1$, the maximum of S on $[p, p']$ exceeds 1 and is attained at an interior point x^* , where $S'(x^*) = 0$ and $S''(x^*) \leq 0$. The equivalent form follows from Lemma 2.9(d) with $K = m/\lambda$ and $\tilde{B} > 0$.

(b) By Proposition 2.11(c), $D = 0$ on $[c_1, c_2]$; hence $D' = 0$ and $S = 1$ there. For $q \in [c_1, c_2)$ the right derivative of S vanishes, and Lemma 2.9(c)(iv) gives $\mathcal{N}(q) = 2 \frac{\gamma(q)}{\lambda(q)} \tilde{B}(q)$, that is, $\gamma(q) = \tilde{x}(q)$. The left derivative at $q \in (c_1, c_2]$ gives $\gamma(q-) = \tilde{x}(q)$ in the same way. Since $\tilde{x}$ is continuous, so is γ on (c_1, c_2) . $\square$

2.6 The value identity, the slope in the margin, and weak top stability

The rest of this chapter studies the variational problem $\mathcal{P}_*(\kappa) = \inf_{\gamma \in \mathcal{U}} \mathcal{P}_\kappa(\gamma)$. §§2.6 and 2.7 contain the value identity (Lemma 2.13), the universal slope bound (Lemma 2.14), weak top stability (Lemma 2.15), the existence of minimizers (Theorem 2.16), the finiteness of the critical margin (Proposition 2.17) and the divergence of $\mathcal{P}_*$ at the critical margin (Theorem 2.18). They appear with the same statements in Part I, Lemmas 6.2, 6.4, and 6.5, Theorems 6.3 and 6.8, and Proposition 6.7, and we refer to Part I, Chapter 6, for their proofs. §2.8 introduces the zero-temperature functional J^k , compares $\mathcal{P}_\kappa$ with it through an order-preserving rescaling of the Cole–Hopf recursion (Lemma 2.20), identifies κ_c with the threshold of J^k (Theorem 2.21), and records the slack of J^k below κ_c (Corollary 2.22), which enters the top-gap bound of Chapter 4. §2.9 proves that $\mathcal{P}_*(\kappa_c - \eta)$ is of logarithmic order in η for $\alpha > 2$ (Proposition 2.24); no later chapter uses it, and it enters only the consequences for the random model stated in §1.2.

Conventions. Throughout the rest of this chapter $\alpha > 0$ is fixed. For $\kappa \in \mathbb{R}$ and $\gamma \in \mathcal{U}$ we use the notation of Definition 2.2: $Q = Q_\gamma$, $\delta = 1 - Q$, $\lambda = \lambda_\gamma$, $u = u_\gamma(\cdot, \cdot; \kappa)$, $b = \partial_x u$, $c = -\partial_{xx} u$, and $v = \lambda c$. The diffusion X of (2.14) is driven by the Brownian motion B , and D and S are the functions of (2.22). When the order parameter or the margin varies, we write $u_\gamma(q, x; \kappa)$ as in Definition 2.2, and $b_{\gamma, \kappa} = \partial_x u_\gamma(\cdot, \cdot; \kappa)$, $c_{\gamma, \kappa}$, and so on. The *normalized top position* is

$$Z = \frac{X_Q - \kappa}{\sqrt{\delta}}, \quad \text{so that} \quad u(Q, X_Q) = \log \Phi(Z), \quad b(Q, X_Q) = \frac{R(Z)}{\sqrt{\delta}}$$

by (2.9) and (2.10), and $\sqrt{\delta} b(0, 0) = \mathbb{E} R(Z)$ by Lemma 2.9(b). Here and below, G is a standard Gaussian variable independent of B . Besides Lemma 2.1(c) we use the following shifted form of the Gaussian tail bound: for every $t \in \mathbb{R}$ and $\theta > 0$, with

$$\log_+ y = \max\{\log y, 0\},$$

$$\log \Phi(-t) \geq -\frac{1}{2}(t + \theta)_+^2 - \log_+ \frac{1}{\theta} - 2. \quad (2.29)$$

Indeed, if $t \leq 0$ the left side is at least $\log \frac{1}{2} > -2$. If $t > 0$, then $\frac{1}{2}(t + \theta)^2 \geq \frac{1}{2}t^2 + \theta t$, and by Lemma 2.1(c) it suffices to show $\log(1 + t) \leq \theta t + \log_+(1/\theta)$. For $\theta \geq 1$ this follows from $\log(1 + t) \leq t$; for $\theta < 1$ the concave function $t \mapsto \log(1 + t) - \theta t$ is maximal at $t = 1/\theta - 1$, where it equals $\log(1/\theta) - 1 + \theta < \log(1/\theta)$. Part I, (6.4), states the same bound with $-\log \Phi(-1) < 2$ in place of 2.

The next identity expresses $\mathcal{P}_\kappa(\gamma)$ through the terminal law of X and the function D . For a minimizer the D -terms vanish or have a sign, and the identity then bounds the top gap δ from below by the value.

Lemma 2.13. *For every $\kappa \in \mathbb{R}$ and $\gamma \in \mathcal{U}$,*

$$\mathcal{P}_\kappa(\gamma) = \frac{1}{2} \log \delta + \alpha \mathbb{E}h(Z) - \frac{\delta}{2} D(Q) - \frac{1}{2} \int_0^Q \gamma(q) D(q) dq. \quad (2.30)$$

If $r \in [0, 1)$ and γ minimizes $\mathcal{P}_\kappa$ over $\mathcal{U}_r$, then $\int_0^Q \gamma D dq = 0$ and $D(Q) \geq 0$, with $D(Q) = 0$ if $Q < r$; consequently

$$\mathcal{P}_\kappa(\gamma) < \frac{1}{2} \log \delta. \quad (2.31)$$

If γ minimizes $\mathcal{P}_\kappa$ over $\mathcal{U}$, then $D(Q) = 0$ and $\mathcal{P}_\kappa(\gamma) = \frac{1}{2} \log \delta + \alpha \mathbb{E}h(Z)$.

Proof. This is Part I, Lemma 6.2. □

We next study the dependence on the margin. By Lemma 2.4(g), changing κ translates u , so ∂_κ acts as $-\partial_x$. The next lemma bounds the slope $b(0, 0)$ for every order parameter, which is what makes it usable for near-minimizers and constrained minimizers alike. Its upper bound follows from the inequality $F \geq 0$ of Lemma 2.10(a) at $(q, x) = (0, 0)$ and from $v < 1$.

Lemma 2.14. *Let $C_\kappa = \frac{1}{2}(\kappa + \sqrt{\kappa^2 + 4})$. For every $\gamma \in \mathcal{U}$ and $\kappa \in \mathbb{R}$,*

$$\mathbb{E}(\kappa - B_Q)_+ \leq \lambda(0) b(0, 0) < C_\kappa. \quad (2.32)$$

The map $\kappa \mapsto \mathcal{P}_\kappa(\gamma)$ is twice continuously differentiable, with $\partial_\kappa \mathcal{P}_\kappa(\gamma) = -\alpha b(0, 0) < 0$ and $\partial_\kappa^2 \mathcal{P}_\kappa(\gamma) = -\alpha c(0, 0) < 0$; in particular it is strictly decreasing and strictly concave. Consequently $\mathcal{P}_*$, as a function with values in $[-\infty, 0)$, is nonincreasing and concave, and $\kappa \mapsto b_{\gamma, \kappa}(0, 0)$ is increasing.

Proof. This is Part I, Lemma 6.4. □

We use the lemma mostly through the tangent-line inequality: by concavity, for every $\gamma \in \mathcal{U}$ and all margins k, k' ,

$$\mathcal{P}_{k'}(\gamma) \leq \mathcal{P}_k(\gamma) - \alpha b_{\gamma,k}(0,0) (k' - k). \quad (2.33)$$

When the support of μ_γ has more than one point, the curvature S at the top cannot be small; this is the content of the next lemma. Together with Lemma 2.14 it bounds the mass $\lambda(0)$ of a minimizer by a multiple of $\sqrt{\delta}$.

Lemma 2.15. *Let $\kappa \in \mathbb{R}$ and $r \in [0, 1)$, and let γ minimize $\mathcal{P}_\kappa$ over $\mathcal{U}_r$. If μ_γ has more than one support point, then $\alpha > 1$ and*

$$S(Q) \geq s_\alpha := \min \left\{ \frac{1}{2}, \frac{1}{16\alpha} \right\}.$$

Moreover, every $\gamma \in \mathcal{U}$ satisfies, at every margin κ ,

$$\sqrt{\delta} b(0,0) \geq \frac{S(Q)}{2\alpha}. \quad (2.34)$$

Consequently, every minimizer γ of $\mathcal{P}_\kappa$ over $\mathcal{U}_r$ satisfies

$$\lambda(0) \leq \max \left\{ 1, \frac{2\alpha C_\kappa}{s_\alpha} \right\} \sqrt{\delta}. \quad (2.35)$$

Proof. This is Part I, Lemma 6.5. □

2.7 The critical margin and existence of minimizers

By Lemma 2.14, $\mathcal{P}_*(\cdot; \alpha)$ is nonincreasing. By (2.11) and $u \leq 0$, $\mathcal{P}_\kappa(\gamma; \alpha)$ and $\mathcal{P}_*(\kappa; \alpha)$ are nonincreasing in α . The *critical margin* is

$$\kappa_c(\alpha) = \sup \{ \kappa \in \mathbb{R} : \mathcal{P}_*(\kappa; \alpha) > -\infty \}. \quad (2.36)$$

Thus $\mathcal{P}_* > -\infty$ on $(-\infty, \kappa_c)$ and $\mathcal{P}_* = -\infty$ on (κ_c, ∞) , and $\alpha \mapsto \kappa_c(\alpha)$ is nonincreasing. For $r \in [0, 1)$ we write

$$\mathcal{P}_*^r(\kappa) = \min_{\gamma \in \mathcal{U}_r} \mathcal{P}_\kappa(\gamma) \quad (2.37)$$

for the constrained minimum, which is attained because $\mathcal{U}_r$ is compact and $\mathcal{P}_\kappa$ is continuous on it (Lemma 2.4(f)). Since the classes $\mathcal{U}_r$ increase with r and their union is $\mathcal{U}$, $\mathcal{P}_*^r(\kappa)$ decreases to $\mathcal{P}_*(\kappa)$ as $r \uparrow 1$. Also $\mathcal{P}_*^0(\kappa) = \alpha \log \Phi(-\kappa) < 0$, since $\mathcal{U}_0 = \{\mathbf{1}_{[0,1]}\}$; in particular $\mathcal{P}_*(\kappa) < 0$.

Theorem 2.16. *Let $\kappa \in \mathbb{R}$ with $\mathcal{P}_*(\kappa) > -\infty$. Then the infimum defining $\mathcal{P}_*(\kappa)$ is attained. Every minimizer γ satisfies*

$$1 - Q_\gamma > e^{2\mathcal{P}_*(\kappa)},$$

that is, every minimizer lies in $\mathcal{U}_r$ with $r = 1 - e^{2\mathcal{P}_*(\kappa)} \in (0, 1)$, and the set of minimizers is compact in $L^1([0, 1])$.

Proof. This is Part I, Theorem 6.3. $\square$

The theorem concerns minimizers. It does not assert that the tops of an arbitrary minimizing sequence stay away from one: adding a vanishing mass near one changes the top without changing the L^1 limit.

For the replica-symmetric order parameter $\mathbf{1}_{[q,1]}$, $0 \leq q < 1$, the order parameter vanishes on $[0, q)$, so $X = B$ there and $\lambda = 1 - q$ on $[0, q]$. Hence

$$\mathcal{P}_\kappa^{\text{RS}}(q) := \mathcal{P}_\kappa(\mathbf{1}_{[q,1]}) = \alpha \mathbb{E} \log \Phi\left(\frac{\sqrt{q} G - \kappa}{\sqrt{1-q}}\right) + \frac{q}{2(1-q)} + \frac{1}{2} \log(1-q). \quad (2.38)$$

Proposition 2.17. *For every $\alpha > 0$,*

$$-\infty < \kappa_c(\alpha) \leq \kappa_{\text{RS}}(\alpha) < \infty,$$

where $\kappa_{\text{RS}}(\alpha)$ is the unique solution of $\alpha m(\kappa_{\text{RS}}) = 1$. In particular $\kappa_c(\alpha) < 0$ if $\alpha > 2$.

Proof. This is Part I, Proposition 6.7. $\square$

The bound $\kappa_c \leq \kappa_{\text{RS}}$ comes from the limit $q \uparrow 1$ in (2.38): if $\alpha m(\kappa) > 1$, then $\mathcal{P}_\kappa^{\text{RS}}(q) \rightarrow -\infty$ as $q \uparrow 1$. This is the limit in which Gardner computed the replica-symmetric capacity [62]. For this functional, El Alaoui and Sellke made the same observation in the remark after [54, Conjecture 1].

Theorem 2.18. *For every $\alpha > 0$ the function $\mathcal{P}_*$ is finite, nonincreasing, concave, and continuous on $(-\infty, \kappa_c)$, and $\mathcal{P}_*(\kappa) = -\infty$ for $\kappa \geq \kappa_c$. More precisely, with $C_* = C_{\kappa_c}$, for all sufficiently small $\eta > 0$,*

$$\mathcal{P}_*(\kappa_c - \eta) \leq \frac{1}{2} \log \eta + \frac{1}{2} \log(2\alpha C_*) + \frac{1}{2}. \quad (2.39)$$

Proof. This is Part I, Theorem 6.8. $\square$

2.8 The zero-temperature functional

The critical margin also has a zero-temperature characterization, which is used to certify bounds on κ_c in Chapter 3. It involves the Parisi equation with a quadratic terminal penalty in place of $\log \Phi$. Zero-temperature Parisi functionals were introduced for the ground state energy of mixed p -spin models by Auffinger and W.-K. Chen [9], and for spherical models by W.-K. Chen and Sen [40] and by Jagannath and Tobasco [78]; see also [75]. For the negative perceptron, the quadratic penalty is the boundary condition of the jamming limit in the replica analysis of Franz and Parisi [58, (15)]; see also [59]. For $k \in \mathbb{R}$ let

$$f_k(x) = -\frac{1}{2}(k - x)_+^2,$$

which belongs to the class $\mathfrak{F}$ of Lemma 2.8, with $f'_k(x) = (k - x)_+$. For $T > 0$ let $\mathfrak{X}_T$ be the set of bounded, nondecreasing, right-continuous functions $\xi : [0, T) \rightarrow [0, \infty)$, and put $\mathfrak{X} = \mathfrak{X}_1$; the values of ξ may exceed one. For a step function $\xi \in \mathfrak{X}_T$ let $Y = Y_{T,\xi}^k = Y_{\xi,f_k}$ be the solution of the recursion of Lemma 2.8, that is, of

$$Y_q = -\frac{1}{2}(Y_{xx} + \xi Y_x^2), \quad Y(T, x) = f_k(x), \quad (2.40)$$

and for general $\xi \in \mathfrak{X}_T$ let $Y_{T,\xi}^k$ be the L^1 -limit of step approximations (Lemma 2.19(a)). Define

$$J_T^k(\xi) = \alpha Y_{T,\xi}^k(0, 0) + \frac{1}{2} \int_0^T \frac{dq}{\Lambda_\xi(q)}, \quad \Lambda_\xi(q) = 1 + \int_q^T \xi(t) dt, \quad J^k = J_1^k, \quad (2.41)$$

and the thresholds

$$\begin{aligned} \kappa_J(\alpha) &= \inf\{k : J^k(\xi) < 0 \text{ for some } \xi \in \mathfrak{X}\}, \\ \kappa_c^0(\alpha) &= \inf\{k : J^k(\xi) < 0 \text{ for some step function } \xi \in \mathfrak{X}\}. \end{aligned}$$

If $k < \kappa_J$, then $J^k(\xi) \geq 0$ for every $\xi \in \mathfrak{X}$.

Two elementary properties of the Cole–Hopf step $\mathcal{T}_{m,s}$ of (2.5) drive this section. Let $m \geq 0$, $s > 0$, and let g be bounded above with $g(x) \geq -C(1 + x^2)$. For $K > 0$ and $T > 0$,

$$K \mathcal{T}_{m,s} g = \mathcal{T}_{m/K,s}(Kg), \quad \frac{1}{T}(\mathcal{T}_{m,s} g)(\sqrt{T} y) = \mathcal{T}_{Tm,s/T} g^\sharp(y), \quad g^\sharp(y) = \frac{1}{T} g(\sqrt{T} y). \quad (2.42)$$

The first identity is (2.7). For the second, if $m > 0$ the right side equals $(Tm)^{-1} \log \mathbb{E} \exp\{m g(\sqrt{T} y + \sqrt{s} G)\}$, which is the left side, and if $m = 0$ both sides equal $T^{-1} \mathbb{E} g(\sqrt{T} y + \sqrt{s} G)$. By the first identity, multiplying the terminal datum by K and dividing all coefficients by K multiplies the solution of the recursion by K . By the second, rescaling time by T and space by $\sqrt{T}$ maps a recursion on $[0, T]$ to a recursion on $[0, 1]$. Together with the monotonicity in (2.7), these identities are all that the comparison of $\mathcal{P}_\kappa$ with J^k requires.

Lemma 2.19. *Let $T > 0$, $M > 0$, and $K_0 \geq 0$.*

(a) *For step functions $\xi, \tilde{\xi} \in \mathfrak{X}_T$ with $\xi, \tilde{\xi} \leq M$ and for $|k| \leq K_0$,*

$$\begin{aligned} |Y_{T,\tilde{\xi}}^k(q, x) - Y_{T,\xi}^k(q, x)| &\leq C(1 + x^2) \|\tilde{\xi} - \xi\|_{L^1(0,T)}, \\ |J_T^k(\tilde{\xi}) - J_T^k(\xi)| &\leq C(1 + \alpha) \|\tilde{\xi} - \xi\|_{L^1(0,T)}, \end{aligned}$$

with $C = C(T, M, K_0)$. Hence $Y_{T,\xi}^k$ and $J_T^k(\xi)$ extend uniquely to all $\xi \in \mathfrak{X}_T$, continuously in L^1 on $\{\xi \leq M\}$.

(b) For $\xi \in \mathfrak{X}$ and $k < k'$,

$$\frac{\alpha}{\Lambda_{\xi}(0)} \int_k^{k'} m_1(t) dt \leq J^k(\xi) - J^{k'}(\xi) \leq \alpha \int_k^{k'} m_1(t) dt.$$

In particular $k \mapsto J^k(\xi)$ is continuous and strictly decreasing. For step functions ξ , $k \mapsto J^k(\xi)$ is continuously differentiable with $\partial_k J^k(\xi) = -\alpha \partial_x Y_{1,\xi}^k(0,0)$.

(c) Brownian scaling. For $\xi \in \mathfrak{X}_T$ put $\tilde{\xi}(t) = T\xi(Tt)$ on $[0,1)$. Then $J_T^k(\xi) = T J^{k/\sqrt{T}}(\tilde{\xi})$.

(d) $\kappa_J(\alpha) = \kappa_c^0(\alpha)$.

Proof. (a) The datum f_k belongs to $\mathfrak{F}$, with $f_k(x) \geq -(K_0^2 + x^2)$ and $\text{Lip}(f'_k) = 1$. Hence Lemma A.2 gives the first bound, with C depending only on T , M , and K_0 . Since $\Lambda_{\tilde{\xi}}, \Lambda_{\xi} \geq 1$ and $|\Lambda_{\tilde{\xi}} - \Lambda_{\xi}| \leq \|\tilde{\xi} - \xi\|_{L^1}$, the entropy terms differ by at most $\frac{T}{2} \|\tilde{\xi} - \xi\|_{L^1}$. Every $\xi \in \mathfrak{X}_T$ with $\xi \leq M$ is the L^1 -limit of the nondecreasing step functions $\xi_n \leq M$ that equal $\xi(jT/n)$ on $[jT/n, (j+1)T/n)$; the two bounds show that Y_{T,ξ_n}^k and $J_T^k(\xi_n)$ are Cauchy sequences whose limits do not depend on the approximating sequence, and the bounds pass to the limit.

(b) Let ξ be a step function, and let X^* be the optimal process of Lemma 2.8 with $T = 1$, $f = f_k$, and $x = 0$. Then $M_q = \partial_x Y_{1,\xi}^k(q, X_q^*) \geq 0$ is a martingale with $M_1 = (k - X_1^*)_+$. Write $X_1^* = B_1 + \Delta$ with $\Delta = \int_0^1 \xi(q) M_q dq \geq 0$, so that $\mathbb{E}\Delta = M_0 \int_0^1 \xi = M_0(\Lambda_{\xi}(0) - 1)$. Since $(k - X_1^*)_+ \leq (k - B_1)_+ \leq (k - X_1^*)_+ + \Delta$, taking expectations gives $M_0 \leq m_1(k) \leq \Lambda_{\xi}(0) M_0$. The recursion commutes with translations and $f_k(x) = f_0(x - k)$, so $Y_{1,\xi}^k(0,0) = Y_{1,\xi}^0(0,-k)$; since $\partial_x Y_{1,\xi}^0$ is continuous (Lemma 2.8), $k \mapsto Y_{1,\xi}^k(0,0)$ is continuously differentiable with derivative $-M_0 = -\partial_x Y_{1,\xi}^k(0,0)$. Hence $-\alpha m_1(k) \leq \partial_k J^k(\xi) \leq -\alpha m_1(k)/\Lambda_{\xi}(0)$, and integrating over $[k, k']$ gives (b) for step functions. By (a) and the continuity of $\xi \mapsto \Lambda_{\xi}(0)$ in L^1 , (b) extends to all $\xi \in \mathfrak{X}$.

(c) Let ξ be a step function, equal to m_j on $[t_j, t_{j+1})$. Then $\tilde{\xi}$ equals Tm_j on $[t_j/T, t_{j+1}/T)$, and $f_k^\sharp(y) = T^{-1}f_k(\sqrt{T}y) = f_{k/\sqrt{T}}(y)$. Applying the second identity in (2.42) to the steps of the recursion, from the last to the first, gives $T^{-1}Y_{T,\xi}^k(Tt, \sqrt{T}y) = Y_{1,\tilde{\xi}}^{k/\sqrt{T}}(t, y)$ at the breakpoints $t = t_j/T$, and in particular at $t = 0$. The substitution $q = Tt$ and $\Lambda_{\xi}(Tt) = 1 + \int_t^1 T\tilde{\xi}(Ts) ds = \Lambda_{\tilde{\xi}}(t)$ give $\frac{1}{2} \int_0^T \Lambda_{\xi}^{-1} dq = \frac{T}{2} \int_0^1 \Lambda_{\tilde{\xi}}^{-1} dt$. This proves (c) for step functions. For general ξ it follows from (a), since $\|\tilde{\xi}_1 - \tilde{\xi}_2\|_{L^1(0,1)} = \|\xi_1 - \xi_2\|_{L^1(0,T)}$ and $\tilde{\xi} \leq T\|\xi\|_\infty$.

(d) Since step functions belong to $\mathfrak{X}$, $\kappa_J \leq \kappa_c^0$. Conversely, if $J^k(\xi) < 0$ for some $\xi \in \mathfrak{X}$ with $\xi \leq M$, then by (a) the same holds for a nondecreasing step function $\xi_n \leq M$ close to ξ in L^1 , and ξ_n can be taken right-continuous. $\square$

The comparison with $\mathcal{P}_\kappa$ rests on the terminal data. By Lemma 2.1(c) and (2.29) with $t = (\kappa - x)/\sqrt{\delta}$ and $\theta = \rho/\sqrt{\delta}$, for $\delta > 0$, $\rho > 0$, and $x \in \mathbb{R}$,

$$f_{\kappa+\rho}(x) - \delta \left(2 + \log_+ \frac{\sqrt{\delta}}{\rho} \right) \leq \delta \log \Phi \left(\frac{x - \kappa}{\sqrt{\delta}} \right) \leq f_\kappa(x), \quad (2.43)$$

because $\delta t_+^2 = (\kappa - x)_+^2$ and $\delta(t + \theta)_+^2 = (\kappa + \rho - x)_+^2$. Multiplying the hard-wall datum by δ therefore turns it into the quadratic penalty, up to a shift of the margin and an additive error of order δ .

Lemma 2.20. *Let $\kappa \in \mathbb{R}$ and $\gamma \in \mathcal{U}$ with $Q > 0$, and put $\xi_\gamma = \gamma/\delta$ on $[0, Q]$. Then*

$$\mathcal{P}_\kappa(\gamma) \leq \frac{J_Q^\kappa(\xi_\gamma)}{\delta} - \frac{1}{2} \log \frac{1}{\delta}, \quad (2.44)$$

and, for every $\rho > 0$,

$$\mathcal{P}_\kappa(\gamma) \geq \frac{J_Q^{\kappa+\rho}(\xi_\gamma)}{\delta} + \frac{1}{2} \log \delta - \alpha \left(2 + \log_+ \frac{\sqrt{\delta}}{\rho} \right). \quad (2.45)$$

In particular,

$$\mathcal{P}_\kappa(\gamma) \geq \frac{J_Q^{\kappa+\rho}(\xi_\gamma)}{\delta} + \frac{1}{2} \log \delta - 2\alpha \quad \text{for every } \rho \geq \sqrt{\delta}. \quad (2.46)$$

Moreover $J_Q^k(\xi_\gamma) = Q J^{k/\sqrt{Q}}(\tilde{\xi}_\gamma)$ for every $k \in \mathbb{R}$, where $\tilde{\xi}_\gamma(t) = Q\gamma(Qt)/\delta$ for $t \in [0, 1]$.

Proof. The function ξ_γ belongs to $\mathfrak{X}_Q$ and is bounded by $1/\delta$. Since $\Lambda_{\xi_\gamma}(q) = 1 + \delta^{-1} \int_q^Q \gamma = \lambda(q)/\delta$, the entropy terms satisfy exactly

$$\frac{1}{2} \int_0^Q \frac{dq}{\lambda(q)} = \frac{1}{\delta} \cdot \frac{1}{2} \int_0^Q \frac{dq}{\Lambda_{\xi_\gamma}(q)}. \quad (2.47)$$

It remains to prove

$$Y_{Q, \xi_\gamma}^{\kappa+\rho}(0, 0) - \delta \left(2 + \log_+ \frac{\sqrt{\delta}}{\rho} \right) \leq \delta u(0, 0) \leq Y_{Q, \xi_\gamma}^\kappa(0, 0); \quad (2.48)$$

multiplying by α/δ and adding (2.47) and $\frac{1}{2} \log \delta$ gives (2.44) and (2.45), and (2.46) is the case $\sqrt{\delta}/\rho \leq 1$ of (2.45).

Suppose first that γ is a step order parameter. Then $\gamma \in \mathcal{U}_Q$, and u is obtained on $[0, Q]$ from the datum $f_Q(x) = \log \Phi((x - \kappa)/\sqrt{\delta})$ at time Q by the recursion (2.6) with the blocks of γ . By the first identity in (2.42) with $K = \delta$, applied to the steps from the last to the first, $\delta u(0, \cdot)$ is obtained from the datum δf_Q at time Q by the recursion with the coefficients $\gamma/\delta = \xi_\gamma$. The recursion with these coefficients is nondecreasing

in its datum and commutes with the addition of constants (2.7), and $Y_{Q,\xi\gamma}^k$ is the same recursion started from f_k . Hence (2.43) gives (2.48).

For general γ , choose partitions of $[0, Q]$ containing Q with mesh tending to zero, and let $\gamma_n(q) = \gamma(t_j)$ on $[t_j, t_{j+1})$ and $\gamma_n = 1$ on $[Q, 1]$. Then γ_n is a step order parameter with $\gamma_n \leq \gamma$ and the same top Q , because $\gamma < 1$ on $[0, Q]$; moreover $\gamma_n \rightarrow \gamma$ in L^1 , and $\xi_{\gamma_n} \rightarrow \xi_\gamma$ in $L^1(0, Q)$ with $\xi_{\gamma_n} \leq 1/\delta$. Lemma 2.4(f) and Lemma 2.19(a) pass (2.44) and (2.45) to the limit. The last assertion is Lemma 2.19(c) with $T = Q$, since $Q\xi_\gamma(Qt) = \tilde{\xi}_\gamma(t)$. $\square$

We can now state the main result of this section, the zero-temperature characterization of κ_c .

Theorem 2.21. *For every $\alpha > 0$, $\kappa_c(\alpha) = \kappa_J(\alpha) = \kappa_c^0(\alpha)$.*

The two directions of Theorem 2.21 play different roles in Chapter 3. The inequality $\kappa_c \leq \kappa_c^0$ says that a single negative value of J^k at a step coefficient forces $\kappa_c \leq k$; this gives the upper bound on κ_c in Theorem 3.13. The inequality $\kappa_J \leq \kappa_c$ is the converse: if $J^k \geq 0$ for all coefficients, then $\kappa_c \geq \kappa_J \geq k$; this turns the dual certificate of Chapter 3 into a lower bound on κ_c .

Proof. We divide the proof into three steps.

STEP 1. *The inequality $\kappa_c \leq \kappa_c^0$.* Let $\xi \in \mathfrak{X}$ be a step function with $\xi \leq M$ and $J^k(\xi) < 0$, and let $k' > k$. For $\delta \in (0, \frac{1}{2})$ with $2\delta M < 1$ put $Q = 1 - \delta$ and $\gamma(q) = \delta Q^{-1}\xi(q/Q)$ for $q < Q$, $\gamma = 1$ on $[Q, 1]$. Then γ is a step order parameter with values in $[0, 2\delta M] \subset [0, 1)$ on $[0, Q)$, so its top is Q , and $\tilde{\xi}_\gamma(t) = Q\gamma(Qt)/\delta = \xi(t)$. By (2.44) and the last assertion of Lemma 2.20,

$$\mathcal{P}_{k'}(\gamma) \leq \frac{Q}{\delta} J^{k'/\sqrt{Q}}(\xi) + \frac{1}{2} \log \delta.$$

As $\delta \downarrow 0$, $k'/\sqrt{Q} \rightarrow k' > k$, so $k'/\sqrt{Q} \geq k$ for small δ , and then $J^{k'/\sqrt{Q}}(\xi) \leq J^k(\xi) < 0$ by Lemma 2.19(b). Hence $\mathcal{P}_{k'}(\gamma) \rightarrow -\infty$, and $\mathcal{P}_*(k') = -\infty$ for every $k' > k$. Thus $\kappa_c \leq k$, and taking the infimum over k gives $\kappa_c \leq \kappa_c^0$.

STEP 2. *The inequality $\kappa_J \leq \kappa_c$.* Fix margins $\kappa_c < k < k''$. We produce negative values of J at margins close to k'' from constrained minimizers at the margin k , where $\mathcal{P}_*(k) = -\infty$. For $r \in [0, 1)$ let γ_r minimize $\mathcal{P}_k$ over $\mathcal{U}_r$, and write $w_r = \mathcal{P}_*^r(k)$, $Q_r = Q_{\gamma_r}$, $\delta_r = 1 - Q_r$, and $b_r = b_{\gamma_r, k}(0, 0)$. We use four facts.

- (i) $w_r \rightarrow -\infty$ as $r \uparrow 1$, because $\mathcal{P}_*^r(k) \downarrow \mathcal{P}_*(k)$.
- (ii) $\frac{1}{2} \log \delta_r > w_r$, by (2.31).
- (iii) $\delta_r \rightarrow 0$ as $r \uparrow 1$. Indeed, for fixed $r_0 < 1$ the minimum $\mathcal{P}_*^{r_0}(k)$ is finite; once $w_r < \mathcal{P}_*^{r_0}(k)$, the minimizer γ_r does not belong to $\mathcal{U}_{r_0}$, that is, $Q_r > r_0$.

(iv) The slope b_r is large: by (2.33) at the margin $\kappa_c - 1$, where $\mathcal{P}_*$ is finite, $\mathcal{P}_*(\kappa_c - 1) \leq \mathcal{P}_{\kappa_c-1}(\gamma_r) \leq w_r + \alpha(k - \kappa_c + 1)b_r$.

By (2.33) at the margin k'' and by (iv), with $\vartheta = (k'' - k)/(k - \kappa_c + 1) > 0$,

$$\mathcal{P}_{k''}(\gamma_r) \leq w_r - \alpha(k'' - k)b_r \leq w_r - \vartheta(\mathcal{P}_*(\kappa_c - 1) - w_r).$$

For r close to one, $Q_r > 0$ by (iii), and (2.46) at the margin k'' with $\rho = \sqrt{\delta_r}$, followed by (ii), gives

$$\frac{J_{Q_r}^{k''+\sqrt{\delta_r}}(\xi_{\gamma_r})}{\delta_r} \leq \mathcal{P}_{k''}(\gamma_r) - \frac{1}{2} \log \delta_r + 2\alpha < -\vartheta(\mathcal{P}_*(\kappa_c - 1) - w_r) + 2\alpha.$$

By (i) the right side tends to $-\infty$ as $r \uparrow 1$. For r close to one it is negative, and the last assertion of Lemma 2.20 gives $J^{(k''+\sqrt{\delta_r})/\sqrt{Q_r}}(\tilde{\xi}_{\gamma_r}) < 0$ with $\tilde{\xi}_{\gamma_r} \in \mathfrak{X}$. Hence $\kappa_J \leq (k'' + \sqrt{\delta_r})/\sqrt{Q_r}$, and by (iii) this margin tends to k'' . Thus $\kappa_J \leq k''$ for every $k'' > \kappa_c$, that is, $\kappa_J \leq \kappa_c$.

STEP 3. Conclusion. By Steps 1 and 2 and Lemma 2.19(d), $\kappa_c \leq \kappa_c^0 = \kappa_J \leq \kappa_c$. This proves the theorem. $\square$

Since $J^{\kappa_c}(\xi) \geq 0$ for every $\xi \in \mathfrak{X}$, Lemma 2.19(b) yields a quantitative lower bound on J below the critical margin. Combined with (2.46) it bounds $\mathcal{P}_\kappa(\gamma)$ from below in terms of $\lambda(0)$ and δ ; this is the form used for the top-gap lower bound in Chapter 4.

Corollary 2.22. *For $k < \kappa_c(\alpha)$ and every $\xi \in \mathfrak{X}$,*

$$J^k(\xi) \geq \frac{A(k)}{\Lambda_\xi(0)}, \quad A(k) = \alpha \int_k^{\kappa_c} m_1(t) dt > 0.$$

Consequently, for $\gamma \in \mathcal{U}$ with $Q > 0$, $\rho \geq \sqrt{\delta}$, and $(\kappa + \rho)/\sqrt{Q} \leq k < \kappa_c$,

$$\mathcal{P}_\kappa(\gamma) \geq \frac{Q A(k)}{\lambda(0)} + \frac{1}{2} \log \delta - 2\alpha.$$

Proof. If $J^{\kappa_c}(\xi) < 0$ for some ξ , the continuity in Lemma 2.19(b) gives $J^{k'}(\xi) < 0$ for some $k' < \kappa_c = \kappa_J$, contradicting the definition of κ_J . Hence $J^{\kappa_c}(\xi) \geq 0$, and the lower bound in Lemma 2.19(b) with $k' = \kappa_c$ gives the first claim. For the second, (2.46), the last assertion of Lemma 2.20, the monotonicity in k , and $\Lambda_{\tilde{\xi}_\gamma}(0) = \Lambda_{\xi_\gamma}(0) = \lambda(0)/\delta$ give $J_Q^{\kappa+\rho}(\xi_\gamma) = Q J^{(\kappa+\rho)/\sqrt{Q}}(\tilde{\xi}_\gamma) \geq Q J^k(\tilde{\xi}_\gamma) \geq Q A(k) \delta/\lambda(0)$. $\square$

2.9 Logarithmic divergence at the critical margin

By Theorem 2.18, $\mathcal{P}_*(\kappa_c - \eta) \leq \frac{1}{2} \log \eta + O(1)$. For $\alpha > 2$ the comparison with J^k gives a lower bound of the same logarithmic order. The reason is that $\kappa_c < 0$ for $\alpha > 2$: dividing

a margin below κ_c by $\sqrt{Q} \leq 1$, as in the last assertion of Lemma 2.20, moves it further below $\kappa_c = \kappa_J$, where $J \geq 0$. The first lemma records the resulting lower bound, which holds for every order parameter.

Lemma 2.23. *Let $\alpha > 2$. Let $\gamma \in \mathcal{U}$, $\kappa \in \mathbb{R}$, and $\rho > 0$ with $\kappa + \rho < \kappa_c$. Then*

$$\mathcal{P}_\kappa(\gamma) \geq \frac{1}{2} \log \delta - \alpha \left(2 + \log_+ \frac{\sqrt{\delta}}{\rho} \right).$$

Proof. By Proposition 2.17, $\kappa + \rho < \kappa_c < 0$. If $Q = 0$, then $\gamma = \mathbf{1}_{[0,1]}$, $\delta = 1$, and $\mathcal{P}_\kappa(\gamma) = \alpha \log \Phi(-\kappa) \geq \alpha \log \frac{1}{2} > -2\alpha$ because $-\kappa > \rho > 0$. If $Q > 0$, then $0 < \sqrt{Q} \leq 1$ gives $(\kappa + \rho)/\sqrt{Q} \leq \kappa + \rho < \kappa_c = \kappa_J$ (Theorem 2.21), so $J^{(\kappa+\rho)/\sqrt{Q}}(\tilde{\xi}_\gamma) \geq 0$. By the last assertion of Lemma 2.20, $J_Q^{\kappa+\rho}(\xi_\gamma) \geq 0$, and (2.45) gives the claim. $\square$

Proposition 2.24. *Let $\alpha > 2$. There are constants $C, \eta_0 > 0$, depending only on α , such that for $0 < \eta \leq \eta_0$*

$$\alpha \log \eta - C \leq \mathcal{P}_*(\kappa_c - \eta) \leq \frac{1}{2} \log \eta + C, \quad (2.49)$$

and every $\gamma \in \mathcal{U}$ with $\mathcal{P}_{\kappa_c - \eta}(\gamma) \leq \mathcal{P}_*(\kappa_c - \eta) + 1$ satisfies $1 - Q_\gamma \leq C\eta^{1/(1+\alpha)}$.

Proof. We divide the proof into four steps. First we bound $\mathcal{P}_\kappa(\gamma)$ below in terms of the top of γ , for every order parameter. This bounds the tops of near-minimizers. Then we bound the mass and the slope of minimizers, and finally we bound the value below by distinguishing two cases for the top gap.

Put $C_* = C_{\kappa_c}$; since C_κ is increasing in κ , $C_\kappa \leq C_*$ for $\kappa \leq \kappa_c$. By Theorem 2.18 there is $\eta_1 \in (0, 1]$ such that $\mathcal{P}_*(\kappa_c - \eta) \leq \frac{1}{2} \log \eta + C_1$ for $0 < \eta \leq \eta_1$, with $C_1 = \frac{1}{2} \log(2\alpha C_*) + \frac{1}{2}$. This is the upper bound in (2.49).

STEP 1. *A lower bound in terms of the top.* We show that for all $\gamma \in \mathcal{U}$ and $\kappa \leq \kappa_c$,

$$\mathcal{P}_\kappa(\gamma) \geq \frac{1+\alpha}{2} \log \delta - C_2, \quad C_2 = \alpha(2 + \log 2 + C_*). \quad (2.50)$$

Lemma 2.23 at the margin $\kappa_c - \delta$ with $\rho = \delta/2$ gives, since $\sqrt{\delta}/\rho = 2/\sqrt{\delta} \geq 1$,

$$\mathcal{P}_{\kappa_c - \delta}(\gamma) \geq \frac{1}{2} \log \delta - 2\alpha - \alpha \log \frac{2}{\sqrt{\delta}} = \frac{1+\alpha}{2} \log \delta - \alpha(2 + \log 2).$$

If $\kappa \leq \kappa_c - \delta$, then $\mathcal{P}_\kappa(\gamma) \geq \mathcal{P}_{\kappa_c - \delta}(\gamma)$ because $\mathcal{P}_\cdot(\gamma)$ is decreasing. If $\kappa_c - \delta < \kappa \leq \kappa_c$, then $\mathcal{P}_\kappa(\gamma) \geq \mathcal{P}_{\kappa_c}(\gamma)$, and on $[\kappa_c - \delta, \kappa_c]$ the derivative of $\mathcal{P}_\cdot(\gamma)$ is at least $-\alpha C_*/\lambda(0) \geq -\alpha C_*/\delta$ by Lemma 2.14; integrating over this interval of length δ gives $\mathcal{P}_{\kappa_c}(\gamma) \geq \mathcal{P}_{\kappa_c - \delta}(\gamma) - \alpha C_*$. This proves (2.50).

STEP 2. *Tops of near-minimizers.* Let $0 < \eta \leq \eta_1$, put $k = \kappa_c - \eta$, and let $\gamma \in \mathcal{U}$ satisfy $\mathcal{P}_k(\gamma) \leq \mathcal{P}_*(k) + 1$. By (2.50) and the upper bound in (2.49), $\frac{1+\alpha}{2} \log \delta - C_2 \leq \frac{1}{2} \log \eta + C_1 + 1$, that is,

$$1 - Q_\gamma = \delta \leq C_3 \eta^{1/(1+\alpha)}, \quad C_3 = e^{2(C_1+C_2+1)/(1+\alpha)}.$$

We fix $\eta_0 \in (0, \eta_1]$ with $C_3 \eta_0^{1/(1+\alpha)} \leq \frac{1}{2}$. From now on $0 < \eta \leq \eta_0$, $k = \kappa_c - \eta$, and γ is a minimizer of $\mathcal{P}_k$ over $\mathcal{U}$, which exists by Theorem 2.16. We write δ, λ, b for its quantities at the margin k . By the bound just proved, $\delta \leq \frac{1}{2}$, that is, $Q \geq \frac{1}{2}$.

STEP 3. *The mass and the slope.* The minimizer γ minimizes $\mathcal{P}_k$ over $\mathcal{U}_r$ for every $r \in (Q, 1)$, so (2.35) and $C_k \leq C_*$ give $\lambda(0) \leq C_4 \sqrt{\delta}$ with $C_4 = \max\{1, 2\alpha C_*/s_\alpha\}$. By (2.32), $\lambda(0)b(0,0) \geq \mathbb{E}(k - \sqrt{Q}G)_+$. Since $k \geq \kappa_c - 1$, $\kappa_c - 1 < 0$, and $\frac{1}{2} \leq Q \leq 1$, we have $(\kappa_c - 1)/\sqrt{Q} \geq \sqrt{2}(\kappa_c - 1)$, and therefore

$$\lambda(0)b(0,0) \geq \sqrt{Q} m_1\left(\frac{\kappa_c - 1}{\sqrt{Q}}\right) \geq \frac{1}{\sqrt{2}} m_1(\sqrt{2}(\kappa_c - 1)) =: c_4 > 0.$$

STEP 4. *The lower bound in (2.49).* Put $s = \sqrt{\delta} \leq 1$. We distinguish two cases.

Case A: $s \geq \eta/4$. Lemma 2.23 at the margin k with $\rho = \eta/2$ gives $\mathcal{P}_*(k) = \mathcal{P}_k(\gamma) \geq \log s - 2\alpha - \alpha \log_+(2s/\eta)$. If $2s \geq \eta$, the right side equals $\alpha \log \eta + (1 - \alpha) \log s - \alpha(2 + \log 2) \geq \alpha \log \eta - \alpha(2 + \log 2)$, because $s \leq 1$ and $\alpha > 1$. If $2s < \eta$, it equals $\log s - 2\alpha \geq \log \eta - \log 4 - 2\alpha \geq \alpha \log \eta - \log 4 - 2\alpha$, because $\eta \leq 1$.

Case B: $s < \eta/4$. Here the margin shift in Lemma 2.23 costs nothing, but $\frac{1}{2} \log \delta$ may be very negative; the slope compensates. By (2.33) and Step 3,

$$\mathcal{P}_k(\gamma) \geq \mathcal{P}_{k+\eta/2}(\gamma) + \frac{\alpha\eta}{2} b(0,0) \geq \mathcal{P}_{k+\eta/2}(\gamma) + \frac{\alpha c_4 \eta}{2\lambda(0)}.$$

Lemma 2.23 at the margin $k + \eta/2 = \kappa_c - \eta/2$ with $\rho = \eta/4$ gives $\mathcal{P}_{k+\eta/2}(\gamma) \geq \log s - 2\alpha$, because $\sqrt{\delta}/\rho = 4s/\eta < 1$. With $\lambda(0) \leq C_4 s$ we obtain, with $c_5 = \alpha c_4/(2C_4)$,

$$\mathcal{P}_*(k) \geq \log s + \frac{c_5 \eta}{s} - 2\alpha \geq \log \eta + 1 + \log c_5 - 2\alpha,$$

since $s \mapsto \log s + c_5 \eta/s$ is minimal at $s = c_5 \eta$. As $\log \eta \geq \alpha \log \eta$, both cases give the lower bound in (2.49) with a constant depending only on α . $\square$

The exponent $1/(1+\alpha)$ for near-minimizers comes from the factor $\frac{1+\alpha}{2}$ in (2.50). The proposition does not identify the coefficient of $\log \eta$ in $\mathcal{P}_*(\kappa_c - \eta)$. For exact minimizers, Part I, Lemma 6.10 and Proposition 6.11, prove the bound on $1 - Q_\gamma$ for every $\alpha > 0$, directly from the value identity.

Replica-symmetric order parameters, the de Almeida–Thouless line, and the ordering of thresholds

This chapter studies the replica-symmetric order parameters $\mathbf{1}_{[q,1]}$ and the three margins κ_{RS} , κ_{d} , and κ_{c} for $\alpha = 2 + e$ with $e > 0$ small. The margin κ_{RS} is Gardner’s prediction, and Proposition 2.17 gives $\kappa_{\text{c}} \leq \kappa_{\text{RS}}$. §3.1 expands κ_{RS} near $\alpha = 2$, and §3.2 computes the functional, the function D , and the replicon eigenvalue at a replica-symmetric order parameter. In §3.3 an exact formula for the derivative of D along the replica-symmetric family (Lemma 3.5) shows that, for $\alpha < \pi^2/4$, every margin below κ_{RS} has exactly one stationary replica-symmetric order parameter, and Proposition 3.8 shows that its replicon eigenvalue changes sign at a single margin κ_{d} , with $\kappa_{\text{RS}} - \kappa_{\text{d}} = \theta_{\text{d}} e^2 + O(e^3 \log(1/e))$. §3.4 proves a duality bound for the zero-temperature functional J^k of Chapter 2 (Theorem 3.10). In §3.5 we construct an explicit dual certificate and an explicit negative value of J^k . This gives $\kappa_{\text{RS}} - C_{\text{ord}} e^3 \leq \kappa_{\text{c}} \leq \kappa_{\text{RS}} - c_{\text{cub}} e^3$ (Theorems 3.11 and 3.13); on the interval $(\kappa_{\text{d}}, \kappa_{\text{c}})$ no minimizer is replica symmetric (Corollary 3.14). In the terms of §1.3, the certified margin k_* of Theorem 3.11 enters Step 1 through the top-gap lower bound of Chapter 4; Theorem 3.11(iii) places every compact window Θ of rescaled margins below κ_{c} (Remark 3.12); and Steps 4 and 5 use the order parameter $\gamma_{\text{d}} = \mathbf{1}_{[q_{\text{d}},1]}$ of Proposition 3.8 and the asymptotics of $1 - q_{\text{d}}$.

3.1 Gardner’s prediction near $\alpha = 2$

We use the Gaussian functions R , $\bar{R}$, V , m , m_1 , and ϱ of §2.1 and the facts collected in Lemma 2.1. The next lemma expands Gardner’s prediction κ_{RS} near $\alpha = 2$. Its part (b) identifies the quantity $\alpha\Phi(\kappa_{\text{RS}}) - 1 \approx e/4$, which drives both the de Almeida–Thouless crossing and the cubic separation below.

Lemma 3.1. *There is $e_{\text{g}} > 0$ such that for $0 < e \leq e_{\text{g}}$ and $\alpha = 2 + e$:*

- (a) $-\frac{1}{2} < \kappa_{\text{RS}} < 0$ and $\frac{e}{4\alpha\phi_0} \leq |\kappa_{\text{RS}}| \leq \frac{e}{8m_1(-1/2)}$; moreover $\kappa_{\text{RS}} = -\frac{e}{8\phi_0} + O(e^2)$;
- (b) $\alpha\Phi(\kappa_{\text{RS}}) - 1 = \alpha|\kappa_{\text{RS}}| m_1(\kappa_{\text{RS}})$; in particular $0 < \alpha\Phi(\kappa_{\text{RS}}) - 1 \leq \alpha\phi_0|\kappa_{\text{RS}}|$, $\alpha\Phi(\kappa_{\text{RS}}) - 1 \geq \frac{m_1(-1/2)}{4\phi_0} e$, and $\alpha\Phi(\kappa_{\text{RS}}) - 1 = \frac{e}{4} + O(e^2)$;

$$(c) \quad \alpha m'(\kappa_{\text{RS}}) = 4\phi_0 + e(2\phi_0 - \frac{1}{4\phi_0}) + O(e^2).$$

Proof. (a) From $\alpha m(\kappa_{\text{RS}}) = 1$ and $m(0) = \frac{1}{2}$ we get $m(0) - m(\kappa_{\text{RS}}) = \frac{1}{2} - \frac{1}{2+e} = \frac{e}{2\alpha} > 0$, so $\kappa_{\text{RS}} < 0$. Since $m(-\frac{1}{2}) < \frac{1}{2}$, for small e we have $m(\kappa_{\text{RS}}) = \frac{1}{2+e} > m(-\frac{1}{2})$, that is, $\kappa_{\text{RS}} > -\frac{1}{2}$. By convexity of m and Lemma 2.1(g),

$$2m_1(-\frac{1}{2})|\kappa_{\text{RS}}| \leq m'(\kappa_{\text{RS}})|\kappa_{\text{RS}}| \leq m(0) - m(\kappa_{\text{RS}}) = \frac{e}{2\alpha} \leq m'(0)|\kappa_{\text{RS}}| = 2\phi_0|\kappa_{\text{RS}}|,$$

which gives the two-sided bound, since $\alpha \geq 2$. Taylor's formula with $0 < m'' \leq 2$ gives $|m(\kappa_{\text{RS}}) - \frac{1}{2} - 2\phi_0\kappa_{\text{RS}}| \leq \kappa_{\text{RS}}^2$, and $\frac{e}{2\alpha} = \frac{e}{4} + O(e^2)$; hence $\kappa_{\text{RS}} = -e/(8\phi_0) + O(e^2)$.

(b) The identity follows from $\alpha m(\kappa_{\text{RS}}) = 1$ and $\Phi - m = -km_1$. Since m_1 is increasing, $m_1(-\frac{1}{2}) \leq m_1(\kappa_{\text{RS}}) \leq m_1(0) = \phi_0$; combine with (a). For the expansion, $m_1(\kappa_{\text{RS}}) = \phi_0 + O(e)$ and $\alpha|\kappa_{\text{RS}}| = (2+e)(\frac{e}{8\phi_0} + O(e^2))$.

(c) Since $m'_1(0) = \Phi(0) = \frac{1}{2}$ and $m''_1 = \phi$ is bounded, $m'(\kappa_{\text{RS}}) = 2m_1(\kappa_{\text{RS}}) = 2\phi_0 + \kappa_{\text{RS}} + O(\kappa_{\text{RS}}^2)$; multiply by $\alpha = 2+e$ and insert (a). $\square$

3.2 Replica-symmetric order parameters

Definition 3.2. For $q \in [0, 1)$ the *replica-symmetric* (RS) order parameter is $\gamma_q := \mathbf{1}_{[q, 1]} \in \mathcal{U}$. Its measure is the point mass at q , so $q_0 = Q = q$ and $\delta = 1 - q =: \delta_q$. The *RS value* is $\mathcal{P}_\kappa^{\text{RS}}(q) := \mathcal{P}_\kappa(\gamma_q)$. For $q \in (0, 1)$ we use the parameters

$$k := \frac{\kappa}{\sqrt{q}}, \quad \beta := \sqrt{\frac{1-q}{q}}, \quad Z^q := \frac{\sqrt{q}G - \kappa}{\sqrt{\delta_q}} = \frac{G - k}{\beta},$$

and we define, for $k \in \mathbb{R}$ and $\beta > 0$,

$$F_{\text{RS}}(k, \beta) := \beta^2 \mathbb{E}R\left(\frac{G - k}{\beta}\right)^2, \quad H_{\text{RS}}(k, \beta) := \mathbb{E}V\left(\frac{G - k}{\beta}\right)^2.$$

We call $q \in (0, 1)$ *RS-stationary* at κ if $D_{\gamma_q}(q) = 0$. At an RS-stationary q the *replicon eigenvalue* is

$$\Lambda_{\text{R}} := \frac{1 - \alpha H_{\text{RS}}(k, \beta)}{4(1 - q)^2}.$$

When κ admits exactly one RS-stationary point, we call it the *stationary RS point* and denote it by $q^*(\kappa)$.

For γ_q the diffusion is a Brownian motion on $[0, q]$, so Z^q is the normalized top position $Z = (X_Q - \kappa)/\sqrt{\delta}$ of γ_q . The quantity H_{RS} is the Gaussian second moment of V ; it should not be confused with the obstacle potential $\mathcal{H}$. The next lemma identifies Λ_{R} with one half of the curvature of $\mathcal{H}_{\gamma_q}$ at the atom, as in (1.7). Its sign decides whether

moving a small part of the mass of the atom to a nearby point lowers the functional. At an RS-stationary point, $\alpha H_{\text{RS}}(k, \beta) = 1$ is the de Almeida–Thouless condition recalled in §1.1, and the stationarity condition $\alpha F_{\text{RS}}(k, \beta) = 1$ is [59, (43)].

Lemma 3.3. *Let $\alpha > 0$, $\kappa \in \mathbb{R}$, $q \in (0, 1)$, $\gamma = \gamma_q$, and $r \in (q, 1)$.*

- (a) $\mathcal{P}_\kappa^{\text{RS}}(q) = \alpha \mathbb{E} \log \Phi\left(\frac{\sqrt{q}G - \kappa}{\sqrt{1-q}}\right) + \frac{q}{2(1-q)} + \frac{1}{2} \log(1-q)$. The same formula holds at $q = 0$.
- (b) $D_\gamma(q) = \frac{q(\alpha F_{\text{RS}}(k, \beta) - 1)}{(1-q)^2}$ and $S_\gamma(q) = \alpha H_{\text{RS}}(k, \beta)$. Moreover $D_\gamma \in C^1([0, r])$ and $D'_\gamma = (S_\gamma - 1)/\lambda_\gamma^2$.
- (c) $q \mapsto \mathcal{P}_\kappa^{\text{RS}}(q)$ is C^1 on $(0, 1)$ with $\frac{d}{dq} \mathcal{P}_\kappa^{\text{RS}}(q) = -\frac{1}{2} D_{\gamma_q}(q)$.
- (d) If q is RS-stationary, then $\mathcal{H}_\gamma(s) - \mathcal{H}_\gamma(q) = \Lambda_R(s - q)^2 + o((s - q)^2)$ as $s \rightarrow q$.
- (e) If q is RS-stationary and $\Lambda_R < 0$, then q is a strict local maximum of $\mathcal{H}_\gamma$. For every $q' \neq q$ sufficiently close to q and all sufficiently small $t > 0$,

$$\mathcal{P}_\kappa((1-t)\gamma_q + t\gamma_{q'}) < \mathcal{P}_\kappa(\gamma_q).$$

In particular γ_q is neither a global minimizer of $\mathcal{P}_\kappa$ nor a local minimizer for the L^1 distance.

Proof. (a) Since $\gamma = 0$ on $[0, q]$, Lemma 2.5(a) with $m = 0$ and (2.9) give $u_\gamma(0, 0; \kappa) = \mathbb{E} u_\gamma(q, B_q; \kappa) = \mathbb{E} \log \Phi(Z^q)$. Since $\lambda_\gamma = 1 - q$ on $[0, q]$ and $\lambda_\gamma(t) = 1 - t$ on $[q, 1]$,

$$\text{CS}(\gamma) = \frac{1}{2} \int_0^q \left(\frac{1}{1-q} - \frac{1}{1-t} \right) dt = \frac{q}{2(1-q)} + \frac{1}{2} \log(1-q).$$

At $q = 0$ both sides equal $\alpha \log \Phi(-\kappa)$.

(b) By (2.10) with $z = (x - \kappa)/\sqrt{1-q}$ we have $b(q, x) = R(z)/\sqrt{1-q}$ and $c(q, x) = V(z)/(1-q)$, and $X_q = B_q$ has the law of $\sqrt{q}G$. Hence, as $\lambda_\gamma = 1 - q$ on $[0, q]$,

$$D_\gamma(q) = \frac{\alpha \mathbb{E} R(Z^q)^2}{1-q} - \frac{q}{(1-q)^2} = \frac{\alpha(1-q) \mathbb{E} R(Z^q)^2 - q}{(1-q)^2},$$

and $\alpha(1-q) \mathbb{E} R(Z^q)^2 = q \alpha F_{\text{RS}}(k, \beta)$ because $\beta^2 = (1-q)/q$. Similarly $S_\gamma(q) = \alpha(1-q)^2 \mathbb{E} c(q, X_q)^2 = \alpha \mathbb{E} V(Z^q)^2$. The last claim is Lemma 2.9(c)(i).

(c) The derivative of $\frac{q}{2(1-q)} + \frac{1}{2} \log(1-q)$ is $\frac{q}{2(1-q)^2}$. Next

$$\frac{dZ^q}{dq} = \frac{G}{2\sqrt{q(1-q)}} + \frac{Z^q}{2(1-q)},$$

and Gaussian integration by parts with $R' = -V$ gives $\mathbb{E}[R(Z^q)G] = -\sqrt{q/(1-q)} \mathbb{E} V(Z^q)$. Since $V(z) = zR(z) + R(z)^2$,

$$\frac{d}{dq} \mathbb{E} \log \Phi(Z^q) = \mathbb{E} \left[R(Z^q) \frac{dZ^q}{dq} \right] = \frac{\mathbb{E}[R(Z^q)Z^q - V(Z^q)]}{2(1-q)} = -\frac{\mathbb{E} R(Z^q)^2}{2(1-q)}.$$

Differentiation under $\mathbb{E}$ is justified by $0 < R(z) \leq |z| + 1$. Summing,

$$\frac{d}{dq} \mathcal{P}_\kappa^{\text{RS}}(q) = -\frac{\alpha(1-q)\mathbb{E}R(Z^q)^2 - q}{2(1-q)^2} = -\frac{1}{2}D_{\gamma_q}(q).$$

(d) By (b), $\mathcal{H}'_\gamma = -D_\gamma/2$ is C^1 near q , $\mathcal{H}'_\gamma(q) = 0$, and $\mathcal{H}''_\gamma(q) = -\frac{1}{2}D'_\gamma(q) = \frac{1-S_\gamma(q)}{2(1-q)^2} = 2\Lambda_R$. Taylor's formula gives the claim.

(e) By (d), $\mathcal{H}_\gamma(s) < \mathcal{H}_\gamma(q)$ for $s \neq q$ near q . For $q' < q$ close to q , $\gamma_{q'} - \gamma_q = \mathbf{1}_{[q',q]}$, and the first-variation formula (2.28) of Proposition 2.11(a) (both order parameters lie in $\mathcal{U}_{r'}$ for $r' \in (\max\{q, q'\}, 1)$) gives the right derivative

$$\left. \frac{d}{dt} \mathcal{P}_\kappa((1-t)\gamma_q + t\gamma_{q'}) \right|_{t=0+} = \frac{1}{2} \int_{q'}^q D_\gamma(s) ds = \mathcal{H}_\gamma(q') - \mathcal{H}_\gamma(q) < 0;$$

the case $q' > q$ is identical with $\gamma_{q'} - \gamma_q = -\mathbf{1}_{[q,q']}$. Moreover $\|(1-t)\gamma_q + t\gamma_{q'} - \gamma_q\|_{L^1} = t|q' - q| \rightarrow 0$. $\square$

Part (e) is the analogue for $\mathcal{P}_\kappa$ of the instability of the replica-symmetric solution beyond the de Almeida–Thouless line for SK-type models [129], [77, Theorem 1.8], recalled in Chapter 1.

3.3 The stationary curve and the crossing

We now locate the RS-stationary points. By Lemma 3.3(b), q is RS-stationary exactly when $\alpha F_{\text{RS}}(k, \beta) = 1$. Uniqueness of the stationary point comes from an exact formula for the derivative of $D_{\gamma_q}(q)$ in q (Lemma 3.5); the crossing is then located by following the replicon eigenvalue along the stationary curve, parametrized by β .

Lemma 3.4. *Let $k \in \mathbb{R}$ and $\beta > 0$, and write $Z := (G - k)/\beta$.*

- (a) $F_{\text{RS}}(k, \beta) = m(k) + \beta^2 \mathbb{E}q(Z)$; hence $m(k) < F_{\text{RS}}(k, \beta) \leq m(k) + M_q \beta^2$.
- (b) $|F_{\text{RS}}(k, \beta) - m(k) - 2\beta^2 \Phi(k)| \leq C_F \beta^3$, with $C_F := \phi_0 \|q - 2 \cdot \mathbf{1}_{(-\infty, 0)}\|_{L^1}$.
- (c) F_{RS} is continuous on $\mathbb{R} \times (0, \infty)$ and strictly increasing in each variable. For fixed β , $F_{\text{RS}}(\cdot, \beta)$ increases from 0 to ∞ .
- (d) For $\alpha > 0$ and $\beta > 0$ there is a unique $K_\alpha(\beta)$ with $\alpha F_{\text{RS}}(K_\alpha(\beta), \beta) = 1$. The function K_α is continuous and strictly decreasing, and $m(K_\alpha(\beta)) \in [\frac{1}{\alpha} - M_q \beta^2, \frac{1}{\alpha}]$. In particular $K_\alpha(\beta) < \kappa_{\text{RS}}(\alpha)$, and $K_\alpha(\beta) \uparrow \kappa_{\text{RS}}(\alpha)$ as $\beta \downarrow 0$.

Proof. (a) Since $\beta Z_- = (k - G)_+$, we have $\beta^2 \mathbb{E}Z_-^2 = m(k)$ and $\beta^2 \mathbb{E}R(Z)^2 = m(k) + \beta^2 \mathbb{E}q(Z)$. Use $0 < q \leq M_q$.

(b) Z has density $z \mapsto \beta \phi(k + \beta z) \leq \beta \phi_0$, and $\mathbb{P}(Z < 0) = \Phi(k)$. Hence $|\mathbb{E}q(Z) - 2\Phi(k)| \leq \beta \phi_0 \|q - 2 \cdot \mathbf{1}_{(-\infty, 0)}\|_{L^1}$, which is finite by Lemma 2.1(f).

(c) $F_{\text{RS}}(k, \beta) = \mathbb{E} f_{k, \beta}(G)^2$ with $f_{k, \beta}(g) := \beta R((g - k)/\beta) > 0$. For fixed g , $f_{k, \beta}(g)$ is strictly increasing in k because R is decreasing. It is strictly increasing in β because, with $z = (g - k)/\beta$,

$$\partial_{\beta} f_{k, \beta}(g) = R(z) - zR'(z) = R(z)(1 + z\bar{R}(z)) = R(z) \mathbb{E} T_z^2 > 0$$

by Lemma 2.1(a). Since $R(z) \leq z_- + R(0)$ (Lemma 2.1(a)), we have $f_{k, \beta}(g) \leq (k - g)_+ + \beta R(0)$, and dominated convergence gives continuity and $F_{\text{RS}} \rightarrow 0$ as $k \rightarrow -\infty$, where $f_{k, \beta} \rightarrow 0$ pointwise. Finally $F_{\text{RS}} \geq m(k) \rightarrow \infty$ as $k \rightarrow \infty$.

(d) Existence and uniqueness follow from (c). If $\beta < \beta'$, then $\alpha F_{\text{RS}}(K_{\alpha}(\beta), \beta') > 1$, so $K_{\alpha}(\beta') < K_{\alpha}(\beta)$. For continuity at β , let $\varepsilon > 0$; the strict inequalities $\alpha F_{\text{RS}}(K_{\alpha}(\beta) - \varepsilon, \cdot) < 1 < \alpha F_{\text{RS}}(K_{\alpha}(\beta) + \varepsilon, \cdot)$ hold at β , hence near β , and there $|K_{\alpha} - K_{\alpha}(\beta)| < \varepsilon$. By (a), $m(K_{\alpha}(\beta)) = \frac{1}{\alpha} - \beta^2 \mathbb{E} \varrho(Z) \in [\frac{1}{\alpha} - M_{\varrho} \beta^2, \frac{1}{\alpha})$. Since m is continuous and strictly increasing and $\alpha m(\kappa_{\text{RS}}) = 1$, this gives $K_{\alpha}(\beta) < \kappa_{\text{RS}}$ and $K_{\alpha}(\beta) \rightarrow \kappa_{\text{RS}}$ as $\beta \rightarrow 0$. $\square$

The next lemma is the structural input for the uniqueness of the stationary RS point: along the RS family, every zero of D is a strict down-crossing. It holds for every margin below κ_{RS} , as soon as $\alpha < \pi^2/4$.

Lemma 3.5. *Let $0 < \alpha < \pi^2/4$ and $\kappa < \kappa_{\text{RS}}(\alpha)$. For $q \in [0, 1)$ put $\delta_q = 1 - q$, $Z^q = (\sqrt{q} G - \kappa)/\sqrt{\delta_q}$,*

$$D_{\text{RS}}(q; \kappa) := \frac{\alpha}{\delta_q} \mathbb{E} R(Z^q)^2 - \frac{q}{\delta_q^2}, \quad S_{\text{RS}}(q; \kappa) := \alpha \mathbb{E} V(Z^q)^2. \quad (3.1)$$

Then $D_{\text{RS}}(q; \kappa) = D_{\gamma_q}(q)$ and $S_{\text{RS}}(q; \kappa) = S_{\gamma_q}(q)$.

(a) $q \mapsto D_{\text{RS}}(q; \kappa)$ is C^1 on $[0, 1)$, and

$$\delta_q^2 \partial_q D_{\text{RS}} = \alpha \mathbb{E} [V^2 - 2R^2(1 - V)](Z^q) - 1 + 2\delta_q D_{\text{RS}}. \quad (3.2)$$

(b) *There is exactly one $q^*(\kappa) \in (0, 1)$ with $D_{\text{RS}}(q^*(\kappa); \kappa) = 0$. Moreover $D_{\text{RS}} > 0$ on $[0, q^*(\kappa))$ and $D_{\text{RS}} < 0$ on $(q^*(\kappa), 1)$.*

(c) $\kappa \mapsto q^*(\kappa)$ is continuous and strictly increasing on $(-\infty, \kappa_{\text{RS}})$.

Proof. For $q > 0$ the identities $D_{\text{RS}} = D_{\gamma_q}(q)$ and $S_{\text{RS}} = S_{\gamma_q}(q)$ are contained in the proof of Lemma 3.3(b). For $q = 0$ the order parameter $\gamma_0 = \mathbf{1}_{[0, 1]}$ has $u(q, x) = \log \Phi((x - \kappa)/\sqrt{1 - q})$ on $[0, 1)$ by (2.9), so $D_{\gamma_0}(0) = \alpha b(0, 0)^2 = \alpha R(-\kappa)^2$ and $S_{\gamma_0}(0) = \alpha V(-\kappa)^2$, as claimed.

(a) Let $\bar{b}(q, x) := \partial_x \log \Phi((x - \kappa)/\sqrt{\delta_q}) = R((x - \kappa)/\sqrt{\delta_q})/\sqrt{\delta_q}$, so that $D_{\text{RS}} = \alpha \mathbb{E} \bar{b}(q, B_q)^2 - q/\delta_q^2$ for a Brownian motion B . The function $\Phi((x - \kappa)/\sqrt{\delta_q}) = \mathbb{P}(x + \sqrt{\delta_q} G \geq \kappa)$ solves the backward heat equation, so its logarithm $\bar{u}$ solves $\bar{u}_q = -\frac{1}{2}(\bar{u}_{xx} + \bar{u}_x^2)$, and $\bar{b} = \bar{u}_x$ solves $\bar{b}_q = -\frac{1}{2}(\bar{b}_{xx} + 2\bar{b}\bar{b}_x)$; also $\bar{b}_x = -V((x - \kappa)/\sqrt{\delta_q})/\delta_q$. On

$[0, r] \times \mathbb{R}$ with $r < 1$ these functions are smooth, $\bar{b}$ grows at most linearly in x (as $R(z) \leq z_- + R(0)$), and $\bar{b}_x, \bar{b}_{xx}$ are bounded, because V and $V' = R(1 - V - \bar{R}^2)$ are bounded by Lemma 2.1(a),(b),(e). By Itô's formula,

$$\frac{d}{dq} \mathbb{E} \bar{b}(q, B_q)^2 = \mathbb{E} [2\bar{b}\bar{b}_q + \bar{b}_x^2 + \bar{b}\bar{b}_{xx}] = \mathbb{E} [\bar{b}_x^2 - 2\bar{b}^2\bar{b}_x] = \delta_q^{-2} \mathbb{E} [V^2 + 2R^2V](Z^q),$$

which is continuous in q . Since $\partial_q(q/\delta_q^2) = \delta_q^{-2} + 2q\delta_q^{-3}$ and $2q\delta_q^{-3} = 2\delta_q^{-2}(\alpha\mathbb{E}R(Z^q)^2 - \delta_q D_{\text{RS}})$ by the definition of D_{RS} , (3.2) follows.

(b) Put $\mathfrak{p} := V^2 - 2R^2(1 - V)$. For $z \geq 0$, $\mathfrak{p} \leq V^2 \leq V(0)^2 = 4/\pi^2$, because V is decreasing and $V(0) = R(0)^2 = 2/\pi$. For $z < 0$, (2.3) gives $2R^2(1 - V) \geq \frac{2}{3}R^2\bar{R}^2 = \frac{2}{3}V^2$, so $\mathfrak{p} \leq \frac{1}{3}V^2 < \frac{1}{3} < \frac{4}{\pi^2}$. Hence at every zero q of $D_{\text{RS}}(\cdot; \kappa)$, (a) gives $\delta_q^2 \partial_q D_{\text{RS}} \leq 4\alpha/\pi^2 - 1 < 0$. A C^1 function on an interval whose zeros are all strict down-crossings has at most one zero. Indeed, suppose $q_1 < q_2$ are zeros, and let q_3 be the smallest zero in $(q_1, q_2]$, which exists because the function is negative on a right neighborhood of q_1 . Then the function is negative on (q_1, q_3) , so its derivative at q_3 is nonnegative, a contradiction. For existence, $D_{\text{RS}}(0; \kappa) = \alpha R(-\kappa)^2 > 0$, while by Lemma 3.3(b) and Lemma 3.4(a), $\delta_q^2 D_{\text{RS}} = q(\alpha F_{\text{RS}}(\kappa/\sqrt{q}, \beta) - 1)$ with $\beta^2 = \delta_q/q$ and $m(\kappa/\sqrt{q}) < F_{\text{RS}}(\kappa/\sqrt{q}, \beta) \leq m(\kappa/\sqrt{q}) + M_q\beta^2$. Thus $\delta_q^2 D_{\text{RS}} \rightarrow \alpha m(\kappa) - 1 < 0$ as $q \uparrow 1$. The intermediate value theorem and uniqueness give the sign pattern.

(c) For fixed q , Z^q decreases in κ and R is positive and decreasing, so $D_{\text{RS}}(q; \kappa)$ is strictly increasing in κ ; it is continuous in κ by dominated convergence. If $\kappa < \kappa' < \kappa_{\text{RS}}$, then $D_{\text{RS}}(q^*(\kappa); \kappa') > 0$, so $q^*(\kappa) < q^*(\kappa')$ by (b). For continuity at κ , let $\varepsilon > 0$. The strict inequalities $D_{\text{RS}}(q^*(\kappa) - \varepsilon; \cdot) > 0 > D_{\text{RS}}(q^*(\kappa) + \varepsilon; \cdot)$ (the first omitted if $q^*(\kappa) \leq \varepsilon$) hold at κ , hence near κ , and there $|q^* - q^*(\kappa)| < \varepsilon$ by (b). $\square$

The stability function H_{RS} is expanded next. Its first-order term in β comes from the signed area $\int (V^2 - \mathbf{1}_{\{z < 0\}}) = -c_0$ of Lemma 2.1(f).

Lemma 3.6. *There is a numerical constant $C_H < \infty$ such that, for $|k| \leq 1$ and $0 < \beta \leq \frac{1}{2}$,*

$$\begin{aligned} |H_{\text{RS}}(k, \beta) - \Phi(k) + c_0\beta\phi(k)| &\leq C_H\beta^2 \log(1/\beta), \\ |\partial_\beta H_{\text{RS}}(k, \beta) + c_0\phi(k)| &\leq C_H\beta \log(1/\beta). \end{aligned}$$

Moreover $H_{\text{RS}}(\cdot, \beta)$ is nondecreasing for every $\beta > 0$.

Proof. Put $\mathfrak{v} := V^2 - \mathbf{1}_{\{z < 0\}}$, so that $|\mathfrak{v}(z)| \leq C_g/(1+z^2)$ and $\int \mathfrak{v} = -c_0$ by Lemma 2.1(f). Since $Z = (G - k)/\beta$ has density $\beta\phi(k + \beta z)$ and $\mathbb{P}(Z < 0) = \Phi(k)$,

$$\begin{aligned} H_{\text{RS}}(k, \beta) &= \Phi(k) + \beta \int_{\mathbb{R}} \phi(k + \beta z) \mathfrak{v}(z) dz \\ &= \Phi(k) - c_0\beta\phi(k) + \beta \int_{\mathbb{R}} (\phi(k + \beta z) - \phi(k)) \mathfrak{v}(z) dz. \end{aligned}$$

Since $|\phi'| \leq \phi_0$, $|\phi(k + \beta z) - \phi(k)| \leq \phi_0 \min\{1, \beta|z|\}$, and

$$\begin{aligned} \int_{\mathbb{R}} \frac{\min\{1, \beta|z|\}}{1 + z^2} dz &\leq \beta \int_{|z| \leq 1/\beta} \frac{|z|}{1 + z^2} dz + \int_{|z| > 1/\beta} \frac{dz}{z^2} \\ &= \beta \log(1 + \beta^{-2}) + 2\beta \leq C\beta \log(1/\beta) \end{aligned}$$

for $\beta \leq \frac{1}{2}$. This proves the first bound. Differentiating under the integral, which is justified by the same majorants,

$$\partial_{\beta} H_{\text{RS}} = \int_{\mathbb{R}} \phi(k + \beta z) \mathfrak{v}(z) dz + \beta \int_{\mathbb{R}} z \phi'(k + \beta z) \mathfrak{v}(z) dz.$$

The first integral is $-c_0 \phi(k) + O(\beta \log(1/\beta))$ by the estimate above. Since $|z|/(1 + z^2) \leq \beta$ for $|z| > 1/\beta$ and $\int_{\mathbb{R}} |\phi'(k + \beta z)| dz = \beta^{-1} \|\phi'\|_{L^1} = 2\phi_0 \beta^{-1}$, the second is at most

$$\beta C_g \left[\phi_0 \int_{|z| \leq 1/\beta} \frac{|z|}{1 + z^2} dz + \beta \int_{|z| > 1/\beta} |\phi'(k + \beta z)| dz \right] \leq C\beta \log(1/\beta).$$

Finally, $Z = (G - k)/\beta$ decreases in k and V is positive and decreasing, so $V(Z)^2$ increases in k . $\square$

For $\alpha > 0$ and $\beta > 0$ put

$$\kappa_{\alpha}(\beta) := \frac{K_{\alpha}(\beta)}{\sqrt{1 + \beta^2}}, \quad \omega_{\alpha}(\beta) := \alpha H_{\text{RS}}(K_{\alpha}(\beta), \beta) - 1.$$

For $q \in (0, 1)$ write $\beta(q) = \sqrt{(1 - q)/q}$, so that $\kappa/\sqrt{q} = \kappa\sqrt{1 + \beta(q)^2}$. By Lemma 3.3(b) and part (a) of the next lemma, $\omega_{\alpha}(\beta(q)) = S_{\gamma_q}(q) - 1 = -4(1 - q)^2 \Lambda_{\text{R}}$ whenever q is RS-stationary.

Lemma 3.7. (a) For every $\alpha > 0$, $\kappa \in \mathbb{R}$, and $q \in (0, 1)$, $\text{sign } D_{\gamma_q}(q) = \text{sign}(\kappa - \kappa_{\alpha}(\beta(q)))$.

Thus q is RS-stationary at κ exactly when $\kappa = \kappa_{\alpha}(\beta(q))$, and then $k = K_{\alpha}(\beta)$ and $\alpha H_{\text{RS}}(k, \beta) - 1 = \omega_{\alpha}(\beta)$.

(b) There are numerical constants $\beta_0 \in (0, \frac{1}{4}]$, $e_a > 0$, and $C_c < \infty$ such that for $0 < e \leq e_a$, $\alpha = 2 + e$, and $0 < \beta' < \beta \leq \beta_0$: $|K_{\alpha}(\beta)| \leq \frac{1}{4}$,

$$\begin{aligned} \left| \kappa_{\text{RS}} - \kappa_{\alpha}(\beta) - \frac{\beta^2}{2\phi_0} \right| &\leq C_c(e\beta^2 + \beta^3) \leq \frac{\beta^2}{4\phi_0}, \\ \left| \omega_{\alpha}(\beta) - \frac{e}{4} + 2c_0\phi_0\beta \right| &\leq C_c(e^2 + e\beta + \beta^2 \log(1/\beta)), \end{aligned} \tag{3.3}$$

and

$$\omega_{\alpha}(\beta) \leq \omega_{\alpha}(\beta') - \frac{1}{2}c_0\phi_0(\beta - \beta'). \tag{3.4}$$

Moreover ω_{α} is continuous on $(0, \infty)$ and $\omega_{\alpha}(0+) = \alpha\Phi(\kappa_{\text{RS}}) - 1$.

Proof. (a) By Lemma 3.3(b), $\text{sign } D_{\gamma_q}(q) = \text{sign}(\alpha F_{\text{RS}}(k, \beta) - 1)$ with $k = \kappa \sqrt{1 + \beta^2}$. Since $F_{\text{RS}}(\cdot, \beta)$ is strictly increasing and $\alpha F_{\text{RS}}(K_\alpha(\beta), \beta) = 1$, this is the sign of $\kappa \sqrt{1 + \beta^2} - K_\alpha(\beta)$.

(b) *Size of K .* Fix $\beta_1 \in (0, \frac{1}{4}]$ with $M_\varrho \beta_1^2 \leq \frac{1}{2}(\frac{1}{2} - m(-\frac{1}{2}))$ and $M_\varrho \beta_1^2 \leq \frac{1}{4}m_1(-\frac{1}{2})$, and let $e_a \leq e_g$ be so small that $\frac{1}{2+e} - \frac{1}{2} \geq -\frac{1}{2}(\frac{1}{2} - m(-\frac{1}{2}))$ and $|\kappa_{\text{RS}}| \leq \frac{1}{8}$ for $e \leq e_a$ (Lemma 3.1). Let $0 < \beta \leq \beta_1$ and $K = K_\alpha(\beta)$. By Lemma 3.4(d), $m(K) \geq \frac{1}{\alpha} - M_\varrho \beta^2 \geq m(-\frac{1}{2})$, so $K \geq -\frac{1}{2}$, and $m(\kappa_{\text{RS}}) - m(K) \leq M_\varrho \beta^2$. Since $m' = 2m_1 \geq 2m_1(-\frac{1}{2})$ on $[-\frac{1}{2}, 0]$,

$$0 \leq \kappa_{\text{RS}} - K \leq \frac{M_\varrho \beta^2}{2m_1(-\frac{1}{2})} \leq \frac{1}{8};$$

hence $|K| \leq \frac{1}{4}$.

Expansions. Subtract $\alpha m(\kappa_{\text{RS}}) = 1$ from $\alpha F_{\text{RS}}(K, \beta) = 1$ and use Lemma 3.4(b):

$$\alpha(m(\kappa_{\text{RS}}) - m(K)) = 2\alpha\beta^2\Phi(K) + O(\beta^3).$$

The left side equals $\alpha m'(\xi)(\kappa_{\text{RS}} - K)$ for some $\xi \in [K, \kappa_{\text{RS}}]$, and $|\xi - \kappa_{\text{RS}}| \leq C\beta^2$. By Lemma 3.1(b),(c) and $|m''|, |\Phi'| \leq 2$, $\alpha m'(\xi) = 4\phi_0 + O(e + \beta^2)$ and $\alpha\Phi(K) = 1 + O(e + \beta^2)$. Dividing, $\kappa_{\text{RS}} - K = \frac{\beta^2}{2\phi_0} + O(e\beta^2 + \beta^3)$. Finally $K - \kappa_\alpha(\beta) = K(1 - (1 + \beta^2)^{-1/2})$ has absolute value at most $|K|\beta^2/2$, and $|K| \leq |\kappa_{\text{RS}}| + C\beta^2 \leq C(e + \beta^2)$. For ω_α , Lemma 3.6 (applicable as $|K| \leq \frac{1}{4}$ and $\beta \leq \frac{1}{4}$) gives $\omega_\alpha(\beta) = \alpha\Phi(K) - 1 - \alpha c_0\beta\phi(K) + O(\beta^2 \log(1/\beta))$. Here $\alpha\Phi(K) - 1 = \alpha\Phi(\kappa_{\text{RS}}) - 1 + O(\beta^2) = \frac{e}{4} + O(e^2 + \beta^2)$ by Lemma 3.1(b), and $\alpha c_0\phi(K) = 2c_0\phi_0 + O(e + K^2) = 2c_0\phi_0 + O(e + \beta^2)$. This defines C_c , for $\beta \leq \beta_1$ and $e \leq e_a$. Now fix $\beta_0 \leq \beta_1$ with

$$C_c\beta_0 \leq \frac{1}{8\phi_0}, \quad C_H\beta_0 \log(1/\beta_0) \leq \frac{1}{4}c_0\phi_0,$$

and decrease e_a so that $C_ce_a \leq \frac{1}{8\phi_0}$; this gives the second inequality in the first display of (3.3).

Monotonicity and limits. Let $0 < \beta' < \beta \leq \beta_0$ and $K' = K_\alpha(\beta')$. By Lemma 3.4(d) and Lemma 3.6, $K_\alpha(\beta) < K'$ and $H_{\text{RS}}(\cdot, \beta)$ is nondecreasing, so

$$\begin{aligned} H_{\text{RS}}(K_\alpha(\beta), \beta) &\leq H_{\text{RS}}(K', \beta) = H_{\text{RS}}(K', \beta') + \int_{\beta'}^{\beta} \partial_b H_{\text{RS}}(K', b) db \\ &\leq H_{\text{RS}}(K', \beta') - \frac{1}{4}c_0\phi_0(\beta - \beta'). \end{aligned}$$

Indeed, for $b \leq \beta_0$, $\partial_b H_{\text{RS}}(K', b) \leq -c_0\phi(K') + C_H b \log(1/b) \leq -\frac{1}{2}c_0\phi_0 + \frac{1}{4}c_0\phi_0$ by Lemma 3.6, since $\phi(K') \geq \phi(\frac{1}{4}) \geq \frac{1}{2}\phi_0$, $b \log(1/b)$ is increasing on $(0, \frac{1}{4}]$, and β_0 was chosen accordingly. Multiplying by $\alpha \geq 2$ gives (3.4). Since $K_\alpha(\beta) \rightarrow \kappa_{\text{RS}}$ (Lemma 3.4(d)) and $H_{\text{RS}}(k, \beta) \rightarrow \Phi(k)$ uniformly for $|k| \leq 1$ (Lemma 3.6), $\omega_\alpha(0+) = \alpha\Phi(\kappa_{\text{RS}}) - 1$. Continuity of ω_α follows from that of K_α and of H_{RS} , the latter by dominated convergence. $\square$

We can now describe the replica-symmetric family near $\alpha = 2$. Part (i) is global in the margin; parts (ii)–(iv) locate the crossing and expand the stationary point and the replicon eigenvalue in the window of margins $\kappa_e(\theta) = \kappa_{\text{RS}}(2 + e) - e^2\theta$ used in the rest of this part.

Proposition 3.8. *There are numerical constants $e_1 > 0$ and $c_2 > 0$ such that for every $0 < e < e_1$ and $\alpha = 2 + e$ the following hold.*

- (i) *Stationary points. For every $\kappa < \kappa_{\text{RS}}$, in particular for $\kappa \in [\kappa_{\text{RS}} - c_2, \kappa_{\text{RS}})$, there is exactly one RS-stationary point $q^*(\kappa) \in (0, 1)$. Moreover $D_{\gamma_q}(q) > 0$ for $q \in [0, q^*)$ and $D_{\gamma_q}(q) < 0$ for $q \in (q^*, 1)$; $q^*(\kappa)$ is the unique minimizer of $\mathcal{P}_{\kappa}^{\text{RS}}$ on $[0, 1]$; and $\kappa \mapsto q^*(\kappa)$ is continuous and strictly increasing.*
- (ii) *The crossing and the replicon eigenvalue. There is exactly one $\kappa_d = \kappa_d(\alpha) \in (\kappa_{\text{RS}} - c_2, \kappa_{\text{RS}})$ at which $\alpha H_{\text{RS}} = 1$ at $q^*(\kappa_d)$. Put $q_d := q^*(\kappa_d)$ and $\gamma_d := \mathbf{1}_{[q_d, 1]}$. The replicon eigenvalue $\Lambda_{\text{R}}(\kappa)$ of Definition 3.2 at $q^*(\kappa)$ satisfies*

$$\Lambda_{\text{R}}(\kappa) < 0 \quad (\kappa_d < \kappa < \kappa_{\text{RS}}), \quad \Lambda_{\text{R}}(\kappa_d) = 0, \quad \Lambda_{\text{R}}(\kappa) > 0 \quad (\kappa_{\text{RS}} - c_2 \leq \kappa < \kappa_d).$$

Consequently, for $\kappa \in (\kappa_d, \kappa_{\text{RS}})$ the stationary replica-symmetric order parameter $\gamma_{q^*(\kappa)}$ is unstable in the sense of Lemma 3.3(e).

- (iii) *Asymptotics. As $e \downarrow 0$, with $\beta_d := \beta(q_d)$,*

$$\begin{aligned} \kappa_{\text{RS}} - \kappa_d &= \theta_d e^2 + O(e^3 \log(1/e)), \quad \theta_d = \frac{(2\pi)^{3/2}}{128 c_0^2}, \\ \beta_d &= \frac{e}{8c_0\phi_0} + O(e^2 \log(1/e)), \end{aligned}$$

and the rescaled top gap of γ_d satisfies

$$d_d := \frac{1 - q_d}{e^2} = \frac{1}{64c_0^2\phi_0^2} + O(e \log(1/e)) = 2\phi_0\theta_d + O(e \log(1/e)).$$

- (iv) *The stationary point in the window. For every compact $\Theta \subset (0, \infty)$ there are $e_{\Theta} > 0$ and $C_{\Theta} < \infty$ such that, for $0 < e < e_{\Theta}$ and $\theta \in \Theta$, $\kappa_e(\theta) \in [\kappa_{\text{RS}} - c_2, \kappa_{\text{RS}})$ and*

$$\left| \frac{1 - q^*(\kappa_e(\theta))}{e^2} - 2\phi_0\theta \right| \leq C_{\Theta}e.$$

Moreover, with $t(\theta) := 2\phi_0\sqrt{2\phi_0\theta}$, the value H_{RS} at $q^*(\kappa_e(\theta))$ satisfies

$$\left| (1 - \alpha H_{\text{RS}}) - e \left(c_0 t(\theta) - \frac{1}{4} \right) \right| \leq C_{\Theta}e^2 \log(1/e).$$

In particular $c_0 t(\theta_d) = \frac{1}{4}$, that is, $t(\theta_d) = t_d = 1/(4c_0)$.

Proof. Let β_0, e_a, C_c be as in Lemma 3.7(b), put $c_2 := \beta_0^2/(4\phi_0)$, and let $e_1 \leq \min\{e_a, \frac{1}{4}\}$, to be decreased finitely often below. Then $\alpha = 2 + e < \pi^2/4$.

(i) By Lemma 3.5, applicable because $\alpha < \pi^2/4$, every $\kappa < \kappa_{RS}$ has exactly one RS-stationary point $q^*(\kappa)$, with the stated sign pattern, and q^* is continuous and strictly increasing. By Lemma 3.3(c), $\mathcal{P}_\kappa^{RS}$ strictly decreases on $[0, q^*]$ and strictly increases on $[q^*, 1]$; it is continuous at 0 by dominated convergence in Lemma 3.3(a).

(ii) For $\kappa < \kappa_{RS}$ put $\beta^*(\kappa) := \beta(q^*(\kappa))$. By (i) it is continuous and strictly decreasing, and $\kappa_\alpha(\beta^*(\kappa)) = \kappa$ by Lemma 3.7(a). Conversely, for $\beta \in (0, \beta_0]$, (3.3) gives $\kappa_{RS} - \kappa_\alpha(\beta) \geq \beta^2/(4\phi_0) > 0$, and $1/(1 + \beta^2)$ is RS-stationary at $\kappa_\alpha(\beta)$ by Lemma 3.7(a); so $\beta^*(\kappa_\alpha(\beta)) = \beta$ by (i). In particular $\kappa_\alpha(\beta_0) \leq \kappa_{RS} - c_2$, hence $\beta^*(\kappa) \in (0, \beta_0]$ for $\kappa \in [\kappa_{RS} - c_2, \kappa_{RS}]$. By Lemma 3.7(b), ω_α is continuous and strictly decreasing on $(0, \beta_0]$, $\omega_\alpha(0+) = \alpha\Phi(\kappa_{RS}) - 1 > 0$ by Lemma 3.1(b), and

$$\omega_\alpha(\beta_0) \leq \alpha\Phi(\kappa_{RS}) - 1 - \frac{1}{2}c_0\phi_0\beta_0 \leq \alpha\phi_0|\kappa_{RS}| - \frac{1}{2}c_0\phi_0\beta_0 < 0$$

after decreasing e_1 (Lemma 3.1(a),(b)). Hence ω_α has exactly one zero $\beta_d \in (0, \beta_0)$, with $\omega_\alpha > 0$ on $(0, \beta_d)$ and $\omega_\alpha < 0$ on $(\beta_d, \beta_0]$. Put $\kappa_d := \kappa_\alpha(\beta_d)$, so that $\beta^*(\kappa_d) = \beta_d$. By (iii) below, $\beta_d \leq \beta_0/2$ after decreasing e_1 , so $\kappa_{RS} - \kappa_d \leq \frac{3\beta_d^2}{4\phi_0} < c_2$ by (3.3). At $q^*(\kappa)$ we have $\alpha H_{RS} - 1 = \omega_\alpha(\beta^*(\kappa))$ by Lemma 3.7(a), so $\Lambda_R(\kappa) = -\omega_\alpha(\beta^*)/(4(1 - q^*)^2)$. For $\kappa \in [\kappa_{RS} - c_2, \kappa_{RS})$, $\beta^*(\kappa) < \beta_d$ exactly when $\kappa > \kappa_d$. This gives the signs and the uniqueness of κ_d , and Lemma 3.3(e) applies for $\kappa \in (\kappa_d, \kappa_{RS})$.

(iii) Letting $\beta' \downarrow 0$ in (3.4) and using Lemma 3.1(b), $0 = \omega_\alpha(\beta_d) \leq \alpha\phi_0|\kappa_{RS}| - \frac{1}{2}c_0\phi_0\beta_d$, so $\beta_d \leq 2\alpha|\kappa_{RS}|/c_0 \leq Ce$ by Lemma 3.1(a). As $x^2 \log(1/x)$ is increasing on $(0, \frac{1}{2}]$, $\beta_d^2 \log(1/\beta_d) \leq Ce^2 \log(1/e)$, and the expansion (3.3) of ω_α at β_d gives $|\frac{e}{4} - 2c_0\phi_0\beta_d| \leq Ce^2 \log(1/e)$. Hence

$$\beta_d = \frac{e}{8c_0\phi_0} + O(e^2 \log(1/e)), \quad \beta_d^2 = \frac{e^2}{64c_0^2\phi_0^2} + O(e^3 \log(1/e)).$$

By (3.3) at $\beta = \beta_d$, $\kappa_{RS} - \kappa_d = \beta_d^2/(2\phi_0) + O(e^3) = e^2/(128c_0^2\phi_0^3) + O(e^3 \log(1/e))$, and $1/(128\phi_0^3) = (2\pi)^{3/2}/128$. Finally $1 - q_d = \beta_d^2/(1 + \beta_d^2)$, so $d_d = \beta_d^2 e^{-2}(1 + O(e^2))$, and $2\phi_0\theta_d = 1/(64c_0^2\phi_0^2)$.

(iv) Let $e_\Theta \leq e_1$ be so small that $e_\Theta^2 \max \Theta \leq c_2$; then $\kappa_e(\theta) \in [\kappa_{RS} - c_2, \kappa_{RS}]$, and $\beta^* := \beta^*(\kappa_e(\theta)) \in (0, \beta_0]$ by the proof of (ii). By Lemma 3.7(a) and (3.3),

$$e^2\theta = \kappa_{RS} - \kappa_\alpha(\beta^*) = \frac{\beta^{*2}}{2\phi_0} + O(e\beta^{*2} + \beta^{*3}), \quad \frac{\beta^{*2}}{4\phi_0} \leq e^2\theta \leq \frac{3\beta^{*2}}{4\phi_0}.$$

Hence $\frac{4}{3}\phi_0\theta e^2 \leq \beta^{*2} \leq 4\phi_0\theta e^2$, and then $\beta^{*2} = 2\phi_0\theta e^2 + O(e^3)$, with constants depending on Θ . Since $1 - q^* = \beta^{*2}/(1 + \beta^{*2})$, the first claim follows. For the second, $\beta^* = e\sqrt{2\phi_0\theta} + O(e^2)$, $\beta^{*2} \log(1/\beta^*) \leq Ce^2 \log(1/e)$, and $1 - \alpha H_{RS} = -\omega_\alpha(\beta^*)$ by Lemma 3.7(a); insert these into (3.3). Finally $2\phi_0\theta_d = 1/(64c_0^2\phi_0^2)$ gives $c_0 t(\theta_d) = c_0 \cdot 2\phi_0 \cdot \frac{1}{8c_0\phi_0} = \frac{1}{4}$. $\square$

At the stationary RS point, $\mathcal{H}_{\gamma_{q^*}}(s) - \mathcal{H}_{\gamma_{q^*}}(q^*) = \Lambda_R(s - q^*)^2 + o((s - q^*)^2)$ by Lemma 3.3(d). The instability for $\kappa > \kappa_d$ concerns mass split off near the atom. The longitudinal direction, in which the whole atom moves, is stable on both sides of κ_d , since q^* minimizes $\mathcal{P}_\kappa^{\text{RS}}$. Proposition 3.8 does not assert that γ_{q^*} is a global minimizer for $\kappa \leq \kappa_d$; for $\kappa \in [\kappa_d - ce^2, \kappa_d]$ this is proved in Chapter 8 (Theorem 8.13).

3.4 The zero-temperature functional and a dual certificate

The critical margin is defined through the divergence of $\mathcal{P}_*$, which is not directly computable. Theorem 2.21 characterizes it through the zero-temperature functional J^k of (2.41), in which the terminal datum $\log \Phi((x - \kappa)/\sqrt{\delta})$ is replaced by its Gaussian-tail approximation $-(\kappa - x)_+^2/(2\delta)$ and the top gap is scaled out. We first record the consequences of Chapter 2 used below. We then prove a duality bound: a single function $g \geq k - x$ satisfying two Gaussian moment constraints forces $J^k \geq 0$ for all coefficients at once, and hence $\kappa_c \geq k$.

For $T \in (0, 1]$ let $\mathcal{S}_T$ be the set of nondecreasing, right-continuous step functions $\xi : [0, T) \rightarrow [0, \infty)$ with finitely many values. For $\xi \in \mathcal{S}_T$ and $k \in \mathbb{R}$, $Y = Y_{T, \xi}^k$ denotes the solution of (2.40), given by the Cole–Hopf recursion run backwards from the terminal datum $f_k(x) = -\frac{1}{2}(k - x)_+^2$ at time T with coefficient ξ . As in (2.41), $\Lambda_\xi(q) = 1 + \int_q^T \xi$, $L_\xi = \Lambda_\xi(0)$, $J_T^k(\xi) = \alpha Y(0, 0) + \frac{1}{2} \int_0^T \Lambda_\xi(q)^{-1} dq$ and $J^k = J_1^k$. The coefficient ξ may exceed one. For $\xi = 0$, $Y(0, 0) = -\frac{1}{2} \mathbb{E}(k - \sqrt{T}G)_+^2$, so

$$J^k(0) = \frac{1}{2}(1 - \alpha m(k)), \quad (3.5)$$

which vanishes exactly at $k = \kappa_{\text{RS}}$.

Lemma 3.9. *Let $\alpha > 0$ and $k \in \mathbb{R}$.*

- (a) *Two criteria. If $J^k(\xi) < 0$ for some $\xi \in \mathcal{S}_1$, then $\kappa_c \leq k$. If $J^k(\xi) \geq 0$ for every $\xi \in \mathcal{S}_1$, then $\kappa_c \geq k$.*
- (b) *The optimal process. Let $\xi \in \mathcal{S}_1$ and $Y = Y_{1, \xi}^k$, let B be a Brownian motion with natural filtration $(\mathcal{F}_q)$, and let X^* be the strong solution of $dX_q^* = \xi(q)Y_x(q, X_q^*)dq + dB_q$, $X_0^* = 0$. Then $M_q := Y_x(q, X_q^*)$ is a square-integrable $(\mathcal{F}_q)$ -martingale on $[0, 1]$ with $M_1 = (k - X_1^*)_+$, and*

$$Y(0, 0) = -\frac{1}{2} \mathbb{E} \left[M_1^2 + \int_0^1 \xi(q) M_q^2 dq \right].$$

Proof. (a) By Theorem 2.21, $\kappa_c = \kappa_c^0$, the infimum of the margins k' for which $J^{k'}(\xi) < 0$ for some nondecreasing step function ξ on $[0, 1]$. The first claim follows. For the second, $k' \mapsto J^{k'}(\xi)$ is nonincreasing by Lemma 2.19(b), so $J^{k'}(\xi) \geq J^k(\xi) \geq 0$ for all $k' \leq k$ and $\xi \in \mathcal{S}_1$; hence $\kappa_c^0 \geq k$.

(b) The terminal datum f_k belongs to the class of terminal data of Lemma 2.8, with $f'_k(x) = (k - x)_+$, and Y is the solution considered there. By Lemma 2.8, X^* exists and M is a square-integrable martingale with $M_1 = f'_k(X_1^*) = (k - X_1^*)_+$; it is adapted to $(\mathcal{F}_q)$ because X^* is a strong solution. Equality holds in the control inequality (2.20) for the feedback control $\theta_q^* = \zeta(q)M_q$, whose cost is $\theta_q^{*2}/\zeta(q) = \zeta(q)M_q^2$. Hence $Y(0, 0) = \mathbb{E}f_k(X_1^*) - \frac{1}{2}\mathbb{E} \int_0^1 \zeta M_q^2 dq$, and $f_k(X_1^*) = -\frac{1}{2}M_1^2$. $\square$

We now prove the duality bound. A trial control in the variational representation of $Y(0, 0)$ (Lemma 2.8) gives a lower bound on $J^k(\zeta)$ for a fixed ζ . To obtain a nonnegative lower bound uniformly in ζ , we use a dual object: a function g lying above $k - x$ that is small in L^2 of the Gaussian measure and has small slope in the same space. Neither condition involves ζ .

Theorem 3.10. *Let $\alpha > 0$ and $k \in \mathbb{R}$, and let $g : \mathbb{R} \rightarrow \mathbb{R}$ be Lipschitz with $g(x) \geq k - x$ for all $x \in \mathbb{R}$. Assume*

$$(C1) \quad \alpha \mathbb{E}g(G)^2 \leq 1, \quad (C2) \quad \alpha \mathbb{E}g'(G)^2 \leq 1. \quad (3.6)$$

Then $J^k(\zeta) \geq 0$ for every $\zeta \in \mathcal{S}_1$.

Proof. Fix $\zeta \in \mathcal{S}_1$, write $\Lambda = \Lambda_\zeta$ and $L = L_\zeta$, and let B , $(\mathcal{F}_t)$, Y , X^* , and M be as in Lemma 3.9(b). Put $\nu := M_1 = (k - X_1^*)_+ \geq 0$, so that $M_s = \mathbb{E}[\nu \mid \mathcal{F}_s]$. We divide the proof into four steps.

STEP 0. Martingale representations. Let (P_t) denote the heat semigroup. Since g is Lipschitz, $\partial_x P_t g = P_t g'$ for $t > 0$ and $|P_t g'| \leq \text{Lip}(g)$. The martingale $\mathbb{E}[g(B_1) \mid \mathcal{F}_r] = (P_{1-r}g)(B_r)$ has terminal value $g(B_1)$, and for $r' < 1$ Itô's formula, with $\partial_r(P_{1-r}g) = -\frac{1}{2}\partial_{xx}P_{1-r}g$, gives

$$(P_{1-r'}g)(B_{r'}) = \mathbb{E}g(B_1) + \int_0^{r'} (P_{1-r}g')(B_r) dB_r.$$

Letting $r' \uparrow 1$ in L^2 , we obtain $g(B_1) = \mathbb{E}g(B_1) + \int_0^1 \varsigma_r dB_r$ with $\varsigma_r := (P_{1-r}g')(B_r)$, which is the Clark–Ocone formula for $g(B_1)$. By Jensen's inequality and (C2),

$$\mathbb{E}\varsigma_r^2 \leq \mathbb{E}(P_{1-r}g'^2)(B_r) = \mathbb{E}g'(G)^2 \leq \frac{1}{\alpha} \quad (0 \leq r < 1). \quad (3.7)$$

Similarly write $\nu = \mathbb{E}\nu + \int_0^1 n_r dB_r$ (by the martingale representation theorem; see [79]), so that $\mathbb{E}M_s^2 = (\mathbb{E}\nu)^2 + \int_0^s \mathbb{E}n_r^2 dr$.

STEP 1. J in terms of ν . Put

$$\mathcal{Q} := \mathbb{E}\nu^2 + \int_0^1 \zeta(s) \mathbb{E}M_s^2 ds = (\mathbb{E}\nu)^2 L + \int_0^1 \Lambda(r) \mathbb{E}n_r^2 dr,$$

where the second expression follows from the formula for $\mathbb{E}M_s^2$, Fubini's theorem, and $\Lambda(r) = 1 + \int_r^1 \zeta$. By Lemma 3.9(b), $J^k(\zeta) = \frac{1}{2} \int_0^1 \Lambda^{-1} - \frac{\alpha}{2} \mathcal{Q}$. Moreover $\mathcal{Q} = \mathbb{E}[\nu(k - B_1)]$. Indeed $k - B_1 = (k - X_1^*) + \int_0^1 \zeta M_s ds$ and $\nu(k - X_1^*) = \nu^2$, so

$$\mathbb{E}[\nu(k - B_1)] = \mathbb{E}\nu^2 + \int_0^1 \zeta(s) \mathbb{E}[\nu M_s] ds = \mathbb{E}\nu^2 + \int_0^1 \zeta(s) \mathbb{E}M_s^2 ds = \mathcal{Q}.$$

STEP 2. Comparison with $g(B_1)$. Since $\nu \geq 0$ and $g(B_1) \geq k - B_1$, the Itô isometry and the Cauchy–Schwarz inequality, applied to $(\sqrt{L} \mathbb{E}\nu)(\mathbb{E}g(B_1)/\sqrt{L})$ and to $(\sqrt{\Lambda(r)} n_r)(\zeta_r/\sqrt{\Lambda(r)})$, give

$$\begin{aligned} \mathcal{Q} &\leq \mathbb{E}[\nu g(B_1)] = \mathbb{E}\nu \mathbb{E}g(B_1) + \int_0^1 \mathbb{E}[n_r \zeta_r] dr \leq \mathcal{Q}^{1/2} \mathcal{R}^{1/2}, \\ \mathcal{R} &:= \frac{(\mathbb{E}g(B_1))^2}{L} + \int_0^1 \frac{\mathbb{E}\zeta_r^2}{\Lambda(r)} dr. \end{aligned}$$

Hence $\mathcal{Q} \leq \mathcal{R}$ and $J^k(\zeta) \geq \frac{1}{2} [\int_0^1 \Lambda^{-1} - \alpha \mathcal{R}]$.

STEP 3. The sign. The function Λ^{-1} is nondecreasing with $\Lambda(0)^{-1} = 1/L$, and $\int_0^1 \mathbb{E}\zeta_r^2 dr = \text{Var } g(B_1)$ by the Itô isometry. Hence

$$\begin{aligned} \int_0^1 \frac{dr}{\Lambda(r)} - \alpha \mathcal{R} &= \frac{1}{L} [1 - \alpha \text{Var } g(B_1) - \alpha (\mathbb{E}g(B_1))^2] + \int_0^1 \left(\frac{1}{\Lambda(r)} - \frac{1}{L} \right) (1 - \alpha \mathbb{E}\zeta_r^2) dr \\ &\geq \frac{1}{L} (1 - \alpha \mathbb{E}g(G)^2) \geq 0, \end{aligned}$$

by (3.7) for the integral and (C1) for the last inequality. This proves the theorem. $\square$

3.5 Ordering of the thresholds

We now construct an explicit certificate. A natural candidate is $g(x) = \beta R((x - k)/\beta)$, the terminal value of λb for a replica-symmetric order parameter. It satisfies $g(x) \geq (k - x)_+$, $\mathbb{E}g(G)^2 = F_{\text{RS}}(k, \beta)$ and $\mathbb{E}g'(G)^2 = H_{\text{RS}}(k, \beta)$, so (C1) and (C2) read $\alpha F_{\text{RS}} \leq 1$ and $\alpha H_{\text{RS}} \leq 1$. Along the stationary curve $\alpha F_{\text{RS}} = 1$ they hold only for $\beta \geq \beta_d$ (on the range of Lemma 3.7), and the certified margins $K_\alpha(\beta) \leq K_\alpha(\beta_d) = \kappa_d \sqrt{1 + \beta_d^2}$ lie below κ_d . The loss comes from the left tail: $R(z)^2 - z^2 \rightarrow 2$ as $z \rightarrow -\infty$, which costs $2\beta^2 \Phi(k)$ in (C1) (Lemma 3.4(b)). A certificate that equals $k - x$ exactly outside a window of width $O(e)$ around k avoids this cost. Its price in (C1) is only cubic in the width, while it still lowers the slope enough to satisfy (C2).

Theorem 3.11. *There are numerical constants $e_2 > 0$ and $C_{\text{ord}} < \infty$ such that for $0 < e < e_2$ and $\alpha = 2 + e$ the following hold.*

- (i) There is $k_* = k_*(\alpha)$ with $\kappa_{\text{RS}} - C_{\text{ord}}e^3 \leq k_* < \kappa_{\text{RS}}$ such that $J^{k_*}(\xi) \geq 0$ for every $\xi \in \mathcal{S}_1$.
Consequently

$$\kappa_c(\alpha) \geq k_*(\alpha) \geq \kappa_{\text{RS}}(\alpha) - C_{\text{ord}}e^3. \quad (3.8)$$

- (ii) $\kappa_c(\alpha) - \kappa_d(\alpha) \geq \frac{1}{2}\theta_d e^2$.
(iii) Subcritical window. For every compact $\Theta \subset (0, \infty)$ there is $e_\Theta > 0$ such that, for $0 < e < e_\Theta$ and $\theta \in \Theta$,

$$k_*(2+e) - \kappa_e(\theta) \geq \frac{1}{2}e^2 \min \Theta, \quad \text{in particular} \quad \kappa_c(2+e) - \kappa_e(\theta) \geq \frac{1}{2}e^2 \min \Theta.$$

- (iv) Distance from the crossing. There is a numerical constant $C_{\text{dc}} < \infty$ such that

$$\begin{aligned} |\kappa_c(\alpha) - \kappa_d(\alpha) - \theta_d e^2| &\leq C_{\text{dc}} e^3 \log(1/e), \quad \text{that is,} \\ \kappa_c - \kappa_d &= \theta_d e^2 + O(e^3 \log(1/e)). \end{aligned}$$

Proof. (i) Let $e \leq e_g$ (Lemma 3.1), and put

$$\bar{\beta} := 2\alpha|\kappa_{\text{RS}}|, \quad \Delta := \frac{3\phi_0\bar{\beta}^3}{2m_1(-1/2)}, \quad k_* := \kappa_{\text{RS}} - \Delta.$$

By Lemma 3.1(a), $\bar{\beta} \leq Ce$ and $\Delta \leq C_{\text{ord}}e^3$, and for small e we have $k_* \geq -\frac{1}{2}$ and $|k_*| + \bar{\beta} \leq \frac{1}{2}$. Define $\mathfrak{g} : \mathbb{R} \rightarrow \mathbb{R}$ by $\mathfrak{g}(z) = -z$ for $z \leq -1$, $\mathfrak{g}(z) = 1$ for $-1 \leq z \leq 0$, and $\mathfrak{g}(z) = (1 - z/2)_+$ for $z \geq 0$, and put

$$g(x) := \bar{\beta} \mathfrak{g}\left(\frac{x - k_*}{\bar{\beta}}\right).$$

Since $\mathfrak{g}(z) \geq -z$ for all z , $g(x) \geq k_* - x$. The function g is Lipschitz with $g' = -1$ on $(-\infty, k_* - \bar{\beta})$, $g' = 0$ on $(k_* - \bar{\beta}, k_*)$, $g' = -\frac{1}{2}$ on $(k_*, k_* + 2\bar{\beta})$ and $g' = 0$ on $(k_* + 2\bar{\beta}, \infty)$.

Condition (C2). Since $\phi \geq \phi(|k_*| + \bar{\beta})$ on $[k_* - \bar{\beta}, k_*]$ and $\phi \leq \phi_0$,

$$\begin{aligned} \mathbb{E}g'(G)^2 &= \Phi(k_* - \bar{\beta}) + \frac{1}{4}(\Phi(k_* + 2\bar{\beta}) - \Phi(k_*)) \\ &\leq \Phi(k_*) - \bar{\beta}\phi(|k_*| + \bar{\beta}) + \frac{1}{2}\bar{\beta}\phi_0 \leq \Phi(\kappa_{\text{RS}}) - \frac{3}{8}\bar{\beta}\phi_0, \end{aligned}$$

because $\phi(|k_*| + \bar{\beta}) \geq \phi(\frac{1}{2}) = \phi_0 e^{-1/8} \geq \frac{7}{8}\phi_0$ and $k_* < \kappa_{\text{RS}}$. By Lemma 3.1(b), $\alpha\Phi(\kappa_{\text{RS}}) \leq 1 + \alpha\phi_0|\kappa_{\text{RS}}|$; with $\bar{\beta} = 2\alpha|\kappa_{\text{RS}}|$ and $\alpha \geq 2$,

$$\alpha\mathbb{E}g'(G)^2 \leq 1 + \alpha\phi_0|\kappa_{\text{RS}}|(1 - \frac{3}{4}\alpha) \leq 1 - \frac{1}{2}\alpha\phi_0|\kappa_{\text{RS}}| < 1.$$

Condition (C1). With $z = (x - k_*)/\bar{\beta}$, we have $0 \leq g(x)^2 - (k_* - x)_+^2 = \bar{\beta}^2(\mathfrak{g}(z)^2 - z_-^2) \leq \bar{\beta}^2\mathbf{1}_{[-1,2]}(z)$: on $[-1, 0]$ the difference is $1 - z^2$, and on $[0, 2]$ it is $(1 - z/2)^2$. Hence

$$\mathbb{E}g(G)^2 \leq m(k_*) + \bar{\beta}^2\mathbb{P}(G \in [k_* - \bar{\beta}, k_* + 2\bar{\beta}]) \leq m(k_*) + 3\phi_0\bar{\beta}^3.$$

By convexity of m and $m' = 2m_1 \geq 2m_1(-\frac{1}{2})$ on $[-\frac{1}{2}, 0]$,

$$\alpha m(k_*) \leq \alpha m(\kappa_{\text{RS}}) - 2\alpha m_1(-\frac{1}{2})\Delta = 1 - 3\alpha\phi_0\bar{\beta}^3.$$

Therefore $\alpha \mathbb{E}g(G)^2 \leq 1$. By Theorem 3.10, $J^{k_*}(\xi) \geq 0$ for every $\xi \in \mathcal{S}_1$, and Lemma 3.9(a) gives $\kappa_c \geq k_*$.

(ii) Take $e_2 \leq e_1$. By (i) and Proposition 3.8(iii), $\kappa_c - \kappa_d \geq \kappa_{\text{RS}} - \kappa_d - C_{\text{ord}}e^3 = \theta_d e^2 - O(e^3 \log(1/e)) \geq \frac{1}{2}\theta_d e^2$ for small e .

(iii) $k_* - \kappa_e(\theta) = e^2\theta - \Delta \geq e^2 \min \Theta - C_{\text{ord}}e^3 \geq \frac{1}{2}e^2 \min \Theta$ for small e .

(iv) By Proposition 2.17, $\kappa_c \leq \kappa_{\text{RS}}$; with (i), $0 \leq \kappa_{\text{RS}} - \kappa_c \leq C_{\text{ord}}e^3$. Write $\kappa_c - \kappa_d = (\kappa_{\text{RS}} - \kappa_d) - (\kappa_{\text{RS}} - \kappa_c)$ and use Proposition 3.8(iii): $|\kappa_c - \kappa_d - \theta_d e^2| \leq |\kappa_{\text{RS}} - \kappa_d - \theta_d e^2| + C_{\text{ord}}e^3 \leq Ce^3 \log(1/e)$. $\square$

Remark 3.12. By Theorem 3.11(iii), every compact interval $\Theta \subset (0, \infty)$ lies in the *subcritical range*: for all sufficiently small e , the margins $\kappa_e(\theta)$, $\theta \in \Theta$, lie below κ_c by at least $\frac{1}{2}e^2 \min \Theta$. This rests on the lower bound (3.8), which is proved with the explicit dual certificate; every later use of the subcriticality of Θ depends on it. From now on $\Theta \subset (0, \infty)$ denotes a fixed compact interval, and we write $\theta_- := \min \Theta > 0$ and $\theta_+ := \max \Theta$. Up to §7.8, Θ is arbitrary, and all constants depend on Θ only through $\theta_{\pm}$. From §7.8 on, we assume in addition that Θ contains θ_d in its interior; Chapter 8 takes $\Theta = [\frac{1}{2}\theta_d, 2\theta_d]$.

We now prove that the replica-symmetric formula overestimates κ_c at order e^3 . The negative witness is a coefficient carried by a top layer of width $\varepsilon \asymp e^2$ and total mass one. At $k = \kappa_{\text{RS}}$ the value $J^{\text{RS}}(0)$ vanishes by (3.5); the layer gains an amount linear in ε with coefficient proportional to $\alpha\Phi(\kappa_{\text{RS}}) - 1 \approx e/4$, at a cost of order $\varepsilon^{3/2}$, and the choice $\varepsilon \asymp e^2$ balances the two. The value of J^k at a one-block coefficient involves exponential moments of the form $\mathbb{E} \exp(-t(a - G)_+^2)$, which also enter the upper bound on the capacity in [90, Definition 2 and Theorem 4.2]. The function $J_\ell(k)$ of Part I, Chapter 7, is $J^k(\xi_\varepsilon)$ at $\varepsilon = \ell$, for the one-block coefficient $\xi_\varepsilon = \varepsilon^{-1}\mathbf{1}_{[1-\varepsilon, 1]}$ of Step 1 below.

Theorem 3.13. *There are numerical constants $e_3 > 0$ and $c_{\text{cub}} > 0$ such that for $0 < e < e_3$ and $\alpha = 2 + e$,*

$$\kappa_{\text{RS}}(\alpha) - \kappa_c(\alpha) \geq c_{\text{cub}}e^3.$$

Together with Theorem 3.11(i), $c_{\text{cub}}e^3 \leq \kappa_{\text{RS}} - \kappa_c \leq C_{\text{ord}}e^3$.

Proof. We divide the proof into four steps.

STEP 1. *An exact formula.* For $0 < \varepsilon \leq \frac{1}{2}$ let $\xi_\varepsilon := \varepsilon^{-1}\mathbf{1}_{[1-\varepsilon, 1]} \in \mathcal{S}_1$ and $p := 1 - \varepsilon$. Put

$$f(z) := \log \mathbb{E} \exp\left(-\frac{1}{2}(z + G)_-^2\right) = \log \left[\Phi(z) + 2^{-1/2}e^{-z^2/4}\Phi(-z/\sqrt{2}) \right],$$

where the second expression follows by splitting the expectation according to the sign of $z + G$ and completing the square, $(z + g)^2 + g^2 = 2(g + z/2)^2 + z^2/2$. On the top

block the coefficient is $1/\varepsilon$, and $(k - x - \sqrt{\varepsilon}G)_+^2/\varepsilon = (z + G)_-^2$ with $z = (x - k)/\sqrt{\varepsilon}$; so the Cole–Hopf step gives $Y(p, x) = \varepsilon f((x - k)/\sqrt{\varepsilon})$. Below p the coefficient vanishes, so $Y(0, 0) = \varepsilon \mathbb{E}f(Z_\varepsilon)$ with $Z_\varepsilon := (\sqrt{p}G - k)/\sqrt{\varepsilon}$. Since $\Lambda_{\xi_\varepsilon} = 2$ on $[0, p)$ and $\Lambda_{\xi_\varepsilon}(q) = 1 + (1 - q)/\varepsilon$ on $[p, 1)$, the entropy term equals $\frac{1}{2}(\frac{p}{2} + \varepsilon \log 2)$. Hence

$$J^k(\xi_\varepsilon) = \alpha \varepsilon \mathbb{E}f(Z_\varepsilon) + \frac{p}{4} + \frac{\varepsilon \log 2}{2}.$$

STEP 2. *The remainder.* Put $\mathfrak{r}(z) := f(z) + \frac{1}{4}z_-^2 + \frac{\log 2}{2}\mathbf{1}_{\{z < 0\}}$; only an upper bound on $\mathfrak{r}$ is needed. For $z \geq 0$, $e^f \leq 1$, so $\mathfrak{r}(z) = f(z) \leq 0$. For $z < 0$ let $\mathfrak{a}(z) := \mathbb{E}e^{-(z+G)^2/2} = 2^{-1/2}e^{-z^2/4}$. Then $\mathfrak{r} = \log(e^f/\mathfrak{a})$, and $e^f - \mathfrak{a} = \mathbb{E}[1 - e^{-(z+G)^2/2}; z + G > 0] \leq \Phi(z)$, so $\mathfrak{r}(z) \leq \Phi(z)/\mathfrak{a}(z)$. The right side is at most $\sqrt{2}e^{1/4}$ on $[-1, 0)$, and at most $\sqrt{2}\phi_0 e^{-z^2/4}/|z|$ for $z \leq -1$ by Lemma 2.1(b). Thus $\mathfrak{r} \leq \bar{\mathfrak{r}} := (\Phi/\mathfrak{a})\mathbf{1}_{(-\infty, 0)} \in L^1(\mathbb{R})$. Using $\mathbb{E}(Z_\varepsilon)_-^2 = m(k; p)/\varepsilon$ with $m(k; p) := \mathbb{E}(k - \sqrt{p}G)_+^2$, so that $m(k; 1) = m(k)$, and $\mathbb{P}(Z_\varepsilon < 0) = \Phi(k/\sqrt{p})$,

$$J^k(\xi_\varepsilon) = \frac{p - \alpha m(k; p)}{4} + \frac{\varepsilon \log 2}{2}(1 - \alpha \Phi(k/\sqrt{p})) + \alpha \varepsilon \mathbb{E}\mathfrak{r}(Z_\varepsilon).$$

STEP 3. *Expansion.* Let $|k| \leq 1$ and $\alpha \leq 3$. The density of Z_ε is $z \mapsto \sqrt{\varepsilon}\phi_p(k + \sqrt{\varepsilon}z) \leq \sqrt{2\varepsilon}\phi_0$, where ϕ_p is the $N(0, p)$ density, so $\alpha \varepsilon \mathbb{E}\mathfrak{r}(Z_\varepsilon) \leq 3\sqrt{2}\phi_0 \|\bar{\mathfrak{r}}\|_{L^1} \varepsilon^{3/2}$. By Gaussian integration by parts, $\partial_t m(k; t) = \Phi(k/\sqrt{t})$. For $t \in [\frac{1}{2}, 1]$, $|\partial_t \Phi(k/\sqrt{t})| = |k|\phi(k/\sqrt{t})/(2t^{3/2}) \leq \sqrt{2}\phi_0|k| \leq |k|$, so $|\Phi(k/\sqrt{t}) - \Phi(k)| \leq |k|(1 - t)$. Since $m(k) - m(k; p) = \int_p^1 \Phi(k/\sqrt{t}) dt$, this gives $|m(k; p) - m(k) + \varepsilon \Phi(k)| \leq \varepsilon^2$ and $|\Phi(k/\sqrt{p}) - \Phi(k)| \leq \varepsilon$. Hence, with $p = 1 - \varepsilon$,

$$\begin{aligned} \frac{p - \alpha m(k; p)}{4} &\leq \frac{1 - \alpha m(k)}{4} + \frac{\varepsilon}{4}(\alpha \Phi(k) - 1) + \frac{\alpha \varepsilon^2}{4}, \\ \frac{\varepsilon \log 2}{2}(1 - \alpha \Phi(k/\sqrt{p})) &\leq -\frac{\varepsilon \log 2}{2}(\alpha \Phi(k) - 1) + \frac{\alpha \log 2}{2}\varepsilon^2. \end{aligned}$$

Adding these bounds and the bound on $\alpha \varepsilon \mathbb{E}\mathfrak{r}(Z_\varepsilon)$, and using $\varepsilon^2 \leq \varepsilon^{3/2}$ and $\alpha \leq 3$, we obtain, with $d := \log 2 - \frac{1}{2}$,

$$J^k(\xi_\varepsilon) \leq \frac{1 - \alpha m(k)}{4} - \frac{\varepsilon}{2}d(\alpha \Phi(k) - 1) + C_J \varepsilon^{3/2}, \quad (3.9)$$

where $C_J := \frac{3}{4} + \frac{3}{2}\log 2 + 3\sqrt{2}\phi_0 \|\bar{\mathfrak{r}}\|_{L^1}$ is a numerical constant. Here $d > 0$ because $\log 2 = \int_1^2 t^{-1} dt > \frac{1}{2}$.

STEP 4. *Choice of ε .* At $k = \kappa_{RS}$, $1 - \alpha m(\kappa_{RS}) = 0$ and, by Lemma 3.1(b), $\alpha \Phi(\kappa_{RS}) - 1 \geq c_\Phi e$ with $c_\Phi := m_1(-\frac{1}{2})/(4\phi_0)$. Let $t_* := (dc_\Phi/(4C_J))^2$ and $\varepsilon := t_* e^2 \leq \frac{1}{2}$. Then (3.9) gives

$$J^{\kappa_{RS}}(\xi_\varepsilon) \leq -\frac{dc_\Phi}{2}t_* e^3 + C_J t_*^{3/2} e^3 = -\frac{dc_\Phi}{4}t_* e^3 =: -c_4 e^3.$$

Let $c_{\text{cub}} := c_4/3$. By Lemma 2.19(b), and since $m_1(t) \leq m_1(0) = \phi_0$ for $t \leq 0$ and $\alpha\phi_0 \leq 3\phi_0 \leq \frac{3}{2}$,

$$J^{\kappa_{\text{RS}} - c_{\text{cub}}e^3}(\xi_\varepsilon) \leq J^{\kappa_{\text{RS}}}(\xi_\varepsilon) + \alpha\phi_0 c_{\text{cub}}e^3 \leq -c_4e^3 + \frac{1}{2}c_4e^3 < 0.$$

By Lemma 3.9(a), $\kappa_c \leq \kappa_{\text{RS}} - c_{\text{cub}}e^3$. □

The two cubic bounds are statements about the zero-temperature functional J only: Theorem 3.11(i) certifies $J^{k_*} \geq 0$ and Theorem 3.13 exhibits $J^{\kappa_{\text{RS}} - c_{\text{cub}}e^3}(\xi_\varepsilon) < 0$; they are transferred to κ_c through the characterization $\kappa_c = \kappa_c^0$ of Theorem 2.21.

Corollary 3.14. *Let $0 < e < \min\{e_1, e_2, e_3\}$ and $\alpha = 2 + e$. Then $\kappa_d < \kappa_c < \kappa_{\text{RS}}$, and for every $\kappa \in (\kappa_d, \kappa_c)$ the functional $\mathcal{P}_\kappa$ has global minimizers, and no global minimizer is replica symmetric: μ_γ is not a point mass. In particular every global minimizer has at least two support points.*

Proof. The inequalities $\kappa_d < \kappa_c$ and $\kappa_c < \kappa_{\text{RS}}$ are Theorem 3.11(ii) and Theorem 3.13. Minimizers exist by Theorems 2.16 and 2.18. Suppose that $\gamma = \gamma_q$ is a global minimizer at some $\kappa \in (\kappa_d, \kappa_c)$. By Proposition 2.11(c), $0 < q < 1$ and $D_{\gamma_q}(q) = 0$, so q is RS-stationary at κ . Since $\kappa \in (\kappa_d, \kappa_{\text{RS}})$, Proposition 3.8(i) gives $q = q^*(\kappa)$, and Proposition 3.8(ii) gives $\Lambda_R(\kappa) < 0$. By Lemma 3.3(e), γ_q is not a global minimizer, a contradiction. □

Corollary 3.14 says nothing about the structure of the minimizers beyond having two support points, and it does not use the uniqueness results of later chapters. Their interval structure for κ immediately above κ_d is established in Chapter 8.

Rescaled variables and a priori bounds for minimizers

For $\alpha = 2 + e$ and margins $\kappa = \kappa_e(\theta) = \kappa_{\text{RS}} - e^2\theta$, the relevant overlaps lie within $O(e^2)$ of one, and the relevant values of $x - \kappa$ are of order e . This chapter introduces variables adapted to these scales and proves the a priori bounds of Step 1 in §1.3 for all exact global minimizers. In §§4.1 and 4.2 we restate the identities of Chapter 2 in rescaled variables (Lemma 4.2) and write the value $\mathcal{P}_\kappa(\gamma)$ exactly in terms of the rescaled solution at the bottom of the support (Lemma 4.3). Proposition 4.5 splits this value into a term $K_e(q_0)$ that depends only on q_0 , three explicit terms that depend on γ only through the bottom q_0 , the rescaled mass ℓ_0 , and an entropy-type quantity $\tilde{\Gamma}$, and a remainder $\alpha \mathbb{E}W(Y_0)$ with W integrable; Lemma 4.6 expands the coefficients to first order in e , and Lemma 4.7 evaluates the replica-symmetric trial order parameters. In §4.3 we read off bounds on the mass and the depth from the first-order condition $D(q_0) = 0$ at the bottom of the support (Lemma 4.9). In §4.4 we prove that every global minimizer satisfies $c e^2 \leq d \leq D_{\text{top}}$, $\ell_0 \leq L$, and $s_0 \leq C_0 e^{-2/3}$, uniformly in $\theta \in \Theta$ (Theorems 4.10 and 4.11). Chapter 5 improves the depth bound to $s_0 = O(1)$, and Chapter 6 uses all of these bounds; Proposition 4.5 and Lemmas 4.6 and 4.7 are also the starting point of the flat limit in Chapter 7.

Rescalings of a similar kind appear in the physics analysis of the jamming limit [59, (67)–(68)], where λ is divided by the distance to the jamming line and $x - \kappa$ is compared with $\sqrt{1 - q}$; see also [30] for hard spheres.

Throughout this chapter, $\alpha = 2 + e$ with $e \in (0, 1]$, $\Theta \subset (0, \infty)$ is a compact interval as in Remark 3.12, $\theta_- = \min \Theta$ and $\theta_+ = \max \Theta$, and $\kappa = \kappa_e(\theta)$ with $\theta \in \Theta$ unless stated otherwise. Constants c, C depend only on Θ unless stated otherwise.

4.1 Rescaled variables

Definition 4.1. Let $e > 0$, $\kappa \in \mathbb{R}$, and let $\gamma \in \mathcal{U}$ have top $Q = 1 - \delta$ and bottom $q_0 = \min \text{supp } \mu_\gamma$; write $\lambda = \lambda_\gamma$ and $u = u_\gamma(\cdot, \cdot; \kappa)$. The *level* and the *rescaled space variable* are

$$s = \frac{1 - q}{e^2}, \quad q(s) := 1 - e^2 s, \quad y = \frac{x - \kappa}{e}.$$

The *rescaled profile* and its primitive are

$$\zeta(s) := \gamma(1 - e^2 s) \quad (s \in [0, e^{-2}]), \quad \ell(s) := \int_0^s \zeta(t) dt,$$

so that $\lambda(q(s)) = e^2 \ell(s)$. The *top gap*, the *depth*, and the *mass* are

$$d := \frac{\delta}{e^2}, \quad s_0 := \frac{1 - q_0}{e^2}, \quad \ell_0 := \ell(s_0) = \frac{\lambda(0)}{e^2}.$$

On $(0, e^{-2}] \times \mathbb{R}$ we put

$$\begin{aligned} U(s, y) &:= u(q(s), \kappa + ey), & v(s, y) &:= -\ell(s)U_{yy}(s, y), & \Psi(s, y) &:= \ell(s)U_y(s, y) + y, \\ F(s, y) &:= s(-U_{yy}(s, y)) - U_y(s, y)\Psi(s, y). \end{aligned}$$

The endpoint $s = 0$ corresponds to the singular hard wall $q = 1$ and is not in the domain of these smooth functions; the profile ζ and its primitive ℓ are nevertheless defined at $s = 0$. For $s \geq d$ also put

$$\Gamma(s) := \frac{1}{2} \log d + \frac{1}{2} \int_d^s \frac{dt}{\ell(t)}, \quad \tilde{\Gamma} := \Gamma(s_0) - \frac{1}{2} \log \ell_0.$$

Finally $Y_s := (X_{q(s)} - \kappa)/e$ for $s \in [d, e^{-2}]$, and $Y_0 := Y_{s_0}$. We call γ (or ζ) a *step profile* if it takes finitely many values.

Levels $s \in [d, e^{-2}]$ correspond to $q \in [0, Q]$. In rescaled variables $v = \lambda c$ becomes $v = -\ell U_{yy}$, and the harmonic function $\psi = \lambda b + x - \kappa$ of (2.21) becomes $\Psi = \psi/e$. The rescaled profile ζ is nonincreasing and left-continuous, $\zeta = 1$ on $[0, d]$, $\zeta < 1$ on $(d, e^{-2}]$ (because $Q \in \text{supp } \mu_\gamma$), $\zeta > 0$ on $[0, s_0)$ (because $q_0 \in \text{supp } \mu_\gamma$) and $\zeta = 0$ on $(s_0, e^{-2}]$; also $\zeta(d+) = \gamma(Q-)$. Hence ℓ is concave and nondecreasing, $\ell(s) = s$ on $[0, d]$, $\ell \leq s$, and $d \leq \ell_0 \leq s_0$. Since $(\log \ell)' = \zeta/\ell \leq 1/\ell$ on $[d, s_0]$,

$$\tilde{\Gamma} = \frac{1}{2} \log \frac{d}{\ell_0} + \frac{1}{2} \int_d^{s_0} \frac{dt}{\ell(t)} \geq 0. \quad (4.1)$$

Lemma 4.2. *Let $\gamma \in \mathcal{U}$, $\kappa \in \mathbb{R}$, and $e > 0$. On $[d, e^{-2}] \times \mathbb{R}$:*

- (a) $U \leq 0$, $U_y > 0$, $0 < v < 1$, $v_y < 0$, $\Psi > 0$, and $\Psi_y = 1 - v$; $\Psi(s, \cdot)$ is convex and increasing, and ℓU_y is decreasing in y , with $(\ell U_y)_y = -v$. At the top level,

$$U(d, y) = \log \Phi(y/\sqrt{d}), \quad U_y(d, y) = R(y/\sqrt{d})/\sqrt{d}, \quad \Psi(d, y) = \sqrt{d} \bar{R}(y/\sqrt{d}).$$

On every open interval of levels on which $\zeta \equiv m$, U is smooth and $U_s = \frac{1}{2}(U_{yy} + mU_y^2)$.

- (b) The sandwich.

$$v \leq U_y \Psi \leq \frac{s}{\ell(s)} v; \quad (4.2)$$

equivalently, $F \geq 0$ and $v \leq U_y \Psi$.

- (c) For $y < 0$: $U_y(s, y) \geq |y|/\ell(s)$ and $\Psi(s, y) \leq \min\{\sqrt{s}, s/|y|\}$. For $y > 0$: $y < \Psi(s, y) \leq y + \min\{\sqrt{s}, s/y\}$.
- (d) Let $q = q(s)$, $x = \kappa + ey$, and let $X^{q,x}$ be the diffusion (2.14) started from x at time q . Then

$$\Psi(s, y) = \mathbb{E} \left[\sqrt{d} \bar{R} \left(\frac{X_Q^{q,x} - \kappa}{e\sqrt{d}} \right) \right] \geq \max \left\{ \mathbb{E} (y + \sqrt{s-d} G)_+, \sqrt{d} \bar{R}(y/\sqrt{d}) \right\}.$$

Proof. Under the change of variables, $b = U_y/e$, $c = -U_{yy}/e^2$, $\lambda = e^2\ell$, $\lambda c = v$, $\psi = e\Psi$, $\partial_x = e^{-1}\partial_y$, and $(1-q)/\lambda = s/\ell$.

(a) The signs are Lemma 2.4(b),(c) and Lemma 2.9(b), and $(\ell U_y)_y = \ell U_{yy} = -v$. The top formulas follow from (2.9), $R' = -V$, and $\bar{R} = z + R$. Since $\partial_s = -e^2\partial_q$, the equation $u_q = -\frac{1}{2}(u_{xx} + \gamma u_x^2)$ becomes $U_s = \frac{1}{2}(U_{yy} + \zeta U_y^2)$ (Lemma 2.4(a)).

(b) The sandwich $v \leq b\psi \leq \frac{1-q}{\lambda}v$ of Lemma 2.10(a) becomes (4.2). Since $s(-U_{yy}) = \frac{s}{\ell}v$, the upper half is $F \geq 0$.

(c) Let $y < 0$. Since $\Psi > 0$, $\ell U_y = \Psi - y > |y|$. By (4.2) and $v < 1$, $\Psi \leq s/(\ell U_y) < s/|y|$. At $y = 0$, $\Psi = \ell U_y$, so $\Psi(s, 0)^2 = \ell U_y \Psi \leq sv < s$; since Ψ increases in y , $\Psi(s, y) \leq \sqrt{s}$ for $y \leq 0$. Let $y > 0$. Then $\Psi > y$ because $U_y > 0$, so $\ell U_y \leq s/\Psi < s/y$ by (4.2); and ℓU_y is decreasing in y , so $\ell U_y(s, y) \leq \ell U_y(s, 0) = \Psi(s, 0) \leq \sqrt{s}$.

(d) By Lemma 2.9(b), $\psi(q, x) = \mathbb{E}\psi(Q, X_Q^{q,x})$, and $\psi(Q, x') = \sqrt{d} \bar{R}((x' - \kappa)/\sqrt{d})$; divide by e . Put $\Xi := (X_Q^{q,x} - \kappa)/e$. Since the drift of X is nonnegative, (2.15) gives $X_Q^{q,x} \geq x + B_Q - B_q$, so $\Xi \geq y + (B_Q - B_q)/e$, a Gaussian variable with mean y and variance $s - d$. Since $\bar{R}(z) > z$, $\bar{R} > 0$, and $\bar{R}$ is increasing, $\sqrt{d} \bar{R}(\Xi/\sqrt{d}) \geq \Xi_+ \geq (y + (B_Q - B_q)/e)_+$, which gives the first lower bound. Since $\bar{R}$ is convex and increasing (Lemma 2.1(a)) and $\mathbb{E}\Xi \geq y$, Jensen's inequality gives the second. $\square$

4.2 The value formula and the exact decomposition

Below its bottom q_0 the order parameter vanishes, so the diffusion X is a Brownian motion there and u solves the heat equation. This makes the input of the rescaled problem an explicit Gaussian.

Lemma 4.3. Let $\alpha > 0$, let $\gamma \in \mathcal{U}$ with $q_0 > 0$, let $\kappa \in \mathbb{R}$ and $e > 0$. Then $Y_0 = (B_{q_0} - \kappa)/e$ has the Gaussian law $N(-\kappa/e, q_0/e^2)$, with density

$$p_0(y) = \frac{e}{\sqrt{q_0}} \phi\left(\frac{ey + \kappa}{\sqrt{q_0}}\right) \leq \frac{e\phi_0}{\sqrt{q_0}},$$

and

$$\mathcal{P}_\kappa(\gamma) = \alpha \mathbb{E}U(s_0, Y_0) + \Gamma(s_0) + \frac{q_0}{2e^2\ell_0} + \log e. \quad (4.3)$$

Proof. Since $\gamma = 0$ on $[0, q_0)$, the drift of X vanishes there, so $X_{q_0} = B_{q_0}$; and by Lemma 2.5(a) with $m = 0$, $u(0, 0) = \mathbb{E}u(q_0, B_{q_0}) = \mathbb{E}U(s_0, Y_0)$. The entropy is $\text{CS}(\gamma) =$

$\frac{1}{2} \int_0^Q \lambda^{-1} + \frac{1}{2} \log \delta$. On $[0, q_0]$ we have $\lambda = \lambda(0) = e^2 \ell_0$, so $\frac{1}{2} \int_0^{q_0} \lambda^{-1} = \frac{q_0}{2e^2 \ell_0}$. Substituting $q = 1 - e^2 t$ on $[q_0, Q]$ gives $\frac{1}{2} \int_{q_0}^Q \lambda^{-1} dq = \frac{1}{2} \int_d^{s_0} \ell^{-1} dt$. Finally $\frac{1}{2} \log \delta = \log e + \frac{1}{2} \log d$. $\square$

The proof uses only that $\gamma = 0$ on $[0, q_0)$. Hence Lemma 4.3, and with it Proposition 4.5 below, remains valid verbatim if q_0 is replaced by any level $q_1 \in (0, q_0]$, with s_0 replaced by $s_1 = (1 - q_1)/e^2$ and Y_0 by $(B_{q_1} - \kappa)/e$; the mass is unchanged, $\ell(s_1) = \ell_0$, because $\zeta = 0$ on $(s_0, s_1]$, and (4.1) continues to hold with s_1 in place of s_0 .

The next lemma records the tails of U . In the far left, the terminal datum $\log \Phi(y/\sqrt{d})$ behaves like $-y^2/(2d) - \log |y|$. Below the top the quadratic coefficient becomes $1/(2\ell(s))$, the logarithm keeps coefficient one, and the constant term records the entropy Γ . This is Lemma 2.4(b),(d),(e) in rescaled variables.

Lemma 4.4. *Let $\gamma \in \mathcal{U}$, let $\kappa \in \mathbb{R}$ and $e > 0$, and let $r \in (Q, 1)$. There are $C, Y_1 < \infty$, depending only on $e, |\kappa|$, and r , such that for $s \in [d, e^{-2}]$:*

- (a) $0 \geq U(s, y) \geq -C(1 + y^2)$ for all $y \in \mathbb{R}$, and $U(s, y) \geq -C e^{-y^2/(8s)}$ for $y \geq 0$;
- (b) with $\omega(s) := \log \ell(s) - \Gamma(s) - \log \sqrt{2\pi}$,

$$\left| U(s, y) + \frac{y^2}{2\ell(s)} + \log |y| - \omega(s) \right| \leq \frac{C}{y^2} \quad (y \leq -Y_1).$$

Proof. Put $q = 1 - e^2 s \in [0, Q]$ and $x = \kappa + ey$, and apply Lemma 2.4(b),(d),(e) with the cutoff r and $K_0 = |\kappa|$. Part (a) follows from $u \leq 0$, the quadratic lower bound of Lemma 2.4(b), Lemma 2.4(d), and $(x - \kappa)^2/(8(1 - q)) = y^2/(8s)$. For (b), we have $(x - \kappa)^2/(2\lambda(q)) = y^2/(2\ell(s))$ and $\log(\kappa - x) = \log e + \log |y|$. Since $\lambda = e^2 \ell$, $\delta = e^2 d$, and the substitution $t = 1 - e^2 t'$ gives $\int_q^Q \lambda(t)^{-1} dt = \int_d^s \ell(t')^{-1} dt'$, the constant (2.13) of Lemma 2.4(e) equals

$$\log(e^2 \ell(s)) - \frac{1}{2} \log(e^2 d) - \frac{1}{2} \int_d^s \frac{dt'}{\ell(t')} - \log \sqrt{2\pi} = \log e + \omega(s).$$

The terms $\log e$ cancel, and the error $C/(\kappa - x)^2$ equals Ce^{-2}/y^2 . $\square$

We now split $\mathcal{P}_\kappa(\gamma)$. The explicit left asymptotics of Lemma 4.4 can be integrated against the Gaussian law exactly. What remains is a function W that is integrable on $\mathbb{R}$. The estimate $|\alpha EW(Y_0)| \leq \alpha e \phi_0 q_0^{-1/2} \|W\|_{L^1}$ makes this contribution $O(e)$ on classes where q_0 is bounded away from zero and $\|W\|_{L^1}$ is uniformly bounded.

Proposition 4.5. *Let $\alpha > 0$, let $\gamma \in \mathcal{U}$ with $q_0 > 0$, let $\kappa \in \mathbb{R}$ and $e > 0$. Define*

$$k_0 := \frac{\kappa}{\sqrt{q_0}}, \quad \rho := \alpha \Phi(k_0), \quad c_e(q_0) := \frac{q_0(1 - \alpha m(k_0))}{2e^2},$$

$$\omega(s_0) := \log \ell_0 - \Gamma(s_0) - \log \sqrt{2\pi},$$

$$K_e(q_0) := \log e - \alpha \mathbb{E}[\log |Y_0|; Y_0 \leq -1] - \rho \log \sqrt{2\pi},$$

$$W(y) := U(s_0, y) + \frac{y^2}{2\ell_0} - \omega(s_0)\mathbf{1}_{\{y \leq 0\}} + \log |y| \mathbf{1}_{\{y \leq -1\}},$$

and the profile-independent constant $K_e := \log e - \alpha \mathbb{E}[\log |Y^*|; Y^* \leq -1] - \alpha \Phi(\kappa) \log \sqrt{2\pi}$, where $Y^* := (G - \kappa)/e$; this is $K_e(q_0)$ at $q_0 = 1$.

(a) $\mathbb{E}|W(Y_0)| < \infty$ and

$$\mathcal{P}_\kappa(\gamma) = K_e(q_0) + \frac{c_e(q_0)}{\ell_0} + \frac{1+\rho}{2} \log \ell_0 + (1-\rho)\tilde{\Gamma} + \alpha \mathbb{E}W(Y_0). \quad (4.4)$$

Here $K_e(q_0)$, $c_e(q_0)$, and ρ depend on γ only through q_0 , and $\tilde{\Gamma} \geq 0$.

(b) W is bounded on compact sets, $W(y) = O(y^{-2})$ as $y \rightarrow -\infty$, and $W(y) = U(s_0, y) \in [-C_\gamma e^{-y^2/(8s_0)}, 0]$ for $y > 0$. In particular $W \in L^1(\mathbb{R})$. The constants here depend on γ (through a cutoff $r \in (Q, 1)$) and on e ; no bound on $\|W\|_{L^1}$ uniform in e is claimed at this stage.

Proof. (a) Put $\mathcal{E}(y) := -\frac{y^2}{2\ell_0} + \omega(s_0)\mathbf{1}_{\{y \leq 0\}} - \log |y| \mathbf{1}_{\{y \leq -1\}}$, so that $U(s_0, \cdot) = \mathcal{E} + W$. Both $U(s_0, \cdot)$ (by Lemma 4.4(a)) and $\mathcal{E}$ have at most quadratic growth, so $W(Y_0)$ is integrable. Since $e(Y_0)_- = (\kappa - B_{q_0})_+$ and $\mathbb{P}(Y_0 \leq 0) = \Phi(k_0)$,

$$\alpha \mathbb{E}\mathcal{E}(Y_0) = -\frac{\alpha q_0 m(k_0)}{2e^2 \ell_0} + \rho \omega(s_0) - \alpha \mathbb{E}[\log |Y_0|; Y_0 \leq -1].$$

Add $\Gamma(s_0) + \frac{q_0}{2e^2 \ell_0} + \log e$ as in (4.3). The first term and $\frac{q_0}{2e^2 \ell_0}$ combine to $c_e(q_0)/\ell_0$. Using $\omega(s_0) = \log \ell_0 - \Gamma(s_0) - \log \sqrt{2\pi}$ and $\Gamma(s_0) = \tilde{\Gamma} + \frac{1}{2} \log \ell_0$,

$$\begin{aligned} \rho \omega(s_0) + \Gamma(s_0) &= \rho \log \ell_0 + (1-\rho) \left(\tilde{\Gamma} + \frac{1}{2} \log \ell_0 \right) - \rho \log \sqrt{2\pi} \\ &= \frac{1+\rho}{2} \log \ell_0 + (1-\rho)\tilde{\Gamma} - \rho \log \sqrt{2\pi}. \end{aligned}$$

Collecting the remaining terms gives $K_e(q_0)$. Nonnegativity of $\tilde{\Gamma}$ is (4.1).

(b) $U(s_0, \cdot)$ is continuous (Lemma 2.4(a)), and $\mathcal{E}$ is bounded on compact sets. The left tail is Lemma 4.4(b) at $s = s_0$, and the right tail is Lemma 4.4(a). $\square$

The coefficients in (4.4) depend on γ only through $q_0 = 1 - e^2 s_0$. The next lemma expands them. The main term of c_e is $2\phi_0\theta$, coming from $\kappa_{RS} - \kappa = e^2\theta$ and $\alpha m'(\kappa_{RS}) \approx 4\phi_0$; the correction $es_0/8$ comes from the shift of the effective margin $\kappa/\sqrt{q_0}$ with depth.

Lemma 4.6. *There are $e_L > 0$ and $C_L < \infty$, depending only on Θ , such that for $0 < e \leq e_L$, $\theta \in \Theta$, and $\kappa = \kappa_e(\theta)$:*

(a) $|\kappa_{RS} + e/(8\phi_0)| \leq C_L e^2$ and $-e/2 \leq \kappa < 0$.

(b) For every $q_0 \in [\frac{1}{2}, 1]$, with $s_0 = (1 - q_0)/e^2$,

$$\left| \alpha \Phi\left(\frac{\kappa}{\sqrt{q_0}}\right) - 1 - \frac{e}{4} \right| \leq C_L(e^2 + e^3 s_0),$$

$$\left| c_e(q_0) - 2\phi_0\theta - e\left[\left(\phi_0 - \frac{1}{8\phi_0}\right)\theta + \frac{s_0}{8}\right] \right| \leq C_L(e^2 + e^2 s_0 + e^3 s_0^2),$$

$$|K_e(q_0) - K_e| \leq C_L e^2 s_0 \log(1/e), \quad |K_e - 2\log e| \leq C_L.$$

(c) For every $q \in (0, 1)$, with $s = (1 - q)/e^2$,

$$0 \leq c_e(q) \leq \frac{5\phi_0}{2} \left(\theta + \frac{1}{2} \min\left\{ es, \frac{\sqrt{q}}{e} \right\} \right).$$

Proof. (a) The first bound is Lemma 3.1(a). Since $\kappa = \kappa_{RS} - e^2\theta < 0$ and $\frac{1}{8\phi_0} = \frac{\sqrt{2\pi}}{8} < \frac{1}{2}$ (as $2\pi < 16$), $\kappa \geq -\frac{e}{8\phi_0} - (C_L + \theta_+)e^2 \geq -\frac{e}{2}$ for small e .

(b) Put $\tau_0 := 1 - q_0 = e^2 s_0 \leq \frac{1}{2}$, $k_0 := \kappa/\sqrt{q_0}$ and $\Delta_0 := \kappa_{RS} - k_0 = e^2\theta + |\kappa|(q_0^{-1/2} - 1)$. For $\tau_0 \leq \frac{1}{2}$, Taylor's formula for $(1 - \tau_0)^{-1/2}$ gives $0 \leq q_0^{-1/2} - 1 - \frac{\tau_0}{2} \leq 3\tau_0^2$; with (a),

$$\Delta_0 = e^2\theta + \frac{e\tau_0}{16\phi_0} + O(e^2\tau_0 + e\tau_0^2), \quad 0 \leq \Delta_0 \leq C(e^2 + e\tau_0).$$

The factor ρ . We have $\alpha\Phi(k_0) - 1 - \frac{e}{4} = [\alpha\Phi(\kappa_{RS}) - 1 - \frac{e}{4}] - \alpha[\Phi(\kappa_{RS}) - \Phi(k_0)]$; the first bracket is $O(e^2)$ by Lemma 3.1(b), and the second lies in $[0, \alpha\phi_0\Delta_0]$. The coefficient c_e . Since $\alpha m(\kappa_{RS}) = 1$ and $0 < m'' \leq 2$, $1 - \alpha m(k_0) = \alpha m'(\kappa_{RS})\Delta_0 + O(\Delta_0^2)$. By Lemma 3.1(c),

$$1 - \alpha m(k_0) = 4\phi_0 e^2\theta + e^3 \left(2\phi_0 - \frac{1}{4\phi_0} \right) \theta + \frac{e\tau_0}{4} + O(e^4 + e^2\tau_0 + e\tau_0^2).$$

Multiplying by $q_0/(2e^2) = (1 - \tau_0)/(2e^2)$ and using $\tau_0 = e^2 s_0$ gives the expansion of $c_e(q_0)$. The constant $K_e(q_0)$. For $q \in [\frac{1}{2}, 1]$ write $\mathbb{E}[\log |Y_q|; Y_q \leq -1] = \mathbb{E}[\mathfrak{l}(\sqrt{q} G)]$, with $Y_q = (\sqrt{q} G - \kappa)/e$ and $\mathfrak{l}(x) := \log((\kappa - x)/e) \mathbf{1}_{\{x \leq \kappa - e\}}$. The function $\mathfrak{l}$ is continuous, with $\mathfrak{l}' = -1/(\kappa - x)$ on $(-\infty, \kappa - e)$ and $\mathfrak{l}' = 0$ on $(\kappa - e, \infty)$. By the heat equation, and since $\int_{-\infty}^{\kappa - e} (\kappa - x)^{-2} dx = 1/e$ equals the jump of $\mathfrak{l}'$ at $\kappa - e$,

$$\frac{d}{dq} \mathbb{E}[\mathfrak{l}(\sqrt{q} G)] = \frac{1}{2} \int_{-\infty}^{\kappa - e} \frac{\phi_q(\kappa - e) - \phi_q(x)}{(\kappa - x)^2} dx,$$

where ϕ_q is the $N(0, q)$ density of §2.1. Since $|\phi_q(\kappa - e) - \phi_q(x)| \leq C \min\{1, \kappa - e - x\}$ for $q \in [\frac{1}{2}, 1]$, this is at most $C \int_e^\infty \min\{1, u - e\} u^{-2} du \leq C(1 + \log(1/e))$. Also $|\frac{d}{dq} \Phi(\kappa/\sqrt{q})| \leq C|\kappa| \leq C$. Integrating over $[q_0, 1]$ gives the bound on $K_e(q_0) - K_e$. Finally,

$$K_e = (1 + \alpha\Phi(\kappa - e)) \log e - \alpha \mathbb{E}[\log(\kappa - G); G \leq \kappa - e] - \alpha\Phi(\kappa) \log \sqrt{2\pi};$$

the middle term is bounded because $\log$ is integrable near zero, and $\alpha\Phi(\kappa - e) = 1 + O(e)$.

(c) Let $k = \kappa / \sqrt{q}$. Since $\kappa < 0$, $k \leq \kappa < \kappa_{\text{RS}}$, so $\alpha m(k) \leq 1$ and $c_e(q) \geq 0$. By convexity of m , $1 - \alpha m(k) \leq \alpha m'(\kappa_{\text{RS}})(\kappa_{\text{RS}} - k)$, and

$$q(\kappa_{\text{RS}} - k) = qe^2\theta + |\kappa|(\sqrt{q} - q), \quad \sqrt{q} - q \leq \min\{1 - q, \sqrt{q}\}.$$

With $|\kappa| \leq e/2$ (by (a)) and $\alpha m'(\kappa_{\text{RS}}) \leq 5\phi_0$ for small e (by Lemma 3.1(c)), this gives the bound. $\square$

The replica-symmetric trial order parameters give an upper bound on the minimal value.

Lemma 4.7. (a) For $e > 0$, $\kappa \in \mathbb{R}$, and $0 < d < e^{-2}$, the replica-symmetric order parameter with rescaled top gap d , namely $\gamma = \mathbf{1}_{[q_0, 1]}$ with $q_0 = 1 - e^2 d$, has $s_0 = \ell_0 = d$, $\tilde{\Gamma} = 0$, $U(d, \cdot) = \log \Phi(\cdot / \sqrt{d})$ and

$$\mathcal{P}_\kappa(\gamma) = \alpha \mathbb{E} \log \Phi(Y_0 / \sqrt{d}) + \frac{1}{2} \log d + \frac{q_0}{2e^2 d} + \log e.$$

Its remainder in (4.4) is

$$W_{\text{RS},d}(y) := \log \Phi(y / \sqrt{d}) + \frac{y_-^2}{2d} - \left(\frac{1}{2} \log d - \log \sqrt{2\pi} \right) \mathbf{1}_{\{y \leq 0\}} + \log |y| \mathbf{1}_{\{y \leq -1\}},$$

and $\|W_{\text{RS},d}\|_{L^1(\mathbb{R})} \leq C_W(d)$ with C_W bounded on compact subsets of $(0, \infty)$.

(b) There are $e_{\text{R}} > 0$ and $C_{\text{RS}} < \infty$, depending only on Θ , such that for $0 < e \leq e_{\text{R}}$ and $\theta \in \Theta$,

$$\mathcal{P}_*(\kappa_e(\theta)) \leq K_e + 1 + \log(2\phi_0\theta) + C_{\text{RS}}e,$$

with K_e as in Proposition 4.5. In particular $\mathcal{P}_*(\kappa_e(\theta)) \leq K_e + C'_{\text{RS}}$ with $C'_{\text{RS}} := 1 + \log(2\phi_0\theta_+) + C_{\text{RS}}$.

Proof. (a) The first formula is (4.3) with $\Gamma(d) = \frac{1}{2} \log d$, and the formula for $W_{\text{RS},d}$ is the definition of W with $\omega(d) = \frac{1}{2} \log d - \log \sqrt{2\pi}$. With $z = y / \sqrt{d}$ and $\mathfrak{w}(z) := \log \Phi(z) + \frac{z_-^2}{2} + \log \sqrt{2\pi} \mathbf{1}_{\{z \leq 0\}} + \log |z| \mathbf{1}_{\{z \leq -1\}}$,

$$W_{\text{RS},d}(y) = \mathfrak{w}(z) + \log |z| (\mathbf{1}_{\{y \leq -1\}} - \mathbf{1}_{\{z \leq -1\}}) + \frac{1}{2} \log d (\mathbf{1}_{\{y \leq -1\}} - \mathbf{1}_{\{y \leq 0\}}).$$

For $z = -t \leq -1$, $\mathfrak{w}(-t) = \log(tM(t)) \in (-t^{-2}, 0)$, where $M(t) = \Phi(-t)/\phi(t)$ and $t^2/(1+t^2) < tM(t) < 1$ by Lemma 2.1(b); for $z \geq 0$, $|\mathfrak{w}| \leq 2(1 - \Phi(z))$ by Lemma 2.1(c); so $\mathfrak{w} \in L^1$. The second term is supported where exactly one of $y \leq -1$, $y \leq -\sqrt{d}$ holds, a set of length at most $|1 - \sqrt{d}|$ on which $|\log |z|| \leq \frac{1}{2} |\log d|$. Hence

$$\|W_{\text{RS},d}\|_{L^1} \leq \sqrt{d} \|\mathfrak{w}\|_{L^1} + \frac{1}{2} |\log d| (|1 - \sqrt{d}| + 1).$$

(b) Take $d = \ell^* := 2\phi_0\theta \in [2\phi_0\theta_-, 2\phi_0\theta_+]$ and $q_0 = 1 - e^2\ell^*$, which lies in $[\frac{1}{2}, 1]$ for small e . By (4.4), (a), and Lemma 4.6(b),

$$\mathcal{P}_*(\kappa) \leq \mathcal{P}_\kappa(\gamma) = K_e + \frac{2\phi_0\theta}{\ell^*} + \log \ell^* + O(e) + \alpha \mathbb{E} W_{\text{RS}, \ell^*}(Y_0),$$

and, since Y_0 has density at most $e\phi_0/\sqrt{q_0} \leq \sqrt{2}\phi_0e$ (Lemma 4.3) and $\alpha \leq 3$, $|\alpha \mathbb{E} W_{\text{RS}, \ell^*}(Y_0)| \leq 3\sqrt{2}\phi_0e C_W(\ell^*)$. The first two explicit terms equal $1 + \log(2\phi_0\theta)$. $\square$

Finally we record a crude lower bound on $\mathcal{P}_\kappa$ at a fixed top. Together with Lemma 4.7(b) it shows that the top gap of a global minimizer is small.

Lemma 4.8. *For every $\gamma \in \mathcal{U}$ with top Q and every $\kappa \in \mathbb{R}$,*

$$\mathcal{P}_\kappa(\gamma) \geq \alpha \mathbb{E} \log \Phi\left(\frac{B_Q - \kappa}{\sqrt{\delta}}\right) + \frac{1}{2} \log \delta.$$

Proof. The entropy satisfies $\text{CS}(\gamma) = \frac{1}{2} \int_0^Q \lambda^{-1} + \frac{1}{2} \log \delta \geq \frac{1}{2} \log \delta$. The heat lower bound of Lemma 2.5(c) with $(q, t, x) = (0, Q, 0)$ and (2.9) give $u_\gamma(0, 0; \kappa) \geq \mathbb{E} u_\gamma(Q, B_Q; \kappa) = \mathbb{E} \log \Phi((B_Q - \kappa)/\sqrt{\delta})$. $\square$

4.3 The stationary bottom

At the bottom q_0 of a global minimizer, $D(q_0) = 0$ (Proposition 2.11(c)). As in §4.2, the diffusion is a Brownian motion below q_0 , so this condition is an exact Gaussian identity at the single level s_0 . For a replica-symmetric order parameter, $D(q_0) = 0$ is the saddle-point equation of Gardner's replica-symmetric computation [62] (see Lemma 3.5). For mixed p -spin models, Auffinger and W.-K. Chen used the same observation, that the order parameter vanishes below the bottom of the support, in the proof of [8, Theorem 1].

Lemma 4.9. *Let $0 < e \leq e_B := \min\{e_L, \frac{1}{8}\}$, $\theta \in \Theta$, and $\kappa = \kappa_e(\theta)$, and let $\gamma \in \mathcal{U}$ have bottom $q_0 \in (0, 1)$ with $D(q_0) = 0$; by Proposition 2.11(c), every global minimizer qualifies. Put $k_0 := \kappa/\sqrt{q_0}$ and*

$$g(y) := (\ell_0 U_y(s_0, y))^2 - y_-^2 = (\Psi(s_0, y) - y)^2 - y_-^2.$$

(a) $g \geq 0$ on $\mathbb{R}$, and $2c_e(q_0) = \alpha \mathbb{E} g(Y_0)$, with c_e as in Proposition 4.5.

(b) Depth. $s_0 \leq \max\{2d, C_1 e^{-2/3}, C_2\}$, where

$$c_G := \frac{\phi(1)\phi_0^3}{2^{9/2}}, \quad C_1 := \left(\frac{10\phi_0\theta_+}{c_G}\right)^{2/3}, \quad C_2 := \left(\frac{5\phi_0}{c_G}\right)^2.$$

(c) Mass. If $e\sqrt{s_0} \leq \frac{1}{5}$, then $\ell_0 \leq 5\phi_0\theta_+ + e s_0 + 4e s_0^{3/2}$.

Proof. (a) Since $\gamma = 0$ on $[0, q_0]$, we have $X_{q_0} = B_{q_0}$ and $\lambda = \lambda(0) = e^2 \ell_0$ on $[0, q_0]$, so $\int_0^{q_0} \lambda^{-2} = q_0 / (e^4 \ell_0^2)$. With $b(q_0, \kappa + ey) = U_y(s_0, y)/e$ and $B_{q_0} = \kappa + eY_0$, the condition $D(q_0) = 0$ reads $\alpha \mathbb{E}(\ell_0 U_y(s_0, Y_0))^2 = q_0 / e^2$. Since $e(Y_0)_- = (\kappa - B_{q_0})_+$, $\mathbb{E}(Y_0)_-^2 = q_0 m(k_0) / e^2$. Subtracting α times this, $\alpha \mathbb{E}g(Y_0) = q_0(1 - \alpha m(k_0)) / e^2 = 2c_e(q_0)$. The second expression for g is the definition of Ψ . For $y \geq 0$, $g = (\Psi - y)^2 \geq 0$; for $y < 0$, $g = \Psi^2 + 2|y|\Psi > 0$ because $\Psi > 0$ (Lemma 4.2(a)).

Two further facts are used in (b) and (c). By Lemma 4.3, Y_0 has density $p_0 \leq e\phi_0 / \sqrt{q_0}$. By Lemma 4.6(a),(c), $-\frac{e}{2} \leq \kappa < 0$ and

$$2c_e(q_0) \leq 5\phi_0\theta_+ + \frac{5}{2}\phi_0 e s_0. \quad (4.5)$$

(b) Suppose that $s_0 > \max\{2d, C_2\}$; we show $s_0 \leq C_1 e^{-2/3}$. Put $\sigma_0 := \sqrt{s_0 - d}$ and $y_1 := \frac{1}{2}\phi_0\sigma_0$. Since $s_0 > 2d$, $\sigma_0^2 \geq s_0/2$. For $y \geq 0$, Lemma 4.2(d) and the monotonicity of $y \mapsto \mathbb{E}(y + \sigma_0 G)_+$ give $\Psi(s_0, y) \geq \mathbb{E}(\sigma_0 G)_+ = \phi_0\sigma_0$; hence $g(y) = (\Psi - y)^2 \geq \frac{1}{4}\phi_0^2\sigma_0^2$ for $0 \leq y \leq y_1$. Put $x_1 := ey_1 = \frac{1}{2}e\phi_0\sigma_0$. Then $x_1 \leq \frac{1}{2}\phi_0 < 1$ because $e^2s_0 = 1 - q_0 < 1$. Also $c_G < \phi_0^3$ gives $C_2 > (5/\phi_0^2)^2 > 16\pi$, so $\sigma_0^2 \geq s_0/2 > 8\pi = 4/\phi_0^2$ and $x_1 \geq e \geq 2|\kappa|$. As $\kappa < 0 < \kappa + x_1 \leq 1$ and $q_0 \leq 1$,

$$\begin{aligned} \mathbb{P}(0 \leq Y_0 \leq y_1) &= \mathbb{P}(\kappa \leq \sqrt{q_0} G \leq \kappa + x_1) \geq \mathbb{P}(0 \leq G \leq \kappa + x_1) \\ &\geq \phi(1)(\kappa + x_1) \geq \frac{1}{2}\phi(1)x_1. \end{aligned}$$

By (a), $g \geq 0$ and $\alpha \geq 2$,

$$2c_e(q_0) \geq \alpha \mathbb{E}[g(Y_0); 0 \leq Y_0 \leq y_1] \geq 2 \cdot \frac{\phi_0^2\sigma_0^2}{4} \cdot \frac{\phi(1)\phi_0 e\sigma_0}{4} = \frac{\phi(1)\phi_0^3}{8} e\sigma_0^3 \geq c_G e s_0^{3/2}.$$

Since $s_0 > C_2$, we have $\frac{5}{2}\phi_0 e s_0 < \frac{1}{2}c_G e s_0^{3/2}$, so (4.5) gives $\frac{1}{2}c_G e s_0^{3/2} < 5\phi_0\theta_+$, that is, $s_0 < C_1 e^{-2/3}$.

(c) Let $e\sqrt{s_0} \leq \frac{1}{5}$. Then $1 - q_0 = e^2s_0 \leq \frac{1}{25}$, so $p_0 \leq \sqrt{2}\phi_0 e$ and $\sqrt{2}\phi_0 e\sqrt{s_0} \leq \frac{\sqrt{2}\phi_0}{5} < \frac{1}{8}$; also $|k_0| \leq \sqrt{2}|\kappa| \leq e \leq \frac{1}{8}$, so $\Phi(k_0) \geq \Phi(-\frac{1}{8}) > \frac{3}{8}$. Let $y < 0$. The lower half of the sandwich (4.2) at level s_0 gives $(\Psi + |y|)\Psi = \ell_0 U_y \Psi \geq \ell_0 v$, so

$$g(y) = 2(\Psi + |y|)\Psi - \Psi^2 \geq 2\ell_0 - 2\ell_0(1 - v) - \Psi^2.$$

By Lemma 4.2(a),(c), $\Psi(s_0, \cdot)$ is increasing with $\Psi_y = 1 - v$ and $\Psi(s_0, y) \leq \min\{\sqrt{s_0}, s_0/|y|\}$ for $y < 0$. Hence

$$\begin{aligned} \int_{-\infty}^0 (1 - v)(s_0, y) dy &= \Psi(s_0, 0) \leq \sqrt{s_0}, \\ \int_{-\infty}^0 \Psi(s_0, y)^2 dy &\leq \int_0^\infty \min\left\{s_0, \frac{s_0^2}{u^2}\right\} du = 2s_0^{3/2}. \end{aligned}$$

Using $g \geq 0$ on $[0, \infty)$, $\mathbb{P}(Y_0 < 0) = \Phi(k_0)$, and $p_0 \leq \sqrt{2}\phi_0 e$,

$$2c_e(q_0) = \alpha \mathbb{E}g(Y_0) \geq \alpha \left[2\ell_0(\Phi(k_0) - \sqrt{2}\phi_0 e \sqrt{s_0}) - 2\sqrt{2}\phi_0 e s_0^{3/2} \right] \geq \ell_0 - 6\sqrt{2}\phi_0 e s_0^{3/2},$$

where we used $\Phi(k_0) - \sqrt{2}\phi_0 e \sqrt{s_0} \geq \frac{1}{4}$ and $2 \leq \alpha \leq 3$. Combining with (4.5), and using $\frac{5}{2}\phi_0 < 1$ and $6\sqrt{2}\phi_0 < 4$, gives (c). $\square$

Only the stationarity $D(q_0) = 0$ of the bottom enters Lemma 4.9, not global minimality. The two parts use disjoint regions of the Gaussian input. On $\{Y_0 < 0\}$ the sandwich makes $g \geq 2\ell_0 v - \Psi^2$, which forces $\ell_0 \lesssim c_e(q_0)$. On $\{0 \leq Y_0 \lesssim \sqrt{s_0}\}$, heat smoothing over the levels $[d, s_0]$ makes g of order s_0 on a set of probability of order $e\sqrt{s_0}$; this gain of order $es_0^{3/2}$ cannot be balanced by $2c_e(q_0) = O(1 + es_0)$ at large depth.

4.4 A priori bounds for global minimizers

The top gap is bounded below by comparison with the certified margin $k_* \leq \kappa_c$ of Theorem 3.11. At a distance $\eta \asymp e^2$ below k_* , the slack bound of Corollary 2.22 makes $\mathcal{P}_\kappa(\gamma)$ large when $\delta \ll \eta^2$, unless $\lambda(0) \gtrsim \eta$; weak top stability (Lemma 2.15) forces $\lambda(0) \lesssim \sqrt{\delta}$ for minimizers; and the value inequality of Lemma 2.13 closes the argument.

Theorem 4.10. *There are $e_D > 0$, depending only on Θ , and $c_D > 0$, depending only on θ_- , such that for $0 < e < e_D$ and $\theta \in \Theta$, every global minimizer γ of $\mathcal{P}_{\kappa_e(\theta)}$ satisfies*

$$\delta = 1 - Q \geq c_D e^4, \quad \text{that is,} \quad d \geq c_D e^2.$$

Proof. Let e be so small that Theorem 3.11 and Lemma 4.6(a) apply, so that $k_* = k_*(2 + e) < \kappa_{RS} < 0$, $k_* \leq \kappa_c$, and $\kappa = \kappa_e(\theta) \in [-\frac{e}{2}, 0)$. Put $\eta := k_* - \kappa$. By Theorem 3.11(iii), which rests on the certificate bound (3.8), $\eta \geq \frac{1}{2}\theta_- e^2$; also $\eta \leq |\kappa| \leq \frac{1}{2}$. If $\delta > \eta^2/9$, then $\delta \geq \theta_-^2 e^4/36$. Assume from now on that $\delta \leq \eta^2/9$, so that $Q \geq \frac{35}{36}$. We divide the rest of the proof into three steps.

STEP 1. *A lower bound on $\lambda(0)$.* Apply Corollary 2.22 with the margin shift $\eta/3 \geq \sqrt{\delta}$ and with $k := k_* - \eta/3 < \kappa_c$. Its hypothesis holds: $\kappa + \eta/3 = k_* - \frac{2}{3}\eta < 0$ and $Q \leq 1$, so $(\kappa + \eta/3)/\sqrt{Q} \leq \kappa + \eta/3 \leq k$. Since $k_* \leq \kappa_c$ and $m_1 > 0$ is increasing, with $k \geq \kappa \geq -1$, the constant of Corollary 2.22 satisfies $A(k) \geq \alpha \int_k^{k_*} m_1(t) dt \geq \alpha m_1(-1)\eta/3$. Hence

$$\mathcal{P}_\kappa(\gamma) \geq \frac{Q \alpha m_1(-1) \eta}{3\lambda(0)} + \frac{1}{2} \log \delta - 2\alpha.$$

Together with $\mathcal{P}_\kappa(\gamma) < \frac{1}{2} \log \delta$ (Lemma 2.13) and $Q \geq \frac{1}{2}$, this gives $\lambda(0) > m_1(-1)\eta/12$.

STEP 2. *An upper bound on $\lambda(0)$.* A global minimizer minimizes $\mathcal{P}_\kappa$ over $\mathcal{U}_r$ for every $r \in (Q, 1)$ (Proposition 2.11(c)), so (2.35) of Lemma 2.15 gives $\lambda(0) \leq \max\{1, 2\alpha C_\kappa/s_\alpha\} \sqrt{\delta} \leq$

$C_\lambda \sqrt{\delta}$ with the numerical constant $C_\lambda := 288$. Indeed $\alpha \leq 3$, $C_\kappa \leq 1$ for $\kappa \leq 0$, and $s_\alpha = \min\{\frac{1}{2}, \frac{1}{16\alpha}\} = \frac{1}{16\alpha}$ for $\alpha \geq 2$, so that $2\alpha C_\kappa / s_\alpha \leq 32\alpha^2 \leq 288$.

STEP 3. Conclusion. By Steps 1 and 2, $\sqrt{\delta} \geq m_1(-1)\eta/(12C_\lambda) \geq m_1(-1)\theta_- e^2/(24C_\lambda)$. Both cases give $\delta \geq c_D e^4$ with $c_D := (m_1(-1)\theta_-/(24C_\lambda))^2$, which is smaller than $\theta_-^2/36$ because $m_1(-1) < 1 < C_\lambda$. $\square$

The constants $m_1(-1)$ and C_λ do not depend on $\alpha \in (2, 3]$: the only α -dependent input of weak top stability is $s_\alpha = \min\{\frac{1}{2}, \frac{1}{16\alpha}\} \geq \frac{1}{48}$ for $\alpha \leq 3$. This is why the bound is uniform as $e \downarrow 0$. The argument uses the value inequality only in the direction $\mathcal{P}_\kappa(\gamma) < \frac{1}{2} \log \delta$; it gives no lower bound on the optimal value $\mathcal{P}_*$, and none is claimed.

The next theorem concerns exact global minimizers. In its proof, global minimality is used once, through comparison with the replica-symmetric trial order parameter of Lemma 4.7, to make the top gap small in original units (Step 1). The upper bounds then follow from the stationarity $D(q_0) = 0$ of the bottom (Lemma 4.9).

Theorem 4.11. *There are $e_U > 0$ and $D_{\text{top}}, L, C_0 < \infty$, depending only on Θ , such that for $0 < e < e_U$, $\theta \in \Theta$, and $\kappa = \kappa_e(\theta)$, every global minimizer γ of $\mathcal{P}_\kappa$ satisfies*

$$d \leq D_{\text{top}}, \quad \ell_0 \leq L, \quad s_0 \leq C_0 e^{-2/3}, \quad q_0 \geq \frac{1}{2}, \quad |\kappa| \leq e.$$

More precisely, $s_0 \leq \max\{2d, C_1 e^{-2/3}, C_2\}$ with C_1, C_2 as in Lemma 4.9, and one may take $D_{\text{top}} = L$.

Proof. Let $e < e_U \leq \min\{e_B, e_R\}$, with e_U further reduced below. The bound $|\kappa| \leq e$ is Lemma 4.6(a). By Proposition 2.11(c), $0 < q_0 \leq Q < 1$ and $D(q_0) = 0$, so Lemma 4.9 applies. Fix $\bar{\delta} := 2^{-12}$. We divide the proof into four steps.

STEP 1. $\delta < \bar{\delta}$. Let $\gamma \in \mathcal{U}$ have $\delta \geq \bar{\delta}$. By Lemma 2.1(c), $\log \Phi(z) \geq -\frac{1}{2}z_-^2 - z_- - 2$ for all z . For $|\kappa| \leq 1$ and $Q \leq 1$ we have $\mathbb{E}(\kappa - B_Q)_+^2 \leq 4$ and $\mathbb{E}(\kappa - B_Q)_+ \leq 2$. Hence Lemma 4.8 and $\alpha \leq 3$ give

$$\mathcal{P}_\kappa(\gamma) \geq -3\left(\frac{2}{\bar{\delta}} + \frac{2}{\sqrt{\bar{\delta}}} + 2\right) + \frac{1}{2} \log \bar{\delta} =: -C(\bar{\delta}).$$

By Lemma 4.7(b) and Lemma 4.6(b), $\mathcal{P}_*(\kappa) \leq K_e + C'_{\text{RS}} \leq 2 \log e + C_L + C'_{\text{RS}} < -C(\bar{\delta})$ for small e . Hence every global minimizer has $\delta < \bar{\delta}$.

STEP 2. Depth. By Lemma 4.9(b), $s_0 \leq \max\{2d, \sigma_e\}$ with $\sigma_e := \max\{C_1 e^{-2/3}, C_2\}$. In both cases $e\sqrt{s_0} \leq \frac{1}{5}$: if $s_0 \leq 2d$, then $e^2 s_0 \leq 2\delta < 2^{-11}$; otherwise $e^2 s_0 \leq \max\{C_1 e^{4/3}, C_2 e^2\} \leq \frac{1}{25}$ for small e . In particular $q_0 = 1 - e^2 s_0 \geq \frac{1}{2}$.

STEP 3. Mass. By Step 2 and Lemma 4.9(c), $\ell_0 \leq 5\phi_0\theta_+ + es_0 + 4es_0^{3/2}$. If $s_0 \leq 2d$, then $s_0 \leq 2\ell_0$ (as $d \leq \ell_0$), so $es_0 \leq 2e\ell_0 \leq \frac{1}{4}\ell_0$ and $4es_0^{3/2} = 4(e\sqrt{s_0})s_0 \leq 4 \cdot 2^{-11/2} \cdot 2\ell_0 < \frac{1}{4}\ell_0$;

hence $\ell_0 \leq 10\phi_0\theta_+$. Otherwise $s_0 \leq \sigma_e$, and $es_0 \leq 1$, $es_0^{3/2} \leq \max\{C_1^{3/2}, C_2^{3/2}e\} \leq C_1^{3/2} + 1$ for small e ; hence $\ell_0 \leq 5\phi_0\theta_+ + 5 + 4C_1^{3/2}$. In both cases

$$\ell_0 \leq L := 10\phi_0\theta_+ + 5 + 4C_1^{3/2}.$$

STEP 4. *Conclusion.* Since $d \leq \ell_0$, $d \leq D_{\text{top}} := L$. By Step 2, $s_0 \leq \max\{2L, C_1e^{-2/3}, C_2\} \leq C_0e^{-2/3}$ with $C_0 := 2L + C_1 + C_2$ (as $e \leq 1$), and $q_0 \geq \frac{1}{2}$. $\square$

The constants are chosen in the following order: Θ (Remark 3.12); the constant C_{ord} and the thresholds e_2, e_{Θ} of Theorem 3.11; C_1, C_2 (Lemma 4.9); then $\bar{\delta}$ and $D_{\text{top}} = L$, C_0 (Theorem 4.11); and c_D (Theorem 4.10), which does not depend on the upper bounds. Each threshold for e is chosen after the constants it involves, and none of the constants depends on the minimizer or on $\theta \in \Theta$.

Bounded depth of minimizers

The coarse bound $s_0 \leq C_0 e^{-2/3}$ of Theorem 4.11 permits thin tails of the support at unbounded rescaled depth. This chapter excludes them: Theorem 5.1 shows that every global minimizer in the window of margins has $s_0 = O(1)$, which is what places all minimizers in one compact class in Chapter 6. The proof is Step 2 of §1.3. In §5.1 the first-order conditions give the balance law of Lemma 5.2, which equates the integral of the width $w = \alpha \mathbb{E}[v(1-v)]$ over the rescaled levels with a single scalar $\tilde{\chi}$ evaluated at the bottom. In §5.2 a signed-area identity (Lemma 5.4) gives $\tilde{\chi}$ to leading order. In §5.3 we bound w from below on a fixed window and show that the two estimates are incompatible at large depth.

The condition $D = 0$ on the support is the analogue for $\mathcal{P}_\kappa$ of the condition satisfied on the support by Parisi measures, and the term $\int_0^q \lambda^{-2}$ in D comes from the Crisanti–Sommers entropy, as for spherical spin glasses; the earlier work is described in the remark after Proposition 2.11. The formula $D' = (S - 1)/\lambda^2$ is the analogue of a derivative formula of Auffinger and W.-K. Chen (see the remark after Lemma 2.9). In the physics literature, $D = 0$ on the support and $S = 1$ on its intervals are the saddle-point equations [59, (34)–(35)] of the full replica symmetry breaking solution of the perceptron.

Throughout this chapter, $\alpha = 2 + e$ with $e \in (0, 1]$ and $\kappa = \kappa_e(\theta)$ with $\theta \in \Theta$, a compact window as in Remark 3.12, unless stated otherwise. We use the rescaled variables of Definition 4.1, and the same letter for a function of q and of $s = (1 - q)/e^2$; thus

$$w(s) = \alpha \mathbb{E}[v(1-v)](s, Y_s) \quad (d \leq s < e^{-2}),$$

and p_s denotes the density of Y_s . Recall that $\gamma \in \mathcal{U}$ is replica symmetric if μ_γ is a point mass, that is, if γ is one of the order parameters $\mathbf{1}_{[q,1]}$ of Chapter 3. Since $Q \in \text{supp } \mu_\gamma$, this happens exactly when $q_0 = Q$, that is, $s_0 = d$; then $\zeta = \mathbf{1}_{[0,d]}$ and $\ell_0 = d = s_0$. Put

$$c_w := \frac{3\phi_0 m_1(-3)}{4} = \frac{3\phi_0}{4} [\phi(3) - 3\Phi(-3)] > 0; \quad (5.1)$$

it is positive by the Mills-ratio bound $\Phi(-3) < \phi(3)/3$ of Lemma 2.1(b).

Theorem 5.1. *There is $e_S > 0$, depending only on Θ , such that for $0 < e < e_S$ and $\theta \in \Theta$ every global minimizer of $\mathcal{P}_{\kappa_e(\theta)}$ satisfies the following.*

- (i) *If the minimizer is not replica symmetric, then $s_0 \leq 1/(4c_w^2)$.*

(ii) In all cases, $s_0 \leq S_* := \max\{L, 1/(4c_w^2)\}$, where L is the mass bound of Theorem 4.11. Thus $1 - q_0 \leq S_* e^2$.

The theorem concerns exact global minimizers; Remark 5.5 below shows that near-minimality of the value alone cannot bound the depth.

5.1 The balance law

The first-order conditions of Proposition 2.11 give $D = 0$ at every support point and $\chi = \chi(0) + \gamma D$. The next lemma combines them into an identity in which the unknown term γD cancels.

Lemma 5.2. *Let $\alpha > 0$, $\kappa \in \mathbb{R}$, and let γ be a KKT point of $\mathcal{P}_\kappa$ in the sense of Proposition 2.11(c); for instance, let γ be a global minimizer. Then λD is absolutely continuous on $[0, r]$ for every $r < 1$, and*

$$(\lambda D)' = \chi(0) - \frac{w}{\lambda} \quad \text{almost everywhere on } [0, 1). \quad (5.2)$$

Consequently $\int_{q_0}^Q w/\lambda \, dq = \chi(0)(Q - q_0)$. If $\alpha = 2 + e$ and $\tilde{\chi} := e^2 \chi(0)$, then

$$\tilde{\chi} = \frac{\alpha \mathbb{E}v(s_0, Y_0) - 1}{\ell_0} \quad (5.3)$$

and

$$\int_d^{s_0} \frac{w(s)}{\ell(s)} \, ds = \tilde{\chi}(s_0 - d). \quad (5.4)$$

Proof. The function λ is Lipschitz with $\lambda' = -\gamma$ almost everywhere, and $D \in C^1([0, 1))$ with $D' = (S - 1)/\lambda^2$ by Lemma 2.9(c)(i). Hence λD is absolutely continuous on $[0, r]$, and almost everywhere

$$(\lambda D)' = -\gamma D + \frac{S - 1}{\lambda} = -\gamma D + \frac{\lambda \chi - w}{\lambda} = -\gamma D + \chi(0) + \gamma D - \frac{w}{\lambda},$$

by the identity $S - 1 = \lambda \chi - w$ of Lemma 2.9(c)(v) and by $\chi = \chi(0) + \gamma D$ from Proposition 2.11(d). The points q_0 and Q lie in $\text{supp } \mu_\gamma$, so $D(q_0) = D(Q) = 0$ by Proposition 2.11(c), and integrating (5.2) over $[q_0, Q]$ gives the integrated identity. By Proposition 2.11(c),(d), χ is constant on $[0, q_0]$, $\lambda = \lambda(0) = e^2 \ell_0$ there, and $X_{q_0} = B_{q_0}$; hence

$$\chi(0) = \chi(q_0) = \frac{\alpha \mathbb{E}v(q_0, X_{q_0}) - 1}{\lambda(0)},$$

and $v(q_0, X_{q_0}) = v(s_0, Y_0)$ gives (5.3). The substitution $q = 1 - e^2 s$, $\lambda = e^2 \ell$ has $dq = -e^2 ds$ and maps q_0, Q to s_0, d ; reversing the limits gives (5.4). $\square$

The identity uses $D = 0$ at every support point, through $\chi = \chi(0) + \gamma D$; the endpoint conditions $D(q_0) = D(Q) = 0$ alone would not suffice.

Remark 5.3. The same argument gives a family of weighted balance laws, which we do not use below. Let $\varphi : [0, 1] \rightarrow \mathbb{R}$ be continuous and monotone. The measure $d(\varphi \circ \gamma)$ is carried by $\text{supp } \mu_\gamma$, because γ is constant on each component of the complement, and λD vanishes on $\text{supp } \mu_\gamma$. Integration by parts therefore gives

$$\int_{q_0}^Q \varphi(\gamma(q)) \left(\chi(0) - \frac{w(q)}{\lambda(q)} \right) dq = - \int_{(q_0, Q]} \lambda D d(\varphi \circ \gamma) = 0.$$

The weight $\varphi \equiv 1$ is Lemma 5.2. The weight $\varphi(t) = 1 - t$, together with the representation of $b\psi - v$ in Lemma 2.10(a), the identity $c(1 - v) = v(1 - v)/\lambda$, and the Markov property of X , gives the identity $s_0/\ell_0 - 1 = \alpha \text{EF}(s_0, Y_0)$ at the bottom. It involves only the solution at the level s_0 and the Gaussian law of Y_0 , and it expresses stationarity under dilations. For c close to one let $\gamma_c \in \mathcal{U}$ have rescaled profile $\zeta(\cdot/c)$. Its rescaled solution at level cs is $U(s, y/\sqrt{c})$, by the Cole–Hopf recursion of Lemma 2.5(a), Brownian scaling, and approximation in L^1 (Lemma 2.4(f)). Differentiating (4.3) for γ_c at $c = 1$ gives

$$\frac{d}{dc} \mathcal{P}_\kappa(\gamma_c) \Big|_{c=1} = \frac{1}{2} \left(\alpha \text{EF}(s_0, Y_0) - \frac{s_0}{\ell_0} + 1 \right) + \frac{1}{2} \lambda(0) D(q_0).$$

Thus, when $D(q_0) = 0$, stationarity under dilations of the rescaled profile carries no information beyond the weighted balance laws. The dilation $\gamma \mapsto \gamma_c$ pushes μ_γ forward by the affine map $q \mapsto 1 - c(1 - q)$, so it is a transport variation of the kind used by Talagrand [126, Lemma 3.7], and by Auffinger and W.-K. Chen to derive the conditions on the support of Parisi measures [8, Section 3].

5.2 Signed-area cancellation

The bottom scalar $\tilde{\chi}$ of (5.3) is a Gaussian average of $v(s_0, \cdot)$. At every level, v passes from one to zero around $y = 0$, and the excess of Ψ over its asymptote measures the signed area between v and the step function $\mathbf{1}_{\{y < 0\}}$. At a level $s \in [d, e^{-2}]$ put

$$\mathfrak{o}_s(y) := \Psi(s, y) - y_+ = \begin{cases} \Psi(s, y), & y \leq 0, \\ \ell(s) U_y(s, y), & y > 0, \end{cases} \quad (5.5)$$

the two expressions agreeing at $y = 0$ because $\Psi = \ell U_y + y$.

Lemma 5.4. *Let $e > 0$, $\kappa \in \mathbb{R}$, $\gamma \in \mathcal{U}$, and $s \in [d, e^{-2}]$, and write $v = v(s, \cdot)$. The function $\mathfrak{o}_s$ is Lipschitz, tends to zero at $\pm\infty$, and satisfies*

$$\mathfrak{o}'_s = \mathbf{1}_{\{y < 0\}} - v \quad \text{almost everywhere,} \quad 0 < \mathfrak{o}_s \leq \sqrt{s}, \quad |y| \mathfrak{o}_s(y) \leq s \quad (y \in \mathbb{R}).$$

In particular $v - \mathbf{1}_{\{y < 0\}}$ is integrable, and

$$\int_{-\infty}^0 (1 - v(y)) dy = \Psi(s, 0) = \int_0^{\infty} v(y) dy, \quad \int_{\mathbb{R}} (v(y) - \mathbf{1}_{\{y < 0\}}) dy = 0. \quad (5.6)$$

For every Gaussian random variable Y' with mean $\bar{y}$ and variance $\sigma^2 > 0$,

$$\mathbb{E}v(Y') - \mathbb{P}(Y' < 0) = -\frac{\mathbb{E}[\mathfrak{o}_s(Y')(Y' - \bar{y})]}{\sigma^2}, \quad |\mathbb{E}v(Y') - \mathbb{P}(Y' < 0)| \leq \frac{s + |\bar{y}|\sqrt{s}}{\sigma^2}. \quad (5.7)$$

Proof. By Lemma 4.2(a), $\Psi_y = 1 - v$ and $(\ell U_y)_y = -v$, with $0 < v < 1$; this gives the formula for $\mathfrak{o}'_s$ and the Lipschitz property. By Lemma 4.2(a),(c), for $y \leq 0$ we have $0 < \Psi(s, y) \leq \min\{\sqrt{s}, s/|y|\}$, and for $y > 0$ we have $0 < \ell U_y(s, y) = \Psi(s, y) - y \leq \min\{\sqrt{s}, s/y\}$. These are the two bounds, and they give the zero limits at $\pm\infty$. Integrating $\mathfrak{o}'_s$ over each half-line gives (5.6). If $p_{Y'}$ is the density of Y' , then $p'_{Y'} = -(y - \bar{y})p_{Y'}/\sigma^2$ and $v - \mathbf{1}_{\{y < 0\}} = -\mathfrak{o}'_s$. Integration by parts, in which the boundary terms vanish because $\mathfrak{o}_s$ is bounded and $p_{Y'}$ decays, gives the identity in (5.7). The bounds $|y|\mathfrak{o}_s \leq s$ and $\mathfrak{o}_s \leq \sqrt{s}$ give $|\mathbb{E}[\mathfrak{o}_s(Y')(Y' - \bar{y})]| \leq s + |\bar{y}|\sqrt{s}$. $\square$

The cancellation (5.6) is essential below. The cruder bound $\|p_{s_0}\|_{\infty} \int_{\mathbb{R}} |v - \mathbf{1}_{\{y < 0\}}| dy = 2\|p_{s_0}\|_{\infty} \Psi(s_0, 0)$ is of order $e\sqrt{s_0}$, which is the same order as the width contribution in (5.4).

5.3 Proof of the bounded-depth theorem

Proof of Theorem 5.1. Fix a global minimizer γ . By Theorem 4.11 and Lemma 4.6(a), for e small enough, depending only on Θ ,

$$q_0 \geq \frac{1}{2}, \quad s_0 \leq C_0 e^{-2/3}, \quad \ell_0 \leq L, \quad -\frac{e}{2} \leq \kappa < 0.$$

If γ is replica symmetric, then $s_0 = d = \ell_0 \leq L \leq S_*$. Assume from now on that $s_0 > d$. We divide the rest of the proof into three steps.

STEP 1. The bottom scalar. By Lemma 4.3, Y_0 is Gaussian with mean $-\kappa/e \in (0, \frac{1}{2}]$ and variance $q_0/e^2 \geq 1/(2e^2)$, and $\mathbb{P}(Y_0 < 0) = \Phi(\kappa/\sqrt{q_0})$. Applying (5.7) at the level s_0 and using $\alpha \leq 3$ gives

$$|\alpha \mathbb{E}v(s_0, Y_0) - \alpha \Phi(\kappa/\sqrt{q_0})| \leq 6e^2(s_0 + \frac{1}{2}\sqrt{s_0}) \leq Ce^2(1 + s_0).$$

Combined with the expansion $\alpha \Phi(\kappa/\sqrt{q_0}) - 1 = \frac{e}{4} + O(e^2 + e^3 s_0)$ of Lemma 4.6(b), (5.3) yields

$$\tilde{\chi} \ell_0 = \frac{e}{4} + O(e^2(1 + s_0)) \leq \frac{e}{3} \quad (5.8)$$

for small e : since $s_0 \leq C_0 e^{-2/3}$, the error is $O(e^2 + e^{4/3}) = o(e)$, uniformly in the minimizer.

STEP 2. *A fixed positive window.* Let $s \in [d, s_0]$, $q = 1 - e^2 s$, and $x = \kappa + ey$. By Lemma 4.2(d) and the pathwise comparison (2.15), since $\bar{R}$ is increasing,

$$\Psi(s, y) \geq \mathbb{E} \left[\sqrt{d} \bar{R} \left(\frac{y + \sqrt{s-d} G}{\sqrt{d}} \right) \right].$$

For every z , $\sqrt{d} \bar{R}(z/\sqrt{d}) \geq \sqrt{d} \Phi(z/\sqrt{d}) \bar{R}(z/\sqrt{d}) = \sqrt{d} m_1(z/\sqrt{d}) = \mathbb{E}(z + \sqrt{d} G')_+$ by Lemma 2.1(g), where G' is a standard Gaussian independent of G . Adding the two independent Gaussian variances,

$$\Psi(s, y) \geq \mathbb{E}(y + \sqrt{s} G)_+ = \sqrt{s} m_1(y/\sqrt{s}) \quad (y \in \mathbb{R}). \quad (5.9)$$

Since $v(s, \cdot)$ is decreasing and $\int_0^\infty v(s, y) dy = \Psi(s, 0) \leq \sqrt{s}$ by (5.6) and Lemma 4.2(c),

$$v(s, 2\sqrt{s}) \leq \frac{1}{2\sqrt{s}} \int_0^{2\sqrt{s}} v(s, y) dy \leq \frac{1}{2}.$$

The upper half of the sandwich (4.2) reads $sv \geq \ell U_y \Psi = (\Psi - y)\Psi$. At $y = 3\sqrt{s}$ we have $\Psi \geq 3\sqrt{s}$ by Lemma 4.2(c), and $\Psi - 3\sqrt{s} \geq \sqrt{s}(m_1(3) - 3) = \sqrt{s} m_1(-3)$ by (5.9), because $m_1(k) - m_1(-k) = \mathbb{E}(k - G) = k$. Hence

$$v(s, 3\sqrt{s}) \geq 3m_1(-3).$$

Since v is decreasing and $3m_1(-3) \leq \frac{1}{2}$, throughout the interval $[2\sqrt{s}, 3\sqrt{s}]$

$$v(s, y)(1 - v(s, y)) \geq 3m_1(-3)(1 - 3m_1(-3)) \geq \frac{3}{2}m_1(-3).$$

On this interval the density floor (2.18), applied to the density p_q^X of X_q , gives

$$p_s(y) = e p_q^X(\kappa + ey) \geq e \Phi(y/\sqrt{s}) \phi_q(\kappa + ey) \geq \frac{e\phi_0}{4}.$$

Indeed $\Phi(y/\sqrt{s}) \geq \Phi(2) \geq \frac{1}{2}$; and $q \geq q_0 \geq \frac{1}{2}$ gives $\phi_q(x) \geq \phi_0 e^{-x^2}$, while $|\kappa + ey| \leq \frac{e}{2} + 3e\sqrt{s_0} \leq \frac{e}{2} + 3\sqrt{C_0}e^{2/3}$, whose square is at most $\log 2$ for small e , so that $\phi_q(\kappa + ey) \geq \phi_0/2$. Integrating over the interval, which has length $\sqrt{s}$, and using $\alpha \geq 2$ proves

$$w(s) \geq 2 \cdot \frac{3}{2}m_1(-3) \cdot \frac{e\phi_0}{4} \cdot \sqrt{s} = c_w e \sqrt{s} \quad (d \leq s \leq s_0). \quad (5.10)$$

STEP 3. *The scalar inequality.* Since ℓ is nondecreasing, $\ell(s) \leq \ell_0$ on $[d, s_0]$. Inserting (5.8) and (5.10) into (5.4) gives

$$\frac{c_w e}{\ell_0} \cdot \frac{2}{3}(s_0^{3/2} - d^{3/2}) \leq \int_d^{s_0} \frac{w(s)}{\ell(s)} ds = \tilde{\chi}(s_0 - d) \leq \frac{e}{3\ell_0}(s_0 - d).$$

Cancel $e/(3\ell_0)$ and use $s_0^{3/2} - d^{3/2} - \sqrt{s_0}(s_0 - d) = d(\sqrt{s_0} - \sqrt{d}) \geq 0$. Since $s_0 > d$, division by $s_0 - d$ gives $2c_w\sqrt{s_0} \leq 1$. This proves (i), and together with the replica-symmetric case it proves (ii). $\square$

The factors e and ℓ_0 cancel in the last display, so the proof uses no lower bound on d or ℓ_0 ; it also uses no bound on the number of steps of γ and no approximation by stationary step profiles. The obstacle inequality $S \leq 1$ is not used either: the balance law uses only $D = 0$ on the support. The coarse depth bound of Theorem 4.11 is used only to make the bottom error in (5.8) of order $o(e)$ and to keep the window $[2\sqrt{s}, 3\sqrt{s}]$ in the region where the density floor is of order e . The conclusion says that the entire rescaled profile of a global minimizer is supported in $[0, S_*]$.

Remark 5.5. For fixed e and θ , let γ be a global minimizer and choose any $S \in (s_0, e^{-2})$. Put $q_1 = 1 - e^2S$ and, for $0 < \varepsilon < 1$, let $\gamma_\varepsilon = \max\{\gamma, \varepsilon \mathbf{1}_{[q_1, 1]}\}$. Then $\gamma_\varepsilon \in \mathcal{U}$ has rescaled depth S , it lies in the same class $\mathcal{U}_r$ as γ for every $r \in (Q, 1)$, and

$$\|\gamma_\varepsilon - \gamma\|_{L^1} \leq \varepsilon e^2(S - d) \longrightarrow 0 \quad (\varepsilon \downarrow 0).$$

By Lemma 2.4(f), $\mathcal{P}_\kappa(\gamma_\varepsilon) \rightarrow \mathcal{P}_\kappa(\gamma) = \mathcal{P}_*(\kappa)$. Thus profiles of every admissible depth have values arbitrarily close to the minimum, and near-minimality of the value alone cannot bound the depth, even for fixed e ; exact stationarity is what excludes a tail whose mass tends to zero.

Compactness of minimizers

This chapter proves that, uniformly for $\theta \in \Theta$ and small e , the rescaled profile of every global minimizer lies in a fixed compact class. Theorem 5.1 bounds the depth by $s_0 \leq S_*$. What remains are lower bounds on the rescaled mass ℓ_0 and on the rescaled top gap d , with constants that do not depend on e or θ .

This is Step 3 of §1.3. The main tool is the jamming identity of Lemma 6.3, which is exact for every $\alpha > 0$, every $\kappa \in \mathbb{R}$, and every order parameter γ with $q_0 > 0$. It writes $\mathcal{P}_\kappa(\gamma)$ as a replica-symmetric expression $\Pi(\ell_0; q_0)$ in the mass and the bottom, plus an integral over the *jammed* levels, where $\zeta < 1$, with weight $(1 - \zeta)/(2\ell)$ and an explicit *jamming coefficient* $\mathcal{J}$. Near $\alpha = 2$ this coefficient is bounded below in Lemma 6.7: a jammed top is expensive.

The proof of Theorem 6.1 has three steps. First, a profile of bounded depth and small mass has a value above the replica-symmetric value, so every global minimizer has $\ell_0 \geq \ell_{\min}$ (Proposition 6.10). Second, the first-order conditions of Proposition 2.11 bound the mass $\gamma(Q-)$ just below the top atom of every global minimizer by a fixed $m_* < 1$ (Theorem 6.13); the long estimates for this step are proved in Chapter B. Third, a profile of bounded depth, with mass at least $\ell_{\min}$, top jump $1 - \gamma(Q-) \geq 1 - m_*$, and a small rescaled top gap, has a larger value than the replica-symmetric order parameter of the same mass (Proposition 6.14). Hence $d > d_1$ for every global minimizer.

The value comparisons in the first and third steps are inequalities for all profiles; only their conclusions use minimality. The third step uses the bounds of the first two as hypotheses, and the first uses the bounded depth. The constants are therefore chosen in the order

$$\Theta \longrightarrow (D_{\text{top}}, L, C_0, c_D, C'_{\text{RS}}) \longrightarrow S_* \longrightarrow \ell_{\min} \longrightarrow m_* \longrightarrow d_1, \quad (6.1)$$

and each is independent of e and θ . The constant m_* is in fact numerical (Theorem 6.13); it appears after $\ell_{\min}$ in (6.1) because d_1 depends on both. For $0 < d_1 \leq S$ let $\mathcal{K}(d_1, S)$ be the set of nonincreasing functions $\zeta: [0, \infty) \rightarrow [0, 1]$ with $\zeta = 1$ almost everywhere on $[0, d_1]$ and $\zeta = 0$ almost everywhere on (S, ∞) , where the rescaled profile of Definition 4.1 is extended by zero to (e^{-2}, ∞) . The rescaled profile of $\gamma \in \mathcal{U}$ lies in $\mathcal{K}(d_1, S)$ exactly when $d \geq d_1$ and $s_0 \leq S$. By Helly's selection theorem and dominated convergence, this class is compact in L^1 . It is the setting of the uniform expansion of Chapter 7 and of the support localization of Chapter 8.

Theorem 6.1. *There are constants $\ell_{\min} > 0$, $m_* \in (0, 1)$, $d_1 > 0$, and $e_* > 0$, depending only on Θ and chosen in the order (6.1), such that the following holds for $0 < e < e_*$ and $\theta \in \Theta$. Every global minimizer γ of $\mathcal{P}_{\kappa_e(\theta)}$ satisfies*

$$s_0 \leq S_*, \quad \ell_0 \geq \ell_{\min}, \quad \gamma(Q-) \leq m_*, \quad d > d_1.$$

In particular the rescaled profile of γ belongs to $\mathcal{K}(d_1, S_*)$.

The proof is given in §6.8, after the three steps. The relation of the second step to earlier work is described in §6.6.

6.1 Conventions and inputs

Throughout this chapter, except in §§6.2 and 6.3, $\alpha = 2 + e$ and $\kappa = \kappa_e(\theta) = \kappa_{\text{RS}}(2 + e) - e^2\theta$ with $\theta \in \Theta$, where $\Theta \subset (0, \infty)$ is a compact interval as in Remark 3.12; as there, $\theta_- = \min \Theta$ and $\theta_+ = \max \Theta$. Unless stated otherwise, c and C denote positive constants that depend only on Θ and on the constants named in the statement at hand. They never depend on e , on θ , or on the profile, and their values may change from line to line.

We use the following inputs from Chapters 4 and 5. Let $e_1 > 0$ be smaller than the thresholds of Theorems 4.10, 4.11, and 5.1, of Lemma 4.6 and of Lemma 4.7(b). For $0 < e < e_1$, $\theta \in \Theta$, and every global minimizer γ of $\mathcal{P}_{\kappa_e(\theta)}$:

- (I1) $d \leq D_{\text{top}}$, $\ell_0 \leq L$, $s_0 \leq C_0 e^{-2/3}$, and $q_0 \geq \frac{1}{2}$ (Theorem 4.11);
- (I2) $d \geq c_D e^2$ for a constant $c_D > 0$ (Theorem 4.10);
- (I3) $s_0 \leq S_*$ (Theorem 5.1);
- (I4) $\mathcal{P}_\kappa(\gamma) = \mathcal{P}_*(\kappa) \leq K_e + C'_{\text{RS}}$, where K_e is the profile-independent constant of Proposition 4.5 and C'_{RS} is the constant of Lemma 4.7(b).

Moreover $-e/2 \leq \kappa < 0$ by Lemma 4.6(a). The resulting bound $|\kappa| \leq e$ does not involve the profile, and we use it for every profile. The constant L chosen in the proof of Theorem 4.11 satisfies $L \geq 5$, and $S_* \geq L$ by Theorem 5.1; in particular $L \geq 1$ and $S_* \geq 1$.

Rescaled variables. We use the rescaled variables $s, y, \zeta, \ell, d, s_0, \ell_0, U, v, \Psi, \Gamma, \tilde{\Gamma}, Y_0$ of Definition 4.1 and the properties of U, v , and Ψ listed in Lemma 4.2. They are defined for every $e > 0$, $\kappa \in \mathbb{R}$, and $\gamma \in \mathcal{U}$. Recall that ζ is nonincreasing and left-continuous, that $\zeta = 1$ on $[0, d]$ and $\zeta < 1$ on (d, ∞) , that $\zeta > 0$ on $[0, s_0)$ and $\zeta = 0$ on (s_0, ∞) , and that $\zeta(d+) = \gamma(Q-)$. Hence ℓ is concave and nondecreasing with $\ell(0) = 0$ and $\ell(d) = d$, so that $\ell \geq d$ on $[d, \infty)$ and

$$\frac{\ell_0}{s_0} t \leq \ell(t) \leq t \quad (0 \leq t \leq s_0). \quad (6.2)$$

Since $(\log \ell)' = \zeta/\ell$ on $[d, s_0]$ and $\Gamma(s) = \frac{1}{2} \log d + \frac{1}{2} \int_d^s \ell^{-1}$,

$$\tilde{\Gamma} = \frac{1}{2} \log \frac{d}{\ell_0} + \frac{1}{2} \int_d^{s_0} \frac{dt}{\ell(t)} = \frac{1}{2} \int_d^{s_0} \frac{1 - \zeta(t)}{\ell(t)} dt. \quad (6.3)$$

If $q_0 > 0$, then Y_0 is Gaussian with mean $-\kappa/e$ and variance q_0/e^2 (Lemma 4.3); we denote its density by p_0 .

Coefficients. The quantities $K_e(q_0)$, $c_e(q_0)$, and $\rho = \alpha\Phi(\kappa/\sqrt{q_0})$ of Proposition 4.5 depend on the profile only through q_0 . For $q \in (0, 1]$ we write $K_e(q)$, $c_e(q)$, and $\rho_q := \alpha\Phi(\kappa/\sqrt{q})$ for the same formulas with q in place of q_0 . Let $S \geq 1$. By Lemma 4.6(b) there are $e_K(S) \in (0, (2S)^{-1/2}]$ and $C'_S < \infty$, depending only on S and Θ , such that for $0 < e \leq e_K(S)$, $\theta \in \Theta$, and $q \in [1 - e^2 S, 1]$,

$$c_e(q) \geq 2\phi_0\theta - C'_S e, \quad K_e(q) \geq K_e - C'_S e^2 \log(1/e). \quad (6.4)$$

We also use the elementary bound

$$|\rho_q - 1| \leq 3e \quad (0 < e \leq 1, |\kappa| \leq e, q \in [\frac{1}{2}, 1]). \quad (6.5)$$

Indeed, $|\Phi(\kappa/\sqrt{q}) - \frac{1}{2}| \leq \phi_0|\kappa|/\sqrt{q} \leq \sqrt{2}\phi_0 e$ and $\rho_q - 1 = 2(\Phi(\kappa/\sqrt{q}) - \frac{1}{2}) + e\Phi(\kappa/\sqrt{q})$, so $|\rho_q - 1| \leq 2\sqrt{2}\phi_0 e + e(\frac{1}{2} + \sqrt{2}\phi_0 e) \leq (3\sqrt{2}\phi_0 + \frac{1}{2})e \leq 3e$.

6.2 Approximation by step profiles at fixed e

The jamming identity of §6.3 is first proved for step profiles, on whose blocks the solution is smooth, and then extended by approximation. The following lemma provides the approximation. It works at a fixed value of e and on one common cutoff, and it preserves exactly the rescaled quantities on which our constants depend. The approximants are never asserted to be minimizers or KKT points, and no limit in e is interchanged with the limit $n \rightarrow \infty$.

Lemma 6.2. Fix $e > 0$, $\kappa \in \mathbb{R}$, and $\gamma \in \mathcal{U}$ with $q_0 > 0$, with rescaled data (d, s_0, ℓ_0) . Let $T \subset [d, s_0]$ be a finite set containing d and s_0 . For $n \geq 1$ let $\mathcal{P}_n$ be the partition of $[d, s_0]$ generated by T and the points $d + k2^{-n}(s_0 - d)$, $0 \leq k \leq 2^n$. Define $\zeta_n = 1$ on $[0, d]$, $\zeta_n = 0$ on (s_0, ∞) , and, on each cell $(t', t'']$ of $\mathcal{P}_n$, $\zeta_n \equiv (t'' - t')^{-1} \int_{t'}^{t''} \zeta$. Let $\gamma_n(q) = \zeta_n((1 - q)/e^2)$ for $q \geq q_0$ and $\gamma_n(q) = 0$ for $q < q_0$. Then:

- (a) $\gamma_n \in \mathcal{U}$ is a step profile with the same q_0 , the same top gap d , and the same depth s_0 as γ . Moreover $\int_0^t \zeta_n = \int_0^t \zeta$ for every point t of $\mathcal{P}_n$ (in particular $\ell_n(s_0) = \ell_0$), $\zeta_n(d+) \leq \zeta(d+)$, and $\zeta_n \geq \zeta(t)$ on $[0, t]$ for every $t \in T$.
- (b) γ_n and γ belong to $\mathcal{U}_r$ with $r := 1 - e^2 d$, and $\|\gamma_n - \gamma\|_{L^1(0,1)} = e^2 \|\zeta_n - \zeta\|_{L^1(0,\infty)} \rightarrow 0$.
- (c) $\sup_{t \geq 0} |\ell_n(t) - \ell(t)| \rightarrow 0$ and $\Gamma_n(t) \rightarrow \Gamma(t)$ for every $t \geq d$.

(d) $\mathcal{P}_\kappa(\gamma_n) \rightarrow \mathcal{P}_\kappa(\gamma)$, and $U_n(t, y) \rightarrow U(t, y)$ for every $t \in [d, e^{-2}]$ and $y \in \mathbb{R}$.

Proof. If $s_0 = d$, the profile is replica symmetric, $\zeta_n = \zeta = \mathbf{1}_{[0, d]}$, and every assertion is immediate. We may therefore assume $s_0 > d$.

(a) Averages of a nonincreasing function over consecutive intervals are nonincreasing, the first average is at most $\zeta(d+)$, and all averages are positive because $\zeta > 0$ on $[0, s_0]$. The first average is smaller than one because $\zeta < 1$ on (d, ∞) . Thus ζ_n is a nonincreasing left-continuous step function with $\zeta_n = 1$ exactly on $[0, d]$ and $\zeta_n > 0$ exactly on $[0, s_0]$. Hence $\gamma_n \in \mathcal{U}$ has the same q_0, d , and s_0 . Integrals over cells are preserved by construction. If $t \in T$ and $(t', t'') \subset [0, t]$ is a cell, then $\zeta \geq \zeta(t)$ on (t', t'') , so the average over this cell is at least $\zeta(t)$. (b) Both profiles equal one on $[1 - e^2 d, 1]$. For every continuity point $s \in (d, s_0)$ of ζ , the cell containing s shrinks to s , so $\zeta_n(s) \rightarrow \zeta(s)$. A monotone function has countably many discontinuities and $0 \leq \zeta_n \leq 1$, so $\zeta_n \rightarrow \zeta$ in L^1 . The identity for $\|\gamma_n - \gamma\|_{L^1}$ is the change of variables $q = 1 - e^2 s$. (c) $|\ell_n(t) - \ell(t)| \leq \|\zeta_n - \zeta\|_{L^1}$. For $t \geq d$ both ℓ_n and ℓ are at least d on $[d, t]$, so $|\Gamma_n(t) - \Gamma(t)| \leq (t - d) \sup |\ell_n - \ell| / (2d^2)$. (d) By (b) and Lemma 2.4(f), $\mathcal{P}_\kappa$ is continuous on $\mathcal{U}_r$ in L^1 , and $u_{\gamma_n}(q, x; \kappa) \rightarrow u_\gamma(q, x; \kappa)$ for each $q \leq r$ and $x \in \mathbb{R}$. $\square$

6.3 The jamming identity

In this section $\alpha > 0$, $\kappa \in \mathbb{R}$, and $e > 0$ are arbitrary, and the relation $\alpha = 2 + e$ is not used; e only fixes the units of the rescaled variables. The next lemma compares the solution U , at every level, with the replica-symmetric solution of the same mass. For $\ell > 0$ and $q \in (0, 1]$ put

$$Y^{(q)} := \frac{\sqrt{q} G - \kappa}{e}, \quad \Pi(\ell; q) := \alpha \mathbb{E} \log \Phi(Y^{(q)} / \sqrt{\ell}) + \frac{q}{2e^2 \ell} + \frac{1}{2} \log(e^2 \ell). \quad (6.6)$$

This is the replica-symmetric expression with the mass ℓ and the bottom q decoupled. If $e^2 \ell < 1$, then $\Pi(\ell; 1 - e^2 \ell) = \mathcal{P}_\kappa^{\text{RS}}(1 - e^2 \ell)$ is the value of the replica-symmetric order parameter $\mathbf{1}_{[1 - e^2 \ell, 1]}$ of rescaled mass ℓ , by Lemma 4.7(a). If $\gamma \in \mathcal{U}$ has $q_0 > 0$, then $Y^{(q_0)}$ has the law of Y_0 . The dependence on the bottom is explicit:

$$\partial_q \Pi(\ell; q) = \frac{1 - \alpha \mathbb{E} V(Y^{(q)} / \sqrt{\ell})}{2e^2 \ell} \quad (0 < q \leq 1). \quad (6.7)$$

Indeed, put $f(x) := \log \Phi((x - \kappa) / (e\sqrt{\ell}))$, so that $\mathbb{E} \log \Phi(Y^{(q)} / \sqrt{\ell}) = \mathbb{E} f(\sqrt{q} G)$. Since $f'' = -V((\cdot - \kappa) / (e\sqrt{\ell})) / (e^2 \ell)$ is bounded and continuous and f has at most quadratic growth (Lemma 2.1(c)), the heat equation $\partial_q \phi_q = \frac{1}{2} \phi_q''$ and two integrations by parts give $\frac{d}{dq} \mathbb{E} f(\sqrt{q} G) = \frac{1}{2} \mathbb{E} f''(\sqrt{q} G)$.

The comparison function is

$$g(s, y) := \log \Phi(y / \sqrt{\ell(s)}) + \frac{1}{2} \log \ell(s) - \Gamma(s) \quad (s \in [d, s_0], y \in \mathbb{R}). \quad (6.8)$$

It satisfies $g(d, \cdot) = U(d, \cdot)$, and it is a subsolution of the rescaled Parisi equation whose defect is proportional to $1 - \zeta$; see (6.12) below.

Lemma 6.3. *Let $\alpha > 0$, $\kappa \in \mathbb{R}$, and $e > 0$, let $\gamma \in \mathcal{U}$ with $q_0 > 0$, and put $w := U - g$ on $[d, s_0] \times \mathbb{R}$. For $(s, y) \in [d, s_0] \times \mathbb{R}$ let $\hat{Y} = \hat{Y}^{s,y}$ be the solution, indexed by the decreasing level t , of*

$$\hat{Y}_t = y + \int_t^s \frac{1}{2} \zeta(r) (U_y + g_y)(r, \hat{Y}_r) dr + B_{s-t}, \quad d \leq t \leq s, \quad (6.9)$$

where $(B_\tau)_{\tau \geq 0}$ is a standard Brownian motion. This equation has a unique strong solution, and the following hold.

(a) For $s \in [d, s_0]$ and $y \in \mathbb{R}$,

$$w(s, y) = \mathbb{E} \int_d^s \frac{1 - \zeta(t)}{2\ell(t)} \left(1 - V(\hat{Y}_t^{s,y} / \sqrt{\ell(t)}) \right) dt \geq 0.$$

(b) $\hat{Y}_t \geq y + B_{s-t}$ and $\hat{Y}_t / \ell(t) \geq y / \ell(s) + N_t$ for $t \in [d, s]$, where $N_t := \int_0^{s-t} dB_\tau / \ell(s - \tau)$ is a centered Gaussian variable.

(c) Define the jamming coefficient

$$\mathcal{J}(t) := \alpha \int_{\mathbb{R}} p_0(y) \mathbb{E} \left[1 - V(\hat{Y}_t^{s_0,y} / \sqrt{\ell(t)}) \right] dy - (\alpha - 1) \quad (d \leq t \leq s_0). \quad (6.10)$$

Then

$$\mathcal{P}_\kappa(\gamma) = \Pi(\ell_0; q_0) + \int_d^{s_0} \frac{1 - \zeta(t)}{2\ell(t)} \mathcal{J}(t) dt. \quad (6.11)$$

Proof. We divide the proof into five steps. Part (a) is first proved for step profiles and then extended to all profiles by Lemma 6.2 (Steps 2 and 3); part (b) is proved directly (Step 4), and part (c) follows from part (a) (Step 5).

STEP 1. Well-posedness and the defect of g . Write $z = y / \sqrt{\ell(s)}$. Then $g_y = R(z) / \sqrt{\ell}$ and $g_{yy} = -V(z) / \ell$, while $U_{yy} = -v / \ell$ with $0 < v < 1$ (Lemma 4.2(a)). Since $\ell \geq d$ on $[d, s_0]$, the drift $\hat{b} := \frac{1}{2} \zeta (U_y + g_y)$ of (6.9) is Lipschitz in y with constant $1/d$, uniformly in the level. It is measurable in the level and nonnegative, and it has linear growth, because $0 < U_y \leq C(1 + |y|)$ by Lemma 2.4(b) and $0 < R(z) \leq |z| + 2$ by Lemma 2.1(b) (on $[-1, \infty)$ one has $R \leq R(-1) < 2$). Hence (6.9) has a unique strong solution, with moments of every order, which depends continuously on y . Let s be a level at which ℓ is differentiable with $\ell'(s) = \zeta(s)$. Then $\partial_s z = -\zeta z / (2\ell)$ and $\Gamma' = 1 / (2\ell)$, so $g_s = -\zeta z R(z) / (2\ell) + (\zeta - 1) / (2\ell)$. Using $V = zR + R^2$,

$$2\ell \left(g_s - \frac{1}{2} g_{yy} - \frac{1}{2} \zeta g_y^2 \right) = -\zeta z R + \zeta - 1 + V - \zeta R^2 = -(1 - \zeta)(1 - V(z)). \quad (6.12)$$

STEP 2. Part (a) for step profiles. Let γ be a step profile, with $\zeta = m_j$ on $(t_{j-1}, t_j]$ for $d = t_0 < t_1 < \dots < t_n = s_0$. On each closed block $[t_{j-1}, t_j] \times \mathbb{R}$ the function U is C^∞ , with

one-sided derivatives at the ends, and $U_s = \frac{1}{2}(U_{yy} + m_j U_y^2)$ (Lemma 2.4(a), Lemma 4.2(a)); since ℓ is affine on the block, the same smoothness holds for g . Subtracting (6.12) from the equation for U and writing $U_y^2 - g_y^2 = (U_y + g_y)w_y$ gives, on each block,

$$w_s = \frac{1}{2}w_{yy} + \hat{b}w_y + \frac{1-\zeta}{2\ell}(1 - V(z)). \quad (6.13)$$

Moreover $w(d, \cdot) = 0$, because $U(d, y) = \log \Phi(y/\sqrt{d})$ (Lemma 4.2(a)), $\ell(d) = d$, and $\Gamma(d) = \frac{1}{2} \log d$; and w is continuous on $[d, s_0] \times \mathbb{R}$. Fix (s, y) . By (6.9), $\tau \mapsto \hat{Y}_{s-\tau}$ has drift $\hat{b}(s - \tau, \cdot)$ and noise dB_τ . On each block, Itô's formula and (6.13) give

$$dw(s - \tau, \hat{Y}_{s-\tau}) = -\left[\frac{1-\zeta}{2\ell}\left(1 - V(\hat{Y}/\sqrt{\ell})\right)\right](s - \tau) d\tau + w_y(s - \tau, \hat{Y}_{s-\tau}) dB_\tau.$$

The stochastic integral is a martingale, since $w_y = U_y - g_y$ has linear growth and $\hat{Y}$ has moments of every order. Summing over the blocks between d and s , which is possible because w is continuous, taking expectations, and using $w(d, \cdot) = 0$ gives the formula in (a). The integrand is nonnegative because $\zeta \leq 1$ and $V < 1$.

STEP 3. *Part (a) for general profiles.* Let $\gamma \in \mathcal{U}$ with $q_0 > 0$, and let γ_n be the step profiles of Lemma 6.2 with $T = \{d, s_0\}$. They have the same q_0 , d , and s_0 as γ . We mark the objects of γ_n by an index n . Fix (s, y) and drive $\hat{Y}^n = \hat{Y}^{n,s,y}$ and $\hat{Y} = \hat{Y}^{s,y}$ by the same Brownian motion. Since $\hat{b}_n$ is Lipschitz in y with constant $1/d$,

$$|\hat{Y}_t^n - \hat{Y}_t| \leq \frac{1}{d} \int_t^s |\hat{Y}_r^n - \hat{Y}_r| dr + \int_d^s |\hat{b}_n - \hat{b}|(r, \hat{Y}_r) dr,$$

and Gronwall's inequality gives $\sup_{t \in [d, s]} |\hat{Y}_t^n - \hat{Y}_t| \leq e^{(s-d)/d} \int_d^s |\hat{b}_n - \hat{b}|(r, \hat{Y}_r) dr$. Now

$$|\hat{b}_n - \hat{b}| \leq \frac{1}{2}|\zeta_n - \zeta|(|U_y| + |g_y|) + \frac{1}{2}|U_{n,y} - U_y| + \frac{1}{2}|g_{n,y} - g_y|.$$

All γ_n and γ lie in the common class $\mathcal{U}_r$ of Lemma 6.2(b). By Lemma 2.4(b),(f), in rescaled form, $|U_y| + |g_y| \leq C(1 + |y|)$ and $|U_{n,y} - U_y| \leq C\|\gamma_n - \gamma\|_{L^1}(1 + |y|)$ on $[d, s_0] \times \mathbb{R}$. Since R is Lipschitz ($|R'| = V \leq 1$) and $\ell_n, \ell \geq d$ on $[d, s_0]$, also $|g_{n,y} - g_y| \leq C(1 + |y|) \sup |\ell_n - \ell|$. Here C depends on γ , e , and κ , but not on n . Since $\hat{Y}$ has moments of every order,

$$\mathbb{E} \sup_{t \in [d, s]} |\hat{Y}_t^n - \hat{Y}_t| \leq C(\|\zeta_n - \zeta\|_{L^1} + \|\gamma_n - \gamma\|_{L^1} + \sup |\ell_n - \ell|) \rightarrow 0$$

by Lemma 6.2(b),(c). As V is bounded and continuous and $\ell_n \rightarrow \ell$ uniformly, $\mathbb{E}V(\hat{Y}_t^n/\sqrt{\ell_n(t)}) \rightarrow \mathbb{E}V(\hat{Y}_t/\sqrt{\ell(t)})$ for every t . The weights satisfy $|(1 - \zeta_n)/\ell_n - (1 - \zeta)/\ell| \leq |\zeta_n - \zeta|/d + |\ell_n - \ell|/d^2$, so they converge in $L^1(d, s_0)$, and they are bounded by $1/d$. By dominated convergence the right side of (a) for γ_n converges to the right side for

γ , while the left side $w_n(s, y) = U_n(s, y) - g_n(s, y)$ converges to $w(s, y)$ by Lemma 6.2(c),(d). This proves (a) for γ .

STEP 4. *Part (b).* The drift $\widehat{b}$ is nonnegative because $U_y > 0$ and $R > 0$; this gives the first inequality. For the second, ℓ is absolutely continuous with $\ell' = \zeta$ almost everywhere, so Itô's formula for the product of $\widehat{Y}_{s-\tau}$ with the absolutely continuous function $1/\ell(s-\tau)$ gives

$$\begin{aligned} d\left(\frac{\widehat{Y}}{\ell}\right)(s-\tau) &= \left[\frac{\ell\widehat{b} + \zeta\widehat{Y}}{\ell^2}\right](s-\tau) d\tau + \frac{dB_\tau}{\ell(s-\tau)}, \\ \ell\widehat{b} + \zeta\widehat{Y} &= \frac{1}{2}\zeta(\ell U_y + \widehat{Y}) + \frac{1}{2}\zeta(\ell g_y + \widehat{Y}) = \frac{1}{2}\zeta\Psi + \frac{1}{2}\zeta\sqrt{\ell}\bar{R}(z), \end{aligned}$$

with $z = \widehat{Y}/\sqrt{\ell}$. This is nonnegative because $\Psi > 0$ (Lemma 4.2(a)) and $\bar{R} > 0$ (Lemma 2.1(a)). Integrating over $\tau \in [0, s-t]$ gives the claim. The variable N_t is centered Gaussian with variance $\int_t^s \ell(r)^{-2} dr$.

STEP 5. *Part (c).* By Lemma 4.3,

$$\mathcal{P}_\kappa(\gamma) = \alpha \mathbb{E}U(s_0, Y_0) + \Gamma(s_0) + \frac{q_0}{2e^2\ell_0} + \log e.$$

Insert $U(s_0, \cdot) = g(s_0, \cdot) + w(s_0, \cdot)$. Since Y_0 has the law of $Y^{(q_0)}$ and $\Gamma(s_0) = \tilde{\Gamma} + \frac{1}{2}\log \ell_0$, the terms not involving w equal

$$\alpha \mathbb{E} \log \Phi(Y_0/\sqrt{\ell_0}) + \frac{\alpha}{2} \log \ell_0 - (\alpha-1)\Gamma(s_0) + \frac{q_0}{2e^2\ell_0} + \log e = \Pi(\ell_0; q_0) - (\alpha-1)\tilde{\Gamma}.$$

By (a) with $s = s_0$, the integrand there being nonnegative, and Tonelli's theorem,

$$\alpha \int_{\mathbb{R}} p_0(y) w(s_0, y) dy = \int_d^{s_0} \frac{1-\zeta(t)}{2\ell(t)} (\mathcal{J}(t) + \alpha - 1) dt.$$

Both sides are finite, since $0 \leq w(s_0, \cdot) \leq \tilde{\Gamma}$ by (a) and (6.3). Together with (6.3) for $(\alpha-1)\tilde{\Gamma}$, this gives (6.11). $\square$

The proof of part (a) follows the argument of [76, Lemma 14]: the difference w solves a linear equation with the averaged drift of (6.9), and the Feynman–Kac formula represents w through the source term of that equation. Jagannath and Tobasco use it to compare the Parisi solutions of two order parameters; the history of the resulting Lipschitz bound is recalled after Lemma 2.4. Here g solves the rescaled Parisi equation only up to the defect (6.12), which is proportional to $1-\zeta$ and is the source term, so the identity isolates the contribution of the jammed levels.

The identity (6.11) is exact and holds for every order parameter with $q_0 > 0$; its proof uses nothing specific to α near two. In the original variables the weight $\frac{1-\zeta(t)}{2\ell(t)} dt$ is $\frac{1-\gamma(q)}{2\lambda(q)} dq$ on $[q_0, Q]$, so only the levels where $\gamma < 1$ contribute. Each of them contributes

with the sign of $\mathcal{J}$. Since $0 < 1 - V < 1$, the jamming coefficient takes values in $(1 - \alpha, 1)$, and (6.11) together with (6.3) gives the two-sided bound

$$\Pi(\ell_0; q_0) - (\alpha - 1)\tilde{\Gamma} \leq \mathcal{P}_\kappa(\gamma) \leq \Pi(\ell_0; q_0) + \tilde{\Gamma}.$$

Part (a) says that $U + \Gamma \geq \log \Phi(y/\sqrt{\ell}) + \frac{1}{2} \log \ell$ at every level: the Cole–Hopf steps never lower $U + \Gamma$ below the replica-symmetric solution of the same mass.

Below we use (6.11), together with a Lipschitz bound for Π in q (Lemma 6.6(c)).

Remark 6.4. The first term $\Pi(\ell_0; q_0)$ depends on the mass and the bottom separately. The replica-symmetric order parameter $\mathbf{1}_{[1-\lambda(0), 1]}$ of the same mass has the bottom $1 - e^2 \ell_0 \geq q_0$, and the difference $e^2(s_0 - \ell_0) = e^2 \int_d^{s_0} (1 - \zeta)$ is again carried by the jammed levels. Integrating (6.7) along $\hat{q}(t) := 1 - e^2(\ell_0 + t - \ell(t))$, which decreases from $1 - e^2 \ell_0$ at $t = d$ to q_0 at $t = s_0$ with $\hat{q}' = -e^2(1 - \zeta)$, turns (6.11) into

$$\mathcal{P}_\kappa(\gamma) = \mathcal{P}_\kappa^{\text{RS}}(1 - \lambda(0)) + \int_d^{s_0} \frac{1 - \zeta(t)}{2\ell(t)} \left[\mathcal{J}(t) - \frac{\ell(t)}{\ell_0} \left(1 - \alpha \mathbb{E} V(Y^{(\hat{q}(t))} / \sqrt{\ell_0}) \right) \right] dt. \quad (6.14)$$

Thus the value of every order parameter is the replica-symmetric value of the same mass plus an integral over its jammed levels. This form is not used below.

6.4 The jamming coefficient near $\alpha = 2$

From now on $\alpha = 2 + e$ again, so that $\alpha - 1 = 1 + e$ and $2 \leq \alpha \leq 3$ for $e \leq 1$. We first record the Gaussian input and the properties of Π that we need.

Lemma 6.5. *Let $0 < e \leq 1/4$, $|\kappa| \leq e$, and $q \in [\frac{1}{2}, 1]$. The density p of $Y^{(q)}$ satisfies $p \leq e/\sqrt{\pi}$ on $\mathbb{R}$ and $p \geq c_p e$ on $[-1/(2e), 1/(2e)]$, where $c_p := \phi(\sqrt{2})$. Moreover $\mathbb{P}(Y^{(q)} \leq 0) = \Phi(\kappa/\sqrt{q})$.*

Proof. We have $p(y) = (e/\sqrt{q}) \phi((ey + \kappa)/\sqrt{q})$. Since $q \geq \frac{1}{2}$, $p \leq \sqrt{2} \phi_0 e = e/\sqrt{\pi}$. If $|y| \leq 1/(2e)$, then $|ey + \kappa| \leq \frac{1}{2} + e \leq \frac{3}{4}$, and hence $|ey + \kappa|/\sqrt{q} \leq 3\sqrt{2}/4 < \sqrt{2}$. Since $q \leq 1$, also $e/\sqrt{q} \geq e$, and therefore $p(y) \geq e\phi(\sqrt{2})$. The last identity holds because $Y^{(q)} \leq 0$ exactly when $\sqrt{q} G \leq \kappa$. $\square$

For $\ell > 0$ let $W_{\text{RS}, \ell}$ be the remainder of Lemma 4.7(a), with d replaced by ℓ :

$$W_{\text{RS}, \ell}(y) := \log \Phi(y/\sqrt{\ell}) + \frac{y^2}{2\ell} - \left(\frac{1}{2} \log \ell - \log \sqrt{2\pi} \right) \mathbf{1}_{\{y \leq 0\}} + \mathbf{1}_{\{y \leq -1\}} \log |y|. \quad (6.15)$$

Lemma 6.6. *Let $0 < e \leq 1/4$, $|\kappa| \leq e$, $\ell > 0$, and $q, q' \in [\frac{1}{2}, 1]$.*

$$(a) \quad \Pi(\ell; q) = K_e(q) + c_e(q)/\ell + \frac{1+\rho_q}{2} \log \ell + \alpha \mathbb{E} W_{\text{RS}, \ell}(Y^{(q)}).$$

(b) If $0 < \ell \leq L'$, then $W_{\text{RS},\ell} \geq -\log(1 + L')$ on $(-\infty, 0]$, and $\alpha \mathbb{E} W_{\text{RS},\ell}(Y^{(q)}) \geq -3\log(1 + L') - 3\sqrt{L'}e$.

(c) $|\Pi(\ell; q) - \Pi(\ell; q')| \leq 3(1 + \sqrt{\ell})|q - q'|/(2e\ell)$.

Proof. (a) Write $Y = Y^{(q)}$ and $k = \kappa/\sqrt{q}$. Since $eY_- = (\kappa - \sqrt{q}G)_+$, we have $\mathbb{E}Y_-^2 = qm(k)/e^2$; also $\mathbb{P}(Y \leq 0) = \Phi(k)$. Hence, by (6.15),

$$\begin{aligned} \alpha \mathbb{E} \log \Phi(Y/\sqrt{\ell}) &= \alpha \mathbb{E} W_{\text{RS},\ell}(Y) - \frac{\alpha q m(k)}{2e^2 \ell} \\ &\quad + \rho_q \left(\frac{1}{2} \log \ell - \log \sqrt{2\pi} \right) - \alpha \mathbb{E} [\log |Y|; Y \leq -1]. \end{aligned}$$

Adding $\frac{q}{2e^2 \ell} + \log e + \frac{1}{2} \log \ell$ gives $\Pi(\ell; q)$ on the left. On the right, $\frac{q(1-\alpha m(k))}{2e^2 \ell} = c_e(q)/\ell$ and $\log e - \alpha \mathbb{E} [\log |Y|; Y \leq -1] - \rho_q \log \sqrt{2\pi} = K_e(q)$ by the definitions in Proposition 4.5, which proves (a).

(b) Put $z = y/\sqrt{\ell}$. Since $\log \Phi(z) + z^2/2 + \log \sqrt{2\pi} = -\log R(z)$, for $y \leq 0$ we have $W_{\text{RS},\ell}(y) = -\log R(z) - \frac{1}{2} \log \ell + \mathbf{1}_{\{y \leq -1\}} \log |y|$. We use the Mills-ratio bounds $t < R(-t) < t + 1/t$ for $t > 0$ of Lemma 2.1(b). If $y \leq -1$, then $t := |z| \geq 1/\sqrt{\ell}$ and $W_{\text{RS},\ell}(y) = -\log(R(-t)/t) \geq -\log(1 + t^{-2}) \geq -\log(1 + \ell)$. If $-1 < y \leq 0$, then $R(z) \leq R(-1/\sqrt{\ell}) < \sqrt{\ell} + 1/\sqrt{\ell}$ because R is decreasing, so $W_{\text{RS},\ell}(y) \geq -\log(1 + \ell)$. For $y > 0$, $W_{\text{RS},\ell}(y) = \log \Phi(z) \geq -2\Phi(-z)$ by Lemma 2.1(c), and $\int_0^\infty 2\Phi(-y/\sqrt{\ell}) dy = 2\phi_0\sqrt{\ell}$. With $\alpha \leq 3$ and the density bound of Lemma 6.5,

$$\alpha \mathbb{E} W_{\text{RS},\ell}(Y^{(q)}) \geq -3\log(1 + L') - \frac{3e}{\sqrt{\pi}} 2\phi_0\sqrt{L'},$$

and $6\phi_0/\sqrt{\pi} < 3$.

(c) By (6.7), it suffices to show $|1 - \alpha \mathbb{E} V(Y^{(q)}/\sqrt{\ell})| \leq 3e(1 + \sqrt{\ell})$. By Lemma 6.5 and (6.5), $|1 - \alpha \mathbb{P}(Y^{(q)} < 0)| = |1 - \rho_q| \leq 3e$. By the density bound of Lemma 6.5 and $\alpha \leq 3$,

$$\alpha |\mathbb{E} [V(Y^{(q)}/\sqrt{\ell}) - \mathbf{1}_{\{Y^{(q)} < 0\}}]| \leq \frac{3e}{\sqrt{\pi}} \sqrt{\ell} \int_{\mathbb{R}} |V(z) - \mathbf{1}_{\{z < 0\}}| dz = \frac{3e}{\sqrt{\pi}} \sqrt{\ell} \cdot 2R(0) \leq 3e\sqrt{\ell}.$$

Here $\int_{-\infty}^0 (1 - V) = \bar{R}(0) = R(0)$ and $\int_0^\infty V = R(0) = 2\phi_0$, because $\bar{R}' = 1 - V$, $R' = -V$, and $\bar{R}(-\infty) = R(\infty) = 0$ (Lemma 2.1(a),(b)); and $12\phi_0/\sqrt{\pi} < 3$. $\square$

In (6.11) every jammed level contributes with the sign of $\mathcal{J}$. The next lemma shows that $\mathcal{J}$ is at least $-Ce$ at every level and is large and positive near the top. The source of the gain is Lemma 6.3(b): the process $\hat{Y}/\ell$ has nonnegative drift, so the diffusion started at $y \in [-\ell_0/\sqrt{\ell(t)}, 0]$ reaches level t within $\sqrt{\ell(t)}$ of the wall, or beyond it, with probability at least $\frac{1}{2}$. The Gaussian input puts mass of order $e\ell_0/\sqrt{\ell(t)}$ on this interval.

Lemma 6.7. *Let $S \geq 1$, let $0 < e \leq 1/4$ with $e^2 S \leq \frac{1}{2}$, and let $|\kappa| \leq e$. Put*

$$c_{\mathcal{J}} := c_p(1 - V(-1)) > 0, \quad C_S := 3 + 3\sqrt{S}.$$

For every $\gamma \in \mathcal{U}$ with $q_0 > 0$ and $s_0 \leq S$, and every $t \in [d, s_0]$,

$$\mathcal{J}(t) \geq c_{\mathcal{J}} e \min\left\{\frac{\ell_0}{\sqrt{\ell(t)}}, \frac{1}{2e}\right\} - C_S e.$$

Proof. We have $q_0 = 1 - e^2 s_0 \in [\frac{1}{2}, 1]$, so Lemma 6.5 applies to p_0 . Write $\hat{Y}^y := \hat{Y}^{s_0, y}$, $\ell = \ell(t)$, and $A := \ell_0/\sqrt{\ell}$. Since $1 - V > 0$, we may restrict the integral in (6.10) to $\{y > 0\} \cup [-A, 0]$. We treat the two regions separately and then combine them.

STEP 1. *The region $y > 0$.* By Lemma 6.3(b) and since V is decreasing, $V(\hat{Y}_t^y/\sqrt{\ell}) \leq V((y + B_{s_0-t})/\sqrt{\ell})$ almost surely, and B_{s_0-t} has the law of σG with $\sigma^2 = s_0 - t \leq S$. For fixed G , since $0 < V < 1$,

$$\int_0^\infty V\left(\frac{y + \sigma G}{\sqrt{\ell}}\right) dy \leq \sigma G_- + \int_0^\infty V(u/\sqrt{\ell}) du = \sigma G_- + 2\phi_0\sqrt{\ell}.$$

Moreover $\mathbb{E}G_- = \phi_0$ and $\ell \leq s_0 \leq S$. With $p_0 \leq e/\sqrt{\pi}$ and $\alpha \leq 3$,

$$\begin{aligned} \alpha \int_0^\infty p_0(y) \mathbb{E}[1 - V(\hat{Y}_t^y/\sqrt{\ell})] dy &\geq \alpha \mathbb{P}(Y_0 > 0) - \frac{3e}{\sqrt{\pi}} 3\phi_0\sqrt{S} \\ &\geq \alpha(1 - \Phi(\kappa/\sqrt{q_0})) - 3\sqrt{S}e, \end{aligned}$$

because $9\phi_0/\sqrt{\pi} < 3$.

STEP 2. *The region $-A \leq y \leq 0$.* By Lemma 6.3(b) with $s = s_0$, $\hat{Y}_t^y \geq \ell y/\ell_0 + \ell N_t \geq -\sqrt{\ell} + \ell N_t$. The centered Gaussian variable N_t satisfies $\mathbb{P}(N_t \geq 0) \geq \frac{1}{2}$, with equality unless its variance vanishes, which happens at $t = s_0$, where $\mathbb{P}(N_t \geq 0) = 1$. On the event $\{N_t \geq 0\}$ we get $\hat{Y}_t^y/\sqrt{\ell} \geq -1$, and hence $1 - V(\hat{Y}_t^y/\sqrt{\ell}) \geq 1 - V(-1)$ because V is decreasing. By Lemma 6.5, $\mathbb{P}(-A \leq Y_0 \leq 0) \geq c_p e \min\{A, 1/(2e)\}$. Since $\alpha/2 \geq 1$,

$$\alpha \int_{-A}^0 p_0(y) \mathbb{E}[1 - V(\hat{Y}_t^y/\sqrt{\ell})] dy \geq c_{\mathcal{J}} e \min\{A, 1/(2e)\}.$$

STEP 3. Conclusion. Since $\alpha - (\alpha - 1) = 1$, we have $\alpha(1 - \Phi(\kappa/\sqrt{q_0})) - (\alpha - 1) = 1 - \rho \geq -3e$ by (6.5). Adding the two regions and subtracting $\alpha - 1$ gives the claim. $\square$

Corollary 6.8. *Let $S \geq 1$, and let $c_{\mathcal{J}}, C_S$ be as in Lemma 6.7. Put $e_{\mathcal{J}} = e_{\mathcal{J}}(S, \Theta) := \min\{e_1, \frac{1}{4}, (2S)^{-1/2}, c_{\mathcal{J}}/(4C_S)\}$. For $0 < e < e_{\mathcal{J}}$ and $\theta \in \Theta$, every $\gamma \in \mathcal{U}$ with $q_0 > 0$ and $s_0 \leq S$ satisfies*

$$\mathcal{P}_{\kappa}(\gamma) \geq \Pi(\ell_0; q_0) + \frac{c_{\mathcal{J}}e}{4} \int_d^{t_1} \frac{1 - \zeta(t)}{\ell(t)} \min\left\{\frac{\ell_0}{\sqrt{\ell(t)}}, \frac{1}{2e}\right\} dt - C_S e \frac{s_0}{2\ell_0} \log \frac{s_0}{t_1}, \quad (6.16)$$

where $t_1 := \min\{s_0, (c_{\mathcal{J}}\ell_0/(2C_S))^2\}$, and the integral is zero if $t_1 \leq d$.

Proof. The choice of e_j gives $|\kappa| \leq e \leq \frac{1}{4}$, $e^2 S \leq \frac{1}{2}$, and $c_{\mathcal{J}}/(2e) \geq 2C_S$, so Lemma 6.7 applies. We use (6.11). For $t \in [d, t_1]$ we have $\ell(t) \leq t \leq t_1$, so $c_{\mathcal{J}}\ell_0/\sqrt{\ell(t)} \geq 2C_S$. Together with $c_{\mathcal{J}}/(2e) \geq 2C_S$, Lemma 6.7 gives $\mathcal{J}(t) \geq \frac{1}{2}c_{\mathcal{J}}e \min\{\ell_0/\sqrt{\ell(t)}, 1/(2e)\}$ there. For $t \in [\max\{d, t_1\}, s_0]$ we use $\mathcal{J}(t) \geq -C_S e$, $1 - \zeta \leq 1$, and $\ell(t) \geq \ell_0 t/s_0$ from (6.2); the integral of $s_0/(2\ell_0 t)$ over $[\max\{d, t_1\}, s_0]$ is at most $\frac{s_0}{2\ell_0} \log \frac{s_0}{t_1}$. $\square$

6.5 The mass lower bound

We now show that the rescaled mass of a global minimizer of bounded depth cannot be small. The mechanism is visible in Lemma 6.6(a): the coefficient $c_e(q_0) \approx 2\phi_0\theta$ of $1/\ell_0$ is positive, so $c_e(q_0)/\ell_0$ dominates $\log \ell_0$ as $\ell_0 \downarrow 0$. By (6.16), the jammed levels lower $\mathcal{P}_{\kappa}(\gamma)$ below $\Pi(\ell_0; q_0)$ by at most $O(e \log(1/e)/\ell_0)$.

Lemma 6.9. *Let $S \geq 1$ and $L \geq 1$. There are $C_A, C'_A < \infty$ and $e_A > 0$, depending only on S, L, c_D , and Θ , with the following property. Let $0 < e < e_A$ and $\theta \in \Theta$, and let $\gamma \in \mathcal{U}$ with $q_0 > 0$ have rescaled data satisfying*

$$s_0 \leq S, \quad \ell_0 \leq L, \quad d \geq c_D e^2.$$

Then

$$\mathcal{P}_{\kappa}(\gamma) \geq K_e + \frac{2\phi_0\theta - C_A e \log(1/e)}{\ell_0} + \frac{1+\rho}{2} \log \ell_0 - C'_A. \quad (6.17)$$

Proof. Let $e_A := \min\{e_j(S, \Theta), e_K(S), 1/e\}$, so that (6.16) and (6.4) apply and $\log(1/e) \geq 1$. Drop the nonnegative middle term in (6.16). If $t_1 < s_0$, then $t_1 = (c_{\mathcal{J}}\ell_0/(2C_S))^2$, and since $\ell_0 \geq d \geq c_D e^2$ and $s_0 \leq S$,

$$\log \frac{s_0}{t_1} \leq \log \frac{4C_S^2 S}{c_{\mathcal{J}}^2 c_D^2 e^4} \leq C \log(1/e);$$

if $t_1 = s_0$, the logarithm vanishes. Hence $\mathcal{P}_{\kappa}(\gamma) \geq \Pi(\ell_0; q_0) - C e \log(1/e)/\ell_0$. Now $q_0 = 1 - e^2 s_0 \in [1 - e^2 S, 1] \subset [\frac{1}{2}, 1]$, so Lemma 6.6(a),(b) with $L' = L$ and (6.4) give

$$\Pi(\ell_0; q_0) \geq K_e - C'_S e^2 \log \frac{1}{e} + \frac{2\phi_0\theta - C'_S e}{\ell_0} + \frac{1+\rho}{2} \log \ell_0 - 3 \log(1+L) - 3\sqrt{L}e.$$

Since $e \leq e \log(1/e)$ and $e^2 \log(1/e) \leq 1$, this gives (6.17) with $C_A := C'_S + C$ and $C'_A := C'_S + 3 \log(1+L) + 3\sqrt{L}$. $\square$

Proposition 6.10. *Let $S \geq 1$. There are $\ell_{\min} = \ell_{\min}(S, \Theta) \in (0, 1]$ and $e_M > 0$ such that for $0 < e < e_M$ and $\theta \in \Theta$, every global minimizer γ of $\mathcal{P}_{\kappa_e(\theta)}$ with $s_0 \leq S$ satisfies $\ell_0 \geq \ell_{\min}$.*

Proof. Let $e_M \leq \min\{e_1, e_A\}$ be so small that $C_A e \log(1/e) \leq \phi_0\theta_-$ for $e < e_M$, where C_A, C'_A, e_A are given by Lemma 6.9 with the constant S and with L the mass bound of

Theorem 4.11. Since $q_0 > 0$ (Proposition 2.11(c)), (I1) and (I2) show that γ satisfies the hypotheses of Lemma 6.9. Combining (6.17) with (I4),

$$\frac{\phi_0\theta_-}{\ell_0} + \frac{1+\rho}{2} \log \ell_0 \leq C'_{\text{RS}} + C'_A =: C''_A.$$

Suppose $\ell_0 < 1$. Since $0 < (1+\rho)/2 \leq 2$, the second term is at least $2 \log \ell_0$. Let $x_0 \in (0,1)$ be such that $2x \log(1/x) \leq \phi_0\theta_-/2$ for $0 < x \leq x_0$. If $\ell_0 \leq x_0$, then $2 \log \ell_0 \geq -\phi_0\theta_-/(2\ell_0)$, hence $\phi_0\theta_-/(2\ell_0) \leq C''_A$. Therefore

$$\ell_0 \geq \ell_{\min} := \min\{x_0, \phi_0\theta_-/(2 \max\{C''_A, 1\}), 1\}.$$

The constant $\ell_{\min}$ depends only on S and Θ , since L and c_D depend only on Θ . □

We apply Proposition 6.10 with $S = S_*$ from Theorem 5.1, which fixes $\ell_{\min}$.

6.6 The top-jump bound

We next bound the top jump of every global minimizer. The argument works in the original variables (q, x) and holds for every $\alpha > 0$; only the smallness of κ and of the mass $\lambda(0)$ enters. For $\gamma \in \mathcal{U}$ let X be its Parisi diffusion. We use the scalar functions $\mathcal{N}$ and $\tilde{B}$ of (2.22) and the *breakpoint ratio* $\tilde{x}$ of (2.23), with a subscript γ when the dependence on the order parameter matters:

$$\tilde{B}_\gamma(q) = \mathbb{E}[v^2(1-v)](q, X_q) > 0, \quad \mathcal{N}_\gamma(q) = \mathbb{E}[v_x(q, X_q)^2], \quad \tilde{x}_\gamma(q) = \frac{\lambda(q)\mathcal{N}_\gamma(q)}{2\tilde{B}_\gamma(q)}. \quad (6.18)$$

The breakpoint ratio governs the derivative of $S_\gamma(q) := \alpha \mathbb{E}v(q, X_q)^2$, the function S of (2.22), for which $D' = (S_\gamma - 1)/\lambda^2$. By Lemma 2.9(c)(iv), S_γ has one-sided derivatives at every $q \in [0, 1)$,

$$S'_\gamma(q+) = \frac{2\alpha\tilde{B}_\gamma(q)}{\lambda(q)}(\tilde{x}_\gamma(q) - \gamma(q)), \quad S'_\gamma(q-) = \frac{2\alpha\tilde{B}_\gamma(q)}{\lambda(q)}(\tilde{x}_\gamma(q) - \gamma(q-)), \quad (6.19)$$

and the functions $\tilde{B}_\gamma$ and $\tilde{x}_\gamma$ are continuous (Lemma 2.9(c)). These are the analogues of the one-sided derivatives in [8, Proposition 3], which also differ only through the value of the distribution function at q or at $q-$. In the physics literature, $\tilde{x}$ is the formula for the order parameter on an interval of the support in the full replica symmetry breaking solution of the perceptron [59, (36)].

The argument of this section follows the strategy of Auffinger and W.-K. Chen in the proof of [8, Theorem 4]. There, continuity of a Parisi measure at the top of its support and these one-sided derivatives identify the mass below the top with a ratio of two

expectations, and Jensen's inequality shows that this ratio is not one under a condition on the mixture. The same strategy is used for the Sherrington–Kirkpatrick model at every $\beta > 1$ in [87, Proposition 9.3], where the log-concavity of a transformed density and a covariance inequality exclude the corresponding relation. The argument here differs from these works as follows. The functional has a hard wall: at the top of the support the solution is $\log \Phi((x - \kappa)/\sqrt{\delta})$, in place of $\log \cosh x$ up to a constant. When the top is isolated in the support, $\gamma(Q-) < 1$ is automatic but a bound uniform in the minimizer is not; in this case the gap principle (Lemma 2.12) produces a level x^* below the top with $\tilde{x}_\gamma(x^*) = \gamma(Q-)$. The conclusion is a bound $\gamma(Q-) \leq m_*$, not only an atom at the top, and m_* is uniform: Proposition 6.12 holds for all order parameters, and m_* does not depend on e , θ , or the minimizer. Finally, the monotone weight comes from the tilted-density representation of Lemma 2.7, not from log-concavity; Auffinger and W.-K. Chen express the corresponding averages through a Brownian motion and an exponential weight [8, (18)]. As in [87, Proposition 9.3], the ratio is then bounded by a covariance inequality, here Chebyshev's association inequality combined with the monotonicity of $(V')^2/(V^2(1 - V))$ (Lemma B.2); Auffinger and W.-K. Chen use Jensen's inequality at this point.

We first translate the first-order conditions at the top into a statement about $\tilde{x}_\gamma$.

Lemma 6.11. *Let γ be a global minimizer of $\mathcal{P}_\kappa$ with top Q and $\mu_\gamma \neq \delta_Q$.*

- (a) *If Q is a limit of points of $\text{supp } \mu_\gamma \cap [0, Q)$, then $\gamma(Q-) = \tilde{x}_\gamma(Q)$.*
- (b) *Otherwise let $p := \max(\text{supp } \mu_\gamma \cap [0, Q))$. Then $\gamma \equiv m := \gamma(Q-)$ on $[p, Q)$, $0 < m < 1$, and there is $x^* \in (p, Q)$ with $\tilde{x}_\gamma(x^*) = m$.*

Proof. (a) Let $t_n \in \text{supp } \mu_\gamma$ with $t_n \uparrow Q$. By Proposition 2.11(c), $D(t_n) = D(Q) = 0$, so Rolle's theorem gives $s_n \in (t_n, Q)$ with $D'(s_n) = 0$, that is, $S_\gamma(s_n) = 1$; by continuity $S_\gamma(Q) = 1$. Hence $(S_\gamma(Q) - S_\gamma(s_n))/(Q - s_n) = 0$ for all n , and since the left derivative $S'_\gamma(Q-)$ exists, it vanishes. By (6.19) and $\tilde{B}_\gamma(Q) > 0$, $\gamma(Q-) = \tilde{x}_\gamma(Q)$. (b) Since Q is isolated from the left in $\text{supp } \mu_\gamma$ and $\mu_\gamma \neq \delta_Q$, the point p exists and $(p, Q) \cap \text{supp } \mu_\gamma = \emptyset$. The gap principle, Lemma 2.12(a) with $p' = Q$, gives $\gamma \equiv m = \gamma(p)$ on $[p, Q)$ with $0 < m < 1$, and a point $x^* \in (p, Q)$ with $\tilde{x}_\gamma(x^*) = m$. As γ is constant on $[p, Q)$, $m = \gamma(Q-)$. $\square$

The first-order conditions thus force $\gamma(Q-)$ to equal the value of $\tilde{x}_\gamma$ at Q or at a point of the last gap. The top-jump bound follows once $\tilde{x}_\gamma$ is bounded away from one there. This is the content of the next proposition, which is a statement about *arbitrary* order parameters and is proved in Chapter B. Let

$$x_\Phi := \frac{\int_{\mathbb{R}} \Phi(V')^2 dz}{2 \int_{\mathbb{R}} \Phi V^2(1 - V) dz}. \quad (6.20)$$

By Lemma B.3, $x_\Phi < 1$.

Proposition 6.12. *For every $\varepsilon > 0$ there is $\bar{\eta} \in (0, 1/4]$ such that the following holds for every $\kappa \in [-\bar{\eta}, \bar{\eta}]$ and every $\gamma \in \mathcal{U}$ with top Q .*

- (a) *If $\delta = 1 - Q \leq \bar{\eta}$, then $\tilde{x}_\gamma(Q) \leq x_\Phi + \varepsilon$.*
- (b) *If $\gamma \equiv m$ on $[p, Q)$ for some $p < Q$, with $1 - \bar{\eta} \leq m < 1$ and $\lambda(p) \leq \bar{\eta}$, then $\tilde{x}_\gamma(q) \leq x_\Phi + \varepsilon$ for every $q \in [p, Q)$.*

The proof, in §B.4, has three ingredients. First, on a last constant piece of mass m the functions v and v_x depend on (q, x) only through the scale-free variable $z = (x - \kappa) / \sqrt{\lambda(q)}$ and the ratio $\delta / \lambda(q) \in (0, 1]$. We prove bounds on them that are uniform in this ratio, including the tails $z \rightarrow \pm\infty$ (Proposition B.5). The left-tail bounds follow by completing the square in a one-dimensional Gaussian convolution and differentiating its rescaled tail integral. Second, by the tilted-density representation of Lemma 2.7, the law of z at level q is a fixed weight multiplied by a nonincreasing function. Third, the ratio $(V')^2 / (V^2(1 - V))$ is nondecreasing (Lemma B.2), so Chebyshev's association inequality (Lemma B.6) bounds the weighted average defining $\tilde{x}_\gamma$ by its unweighted Gaussian value x_Φ , up to errors that vanish as $(\kappa, \lambda(p), 1 - m) \rightarrow 0$.

Theorem 6.13. *There are a numerical constant $m_* \in (0, 1)$ and $e_T > 0$ depending only on Θ such that for $0 < e < e_T$ and $\theta \in \Theta$, every global minimizer γ of $\mathcal{P}_{\kappa_e(\theta)}$ satisfies $\gamma(Q-) \leq m_*$.*

Proof. Apply Proposition 6.12 with $\varepsilon := (1 - x_\Phi)/4$, which is positive by Lemma B.3, obtaining $\bar{\eta}$, and put

$$\eta_0 := \min\{\bar{\eta}, (1 - x_\Phi)/2\}, \quad m_* := 1 - \eta_0.$$

Then $x_\Phi + \varepsilon < x_\Phi + (1 - x_\Phi)/2 \leq m_*$. Let $e_T \leq e_1$ be such that $e_T \leq \bar{\eta}$ and $e_T^2 L \leq \bar{\eta}$. Let $0 < e < e_T$ and $\theta \in \Theta$, and let γ be a global minimizer. By (I1) and the bound $|\kappa| \leq e$ of §6.1, $|\kappa| \leq e \leq \bar{\eta}$ and, since λ is nonincreasing,

$$\lambda(q) \leq \lambda(0) = e^2 \ell_0 \leq e^2 L \leq \bar{\eta} \quad (q \in [0, 1]).$$

If $\mu_\gamma = \delta_Q$, then $\gamma(Q-) = 0$. Otherwise we use Lemma 6.11. In case (a), $\delta = \lambda(Q) \leq \bar{\eta}$, so Proposition 6.12(a) gives $\gamma(Q-) = \tilde{x}_\gamma(Q) \leq x_\Phi + \varepsilon < m_*$. In case (b), suppose that $m = \gamma(Q-) > m_*$. Then $1 - \bar{\eta} \leq 1 - \eta_0 < m < 1$ and $\lambda(p) \leq \bar{\eta}$, so Proposition 6.12(b) at $q = x^*$ gives $m = \tilde{x}_\gamma(x^*) \leq x_\Phi + \varepsilon < m_*$. This contradicts $m > m_*$. Hence $\gamma(Q-) \leq m_*$. $\square$

The theorem is global: it assumes neither closeness of γ to a replica-symmetric order parameter nor any bound on d . The smallness needed in Proposition 6.12 comes only from $\lambda(p) \leq \lambda(0) = e^2 \ell_0 \leq e^2 L$ and $|\kappa| \leq e$. The modulus in Proposition 6.12 is small only when κ , $\lambda(p)$, and $1 - m$ are all small, which is why m_* is chosen through $\bar{\eta}$ and is not explicit. The bound uses stationarity and fails for arbitrary order parameters: for $\gamma = m\mathbf{1}_{[p, Q)} + \mathbf{1}_{[Q, 1]}$ the jump $1 - \gamma(Q-) = 1 - m$ can be arbitrarily small.

6.7 The top-gap bound

If the top gap d is small and $\gamma(Q-) \leq m_* < 1$, then just below the wall the profile is jammed: on $(d, 2d]$ one has $1 - \zeta \geq 1 - m_*$ and $\ell \leq 2d$. By Lemma 6.7 the jamming coefficient there is of order $e \min\{\ell_0/\sqrt{d}, 1/e\}$, which dominates every other term of (6.16) once d is small. The competitor is the replica-symmetric order parameter of the same mass.

Proposition 6.14. *Let $S \geq 1$, $\ell_{\min} \in (0, 1]$ and $m_* \in (0, 1)$. There are $d_1 > 0$ and $e_E > 0$, depending only on S , $\ell_{\min}$, m_* , and Θ , such that the following holds for $0 < e < e_E$ and $\theta \in \Theta$. Let $\gamma \in \mathcal{U}$ with $q_0 > 0$ have rescaled data satisfying*

$$s_0 \leq S, \quad \ell_0 \geq \ell_{\min}, \quad \gamma(Q-) = \zeta(d+) \leq m_*, \quad d \leq d_1,$$

and let $\gamma^\natural := \mathbf{1}_{[1-e^2\ell_0, 1]}$ be the replica-symmetric order parameter of the same mass. Then $\mathcal{P}_\kappa(\gamma) > \mathcal{P}_\kappa(\gamma^\natural)$.

Proof. Let $c_{\mathcal{J}}, C_S$ be as in Lemma 6.7 and put

$$\begin{aligned} t_{\min} &:= \min\{\ell_{\min}, (c_{\mathcal{J}}\ell_{\min}/(2C_S))^2\}, & C_{\sharp} &:= \frac{3(1+\sqrt{S})S}{2\ell_{\min}} + \frac{C_S S}{2\ell_{\min}} \log \frac{S}{t_{\min}}, \\ \Lambda_{\sharp} &:= \frac{8C_{\sharp}}{c_{\mathcal{J}}(1-m_*)\log 2}, & d_1 &:= \min\left\{\frac{t_{\min}}{2}, \frac{\ell_{\min}^2}{2\Lambda_{\sharp}^2}\right\}, & e_E &:= \min\left\{e_{\mathcal{J}}(S, \Theta), \frac{1}{2\Lambda_{\sharp}}\right\}. \end{aligned}$$

Let $0 < e < e_E$ and let γ be as in the statement. Since $e^2\ell_0 \leq e^2S \leq \frac{1}{2}$, we have $\gamma^\natural \in \mathcal{U}$ and $\mathcal{P}_\kappa(\gamma^\natural) = \Pi(\ell_0; 1 - e^2\ell_0)$ by Lemma 4.7(a). Both $q_0 = 1 - e^2s_0$ and $1 - e^2\ell_0$ lie in $[1 - e^2S, 1] \subset [\frac{1}{2}, 1]$, and they differ by $e^2(s_0 - \ell_0) \leq e^2S$. Hence Lemma 6.6(c) and $\ell_{\min} \leq \ell_0 \leq S$ give

$$\Pi(\ell_0; q_0) \geq \mathcal{P}_\kappa(\gamma^\natural) - \frac{3(1+\sqrt{S})S}{2\ell_{\min}} e.$$

In (6.16), $\ell_0 \geq \ell_{\min}$ and $s_0 \geq \ell_0 \geq \ell_{\min}$ give $t_1 \geq t_{\min}$, so the last term is at least $-e C_S S \log(S/t_{\min})/(2\ell_{\min})$. Moreover $2d \leq t_{\min} \leq t_1$. On $(d, 2d]$ we have $\zeta \leq \zeta(d+) \leq m_*$ and $\ell(t) \leq t \leq 2d$, so

$$\int_d^{t_1} \frac{1-\zeta}{\ell} \min\left\{\frac{\ell_0}{\sqrt{\ell}}, \frac{1}{2e}\right\} dt \geq (1-m_*) \log 2 \min\left\{\frac{\ell_{\min}}{\sqrt{2d}}, \frac{1}{2e}\right\} \geq (1-m_*) \log 2 \Lambda_{\sharp},$$

by the choices of d_1 and e_E . Inserting these bounds into (6.16),

$$\mathcal{P}_\kappa(\gamma) - \mathcal{P}_\kappa(\gamma^\natural) \geq e \left[\frac{c_{\mathcal{J}}}{4} (1-m_*) \log 2 \Lambda_{\sharp} - C_{\sharp} \right] = e C_{\sharp} > 0.$$

This proves the proposition. $\square$

The proposition is an inequality for all order parameters, and it is combined with minimality only in the proof of Theorem 6.1. It needs no lower bound on d in terms of e : when $d \lesssim e^2$ the minimum in the gain is $1/(2e)$, and the gain is of order one. The bounds $\ell_0 \geq \ell_{\min}$ and $s_0 \leq S$ enter only through $t_1 \geq t_{\min}$, the Lipschitz bound of Lemma 6.6(c) and the factor $\ell_0/\sqrt{\ell}$ in the gain.

6.8 Proof that minimizers lie in a compact class

Proof of Theorem 6.1. We fix the constants in the order (6.1). The window Θ fixes D_{top} , L and C_0 (Theorem 4.11), c_D (Theorem 4.10), C'_{RS} (Lemma 4.7(b)), and then S_* (Theorem 5.1). Let $\ell_{\min} := \ell_{\min}(S_*, \Theta)$ and e_M be given by Proposition 6.10 with $S = S_*$, and let m_* and e_T be given by Theorem 6.13. Let d_1 and e_E be given by Proposition 6.14 with $(S, \ell_{\min}, m_*) = (S_*, \ell_{\min}, m_*)$. Put $e_* := \min\{e_L, e_M, e_T, e_E\}$.

Let $0 < e < e_*$ and $\theta \in \Theta$, and let γ be a global minimizer of $\mathcal{P}_{\kappa_e(\theta)}$. Then $q_0 > 0$ by Proposition 2.11(c). By Theorem 5.1, $s_0 \leq S_*$. By Proposition 6.10, $\ell_0 \geq \ell_{\min}$. By Theorem 6.13, $\gamma(Q-) \leq m_*$. Suppose that $d \leq d_1$. Then γ satisfies all hypotheses of Proposition 6.14, so $\mathcal{P}_{\kappa}(\gamma^\natural) < \mathcal{P}_{\kappa}(\gamma) = \mathcal{P}_*(\kappa)$, which contradicts the definition of $\mathcal{P}_*(\kappa)$. Hence $d > d_1$, and the rescaled profile of γ lies in $\mathcal{K}(d_1, S_*)$. Each constant depends only on the constants preceding it in (6.1), and none depends on e , θ , or the minimizer. This proves the theorem. $\square$

The flat limit

Chapters 5 and 6 show that, for $\alpha = 2 + e$ with e small and $\theta \in \Theta$, the rescaled profile of every global minimizer lies in a fixed compact class $\mathcal{K}(d_1, S_*)$ (Theorem 6.1). This chapter converts that compactness into closeness to the replica-symmetric crossing order parameter $\gamma_d = \mathbf{1}_{[q_d, 1]}$. This is Step 4 of §1.3.

§7.1 defines the flat functional $F_t = A + tI$ (Definition 7.1). In §7.2 we expand $\mathcal{P}_\kappa(\gamma)$ to order e , uniformly on compact classes (Theorem 7.3). §§7.3–7.5 pass to the inverse profile $T = \ell^{-1}$ and prove that I is convex along every segment from the replica-symmetric profile to a normalized step profile. §7.6 computes the first variation of F_{t_d} at $\mathbf{1}_{[0, 1]}$, and §7.7 combines the two into the cubic modulus of Theorem 7.13, for every $t \geq t_d = 1/(4c_0)$. In §7.8 the expansion and the modulus give a stability estimate (Proposition 7.14), uniform in θ near θ_d , from which Chapter 8 starts.

Throughout this chapter, $\alpha = 2 + e$, Θ is a compact window as in Remark 3.12, with endpoints $\theta_- \leq \theta_+$, and $\kappa = \kappa_e(\theta) = \kappa_{\text{RS}}(2 + e) - e^2\theta$ with $\theta \in \Theta$. We use the rescaled variables of Definition 4.1, and we write $L_e = \log(1/e)$. Constants c, C depend only on Θ and on the parameters named in each statement. The longer proofs are in Chapter C.

7.1 Compact classes and the flat functional

A *profile* is a nonincreasing function $\zeta: [0, \infty) \rightarrow [0, 1]$. Profiles that agree almost everywhere are identified, and we use the distance of $L^1(0, \infty)$. The class $\mathcal{K}(d_1, S)$, $0 < d_1 \leq S$, defined before Theorem 6.1, consists of the profiles that equal one almost everywhere on $[0, d_1]$ and vanish almost everywhere on (S, ∞) ; rescaled profiles ζ_γ are extended by zero to (e^{-2}, ∞) . For a profile ζ we write d for the largest s with $\zeta = 1$ almost everywhere on $[0, s]$ and s_0 for the smallest s with $\zeta = 0$ almost everywhere on (s, ∞) . For $\zeta = \zeta_\gamma$ these are the top gap and the depth of Definition 4.1, and $\zeta \in \mathcal{K}(d_1, S)$ means $d \geq d_1$ and $s_0 \leq S$. The class $\mathcal{K}(d_1, S)$ is compact in L^1 . For $\zeta \in \mathcal{K}(d_1, S)$ we write $\ell(s) = \int_0^s \zeta$ and call $\ell_0 = \int_0^\infty \zeta$ its *mass*. Then

$$\min\{s, d_1\} \leq \ell(s) \leq \min\{s, \ell_0\}, \quad d_1 \leq \ell_0 \leq S. \quad (7.1)$$

The *normalized class* is $\mathcal{K}^1(d_1, S) = \{\zeta \in \mathcal{K}(d_1, S) : \ell_0 = 1\}$. It is closed in $\mathcal{K}(d_1, S)$, hence compact. The *normalization* of $\zeta \in \mathcal{K}(d_1, S)$ is $\tilde{\zeta}(s) = \zeta(\ell_0 s)$, and $\tilde{\zeta} \in \mathcal{K}^1(d_1/\ell_0, S/\ell_0)$. A *step profile* is a profile that takes finitely many values.

Step profiles are dense in every class, in a form that preserves the class and the mass. This is the cell-average construction of Lemma 6.2, which involves only the profile. Let $\zeta \in \mathcal{K}(d_1, S)$, fix $e > 0$ with $e^2 S < 1$, and let $\gamma \in \mathcal{U}$ be the right-continuous function with $\gamma(q) = \zeta((1-q)/e^2)$ for $q \geq 1 - e^2 S$ and $\gamma = 0$ below. Then $\zeta_\gamma = \zeta$, the bottom of the support of μ_γ is $1 - e^2 s_0 > 0$, and γ has rescaled top gap d and depth s_0 . Apply Lemma 6.2(a),(b) with the finite set $\{d, s_0\}$. The rescaled profiles ζ_n of the step profiles γ_n do not depend on e ; they are nonincreasing step functions, equal to one on $[0, d]$ and to zero on (s_0, ∞) , and they preserve the mass. Thus

$$\zeta_n \in \mathcal{K}(d, s_0) \subset \mathcal{K}(d_1, S), \quad \int_0^\infty \zeta_n = \int_0^\infty \zeta, \quad \zeta_n \rightarrow \zeta \text{ in } L^1. \quad (7.2)$$

In particular $\zeta_n \in \mathcal{K}^1(d_1, S)$ if $\zeta \in \mathcal{K}^1(d_1, S)$. We call (ζ_n) the *step approximations* of ζ .

For $\zeta \in \mathcal{K}(d_1, S)$, let $U = U_\zeta$ be the solution on $(0, S] \times \mathbb{R}$ of

$$U_s = \frac{1}{2}(U_{yy} + \zeta U_y^2) \quad (d_1 < s \leq S), \quad U(s, y) = \log \Phi(y/\sqrt{s}) \quad (0 < s \leq d_1). \quad (7.3)$$

By the dilation of Lemma C.2, (7.3) is the finite-cutoff Parisi equation of an order parameter in $\mathcal{U}_r$, $r = 1 - d_1/S$, at margin zero, written in rescaled variables. This defines U_ζ for every $\zeta \in \mathcal{K}(d_1, S)$ (by the Cole–Hopf recursion for step profiles, and as a limit in general), and Lemmas 2.4, 2.9, and 2.10 apply to it with constants that depend only on (d_1, S) . If $\zeta = \zeta_\gamma$ for some $\gamma \in \mathcal{U}$ and $e^2 S < 1$, then $U_\zeta(s, y) = u_\gamma(1 - e^2 s, \kappa + ey)$ on $[d_1, S] \times \mathbb{R}$; so U_ζ is the rescaled solution U of Definition 4.1, and it depends neither on e nor on κ . As in Definition 4.1 we write

$$v = -\ell U_{yy} \in (0, 1), \quad \Psi = \ell U_y + y > 0.$$

The flat functional renormalizes $U(S, \cdot)$ by its behavior as $y \rightarrow -\infty$. By Lemma 7.2 below this behavior is $-y^2/(2\ell_0) - \log|y| + \omega(S)$, where, for $s \geq d_1$,

$$\Gamma(s) = \frac{1}{2} \log d_1 + \frac{1}{2} \int_{d_1}^s \frac{dt}{\ell(t)}, \quad \omega(s) = \log \ell(s) - \Gamma(s) - \log \sqrt{2\pi}. \quad (7.4)$$

Since $\ell(t) = t$ on $[d_1, d]$, $\Gamma(s)$ agrees with $\frac{1}{2} \log d + \frac{1}{2} \int_d^s dt/\ell(t)$, the function Γ of Definition 4.1.

Definition 7.1. Let $\zeta \in \mathcal{K}(d_1, S)$. Put

$$W_\zeta(y) = U_\zeta(S, y) + \frac{y^2}{2\ell_0} - \omega(S) \mathbf{1}_{\{y \leq 0\}} + \log|y| \mathbf{1}_{\{y \leq -1\}},$$

$$\tilde{A}(\zeta) = \frac{1}{8} \left[\int_{d_1}^S \left(\frac{1}{\ell_0} - \frac{1}{\ell(s)} \right) ds + \frac{d_1}{\ell_0} - \log d_1 \right], \quad \mathcal{P}^{(1)}(\zeta) = \tilde{A}(\zeta) + 2\phi_0 \int_{\mathbb{R}} W_\zeta(y) dy.$$

For a normalized profile $\zeta \in \mathcal{K}^1(d_1, S)$ we write $A(\zeta) = \tilde{A}(\zeta)$, $I(\zeta) = \int_{\mathbb{R}} W_\zeta dy - 1$ and, for $t \in \mathbb{R}$,

$$F_t(\zeta) = A(\zeta) + t I(\zeta).$$

We call F_t the *flat functional*. Its *crossing parameter* is $t_d = 1/(4c_0)$, where $c_0 = \int_{\mathbb{R}} V(1 - V)$.

Lemma 7.2 shows that $W_\zeta \in L^1(\mathbb{R})$, and Corollary 7.4 shows that $\mathcal{P}^{(1)}(\zeta)$ does not depend on the pair (d_1, S) used to compute it. The two terms have different roles. The term $\tilde{A}$ collects the order- e part of the entropy and of the bottom Gaussian term. On normalized profiles it is maximized by $\mathbf{1}_{[0,1]}$ (Corollary 7.6), so it favors replica symmetry breaking. The term $\int W_\zeta$ is the order- e part of the energy $\alpha \mathbb{E}U(s_0, Y_0)$. On bounded sets the law of the bottom position Y_0 , multiplied by α , has density $2e\phi_0(1 + o(1))$ by Lemma 4.3, which is why Lebesgue measure appears.

7.2 The compact-class expansion

Let $K_e = \log e - \alpha \mathbb{E}[\log |Y^*|; Y^* \leq -1] - \alpha \Phi(\kappa) \log \sqrt{2\pi}$, with $Y^* = (G - \kappa)/e$, be the profile-independent constant of Proposition 4.5. For $l > 0$ put

$$\Psi_\theta(l) = \frac{2\phi_0\theta}{l} + \log l, \quad g_\theta(l) = \left(\phi_0 - \frac{1}{8\phi_0}\right)\frac{\theta}{l} + \frac{1}{4}\log l. \quad (7.5)$$

The scalar function Ψ_θ of the mass always carries the subscript θ ; it is unrelated to the spatial function $\Psi(s, y) = \ell U_y + y$.

The following uniform description of U_ζ as $y \rightarrow \pm\infty$ is the analytic input of the expansion. Parts (a) and (d) are the asymptotics of Lemma 2.4 read through the dilation; parts (b) and (c) come from Feynman–Kac representations along the rescaled diffusion and a transport estimate for it. The proof is in §C.2.

Lemma 7.2. Fix $0 < d_1 \leq S$. There are $C, y_* < \infty$, depending only on (d_1, S) , such that for every $\zeta \in \mathcal{K}(d_1, S)$ and every $s \in [d_1, S]$ the following hold.

- (a) $|\Psi(s, y) - \ell(s)/|y|| \leq C|y|^{-3}$ for $y \leq -y_*$.
- (b) $0 \leq 1 - v(s, y) \leq C(1 + y^2)^{-1}$ for $y \leq 0$, and $0 \leq v(s, y) \leq Ce^{-y^2/C}$ for $y \geq 0$.
- (c) $|v_y(s, y)| + |v_{yy}(s, y)| \leq C(1 + |y|)^{-3}$ for $y \leq 0$, and $|v_y(s, y)| + |v_{yy}(s, y)| \leq Ce^{-y^2/C}$ for $y \geq 0$.
- (d) $|W_\zeta(y)| \leq C(1 + y^2)^{-1}$ for all $y \in \mathbb{R}$; in particular $W_\zeta \in L^1(\mathbb{R})$.

Moreover $\zeta \mapsto W_\zeta$ is continuous from $(\mathcal{K}(d_1, S), L^1)$ to $L^1(\mathbb{R})$, and $\tilde{A}$ is Lipschitz on $(\mathcal{K}(d_1, S), L^1)$.

Theorem 7.3. Fix $0 < d_1 \leq S$. There are $C, e_x > 0$ such that for $0 < e < e_x$, $\theta \in \Theta$, and every $\gamma \in \mathcal{U}$ whose rescaled profile $\zeta = \zeta_\gamma$ lies in $\mathcal{K}(d_1, S)$,

$$\left| \mathcal{P}_{\kappa_e(\theta)}(\gamma) - K_e - \Psi_\theta(\ell_0) - e\{g_\theta(\ell_0) + \mathcal{P}^{(1)}(\zeta)\} \right| \leq Ce^2 L_e, \quad (7.6)$$

where $\ell_0 = \lambda_\gamma(0)/e^2$ is the mass. The function $\mathcal{P}^{(1)}$ is continuous on $(\mathcal{K}(d_1, S), L^1)$, and for every $\zeta \in \mathcal{K}(d_1, S)$

$$\mathcal{P}^{(1)}(\zeta) = F_{2\phi_0\sqrt{\ell_0}}(\tilde{\zeta}) - \frac{1}{8} \log \ell_0 + 2\phi_0. \quad (7.7)$$

The proof is in §C.3. It applies the exact decomposition of Proposition 4.5 at the declared level $q_1 = 1 - e^2 S$, which may lie below the bottom of the support, and expands the scalar coefficients with Lemma 4.6. The only term that is not explicit is $\alpha \mathbb{E} W_\zeta(Y_1)$, where $Y_1 = (B_{q_1} - \kappa)/e$; it equals $2e\phi_0 \int W_\zeta + O(e^2 L_e)$ because of the uniform majorant of Lemma 7.2(d). The constants depend on Θ only through $\theta_\pm$.

Corollary 7.4. $\mathcal{P}^{(1)}(\zeta)$ does not depend on the choice of (d_1, S) with $\zeta \in \mathcal{K}(d_1, S)$.

Proof. Let $\zeta \in \mathcal{K}(d_1, S) \cap \mathcal{K}(d'_1, S')$. Then $\zeta \in \mathcal{K}(d''_1, S'')$ with $d''_1 = \min\{d_1, d'_1\}$ and $S'' = \max\{S, S'\}$. Fix $\theta \in \Theta$. For $e < (S'')^{-1/2}$ let $\gamma \in \mathcal{U}$ be the right-continuous function with $\gamma(q) = \zeta((1-q)/e^2)$ for $q \geq 1 - e^2 S''$ and $\gamma = 0$ below, so that $\zeta_\gamma = \zeta$. Apply Theorem 7.3 to γ with the pairs (d_1, S) and (d''_1, S'') . Both expansions have the same left side and the same terms K_e , $\Psi_\theta(\ell_0)$, and $g_\theta(\ell_0)$. Subtracting, dividing by e , and letting $e \rightarrow 0$ shows that the two values of $\mathcal{P}^{(1)}(\zeta)$ coincide. The same argument compares (d'_1, S') with (d''_1, S'') . $\square$

The expansion identifies the minimization problem in two layers. The order-one term $\Psi_\theta(\ell_0)$ has a unique nondegenerate minimum at

$$\ell^* = 2\phi_0\theta, \quad \Psi_\theta(l) - \Psi_\theta(\ell^*) = \frac{\ell^*}{l} - 1 - \log \frac{\ell^*}{l} \geq 0, \quad (7.8)$$

so it fixes the mass. At order e the profile is selected by $\mathcal{P}^{(1)}$, that is, by the flat functional of the normalized profile at $t = 2\phi_0\sqrt{\ell_0}$. At the crossing margin κ_d , the parameter $\theta = (\kappa_{RS} - \kappa_d)/e^2$ tends to $\theta_d = (2\pi)^{3/2}/(128c_0^2) = 1/(128c_0^2\phi_0^3)$ as $e \rightarrow 0$ (Proposition 3.8(iii)). At $\theta = \theta_d$ the optimal mass is $\ell^* = d_0 := 2\phi_0\theta_d = 1/(64c_0^2\phi_0^2)$, and, as in Proposition 3.8(iv),

$$2\phi_0\sqrt{2\phi_0\theta_d} = \frac{2\phi_0}{8c_0\phi_0} = \frac{1}{4c_0} = t_d. \quad (7.9)$$

7.3 The inverse parametrization

For normalized profiles we change the independent variable from the depth s to the mass $l = \ell(s) \in [0, 1]$. This makes the entropy term affine.

Let $\zeta \in \mathcal{K}^1(d_1, S)$ and let $s_0 = \inf\{s : \ell(s) = 1\} \leq S$ be the bottom of its support. Since ζ is nonincreasing and $\ell(s) < 1 = \ell(s_0)$ for $s < s_0$, $\zeta > 0$ on $[0, s_0)$, so ℓ is strictly increasing on $[0, s_0]$. Let $T: [0, 1] \rightarrow [0, s_0]$ be its inverse. Then $T(l) = l$ on $[0, d_1]$, $T(1) = s_0$, and $T'(l) = 1/\zeta(T(l)) \geq 1$ for almost every l . Since ζ is nonincreasing,

ℓ is concave, so its increasing inverse T is convex. The inverse T is continuous on $[0, 1]$. Convexity together with continuity at both endpoints implies that T is absolutely continuous on $[0, 1]$. Conversely, every continuous convex T on $[0, 1]$ with $T(0) = 0$, $T' = 1$ on $[0, d_1]$ and $T(1) < \infty$ is the inverse of a unique $\zeta \in \mathcal{K}^1(d_1, T(1))$, namely the derivative of T^{-1} extended by zero. Normalized step profiles correspond to piecewise linear T . Bounded support is used here: a normalized profile with unbounded support has $T(1) = \infty$. We write $\sigma = T'$, a nondecreasing function. In the variable l , σ plays the role of a diffusivity. The replica-symmetric profile $\mathbf{1}_{[0,1]}$ corresponds to $T = \text{id}$ and $\sigma \equiv 1$.

Neither A nor I depends on the pair (d_1, S) with $\zeta \in \mathcal{K}^1(d_1, S)$. Indeed, $\ell(s) = s$ on $[d_1, d]$ and $\ell = \ell_0$ on $[s_0, S]$, so

$$8\tilde{A}(\zeta) = \int_d^{s_0} \left(\frac{1}{\ell_0} - \frac{1}{\ell(s)} \right) ds + \frac{d}{\ell_0} - \log d,$$

which involves neither d_1 nor S ; and $2\phi_0(I + 1) = \mathcal{P}^{(1)} - \tilde{A}$ does not depend on them by Corollary 7.4. We therefore compute A and I with the cutoff $S = s_0 = T(1)$ whenever we work in the variable l .

Proposition 7.5. *Let $\zeta \in \mathcal{K}^1(d_1, S)$, with inverse profile T and $\sigma = T'$. Then*

$$8A(\zeta) = 1 - \int_0^1 (\sigma(l) - 1) \frac{1-l}{l} dl, \quad \Gamma(s_0) = \frac{1}{2} \int_0^1 \frac{\sigma(l) - 1}{l} dl. \quad (7.10)$$

Moreover, $\mathcal{U}(l, y) = U_\zeta(T(l), y)$ satisfies $\mathcal{U}(l, y) = \log \Phi(y/\sqrt{l})$ for $l \leq d_1$, $\mathcal{U}(1, \cdot) = U_\zeta(s_0, \cdot)$, and, for almost every $l \in (d_1, 1)$,

$$\mathcal{U}_l = \frac{1}{2} \mathcal{U}_y^2 + \frac{\sigma(l)}{2} \mathcal{U}_{yy}. \quad (7.11)$$

Proof. Substitute $s = T(l)$, $ds = \sigma dl$, and $\ell(T(l)) = l$. Since $(1 - 1/l)\sigma = -(\sigma - 1)(1 - l)/l + (1 - 1/l)$,

$$\int_{d_1}^{s_0} \left(1 - \frac{1}{\ell(s)} \right) ds = \int_{d_1}^1 \left(1 - \frac{1}{l} \right) \sigma dl = - \int_{d_1}^1 (\sigma - 1) \frac{1-l}{l} dl + 1 - d_1 + \log d_1.$$

Adding $d_1 - \log d_1$ and using $\sigma = 1$ on $[0, d_1]$ gives the formula for $8A$. Similarly, $\Gamma(s_0) = \frac{1}{2} \log d_1 + \frac{1}{2} \int_{d_1}^1 \sigma dl/l = \frac{1}{2} \int_{d_1}^1 (\sigma - 1) dl/l$. Finally, $\mathcal{U}_l = \sigma U_s$ and $\sigma \zeta(T(l)) = 1$ give (7.11). $\square$

From now on, in the variable l , we write $U(l, y)$ for $\mathcal{U}(l, y)$. Since $\ell_0 = 1$, we have $\omega(s_0) = -\Gamma(s_0) - \log \sqrt{2\pi}$, and

$$\begin{aligned} I(\zeta) &= \int_{\mathbb{R}} \left[U(1, y) + \frac{y^2}{2} + (\Gamma_\sigma + \log \sqrt{2\pi}) \mathbf{1}_{\{y \leq 0\}} + \log |y| \mathbf{1}_{\{y \leq -1\}} \right] dy - 1, \\ \Gamma_\sigma &= \frac{1}{2} \int_0^1 \frac{\sigma - 1}{l} dl. \end{aligned} \quad (7.12)$$

Only $U(1, \cdot)$ depends nonlinearly on σ ; the counterterms are affine in σ .

Corollary 7.6. *For every normalized profile ζ with bounded support and $\zeta = 1$ near 0, $A(\zeta) \leq 1/8$, with equality if and only if $\zeta = \mathbf{1}_{[0,1]}$.*

Proof. In (7.10), $\sigma - 1 \geq 0$ and $(1 - l)/l > 0$ on $(0, 1)$. The integral vanishes if and only if $\sigma = 1$ almost everywhere, that is, $T = \text{id}$. $\square$

Corollary 7.6 explains why the flat problem is not trivial: moving away from $\mathbf{1}_{[0,1]}$ always decreases A , and the question is whether tI increases enough to compensate. For t near 0 it does not.

7.4 Monotonicity of the flat density

We now fix a normalized step profile, that is, a piecewise linear convex T . Its breakpoints $d_1 \leq l_0 < l_1 < \dots < l_n = 1$ define *blocks* $(l_{j-1}, l_j]$, on which $\sigma = \sigma_j$ is constant, and $\sigma = 1$ on $(0, l_0]$. The statements of this section hold for every positive step function σ of this form, not only for $\sigma \geq 1$ nondecreasing, and we state them in that generality because the proofs use only $\sigma > 0$.

On a block, $F := e^{U/\sigma_j}$ solves the heat equation $F_l = \frac{\sigma_j}{2} F_{yy}$, by (7.11). Hence, with $P_\tau f(y) = \mathbb{E}f(y + \sqrt{\tau} G)$,

$$U(l, \cdot) = \sigma_j \log P_{\sigma_j(l-l_{j-1})} [e^{U(l_{j-1}, \cdot)/\sigma_j}], \quad l \in (l_{j-1}, l_j]. \quad (7.13)$$

We take (7.13), with $U(l, y) = \log \Phi(y/\sqrt{l})$ for $l \leq l_0$, as the definition of U for a general positive step function σ . Write $b = U_y$ and $c = -U_{yy}$, and let

$$\mathcal{G} = b \partial_y + \frac{\sigma}{2} \partial_{yy}$$

be the operator that appears in the linearization of (7.11).

The following signs are the rescaled counterparts of Lemma 2.4(c); for σ that comes from a profile, the upper bound $c \leq 1/l$ is the rescaled form of the bound $c < 1/\lambda$ there. What the lemma adds is that the signs, including $c \leq 1/l$, hold for every positive step function σ . The bound $c \leq 1/l$ controls the flat density below. For the Sherrington–Kirkpatrick datum $\log \cosh$, the analogous bounds on the first two derivatives are [8, Proposition 2], [7, Proposition 2], and [76, Lemma 16], and the sign of the third derivative on the positive half-line is [87, (7.6)].

Lemma 7.7. *Let $\sigma > 0$ be a step function on a grid of $[l_0, 1]$ with $\sigma = 1$ on $(0, l_0]$, and let U be defined by (7.13). For every $l \in (0, 1]$, $U(l, \cdot) \leq 0$ is smooth, $b > 0$, and*

$$0 < c(l, y) \leq 1/l, \quad c_y(l, y) \leq 0.$$

Proof. For $l \leq l_0$ we have $U = \log \Phi(y/\sqrt{l})$, so $b = R(y/\sqrt{l})/\sqrt{l} > 0$, $c = V(y/\sqrt{l})/l \in (0, 1/l)$ and $c_y = V'(y/\sqrt{l})l^{-3/2} < 0$ by Lemma 2.1(a). We argue by induction over the blocks. Let $l \in (l_{j-1}, l_j]$, $\sigma = \sigma_j$, $\tau = \sigma(l - l_{j-1})$, and $f = e^{U(l_{j-1}, \cdot)/\sigma}$.

We verify the four claims in turn.

STEP 1. $U \leq 0$ and $b > 0$. The function f takes values in $(0, 1]$ and is strictly increasing, and P_τ preserves both properties.

STEP 2. $c \geq 0$. The function f is log-concave because $U(l_{j-1}, \cdot)$ is concave. The Gaussian convolution $P_\tau f$ is a marginal of the log-concave function $(w, y) \mapsto f(w)\phi_\tau(y - w)$, where ϕ_τ is the $N(0, \tau)$ density, and the marginal of a log-concave function is log-concave by Prékopa's theorem [108, Theorem 6].

STEP 3. $c \leq 1/l$. By induction, $-(\log f)'' = c(l_{j-1}, \cdot)/\sigma \leq K_0 := 1/(\sigma l_{j-1})$, so $f(w) = e^{-K_0 w^2/2} g(w)$ with g log-convex. Completing the square,

$$P_\tau f(y) = \frac{e^{-K_0 y^2/(2(1+K_0\tau))}}{\sqrt{1+K_0\tau}} \mathbb{E} g\left(\frac{y}{1+K_0\tau} + \sqrt{\frac{\tau}{1+K_0\tau}} G\right).$$

A mixture of log-convex functions is log-convex, so the expectation is log-convex in y . Hence $-(\log P_\tau f)'' \leq K_0/(1+K_0\tau) = 1/(\sigma l)$, that is, $c \leq 1/l$.

STEP 4. $c > 0$ and $c_y \leq 0$. Differentiating (7.11) twice and three times in y gives, on each block,

$$c_l = \mathcal{G}c - c \cdot c, \quad g_l = \mathcal{G}g - 3c g \quad (g = c_y),$$

and c, g are continuous in l across the breakpoints. The potentials c and $3c$ are nonnegative and bounded by $3/l_0$, the drift b is Lipschitz in y (because $|b_y| = c \leq 1/l_0$), and c, g have polynomial growth. Let Y solve

$$Y_\lambda = y + \int_\lambda^l b(\lambda', Y_{\lambda'}) d\lambda' + \int_\lambda^l \sqrt{\sigma(\lambda')} d\bar{B}_{\lambda'} \quad (l_0 \leq \lambda \leq l),$$

where $\bar{B}$ is a Brownian motion in the reversed time $l - \lambda$. This is the diffusion with generator $\mathcal{G}$ run from (l, y) downward in l , and it is well posed because b is Lipschitz in y . By the Feynman–Kac formula along Y (see [79]),

$$c(l, y) = \mathbb{E} \left[c(l_0, Y_{l_0}) e^{-\int_{l_0}^l c(\lambda, Y_\lambda) d\lambda} \right] > 0, \quad g(l, y) = \mathbb{E} \left[g(l_0, Y_{l_0}) e^{-\int_{l_0}^l 3c(\lambda, Y_\lambda) d\lambda} \right] \leq 0,$$

because $c(l_0, \cdot) > 0$ and $g(l_0, \cdot) < 0$. □

The diffusion in the Feynman–Kac formula is the rescaled Parisi diffusion of Definition 4.1, written in the variable l (see (C.1)). Started at the initial level $l = 1$ (the bottom of the support), it defines the density that governs the second variation.

Definition 7.8. Let σ be as in Lemma 7.7. For $y \in \mathbb{R}$ let $(Y_l^y)_{0 \leq l \leq 1}$ solve, downward in l ,

$$Y_1^y = y, \quad Y_l^y = y + \int_l^1 b(\lambda, Y_\lambda^y) d\lambda + \int_l^1 \sqrt{\sigma(\lambda)} d\bar{B}_\lambda,$$

where $\bar{B}$ is a Brownian motion in the reversed time $1 - l$. The *flat density* p_l is the density of the measure $E \mapsto \int_{\mathbb{R}} \mathbb{P}(Y_l^y \in E) dy$.

For each Brownian path, $y \mapsto Y_l^y$ is an increasing C^1 bijection of $\mathbb{R}$, because b is smooth and Lipschitz in y ; differentiating the equation gives the Jacobian $\partial_y Y_l^y = \exp(-\int_l^1 c(\lambda, Y_\lambda^y) d\lambda)$, which lies in $[l, 1]$ by Lemma 7.7. The image of Lebesgue measure under such a map has density $1/\partial_y Y_l^y$ at the point Y_l^y , which lies in $[1, 1/l]$. Averaging over the Brownian path, p_l exists and

$$1 \leq p_l \leq 1/l, \quad p_1 \equiv 1. \quad (7.14)$$

When σ comes from a profile, p_l plays the role of the law of the rescaled Parisi diffusion at the level l , with Lebesgue measure at $l = 1$ in place of the law of the bottom position Y_0 (see the end of §7.1).

On a block, the diffusion is the Doob transform, by the space-time harmonic function $F = e^{U/\sigma_j}$, of a Brownian motion with variance rate σ_j (see [50, 110]). Its transition density from level l_j to level $l \in [l_{j-1}, l_j]$ is $\phi_\Delta(x - y)F(l, x)/F(l_j, y)$, where $\Delta = \sigma_j(l_j - l)$ and ϕ_Δ is the $N(0, \Delta)$ density. Integrating against p_{l_j} ,

$$p_l(x) = F(l, x) \int_{\mathbb{R}} \phi_\Delta(x - y) \frac{p_{l_j}(y)}{F(l_j, y)} dy, \quad F(l_j, \cdot) = P_\Delta F(l, \cdot). \quad (7.15)$$

The integral converges: $p_{l_j} \leq 1/l_j$; by concavity and $c \leq 1/l_j$, $U(l_j, y) \geq U(l_j, 0) + b(l_j, 0)y - y^2/(2l_j)$, so $1/F(l_j, y) \leq \exp((y^2/(2l_j) + C(1 + |y|))/\sigma_j)$; and $\Delta < \sigma_j l_j$.

For the Sherrington–Kirkpatrick model, monotonicity and log-concavity properties of a density of the Parisi diffusion were proved in [87, Proposition 7.1], for the product of the density with an exponential of the solution of the Parisi equation, by following the heat evolution between successive atoms. The next proposition is proved in the same block-by-block way. Here the initial law is Lebesgue measure, the monotone quantity is the density itself, and σ is an arbitrary positive step function.

Proposition 7.9. *Let $\sigma > 0$ be a step function on a grid of $[l_0, 1]$ with $\sigma = 1$ on $(0, l_0]$. For every $l \in [l_0, 1]$ the function $y \mapsto p_l(y)$ is nonincreasing. In particular $y \mapsto p_l(y)c(l, y)$ is nonincreasing.*

Proof. We argue by induction from the initial level $l = 1$ downward. At $l = 1$, $p_1 \equiv 1$. Suppose that p_{l_j} is nonincreasing, and let $l \in [l_{j-1}, l_j]$. Write $\sigma = \sigma_j$, $\Delta = \sigma(l_j - l)$, and

$f_\lambda = U(\lambda, \cdot)/\sigma = \log F(\lambda, \cdot)$. By Lemma 7.7, at every level λ the function $f'_\lambda = b/\sigma$ is positive, decreasing ($f''_\lambda = -c/\sigma < 0$), and convex ($f'''_\lambda = -c_y/\sigma \geq 0$).

Fix y and let $\mathbb{Q}_y(dx) \propto \phi_\Delta(x - y) p_{l_j}(x) F(l_j, x)^{-1} dx$. Differentiating (7.15) and integrating by parts in x (the boundary terms vanish by the bounds after (7.15)),

$$(\log p_l)'(y) = f'_l(y) + \mathbb{E}_{\mathbb{Q}_y} \left[\frac{X - y}{\Delta} \right] = f'_l(y) + \mathbb{E}_{\mathbb{Q}_y} [(\log p_{l_j})'(X) - f'_{l_j}(X)].$$

Put $\bar{r} = \mathbb{E}_{\mathbb{Q}_y} f'_{l_j}(X)$. Since $(\log p_{l_j})' \leq 0$, the two expressions for the same expectation give

$$(\log p_l)'(y) \leq f'_l(y) - \bar{r}, \quad \mathbb{E}_{\mathbb{Q}_y} X \leq y - \Delta \bar{r}. \quad (7.16)$$

STEP 1. *A lower bound for $\bar{r}$.* By Jensen's inequality (f'_{l_j} is convex), then monotonicity (f'_{l_j} is decreasing), and (7.16),

$$\bar{r} \geq f'_{l_j}(\mathbb{E}_{\mathbb{Q}_y} X) \geq f'_{l_j}(y - \Delta \bar{r}).$$

STEP 2. *A fixed point.* The function $\mathcal{R}(r) = f'_{l_j}(y - \Delta r) - r$ is continuous, and by $c \leq 1/l_j$,

$$\mathcal{R}'(r) = \frac{\Delta}{\sigma} c(l_j, y - \Delta r) - 1 = (l_j - l) c(l_j, y - \Delta r) - 1 \leq \frac{l_j - l}{l_j} - 1 = -\frac{l}{l_j} < 0.$$

Hence $\mathcal{R}$ has a unique zero r^* . Step 1 says $\mathcal{R}(\bar{r}) \leq 0$, so $\bar{r} \geq r^*$. Let $x^* = y - \Delta r^*$. Then $r^* = f'_{l_j}(x^*)$ and $y = x^* + \Delta f'_{l_j}(x^*)$.

STEP 3. *Tweedie's formula.* Since $F(l_j, \cdot) = P_\Delta F(l, \cdot)$, the probability measure $\pi(dw) \propto F(l, w) \phi_\Delta(w - x^*) dw$ satisfies

$$\begin{aligned} \mathbb{E}_\pi W &= x^* + \Delta (\log F(l_j, \cdot))'(x^*) = x^* + \Delta f'_{l_j}(x^*) = y, \\ f'_{l_j}(x^*) &= \frac{(P_\Delta F(l, \cdot))'(x^*)}{P_\Delta F(l, \cdot)(x^*)} = \mathbb{E}_\pi f'_l(W). \end{aligned}$$

By Jensen's inequality (f'_l is convex), $r^* = \mathbb{E}_\pi f'_l(W) \geq f'_l(\mathbb{E}_\pi W) = f'_l(y)$. The formula for $\mathbb{E}_\pi W$ is Tweedie's formula [53, 111].

Combining the three steps with (7.16), $(\log p_l)'(y) \leq f'_l(y) - \bar{r} \leq f'_l(y) - r^* \leq 0$. The flat density is continuous in l at the breakpoints (at $l = l_j$, (7.15) returns p_{l_j}), so the induction proceeds block by block. Since p_l and c are positive and nonincreasing, so is $p_l c$. $\square$

7.5 Convexity along ordered rays

Let T_0, T_1 be the inverse profiles of two normalized step profiles, on a common grid with $\sigma = 1$ on $(0, l_0]$, such that $K := T_1 - T_0 \geq 0$ on $[0, 1]$. For $\vartheta \in [0, 1]$ put $T_\vartheta = T_0 + \vartheta K$,

so that $\sigma_\vartheta = T'_\vartheta = \sigma_0 + \vartheta \dot{\sigma}$ with $\dot{\sigma} = K'$, and let U^ϑ be the corresponding solution and $I(\vartheta) = I(\zeta_\vartheta)$. The main example is the ray from replica symmetry: $T_0 = \text{id}$ and $T_1 = T$ the inverse of an arbitrary normalized step profile, for which $K = T - \text{id} \geq 0$ because $T' \geq 1$. For $\vartheta \in [0, 1]$, T_ϑ is again convex with $T'_\vartheta \geq 1$, $T'_\vartheta = 1$ on $[0, l_0]$ and $T_\vartheta(1) \leq \max\{T_0(1), T_1(1)\}$, so every ζ_ϑ lies in the fixed class $\mathcal{K}^1(l_0, \max\{T_0(1), T_1(1)\})$.

Write $\dot{U} = \partial_\vartheta U^\vartheta$ and $\ddot{U} = \partial_\vartheta^2 U^\vartheta$. Differentiating (7.11) in ϑ , and using that σ_ϑ is affine in ϑ ,

$$\dot{U}_l = \mathcal{G}\dot{U} - \frac{\dot{\sigma}}{2}c, \quad \ddot{U}_l = \mathcal{G}\ddot{U} + \dot{U}_y^2 + \dot{\sigma}\dot{U}_{yy}, \quad \dot{U} = \ddot{U} = 0 \text{ for } l \leq l_0. \quad (7.17)$$

The function that makes the second variation transparent is

$$\mathcal{Z} = \dot{U} + \frac{K}{2}c, \quad \text{which solves} \quad \mathcal{Z}_l = \mathcal{G}\mathcal{Z} - \frac{K}{2}c^2, \quad \mathcal{Z} = 0 \text{ } (l \leq l_0). \quad (7.18)$$

Indeed, $c_l = \mathcal{G}c - c^2$ (proof of Lemma 7.7), $K' = \dot{\sigma}$, and K depends on l only, so $\mathcal{G}(Kc) = K\mathcal{G}c$.

Proposition 7.10. *In the setting above, $\vartheta \mapsto I(\vartheta)$ is of class C^2 on $[0, 1]$ and*

$$I''(\vartheta) = \int_{l_0}^1 \int_{\mathbb{R}} \left[p_l \mathcal{Z}_y^2 - 2K(p_l c)_y \mathcal{Z}_y + \frac{K^2}{2} p_l c_y^2 \right] dy dl, \quad (7.19)$$

where p , c , and $\mathcal{Z}$ are computed at σ_ϑ .

The proof, in §C.4, starts from $I'' = \iint p (\dot{U}_y^2 + \dot{\sigma}\dot{U}_{yy})$ and substitutes $\dot{U} = \mathcal{Z} - Kc/2$. All terms that involve derivatives in l combine into $(\partial_l - \mathcal{G})(K\mathcal{Z}_{yy} - K^2 c_{yy}/4)$, whose integral against p vanishes; one integration by parts in y then gives (7.19).

Corollary 7.11. *In the setting of Proposition 7.10,*

$$I''(\vartheta) \geq \frac{1}{2} \int_{l_0}^1 K(l)^2 \int_{\mathbb{R}} p_l c_y^2 dy dl \geq 0.$$

Consequently, for every $t \geq 0$, $\vartheta \mapsto F_t(\zeta_\vartheta)$ is convex on $[0, 1]$.

Proof. We show that each term in (7.19) is nonnegative. By (7.18) and the Feynman–Kac formula along the diffusion of Definition 7.8,

$$\mathcal{Z}(l, y) = -\frac{1}{2} \mathbb{E} \int_{l_0}^l K(\lambda) c(\lambda, Y_\lambda)^2 d\lambda,$$

where Y starts at (l, y) and runs downward in l . By Lemma 7.7, c^2 is nonincreasing in y , and the flow $y \mapsto Y_\lambda$ is increasing; since $K \geq 0$, $\mathcal{Z}$ is nondecreasing in y , so $\mathcal{Z}_y \geq 0$. By Proposition 7.9, $(p_l c)_y \leq 0$. Hence $-2K(p_l c)_y \mathcal{Z}_y \geq 0$, and the other two terms are nonnegative. The functional A is affine in T by Proposition 7.5, so $F_t = A + tI$ is convex along the segment for $t \geq 0$. $\square$

Only the ordering $K \geq 0$ enters the proof. For directions K of both signs the middle term of (7.19) has no sign, and we do not know whether I is convex along such segments; the ordered rays from the replica-symmetric profile suffice for Theorem 7.13. Lemma 7.7 and Proposition 7.9 hold for every positive step function σ , while the proof of Proposition 7.10 uses the tail estimates of Lemma 7.2, which are proved for profiles.

This convexity differs from the classical convexity of Parisi functionals: of the Crisanti–Sommers functional for spherical models [125], and of the Parisi functional for Ising spins [7, Theorem 2] (see Chapter 1). Those results concern linear interpolation of distribution functions, and the proofs for Ising spins use the convexity of the terminal datum $\log \cosh$ in the spatial variable. The terminal datum $\log \Phi$ of (1.2) is concave (see the opening of Chapter 2), so these arguments do not apply to $\mathcal{P}_\kappa$. Corollary 7.11 concerns only the flat functional, interpolates linearly in the inverse profile T rather than in the profile, and is proved only along ordered segments.

7.6 The first variation at the replica-symmetric profile

By Corollary 7.11, $F_{t_d}(\zeta) - F_{t_d}(\mathbf{1}_{[0,1]})$ is bounded below by the one-sided derivative of F_{t_d} at $T = \text{id}$ in the direction $K = T - \text{id}$. We now compute this derivative and show that it is positive, with cubic contact at the edge of the admissible cone.

At $T = \text{id}$ the solution is $U(l, y) = \log \Phi(y/\sqrt{l})$, and the diffusion of Definition 7.8 is the Doob transform of Brownian motion by $\Phi(\cdot/\sqrt{l})$. In the variables

$$Z = Y_l/\sqrt{l}, \quad \zeta = \log(1/l),$$

Itô's formula gives the time-homogeneous diffusion

$$dZ = (R(Z) + \frac{Z}{2}) d\zeta + dB_\zeta, \quad \mathcal{G}_{\text{top}} = \frac{1}{2} \partial_{zz} + (R(z) + \frac{z}{2}) \partial_z. \quad (7.20)$$

It describes the normalized position of the Parisi diffusion above the atom of a replica-symmetric order parameter, which is why it reappears in Chapter 8; its transition kernel is computed in Lemma D.1. We write Z^z for the solution started at $Z_0 = z$. For $0 < \rho < 1$ define

$$\mathcal{Q}(\rho) = \int_{\mathbb{R}} \left[V(z)^2 - \mathbb{E} V(Z_{\log(1/\rho)}^z)^2 \right] dz. \quad (7.21)$$

This is the decrease of $\int V^2$ when Lebesgue measure is transported by the diffusion (7.20) for the time $\log(1/\rho)$. The integrand is integrable by the Dynkin calculation in the proof below, although V^2 itself is not integrable.

Proposition 7.12. *Let T be the inverse profile of a normalized step profile, let $K = T - \text{id}$, and let ζ_ϑ be the normalized profile with inverse profile $\text{id} + \vartheta K$. The function $\dot{\sigma} = K' = \sigma - 1$ is a nonnegative nondecreasing step function that vanishes on $[0, l_0]$, and*

$$\left. \frac{d}{d\vartheta} F_{t_d}(\zeta_\vartheta) \right|_{\vartheta=0} = \int_{(0,1]} \mathcal{E}(r) d\dot{\sigma}(r), \quad \mathcal{E}(r) = \frac{1}{8c_0} \int_r^1 \mathcal{Q}(\rho) \frac{\rho - r}{\rho^2} d\rho, \quad (7.22)$$

where $d\dot{\sigma}$ is the Stieltjes measure of $\dot{\sigma}$. Moreover $\mathcal{Q} > 0$ on $(0, 1)$, and with $c_{\text{FS}} = 1/96$,

$$\mathcal{E}(r) \geq c_{\text{FS}}(1-r)^3 \quad (0 < r < 1). \quad (7.23)$$

Proof. We divide the proof into three steps. First, we write the derivative as an integral of $\dot{\sigma}$ against the defect of a transported Lebesgue measure. Second, a Dynkin calculation expresses that defect through $\mathcal{Q}$. Third, a comparison with an Ornstein–Uhlenbeck process bounds $\mathcal{Q}$ from below.

STEP 1. The first variation. By Lemma C.4 at $\vartheta = 0$ and (7.12), $I'(0) = \int_{\mathbb{R}} [\dot{U}(1, y) + \dot{\Gamma} \mathbf{1}_{\{y \leq 0\}}] dy$ with $\dot{\Gamma} = \frac{1}{2} \int_0^1 \dot{\sigma}(l) dl/l$ and $\dot{U}(1, y) = -\frac{1}{2} \mathbb{E} \int_0^1 \dot{\sigma}(l) c(l, Y_l^y) dl$. At $T = \text{id}$, $c(l, y) = V(y/\sqrt{l})/l$ and $Y_l^y/\sqrt{l} = Z_{\varsigma_l}^y$ with $\varsigma_l = \log(1/l)$. Hence

$$\dot{U}(1, y) + \dot{\Gamma} \mathbf{1}_{\{y \leq 0\}} = \frac{1}{2} \int_0^1 \frac{\dot{\sigma}(l)}{l} [\mathbf{1}_{\{y \leq 0\}} - \mathbb{E} V(Z_{\varsigma_l}^y)] dl.$$

For fixed y the integrand has a constant sign ($0 < V < 1$), and the left side is bounded by $C(1+y^2)^{-1}$ by Lemma C.4(a). Tonelli's theorem on $\{y \leq 0\}$ and on $\{y > 0\}$ therefore allows us to integrate first in y . Since $\int_{\mathbb{R}} (\mathbf{1}_{\{y \leq 0\}} - V) dy = 0$ by Lemma 2.1(f),

$$I'(0) = \frac{1}{2} \int_0^1 \frac{\dot{\sigma}(l)}{l} M(l) dl, \quad M(l) := \int_{\mathbb{R}} [V(y) - \mathbb{E} V(Z_{\varsigma_l}^y)] dy.$$

By (7.10), $A(\zeta_{\vartheta}) = \frac{1}{8} - \frac{\vartheta}{8} \int_0^1 \dot{\sigma}(l)(1-l)l^{-1} dl$. Since $t_d/2 = 1/(8c_0)$,

$$\left. \frac{d}{d\vartheta} F_{t_d}(\zeta_{\vartheta}) \right|_{\vartheta=0} = \int_0^1 \dot{\sigma}(l) \epsilon(l) dl, \quad \epsilon(l) = \frac{M(l) - c_0(1-l)}{8c_0 l}. \quad (7.24)$$

STEP 2. A Dynkin calculation for the transported Lebesgue measure. The drift $R(z) + z/2$ of (7.20) has derivative $\frac{1}{2} - V \in (-\frac{1}{2}, \frac{1}{2})$, so the flow $z \mapsto Z_{\varsigma}^z$ is an increasing C^1 bijection with Jacobian in $[e^{-\varsigma/2}, e^{\varsigma/2}]$. Let ν_{ς} be the density of the image of Lebesgue measure under this flow; then $\nu_{\varsigma} \leq e^{\varsigma/2}$. Using the identity (2.2) for V'' ,

$$\begin{aligned} \mathcal{G}_{\text{top}} V &= -V(1-V), & \mathcal{G}_{\text{top}}(V^2) &= k_1 := V'^2 - 2V^2(1-V), \\ \mathcal{G}_{\text{top}}(V(1-V)) &= -V(1-V) - k_1. \end{aligned} \quad (7.25)$$

All three right sides are bounded and integrable by Lemma C.1(a). For $f \in \{V, V^2, V(1-V)\}$, Dynkin's formula for each starting point, and Fubini's theorem (justified by $\nu_{\varsigma} \leq e^{\varsigma/2}$ and $\mathcal{G}_{\text{top}} f \in L^1$), give

$$\int_{\mathbb{R}} [\mathbb{E} f(Z_{\varsigma}^z) - f(z)] dz = \int_0^{\varsigma} \int_{\mathbb{R}} \nu_u \mathcal{G}_{\text{top}} f dz du.$$

Put $j(u) = \int \nu_u V(1-V)$ and $\mathfrak{k}(u) = \int \nu_u k_1$, which are bounded on bounded intervals. With $f = V$, $f = V^2$, and $f = V(1-V)$ this gives

$$M(l) = \int_0^{\varsigma_l} j(u) du, \quad \mathcal{Q}(\rho) = - \int_0^{\varsigma_{\rho}} \mathfrak{k}(u) du, \quad j(u) = c_0 - \int_0^u (j + \mathfrak{k}).$$

The last identity is the integral form of $j' = -j - \mathfrak{k}$, $j(0) = c_0$, so $j(u) = c_0 e^{-u} - \int_0^u e^{u'-u} \mathfrak{k}(u') du'$. Integrating over $[0, \zeta_l]$ and applying Fubini's theorem twice (with $e^{-\zeta_l} = l$),

$$M(l) - c_0(1-l) = - \int_0^{\zeta_l} \mathfrak{k}(u') (1 - l e^{u'}) du' = l \int_l^1 \mathcal{Q}(\rho) \rho^{-2} d\rho.$$

The second equality follows by inserting the formula for $\mathcal{Q}$, since $u' \leq \zeta_\rho$ if and only if $\rho \leq e^{-u'}$, and $l \int_l^{e^{-u'}} \rho^{-2} d\rho = 1 - l e^{u'}$. Hence $\mathfrak{e}(l) = (8c_0)^{-1} \int_l^1 \mathcal{Q}(\rho) \rho^{-2} d\rho$. Since $\dot{\sigma}(l) = \int_{(0,l]} d\dot{\sigma}(r)$, Fubini's theorem turns (7.24) into (7.22) with $\mathcal{E}(r) = \int_r^1 \mathfrak{e}(l) dl = (8c_0)^{-1} \int_r^1 \mathcal{Q}(\rho) (\rho - r) \rho^{-2} d\rho$.

STEP 3. *One comparison gives positivity and the cubic bound.* Couple (7.20), started at z , with the Ornstein–Uhlenbeck process $dX = -\frac{1}{2}X d\zeta + dB_\zeta$, $X_0 = z$, driven by the same Brownian motion. The difference $\Xi = Z - X$ satisfies $\Xi' = R(Z) + \frac{1}{2}Z + \frac{1}{2}X = \bar{R}(Z) - \frac{1}{2}\Xi$ and $\Xi_0 = 0$, so

$$\Xi_\zeta = \int_0^\zeta e^{-(\zeta-u)/2} \bar{R}(Z_u) du > 0,$$

because $\bar{R} > 0$ (Lemma 2.1(a)). Thus $Z_\zeta \geq X_\zeta$, and $X_\zeta \sim N(\sqrt{\rho}z, 1-\rho)$ with $\rho = e^{-\zeta}$. This is a pathwise comparison of the type of [74], made explicit here by the linear equation for Ξ . Since V^2 is decreasing, with P_v the convolution with the $N(0, v)$ density,

$$\mathbb{E} V(Z_\zeta^z)^2 \leq P_{1-\rho}(V^2)(\sqrt{\rho}z).$$

This comparison can be integrated despite the nonzero left limit of V^2 . Put $f = V^2 - \mathbf{1}_{\{z < 0\}}$. By Lemma 2.1(f), $f \in L^1$ and $\int f = -c_0$. The function $\mathbf{1}_{\{z < 0\}} - P_{1-\rho} \mathbf{1}_{\{ \cdot < 0 \}}(\sqrt{\rho}z) = \mathbf{1}_{\{z < 0\}} - \Phi(-az)$, $a = \sqrt{\rho/(1-\rho)}$, is integrable with integral $\int_{-\infty}^0 \Phi(az) dz - \int_0^\infty \Phi(-az) dz = 0$. The substitution $w = \sqrt{\rho}z + \sqrt{1-\rho}g$ shows $\int P_{1-\rho} f(\sqrt{\rho}z) dz = \rho^{-1/2} \int f$. Hence

$$\int_{\mathbb{R}} [V(z)^2 - P_{1-\rho}(V^2)(\sqrt{\rho}z)] dz = (1 - \rho^{-1/2}) \int_{\mathbb{R}} f = c_0(\rho^{-1/2} - 1).$$

The integrand of (7.21) is integrable by Step 2 and pointwise at least the integrand on the left. Since $\rho^{-1/2} - 1 \geq 1 - \sqrt{\rho} = (1-\rho)/(1+\sqrt{\rho}) \geq (1-\rho)/2$,

$$\mathcal{Q}(\rho) \geq c_0(\rho^{-1/2} - 1) \geq \frac{c_0}{2}(1-\rho) > 0 \quad (0 < \rho < 1). \quad (7.26)$$

With the formula for $\mathfrak{e}$ from Step 2 and $\rho^{-2} \geq 1$,

$$\mathfrak{e}(l) \geq \frac{1}{16} \int_l^1 (1-\rho) d\rho = \frac{(1-l)^2}{32}, \quad \mathcal{E}(r) = \int_r^1 \mathfrak{e}(l) dl \geq \frac{(1-r)^3}{96}.$$

This proves (7.23). □

The number $\mathcal{E}(r)$ is the derivative of F_{t_d} in the one-step direction $K = (l - r)_+$. Its positivity for every r is the flat counterpart of the statement that the obstacle potential of the replica-symmetric crossing order parameter is strictly increasing above its atom (Proposition 8.4). The contact at $r = 1$ is cubic because at $t = t_d$ the term of order $1 - l$ in $\mathfrak{e}(l)$ cancels, which is the flat analogue of the vanishing of the replicon eigenvalue at κ_d . More precisely, $\mathfrak{e}$ is continuous at 0 with $\mathfrak{e}(0) = \int k_1 = A_V - 2B_V$, where $A_V = \int V'^2$ and $B_V = \int V^2(1 - V)$: for $f \in C_c^2$, Dynkin's formula gives $|\int \nu_u f - \int f| \leq u e^{u/2} \|\mathcal{G}_{\text{top}} f\|_{L^1}$, and $k_1 \in L^1$ is approximated by such f while $\nu_u \leq e^{u/2}$. With $x_f = A_V / (2B_V)$ the flat breakpoint of Definition 8.1, this gives $\mathcal{Q}(\rho) = 2B_V(1 - x_f)(1 - \rho) + o(1 - \rho)$ and

$$\mathcal{E}(r) = \frac{B_V(1 - x_f)}{24c_0} (1 - r)^3 + o((1 - r)^3) \quad (r \uparrow 1).$$

We do not use this expansion. The same computation at a general parameter t shows that t_d is sharp. By Step 1, the derivative of $F_t(\zeta_\vartheta)$ at $\vartheta = 0$ in the direction $K = (l - r)_+$ equals $\int_r^1 [tM(l) - \frac{1}{4}(1 - l)] (2l)^{-1} dl$, and $M(l) = c_0(1 - l) + O((1 - l)^2)$ by Step 2 (because $|\mathfrak{e}| \leq e^{1/2} \|k_1\|_{L^1}$ on $[0, 1]$). For $t < t_d$ this derivative is negative for r close to one, so $\mathbf{1}_{[0,1]}$ is not a minimizer of F_t . This is the flat counterpart of the instability of the replica-symmetric order parameter beyond κ_d (Proposition 3.8(ii)).

7.7 Uniqueness of the flat minimizer

We now combine the ray convexity with the first variation. The cubic contact of (7.23) becomes a cubic L^1 modulus.

Theorem 7.13. *Let ζ be a normalized profile with bounded support and $\zeta = 1$ on a neighborhood of 0, that is, $\zeta \in \bigcup_{0 < d_1 \leq S} \mathcal{K}^1(d_1, S)$. Then*

$$F_{t_d}(\zeta) - F_{t_d}(\mathbf{1}_{[0,1]}) \geq \frac{9}{1024} \|\zeta - \mathbf{1}_{[0,1]}\|_{L^1(0,\infty)}^3. \quad (7.27)$$

Moreover $I(\zeta) \geq I(\mathbf{1}_{[0,1]})$, and for every $t \geq t_d$,

$$F_t(\zeta) - F_t(\mathbf{1}_{[0,1]}) \geq \frac{9c_0 t}{256} \|\zeta - \mathbf{1}_{[0,1]}\|_{L^1(0,\infty)}^3. \quad (7.28)$$

In particular, for every $t \geq t_d$, $\mathbf{1}_{[0,1]}$ is the unique minimizer of F_t on this class.

Proof. We divide the proof into three steps.

STEP 1. *Reduction to the first variation for step profiles.* Let ζ be a normalized step profile, with inverse profile T , and let $K = T - \text{id}$ and $\hat{\sigma} = T' - 1$. By Corollary 7.11 with

$t = t_d > 0$, $\vartheta \mapsto F_{t_d}(\zeta_\vartheta)$ is convex on $[0, 1]$, where ζ_ϑ has inverse profile $\text{id} + \vartheta K$; it is differentiable by Proposition 7.10. By Proposition 7.12,

$$F_{t_d}(\zeta) - F_{t_d}(\mathbf{1}_{[0,1]}) \geq \frac{d}{d\vartheta} F_{t_d}(\zeta_\vartheta) \Big|_{\vartheta=0} \geq \frac{1}{96} \int_{(0,1]} (1-r)^3 d\dot{\sigma}(r).$$

STEP 2. Comparison with the L^1 distance. Let $l_* = \ell(1) \in (0, 1]$, so that $T(l_*) = 1$. Since $\int \zeta = 1 = \int \mathbf{1}_{[0,1]}$,

$$\frac{1}{2} \|\zeta - \mathbf{1}_{[0,1]}\|_{L^1} = \int_0^1 (1 - \zeta) = 1 - l_*.$$

From $T(l) = l + \int_0^l \dot{\sigma}$ and $T(l_*) = 1$, $1 - l_* = \int_0^{l_*} \dot{\sigma}(u) du = \int_{(0,1]} (l_* - r)_+ d\dot{\sigma}(r)$. For $r \leq l_*$ put $a = 1 - l_*$ and $b = l_* - r$. Then $(1 - r)^3 = (a + b)^3 \geq \frac{27}{4} a^2 b$, because $(a + b)^3 - \frac{27}{4} a^2 b = (b - \frac{a}{2})^2 (b + 4a) \geq 0$. Therefore

$$\int_{(0,1]} (1 - r)^3 d\dot{\sigma}(r) \geq \frac{27}{4} (1 - l_*)^2 \int_{(0,1]} (l_* - r)_+ d\dot{\sigma}(r) = \frac{27}{4} (1 - l_*)^3 = \frac{27}{32} \|\zeta - \mathbf{1}_{[0,1]}\|_{L^1}^3.$$

With Step 1 and $\frac{27}{32} \cdot \frac{1}{96} = \frac{9}{1024}$, this proves (7.27) for step profiles.

STEP 3. General profiles and the parameters $t \geq t_d$. Let $\zeta \in \mathcal{K}^1(d_1, S)$. By (7.1), $d_1 \leq 1 \leq S$, so $\mathbf{1}_{[0,1]} \in \mathcal{K}^1(d_1, S)$. Let $\zeta_n \in \mathcal{K}^1(d_1, S)$ be the step approximations (7.2) of ζ . By Lemma 7.2, A and $\zeta \mapsto \int W_\zeta$ are continuous on $\mathcal{K}^1(d_1, S)$, hence so is F_{t_d} . Both sides of (7.27) are continuous in L^1 , so the inequality passes to the limit along $\zeta_n \rightarrow \zeta$. Next, $t_d (I(\zeta) - I(\mathbf{1}_{[0,1]})) = F_{t_d}(\zeta) - F_{t_d}(\mathbf{1}_{[0,1]}) + A(\mathbf{1}_{[0,1]}) - A(\zeta)$, and $A(\zeta) \leq A(\mathbf{1}_{[0,1]})$ by Corollary 7.6; hence $I(\zeta) - I(\mathbf{1}_{[0,1]}) \geq \frac{9}{1024 t_d} \|\zeta - \mathbf{1}_{[0,1]}\|_{L^1}^3 \geq 0$. For $t \geq t_d$,

$$F_t(\zeta) - F_t(\mathbf{1}_{[0,1]}) = F_{t_d}(\zeta) - F_{t_d}(\mathbf{1}_{[0,1]}) + (t - t_d)(I(\zeta) - I(\mathbf{1}_{[0,1]})) \geq \frac{9}{1024} \frac{t}{t_d} \|\zeta - \mathbf{1}_{[0,1]}\|_{L^1}^3,$$

which is (7.28) because $t/t_d = 4c_0 t$. $\square$

The constant in (7.27) is $\frac{27}{32} c_{\text{FS}}$. The factor $\frac{27}{32}$ comes from Step 2 and cannot be improved there: for the one-step directions $\dot{\sigma} = 2 \cdot \mathbf{1}_{(r,1]}$ one has $l_* - r = \frac{1}{2}(1 - l_*)$, and the two sides of the inequality in Step 2 are equal. Without the elementary inequality $(a + b)^3 \geq \frac{27}{4} a^2 b$, the same argument gives the constant $\frac{1}{8} c_{\text{FS}} = \frac{1}{768}$. The theorem is a statement about one explicit variational problem, in which neither e nor θ appears; its link with the functional $\mathcal{P}_\kappa$ at finite e is Theorem 7.3.

7.8 Stability at the crossing

From now on we assume, as announced in Remark 3.12, that θ_d lies in the interior of Θ . Put $\theta_e = (\kappa_{RS} - \kappa_d)/e^2$, so that $\kappa_e(\theta_e) = \kappa_d$. By Proposition 3.8(iii), $\theta_e \rightarrow \theta_d$; hence $\theta_e \in \Theta$ for small e , and Proposition 3.8(iv) applies at $\theta = \theta_e$, where $q^*(\kappa_d) = q_d$. Together,

$$\begin{aligned}\theta_e &= \frac{\kappa_{RS} - \kappa_d}{e^2} = \theta_d + O(eL_e), \\ d_d &= \frac{1 - q_d}{e^2} = 2\phi_0\theta_e + O(e) \longrightarrow d_0 = 2\phi_0\theta_d = \frac{1}{64c_0^2\phi_0^2}.\end{aligned}\tag{7.29}$$

The following estimate controls the mass and the shape of a profile at once. It is uniform in θ near θ_d , which is what the uniform window of Chapter 8 needs, and at $\theta = \theta_e$ it gives the convergence of minimizers at the crossing.

Proposition 7.14. *Fix $0 < m < d_0 < M < \infty$ and put $\Theta(m, M) = \{\theta \in \Theta : m \leq 2\phi_0\theta \leq M\}$. There are $C, e_0 > 0$, depending only on (m, M) and Θ , such that the following holds for $0 < e < e_0$, $\theta \in \Theta(m, M)$ and $\varepsilon \geq 0$. Let $\gamma_\theta^* = \mathbf{1}_{[1-2\phi_0\theta e^2, 1]}$ be the replica-symmetric order parameter of rescaled mass $2\phi_0\theta$. If $\gamma \in \mathcal{U}$ has $\zeta_\gamma \in \mathcal{K}(m, M)$ and*

$$\mathcal{P}_{\kappa_e(\theta)}(\gamma) \leq \mathcal{P}_{\kappa_e(\theta)}(\gamma_\theta^*) + \varepsilon,$$

then

$$(\ell_0(\gamma) - 2\phi_0\theta)^2 + e\|\tilde{\zeta}_\gamma - \mathbf{1}_{[0,1]}\|_{L^1}^3 \leq C(e^2L_e + e(\theta_d - \theta)_+ + \varepsilon).\tag{7.30}$$

In particular the left side is at most $C(e^2L_e + e|\theta - \theta_d| + \varepsilon)$, and every global minimizer γ of $\mathcal{P}_{\kappa_e(\theta)}$ with $\zeta_\gamma \in \mathcal{K}(m, M)$ satisfies (7.30) with $\varepsilon = 0$.

Proof. We divide the proof into five steps.

STEP 1. Constants. Let C_x, e_x be the constants of Theorem 7.3 for the class $\mathcal{K}(m, M)$, let $C_I = \sup\{|I(\zeta)| : \zeta \in \mathcal{K}^1(m/M, M/m)\}$, which is finite by the continuity of I on this compact class (Lemma 7.2), and let $L_g = \sup\{|g'_\theta(l)| : \theta \in \Theta, l \in [m, M]\}$. Put $a_0 = m^2/(2M^4)$ and $C_t = 2\phi_0\sqrt{2\phi_0/\theta_d}$. Take $e_0 \leq \min\{e_x, e^{-1}\}$ with $2\phi_0\theta + e_0^2 < 1$, so that $\gamma_\theta^* \in \mathcal{U}$ and $L_e \geq 1$.

STEP 2. Expansion. Write $\ell^* = 2\phi_0\theta \in [m, M]$, $h = \ell_0 - \ell^*$, $r = \mathbf{1}_{[0,1]}$, $x = \|\tilde{\zeta} - r\|_{L^1}$, $t = 2\phi_0\sqrt{\ell_0}$, and $t^* = 2\phi_0\sqrt{\ell^*}$. The rescaled profile of γ_θ^* is $\mathbf{1}_{[0, \ell^*]} \in \mathcal{K}(m, M)$, and its normalization is r . Applying Theorem 7.3 to γ and to γ_θ^* , subtracting (the terms K_e cancel), and using the hypothesis,

$$\Psi_\theta(\ell_0) - \Psi_\theta(\ell^*) + e[g_\theta(\ell_0) - g_\theta(\ell^*) + \mathcal{P}^{(1)}(\zeta) - \mathcal{P}^{(1)}(\mathbf{1}_{[0, \ell^*]})] \leq 2C_x e^2 L_e + \varepsilon.\tag{7.31}$$

STEP 3. *Mass terms.* Let $x_1 = \ell^*/\ell_0 \in [m/M, M/m]$. The function $u \mapsto u - 1 - \log u$ vanishes with its derivative at $u = 1$ and has second derivative $u^{-2} \geq (m/M)^2$ on $[m/M, M/m]$. By (7.8),

$$\Psi_\theta(\ell_0) - \Psi_\theta(\ell^*) = x_1 - 1 - \log x_1 \geq \frac{m^2}{2M^2}(x_1 - 1)^2 = \frac{m^2}{2M^2} \frac{h^2}{\ell_0^2} \geq a_0 h^2.$$

Also $|g_\theta(\ell_0) - g_\theta(\ell^*)| \leq L_g |h|$.

STEP 4. *Flat terms.* By (7.7) for ζ and for $\mathbf{1}_{[0, \ell^*]}$, and since $F_s = A + sI$ for every s ,

$$\begin{aligned} \mathcal{P}^{(1)}(\zeta) - \mathcal{P}^{(1)}(\mathbf{1}_{[0, \ell^*]}) &= F_{t_d}(\tilde{\zeta}) - F_{t_d}(r) + (t - t^*)I(\tilde{\zeta}) \\ &\quad + (t^* - t_d)[I(\tilde{\zeta}) - I(r)] - \frac{1}{8} \log(\ell_0/\ell^*). \end{aligned}$$

Here $\tilde{\zeta} \in \mathcal{K}^1(m/\ell_0, M/\ell_0) \subset \mathcal{K}^1(m/M, M/m)$ and $r \in \mathcal{K}^1(m/M, M/m)$, so $|I(\tilde{\zeta})|, |I(r)| \leq C_I$. By Theorem 7.13, $F_{t_d}(\tilde{\zeta}) - F_{t_d}(r) \geq \frac{9}{1024}x^3$ and $I(\tilde{\zeta}) - I(r) \geq 0$. By (7.9), $t^* - t_d = 2\phi_0\sqrt{2\phi_0}(\sqrt{\theta} - \sqrt{\theta_d})$. If $\theta \geq \theta_d$, the third term is nonnegative. If $\theta < \theta_d$, it is at least $-2C_I(t_d - t^*) \geq -2C_I C_t(\theta_d - \theta)$, because $\sqrt{\theta_d} - \sqrt{\theta} \leq (\theta_d - \theta)/\sqrt{\theta_d}$. Finally $|t - t^*| = 2\phi_0|h|/(\sqrt{\ell_0} + \sqrt{\ell^*}) \leq \phi_0|h|/\sqrt{m}$ and $|\log(\ell_0/\ell^*)| \leq |h|/m$. Hence

$$\mathcal{P}^{(1)}(\zeta) - \mathcal{P}^{(1)}(\mathbf{1}_{[0, \ell^*]}) \geq \frac{9}{1024}x^3 - \left(\frac{C_I\phi_0}{\sqrt{m}} + \frac{1}{8m}\right)|h| - 2C_I C_t(\theta_d - \theta)_+.$$

STEP 5. *Conclusion.* Inserting the bounds of Steps 3 and 4 into (7.31), with $C_1 = L_g + C_I\phi_0/\sqrt{m} + 1/(8m)$ and $C_2 = 2C_I C_t$,

$$a_0 h^2 + \frac{9e}{1024}x^3 \leq eC_1|h| + eC_2(\theta_d - \theta)_+ + 2C_x e^2 L_e + \varepsilon.$$

Young's inequality gives $eC_1|h| \leq \frac{1}{2}a_0 h^2 + \frac{C_1^2}{2a_0}e^2$, and $e^2 \leq e^2 L_e$. Hence $\frac{1}{2}a_0 h^2 + \frac{9e}{1024}x^3 \leq C_3(e^2 L_e + e(\theta_d - \theta)_+ + \varepsilon)$ with $C_3 = \max\{2C_x + C_1^2/(2a_0), C_2, 1\}$. Dividing by $\min\{\frac{1}{2}a_0, \frac{9}{1024}\}$ gives (7.30). The second statement holds because $(\theta_d - \theta)_+ \leq |\theta - \theta_d|$. A global minimizer satisfies the hypothesis with $\varepsilon = 0$, because $\mathcal{P}_{\kappa_e(\theta)}(\gamma) = \mathcal{P}_*(\kappa_e(\theta)) \leq \mathcal{P}_{\kappa_e(\theta)}(\gamma_\theta^*)$. $\square$

The hypothesis holds whenever $\mathcal{P}_{\kappa_e(\theta)}(\gamma) \leq \mathcal{P}_*(\kappa_e(\theta)) + \varepsilon$. At the crossing, $\kappa = \kappa_d$ and $\theta = \theta_e \in \Theta(m, M)$ for small e , and it also holds whenever $\mathcal{P}_{\kappa_d}(\gamma) \leq \mathcal{P}_{\kappa_d}(\gamma_d) + \varepsilon$, because $q_d = q^*(\kappa_d)$ minimizes $\mathcal{P}_{\kappa_d}^{\text{RS}}$ (Proposition 3.8(i)). The proposition requires $\zeta_\gamma \in \mathcal{K}(m, M)$ as a hypothesis. For global minimizers this follows from Theorem 6.1, and Corollary 8.8 combines the two into closeness to γ_d in L^1 on the scale $1 - q_d$, uniformly for margins within a small multiple of e^2 of κ_d .

Uniqueness at the crossing and full replica symmetry breaking

This chapter proves the local statement behind Theorem 1.2. Theorem 8.13 shows that there is a numerical constant $c > 0$ with the following property for all small $e = \alpha - 2 > 0$. For every margin in $[\kappa_d - ce^2, \kappa_d]$, the replica-symmetric order parameter is the unique global minimizer. For every margin in $(\kappa_d, \kappa_d + ce^2)$, every global minimizer has a Parisi measure supported on one nondegenerate interval, with atoms at both ends and a distribution function that is continuously differentiable on the interior, where its derivative is of order e^{-2} . The case $\kappa = \kappa_d$ is recorded separately in Corollary 8.14.

The proof is Step 5 of §1.3. §8.1 writes the KKT condition through anchored obstacles, which involve only S and λ , and defines the local curvature $\tilde{\Sigma}$. §8.2 treats the obstacle potential of the crossing order parameter $\gamma_d = \mathbf{1}_{[q_d, 1]}$ at κ_d , which has a strict cubic minimum at q_d of depth proportional to e , and the replica-symmetric family near the crossing. §8.3 proves estimates at the natural scale in the crossing window. In §8.4, the stability estimate Proposition 7.14 gives closeness of every global minimizer to γ_d in L^1 , on the scale $1 - q_d \asymp e^2$ (Corollary 8.8), and a comparison of anchored obstacles upgrades this to localization of the supports (Proposition 8.10). §8.5 shows that a KKT point with support in a small window has no gap and satisfies a width estimate (Proposition 8.11), and §8.6 combines the two steps. The relation to earlier work is discussed at the end of the chapter.

Throughout this chapter, $\alpha = 2 + e$, $\kappa_d = \kappa_d(\alpha)$, and $q_d, \delta_d = 1 - q_d$, and $\gamma_d = \mathbf{1}_{[q_d, 1]}$ are as in Proposition 3.8. We write

$$\hat{s}(q) = \frac{1 - q}{\delta_d}, \quad \mathcal{W}_\eta = \{q \in [0, 1] : |q - q_d| \leq \eta e^2\}.$$

Thus $\hat{s}(q_d) = 1$, and $\mathcal{W}_\eta$ is a window of fixed rescaled width around the atom of γ_d . In this chapter we fix the window $\Theta = [\theta_d/2, 2\theta_d]$ of Remark 3.12, which contains θ_d in its interior, so that the results of Chapters 4–7 apply to it, and every constant that depends on Θ is numerical. The distance from θ_d to $\mathbb{R} \setminus \Theta$ is $\Delta_\Theta = \theta_d/2$. Recall from Chapter 7 that $L_e = \log(1/e)$, $\theta_e = (\kappa_{RS} - \kappa_d)/e^2$, $d_d = \delta_d/e^2$, and $d_0 = 2\phi_0\theta_d = 1/(64c_0^2\phi_0^2)$. By (7.29), which rests on Proposition 3.8(iii),(iv), there are numerical constants C_{cr} and C_Θ such that

$$|\theta_e - \theta_d| \leq C_{cr} e L_e, \quad |d_d - 2\phi_0\theta_e| \leq C_\Theta e \quad (8.1)$$

for all small e . We fix $e_{\text{st}} \in (0, 1)$ below the thresholds of Proposition 3.8 and Theorems 3.11 and 6.1 and of (8.1), and such that for $0 < e < e_{\text{st}}$ one has $d_{\text{d}} \in [\frac{1}{2}d_0, 2d_0]$, $|\kappa_{\text{d}}| \leq e$ (by Lemma 3.1(a) and Proposition 3.8(iii)) and $C_{\text{cre}}L_e \leq \frac{1}{2}\Delta_{\Theta}$; in particular $\theta_e \in \Theta$. Every threshold introduced below is at most e_{st} . Quantities of an order parameter γ at a margin κ carry the subscripts γ, κ when both vary, and we write $S_{\text{d}}, D_{\text{d}}, \mathcal{H}_{\text{d}}, \lambda_{\text{d}}$ for those of γ_{d} at κ_{d} . Long proofs are in Chapter D.

8.1 KKT points and the local curvature

Recall from Proposition 2.11(c) that $\gamma \in \mathcal{U}$ is a KKT point of $\mathcal{P}_{\kappa}$ if $\text{supp } \mu_{\gamma} \subset \text{argmin}_{[0, r]} \mathcal{H}_{\gamma}$ for some $r \in (Q, 1)$, and that every global minimizer is a KKT point. The earlier forms of this obstacle condition are discussed after Proposition 2.11; unlike in those settings, here the condition is only necessary (see §1.1). For a KKT point, $0 < q_0 \leq Q < 1$, and every support point t satisfies $D(t) = 0$ and $S(t) \leq 1$. Since $\mathcal{H}_{\gamma}$ is C^2 on $[0, r]$ with $\mathcal{H}'_{\gamma} = -D/2$ and $\mathcal{H}''_{\gamma} = (1 - S)/(2\lambda^2)$, Taylor's formula with integral remainder gives, for all $p, t \in [0, r]$,

$$\mathcal{H}_{\gamma}(t) - \mathcal{H}_{\gamma}(p) = -\frac{D(p)}{2}(t - p) + \mathcal{A}_{\gamma}(p, t), \quad \mathcal{A}_{\gamma}(p, t) = \frac{1}{2} \int_{[p, t]} |t - r'| \frac{1 - S(r')}{\lambda(r')^2} dr', \quad (8.2)$$

where $[p, t]$ denotes the closed interval between p and t . We call $\mathcal{A}_{\gamma}$ the *anchored obstacle*. If p and t are support points of a KKT point, both minimize $\mathcal{H}_{\gamma}$ on $[0, r]$ and $D(p) = 0$, so

$$\mathcal{A}_{\gamma}(p, t) = 0. \quad (8.3)$$

This identity involves only S and λ , and not D . This is what makes it usable below: the natural-scale estimates of §8.3 control S and λ with the required precision, while D , whose derivative is $(S - 1)/\lambda^2$, changes by an amount of order e^{-1} when $S - 1$ changes by an amount of order e over a window of width e^2 .

The local structure of the support is governed by the functions $\mathcal{N}(q) = \mathbb{E}v_x(q, X_q)^2$ and $\tilde{B}(q) = \mathbb{E}[v^2(1 - v)](q, X_q)$, the breakpoint ratio $\tilde{x} = \lambda\mathcal{N}/(2\tilde{B})$ of (2.22), and the quadratic form $\Sigma(q; K)$ of (2.23). They are defined for every order parameter, without any stationarity assumption.

Definition 8.1. Let $\gamma \in \mathcal{U}$, $\kappa \in \mathbb{R}$, and $q \in [0, 1)$. The *local curvature* is

$$\begin{aligned} \tilde{\Sigma}(q) &= \lambda(q)^2 \Sigma(q; \tilde{x}(q)/\lambda(q)) \\ &= \mathbb{E} \left[\lambda^2 v_{xx}^2 + 4\tilde{x} \lambda (3v - 1) v_x^2 + 2\tilde{x}^2 v^2 (1 - v)(1 - 3v) \right] (q, X_q). \end{aligned}$$

The flat constants are $A_V = \int_{\mathbb{R}} V'^2$, $B_V = \int_{\mathbb{R}} V^2(1-V)$, $\mathcal{I}_2 = \int_{\mathbb{R}} V''^2$, the flat breakpoint $x_f = A_V/(2B_V)$, and

$$\Sigma_{\text{flat}} = \int_{\mathbb{R}} [V''^2 + 4x_f(3V-1)V'^2 + 2x_f^2 V^2(1-V)(1-3V)] dz, \quad \sigma_f = \frac{\Sigma_{\text{flat}}}{B_V}.$$

The flat breakpoint x_f and the ratio σ_f are the limits of $\tilde{x}$ and $\tilde{\Sigma}/\tilde{B}$ in the crossing window (Lemma 8.6(b)). Their signs are elementary.

Lemma 8.2. *One has $x_f = \frac{1}{2} - c_0/(4B_V) \in (0, \frac{1}{4})$ and*

$$\Sigma_{\text{flat}} = \frac{1 + (3x_f - 2)^2}{5} \mathcal{I}_2 + \frac{3}{4} x_f^2 c_0 > 0.$$

The proof is in §D.1. The breakpoint and the curvature enter through the derivatives of S . By Lemma 2.9(c)(iv),(d), S is Lipschitz on $[0, r]$ for every $r < 1$, with one-sided derivatives

$$S'(q+) = \frac{2\alpha\tilde{B}(q)}{\lambda(q)}(\tilde{x}(q) - \gamma(q)), \quad S'(q-) = \frac{2\alpha\tilde{B}(q)}{\lambda(q)}(\tilde{x}(q) - \gamma(q-)), \quad (8.4)$$

and on an interval where $\gamma \equiv m$ is constant, $S'' = \alpha\Sigma(q; m/\lambda)$. The functions $\mathcal{N}$ and $\tilde{B}$ are Lipschitz on $[0, r]$, and $\tilde{\Sigma}$ is continuous there, by Lemma 2.6(c); since $\tilde{B} > 0$ is continuous, $\tilde{x}$ is Lipschitz on $[0, r]$. The next lemma restates the gap principle Lemma 2.12 in terms of $\tilde{\Sigma}$ and adds the density formula on support intervals. The right and left derivatives in (8.4) differ only through $\gamma(q)$ and $\gamma(q-)$, as in the analogue of [8, Proposition 3] noted after Lemma 2.9; for spherical models the density on support intervals is given explicitly in [8, Theorem 6(ii)].

Lemma 8.3. *Let γ be a KKT point of $\mathcal{P}_\kappa$.*

- (a) *If $p < p'$ are support points with $(p, p') \cap \text{supp } \mu_\gamma = \emptyset$, then there is $x^* \in (p, p')$ with $\tilde{\Sigma}(x^*) \leq 0$.*
- (b) *If $[p, p'] \subset \text{supp } \mu_\gamma$ with $p < p'$, then $D \equiv 0$ and $S \equiv 1$ on $[p, p']$, $\gamma = \tilde{x}$ on $[p, p']$, γ is continuously differentiable on (p, p') , and*

$$\gamma' = \frac{\tilde{\Sigma}}{2\lambda\tilde{B}} \quad \text{on } (p, p').$$

Proof. (a) By Lemma 2.12(a), $\gamma \equiv m$ on $[p, p']$ for some $m \in (0, 1)$, and there is $x^* \in (p, p')$ with $\tilde{x}(x^*) = m$ and $\Sigma(x^*; m/\lambda(x^*)) \leq 0$. Since $m = \tilde{x}(x^*)$, this says $\tilde{\Sigma}(x^*) = \lambda(x^*)^2 \Sigma(x^*; \tilde{x}(x^*)/\lambda(x^*)) \leq 0$.

(b) By Lemma 2.12(b), $D \equiv 0$ and $S \equiv 1$ on $[p, p']$, $\gamma = \tilde{x}$ on $[p, p']$, and γ is continuous on (p, p') . Put $K = \gamma/\lambda$. By Lemma 2.9(c)(iv), $\mathcal{N}$ and $\tilde{B}$ have right and left derivatives at every q , given by

$$\mathcal{N}' = \mathbb{E}[v_{xx}^2 + 2K(3v-1)v_x^2], \quad \tilde{B}' = \mathbb{E}[(1-3v)v_x^2 - Kv^2(1-v)(2-3v)],$$

with $K = \gamma(q)/\lambda(q)$ and $K = \gamma(q-)/\lambda(q)$, respectively. On (p, p') the two coincide, and they are continuous there by Lemma 2.6(c). Thus $\tilde{x}$, and hence γ , is continuously differentiable on (p, p') . Differentiate the identity $\mathcal{N} = 2(\gamma/\lambda)\tilde{B}$, which is $\gamma = \tilde{x}$, and use $(\gamma/\lambda)' = \gamma'/\lambda + K^2$:

$$\mathbb{E}v_{xx}^2 + 2K\mathbb{E}[(3v-1)v_x^2] = 2\left(\frac{\gamma'}{\lambda} + K^2\right)\tilde{B} + 2K\left(\mathbb{E}[(1-3v)v_x^2] - K\mathbb{E}[v^2(1-v)(2-3v)]\right).$$

Collecting terms, this is $\Sigma(q; K) = 2\gamma'\tilde{B}/\lambda$. Since $K = \tilde{x}/\lambda$, it is $\gamma' = \tilde{\Sigma}/(2\lambda\tilde{B})$. $\square$

8.2 The replica-symmetric obstacle at the crossing

Put $G_d(q) = \mathcal{H}_d(q) - \mathcal{H}_d(q_d)$. By Definition 3.2 and Proposition 3.8 and Lemma 3.3(b), $D_d(q_d) = 0$, because q_d is RS-stationary at κ_d , and $S_d(q_d) = \alpha H_{RS} = 1$, which is the crossing. The next proposition is the source of the cubic separation that confines the supports. It concerns the single order parameter γ_d at the single margin κ_d .

Proposition 8.4. *There are $e_o, c_o, C_o > 0$ such that the following holds for $0 < e < e_o$.*

- (a) S_d is strictly increasing on $[0, q_d]$ and strictly decreasing on $[q_d, 1]$; in particular $S_d \leq 1$ on $[0, 1]$. Hence $D_d > 0$ on $[0, q_d)$, $D_d < 0$ on $(q_d, 1]$, and q_d is the unique minimum point of $\mathcal{H}_d$ on $[0, 1]$.
- (b) For every $q \in [0, 1]$,

$$G_d(q) \geq c_o e \min\{|\hat{s}(q) - 1|, \frac{3}{4}\}^3.$$

- (c) If $|\hat{s}(q) - 1| \leq \frac{3}{4}$, then

$$0 \leq 1 - S_d(q) \leq C_o e |\hat{s}(q) - 1|, \quad G_d(q) \leq C_o e |\hat{s}(q) - 1|^3, \\ |\delta_d D_d(q)| \leq C_o e |\hat{s}(q) - 1|^2.$$

The proof, in §D.2, is explicit. Above q_d the law of $(X_q - \kappa_d)/\sqrt{1-q}$ evolves by the transition kernel of the diffusion (7.20), which is explicit and reversible for the weight $\Phi^2 e^{z^2/2}$ (Lemma D.1). Near $\alpha = 2$ the initial density $\beta_d \phi(k_d + \beta_d z)$, with $k_d = \kappa_d/\sqrt{q_d}$ and $\beta_d = \sqrt{\delta_d/q_d}$, is this weight times a strictly decreasing function, and a stochastic-order comparison shows that $S_d = \alpha \mathbb{E}V^2$ decreases. The lower bounds in (b) come from exact positive formulas for $1 - S_d$ on both sides of q_d . They carry the factor $\beta_d \asymp e$, and the potential is cubic because the replicon eigenvalue vanishes at the crossing.

We also need the replica-symmetric family near the crossing. For $q \in (0, 1)$ and a margin κ let $\gamma_q = \mathbf{1}_{[q, 1]}$, $\delta_q = 1 - q$, and $Z^q = (\sqrt{q}G - \kappa)/\sqrt{\delta_q}$ be as in Definition 3.2, let $D_{RS}(q) = D_{RS}(q; \kappa)$ and $S_{RS}(q) = S_{RS}(q; \kappa)$ be the values of D and S of γ_q at its atom, given explicitly in (3.1), and put

$$A_w(q) = \mathbb{E}V'(Z^q)^2.$$

The order parameter γ_q is a KKT point only if $D_{\text{RS}}(q) = 0$. For $\kappa < \kappa_{\text{RS}}$ let $q^*(\kappa)$ be the stationary point of Proposition 3.8(i), and put $\tau(\kappa) = 1 - S_{\text{RS}}(q^*(\kappa))$, the rescaled replicon eigenvalue. By Lemma 3.3(b) and Definition 3.2, $\tau(\kappa) = 4(1 - q^*(\kappa))^2 \Lambda_{\text{R}}(\kappa)$, so Proposition 3.8(ii) gives

$$\tau(\kappa) > 0 \quad (\kappa_{\text{RS}} - c_2 \leq \kappa < \kappa_{\text{d}}), \quad \tau(\kappa_{\text{d}}) = 0, \quad \tau(\kappa) < 0 \quad (\kappa_{\text{d}} < \kappa < \kappa_{\text{RS}}). \quad (8.5)$$

Lemma 8.5. *There are $\eta_1 \in (0, \frac{1}{4}d_0]$, $\eta'_1 \in (0, \eta_1]$, $e_{\text{r}} > 0$, and $c_{\text{r}}, C_{\text{r}} > 0$ such that the following holds for $0 < e < e_{\text{r}}$.*

- (a) *If $|\kappa - \kappa_{\text{d}}| \leq \eta_1 e^2$, then $\partial_q D_{\text{RS}} \leq -c_{\text{r}}/\delta_q^2$ on $\mathcal{W}_{\eta_1}$. If $|\kappa - \kappa_{\text{d}}| \leq \eta'_1 e^2$, then $\kappa \in [\kappa_{\text{RS}} - c_2, \kappa_{\text{RS}})$, $q^*(\kappa) \in \mathcal{W}_{\eta_1}$, and $|q^*(\kappa) - q_{\text{d}}| \leq C_{\text{r}}|\kappa - \kappa_{\text{d}}|$.*
- (b) *If $|\kappa - \kappa_{\text{d}}| \leq \eta_1 e^2$, then $|\partial_q S_{\text{RS}}| \leq C_{\text{r}}/e$ and $|\partial_{\kappa} S_{\text{RS}}| \leq C_{\text{r}}$ on $\mathcal{W}_{\eta_1}$. Consequently $|\tau(\kappa)| \leq C_{\text{r}}|\kappa - \kappa_{\text{d}}|/e$ when $|\kappa - \kappa_{\text{d}}| \leq \eta'_1 e^2$.*
- (c) *If $|\kappa - \kappa_{\text{d}}| \leq \eta_1 e^2$, then $c_{\text{r}}e \leq A_w(q) \leq C_{\text{r}}e$ on $\mathcal{W}_{\eta_1}$.*

The proof is in §D.2; it uses explicit Gaussian formulas for the derivatives of S_{RS} and D_{RS} .

8.3 Natural-scale estimates in the crossing window

Near the crossing, $S - 1$, $\tilde{\Sigma}$, and the relevant differences of $\mathcal{H}$ are all of order e , the height of the Gaussian bottom density on the scale $\sqrt{\delta_{\text{d}}} \asymp e$. Estimates at a fixed α bound errors in absolute terms and lose this factor. The following lemma keeps it. Both parts come from one compactness argument in the rescaled variables of Definition 4.1: after the dilation of Lemma C.2, the Parisi solution of every order parameter whose support has rescaled depth in a fixed compact interval is an instance of the finite-cutoff problem that depends neither on e nor on the margin.

Lemma 8.6. (a) *Fix $0 < u_- < 1 < u_+$. For every $\omega > 0$ there are $\varepsilon, e_{\text{c}} > 0$ such that for $0 < e < e_{\text{c}}$ and every $\gamma \in \mathcal{U}$ with $\hat{s}(\text{supp } \mu_{\gamma}) \subset [u_-, u_+]$ and $\|\gamma - \gamma_{\text{d}}\|_{L^1} \leq \varepsilon \delta_{\text{d}}$,*

$$\sup\{|S_{\gamma, \kappa_{\text{d}}}(q) - S_{\text{d}}(q)| : \hat{s}(q) \in [u_-, u_+]\} \leq \omega e.$$

(b) *As $\eta \rightarrow 0$ and $e \rightarrow 0$,*

$$\tilde{x}(q) \longrightarrow x_{\text{f}}, \quad \frac{\tilde{\Sigma}(q)}{\tilde{B}(q)} \longrightarrow \sigma_{\text{f}},$$

uniformly over margins with $|\kappa - \kappa_{\text{d}}| \leq \eta e^2$, order parameters $\gamma \in \mathcal{U}$ with $\text{supp } \mu_{\gamma} \subset \mathcal{W}_{\eta}$, and levels $q \in [q_{\text{d}} - \eta e^2, Q]$. Consequently there are $\eta_2, e_{\text{n}} > 0$ such that, in this setting with $\eta = \eta_2$ and $0 < e < e_{\text{n}}$,

$$\frac{x_{\text{f}}}{2} < \tilde{x}(q) < \frac{3}{4}, \quad \frac{\sigma_{\text{f}}}{2} \leq \frac{\tilde{\Sigma}(q)}{\tilde{B}(q)} \leq 2\sigma_{\text{f}}. \quad (8.6)$$

The proof is in §D.3. The consequence (8.6) follows from the first assertion of (b) because $x_f \in (0, \frac{1}{4})$ and $\sigma_f > 0$ (Lemma 8.2). Part (a) is used only at the margin κ_d ; the dependence on the margin is controlled separately by Lemma 8.9 below.

The next lemma compares the bottom level of an order parameter supported in a window with the replica-symmetric order parameter of the same mean. For $\gamma \in \mathcal{U}$ with bottom q_0 write

$$m = 1 - \lambda(q_0) = \int t \, d\mu_\gamma(t), \quad \delta_m = 1 - m, \quad \tilde{\ell} = \frac{Q - q_0}{e^2},$$

and let $\gamma_m = \mathbf{1}_{[m,1]}$ be the replica-symmetric order parameter with the same mean; it has the same value $\lambda(q_0)$.

Lemma 8.7. *There are $\eta_3, e_{bt}, C > 0$ such that for $0 < e < e_{bt}$, $|\kappa - \kappa_d| \leq \eta_3 e^2$, and every $\gamma \in \mathcal{U}$ with $\text{supp } \mu_\gamma \subset \mathcal{W}_{\eta_3}$:*

$$\left| S_\gamma(q_0) - S_{RS}(m) + \alpha(m - q_0) \frac{A_w(m)}{\delta_m} \right| \leq C e \tilde{\ell}^2, \quad \sup_{r \in [q_0, m]} |S_{\gamma_m}(r) - S_{RS}(m)| \leq C e \tilde{\ell},$$

$$|D_\gamma(q_0) - D_{\gamma_m}(q_0)| \leq \frac{C \tilde{\ell}^2}{e}, \quad D_{\gamma_m}(q_0) = D_{RS}(m) - \frac{1}{\delta_m^2} \int_{q_0}^m (S_{\gamma_m}(r) - 1) \, dr.$$

Both comparisons are of second order in $\tilde{\ell}$. In the scaled Parisi equation we expand around the Gaussian reference solution with the current value of λ . The explicit first-order correction depends on the order parameter only through its mean, so it is the same for γ and γ_m , and a weighted parabolic comparison controls the second-order remainder. The proof is in §D.4.

8.4 Support localization in a window of margins

We first show that every global minimizer at a margin within a small multiple of e^2 of κ_d is close to γ_d in L^1 .

Corollary 8.8. *For every $\varepsilon > 0$ there are $\rho_1(\varepsilon) \in (0, \frac{1}{2}\Delta_\Theta]$ and $e_1(\varepsilon) > 0$ such that for $0 < e < e_1$ and every κ with $|\kappa - \kappa_d| \leq \rho_1 e^2$, every global minimizer γ of $\mathcal{P}_\kappa$ satisfies $\|\gamma - \gamma_d\|_{L^1} \leq \varepsilon \delta_d$.*

Proof. Let d_1, S_*, e_* be the constants of Theorem 6.1, and put $\underline{d} = \min\{d_1, \frac{1}{2}d_0\}$ and $\bar{d} = \max\{S_*, 2d_0\}$. Then $0 < \underline{d} < d_0 < \bar{d}$ and $\mathcal{K}(d_1, S_*) \subset \mathcal{K}(\underline{d}, \bar{d})$. Let C_B, e_B be the constants C, e_0 of Proposition 7.14 for $(m, M) = (\underline{d}, \bar{d})$.

Let $|\kappa - \kappa_d| \leq \rho_1 e^2$, with $\rho_1 \leq \frac{1}{2}\Delta_\Theta = \frac{1}{4}\theta_d$ to be chosen, and put $\theta = (\kappa_{RS} - \kappa)/e^2$. Then $\kappa = \kappa_e(\theta)$ and $|\theta - \theta_e| = |\kappa - \kappa_d|/e^2 \leq \rho_1$, so by (8.1)

$$|\theta - \theta_d| \leq \rho_1 + C_{cre} L_e \leq \Delta_\Theta = \frac{1}{2}\theta_d$$

for $e < e_{\text{st}}$. Hence $\theta \in \Theta$ and $2\phi_0\theta \in [\frac{1}{2}d_0, \frac{3}{2}d_0] \subset [\underline{d}, \bar{d}]$.

Let $e < \min\{e_*, e_B\}$ and let γ be a global minimizer of $\mathcal{P}_\kappa$. By Theorem 6.1, its rescaled profile ζ lies in $\mathcal{K}(d_1, S_*) \subset \mathcal{K}(\underline{d}, \bar{d})$. Put $h = \ell_0 - 2\phi_0\theta$ and $x = \|\tilde{\zeta} - \mathbf{1}_{[0,1]}\|_{L^1}$, where ℓ_0 and $\tilde{\zeta}$ are the mass and the normalized profile of γ . Since $\theta \in \Theta$ and $\underline{d} \leq 2\phi_0\theta \leq \bar{d}$, Proposition 7.14 with $\varepsilon = 0$ gives $h^2 + e x^3 \leq C_B(e^2 L_e + e|\theta - \theta_d|)$, so

$$\begin{aligned} x &\leq C_B^{1/3}(\rho_1 + (1 + C_{\text{cr}})eL_e)^{1/3}, \\ |h| &\leq C_B^{1/2}(e\rho_1 + (1 + C_{\text{cr}})e^2 L_e)^{1/2}. \end{aligned}$$

The change of variables $q = 1 - e^2 s$ gives $\|\gamma - \gamma_d\|_{L^1} = e^2 \|\zeta - \mathbf{1}_{[0,d_d]}\|_{L^1}$, and the substitution $s = \ell_0 s'$ gives $\|\zeta - \mathbf{1}_{[0,\ell_0]}\|_{L^1} = \ell_0 x$. Since $\ell_0 \leq S_* \leq \bar{d}$, (8.1) gives

$$\begin{aligned} \|\zeta - \mathbf{1}_{[0,d_d]}\|_{L^1} &\leq \ell_0 x + |\ell_0 - d_d| \leq \bar{d}x + |h| + 2\phi_0|\theta - \theta_e| + |2\phi_0\theta_e - d_d| \\ &\leq \bar{d}x + |h| + 2\phi_0\rho_1 + C_\Theta e. \end{aligned}$$

Since $\delta_d = e^2 d_d$ and $d_d \geq \frac{1}{2}d_0$,

$$\frac{\|\gamma - \gamma_d\|_{L^1}}{\delta_d} \leq \frac{2}{d_0} \left[\bar{d} C_B^{1/3}(\rho_1 + (1 + C_{\text{cr}})eL_e)^{1/3} + 2\phi_0\rho_1 + |h| + C_\Theta e \right].$$

Choose ρ_1 so that $\frac{2}{d_0} [\bar{d} C_B^{1/3}(2\rho_1)^{1/3} + 2\phi_0\rho_1] \leq \frac{1}{2}\varepsilon$. Then choose $e_1 \leq \min\{e_*, e_B, e_{\text{st}}\}$ so small that for $e < e_1$ one has $(1 + C_{\text{cr}})eL_e \leq \rho_1$ and, by the bound on $|h|$, $\frac{2}{d_0} [|h| + C_\Theta e] \leq \frac{1}{2}\varepsilon$. This proves the corollary. $\square$

The localization compares S for a minimizer at the margin κ with S for the crossing order parameter at the margin κ_d . Lemma 8.6(a) compares different order parameters at the same margin κ_d . The next lemma shows that changing only the margin moves S by $O(|\kappa - \kappa_d|)$, which is $O(e^2)$ in the window and therefore negligible against the tolerance ωe of Lemma 8.6(a).

Lemma 8.9. *Let $\alpha \leq 3$, let $\gamma \in \mathcal{U}$ have bottom $q_0 \geq \frac{1}{2}$, and let $\kappa, \kappa' \in \mathbb{R}$. Then for every $q \in [\frac{1}{2}, 1)$,*

$$|S_{\gamma,\kappa}(q) - S_{\gamma,\kappa'}(q)| \leq \frac{\alpha|\kappa - \kappa'|}{\sqrt{\pi}} \leq 2|\kappa - \kappa'|.$$

Proof. By Lemma 2.4(g), $u_{\gamma,\kappa}(q, x) = u_{\gamma,0}(q, x - \kappa)$, so $b_{\gamma,\kappa}(q, x) = b_{\gamma,0}(q, x - \kappa)$ and $v_{\gamma,\kappa}(q, x) = v_{\gamma,0}(q, x - \kappa)$. Hence $Y = X^{\gamma,\kappa} - \kappa$ solves $dY_q = \gamma(q)b_{\gamma,0}(q, Y_q)dq + dB_q$ with $Y_0 = -\kappa$, and

$$S_{\gamma,\kappa}(q) = \alpha \mathbb{E}g(Y_q), \quad g = v_{\gamma,0}(q, \cdot)^2 \in [0, 1],$$

by Lemma 2.4(c). The function g does not depend on κ . Put $q' = \min\{q, q_0\} \geq \frac{1}{2}$. Since $\gamma = 0$ on $[0, q_0)$, the drift vanishes on $[0, q')$, and $Y_{q'} = -\kappa + B_{q'}$ has the law $N(-\kappa, q')$.

By the Markov property of the diffusion, $\mathbb{E}g(Y_q) = \mathbb{E}\psi(Y_{q'})$, where $\psi(y) \in [0, 1]$ is the expectation of g at level q for the diffusion of $(\gamma, 0)$ started at level q' from y ; ψ does not depend on κ . Therefore

$$\begin{aligned} |S_{\gamma, \kappa}(q) - S_{\gamma, \kappa'}(q)| &= \alpha |\mathbb{E}\psi(-\kappa + \sqrt{q'}G) - \mathbb{E}\psi(-\kappa' + \sqrt{q'}G)| \\ &\leq \alpha d_{\text{TV}}(N(-\kappa, q'), N(-\kappa', q')). \end{aligned}$$

For two Gaussian laws with the same variance σ^2 and means at distance Δ , the total variation distance is $2\Phi(\Delta/(2\sigma)) - 1 \leq \Delta/(\sigma\sqrt{2\pi})$. With $\sigma^2 = q' \geq \frac{1}{2}$ this proves the lemma. $\square$

The next proposition localizes the supports of all global minimizers at margins within a small multiple of e^2 of κ_d . The anchored obstacles are always compared with that of the crossing order parameter γ_d at the margin κ_d , so the obstacle estimate Proposition 8.4, which concerns κ_d only, suffices.

Proposition 8.10. *For every $\eta \in (0, \frac{3}{4}]$ there are $\rho_l(\eta) > 0$ and $e_l(\eta) > 0$ such that for $0 < e < e_l$ and every κ with $|\kappa - \kappa_d| \leq \rho_l e^2$, every global minimizer γ of $\mathcal{P}_\kappa$ satisfies*

$$\text{supp } \mu_\gamma \subset \{q : |\hat{s}(q) - 1| < \eta\}.$$

Proof. We divide the proof into four steps.

STEP 1. Choice of constants. Fix $\eta \in (0, \frac{3}{4}]$. Let d_1, S_*, e_* be the constants of Theorem 6.1, and put

$$u_- = \min\left\{\frac{d_1}{2d_0}, \frac{1}{2}\right\}, \quad u_+ = \max\left\{\frac{2S_*}{d_0}, 2\right\}.$$

Let c_0, C_0, e_0 be the constants of Proposition 8.4. Choose $\omega > 0$ with

$$\frac{8\omega(u_+ - u_-)^2}{u_-^2} \leq \frac{c_0\eta^3}{8}. \quad (8.7)$$

Let ε_0, e_c be given by Lemma 8.6(a) for the interval $[u_-, u_+]$ and the tolerance $\frac{1}{2}\omega$. Decrease ε_0 so that $\varepsilon_0 \leq \frac{1}{2}u_-$, $(1 + \varepsilon_0/u_-)^{-2} \geq \frac{1}{2}$, and $C_0(\varepsilon_0^3 + (u_+ - u_-)\varepsilon_0^2) \leq \frac{1}{4}c_0\eta^3$. Let $\rho_1 = \rho_1(\varepsilon_0)$ and $e_1 = e_1(\varepsilon_0)$ be given by Corollary 8.8, and put $\rho_l = \min\{\rho_1, 1\}$. Finally let $e_l \leq \min\{e_*, e_0, e_c, e_1, \frac{1}{4}\omega\}$ be so small that $u_+\delta_d \leq \frac{1}{2}$ and $S_*e^2 \leq \frac{1}{2}$ for $e < e_l$.

STEP 2. The minimizer. Let $0 < e < e_l$, let $|\kappa - \kappa_d| \leq \rho_l e^2$, and let γ be a global minimizer of $\mathcal{P}_\kappa$. As in the proof of Corollary 8.8, $\kappa = \kappa_e(\theta)$ with $\theta \in \Theta$. By Theorem 6.1, every support point q of μ_γ satisfies $(1 - q)/e^2 \in [d_1, S_*]$, hence

$$\hat{s}(q) = \frac{1 - q}{e^2 d_d} \in \left[\frac{d_1}{2d_0}, \frac{2S_*}{d_0}\right] \subset [u_-, u_+],$$

and $q_0 \geq 1 - S_* e^2 \geq \frac{1}{2}$. By Corollary 8.8, $\|\gamma - \gamma_d\|_{L^1} \leq \varepsilon_0 \delta_d$. Every level q with $\hat{s}(q) \in [u_-, u_+]$ satisfies $q \geq 1 - u_+ \delta_d \geq \frac{1}{2}$. Hence Lemma 8.6(a), applied at the margin κ_d , and Lemma 8.9 give

$$\sup_{\hat{s}(q) \in [u_-, u_+]} |S_{\gamma, \kappa}(q) - S_d(q)| \leq \frac{1}{2} \omega e + 2|\kappa - \kappa_d| \leq \frac{1}{2} \omega e + 2e^2 \leq \omega e, \quad (8.8)$$

because $\rho_l \leq 1$ and $e \leq \frac{1}{4} \omega$.

STEP 3. An anchor. Since γ and $\gamma_d = \mathbf{1}_{[q_d, 1]}$ are distribution functions of probability measures on $[0, 1]$, Fubini's theorem gives

$$\|\gamma - \gamma_d\|_{L^1} = \int_0^{q_d} \gamma + \int_{q_d}^1 (1 - \gamma) = \int |t - q_d| d\mu_\gamma(t) \leq \varepsilon_0 \delta_d.$$

Hence there is a support point p with $|\hat{s}(p) - 1| \leq \varepsilon_0$. Also $|\lambda_\gamma(q) - \lambda_d(q)| \leq \|\gamma - \gamma_d\|_{L^1}$. In the normalized variable $\hat{\lambda} = \lambda / \delta_d$ this reads $|\hat{\lambda}_\gamma - \hat{\lambda}_d| \leq \varepsilon_0$, where $\hat{\lambda}_d(q) = \min\{\hat{s}(q), 1\} \geq u_-$ when $\hat{s}(q) \in [u_-, u_+]$. Consequently, for such q ,

$$\hat{\lambda}_\gamma^{-2} \leq \frac{4}{u_-^2}, \quad \hat{\lambda}_\gamma^{-2} \geq \left(1 + \frac{\varepsilon_0}{u_-}\right)^{-2} \hat{\lambda}_d^{-2} \geq \frac{1}{2} \hat{\lambda}_d^{-2}.$$

STEP 4. Comparison of anchored obstacles. Suppose that t is a support point with $|\hat{s}(t) - 1| \geq \eta$. Since γ is a KKT point of $\mathcal{P}_\kappa$ and p, t are support points, (8.3) gives $\mathcal{A}_\gamma(p, t) = 0$, where $\mathcal{A}_\gamma$ is computed at the margin κ . For the crossing order parameter at the margin κ_d , (8.2) gives

$$\mathcal{A}_d(p, t) = G_d(t) - G_d(p) + \frac{D_d(p)}{2}(t - p).$$

The linear term is present because p is a stationary point of γ , not of γ_d . By Proposition 8.4(b),(c), since $|\hat{s}(p) - 1| \leq \varepsilon_0 \leq \frac{3}{4}$ and $|t - p| \leq \delta_d(u_+ - u_-)$,

$$\mathcal{A}_d(p, t) \geq c_o e \eta^3 - C_o e \varepsilon_0^3 - \frac{1}{2} C_o e \varepsilon_0^2 (u_+ - u_-) \geq \frac{3}{4} c_o e \eta^3.$$

Change variables $r' = 1 - \delta_d \hat{s}'$ in both anchored obstacles. Each becomes

$$\mathcal{A}(p, t) = \frac{1}{2} \int |\hat{s}(t) - \hat{s}'| (1 - S(\hat{s}')) \hat{\lambda}(\hat{s}')^{-2} d\hat{s}'$$

over the interval between $\hat{s}(p)$ and $\hat{s}(t)$, which lies in $[u_-, u_+]$ by Step 2. On this interval $1 - S_d \geq 0$ by Proposition 8.4(a), and $1 - S_{\gamma, \kappa} \geq 1 - S_d - \omega e$ by (8.8). With Step 3,

$$(1 - S_{\gamma, \kappa}) \hat{\lambda}_\gamma^{-2} \geq \frac{1}{2} (1 - S_d) \hat{\lambda}_d^{-2} - \frac{4\omega e}{u_-^2}.$$

Integrate against $\frac{1}{2} |\hat{s}(t) - \hat{s}'| d\hat{s}'$ over an interval of length at most $u_+ - u_-$. By (8.7),

$$0 = \mathcal{A}_\gamma(p, t) \geq \frac{1}{2} \mathcal{A}_d(p, t) - \frac{2\omega e}{u_-^2} (u_+ - u_-)^2 \geq \frac{3}{8} c_o e \eta^3 - \frac{1}{32} c_o e \eta^3 > 0.$$

This is a contradiction. Therefore every support point satisfies $|\hat{s} - 1| < \eta$. This proves the proposition. $\square$

The constants are chosen in the order $\eta, \omega, (\varepsilon_0, e_c), (\rho_1, e_1), (\rho_l, e_l)$. The margin enters only through the KKT identity $\mathcal{A}_\gamma(p, t) = 0$ at κ and through the error term $2|\kappa - \kappa_d|$ in (8.8). In particular no estimate of the replica-symmetric obstacle at margins other than κ_d is needed.

8.5 Window rigidity

We now analyze KKT points whose support lies in a window $\mathcal{W}_\eta$ of small rescaled width. Let η_1, η'_1, η_2 , and η_3 be the window sizes of Lemmas 8.5–8.7; recall that $\eta_1 \leq \frac{1}{4}d_0$.

Proposition 8.11. *There are $\eta_0 \in (0, \min\{\eta'_1, \eta_2, \eta_3, \frac{1}{4}d_0, \frac{1}{2}\})$, $e_{\text{win}} > 0$, and $c', C' > 0$ such that for $0 < e < e_{\text{win}}$ and $|\kappa - \kappa_d| \leq \eta_0 e^2$ the following holds for every KKT point γ of $\mathcal{P}_\kappa$ with $\text{supp } \mu_\gamma \subset \mathcal{W}_{\eta_0}$.*

- (a) $\text{supp } \mu_\gamma = [q_0, Q]$ is an interval, possibly a single point.
- (b) If $q_0 < Q$, then $D \equiv 0$ and $S \equiv 1$ on $[q_0, Q]$; $\gamma = \tilde{x}$ on $[q_0, Q]$; γ is continuously differentiable on (q_0, Q) with $\gamma' = \tilde{\Sigma}/(2\lambda\tilde{B}) > 0$; and $\mu_\gamma(\{q_0\}) = \tilde{x}(q_0) > \frac{1}{2}x_f$, $\mu_\gamma(\{Q\}) = 1 - \tilde{x}(Q) > \frac{1}{4}$. Moreover, with $\tilde{\ell} = (Q - q_0)/e^2$ and m the mean of μ_γ ,

$$c'e\tilde{\ell} \leq -\tau(\kappa), \quad |m - q^*(\kappa)| \leq C'e^3\tilde{\ell}. \quad (8.9)$$

- (c) If $q_0 = Q$, then $q_0 = q^*(\kappa)$ and $\tau(\kappa) \geq 0$.
- (d) If $\kappa \leq \kappa_d$, then $\gamma = \mathbf{1}_{[q^*(\kappa), 1]}$. If $\kappa > \kappa_d$, then $q_0 < Q$.

Proof. Let c_r, C_r be the constants of Lemma 8.5 and C_{bt} the constant of Lemma 8.7. Take $\eta_0 \leq \min\{\eta'_1, \eta_2, \eta_3, \frac{1}{4}d_0, \frac{1}{2}\}$, to be decreased once at the end of the proof of (b), and $e_{\text{win}} \leq \min\{e_r, e_n, e_{\text{bt}}\}$. Let $|\kappa - \kappa_d| \leq \eta_0 e^2$. By Lemma 8.5(a), $\kappa \in [\kappa_{\text{RS}} - c_2, \kappa_{\text{RS}})$, and $q^*(\kappa)$ exists and lies in $\mathcal{W}_{\eta_1}$.

(a) Let $p < p'$ be support points with $(p, p') \cap \text{supp } \mu_\gamma = \emptyset$. By Lemma 8.3(a) there is $x^* \in (p, p')$ with $\tilde{\Sigma}(x^*) \leq 0$. But $x^* \in [q_d - \eta_0 e^2, Q]$, so (8.6) gives $\tilde{\Sigma}(x^*) \geq \frac{1}{2}\sigma_f \tilde{B}(x^*) > 0$. Hence the support has no gap. It is a closed set that contains every point between its minimum and its maximum, that is, $[q_0, Q]$.

(b) The structure statements are Lemma 8.3(b) with $p = q_0$ and $p' = Q$, together with $\tilde{\Sigma} > 0$ and $\frac{1}{2}x_f < \tilde{x} < \frac{3}{4}$ from (8.6). Since $\gamma = 0$ on $[0, q_0)$, $\mu_\gamma(\{q_0\}) = \gamma(q_0) = \tilde{x}(q_0)$; since $\tilde{x}$ is continuous, $\mu_\gamma(\{Q\}) = 1 - \gamma(Q-) = 1 - \tilde{x}(Q)$.

For the width estimate we record three bounds and then proceed in three steps. First, $m = 1 - \lambda(q_0) = q_0 + \int_{q_0}^Q (1 - \gamma)$ and $\gamma < \frac{3}{4}$ on $[q_0, Q)$, so

$$\frac{1}{4}e^2\tilde{\ell} \leq m - q_0 \leq e^2\tilde{\ell}.$$

Second, $m \in [q_0, Q] \subset \mathcal{W}_{\eta_0}$ and $\eta_0 \leq \frac{1}{4}d_0$, so $\delta_m \in [\frac{1}{4}d_0e^2, 3d_0e^2]$ and $m \in \mathcal{W}_{\eta_1}$. Third, $\tilde{\ell} \leq 2\eta_0 \leq 1$.

STEP 1. *The S-sector.* Since q_0 is a support point, $S_\gamma(q_0) = 1$. By Lemma 8.7, $|S_{\text{RS}}(m) - 1 - \alpha(m - q_0)A_w(m)/\delta_m| \leq C_{\text{bt}}e\tilde{\ell}^2$. By Lemma 8.5(c) and the three bounds, with $2 \leq \alpha \leq 3$,

$$\frac{c_r}{6d_0}e\tilde{\ell} \leq \alpha(m - q_0)\frac{A_w(m)}{\delta_m} \leq \frac{12C_r}{d_0}e\tilde{\ell}.$$

Hence, with $c_1 = c_r/(6d_0)$ and $C_1 = 12C_r/d_0 + C_{\text{bt}}$,

$$S_{\text{RS}}(m) - 1 \geq e(c_1\tilde{\ell} - C_{\text{bt}}\tilde{\ell}^2), \quad |S_{\text{RS}}(m) - 1| \leq C_1e\tilde{\ell}. \quad (8.10)$$

STEP 2. *The mean.* Since $D_\gamma(q_0) = 0$, Lemma 8.7 gives $|D_{\gamma_m}(q_0)| \leq C_{\text{bt}}\tilde{\ell}^2/e$ and $\sup_{[q_0, m]} |S_{\gamma_m} - 1| \leq (C_{\text{bt}} + C_1)e\tilde{\ell}$. By the formula for $D_{\gamma_m}(q_0)$ in that lemma,

$$|D_{\text{RS}}(m)| \leq \frac{C_{\text{bt}}\tilde{\ell}^2}{e} + \frac{m - q_0}{\delta_m^2}(C_{\text{bt}} + C_1)e\tilde{\ell} \leq \frac{C_2\tilde{\ell}}{e}, \quad C_2 = C_{\text{bt}} + \frac{16(C_{\text{bt}} + C_1)}{d_0^2}.$$

Both m and $q^*(\kappa)$ lie in the interval $\mathcal{W}_{\eta_1}$, on which $\partial_q D_{\text{RS}} \leq -c_r/\delta_q^2 \leq -c_r/(3d_0e^2)^2$, because $\delta_q \leq \delta_d + \eta_1e^2 \leq 3d_0e^2$ there; and $D_{\text{RS}}(q^*(\kappa)) = 0$. By the mean value theorem,

$$|m - q^*(\kappa)| \leq \frac{9d_0^2e^4}{c_r}|D_{\text{RS}}(m)| \leq C'e^3\tilde{\ell}, \quad C' = \frac{9d_0^2C_2}{c_r}.$$

STEP 3. *Conclusion.* By Lemma 8.5(b), $|S_{\text{RS}}(m) - S_{\text{RS}}(q^*(\kappa))| \leq (C_r/e)|m - q^*(\kappa)| \leq C_rC'e^2\tilde{\ell}$. With (8.10),

$$-\tau(\kappa) = S_{\text{RS}}(q^*(\kappa)) - 1 \geq e\tilde{\ell}(c_1 - C_{\text{bt}}\tilde{\ell} - C_rC'e).$$

The constants $c_1, C_{\text{bt}}, C_r, C'$ do not depend on η_0 . Now decrease η_0 so that $2C_{\text{bt}}\eta_0 \leq \frac{1}{4}c_1$, and e_{win} so that $C_rC'e_{\text{win}} \leq \frac{1}{4}c_1$. Since $\tilde{\ell} \leq 2\eta_0$, this gives (8.9) with $c' = \frac{1}{2}c_1$.

(c) If $\text{supp } \mu_\gamma = \{q_0\}$, then $\gamma = \mathbf{1}_{[q_0, 1]}$, and $D(q_0) = 0$ says $D_{\text{RS}}(q_0) = 0$. Thus q_0 is RS-stationary at $\kappa \in [\kappa_{\text{RS}} - c_2, \kappa_{\text{RS}})$, and Proposition 3.8(i) gives $q_0 = q^*(\kappa)$. Finally $S(q_0) \leq 1$ says $\tau(\kappa) \geq 0$.

(d) Let $\kappa \leq \kappa_d$. Then $\tau(\kappa) \geq 0$ by (8.5). If $q_0 < Q$, then $\tilde{\ell} > 0$ and (8.9) gives $0 < c'e\tilde{\ell} \leq -\tau(\kappa) \leq 0$, a contradiction. Hence $q_0 = Q$, and (c) gives $\gamma = \mathbf{1}_{[q^*(\kappa), 1]}$. Let $\kappa > \kappa_d$. Then $\tau(\kappa) < 0$ by (8.5), since $\kappa < \kappa_{\text{RS}}$. If $q_0 = Q$, then (c) gives $\tau(\kappa) \geq 0$, a contradiction. Hence $q_0 < Q$. $\square$

The proof of (d) uses the exact bottom conditions $S(q_0) = 1$ and $D(q_0) = 0$, the stationarity equation of the replica-symmetric family, and the sign of the replicon eigenvalue. It does not use an expansion of the value of the functional. The window constants of Lemma 8.6(b) also fix the scale of the continuous part.

Corollary 8.12. *Let e , κ , and γ be as in Proposition 8.11, with $q_0 < Q$. Then*

$$\frac{\sigma_f}{9d_0} \frac{1}{e^2} \leq \gamma'(q) \leq \frac{4\sigma_f}{d_0} \frac{1}{e^2} \quad (q_0 < q < Q),$$

and consequently $\frac{\sigma_f}{9d_0} \tilde{\ell} \leq \mu_\gamma((q_0, Q)) \leq \frac{4\sigma_f}{d_0} \tilde{\ell}$.

Proof. By Proposition 8.11(b), $\gamma' = \tilde{\Sigma}/(2\lambda\tilde{B})$ on (q_0, Q) , and $\frac{1}{2}\sigma_f \leq \tilde{\Sigma}/\tilde{B} \leq 2\sigma_f$ there by (8.6). Since $d_d \in [\frac{1}{2}d_0, 2d_0]$, $\eta_0 \leq \frac{1}{4}d_0$, and $[q_0, Q] \subset \mathcal{W}_{\eta_0}$,

$$\frac{1}{4}d_0e^2 \leq 1 - Q \leq \lambda(q) \leq 1 - q_0 \leq \frac{9}{4}d_0e^2 \quad (q \in [q_0, Q]).$$

These bounds give the bounds on γ' . Integrating γ' over compact subintervals of (q_0, Q) and passing to the endpoints gives $\mu_\gamma((q_0, Q)) = \gamma(Q-) - \gamma(q_0) = \int_{q_0}^Q \gamma'$, which proves the bounds on the mass. $\square$

8.6 A uniform window around the crossing

We can now describe all global minimizers for margins within a fixed multiple of e^2 of κ_d . Recall from Lemma 8.2 that $x_f \in (0, \frac{1}{4})$.

Theorem 8.13. *There are numerical constants $e_0, c > 0$ and $C < \infty$ such that the following holds for $0 < e < e_0$ and $\alpha = 2 + e$.*

- (i) *For every $\kappa \in [\kappa_d - ce^2, \kappa_d]$, the replica-symmetric order parameter $\mathbf{1}_{[q^*(\kappa), 1]}$ is the unique global minimizer of $\mathcal{P}_\kappa$, and $|q^*(\kappa) - q_d| \leq C(\kappa_d - \kappa)$.*
- (ii) *For every $\kappa \in (\kappa_d, \kappa_d + ce^2)$, global minimizers of $\mathcal{P}_\kappa$ exist, and every global minimizer γ satisfies the following. Its support is an interval $[q_0, Q]$ with $0 < q_0 < Q < 1$; $D \equiv 0$ and $S \equiv 1$ on $[q_0, Q]$; $\gamma = \tilde{x}$ on $[q_0, Q]$; γ is continuously differentiable on (q_0, Q) with*

$$\gamma' = \frac{\tilde{\Sigma}}{2\lambda\tilde{B}}, \quad \frac{c}{e^2} \leq \gamma' \leq \frac{C}{e^2} \quad \text{on } (q_0, Q); \quad (8.11)$$

and

$$\mu_\gamma = \tilde{x}(q_0) \delta_{q_0} + \gamma'(q) dq|_{(q_0, Q)} + (1 - \tilde{x}(Q)) \delta_Q, \quad \tilde{x}(q_0) > \frac{1}{2}x_f, \quad 1 - \tilde{x}(Q) > \frac{1}{4}.$$

Moreover,

$$Q - q_0 \leq C(\kappa - \kappa_d), \quad \sup_{t \in \text{supp } \mu_\gamma} |t - q_d| \leq C(\kappa - \kappa_d), \quad \|\gamma - \gamma_d\|_{L^1} \leq C(\kappa - \kappa_d). \quad (8.12)$$

Finally, for every $\varepsilon > 0$ there is $\rho_a(\varepsilon) > 0$ such that if also $e < \rho_a(\varepsilon)$ and $\kappa - \kappa_d < \rho_a(\varepsilon)e^2$, then $|\mu_\gamma(\{q_0\}) - x_f| \leq \varepsilon$ and $|\mu_\gamma(\{Q\}) - (1 - x_f)| \leq \varepsilon$.

Proof. We divide the proof into four steps.

STEP 1. Choice of constants. Let $\eta_0, e_{\text{win}}, c', C'$ be the constants of Proposition 8.11, and let η'_1, c_r, C_r be those of Lemma 8.5. Put $\eta' = \eta_0/(2d_0) \in (0, \frac{1}{8}]$, and let $\rho_l = \rho_l(\eta')$ and $e_l = e_l(\eta')$ be given by Proposition 8.10. Put

$$c_W = \min\{\rho_l, \eta_0, \frac{1}{4}\theta_d\},$$

and let $e_0 \leq \min\{e_{\text{win}}, e_l, e_{\text{st}}\}$. All these constants are numerical.

STEP 2. Localization in the window. Let $0 < e < e_0$ and $|\kappa - \kappa_d| \leq c_W e^2$. By Theorem 3.11(ii), $\kappa_c - \kappa_d \geq \frac{1}{2}\theta_d e^2 > c_W e^2$, so $\kappa < \kappa_c$ and $\mathcal{P}_*(\kappa) > -\infty$ by (2.36), and global minimizers of $\mathcal{P}_\kappa$ exist by Theorem 2.16. Let γ be one. By Proposition 8.10 with η' , every support point q satisfies $|\hat{s}(q) - 1| < \eta'$, that is, $|q - q_d| < \eta' \delta_d \leq 2d_0 \eta' e^2 = \eta_0 e^2$. Hence

$$\text{supp } \mu_\gamma \subset \mathcal{W}_{\eta_0}, \quad |\kappa - \kappa_d| \leq \eta_0 e^2. \quad (8.13)$$

By Proposition 2.11(c), γ is a KKT point of $\mathcal{P}_\kappa$. Thus Proposition 8.11 applies to γ .

STEP 3. Replica symmetry below the crossing. Let $\kappa \in [\kappa_d - c_W e^2, \kappa_d]$. By Step 2 and Proposition 8.11(d), every global minimizer equals $\mathbf{1}_{[q^*(\kappa), 1]}$, and a global minimizer exists. Hence it is the unique global minimizer. Since $|\kappa - \kappa_d| \leq \eta_0 e^2 \leq \eta'_1 e^2$, Lemma 8.5(a) gives $|q^*(\kappa) - q_d| \leq C_r(\kappa_d - \kappa)$. This proves (i).

STEP 4. Full replica symmetry breaking above the crossing. Let $\kappa \in (\kappa_d, \kappa_d + c_W e^2)$ and let γ be a global minimizer. By Proposition 8.11(d), $q_0 < Q$, and by Proposition 2.11(c), $0 < q_0$ and $Q < 1$. By (8.13) and Proposition 8.11(a),(b), the support is $[q_0, Q]$, $D \equiv 0$ and $S \equiv 1$ on $[q_0, Q]$, $\gamma = \tilde{x}$ on $[q_0, Q)$, γ is continuously differentiable on (q_0, Q) with $\gamma' = \tilde{\Sigma}/(2\lambda\tilde{B})$, and

$$\mu_\gamma(\{q_0\}) = \tilde{x}(q_0) > \frac{1}{2}x_f, \quad \mu_\gamma(\{Q\}) = 1 - \tilde{x}(Q) > \frac{1}{4}.$$

Since γ is continuously differentiable on (q_0, Q) , the restriction of μ_γ to (q_0, Q) has the density γ' , and μ_γ has no mass outside $[q_0, Q]$. This proves the decomposition of μ_γ . The bounds on γ' in (8.11) are Corollary 8.12.

For the rates, (8.9) and Lemma 8.5(b) give

$$c'e\tilde{\ell} \leq -\tau(\kappa) \leq \frac{C_r(\kappa - \kappa_d)}{e}, \quad \text{hence} \quad Q - q_0 = e^2\tilde{\ell} \leq \frac{C_r}{c'}(\kappa - \kappa_d).$$

The mean m of μ_γ lies in $[q_0, Q]$. For $t \in [q_0, Q]$, (8.9) and Lemma 8.5(a) give, since $e^3\tilde{\ell} = e(Q - q_0) \leq Q - q_0$,

$$\begin{aligned} |t - q_d| &\leq (Q - q_0) + |m - q^*(\kappa)| + |q^*(\kappa) - q_d| \\ &\leq \frac{(1 + C')C_r}{c'}(\kappa - \kappa_d) + C_r(\kappa - \kappa_d) =: C_4(\kappa - \kappa_d). \end{aligned}$$

Since $\|\gamma - \gamma_d\|_{L^1} = \int |t - q_d| d\mu_\gamma(t)$ (Step 3 of the proof of Proposition 8.10), this proves (8.12) with $C \geq C_4$.

It remains to treat the atoms. By (8.12), $\text{supp } \mu_\gamma \subset \mathcal{W}_\eta$ with $\eta = C_4(\kappa - \kappa_d)/e^2$. Let $\varepsilon > 0$. By Lemma 8.6(b) there are $\eta(\varepsilon) \in (0, \eta_0]$ and $e(\varepsilon) > 0$ such that $|\tilde{x}(q) - x_f| \leq \varepsilon$ whenever $e < e(\varepsilon)$, the margin satisfies $|\kappa - \kappa_d| \leq \eta(\varepsilon)e^2$, the order parameter satisfies $\text{supp } \mu_\gamma \subset \mathcal{W}_{\eta(\varepsilon)}$, and $q \in [q_d - \eta(\varepsilon)e^2, Q]$. Put $\rho_a(\varepsilon) = \min\{\eta(\varepsilon)/\max\{C_4, 1\}, e(\varepsilon)\}$. If $e < \rho_a(\varepsilon)$ and $\kappa - \kappa_d < \rho_a(\varepsilon)e^2$, then $\eta \leq \eta(\varepsilon)$ and $|\kappa - \kappa_d| \leq \eta(\varepsilon)e^2$, so these conditions hold, and $q_0, Q \in [q_d - \eta(\varepsilon)e^2, Q]$. Hence $|\mu_\gamma(\{q_0\}) - x_f| = |\tilde{x}(q_0) - x_f| \leq \varepsilon$ and $|\mu_\gamma(\{Q\}) - (1 - x_f)| = |\tilde{x}(Q) - x_f| \leq \varepsilon$. This proves (ii) with $c = \min\{c_W, \sigma_f/(9d_0)\}$ and $C = \max\{C_r, C_4, 4\sigma_f/d_0\}$. $\square$

The case $\kappa = \kappa_d$ of part (i) is the uniqueness statement of Theorem 1.2(i) at the crossing itself. We record it separately.

Corollary 8.14. *Let e_0 be as in Theorem 8.13. For $0 < e < e_0$, the crossing order parameter $\gamma_d = \mathbf{1}_{[q_d, 1]}$ is the unique global minimizer of $\mathcal{P}_{\kappa_d(2+e)}$.*

Proof. Apply Theorem 8.13(i) with $\kappa = \kappa_d$; here $q^*(\kappa_d) = q_d$ by Proposition 3.8(ii). $\square$

Part (i) of Theorem 8.13 is the analogue of the replica symmetry of the Parisi measure on the replica-symmetric side of the de Almeida–Thouless line of the SK model, which was proved in various regions in the works cited in Chapter 1. Those proofs use the strict convexity of the Parisi functional, through which the first-order conditions characterize the minimizer. Without convexity (see §1.1), part (i) rests instead on the localization of every global minimizer (Proposition 8.10) and on Proposition 8.11(d).

As noted in §1.3, the argument of this chapter follows the near-critical argument of Zhou for the SK model [134]. There the Parisi measure is close to δ_0 for inverse temperatures slightly above the critical value, and this closeness makes an auxiliary function built from the self-consistency function positive at the right end of every gap that starts near the origin, which the optimality conditions forbid. In [87], the first-order conditions at the endpoints of a hypothetical gap are combined with a sign condition inside the gap ([87, Propositions 4.1 and 5.1] for a gap starting at the origin, and [87, Propositions 9.1 and 9.2] for the other gaps). Here the sign condition concerns the local curvature $\tilde{\Sigma}$ (Lemma 8.3(a)). The three differences from these works are listed in §1.3, and all three concern this chapter. In the SK model at zero external field, to which the third refers, the support starts at 0, which carries no mass, and the only atom is at the top of the support [87, 134].

Proofs of the main theorems

We now assemble the results of Chapters 2–8. In the proofs of Theorems 1.3 and 1.4, Θ is an arbitrary compact interval as in Remark 3.12; by Theorem 3.11(iii) the margins $\kappa_e(\theta)$, $\theta \in \Theta$, lie below $\kappa_c(2 + e)$ by at least $\frac{1}{2}e^2 \min \Theta$ for all small e . The proof of Theorem 1.2 uses the window $\Theta = [\theta_d/2, 2\theta_d]$ fixed in Chapter 8. We first record the order in which the results are proved, to make clear that no step uses a later conclusion.

9.1 The logical order

The argument proceeds in the following order.

- (1) *Foundations* (Chapter 2). The finite-cutoff Parisi equation and its regularity, the generator identities, the law of the diffusion, the first-order (obstacle) conditions and the gap principle, the existence of minimizers below κ_c , and the zero-temperature functional. These hold for each fixed α and involve no asymptotics. (§§2.6 and 2.7 are proved in Part I, Chapter 6, which uses only §§2.1–2.5 and Chapter A of this part.)
- (2) *Replica-symmetric analysis and thresholds* (Chapter 3). The crossing asymptotics of Proposition 3.8, the dual certificate and the ordering of Theorem 3.11, and the cubic lower bound on $\kappa_{RS} - \kappa_c$ of Theorem 3.13. These concern replica-symmetric order parameters and explicit trial functions only.
- (3) *A priori bounds* (Chapter 4). For exact global minimizers at $\kappa_e(\theta)$, $\theta \in \Theta$: the top-gap lower bound Theorem 4.10 and the upper bounds Theorem 4.11.
- (4) *Bounded depth* (Chapter 5). Theorem 5.1 uses (1) and (3).
- (5) *Compactness* (Chapter 6). Theorem 6.1 uses (1), (3), and (4). The top-jump bound Theorem 6.13 is global; it does not presuppose that the minimizer is close to a replica-symmetric order parameter.
- (6) *Flat limit* (Chapter 7). The compact-class expansion Theorem 7.3 holds on $\mathcal{K}(d_1, S)$ for fixed d_1, S , uniformly in $\theta \in \Theta$. Flat uniqueness, Theorem 7.13, is a statement about the limiting functional alone. The stability estimate Proposition 7.14 combines the two, uniformly in θ near θ_d .
- (7) *Local analysis* (Chapter 8), with $\Theta = [\theta_d/2, 2\theta_d]$. The obstacle of the crossing order parameter and the replica-symmetric family use (2). Natural-scale continuity uses

the rescaled theory of (6). Localization (Proposition 8.10) uses (5) and, through Corollary 8.8, (6). Window rigidity (Proposition 8.11) is a statement about KKT points in a small window. Together they give Theorem 8.13.

Every depth, mass, and top-gap bound in (3)–(5) is proved for exact global minimizers. None is claimed for near-minimizers; see Remark 5.5.

9.2 Proof of the threshold theorem

Proof of Theorem 1.1. The expansion of $\kappa_{\text{RS}} - \kappa_{\text{d}}$ is Proposition 3.8(iii). The upper bound $\kappa_{\text{RS}} - \kappa_{\text{c}} \leq Ce^3$ is Theorem 3.11(i), and the lower bound $\kappa_{\text{RS}} - \kappa_{\text{c}} \geq ce^3$ is Theorem 3.13. The expansion of $\kappa_{\text{c}} - \kappa_{\text{d}}$ is Theorem 3.11(iv). Taking e_0 to be the smallest of the three thresholds proves the theorem. $\square$

9.3 Proof of the uniform compactness theorem

Proof of Theorem 1.3. Fix a compact interval $\Theta \subset (0, \infty)$. As recorded in Remark 3.12, the results of Chapters 4–6 hold for every such interval, with constants depending on Θ only through $\theta_{\pm}$; the condition that θ_{d} lies in the interior of Θ is used only from §7.8 on. Let e_0 be the smallest of the thresholds in Theorems 4.11, 5.1, and 6.1, and let γ be a global minimizer of $\mathcal{P}_{\kappa_e(\theta)}$ with $\theta \in \Theta$ and $0 < e < e_0$. Theorem 4.11 gives $d \leq D_{\text{top}}$ and $\ell_0 \leq L$. Theorem 5.1 gives $s_0 \leq S_*$, and $s_0 \leq 1/(4c_w^2)$ if γ is not replica symmetric. Theorem 6.1 gives $\ell_0 \geq \ell_{\min}$, $\gamma(Q-) \leq m_*$, and $d > d_1$. All constants depend only on Θ . $\square$

9.4 Proof of the flat-limit theorem

Proof of Theorem 1.4. The uniqueness statement and the cubic modulus are Theorem 7.13. They hold for every normalized profile with bounded support that equals one on a neighborhood of zero. The expansion is Theorem 7.3, combined with the scaling identity (7.7), which expresses the first-order correction $\mathcal{P}^{(1)}(\zeta)$ as the flat functional of the normalized profile at the parameter $t = 2\phi_0\sqrt{\ell_0}$, plus an explicit function of ℓ_0 . By Remark 3.12, Theorem 7.3 holds for every compact interval $\Theta \subset (0, \infty)$. $\square$

9.5 Proof of the full replica symmetry breaking theorem

Proof of Theorem 1.2. Let e_0, c, C be the numerical constants of Theorem 8.13. Part (i) of Theorem 1.2 is Theorem 8.13(i); at $\kappa = \kappa_{\text{d}}$ it is Corollary 8.14. For part (ii), let $\kappa \in (\kappa_{\text{d}}, \kappa_{\text{d}} + ce^2)$. By Theorem 8.13(ii), global minimizers of $\mathcal{P}_{\kappa}$ exist, and every global minimizer γ has support $[q_0, Q]$ with $0 < q_0 < Q < 1$, which is (a). It is continuously differentiable on (q_0, Q) with $ce^{-2} \leq \gamma' \leq Ce^{-2}$ by (8.11), which is (b); consequently

$\mu_\gamma((q_0, Q))$ is of order $(Q - q_0)/e^2$ (Corollary 8.12). Its atoms are $\mu_\gamma(\{q_0\}) = \tilde{x}(q_0) > \frac{1}{2}x_f$ and $\mu_\gamma(\{Q\}) = 1 - \tilde{x}(Q) > \frac{1}{4}$, and the last assertion of Theorem 8.13(ii) gives their convergence to x_f and $1 - x_f$ as $e \rightarrow 0$ and $(\kappa - \kappa_d)/e^2 \rightarrow 0$, uniformly over global minimizers. The formula $x_f = \frac{1}{2} - c_0/(4B_V) \in (0, \frac{1}{4})$ is Lemma 8.2. This proves (c). Finally (d) is (8.12). $\square$

The constants in Theorems 1.1 and 1.2, and the function ρ_a of Theorem 8.13, are numerical; those in Theorem 1.3 depend only on Θ .

Regularity of the Parisi equation

This appendix proves Lemmas 2.4, 2.6, and 2.8, together with the coefficient-continuity identity of Lemma A.2, which is used in Lemma 2.5, Proposition 2.11 and Lemma 2.19. The argument has three stages. First, for step coefficients, the solution is an explicit composition of Gaussian integrals (§§A.1 and A.2); this gives smoothness, the signs, and polynomial bounds whose constants do not depend on the number of steps. Second, a Feynman–Kac representation of the difference of two solutions gives Lipschitz dependence on the coefficient in L^1 for every spatial derivative. This defines the solution for general coefficients and transfers all properties to it (§§A.3 and A.4); §A.5 proves the control representation. Third, the expansion at $-\infty$ is derived from the martingale ψ (§A.6), using Lemmas 2.9 and 2.10, whose proofs do not use Lemma 2.4(e). Throughout, C and p denote constants that may change from line to line; their dependence is stated in each section.

For the Sherrington–Kirkpatrick datum $\log \cosh$, the construction through step coefficients and the Cole–Hopf transformation, and the passage to general coefficients by a limit in L^1 , are standard [8, 68, 76, 98, 105, 124, 128]. El Alaoui and Sellke carried them out for the present terminal datum $\log \Phi$, which is concave and decreases quadratically at $-\infty$; their results are recalled after Lemma 2.4. This appendix adds three things: the strict signs in Lemma 2.4(c), including $v_x < 0$; the expansion (e) at $-\infty$; and the bound (f) for every derivative with a polynomial weight on the whole line, where [54, Proposition 13] bounds the difference for x in bounded sets. We therefore give the whole argument.

We use three standard facts (see [79]). The first is Itô’s formula for a function that is $C^{1,2}$ on $[s, t] \times \mathbb{R}$, applied to a continuous semimartingale. The second is the Feynman–Kac formula. Let Y solve $dY_q = \beta(q, Y_q) dq + dB_q$ on $[s, t]$ with β Lipschitz in y and of linear growth, so that $\mathbb{E} \sup_{[s, t]} |Y|^p < \infty$ for all p . Let φ be continuous on $[s, t] \times \mathbb{R}$ and $C^{1,2}$ on each $[t_{i-1}, t_i] \times \mathbb{R}$ for a finite partition of $[s, t]$, and suppose that $\varphi_q + \frac{1}{2}\varphi_{yy} + \beta\varphi_y + \mathcal{V}\varphi = \mathcal{F}$ there, with $\mathcal{V}$ bounded and with $\varphi, \varphi_y, \mathcal{F}$ of polynomial growth. Then

$$\varphi(s, y) = \mathbb{E} \left[\varphi(t, Y_t) e^{\int_s^t \mathcal{V}(q, Y_q) dq} - \int_s^t e^{\int_s^q \mathcal{V}(q', Y_{q'}) dq'} \mathcal{F}(q, Y_q) dq \right], \quad Y_s = y. \quad (\text{A.1})$$

This follows from Itô’s formula for $e^{\int_s^q \mathcal{V}(q, Y_q)} \varphi(q, Y_q)$ on each $[t_{i-1}, t_i]$, since the stochastic integral has an integrand of polynomial growth and is therefore a true martingale. The third fact is a Gronwall comparison of two diffusions. Let Y and Y' solve $dY =$

$\beta(q, Y) dq + dB$ and $dY' = \beta'(q, Y') dq + dB$ on $[s, t]$ with $Y_s = Y'_s$, where β' is Lipschitz in y with constant L and $|\beta'(q, y) - \beta(q, y)| \leq \Delta(q)(1 + |y|)$. Then $|Y'_q - Y_q| \leq \int_s^q L|Y' - Y| + \int_s^q \Delta(1 + |Y|)$, and Gronwall's inequality gives

$$\sup_{[s, t]} |Y' - Y| \leq e^{L(t-s)} \left(1 + \sup_{[s, t]} |Y|\right) \int_s^t \Delta(q) dq. \quad (\text{A.2})$$

A.1 One Gaussian step

Lemma A.1. *Let $g \in C^1(\mathbb{R})$ be concave and nondecreasing, with g' Lipschitz with constant K , $g \leq 0$, and $g(x) \geq -C_g(1 + x^2)$. Let $m \geq 0$, $s > 0$, $w = \mathcal{T}_{m,s}g$ and $Y = x + \sqrt{s}G$.*

- (i) *$w \in C^\infty(\mathbb{R})$, $\mathbb{E}g(Y) \leq w(x) \leq 0$, and w is concave and nondecreasing with w' Lipschitz with constant K . If $g' > 0$, then $w' > 0$.*
- (ii) *Let ν_x be the law with density proportional to $e^{mg(y)}\phi_s(y - x)$, and write $\mathbb{E}_\nu$ and Var_ν . Then $w'(x) = \mathbb{E}_\nu[g'(Y)]$ and $-w''(x) = \mathbb{E}_\nu[-g''(Y)] - m \text{Var}_\nu[g'(Y)]$, where g'' is the almost everywhere derivative of g' . Moreover $0 \leq w'(x) \leq C_g(3 + 2s) + K|x|$.*
- (iii) *If $g \in C^k$ and $|g^{(j)}(y)| \leq L(1 + |y|)^p$ for $1 \leq j \leq k$, then $|w^{(j)}(x)| \leq L'(1 + |x|)^{p'}$ for $1 \leq j \leq k$, where L', p' depend only on L, p, k, m, s, K, C_g and remain bounded for m and s in bounded sets. As $s \downarrow 0$, $\partial_x^j \mathcal{T}_{m,s}g \rightarrow g^{(j)}$ locally uniformly for $0 \leq j \leq k$; for $j \leq 1$ this holds without the extra hypothesis.*
- (iv) *The function $(s, x) \mapsto \mathcal{T}_{m,s}g(x)$ is C^∞ on $(0, \infty) \times \mathbb{R}$ and satisfies $\partial_s \mathcal{T}_{m,s}g = \frac{1}{2}(\partial_{xx} + m(\partial_x \cdot)^2) \mathcal{T}_{m,s}g$.*

Proof. Take $m > 0$; the case $m = 0$ is the heat equation and is simpler. Since $0 < e^{mg} \leq 1$, the function $e^{mw(x)} = \int e^{mg(y)}\phi_s(y - x) dy$ is positive and C^∞ in (s, x) for $s > 0$, and it solves the heat equation $\partial_s e^{mw} = \frac{1}{2}\partial_{xx} e^{mw}$; this gives (iv) and smoothness. Jensen's inequality gives $w \geq \mathbb{E}g(Y)$, and $g \leq 0$ gives $w \leq 0$. Concavity of w follows from Prékopa's theorem that the marginal of a log-concave function is log-concave [108, Theorem 6], because $(x, y) \mapsto mg(y) + \log \phi_s(y - x)$ is jointly concave. Integrating by parts in y (the function e^{mg} is C^1 with $e^{mg} \leq 1$, and $mg'e^{mg}$ has linear growth) gives $\partial_x e^{mw} = \mathbb{E}[mg'(Y)e^{mg(Y)}]$, which is the formula for w' . A second integration by parts, valid since $g'e^{mg}$ is locally Lipschitz, gives the formula for w'' . Then $w' \geq 0$, $w' > 0$ if $g' > 0$, and $-w'' \leq \mathbb{E}_\nu[-g''] \leq K$; also $-w'' \geq 0$ by concavity. By concavity and $w \leq 0$,

$$w'(0) \leq w(0) - w(-1) \leq -\mathbb{E}g(-1 + \sqrt{s}G) \leq C_g(2 + s),$$

and $w'(x) \leq w'(0) + K|x|$ for $x < 0$, while $w'(x) \leq w'(0)$ for $x \geq 0$.

For (iii), the density of ν_x is e^{-U} with $U(y) = -mg(y) + (y - x)^2/(2s) + \text{const}$, and U' is Lipschitz with $s^{-1} \leq U'' \leq s^{-1} + mK$ almost everywhere. Let y^* be the minimizer of U . Then $y^* = x + smg'(y^*)$, so $x \leq y^* \leq x + smg'(x)$ by monotonicity of g' , and

$|y^*| \leq C(1 + |x|)$, since $0 \leq g'(x) \leq 2C_g + K|x|$ by concavity of g , as in (ii). The bounds on U'' give $e^{-U(y^*)} \sqrt{2\pi/(s^{-1} + mK)} \leq 1$ and $e^{-U(y)} \leq e^{-U(y^*)} e^{-(y-y^*)^2/(2s)}$, hence $\mathbb{E}_\nu |Y - y^*|^p \leq C_p(1 + mKs)^{1/2} s^{p/2}$ and $\mathbb{E}_\nu |Y|^p \leq C(1 + |x|)^p$. Differentiating $\mathbb{E}_\nu[F(Y)] = \mathbb{E}[F(Y)e^{mg(Y)}]/\mathbb{E}[e^{mg(Y)}]$ in x gives

$$\frac{d}{dx} \mathbb{E}_\nu[F(Y)] = \mathbb{E}_\nu[F'(Y)] + m \text{Cov}_\nu(F(Y), g'(Y)).$$

Starting from $w' = \mathbb{E}_\nu[g']$ and iterating, $w^{(j)}$ is a finite sum of products of ν_x -expectations of monomials in $g'(Y), \dots, g^{(j)}(Y)$, with coefficients that are polynomials in m . The moment bound gives the claimed polynomial bound. As $s \downarrow 0$, ν_x converges to the point mass at x locally uniformly in x , by the bounds on y^* and on $\mathbb{E}_\nu |Y - y^*|^2$, and dominated convergence in the same representations gives the convergence of derivatives; when ν_x is replaced by the point mass, the representation of $w^{(j)}$ becomes $g^{(j)}(x)$. $\square$

A.2 Step order parameters

Let $\gamma \in \mathcal{U}_r$ be a step order parameter with grid $0 = t_0 < \dots < t_n = r$, and let $|\kappa| \leq K_0$. In this section C and p depend only on r, K_0 , and the order of differentiation, unless stated otherwise.

Smoothness and the equation. The datum f_r satisfies the hypotheses of Lemma A.1 with $K = (1 - r)^{-1}$. Indeed, with $z = (x - \kappa)/\sqrt{1 - r}$, $f'_r = R(z)/\sqrt{1 - r} > 0$ and $-f''_r = V(z)/(1 - r) \in (0, (1 - r)^{-1})$ by Lemma 2.1(a), and $f_r(x) \geq -C(1 + x^2)$ by Lemma 2.1(c). Moreover all derivatives of f_r have polynomial growth, because $(\log \Phi)' = R$, $R' = -R(z + R)$, and $0 < R(z) \leq |z| + 2$ by Lemma 2.1(b). Applying Lemma A.1 block by block from $t_n = r$ backwards, $u(t_j, \cdot)$ satisfies the same hypotheses with the same K , and $u(q, \cdot)$ is C^∞ with polynomially bounded derivatives, with constants that may depend on the grid. By Lemma A.1(iii),(iv), u is C^∞ on each $[t_{j-1}, t_j) \times \mathbb{R}$, all its x -derivatives are continuous on $[0, r] \times \mathbb{R}$ (including across grid points), and (2.8) holds classically on each open block. Integrating in q gives the integrated equation of Lemma 2.4(a).

Zeroth and first order bounds. By Lemma A.1(i) and the semigroup property of the heat kernel, $\mathbb{E}f_r(x + \sqrt{r - q}G) \leq u(q, x) \leq 0$ and $0 \leq c \leq (1 - r)^{-1}$. Since $\mathbb{E}f_r(x + \sqrt{r - q}G) \geq -C(1 + x^2)$, the quadratic lower bound holds at every grid point with a constant independent of the grid, and Lemma A.1(ii) applied at each step gives $0 < b(q, x) \leq C(1 + |x|)$. Hence the drift γb is Lipschitz in x with constant $(1 - r)^{-1}$ and of linear growth with a constant C , and the diffusion $X^{s,x}$ satisfies $\mathbb{E} \sup_{[s,r]} |X^{s,x}|^p \leq C(1 + |x|^p)$ by Gronwall's inequality.

Signs. On each constant block the identities of Lemma 2.9(a) hold classically; their derivation there uses only (2.8) and $\lambda' = -\gamma$. In particular

$$\mathcal{L}b = 0, \quad \mathcal{L}c_x = 3\gamma c c_x, \quad \mathcal{L}(1 - v) - \frac{\gamma}{\lambda} v(1 - v) = 0, \quad \mathcal{L}v + \frac{\gamma}{\lambda} (1 - v)v = 0.$$

We apply (A.1) along $X^{q,x}$ on $[q, r]$ to these identities; the potentials are bounded because $0 \leq c \leq (1-r)^{-1}$, $\lambda \geq 1-r$, and hence $0 \leq v \leq (1-r)^{-1}$. For b , $\mathcal{L}b = 0$ gives $b(q, x) = \mathbb{E}b(r, X_r^{q,x})$. Put $Z^{q,x} = (X_r^{q,x} - \kappa) / \sqrt{1-r}$. Since $(1-v)(r, \cdot) = 1 - V(Z^{q,x}) > 0$, the identity for $1-v$ gives $1-v(q, x) = \mathbb{E}[(1-V(Z^{q,x})) \exp(-\int_q^r \frac{\gamma}{\lambda} v)] > 0$. Consequently $0 \leq v < 1$ on $[0, r] \times \mathbb{R}$, so that $\int_q^r \frac{\gamma}{\lambda} v \leq \frac{r}{1-r}$ along every path, and the identity for v has the nonnegative potential $\frac{\gamma}{\lambda}(1-v)$. Hence

$$v(q, x) \geq \mathbb{E} V(Z^{q,x}), \quad 1-v(q, x) \geq e^{-r/(1-r)} \mathbb{E}[1-V(Z^{q,x})]. \quad (\text{A.3})$$

For c_x , the identity $\mathcal{L}c_x - 3\gamma c c_x = 0$ with the potential $-3\gamma c \in [-3(1-r)^{-1}, 0]$ and the terminal value $c_x(r, \cdot) = V'(z)(1-r)^{-3/2} < 0$ give

$$\mathbb{E} c_x(r, X_r^{q,x}) \leq c_x(q, x) \leq e^{-3r/(1-r)} \mathbb{E} c_x(r, X_r^{q,x}) < 0. \quad (\text{A.4})$$

Uniform polynomial bounds. Write $u^{(k)} = \partial_x^k u$. Differentiating (2.8) $k \geq 3$ times gives

$$\partial_q u^{(k)} + \frac{1}{2} u^{(k+2)} + \gamma b u^{(k+1)} - k \gamma c u^{(k)} + \frac{\gamma}{2} \sum_{j=2}^{k-2} \binom{k}{j} u^{(j+1)} u^{(k-j+1)} = 0,$$

because in $\partial_x^k(u_x^2) = \sum_{j=0}^k \binom{k}{j} u^{(j+1)} u^{(k-j+1)}$ the terms $j=0$ and $j=k$ equal $2b u^{(k+1)}$ together, and the terms $j=1$ and $j=k-1$ equal $-2kc u^{(k)}$ together. The potential $-k\gamma c$ lies in $[-k(1-r)^{-1}, 0]$, and the sum involves only the orders $3, \dots, k-1$. Since $u^{(2)} = -c$ is bounded, and $|u^{(3)}| = |c_x| \leq \mathbb{E}|c_x(r, X_r)| \leq C(1+|x|^p)$ by (A.4), induction on k with (A.1) and the moment bounds gives $|u^{(k)}(q, x)| \leq C_k(1+|x|^{p_k})$ with C_k, p_k independent of the grid. The growth bound of Lemma 2.4(b) for u follows from $u \geq \mathbb{E}f_r(x + \sqrt{r-q} G)$ and $f_r(y) \geq -C(1+y^2)$.

Behavior at $+\infty$. Let $t = x - \kappa \geq 0$ and $Z = (t + \sqrt{r-q} G) / \sqrt{1-r}$. By Lemma 2.1(c), $\log \Phi(z) \geq -2\Phi(-z) - (z^2 + 2)\mathbf{1}_{\{z < 0\}}$ for all z ; for $z < 0$ we also use $\log(1+|z|) \leq \frac{1}{2}(1+z^2)$ and $\Phi(-z) \geq \frac{1}{2}$. Now $\mathbb{E}\Phi(-Z) = \Phi(-t/\sqrt{1-q}) \leq e^{-t^2/(2(1-q))}$, and by the Cauchy-Schwarz inequality

$$\mathbb{E}[(Z^2 + 2)\mathbf{1}_{\{Z < 0\}}] \leq C(1+t^2)\mathbb{P}(Z < 0)^{1/2} \leq C(1+t^2)e^{-t^2/(4(1-q))},$$

because $\mathbb{P}(Z < 0) = \Phi(-t/\sqrt{r-q}) \leq e^{-t^2/(2(1-q))}$. Since $(1+t^2)e^{-t^2/8}$ is bounded and $1-q \leq 1$, we obtain $0 \geq u(q, x) \geq \mathbb{E}f_r(x + \sqrt{r-q} G) \geq -Ce^{-t^2/(8(1-q))}$.

Translation. The recursion (2.5) commutes with translations in x , and f_r depends on $x - \kappa$ only.

This proves Lemma 2.4(a)–(d) and (g) for step order parameters.

A.3 Continuity in the order parameter and the general case

The next identity is the basic tool: it expresses the difference of two solutions with the same terminal datum through the difference of the coefficients. We state it for the class $\mathfrak{F}$ of terminal data of Lemma 2.8. For models of Sherrington–Kirkpatrick type it appears in the proof of [76, Lemma 14], and for the present datum in the proof of [54, Proposition 13].

Lemma A.2. *Let $f \in \mathfrak{F}$, let $\xi, \tilde{\xi} \leq M$ be nonnegative nondecreasing step functions on $[0, T)$, and put $U = Y_{\xi, f}$ and $\tilde{U} = Y_{\tilde{\xi}, f}$. Let Y solve*

$$dY_t = \beta(t, Y_t) dt + dB_t \quad (q \leq t \leq T), \quad Y_q = x, \quad \beta = \frac{1}{2}\tilde{\xi}(\tilde{U}_x + U_x).$$

Then

$$\begin{aligned} \tilde{U}(q, x) - U(q, x) &= \frac{1}{2} \mathbb{E} \int_q^T (\tilde{\xi} - \xi)(t) U_x(t, Y_t)^2 dt, \\ |\tilde{U}(q, x) - U(q, x)| &\leq C(1 + x^2) \|\tilde{\xi} - \xi\|_{L^1(0, T)}, \end{aligned} \tag{A.5}$$

where C depends only on T, M, C_f , and $\text{Lip}(f')$. The identity also holds for $T = r, f = f_r$, and general $\xi = \gamma, \tilde{\xi} = \tilde{\gamma}$ in $\mathcal{U}_r$, with $U = u_\gamma$ and $\tilde{U} = u_{\tilde{\gamma}}$ (see the end of this section). For the quadratic datum $f_k(x) = -\frac{1}{2}(k - x)_+^2$, which belongs to $\mathfrak{F}$ with $\text{Lip}(f'_k) = 1$ and $C_{f_k} \leq 1 + k^2$, the bound shows that $\xi \mapsto Y_{\xi, f_k}(q, x)$ is Lipschitz for the L^1 distance on nondecreasing step functions bounded by M , uniformly for k in bounded sets, and therefore extends uniquely and continuously to all nondecreasing $\xi : [0, T) \rightarrow [0, M]$.

Proof. The one-step bounds of Lemma A.1 give $0 \leq -U_{xx}, -\tilde{U}_{xx} \leq \text{Lip}(f')$ for $q < T$ and $0 \leq U_x, \tilde{U}_x \leq C(1 + |x|)$, uniformly in the grids. Indeed, both solutions lie between the heat evolution of f and zero, so the quadratic lower bound, and with it the constant C_g of Lemma A.1(ii), remains uniform through the recursion. Thus β is Lipschitz in y with constant $M \text{Lip}(f')$, has linear growth, and $\mathbb{E} \sup_{q \leq t \leq T} |Y_t|^p \leq C_p(1 + |x|^p)$. On each open block of a common refinement of the two grids, subtracting the equations and using $\tilde{U}_x^2 - U_x^2 = (\tilde{U}_x + U_x)(\tilde{U}_x - U_x)$ gives

$$w_q + \frac{1}{2}w_{xx} + \beta w_x = -\frac{1}{2}(\tilde{\xi} - \xi)U_x^2, \quad w = \tilde{U} - U, \quad w(T, \cdot) = 0.$$

Apply (A.1) with $\mathcal{V} = 0$ on $[q, T - \varepsilon]$, and let $\varepsilon \downarrow 0$. The common terminal datum, the continuity of w up to T , and the quadratic bounds make the terminal term vanish by dominated convergence. This gives the identity, and the linear bound on U_x together with the second-moment bound on Y gives the estimate. For f_k , the bounds $\text{Lip}(f'_k) = 1$ and $f_k(x) \geq -\frac{1}{2}(|k| + |x|)^2 \geq -(1 + k^2)(1 + x^2)$ are direct, and the extension follows from the density of nondecreasing step functions bounded by M among nondecreasing functions bounded by M (left-endpoint approximations over fine partitions). $\square$

Proof of Lemma 2.4(f) for step order parameters. Let $\gamma, \tilde{\gamma} \in \mathcal{U}_r$ be step order parameters and $\varepsilon = \|\tilde{\gamma} - \gamma\|_{L^1}$. The case $k = 0$ is Lemma A.2 with $T = r$ and $f = f_r$. Keep its polarized drift $\beta = \frac{1}{2}\tilde{\gamma}(b_{\tilde{\gamma}} + b_{\gamma})$, and write $W_k = \partial_x^k(u_{\tilde{\gamma}} - u_{\gamma})$, $b = b_{\gamma}$, $c = c_{\gamma}$, and $\tilde{c} = c_{\tilde{\gamma}}$. Differentiating the equation for W_0 $k \geq 1$ times with the Leibniz rule, the term $\beta \partial_x W_0$ produces $\beta W_{k+1} + k\beta_x W_k + \sum_{j=2}^k \binom{k}{j} (\partial_x^j \beta) W_{k-j+1}$. This gives the single induction formula

$$(\partial_q + \frac{1}{2}\partial_{xx} + \beta\partial_x + k\beta_x)W_k = -\frac{1}{2}(\tilde{\gamma} - \gamma)\partial_x^k(b^2) - \sum_{j=2}^k \binom{k}{j} (\partial_x^j \beta) W_{k-j+1}. \quad (\text{A.6})$$

The sum is empty for $k = 1$ and otherwise contains only W_i with $i < k$. Moreover $\beta_x = -\frac{1}{2}\tilde{\gamma}(\tilde{c} + c) \leq 0$. The uniform polynomial bounds of §A.2 therefore imply that the right-hand side of (A.6) is bounded by

$$C(1 + |x|^p) \left(|\tilde{\gamma} - \gamma|(q) + \sum_{i < k} |W_i(q, x)| \right).$$

Since $W_k(r, \cdot) = 0$, (A.1) along the diffusion with drift β applies with the potential $k\beta_x \leq 0$, so the exponential weight is at most one. The moment bound and induction on k give

$$|W_k(q, x)| \leq C_k \varepsilon (1 + |x|^{p_k}).$$

For $k = 1$ the forcing is $-(\tilde{\gamma} - \gamma)bb_x = (\tilde{\gamma} - \gamma)bc$, which has linear growth because c is bounded; thus $p_1 = 1$, while $p_0 = 2$ comes from Lemma A.2.

Diffusions. For step order parameters $\gamma, \tilde{\gamma} \in \mathcal{U}_r$ and diffusions driven by the same Brownian motion, the drift $\tilde{\gamma}b_{\tilde{\gamma}}$ is Lipschitz in y with constant $(1 - r)^{-1}$, and

$$|\tilde{\gamma}b_{\tilde{\gamma}} - \gamma b_{\gamma}|(t, y) \leq |\tilde{\gamma} - \gamma|(t) |b_{\tilde{\gamma}}(t, y)| + |b_{\tilde{\gamma}} - b_{\gamma}|(t, y) \leq C(|\tilde{\gamma} - \gamma|(t) + \varepsilon)(1 + |y|)$$

by the growth bound on $b_{\tilde{\gamma}}$ and the case $k = 1$ of (f). By (A.2), $\sup_{[s, r]} |\tilde{X}^{s, x} - X^{s, x}| \leq C\varepsilon(1 + \sup_{[s, r]} |X^{s, x}|)$, and Lemma 2.6(b) follows for step order parameters.

General order parameters. Step order parameters are dense in $\mathcal{U}_r$ for the L^1 distance. For instance, if $0 = t_0 < \dots < t_N = 1$ is a partition containing r and $\gamma_n(q) = \gamma(t_j)$ on $[t_j, t_{j+1})$, then $\gamma_n \in \mathcal{U}_r$ and $\|\gamma_n - \gamma\|_{L^1} \leq \sum_j (t_{j+1} - t_j)(\gamma(t_{j+1}) - \gamma(t_j))$, which is at most the mesh. By (f) for step order parameters, for every sequence of step order parameters $\gamma_n \rightarrow \gamma$ in L^1 the functions $\partial_x^k u_{\gamma_n}$ form a Cauchy sequence for the weighted norms $\sup_{[0, r] \times \mathbb{R}} |\cdot| / (1 + |x|^{p_k})$, and two such sequences can be interleaved, so the limit does not depend on the sequence. This defines $u_{\gamma}(\cdot, \cdot; \kappa)$, shows that its x -derivatives are the limits of the x -derivatives and are continuous on $[0, r] \times \mathbb{R}$, and proves (f) for general order parameters. The pointwise bounds $0 \leq c \leq (1 - r)^{-1}$ and $0 \leq b \leq C(1 + |x|)$ pass to the limit, so the drift γb_{γ} is Lipschitz in x with constant $(1 - r)^{-1}$ and has linear growth. Together with (f) for $k \leq 1$, the argument given above for step order parameters therefore proves Lemma 2.6(b) for all $\gamma, \tilde{\gamma} \in \mathcal{U}_r$; in particular the diffusions of step approximations

converge to the diffusion of the limit. The estimates of (b), (d), and (g), the bounds (A.3)–(A.4) and the identity $b(q, x) = \mathbb{E}b(r, X_r^{q,x})$ pass to the limit, because the right-hand sides are continuous under the convergence of the diffusions and the uniform moment bounds give uniform integrability. This proves (b)–(d), (g), and the strict signs in (c) for general γ : $b(q, x) = \mathbb{E}f'_r(X_r^{q,x}) > 0$, $v > 0$ and $1 - v > 0$ by (A.3), and $c_x < 0$ by (A.4). The integrated equation in (a) passes to the limit since $\int_q^r |\gamma_n - \gamma| u_x^2 \leq C \|\gamma_n - \gamma\|_{L^1} (1 + x^2)$. If $\gamma = m$ on $[s, t]$, choose the partitions to contain s and t , so that $\gamma_n = m$ on $[s, t]$ and $u_{\gamma_n}(q, \cdot) = \mathcal{T}_{m,t-q} u_{\gamma_n}(t, \cdot)$ for $q \in [s, t]$. Since $u_{\gamma_n} \rightarrow u$ locally uniformly, $0 < e^{mu_{\gamma_n}} \leq 1$, and $u_{\gamma_n} \geq -C(1 + x^2)$ uniformly in n , dominated convergence gives $u(q, \cdot) = \mathcal{T}_{m,t-q} u(t, \cdot)$ on $[s, t]$. This proves Lemma 2.5(a) for general γ , and Lemma A.1(iii),(iv), applied to the function $u(t, \cdot)$, which is smooth with polynomially bounded derivatives, gives the smoothness statement of (a).

The identity (A.5) extends to $\gamma, \tilde{\gamma} \in \mathcal{U}_r$ in the same way. Let $\gamma_n \rightarrow \gamma$ and $\tilde{\gamma}_n \rightarrow \tilde{\gamma}$ be step approximations in $\mathcal{U}_r$, and let Y^n be the diffusion of Lemma A.2 for the pair $(\gamma_n, \tilde{\gamma}_n)$. The left sides converge by (f). The drifts of Y^n are Lipschitz with constant $(1 - r)^{-1}$ and, by (b) and (f) with $k = 1$, differ from $\beta = \frac{1}{2}\tilde{\gamma}(b_{\tilde{\gamma}} + b_{\gamma})$ by at most $C(|\tilde{\gamma}_n - \tilde{\gamma}|(t) + \varepsilon_n)(1 + |y|)$, where $\varepsilon_n = \|\gamma_n - \gamma\|_{L^1} + \|\tilde{\gamma}_n - \tilde{\gamma}\|_{L^1}$. By (A.2), $Y^n \rightarrow Y$ in every moment, uniformly on $[q, r]$. Since $|b_{\gamma_n}^2 - b_{\gamma}^2| \leq C\varepsilon_n(1 + y^2)$ and $|\partial_y(b_{\gamma}^2)| \leq C(1 + |y|)$, the expectations $\mathbb{E}b_{\gamma_n}(t, Y_t^n)^2$ converge to $\mathbb{E}b_{\gamma}(t, Y_t)^2$ uniformly in t , and they are bounded uniformly in n and t . Together with $\int_q^r |(\tilde{\gamma}_n - \gamma_n) - (\tilde{\gamma} - \gamma)| \leq \varepsilon_n$, this shows that the right sides converge.

It remains to prove the Lipschitz bound on $\mathcal{P}_\kappa$ and compactness. By (2.12), on $\mathcal{U}_r$ the entropy term is $\frac{1}{2} \int_0^r \lambda^{-1} + \frac{1}{2} \log(1 - r)$ with $\lambda \geq 1 - r$ and $|\lambda_{\tilde{\gamma}} - \lambda_{\gamma}| \leq \|\tilde{\gamma} - \gamma\|_{L^1}$, so its variation is at most $\frac{r}{2(1-r)^2} \|\tilde{\gamma} - \gamma\|_{L^1}$; the energy term is controlled by (f) with $k = 0$ at $(0, 0)$. Compactness of $\mathcal{U}_r$ follows from Helly's selection theorem: a sequence in $\mathcal{U}_r$ has a subsequence converging at every point to a nondecreasing function, which after right-continuous modification lies in $\mathcal{U}_r$, and the convergence holds in L^1 by dominated convergence.

A.4 Itô's formula along the diffusion

Proof of Lemma 2.6. Part (a) was shown in §A.2 for step order parameters, and the bounds pass to the limit; (b) was shown in §A.3. For (c), let first γ be a step order parameter. Then every $g \in \mathcal{D}$ is $C^{1,2}$ on each closed block, where $\partial_q g$ is the formal expression with γ equal to its value on the block, and g is continuous across grid points. Itô's formula on each block, the polynomial growth of g_x , and the moment bounds give the integrated identity. For general $\gamma \in \mathcal{U}_r$ let $\gamma_n(q) = \gamma(\min\{(\lfloor nq \rfloor + 1)/n, 1\})$ for $0 \leq q < 1$, and $\gamma_n(1) = 1$. Then $\gamma_n \geq \gamma$ is a step order parameter in $\mathcal{U}_Q \subset \mathcal{U}_r$, $\gamma_n(q) \rightarrow \gamma(q)$ for every q by right-continuity, and $\|\gamma_n - \gamma\|_{L^1} \rightarrow 0$. The identity holds for γ_n and the

corresponding $g_n \in \mathcal{D}$, given by the same formal expression with u_{γ_n} and λ_{γ_n} in place of u_γ and λ_γ . By Lemma 2.4(b),(f), the bounds $|\lambda_{\gamma_n} - \lambda_\gamma| \leq \|\gamma_n - \gamma\|_{L^1}$ and $\lambda_{\gamma_n} \geq 1 - r$, and Lemma 2.6(b), we have $g_n(q, X_q^n) \rightarrow g(q, X_q)$ in probability with uniformly bounded moments of every order, and likewise for $\mathcal{L}_0 g_n$ and $\mathcal{L}_1 g_n$. Since $\gamma_n(q) \rightarrow \gamma(q)$ for every q and the integrands are bounded uniformly in n and q , dominated convergence gives the identity for γ . The functions $h \in \mathcal{D}$ are continuous on $[0, r] \times \mathbb{R}$ with polynomial growth, and X has continuous paths, so $q \mapsto \mathbb{E}h(q, X_q)$ is continuous. Hence e_g is Lipschitz, and the one-sided derivatives follow from the right-continuity of γ and the existence of $\gamma(q-)$. If $\gamma = m$ on $[s', t')$, the integrand equals $e_{(\mathcal{L}_0 + m\mathcal{L}_1)g}$ on $[s', t')$, which is continuous on $[s', t']$; iterating gives $e_g \in C^\infty([s', t'])$.

A.5 The control representation

The control representation is credited after Lemma 2.8. The proof below is the verification argument of [76, Lemma 18].

Proof of Lemma 2.8. The regularity statements and the bounds on Y , Y_x , and Y_{xx} follow from Lemma A.1, applied block by block from T backwards. Each $Y(t_j, \cdot)$ satisfies the hypotheses of that lemma with the Lipschitz constant $K = \text{Lip}(f')$ and with $Y(t_j, x) \geq \mathbb{E}f(x + \sqrt{T - t_j}G) \geq -C(1 + x^2)$. The bound $Y_x \leq C(1 + |x|)$ is Lemma A.1(ii), and the continuity of Y and Y_x up to $q = T$ is Lemma A.1(iii) with $k \leq 1$.

Let θ be a control and $X_q^\theta = x + B_q + \int_0^q \theta$. Since $\xi \leq \|\xi\|_\infty$, $\mathbb{E} \int_0^T \theta^2 \leq \|\xi\|_\infty \mathbb{E} \int_0^T \theta^2 / \xi < \infty$, so $\mathbb{E} \sup_{[0, T]} |X^\theta|^2 < \infty$. On each closed block $[t_{j-1}, t_j]$ with $t_j < T$, and on $[t_{n-1}, T - \varepsilon]$, Itô's formula gives

$$dY(q, X_q^\theta) = (\theta_q Y_x - \frac{1}{2} \xi Y_x^2) dq + Y_x dB_q \leq \frac{\theta_q^2}{2\xi(q)} dq + Y_x dB_q,$$

because $\theta Y_x - \frac{1}{2} \xi Y_x^2 = \frac{\theta^2}{2\xi} - \frac{\xi}{2} (Y_x - \theta/\xi)^2$ when $\xi > 0$, and both sides vanish when $\xi = \theta = 0$. The stochastic integral has a square-integrable integrand, since Y_x has linear growth, and therefore has mean zero. Taking expectations, summing over blocks, and letting $\varepsilon \downarrow 0$, using the continuity of Y on $[0, T] \times \mathbb{R}$, its quadratic growth, and dominated convergence, gives (2.20). For $\theta^* = \xi Y_x(q, X_q^*)$ the inequality is an equality at every step. The drift ξY_x is Lipschitz in x with constant $\|\xi\|_\infty \text{Lip}(f')$ and has linear growth, so X^* exists, is unique, and has moments of all orders, and θ^* is a control. Finally, differentiating the equation of Y in x on each block shows that Y_x solves $\partial_q Y_x + \frac{1}{2} \partial_{xx} Y_x + \xi Y_x \partial_x Y_x = 0$, so by Itô's formula $dY_x(q, X_q^*) = Y_{xx}(q, X_q^*) dB_q$ on $[0, T]$. Since $|Y_{xx}| \leq \text{Lip}(f')$, this is a square-integrable martingale, and its limit as $q \uparrow T$ is $f'(X_T^*)$ by the continuity of Y_x up to T .

A.6 The expansion at $-\infty$

Proof of Lemma 2.4(e). Constants depend only on r and K_0 . Write $y = \kappa - x > 0$.

STEP 1. $q \in [Q, r]$. Here $u = \log \Phi(-t)$ and $\psi = \sqrt{1-q} \bar{R}(-t)$ with $t = y/\sqrt{1-q}$, by (2.9) and Lemma 2.9(b). Since $\phi(t) = e^{-t^2/2}/\sqrt{2\pi}$,

$$\log \Phi(-t) = -\frac{1}{2}t^2 - \log t - \log \sqrt{2\pi} + \log \frac{t\Phi(-t)}{\phi(t)},$$

and the last term lies in $(-t^{-2}, 0)$, because $\frac{t^2}{1+t^2} < \frac{t\Phi(-t)}{\phi(t)} < 1$ by Lemma 2.1(b). Since $\lambda(q) = 1 - q$ and $\omega(q) = \frac{1}{2} \log(1 - q) - \frac{1}{2} \log(2\pi)$ on $[Q, 1]$, this is the first claim. By Lemma 2.1(b), $\frac{1-q}{y} - \frac{3(1-q)^2}{y^3} < \psi < \frac{1-q}{y}$ for $t \geq 2$, which is the second claim, because $\lambda b - (\kappa - x) = \psi$.

STEP 2. *The estimate for ψ when $q < Q$.* Let $\bar{y} = y/\lambda(q) \geq y$ and let $M_t = (X_t^{q,x} - \kappa)/\lambda(t)$ for $t \in [q, Q]$. By (2.25), $M_t = -\bar{y} + A_t + N_t$ with

$$A_t = \int_q^t \frac{\gamma}{\lambda^2} \psi(s, X_s^{q,x}) ds \geq 0, \quad N_t = \int_q^t \frac{dB_s}{\lambda(s)},$$

and N is a Gaussian martingale with $\langle N \rangle_Q \leq \delta^{-2} \leq (1-r)^{-2}$. By (2.24), $\psi(q, x) = \mathbb{E} \sqrt{\delta} \bar{R}(\sqrt{\delta} M_Q)$.

Lower bound. By (2.24) and Lemma 2.1(b), for $\sqrt{\delta} \bar{y} \geq 2$,

$$\psi(q, x) \geq \sqrt{\delta} \bar{R}(-\sqrt{\delta} \bar{y}) \geq \frac{1}{\bar{y}} - \frac{3}{\delta \bar{y}^3} \geq \frac{\lambda(q)}{y} - \frac{C}{y^3}.$$

Upper bound. Let $E = \{\sup_{[q, Q]} |N| \leq \bar{y}/4\}$; the reflection principle gives $\mathbb{P}(E^c) \leq 4 \exp\{-(1-r)^2 \bar{y}^2/32\}$. Let $T_\star = \inf\{t \geq q : M_t \geq -\bar{y}/2\} \wedge Q$. For $t \leq T_\star$ we have $X_t \leq \kappa - \lambda(t) \bar{y}/2 < \kappa$, so Lemma 2.10(b) gives $\psi(t, X_t) \leq (1-t)/(\kappa - X_t) \leq 2/(\lambda(t) \bar{y})$, and hence

$$A_{T_\star} \leq \frac{2}{\bar{y}} \int_0^Q \lambda^{-3} \leq \frac{C_1}{\bar{y}}, \quad C_1 = 2(1-r)^{-3}.$$

If $\bar{y}^2 \geq 8C_1$, then on E we have $M_t \leq -\bar{y} + \bar{y}/8 + \bar{y}/4 < -\bar{y}/2$ for $t \leq T_\star$, so by continuity $T_\star = Q$, and $M_Q = -\bar{y}(1-\eta)$ with $\eta = (A_Q + N_Q)/\bar{y} \in [-\frac{1}{4}, \frac{3}{8}]$. Since $0 < \bar{R}(z) < 1/|z|$ for $z < 0$, on E

$$\sqrt{\delta} \bar{R}(\sqrt{\delta} M_Q) \leq \frac{1}{|M_Q|} = \frac{1}{\bar{y}(1-\eta)} \leq \frac{1}{\bar{y}} (1 + \eta + 2\eta^2).$$

The event E is invariant under $B \mapsto -B$, which maps N to $-N$; hence $\mathbb{E}[N_Q \mathbf{1}_E] = 0$. Using $A_Q \leq C_1/\bar{y}$ on E and $\mathbb{E} N_Q^2 \leq (1-r)^{-2}$, we get $\mathbb{E}[\eta \mathbf{1}_E] \leq C_1 \bar{y}^{-2}$ and $\mathbb{E}[\eta^2 \mathbf{1}_E] \leq C \bar{y}^{-2}$. On

E^c we use $\bar{R}(z) \leq |z| + 1$, the Cauchy–Schwarz inequality, and the moment bound for X_Q :

$$\mathbb{E}[\sqrt{\delta} \bar{R}(\sqrt{\delta} M_Q) \mathbf{1}_{E^c}] \leq \mathbb{E}[(|X_Q - \kappa| + 1) \mathbf{1}_{E^c}] \leq C(1 + \bar{y}) e^{-(1-r)^2 \bar{y}^2 / 64}.$$

Altogether $\psi(q, x) \leq \bar{y}^{-1} + C\bar{y}^{-3} \leq \lambda(q)/y + Cy^{-3}$ for y large. This proves the second estimate of (e).

STEP 3. *The expansion of u .* Put $\Xi_q(x) = u(q, x) + \frac{y^2}{2\lambda(q)} + \log y$. Since $b = (\psi + y)/\lambda$, we have $\Xi'_q(x) = (\psi - \lambda/y)/\lambda$, so $|\Xi'_q(x)| \leq C(1-r)^{-1}y^{-3}$ by Steps 1 and 2. Hence $\tilde{\omega}(q) = \lim_{x \rightarrow -\infty} \Xi_q(x)$ exists and $|\Xi_q(x) - \tilde{\omega}(q)| \leq Cy^{-2}$. It remains to show $\tilde{\omega} = \omega$. By Steps 1 and 2 and Lemma 2.10(b), uniformly for $s \in [0, r]$ and all large y ,

$$u_x^2 = \frac{(y + \psi)^2}{\lambda^2} = \frac{y^2}{\lambda^2} + \frac{2}{\lambda} + O(y^{-2}), \quad u_{xx} = -\frac{v}{\lambda} = -\frac{1 - \psi_x}{\lambda} = -\frac{1}{\lambda} + O(y^{-2}),$$

because $y\psi = \lambda + O(y^{-2})$, $\psi^2 = O(y^{-2})$, and $0 < \psi_x = 1 - v \leq 4(1-s)/y^2$. For $0 \leq q < q' \leq r$, the difference of the integrated equations of Lemma 2.4(a) at q and q' , together with $\frac{d}{ds}\lambda^{-1} = \gamma/\lambda^2$, gives

$$\begin{aligned} u(q, x) - u(q', x) &= \frac{1}{2} \int_q^{q'} (u_{xx} + \gamma u_x^2)(s, x) ds \\ &= \frac{y^2}{2} \left(\frac{1}{\lambda(q')} - \frac{1}{\lambda(q)} \right) + \int_q^{q'} \frac{\gamma(s) - \frac{1}{2}}{\lambda(s)} ds + O(y^{-2}). \end{aligned}$$

Adding $\frac{y^2}{2\lambda(q)} + \log y$, subtracting $\frac{y^2}{2\lambda(q')} + \log y$, and letting $x \rightarrow -\infty$, we obtain

$$\tilde{\omega}(q) - \tilde{\omega}(q') = \int_q^{q'} \frac{\gamma - \frac{1}{2}}{\lambda} = \omega(q) - \omega(q'),$$

since $\omega' = (1 - 2\gamma)/(2\lambda)$ almost everywhere by (2.13). Taking $q' = r \geq Q$ and using Step 1, $\tilde{\omega} = \omega$ on $[0, r]$. This proves (e).

Estimates for the top-jump bound

This appendix proves Proposition 6.12, which bounds the breakpoint ratio $\tilde{x}_\gamma$ on the last constant piece of an arbitrary order parameter; it is the input to the top-jump bound of Theorem 6.13. §B.1 collects one-variable facts about R , $\bar{R}$, and V beyond those of Lemma 2.1: the monotonicity of $(V')^2/(V^2(1-V))$ (Lemma B.2) and the bound $x_\Phi < 1$ (Lemma B.3). §B.2 shows that on a last constant piece of mass m the functions v and v_x depend only on a scale-free variable z and on the ratio $\beta = \delta/\lambda \in (0, 1]$. §B.3 proves bounds on these functions that are uniform in $\beta \in [0, 1]$ and $m \in [1/2, 1]$, including their tails (Proposition B.5). §B.4 combines them with the tilted-density representation of Lemma 2.7 and Chebyshev's association inequality (Lemma B.6).

Throughout, G denotes a standard Gaussian variable and ϕ_s the $N(0, s)$ density, as in Chapter 2. Constants c, C, C_j are numerical unless stated otherwise.

B.1 Gaussian facts

We use the following facts from Lemma 2.1: $R > 0$, $\bar{R} > 0$, $0 < V < 1$, $R' = -V$, $\bar{R}' = 1 - V$, the identities (2.2) for $V' < 0$ and V'' , and $\mathbb{E}T_z = \bar{R}(z)$, $\text{Var } T_z = 1 - V(z)$ for the variable T_z with density proportional to $e^{zw - w^2/2}$ on $[0, \infty)$ (part (a)); $t < R(-t) < t + 1/t$ and $0 < \bar{R}(-t) < 1/t$ for $t > 0$, and $0 < 1 - V(-t) < 3t^{-2}$ for $t \geq 2$ (part (b)); $\Phi(-t) \leq e^{-t^2/2}$ for $t \geq 0$ (part (c)); and $R(z) \leq 2\phi(z)$ and $V(z) \leq 2(1+z)\phi(z)$ for $z \geq 0$ (part (e)). Put

$$r(z) := \frac{V'(z)^2}{V(z)^2(1-V(z))} \quad (z \in \mathbb{R}).$$

Lemma B.1. (a) For $z \geq 0$: $|V'(z)| \leq 2(2+z)^2\phi(z)$ and $r(z) \leq C(1+z)^2$.

(b) For $t \geq 2$,

$$|1 - V(-t) - t^{-2}| \leq Ct^{-4}, \quad |V'(-t) + 2t^{-3}| \leq Ct^{-5}.$$

Consequently, for $z \leq 0$: $c/(1+z^2) \leq 1 - V(z) \leq C/(1+z^2)$, $|V'(z)| \leq C(1+|z|)^{-3}$, $|V''(z)| \leq C(1+|z|)^{-2}$, and $r(z) \leq C(1+z^4)^{-1}$.

Proof. (a) Let $z \geq 0$. Since $R \leq R(0) < 1$, $\bar{R} = z + R \leq 1 + z$. As $0 < V < 1$, $|V'| \leq R(1 + \bar{R}^2) \leq 2\phi(z)(1 + (1+z)^2) \leq 2(2+z)^2\phi(z)$. Moreover $|V'/V| = |1 - V - \bar{R}^2|/\bar{R} \leq 1/\bar{R} + \bar{R} \leq 1/R(0) + 1 + z$, because $\bar{R}$ is increasing and $\bar{R}(0) = R(0)$; and $1 - V \geq 1 - V(0) > 0$ because V is decreasing. Hence $r(z) \leq C(1+z)^2$.

(b) Let $t \geq 2$ and $I_k := \int_0^\infty w^k \exp(-tw - w^2/2) dw$. Then $\mathbb{E}T_{-t} = I_1/I_0$ and $\mathbb{E}T_{-t}^2 = I_2/I_0$, so that

$$1 - V(-t) = \frac{I_2 I_0 - I_1^2}{I_0^2}, \quad V'(-t) = R(-t)(\text{Var } T_{-t} - (\mathbb{E}T_{-t})^2) = R(-t) \frac{I_2 I_0 - 2I_1^2}{I_0^2},$$

by (2.2). Since $1 - w^2/2 \leq \exp(-w^2/2) \leq 1 - w^2/2 + w^4/8$ and $\int_0^\infty w^j e^{-tw} dw = j! t^{-j-1}$,

$$\begin{aligned} I_0 &= \frac{1}{t}(1 - t^{-2} + O(t^{-4})), \\ I_1 &= \frac{1}{t^2}(1 - 3t^{-2} + O(t^{-4})), \\ I_2 &= \frac{2}{t^3}(1 - 6t^{-2} + O(t^{-4})), \end{aligned}$$

where here and below $O(t^{-j})$ denotes a quantity bounded by a numerical constant times t^{-j} for $t \geq 2$. Hence

$$\begin{aligned} I_2 I_0 &= \frac{2}{t^4}(1 - 7t^{-2} + O(t^{-4})), \\ I_1^2 &= \frac{1}{t^4}(1 - 6t^{-2} + O(t^{-4})), \\ I_0^2 &= \frac{1}{t^2}(1 - 2t^{-2} + O(t^{-4})), \end{aligned}$$

and $I_0^2 \geq \frac{9}{16}t^{-2}$ because $I_0 \geq t^{-1}(1 - t^{-2})$. Therefore

$$\begin{aligned} 1 - V(-t) &= \frac{1 - 8t^{-2} + O(t^{-4})}{t^2(1 - 2t^{-2} + O(t^{-4}))} = t^{-2} + O(t^{-4}), \\ \frac{I_2 I_0 - 2I_1^2}{I_0^2} &= \frac{-2t^{-2} + O(t^{-4})}{t^2(1 + O(t^{-2}))} = -2t^{-4} + O(t^{-6}). \end{aligned}$$

With $R(-t) = t + O(t^{-1})$ this gives $V'(-t) = -2t^{-3} + O(t^{-5})$. For the consequences, let $t_0 \geq 2$ be a large numerical constant. For $z \leq -t_0$ the bounds on $1 - V$ and V' follow from the expansions. On $[-t_0, 0]$ the functions $1 - V$ and $|V'|$ are continuous, and $1 - V \geq 1 - V(-t_0) > 0$. The bound on V'' follows from (2.2), since $0 < R(z) \leq |z| + 1/|z|$ for $z \leq -1$ and $V(1 - V) \leq 1 - V$. For r , use in addition $V \geq V(0) > 0$ on $(-\infty, 0]$. $\square$

The next lemma is the monotonicity that drives the Chebyshev argument of §B.4.

Lemma B.2. *The function $r = (V')^2/(V^2(1 - V))$ is strictly increasing on $\mathbb{R}$.*

Proof. Write $a = 1 - V > 0$ and $b = V'/R = 1 - V - \bar{R}^2$, so that $\bar{R}^2 = a - b$ and $\bar{R}(R + \bar{R}) = V + \bar{R}^2 = 1 - b$.

STEP 1. Reduction. Since $(\log r)' = 2V''/V' - 2V'/V + V'/(1 - V)$, we have $(\log r)' \cdot V'V(1 - V) = -\mathcal{C}$ with

$$\mathcal{C} := (V')^2(2 - 3V) - 2V''V(1 - V).$$

As $V'V(1-V) < 0$, the derivative $(\log r)'$ has the sign of $\mathcal{C}$. Inserting $V'' = -(R + \bar{R})V' - 2V(1-V)$ from (2.2), $V' = Rb$, and $V = R\bar{R}$,

$$\mathcal{C} = R^2[b^2(3a-1) + 2(1-b)ab + 4a^2(a-b)] = R^2(2a-b)(2a^2-ab+b),$$

where the factorization is verified by expanding both sides to $4a^3 - 4a^2b + ab^2 + 2ab - b^2$. Since $2a-b = a + \bar{R}^2 > 0$, it suffices to show that

$$\mathcal{Q} := 2a^2 - ab + b = (1-V)(2-V) - V\bar{R}^2$$

is positive on $\mathbb{R}$.

STEP 2. An equation for $\mathcal{Q}$. We have $\mathcal{Q}' = -V'(3-2V+\bar{R}^2) - 2V\bar{R}(1-V)$ and $V' = Rb = (V/\bar{R})(1-V-\bar{R}^2)$. Eliminating $\bar{R}^2 = ((1-V)(2-V) - \mathcal{Q})/V$, so that $1-V-\bar{R}^2 = (\mathcal{Q} - 2(1-V)^2)/V$ and $3-2V+\bar{R}^2 = (2-V^2-\mathcal{Q})/V$, and expanding,

$$V\bar{R}\mathcal{Q}' = 4(1-V)^3 - \mathcal{Q}(1+3(1-V)^2 - \mathcal{Q}). \quad (\text{B.1})$$

STEP 3. Barrier. The function $\mathfrak{k} := (1+3(1-V)^2 - \mathcal{Q})/(V\bar{R})$ is continuous, since $V\bar{R} > 0$. Fix $z_* \in \mathbb{R}$ and put $\mathcal{F}(z) := \mathcal{Q}(z) \exp \int_{z_*}^z \mathfrak{k}$. By (B.1), $\mathcal{F}' = 4(1-V)^3(V\bar{R})^{-1} \exp \int_{z_*}^z \mathfrak{k} > 0$, so $\mathcal{F}$ is strictly increasing. As $z = -t \rightarrow -\infty$ we have $\bar{R}^2 \leq t^{-2}$ and $1-V \leq 3t^{-2}$; hence $|\mathcal{Q}| \leq 2(1-V) + \bar{R}^2 \rightarrow 0$, and the numerator of $\mathfrak{k}$ tends to one. Choosing z_* sufficiently negative, $\mathfrak{k} > 0$ on $(-\infty, z_*]$, and then $|\mathcal{F}(z)| \leq |\mathcal{Q}(z)| \rightarrow 0$ as $z \rightarrow -\infty$. An increasing function with limit zero at $-\infty$ is positive, so $\mathcal{Q} > 0$. $\square$

Lemma B.3. *The constant x_Φ of (6.20) satisfies*

$$x_\Phi = 1 - \frac{\int_{\mathbb{R}} \Phi V^2 dz}{4 \int_{\mathbb{R}} \Phi V^2 (1-V) dz} < 1.$$

Proof. Let $\mathcal{G}_{\text{top}} f := \frac{1}{2}f'' + (R + \frac{z}{2})f'$. This is the generator of the diffusion (7.20) of Chapter 7, but only the formula is used here. By (2.2),

$$\mathcal{G}_{\text{top}}(V^2) = (V')^2 + VV'' + (2R+z)VV' = (V')^2 - 2V^2(1-V).$$

The formal adjoint is $\mathcal{G}_{\text{top}}^* w = \frac{1}{2}w'' - ((R + \frac{z}{2})w)'$. With $\Phi' = R\Phi$, $\Phi'' = -zR\Phi$, $(R + \frac{z}{2})' = \frac{1}{2} - V$, and $V = zR + R^2$,

$$\mathcal{G}_{\text{top}}^* \Phi = -\frac{1}{2}zR\Phi - \Phi(\frac{1}{2} - V + R^2 + \frac{1}{2}zR) = -\frac{1}{2}\Phi.$$

For $N > 0$,

$$\int_{-N}^N (\Phi \mathcal{G}_{\text{top}}(V^2) - V^2 \mathcal{G}_{\text{top}}^* \Phi) dz = \left[\frac{1}{2}(\Phi(V^2)' - \Phi'V^2) + (R + \frac{z}{2})\Phi V^2 \right]_{-N}^N.$$

As $N \rightarrow \infty$ the boundary terms vanish: at $+N$ because V and V' are $O(N^2\phi(N))$ by Lemma B.1(a) while Φ and R are bounded, and at $-N$ because $\Phi(-N) \leq \phi(N)/N$, $|V| \leq 1$, $|V'| \leq C$, and $R(-N) \leq N+1$. All integrands are integrable. Hence $\int \Phi(V')^2 - 2 \int \Phi V^2(1-V) = -\frac{1}{2} \int \Phi V^2$, which is the formula, and $\int \Phi V^2 > 0$. $\square$

B.2 The last constant piece in scale-free form

For $m \in (0, 1]$ let M_m be a random variable with distribution function Φ^m , independent of G . Its density is $f_M := m\phi\Phi^{m-1}$. For $\beta \in [0, 1]$ put $\sigma_\beta := \sqrt{(1-\beta)/m}$ and

$$\begin{aligned} Y &:= \sqrt{\beta} M_m + \sigma_\beta G, & F_m(z; \beta) &:= \mathbb{P}(Y \leq z), & f &:= \partial_z F_m, \\ R_m(z; \beta) &:= \partial_z \log F_m(z; \beta), & v_m(z; \beta) &:= -\frac{1}{m} \partial_z^2 \log F_m(z; \beta). \end{aligned}$$

When m and β are fixed we write F, v , and so on.

Lemma B.4. *Let $\gamma \in \mathcal{U}$ and $\kappa \in \mathbb{R}$, and suppose that $\gamma \equiv m \in (0, 1]$ on $[p, Q]$ for some $p < Q$. For $q \in [p, Q]$ put $\lambda = \lambda(q)$, $\beta := \delta/\lambda \in (0, 1]$ and $z := (x - \kappa)/\sqrt{\lambda}$. Then*

$$\begin{aligned} e^{mu(q,x)} &= F_m(z; \beta), & b(q, x) &= \frac{R_m(z; \beta)}{m\sqrt{\lambda}}, \\ v(q, x) &= v_m(z; \beta), & v_x(q, x) &= \frac{\partial_z v_m(z; \beta)}{\sqrt{\lambda}}. \end{aligned}$$

Proof. On $[q, Q]$ the order parameter equals m , and $u(Q, x) = \log \Phi((x - \kappa)/\sqrt{\delta})$. By Lemma 2.5(a),

$$e^{mu(q,x)} = \mathbb{E} \Phi\left(\frac{x - \kappa + \sqrt{Q - q} G}{\sqrt{\delta}}\right)^m = \mathbb{P}(\sqrt{\delta} M_m \leq x - \kappa + \sqrt{Q - q} G),$$

and by the symmetry of G this is $\mathbb{P}(\sqrt{\delta} M_m + \sqrt{Q - q} G \leq x - \kappa)$. Divide by $\sqrt{\lambda}$ and use $\lambda = \delta + m(Q - q)$, so that $(Q - q)/\lambda = (1 - \beta)/m$; this gives $e^{mu} = F_m(z; \beta)$. The other formulas follow from $b = u_x$, $v = -\lambda u_{xx}$, and $\partial_x = \lambda^{-1/2} \partial_z$. $\square$

In particular $v_m(\cdot; 1) = V$ (then $F_m = \Phi^m$), $v_m(\cdot; 0) = V(\sqrt{m} \cdot)$ (then $Y = G/\sqrt{m}$), and $v_1(\cdot; \beta) = V$ for every β (then Y is standard Gaussian).

Realization. Some estimates below use the Parisi diffusion of a two-level order parameter. Given $m \in (0, 1)$ and $\beta \in (0, 1)$, choose $\delta \in (0, \frac{1}{2})$ so small that $\delta(1/\beta - 1)/m \leq 1 - \delta$, and put $Q := 1 - \delta$, $q_1 := Q - \delta(1/\beta - 1)/m \in [0, Q)$, $\kappa := 0$ and

$$\gamma^{m,\beta} := m\mathbf{1}_{[q_1, Q)} + \mathbf{1}_{[Q, 1]} \in \mathcal{U}. \quad (\text{B.2})$$

Then $\lambda(q) = \delta + m(Q - q)$ on $[q_1, Q]$ and $\lambda(q_1) = \delta/\beta$. By Lemma B.4, $v(q, x) = v_m(x/\sqrt{\lambda(q)}; \beta_q)$ for $q \in [q_1, Q]$, with $\beta_q := \delta/\lambda(q) \in [\beta, 1]$.

B.3 Uniform bounds on the scale-free functions

Fix $m_0 := 1/2$. All constants in this section are numerical; in particular they are uniform in $m \in [m_0, 1]$ and $\beta \in [0, 1]$.

Proposition B.5. For all $m \in [m_0, 1]$ and $\beta \in [0, 1]$:

(V0) The maps $(m, \beta, z) \mapsto F_m(z; \beta), v_m(z; \beta), \partial_z v_m(z; \beta)$ are continuous on $[m_0, 1] \times [0, 1] \times \mathbb{R}$, and $0 < v_m < 1$.

(V1) $F_m(0; \beta) \geq 1/4$, and $F_m(z; \beta) \leq 2 \exp(-z^2/16)$ for $z \leq 0$.

(V2) For $z \geq 0$, $v_m(z; \beta) + |\partial_z v_m(z; \beta)| \leq C \exp(-cz^2)$.

(V3) For $z \leq 0$, $1 - v_m(z; \beta) \leq C/(1 + z^2)$.

(V4) For $z \leq 0$, $1 - v_m(z; \beta) \geq c/(1 + z^2)$.

(V5) For $z \leq 0$, $|\partial_z v_m(z; \beta)| \leq C(1 + |z|)^{-3}$.

Consequently $r_m := (\partial_z v_m)^2 / (v_m^2(1 - v_m))$ satisfies $r_m(z; \beta) \leq C(1 + z^4)^{-1}$ for $z \leq 0$.

Proof. We prove (V0)–(V5) in five steps and deduce the bound on r_m in a sixth.

STEP 1. *Proof of (V0).* For $\beta \in [0, \frac{1}{2}]$ we have $\sigma_\beta^2 \in [\frac{1}{2}, 2]$ and $f^{(k)}(z) = \mathbb{E} \phi_{\sigma_\beta^2}^{(k)}(z - \sqrt{\beta} M_m)$ for $k \leq 2$. Realizing $M_m = \Phi^{-1}(U^{1/m})$ with U uniform on $(0, 1)$ couples these variables continuously in m , and $\phi_s^{(k)}$ is bounded and jointly continuous for $s \in [\frac{1}{2}, 2]$; dominated convergence gives joint continuity. For $\beta \in [\frac{1}{2}, 1]$ we use $f^{(k)}(z) = \mathbb{E} f_{\sqrt{\beta} M}^{(k)}(z - \sigma_\beta G)$ with $f_{\sqrt{\beta} M}^{(k)}(t) = \beta^{-(k+1)/2} f_M^{(k)}(t/\sqrt{\beta})$. Here $f'_M = f_M \cdot (-t + (m-1)R)$ and $f''_M = f_M \cdot ((-t + (m-1)R)^2 - 1 - (m-1)V)$. For $t \leq 0$, $f_M = m\phi^m R^{1-m} \leq C(1 + |t|) \exp(-mt^2/2)$, while $f_M \leq 2\phi$ for $t \geq 0$. Hence $f_M^{(k)}$ is bounded and jointly continuous in (m, t) , and dominated convergence again gives joint continuity, including at $\sigma_\beta = 0$. Also F_m is continuous: under the coupling above, $\sqrt{\beta} M_m + \sigma_\beta G$ varies continuously in (m, β) almost surely, and its law has no atom, so bounded convergence applies to its distribution function. Its density is positive, hence $F_m > 0$, and therefore

$$v_m = \frac{1}{m} \left(\frac{f^2}{F^2} - \frac{f'}{F} \right), \quad \partial_z v_m = -\frac{1}{m} \left(\frac{f''}{F} - \frac{3ff'}{F^2} + \frac{2f^3}{F^3} \right)$$

are continuous. For $m = 1$ or $\beta \in \{0, 1\}$, $v_m \in (0, 1)$ because $0 < V < 1$. For $m \in [m_0, 1]$ and $\beta \in (0, 1)$, $v_m(z; \beta)$ is the value of v at level q_1 of the order parameter (B.2), so $0 < v_m < 1$ by Lemma 2.4(c).

STEP 2. *Proof of (V1).* Write $Y = X_1 + X_2$ with $X_1 = \sqrt{\beta} M_m$ and $X_2 = \sigma_\beta G$. Independence gives $F_m(0; \beta) \geq \mathbb{P}(X_1 \leq 0) \mathbb{P}(X_2 \leq 0) \geq 2^{-m}/2 \geq 1/4$. For $z \leq 0$ and $0 < \beta < 1$, the union bound gives

$$F_m(z; \beta) \leq \mathbb{P}(X_1 \leq z/2) + \mathbb{P}(X_2 \leq z/2) \leq 2 \exp(-z^2/16).$$

Indeed, $\mathbb{P}(X_1 \leq z/2) = \Phi(z/(2\sqrt{\beta}))^m \leq \Phi(z/2)^{1/2} \leq \exp(-z^2/16)$ by Lemma 2.1(c), and X_2 is centered Gaussian with variance at most 2. The cases $\beta \in \{0, 1\}$ follow by continuity.

STEP 3. *A Gaussian factorization; proof of (V3) and (V5).* Put $\bar{m} := 1 - m \in [0, \frac{1}{2}]$ and $\mathfrak{g} := R^{\bar{m}}$. We first record the elementary estimates

$$c(1 + |x|)^{\bar{m}} \leq \mathfrak{g}(x) \leq C(1 + |x|)^{\bar{m}}, \quad |\mathfrak{g}^{(j)}(x)| \leq C_j(1 + |x|)^{\bar{m}-j} \quad (x \leq 0, 1 \leq j \leq 3), \quad (\text{B.3})$$

with constants uniform in $\bar{m}$. Indeed, the Mills-ratio bounds of Lemma 2.1(b) and the monotonicity of R give $c(1 + |x|) \leq R(x) \leq C(1 + |x|)$ on $(-\infty, 0]$, while $R' = -V = O(1)$, $R'' = -V' = O((1 + |x|)^{-3})$, and $R''' = -V'' = O((1 + |x|)^{-2})$ there by Lemma B.1(b). Differentiating $R^{\bar{m}}$ three times proves (B.3). On $[0, \infty)$ we have $\mathfrak{g} \leq 1$ and $(\log R)' = -z - R$, whose first two derivatives are $-1 + V$ and V' ; hence $|\mathfrak{g}^{(j)}(x)| \leq C(1 + x)^j$ for $x \geq 0$ and $j \leq 3$. These observations include $\bar{m} = 0$.

For $\mu \leq 0$ and $0 \leq \sigma \leq \sqrt{2}$, Gaussian averaging preserves these estimates:

$$\begin{aligned} c(1 + |\mu|)^{\bar{m}} &\leq \mathbb{E}\mathfrak{g}(\mu + \sigma G) \leq C(1 + |\mu|)^{\bar{m}}, \\ |\mathbb{E}\mathfrak{g}^{(j)}(\mu + \sigma G)| &\leq C_j(1 + |\mu|)^{\bar{m}-j} \quad (1 \leq j \leq 3). \end{aligned} \quad (\text{B.4})$$

For the lower bound, restrict to $\{G \leq 0\}$ and use that $\mathfrak{g}$ is nonincreasing. For the upper bound, use $\mathfrak{g}(x) \leq C(1 + x_-)^{\bar{m}}$ and Gaussian moments. For the derivative bounds with $|\mu| \geq 1$, split at $\mu + \sigma G = \mu/2$. On the lower side $|\mu + \sigma G| \geq |\mu|/2$, and (B.3) applies; on the upper side use the polynomial bounds and the Gaussian tail $\mathbb{P}(\sigma G > |\mu|/2) \leq \exp(-\mu^2/16)$. For $|\mu| \leq 1$ uniform Gaussian moments suffice. Thus all constants in (B.4) are independent of $\bar{m}$, μ , and σ .

Since $f_M = m\phi^m R^{\bar{m}}$ and $\phi^m = (2\pi)^{\bar{m}/2} m^{-1/2} \phi_{1/m}$, completing the square in the Gaussian convolution defining f gives

$$f(z) = C_m \phi_{1/m}(z) H(z), \quad C_m := \sqrt{m}(2\pi)^{\bar{m}/2}, \quad H(z) := \mathbb{E} \mathfrak{g}(\sqrt{\beta}z + \sigma_\beta G). \quad (\text{B.5})$$

For $0 < \beta < 1$ this follows by writing the density of $\sqrt{\beta}M_m$ as $C_m \phi_{\beta/m}(x) \mathfrak{g}(x/\sqrt{\beta})$ and using $\phi_{\beta/m}(x) \phi_{(1-\beta)/m}(z-x) = \phi_{1/m}(z) \phi_{\beta(1-\beta)/m}(x - \beta z)$. The same identity holds at $\beta \in \{0, 1\}$ by continuity. By (B.4) with $\mu = \sqrt{\beta}z$ and $\sigma = \sigma_\beta \leq \sqrt{2}$, and since $\sqrt{\beta}/(1 + \sqrt{\beta}t) \leq 1/(1+t)$ for $t \geq 0$, we have for $z \leq 0$ and $t, u \geq 0$

$$|H^{(j)}(z)| \leq \frac{C_j H(z)}{(1 + |z|)^j} \quad (1 \leq j \leq 3), \quad \frac{H(-t-u)}{H(-t)} \leq C \left(1 + \frac{u}{1+t}\right)^{\bar{m}}. \quad (\text{B.6})$$

Also H is nonincreasing.

For $t \geq 1$, integrate (B.5) over $z \leq -t$ and substitute $z = -t - \rho/(mt)$. Since $\frac{m}{2}(t + \frac{\rho}{mt})^2 = \frac{m}{2}t^2 + \rho + \frac{\rho^2}{2mt^2}$,

$$\begin{aligned} F(-t) &= \frac{C_m \sqrt{m/(2\pi)}}{mt} \exp(-mt^2/2) \Xi(t), \\ \Xi(t) &:= \int_0^\infty \exp\left(-\rho - \frac{\rho^2}{2mt^2}\right) H\left(-t - \frac{\rho}{mt}\right) d\rho. \end{aligned} \quad (\text{B.7})$$

We claim that

$$cH(-t) \leq \Xi(t) \leq CH(-t), \quad |\Xi^{(j)}(t)| \leq C_j t^{-j} H(-t) \quad (1 \leq j \leq 3). \quad (\text{B.8})$$

For the lower bound, restrict the integral to $0 \leq \rho \leq 1$, where the exponential is at least e^{-2} , and use $H(-t - \rho/(mt)) \geq H(-t)$. For the upper bounds let $\xi(t) = -t - \rho/(mt)$. Then $|\xi'| \leq 1 + 2\rho$, $|\xi''| \leq 4\rho t^{-3}$, and $|\xi'''| \leq 12\rho t^{-4}$. Together with $|\xi| \geq t$ and (B.6), the chain rule bounds the j th derivative of $H(\xi(t))$ by $C_j t^{-j} P_j(\rho) H(\xi(t))$, where P_j is a fixed polynomial. The j th derivative of $\exp(-\rho^2/(2mt^2))$ has the same type of bound, with that exponential in place of $H(\xi(t))$. Finally (B.6) bounds $H(\xi(t))/H(-t)$ by $C(1 + 2\rho)^{\bar{m}}$. The integrable majorants $Ce^{-\rho} P(\rho)$ justify each differentiation and prove (B.8). Consequently $|(\log \Xi)^{(j)}(t)| \leq C_j t^{-j}$ for $j = 1, 2, 3$. Taking logarithms in (B.7),

$$\log F(-t) = \text{constant} - \frac{m}{2} t^2 - \log t + \log \Xi(t).$$

Its second derivative in t differs from $-m$ by $O(t^{-2})$, and its third derivative is $O(t^{-3})$, uniformly in m and β . Since $v_m = -m^{-1}(\log F)''$, this proves (V3) and (V5) for $z \leq -1$. The compact interval $[-1, 0]$ is covered by (V0).

STEP 4. Proof of (V2). The bounds on $\mathbf{g}, \mathbf{g}', \mathbf{g}''$ above give $|\mathbf{g}^{(j)}(x)| \leq C(1 + |x|)^2$ for all x and $j \leq 2$, hence $|H^{(j)}(z)| \leq C(1 + z)^2$ for $z \geq 0$ and $j \leq 2$. Thus (B.5) gives $|f^{(j)}(z)| \leq C \exp(-cz^2)$ for $z \geq 0$ and $j \leq 2$. Since $F \geq F(0) \geq 1/4$ there by (V1), the formulas for v_m and $\partial_z v_m$ in the proof of (V0) prove (V2).

STEP 5. Proof of (V4). By (V0) and compactness, $1 - v_m \geq c_1 > 0$ on $[-1, 0]$, uniformly in m and β . Let $z < -1$, and first let $m \in [m_0, 1)$ and $\beta \in (0, 1)$. We use the realization (B.2) at level $q = q_1$, where $\beta_q = \beta$, and the point $x = \sqrt{\lambda(q)} z$. Choose $q' \in [q, Q]$ with

$$\bar{\beta} := \lambda(q')/\lambda(q) = \max\{\beta, z^{-2}\},$$

which is possible because $\lambda(\cdot)/\lambda(q)$ decreases continuously from 1 to β on $[q, Q]$. Let $X = X^{q,x}$ be the Parisi diffusion started at (q, x) . The generator identity $\mathcal{L}(1 - v) = (m/\lambda)v(1 - v)$ (Lemma 2.9(a)) and Itô's formula on the constant block $[q, q']$ show that $(1 - v(t, X_t)) \exp(-\int_q^t (m/\lambda)v)$, $t \in [q, q']$, is a local martingale bounded by one, hence a martingale. Since $v < 1$ and $\int_q^{q'} m/\lambda = \log(1/\bar{\beta})$,

$$1 - v(q, x) = \mathbb{E} \left[(1 - v(q', X_{q'})) \exp \left(- \int_q^{q'} \frac{m}{\lambda} v dt \right) \right] \geq \bar{\beta} \mathbb{E} [1 - v(q', X_{q'})].$$

Moreover $\psi = \lambda b + x > 0$ (Lemma 2.9(b), with $\kappa = 0$), and Itô's formula with $\lambda' = -m$ gives

$$d\left(\frac{X_t}{\lambda(t)}\right) = \frac{m\psi(t, X_t)}{\lambda(t)^2} dt + \frac{dB_t}{\lambda(t)}, \quad \text{hence} \quad Z' := \frac{X_{q'}}{\sqrt{\lambda(q')}} \geq \sqrt{\bar{\beta}} z + N,$$

where $N = \sqrt{\lambda(q')} \int_q^{q'} dB_t / \lambda(t)$ is centered Gaussian with variance $(1 - \bar{\beta})/m$. By Lemma B.4, $v(q', \cdot) = v_m(\cdot / \sqrt{\lambda(q')}; \beta / \bar{\beta})$, and v_m is decreasing in z (Lemma 2.4(c)). On $\{N \geq 0\}$, whose probability is at least $\frac{1}{2}$,

$$1 - v_m(Z'; \beta / \bar{\beta}) \geq 1 - v_m(\sqrt{\bar{\beta}} z; \beta / \bar{\beta}).$$

If $\bar{\beta} = z^{-2}$, then $\sqrt{\bar{\beta}} z = -1$ and the right side is at least c_1 , which gives $1 - v_m(z; \beta) \geq c_1 / (2z^2)$. Otherwise $\bar{\beta} = \beta > z^{-2}$ and $q' = Q$, so $v_m(\cdot; 1) = V$ and Lemma B.1(b) gives

$$1 - v_m(z; \beta) \geq \frac{\beta}{2} (1 - V(\sqrt{\beta} z)) \geq \frac{c\beta}{2(1 + \beta z^2)} \geq \frac{c}{4z^2},$$

because $1 + \beta z^2 \leq 2\beta z^2$. In both cases $1 - v_m(z; \beta) \geq c / (1 + z^2)$. The cases $m = 1$ or $\beta \in \{0, 1\}$, where v_m is V or $V(\sqrt{m} \cdot)$, follow from Lemma B.1(b). This proves (V4).

STEP 6. *The bound on r_m .* By (V3), (V0), and compactness, $v_m \geq c > 0$ on $(-\infty, 0]$. Together with (V4) and (V5), this gives $r_m \leq C(1 + |z|)^{-6}(1 + z^2) \leq C(1 + z^4)^{-1}$ for $z \leq 0$, and completes the proof of Proposition B.5. $\square$

B.4 Proof of the breakpoint bound

We first record Chebyshev's association inequality in the form needed here: an average against a nonincreasing weight is bounded by the largest average over half-lines. For a nondecreasing integrand it reduces to the classical inequality (see [21]).

Lemma B.6. *Let ν be a finite measure on $\mathbb{R}$ with a positive density, let $\tilde{g} \geq 0$ be nonincreasing with $0 < \int \tilde{g} d\nu < \infty$, and let $h \geq 0$ be measurable with $\int h d\nu < \infty$. For $\tau \in (-\infty, \infty]$ put $\mathcal{R}_h(\tau) := \nu((-\infty, \tau))^{-1} \int_{(-\infty, \tau)} h d\nu$. Then*

$$\frac{\int h \tilde{g} d\nu}{\int \tilde{g} d\nu} \leq \sup_{\tau \in (-\infty, \infty]} \mathcal{R}_h(\tau).$$

If h is nondecreasing, then $\mathcal{R}_h(\tau) \leq \mathcal{R}_h(\infty)$ for all τ , so the left side is at most $\int h d\nu / \nu(\mathbb{R})$.

Proof. Changing $\tilde{g}$ on a countable set does not affect the integrals, so we may assume that $\tilde{g}$ is right-continuous. Then $\tilde{g}(z) = \tilde{g}(\infty) + \varsigma((z, \infty))$ for the measure $\varsigma \geq 0$ with $\varsigma((z_1, z_2]) = \tilde{g}(z_1) - \tilde{g}(z_2)$. By Tonelli's theorem, for every measurable $k \geq 0$,

$$\int k \tilde{g} d\nu = \tilde{g}(\infty) \int k d\nu + \int \varsigma(d\tau) \int_{(-\infty, \tau)} k d\nu.$$

Taking $k = h$ and $k = 1$ shows that the ratio is a weighted average of the values $\mathcal{R}_h(\tau)$, $\tau \in \mathbb{R}$, with weights $\nu((-\infty, \tau)) \zeta(d\tau)$, and of $\mathcal{R}_h(\infty)$, with weight $\tilde{g}(\infty)\nu(\mathbb{R})$. If h is nondecreasing, then $h(z) \leq h(\tau)$ for $z < \tau$ and $h(z) \geq h(\tau)$ for $z \geq \tau$, so

$$\nu([\tau, \infty)) \int_{(-\infty, \tau)} h d\nu \leq \nu((-\infty, \tau)) \int_{[\tau, \infty)} h d\nu,$$

which is equivalent to $\mathcal{R}_h(\tau) \leq \mathcal{R}_h(\infty)$. $\square$

Proof of Proposition 6.12. We divide the proof into four steps.

STEP 1. The law of the scale-free variable. Let $\varepsilon > 0$ be given; $\bar{\eta} \in (0, 1/4]$ will be chosen at the end. Let $|\kappa| \leq \bar{\eta}$ and $\gamma \in \mathcal{U}$.

Case (b). Suppose that $\gamma \equiv m$ on $[p, Q)$ with $1 - \bar{\eta} \leq m < 1$ and $\lambda(p) \leq \bar{\eta}$, and let $q \in [p, Q)$. Then $m \geq 3/4 \geq m_0$. Since $\gamma \geq m$ on $[p, 1]$, we have $\lambda(p) \geq m(1 - p)$, so $1 - q \leq 1 - p \leq 2\bar{\eta} \leq \frac{1}{2}$ and $q \geq \frac{1}{2}$; also $\lambda := \lambda(q) \leq \bar{\eta}$. Put $\beta := \delta/\lambda$. By Lemma 2.7, X_q has density $e^{mu(q,x)}\phi_q(x)g_q(x)$ with g_q nonincreasing, and by Lemma B.4, $e^{mu(q,x)} = F_m(z; \beta)$ with $z = (x - \kappa)/\sqrt{\lambda}$. Hence $Z_q := (X_q - \kappa)/\sqrt{\lambda}$ has density proportional to $F_m(z; \beta)\mathcal{G}(z)\tilde{g}(z)$, where

$$\mathcal{G}(z) := \frac{\phi_q(\kappa + \sqrt{\lambda}z)}{\phi_q(\kappa)} = \exp\left(-\frac{2\kappa\sqrt{\lambda}z + \lambda z^2}{2q}\right), \quad \tilde{g}(z) := g_q(\kappa + \sqrt{\lambda}z),$$

and $\tilde{g}$ is nonincreasing. Since $2|\kappa|\sqrt{\lambda} \leq 2\bar{\eta}^{3/2} \leq 1 \leq 2q$ and $\lambda \leq 2q$, we have $\exp(-|z| - z^2) \leq \mathcal{G}(z) \leq \exp(|z|)$, and $\mathcal{G} \rightarrow 1$ locally uniformly as $\bar{\eta} \rightarrow 0$. By Lemma B.4 and (6.18),

$$\begin{aligned} \tilde{x}_\gamma(q) &= \frac{\mathbb{E}(\partial_z v_m)^2(Z_q; \beta)}{2\mathbb{E}[v_m^2(1 - v_m)](Z_q; \beta)} = \frac{1}{2} \frac{\int r_m \tilde{g} d\nu}{\int \tilde{g} d\nu}, \\ \nu(dz) &:= F_m(z; \beta) \mathcal{G}(z) (v_m^2(1 - v_m))(z; \beta) dz. \end{aligned}$$

The measure ν is finite and has a positive density: $F_m \leq 2\exp(-z^2/16)$ for $z \leq 0$ and $v_m \leq C\exp(-cz^2)$ for $z \geq 0$ by Proposition B.5, while $\mathcal{G} \leq \exp(|z|)$. Also $0 < \int \tilde{g} d\nu < \infty$, since $\int \tilde{g} F_m \mathcal{G} dz$ is a normalizing constant of the density of Z_q .

Case (a). Suppose that $\delta \leq \bar{\eta}$ and take $q = Q$. Then $Q \geq \frac{1}{2}$, $\gamma(Q) = 1$, $u(Q, x) = \log \Phi(z)$ with $z = (x - \kappa)/\sqrt{\delta}$, and $v(Q, x) = V(z)$. By Lemma 2.7, the same computation gives $\tilde{x}_\gamma(Q) = \frac{1}{2} \int r \tilde{g} d\nu_Q / \int \tilde{g} d\nu_Q$ with $\nu_Q(dz) := \Phi(z)\mathcal{G}(z)V^2(1 - V)(z) dz$, where $\lambda = \delta$ in $\mathcal{G}$.

STEP 2. The main term. Recall $r = r_1 = (V')^2/(V^2(1 - V))$. In case (b), $\tilde{x}_\gamma(q) \leq \frac{1}{2}\mathcal{M} + \frac{1}{2}\mathcal{E}$, where

$$\mathcal{M} := \frac{\int r \tilde{g} d\nu}{\int \tilde{g} d\nu}, \quad \mathcal{E} := \frac{\int |r_m - r| \tilde{g} d\nu}{\int \tilde{g} d\nu}.$$

In case (a), $\tilde{x}_\gamma(Q) = \frac{1}{2}\mathcal{M}$ with ν_Q in place of ν , and there is no error term. Since r is nondecreasing (Lemma B.2) and $\tilde{g}$ is nonincreasing, Lemma B.6 gives $\mathcal{M} \leq \int r \, d\nu / \nu(\mathbb{R})$. We claim that

$$\frac{\int r \, d\nu}{\nu(\mathbb{R})} \longrightarrow \frac{\int \Phi(V')^2}{\int \Phi V^2(1-V)} = 2x_\Phi \quad \text{as } \bar{\eta} \rightarrow 0,$$

uniformly over all admissible (κ, γ, m, q) , and likewise for ν_Q . Otherwise there are $\varepsilon_1 > 0$ and admissible sequences with $\bar{\eta}_n \rightarrow 0$ that violate the claim by ε_1 . Along a subsequence $\beta_n \rightarrow \beta_\infty \in [0, 1]$, while $m_n \rightarrow 1$, $\kappa_n \rightarrow 0$, $\lambda_n \rightarrow 0$, and $q_n \in [\frac{1}{2}, 1]$. By (V0), $F_{m_n}(\cdot; \beta_n) \rightarrow F_1(\cdot; \beta_\infty) = \Phi$ and $v_{m_n}(\cdot; \beta_n) \rightarrow V$ pointwise, and $\mathcal{G}_n \rightarrow 1$. The densities of $r\nu_n$ and ν_n are dominated by an integrable function independent of n : for $z \leq 0$ by $C \exp(|z| - z^2/16)$, using $r \leq C$ there (Lemma B.1(b)); for $z \geq 0$ by $C(1+z)^2 \exp(|z| - cz^2)$, using $r \leq C(1+z)^2$ (Lemma B.1(a)) and (V2). Dominated convergence gives the claim, since $\int \Phi V^2(1-V) > 0$ and $\int \Phi r V^2(1-V) = \int \Phi(V')^2$. The case of ν_Q is the same with $m = 1$.

STEP 3. The error term. By Lemma B.6 with $h = |r_m - r|$, $\mathcal{E} \leq \sup_\tau \mathcal{R}_{|r_m - r|}(\tau)$. Fix $T \geq 1$ and $T' \geq 1$.

Far left. By Proposition B.5 and Lemma B.1(b),

$$\mathfrak{t}_1(T) := \sup\{r_{m'}(z; \beta) + r(z) : z \leq -T, m' \in [m_0, 1], \beta \in [0, 1]\} \leq CT^{-4}.$$

For $\tau \leq -T$, bounding the numerator of $\mathcal{R}_{|r_m - r|}(\tau)$ by $\mathfrak{t}_1(T) \nu((-\infty, \tau))$ gives $\mathcal{R}_{|r_m - r|}(\tau) \leq \mathfrak{t}_1(T)$. For $\tau > -T$, the part of the numerator over $(-\infty, -T)$ is at most $\mathfrak{t}_1(T) \nu((-\infty, -T)) \leq \mathfrak{t}_1(T) \nu((-\infty, \tau))$, so after division its contribution is again at most $\mathfrak{t}_1(T)$. No reciprocal of a small Gaussian mass appears.

Bulk and right tail. For $\tau > -T$ the denominator satisfies $\nu((-\infty, \tau)) \geq \nu((-T - 1, -T)) \geq c_T > 0$, uniformly: on $[-T - 1, -T]$ the factors F_m and $v_m^2(1 - v_m)$ are bounded below by (V0) and compactness, and $\mathcal{G} \geq \exp(-|z| - z^2)$. The remaining part of the numerator is at most

$$\omega_{T, T'}(1 - m) \nu(\mathbb{R}) + \int_{T'}^\infty (r_m + r) \, d\nu,$$

where

$$\omega_{T, T'}(\varepsilon') := \sup\{|r_{m'}(z; \beta) - r(z)| : 1 - \varepsilon' \leq m' \leq 1, \beta \in [0, 1], z \in [-T, T']\}.$$

Here $\nu(\mathbb{R}) \leq C$. By (V0), $r_{m'}$ is jointly continuous on the compact set $[m_0, 1] \times [0, 1] \times [-T, T']$, its denominator being positive there, and $r_1(\cdot; \beta) = r$ for every β ; hence $\omega_{T, T'}(\varepsilon') \rightarrow 0$ as $\varepsilon' \rightarrow 0$. The exact cancellation $r_m \, d\nu = F_m \mathcal{G}(\partial_z v_m)^2 \, dz$ and (V2) give a uniform Gaussian majorant for the r_m term, and Lemma B.1(a) and (V2) give one for the r term. Hence

$$\int_{T'}^\infty (r_m + r) \, d\nu \leq C \int_{T'}^\infty (1 + z)^2 \exp(z - cz^2) \, dz =: \mathfrak{t}_2(T') \longrightarrow 0 \quad (T' \rightarrow \infty).$$

Consequently

$$\mathcal{E} \leq \mathfrak{t}_1(T) + c_T^{-1}(C\omega_{T,T'}(\bar{\eta}) + \mathfrak{t}_2(T')).$$

STEP 4. *Choice of $\bar{\eta}$.* Given $\varepsilon > 0$, first choose T with $\mathfrak{t}_1(T) \leq \varepsilon/2$, then T' with $\mathfrak{t}_2(T') \leq c_T\varepsilon/4$, and then $\bar{\eta} \in (0, 1/4]$ so small that $C\omega_{T,T'}(\bar{\eta}) \leq c_T\varepsilon/4$ and, by the claim above, $\int r \, d\nu/\nu(\mathbb{R}) \leq 2x_\Phi + \varepsilon$ and $\int r \, d\nu_Q/\nu_Q(\mathbb{R}) \leq 2x_\Phi + \varepsilon$ for all admissible data. Then $\mathcal{E} \leq \varepsilon$ and $\mathcal{M} \leq 2x_\Phi + \varepsilon$. Hence $\tilde{x}_\gamma(q) \leq x_\Phi + \varepsilon$ in case (b), and $\tilde{x}_\gamma(Q) \leq x_\Phi + \varepsilon/2$ in case (a). This proves Proposition 6.12. $\square$

Proofs for the flat limit

This appendix proves the technical statements of Chapter 7. §C.1 records the decay of the derivatives of V , which Chapters 7 and 8 use in addition to Lemma 2.1. §C.2 proves the uniform tail estimates of Lemma 7.2: a dilation turns every profile of a compact class into one fixed instance of the finite-cutoff problem of Chapter 2, whose asymptotics give parts (a) and (d), and a transport estimate for the rescaled diffusion gives parts (b) and (c). §C.3 proves the compact-class expansion Theorem 7.3, and §C.4 proves the differentiability of I along ordered rays and the second-variation identity of Proposition 7.10.

C.1 Tails of V and its derivatives

We use the facts of Lemma 2.1, in particular $R' = -V$, $\bar{R}' = 1 - V$, $z_- < R(z) \leq z_- + R(0)$, and the identity (2.2) for V'' . We also need the decay of all derivatives of V .

Lemma C.1. (a) *There is C with $0 < 1 - V(z) \leq C(1 + z^2)^{-1}$ for $z \leq 0$, and for every $k \geq 1$ there is C_k with $|V^{(k)}(z)| \leq C_k(1 + |z|)^{-2-k}$ for $z \leq 0$. For $z \geq 0$ and $k \geq 0$, $|V^{(k)}(z)| \leq C_k(1 + z)^{k+1}\phi(z)$.*

(b) *The numbers $A_V = \int V'^2$, $B_V = \int V^2(1 - V)$, and $\int V''^2$ are finite and positive.*

Proof. (a) Substituting $u = z - w$ in $\Phi(z) = \int_{-\infty}^z \phi(u) du$ gives $\Phi(z) = \phi(z) \int_0^\infty e^{zw - w^2/2} dw$. Hence $\log \Phi = \log \phi + \Lambda$ with $\Lambda(z) = \log \int_0^\infty e^{zw - w^2/2} dw$. The function $u \mapsto \Lambda(z + u) - \Lambda(z)$ is the cumulant generating function of the variable T_z of §2.1, whose density is proportional to $e^{zw - w^2/2}$ on $[0, \infty)$, so $\Lambda^{(j)}(z)$ is the j -th cumulant $\text{cum}_j(T_z)$. Since $(\log \Phi)'' = R' = -V$ and $(\log \phi)'' = -1$,

$$1 - V = \text{cum}_2(T_z), \quad V^{(k)} = -\text{cum}_{k+2}(T_z) \quad (k \geq 1).$$

For $z = -t \leq -1$, T_{-t} has the law of S_t/t , where S_t has density proportional to $e^{-s - s^2/(2t^2)}$ on $[0, \infty)$. The moments of S_t are bounded uniformly in $t \geq 1$: the numerator of $\mathbb{E}S_t^i$ is at most $\int_0^\infty s^i e^{-s} ds = i!$, and the normalizing integral is at least $\int_0^1 e^{-s - 1/2} ds \geq e^{-1}$. Cumulants are polynomials in the moments, and $\text{cum}_j(T_{-t}) = t^{-j} \text{cum}_j(S_t)$; hence $|\text{cum}_j(T_{-t})| \leq C_j t^{-j}$ for $t \geq 1$. This gives the bounds for $z \leq -1$; on $[-1, 0]$ the functions are bounded, and $1 - V > 0$. For $z \geq 0$ we use $R \leq 2\phi$ and $R \leq R(0) < 1$ (Lemma 2.1(a),(e)). By induction, using $R' = -zR - R^2$, $V^{(k)} = P_k(z, R)$ for a polynomial

P_k of total degree at most $k + 2$ each of whose monomials contains R : this holds for $V = zR + R^2$, and $\frac{d}{dz}P(z, R) = \partial_z P - (zR + R^2)\partial_R P$. A monomial $z^i R^j$ of P_k has $j \geq 1$ and $i \leq k + 1$, so $|z^i R^j| \leq z^i R \leq 2(1 + z)^{k+1} \phi(z)$.

(b) The three integrands are nonnegative and continuous, and integrable by (a): on $(-\infty, 0]$ they are bounded by $C(1 + |z|)^{-6}$, $1 - V$, and $C(1 + |z|)^{-8}$, and on $[0, \infty)$ they have Gaussian decay. They are not identically zero, because $0 < V < 1$, $V' < 0$, and V' is not constant (it is negative and tends to zero at $\pm\infty$). $\square$

C.2 Dilation and uniform tails

The rescaled equation (7.3) does not involve e . The following dilation identifies it with the finite-cutoff problem of Lemma 2.4 at margin zero, with a cutoff that depends only on the class.

Lemma C.2. *Let $0 < s_{\text{lo}} \leq s_{\text{hi}}$ and let ζ be a profile with $\zeta = 1$ on $[0, s_{\text{lo}}]$. Put $r = 1 - s_{\text{lo}}/s_{\text{hi}}$ and let $\tilde{\gamma}(\tilde{q}) = \zeta(s_{\text{hi}}(1 - \tilde{q}))$ (right-continuous version). Then $\tilde{\gamma} \in \mathcal{U}_r$, and $\|\tilde{\gamma} - \tilde{\gamma}'\|_{L^1(0,1)} = s_{\text{hi}}^{-1} \|\zeta - \zeta'\|_{L^1(0,s_{\text{hi}})}$ for two such profiles. If $\tilde{u}$ is the solution of Lemma 2.4 for $\tilde{\gamma}$ at margin 0, then*

$$U(s, y) = \tilde{u}(1 - s/s_{\text{hi}}, y/\sqrt{s_{\text{hi}}})$$

solves (7.3) on $[s_{\text{lo}}, s_{\text{hi}}]$, with $U(s_{\text{lo}}, y) = \log \Phi(y/\sqrt{s_{\text{lo}}})$. The Parisi diffusion $\tilde{X}$ of $\tilde{\gamma}$ corresponds, through $Y = \sqrt{s_{\text{hi}}} \tilde{X}$, to the rescaled diffusion

$$dY = \zeta(s) U_y(s, Y) d\zeta + dB_\zeta, \quad \zeta = s_{\text{hi}} - s, \quad (\text{C.1})$$

run in the direction of decreasing s . Finally, if $\gamma \in \mathcal{U}$, $e > 0$, $\kappa \in \mathbb{R}$, $e^2 s_{\text{hi}} < 1$, and $\zeta_\gamma = \zeta$ on $[0, s_{\text{hi}}]$, then $U(s, y) = u_\gamma(1 - e^2 s, \kappa + ey)$ for $s \in [s_{\text{lo}}, s_{\text{hi}}]$.

Proof. With $\tilde{q} = 1 - s/s_{\text{hi}}$ and $\tilde{x} = y/\sqrt{s_{\text{hi}}}$ we have $\partial_{\tilde{q}} \tilde{u} = -s_{\text{hi}} U_s$, $\tilde{u}_{\tilde{x}} = \sqrt{s_{\text{hi}}} U_y$ and $\tilde{u}_{\tilde{x}\tilde{x}} = s_{\text{hi}} U_{yy}$, so the Parisi equation $\tilde{u}_{\tilde{q}} = -\frac{1}{2}(\tilde{u}_{\tilde{x}\tilde{x}} + \tilde{\gamma} \tilde{u}_{\tilde{x}}^2)$ is (7.3). The function $\tilde{\gamma}$ is nondecreasing because ζ is nonincreasing, and $\tilde{\gamma} = 1$ on $[r, 1]$ because $\zeta = 1$ on $[0, s_{\text{lo}}]$. At $\tilde{q} = r$ the terminal datum of Lemma 2.4 is $\log \Phi(\tilde{x}/\sqrt{1-r}) = \log \Phi(y/\sqrt{s_{\text{lo}}})$. The identity for the L^1 distance is the substitution $\tilde{q} = 1 - s/s_{\text{hi}}$, and the diffusion statement is Brownian scaling. For the last statement, the affine map $(\tilde{q}, \tilde{x}) \mapsto (1 - e^2 s_{\text{hi}}(1 - \tilde{q}), \kappa + e\sqrt{s_{\text{hi}}} \tilde{x})$ carries each Gaussian step of the Cole–Hopf recursion for $\tilde{\gamma}$ at margin 0 into the corresponding step for γ at margin κ on $[1 - e^2 s_{\text{hi}}, 1)$, and at level $1 - e^2 s_{\text{lo}} \geq Q_\gamma$ both solutions are $\log \Phi$ of the normalized position, by (2.9). This proves the claim for step profiles, and the general case follows by the L^1 limit that defines the solution (Definition 2.2(iii)). $\square$

Consequently Lemmas 2.4, 2.9, and 2.10 apply to U_ζ , with constants that depend only on $(s_{\text{lo}}, s_{\text{hi}})$. We use them in the following rescaled form, for $\zeta \in \mathcal{K}(d_1, S)$ (so $s_{\text{lo}} = d_1$

and $s_{\text{hi}} = S$), levels $s \in [d_1, S]$, and Y the rescaled diffusion (C.1) started at (s, y) and run down to level d_1 . Expectations are over Y . Under the dilation, $\tilde{\lambda} = \ell/S$, $\tilde{b} = \sqrt{S} U_y$, $\tilde{v} = v$, and $\tilde{\psi} = \Psi/\sqrt{S}$, where $\tilde{\psi} = \tilde{\lambda}\tilde{b} + \tilde{x}$ is the function of (2.21).

- (R1) $U_y(s', Y_{s'})$ and $\Psi(s', Y_{s'})$ are martingales. At the wall, $U_y(d_1, y) = R(y/\sqrt{d_1})/\sqrt{d_1}$ and $\Psi(d_1, y) = \sqrt{d_1} \bar{R}(y/\sqrt{d_1})$.
- (R2) $\Psi > 0$, $\Psi_y = 1 - v \in (0, 1)$, $\Psi_{yy} = -v_y \geq 0$, and $v \leq U_y \Psi \leq (s/\ell)v$.
- (R3) $\hat{Y} = Y/\ell$ satisfies $d\hat{Y} = (\zeta\Psi/\ell^2) d\zeta + \ell^{-1} dB_\zeta$.
- (R4) If ζ is a step profile, then

$$\begin{aligned} 1 - v(s, y) &= \mathbb{E} \left[\left(1 - V(Y_{d_1}/\sqrt{d_1}) \right) e^{-\int_{d_1}^s \frac{\zeta}{\ell} v} \right], & v(s, y) &\leq \frac{\ell(s)}{d_1} \mathbb{E} V(Y_{d_1}/\sqrt{d_1}), \\ v_y(s, y) &= d_1^{-1/2} \mathbb{E} \left[V'(Y_{d_1}/\sqrt{d_1}) e^{-\int_{d_1}^s \frac{\zeta}{\ell} (3v-1)} \right], \\ v_{yy}(s, y) &= d_1^{-1} \mathbb{E} \left[V''(Y_{d_1}/\sqrt{d_1}) e^{-\int_{d_1}^s \frac{\zeta}{\ell} (4v-1)} \right] \\ &\quad - 3 \mathbb{E} \int_{d_1}^s \frac{\zeta(s')}{\ell(s')} v_y(s', Y_{s'})^2 e^{-\int_{s'}^s \frac{\zeta}{\ell} (4v-1)} ds'. \end{aligned}$$

Here each integral in an exponent is taken along the path, over the levels indicated. Since $\int_{d_1}^s \zeta/\ell = \log(\ell(s)/d_1) \leq \log(S/d_1)$ and $0 < v < 1$, every weight lies in $[(d_1/S)^3, S/d_1]$.

(R1) and (R2) are $\mathcal{L}b = 0$, $\mathcal{L}\psi = 0$, $\psi_x = 1 - v$, $\psi > 0$, and $v_x < 0$ from Lemma 2.9(a),(b) and Lemma 2.4(c), together with the sandwich (2.26), read through the dilation; the wall values come from $U(d_1, y) = \log \Phi(y/\sqrt{d_1})$ and $\ell(d_1) = d_1$. (R3) is Itô's formula, since $d\ell/d\zeta = -\zeta$. For (R4), write $\mathcal{G}_s = \frac{1}{2}\partial_{yy} + \zeta U_y \partial_y$. Differentiating (7.3) gives, on each block of a step profile,

$$(\partial_s - \mathcal{G}_s)v = \frac{\zeta}{\ell} v(1 - v), \quad (\partial_s - \mathcal{G}_s)(-U_{yy}) = -\zeta U_{yy}^2 \leq 0,$$

and, using $\partial_y(\mathcal{G}_s f) = \mathcal{G}_s f_y - (\zeta v/\ell) f_y$,

$$(\partial_s - \mathcal{G}_s)v_y = -\frac{\zeta}{\ell}(3v - 1)v_y, \quad (\partial_s - \mathcal{G}_s)v_{yy} = -\frac{\zeta}{\ell}(4v - 1)v_{yy} - 3\frac{\zeta}{\ell}v_y^2.$$

These are the rescaled forms of Lemma 2.9(a). On each block U is smooth with derivatives of polynomial growth, the drift ζU_y is Lipschitz, and v, v_y, v_{yy} are continuous in s at the breakpoints. Itô's formula on each block and the Markov property give the Feynman–Kac representations, with the values $v(d_1, y) = V(y/\sqrt{d_1})$, $v_y(d_1, y) = V'(y/\sqrt{d_1})/\sqrt{d_1}$ and $v_{yy}(d_1, y) = V''(y/\sqrt{d_1})/d_1$ at the wall, and with $-U_{yy}(d_1, \cdot) = V(\cdot/\sqrt{d_1})/d_1$ for the upper bound on $v = -\ell U_{yy}$. The analogous bounds for the Sherrington–Kirkpatrick model are cited before Lemma 7.7.

The next lemma controls where the rescaled diffusion travels. Paths started far to the left stay far to the left, and paths started to the right stay to the right, with Gaussian probability.

Lemma C.3. *There is $C = C(d_1, S)$ such that for $\zeta \in \mathcal{K}(d_1, S)$, $s \in [d_1, S]$ and Y started at (s, y) the following hold.*

- (a) *If $y \leq -C$, then with probability at least $1 - 2e^{-y^2/C}$, $Y_{s'} \leq y/C$ for all $s' \in [d_1, s]$.*
- (b) *If $y \geq 0$, then with probability at least $1 - e^{-y^2/C}$, $Y_{s'} \geq y/2$ for all $s' \in [d_1, s]$.*
- (c) *$U_y(s', 0) \leq C$ for all $s' \in [d_1, S]$, and $0 < \Psi(s', x)/\ell(s') \leq C$ for $x \leq 0$.*

Proof. (c) Start Y at $(s', 0)$. The drift ζU_y is nonnegative, so Y_{d_1} is at least a centered Gaussian variable N of variance $s' - d_1 \leq S$. By (R1), since R is decreasing and $R(z) \leq R(0) + z_-$ (Lemma 2.1(a)),

$$U_y(s', 0) = \mathbb{E} \frac{R(Y_{d_1}/\sqrt{d_1})}{\sqrt{d_1}} \leq \frac{R(0) + \mathbb{E}N_-/\sqrt{d_1}}{\sqrt{d_1}} \leq \frac{R(0) + \sqrt{S/d_1}}{\sqrt{d_1}} =: C_\Psi.$$

Next, $\partial_x(\Psi/\ell) = (1 - v)/\ell \geq 0$ by (R2), and $\Psi(s', 0) = \ell U_y(s', 0)$; so $0 < \Psi(s', x)/\ell(s') \leq U_y(s', 0) \leq C_\Psi$ for $x \leq 0$.

(a) By (R3) and (c), as long as $Y \leq 0$ the drift of $\hat{Y}$ is $(\zeta/\ell)(\Psi/\ell) \leq C_\Psi \zeta/\ell$. Hence $\hat{Y}_{s'} \leq y/\ell(s) + C_\Psi \int_{s'}^s \zeta/\ell + N_{s'}$ up to the first level at which Y reaches 0, where $N_{s'} = \int \ell^{-1} dB$ is a continuous Gaussian martingale with quadratic variation at most $\int_{d_1}^S \ell^{-2} \leq S/d_1^2$, and $\int \zeta/\ell \leq \log(S/d_1)$. By the reflection principle, the event $\{\sup |N| \leq |y|/(4\ell(s))\}$ has probability at least $1 - 2\exp(-y^2 d_1^2/(32S^3))$, because $\ell(s) \leq S$. On this event, and if $|y| \geq 4SC_\Psi \log(S/d_1)$, we get $\hat{Y} \leq y/(2\ell(s)) < 0$ up to that level; hence Y never reaches 0, and $Y_{s'} = \ell(s')\hat{Y}_{s'} \leq d_1 y/(2S)$ for all $s' \in [d_1, s]$.

(b) The drift is nonnegative, so $Y_{s'}$ is at least y plus a Brownian increment over a time at most S . By the reflection principle, the running minimum of that increment is below $-y/2$ with probability at most $2\Phi(-y/(2\sqrt{S})) \leq e^{-y^2/(8S)}$. $\square$

Proof of Lemma 7.2. All constants depend only on (d_1, S) . Let $\zeta \in \mathcal{K}(d_1, S)$ and let $\tilde{\gamma} \in \mathcal{U}_r$, $r = 1 - d_1/S$, be its dilation (Lemma C.2 with $s_{\text{lo}} = d_1$, $s_{\text{hi}} = S$). Level $s \in [d_1, S]$ corresponds to $\tilde{q} = 1 - s/S \in [0, r]$, and $\tilde{x} = y/\sqrt{S}$. Under the dilation $\tilde{\lambda}(\tilde{q}) = \ell(s)/S$, the top of $\tilde{\gamma}$ is $1 - d/S$ with d as in §7.1, and the substitution $t = 1 - u/S$ gives $\int_{\tilde{q}}^{1-d/S} \tilde{\lambda}^{-1} = \int_d^s \ell^{-1}$. Hence the function ω of Lemma 2.4(e), computed for $\tilde{\gamma}$, is

$$\log \frac{\ell(s)}{S} - \frac{1}{2} \log \frac{d}{S} - \frac{1}{2} \int_d^s \frac{du}{\ell(u)} - \log \sqrt{2\pi} = \omega(s) - \frac{1}{2} \log S,$$

with $\omega(s)$ from (7.4).

(a) By Lemma 2.4(e) for $\tilde{\gamma}$ at margin 0, there are C, x_∞ such that $|\tilde{\lambda}\tilde{b} - (-\tilde{x}) - \tilde{\lambda}/(-\tilde{x})| \leq C|\tilde{x}|^{-3}$ for $\tilde{x} \leq -x_\infty$. Here $\tilde{\lambda}\tilde{b} = \ell U_y/\sqrt{S}$ and $-\tilde{x} = |y|/\sqrt{S}$, and

$\Psi = \ell U_y - |y|$ for $y < 0$. The inequality becomes $|\Psi(s, y) - \ell(s)/|y|| \leq CS^2|y|^{-3}$ for $y \leq -\sqrt{S}x_\infty$.

(d) *Left tail.* The first estimate of Lemma 2.4(e) at $\tilde{q} = 0$, together with $\tilde{x}^2/(2\tilde{\lambda}(0)) = y^2/(2\ell_0)$ and $\log(-\tilde{x}) = \log|y| - \frac{1}{2}\log S$, gives

$$\left| U(S, y) + \frac{y^2}{2\ell_0} + \log|y| - \omega(S) \right| \leq \frac{CS}{y^2} \quad (y \leq -\sqrt{S}x_\infty).$$

For $y \leq -1$ the left side is $|W_\zeta(y)|$. *Bounded range.* By Lemma 2.4(b), $-C(1 + y^2/S) \leq U(S, y) \leq 0$, and $|\omega(S)|$ is bounded by (7.1); hence W_ζ is bounded on $[-\max\{1, \sqrt{S}x_\infty\}, 0]$. *Right tail.* For $y > 0$, $W_\zeta(y) = U(S, y)$, and Lemma 2.4(d) at $\tilde{q} = 0$ gives $0 \geq U(S, y) \geq -Ce^{-y^2/(8S)}$. Together, $|W_\zeta(y)| \leq C(1 + y^2)^{-1}$.

(b) Let first ζ be a step profile, and let C_T be the constant of Lemma C.3. For $y \leq -C_T$, let E be the event of Lemma C.3(a). On E , $Y_{d_1}/\sqrt{d_1} \leq y/(C_T\sqrt{d_1})$, and since $1 - V$ is increasing, Lemma C.1(a) gives $1 - V(Y_{d_1}/\sqrt{d_1}) \leq C(1 + y^2)^{-1}$. The weight in (R4) is at most one, and $\mathbb{P}(E^c) \leq 2e^{-y^2/C_T}$. Hence $1 - v(s, y) \leq C(1 + y^2)^{-1}$ for $y \leq -C_T$, and the same bound holds for $-C_T \leq y \leq 0$ because $1 - v \leq 1$. For $y \geq 0$, the second formula in (R4), Lemma C.3(b) and $V(z) \leq 2(1 + z)\phi(z) \leq Ce^{-z^2/4}$ for $z \geq 0$ give $v(s, y) \leq Ce^{-y^2/C}$.

(c) The same argument applies to the representations of v_y and v_{yy} in (R4), with the weights bounded by S/d_1 , with $|V'(z)| + |V''(z)| \leq C(1 + |z|)^{-3}$ for $z \leq 0$ and Gaussian decay for $z \geq 0$ (Lemma C.1(a)), and, in the source term of v_{yy} , with the bound on v_y just obtained at every level s' : on the event E every $Y_{s'}$ lies below y/C_T .

General profiles in (b) and (c). Let $\zeta_n \in \mathcal{K}(d_1, S)$ be the step approximations (7.2) of ζ , so that $\zeta_n \rightarrow \zeta$ in L^1 . By Lemma C.2 their dilations converge in $L^1(0, 1)$, so by Lemma 2.4(f) the derivatives $\partial_y^k U_{\zeta_n}(s, y)$, $k \leq 4$, converge pointwise; also $\ell_n \rightarrow \ell$ uniformly. Hence v , v_y , v_{yy} of ζ_n converge pointwise to those of ζ , and the bounds, whose constants do not depend on n , pass to the limit.

Continuity. Let $\zeta_n \rightarrow \zeta$ in $\mathcal{K}(d_1, S)$. As in the previous paragraph, $U_{\zeta_n}(S, y) \rightarrow U_\zeta(S, y)$ for every y , and $\omega_n(S) \rightarrow \omega(S)$ because $\ell_n \rightarrow \ell$ uniformly and $\ell_n, \ell \geq d_1$ on $[d_1, S]$. Thus $W_{\zeta_n} \rightarrow W_\zeta$ pointwise, and the majorant $C(1 + y^2)^{-1}$ of (d) gives convergence in $L^1(\mathbb{R})$ by dominated convergence. Finally, $|\ell_n - \ell| \leq \|\zeta_n - \zeta\|_{L^1}$, so $|\tilde{A}(\zeta_n) - \tilde{A}(\zeta)| \leq \frac{1}{8}(2S + d_1)d_1^{-2}\|\zeta_n - \zeta\|_{L^1}$. $\square$

C.3 Proof of the compact-class expansion

Proof of Theorem 7.3. Constants below depend only on (d_1, S) and on $\theta_\pm$. Let e be so small that $e^2S \leq \frac{1}{2}$ and that Lemma 4.6 applies; in particular $|\kappa| \leq e/2$ by Lemma 4.6(a). Let $\gamma \in \mathcal{U}$ have rescaled profile $\zeta \in \mathcal{K}(d_1, S)$, and use the declared level $q_1 = 1 - e^2S \geq \frac{1}{2}$, with $Y_1 = (B_{q_1} - \kappa)/e$ and $k_1 = \kappa/\sqrt{q_1}$. The level q_1 may lie below $q_0 = \min \text{supp } \mu_\gamma$; since $\zeta = 0$ on (S, ∞) , $\gamma = 0$ on $[0, q_1]$. We divide the proof into five steps.

STEP 1. *An exact decomposition.* By Lemma C.2, $U_\zeta(S, y) = u_\gamma(q_1, \kappa + ey)$, so $U_\zeta(S, \cdot)$ is the function $U(S, \cdot)$ of Definition 4.1. Also $\Gamma(S)$ is the function Γ of Definition 4.1 at level S (see after (7.4)), so $\omega(S)$ is the constant $\log \ell_0 - \Gamma(S) - \log \sqrt{2\pi}$ of Proposition 4.5 at that level. Hence Proposition 4.5(a), applied at the declared level q_1 (as permitted by the paragraph after Lemma 4.3), holds with $W = W_\zeta$ and with Y_1 in place of Y_0 . Put $\rho = \alpha\Phi(k_1)$; this is the number ρ of Proposition 4.5 at the declared level. Since $\tilde{\Gamma} = \Gamma(S) - \frac{1}{2} \log \ell_0$ at the level S , $\frac{1}{2}(1 + \rho) \log \ell_0 + (1 - \rho)\tilde{\Gamma} = \rho \log \ell_0 + (1 - \rho)\Gamma(S)$, and Proposition 4.5(a) reads

$$\mathcal{P}_\kappa(\gamma) = K_e(q_1) + \frac{c_e(q_1)}{\ell_0} + \rho \log \ell_0 + (1 - \rho)\Gamma(S) + \alpha \mathbb{E} W_\zeta(Y_1), \quad (\text{C.2})$$

where $c_e(q_1) = q_1(1 - \alpha m(k_1))/(2e^2)$ and $K_e(q_1) = \log e - \alpha \mathbb{E}[\log |Y_1|; Y_1 \leq -1] - \rho \log \sqrt{2\pi}$ are the coefficients of Proposition 4.5 at the level q_1 .

STEP 2. *The coefficients.* By Lemma 4.6(b) at the level q_1 , that is, with s_0 replaced by S :

- (i) $c_e(q_1) = 2\phi_0\theta + e[(\phi_0 - \frac{1}{8\phi_0})\theta + \frac{S}{8}] + O(e^2)$, and $\ell_0 \geq d_1$;
- (ii) $\rho = 1 + \frac{e}{4} + O(e^2)$; since $\ell_0 \in [d_1, S]$ and $\Gamma(S) \in [\frac{1}{2} \log d_1, \frac{1}{2} \log d_1 + S/(2d_1)]$, $\rho \log \ell_0 = \log \ell_0 + \frac{e}{4} \log \ell_0 + O(e^2)$ and $(1 - \rho)\Gamma(S) = -\frac{e}{4}\Gamma(S) + O(e^2)$;
- (iii) $|K_e(q_1) - K_e| \leq Ce^2 L_e$.

STEP 3. *The remainder.* The density of Y_1 is $p_e(y) = (e/\sqrt{q_1})\phi((ey + \kappa)/\sqrt{q_1})$. Since $q_1^{-1/2} \leq 1 + e^2 S$, $\phi_0 - \phi(w) \leq \phi_0 \min\{1, w^2/2\}$ and $(ey + \kappa)^2/q_1 \leq 2e^2(1 + |y|)^2$,

$$\begin{aligned} |\alpha p_e(y) - 2e\phi_0| &\leq 2e\phi_0 \left| (1 + \frac{e}{2})q_1^{-1/2} - 1 \right| + \frac{(2+e)e\phi_0}{\sqrt{q_1}} \min\{1, e^2(1 + |y|)^2\} \\ &\leq Ce[e + \min\{1, e(1 + |y|)\}]. \end{aligned}$$

By Lemma 7.2(d), splitting the integral at $|y| = 1/e$,

$$\left| \alpha \mathbb{E} W_\zeta(Y_1) - 2e\phi_0 \int_{\mathbb{R}} W_\zeta \right| \leq Ce \int_{\mathbb{R}} \frac{e + \min\{1, e(1 + |y|)\}}{1 + y^2} dy \leq Ce^2 L_e.$$

STEP 4. *Summation.* Summing Steps 2 and 3 in (C.2), the terms of order one are $K_e + 2\phi_0\theta/\ell_0 + \log \ell_0 = K_e + \Psi_\theta(\ell_0)$, the other terms are $O(e^2 L_e)$, and the terms of order e are

$$\begin{aligned} g_\theta(\ell_0) + \frac{S}{8\ell_0} - \frac{\Gamma(S)}{4} + 2\phi_0 \int W_\zeta, \\ \frac{S}{8\ell_0} - \frac{\Gamma(S)}{4} = \frac{1}{8} \left[\int_{d_1}^S \left(\frac{1}{\ell_0} - \frac{1}{\ell(s)} \right) ds + \frac{d_1}{\ell_0} - \log d_1 \right] = \tilde{A}(\zeta). \end{aligned}$$

This proves (7.6). The continuity of $\mathcal{P}^{(1)}$ is part of Lemma 7.2.

STEP 5. *The scaling identity.* Let $c = \ell_0$ and $\tilde{\zeta} = \zeta(c \cdot) \in \mathcal{K}^1(d_1/c, S/c)$. The equation (7.3), including its initial condition, is invariant under $s \mapsto cs$, $y \mapsto \sqrt{c}y$, $\zeta \mapsto \zeta(c \cdot)$, so $U_{\tilde{\zeta}}(s, y) = U_{\tilde{\zeta}}(s/c, y/\sqrt{c})$. Also $\ell_{\tilde{\zeta}}(s) = c \ell_{\tilde{\zeta}}(s/c)$, and the substitution $s = cs'$ in (7.4) gives $\Gamma_{\tilde{\zeta}}(S) = \Gamma_{\tilde{\zeta}}(S/c) + \frac{1}{2} \log c$ and $\omega_{\tilde{\zeta}}(S) = \omega_{\tilde{\zeta}}(S/c) + \frac{1}{2} \log c$, where the functions of $\tilde{\zeta}$ are computed with the class parameters $(d_1/c, S/c)$. Hence, with $y = \sqrt{c}y'$,

$$W_{\tilde{\zeta}}(y) = W_{\tilde{\zeta}}(y') + \frac{1}{2} \log c [\mathbf{1}_{\{y' \leq -c^{-1/2}\}} - \mathbf{1}_{\{y' \leq 0\}}] + \log |y'| [\mathbf{1}_{\{y' \leq -c^{-1/2}\}} - \mathbf{1}_{\{y' \leq -1\}}].$$

Put $b = c^{-1/2}$, so that $\frac{1}{2} \log c = -\log b$. The first bracket equals $-\mathbf{1}_{(-b, 0]}$, and its term integrates to $b \log b$. If $b \geq 1$, the second bracket equals $-\mathbf{1}_{(-b, -1]}$, and its term integrates to $-\int_1^b \log u \, du = -(b \log b - b + 1)$; if $b < 1$, it equals $\mathbf{1}_{(-1, -b]}$, and its term integrates to $\int_b^1 \log u \, du = -1 - b \log b + b$. In both cases the two terms integrate to $b - 1$, so $\int W_{\tilde{\zeta}} = \sqrt{c} \int W_{\tilde{\zeta}} + 1 - \sqrt{c}$. The substitution $s = cs'$ also gives $\tilde{A}(\tilde{\zeta}) = A(\tilde{\zeta}) - \frac{1}{8} \log c$. Therefore $\mathcal{P}^{(1)}(\tilde{\zeta}) = A(\tilde{\zeta}) + 2\phi_0 \sqrt{c} (\int W_{\tilde{\zeta}} - 1) - \frac{1}{8} \log c + 2\phi_0$, which is (7.7). $\square$

C.4 Differentiability along ordered rays and the second variation

We work in the setting of §7.5: $T_{\vartheta} = T_0 + \vartheta K$, with T_0, T_1 the inverse profiles of normalized step profiles on a common grid and $K \geq 0$. For $\vartheta \in [0, 1]$ all ζ_{ϑ} lie in the fixed class $\mathcal{K}^1(l_0, \max\{T_0(1), T_1(1)\})$, so the constants of Lemmas 7.2 and C.3 are uniform in ϑ . In the variable l we have $v = lc$, $v_y = lc_y$, and $v_{yy} = lc_{yy}$, and the diffusion of Definition 7.8 is the rescaled diffusion (C.1) after the change of variable $s = T_{\vartheta}(l)$, because $\sigma_{\vartheta} \zeta_{\vartheta}(T_{\vartheta}(l)) = 1$.

Lemma C.4. *For each (l, y) , $\vartheta \mapsto U^{\vartheta}(l, y)$ is of class C^2 , and $\dot{U}, \ddot{U}$ solve (7.17), with*

$$\dot{U}(l, y) = -\frac{1}{2} \mathbb{E} \int_{l_0}^l \dot{\sigma}(\lambda) c(\lambda, Y_{\lambda}) \, d\lambda, \quad \ddot{U}(l, y) = \mathbb{E} \int_{l_0}^l (\ddot{U}_y^2 + \dot{\sigma} \ddot{U}_{yy})(\lambda, Y_{\lambda}) \, d\lambda,$$

where Y starts at (l, y) and runs downward in l . Put $\dot{\Gamma}(l) = \frac{1}{2} \int_0^l \dot{\sigma}(\lambda) \lambda^{-1} \, d\lambda$. There is C , uniform in $\vartheta \in [0, 1]$ and $l \in [l_0, 1]$, such that:

- (a) $|\dot{U}(l, y) + \dot{\Gamma}(l) \mathbf{1}_{\{y \leq 0\}}| + |\ddot{U}(l, y)| \leq C(1 + y^2)^{-1}$;
- (b) $|c_y| + |c_{yy}| + |\dot{U}_y| + |\dot{U}_{yy}| + |\mathcal{Z}_y| + |\mathcal{Z}_{yy}| \leq C(1 + |y|)^{-3}$ for $y \leq 0$, and $\leq C e^{-y^2/c}$ for $y \geq 0$, where $\mathcal{Z} = \dot{U} + \frac{K}{2} c$.

Consequently $I \in C^2([0, 1])$, and

$$\begin{aligned} I'(\vartheta) &= \int_{\mathbb{R}} [\dot{U}(1, y) + \dot{\Gamma}(1) \mathbf{1}_{\{y \leq 0\}}] \, dy, \\ I''(\vartheta) &= \int_{\mathbb{R}} \ddot{U}(1, y) \, dy = \int_{l_0}^1 \int_{\mathbb{R}} p_l (\ddot{U}_y^2 + \dot{\sigma} \ddot{U}_{yy}) \, dy \, dl. \end{aligned}$$

Proof. Smoothness. On each block, (7.13) expresses U^θ as $\sigma_j \log$ of a Gaussian integral of $\exp(U^\theta(l_{j-1}, \cdot)/\sigma_j)$, where $\sigma_j = \sigma_j(\theta)$ is affine in θ and at least one on $[0, 1]$, hence positive on a neighborhood of $[0, 1]$. The integrand is at most one, and derivatives in θ and y produce polynomial factors. The Gaussian integral is at least $\exp(-C(1+y^2))$, because $U^\theta(l_{j-1}, x) \geq U^\theta(l_{j-1}, 0) + b(l_{j-1}, 0)x - x^2/(2l_{j-1})$ by concavity and $c \leq 1/l$ (Lemma 7.7, which holds for every positive σ). To control derivatives of the logarithm, use the tilted Gaussian moment argument of Lemma A.1(iii): the negative log density has curvature between τ^{-1} and $\tau^{-1} + (\sigma_j l_{j-1})^{-1}$, where $\tau = \sigma_j(l - l_{j-1})$, and its mode has at most linear growth in y . Its polynomial moments therefore have polynomial growth in y , uniformly for block parameters in compact sets. Parameter derivatives of the Gaussian variance can be taken as $\partial_\tau P_\tau f = \frac{1}{2} P_\tau f_{yy}$, so this argument remains valid as $\tau \downarrow 0$. Induction over the finite set of blocks now shows that U^θ is smooth in (θ, y) on each block, with derivatives of polynomial growth in y , locally uniformly in θ , and continuous in l . Differentiating (7.11) in θ gives (7.17). The drift b is Lipschitz in y , and the equations for $\dot{U}, \ddot{U}$ are linear, without potential, with sources of polynomial growth and zero data at l_0 ; Itô's formula on each block and the Markov property give the representations.

Tails. The bounds (b) for c_y, c_{yy} are Lemma 7.2(c), because $l \geq l_0$. Differentiate the representation of $\dot{U}$ in y along the flow. Its Jacobian $J_\lambda = \partial_y Y_\lambda = \exp(-\int_\lambda^l c(\lambda', Y_{\lambda'}) d\lambda')$ lies in $(0, 1]$ and satisfies $\partial_y J_\lambda = -J_\lambda \int_\lambda^l c_y(\lambda', Y_{\lambda'}) J_{\lambda'} d\lambda'$. Since c_y and c_{yy} are bounded, differentiation under the expectation is legitimate, and

$$|\dot{U}_y(l, y)| \leq C \mathbb{E} \int_{l_0}^l |c_y(\lambda, Y_\lambda)| d\lambda, \quad |\dot{U}_{yy}(l, y)| \leq C \mathbb{E} \int_{l_0}^l (|c_{yy}| + |c_y| \int_\lambda^l |c_y|)(\lambda, Y_\lambda) d\lambda.$$

By Lemma C.3(a),(b), read in the variable l , for $y \leq -C$ the path stays below y/C with probability at least $1 - 2e^{-y^2/C}$, and for $y \geq 0$ it stays above $y/2$ with probability at least $1 - e^{-y^2/C}$. Since c_y and c_{yy} are bounded, this transfers their tails to $\dot{U}_y$ and $\dot{U}_{yy}$, and hence to $\mathcal{Z}_y = \dot{U}_y + \frac{K}{2}c_y$ and $\mathcal{Z}_{yy} = \dot{U}_{yy} + \frac{K}{2}c_{yy}$. For (a), when $y \leq 0$,

$$\dot{U}(l, y) + \dot{\Gamma}(l) = \frac{1}{2} \mathbb{E} \int_{l_0}^l \frac{\dot{\sigma}(\lambda)}{\lambda} (1 - v(\lambda, Y_\lambda)) d\lambda,$$

and $1 - v$ is bounded by one and by $C(1 + Y_\lambda^2)^{-1}$ when $Y_\lambda \leq 0$ (Lemma 7.2(b)); bounding $|\dot{\sigma}|$ and using the transport estimate gives $|\dot{U} + \dot{\Gamma}| \leq C(1 + y^2)^{-1}$. No sign of $\dot{\sigma} = K'$ is required here: an ordered segment $K \geq 0$ can have K' of both signs. When $y > 0$, $|\dot{U}| \leq C \mathbb{E} \int c(\lambda, Y_\lambda) d\lambda$, and $c = v/\lambda$ has Gaussian decay on the right. The bound on $\ddot{U}$ follows from its representation, (b), and the transport estimate.

Derivatives of I . In (7.12), the counterterms are affine in σ , so the first and second θ -derivatives of the integrand are $\dot{U}(1, y) + \dot{\Gamma}(1)\mathbf{1}_{\{y \leq 0\}}$ and $\ddot{U}(1, y)$. They are continuous in θ and dominated by $C(1 + y^2)^{-1}$ uniformly in θ , so we may differentiate under the integral. Finally, the representation of $\ddot{U}$ and Fubini's theorem give $\int \ddot{U}(1, y) dy =$

$\int_{l_0}^1 \int p_l (\dot{U}_y^2 + \dot{\sigma} \dot{U}_{yy})$, because p_l is the density of the image of Lebesgue measure under the flow from level 1 to level l . The integrand is absolutely integrable by (b) and $p_l \leq 1/l_0$. $\square$

The same argument gives a transport identity for the flat density. Let $l_0 \leq l' < l \leq 1$, and let g be $C^{1,2}$ on each block and continuous in the level, such that g and $f := g_l - \mathcal{G}g$ are bounded on $[l', l] \times \mathbb{R}$, $g(l, \cdot)$ and $g(l', \cdot)$ are integrable, and $\int_{l'}^l \int_{\mathbb{R}} |f| < \infty$. Integrating the Feynman–Kac formula $g(l, y) = \mathbb{E}g(l', Y_{l'}) + \mathbb{E} \int_{l'}^l f(\lambda, Y_\lambda) d\lambda$, for Y started at (l, y) , against $p_l(y) dy$, and using that Y_λ then has density p_λ for $\lambda \leq l$, we obtain

$$\int_{\mathbb{R}} p_l g(l, \cdot) - \int_{\mathbb{R}} p_{l'} g(l', \cdot) = \int_{l'}^l \int_{\mathbb{R}} p_\lambda f(\lambda, \cdot) d\lambda. \quad (\text{C.3})$$

Proof of Proposition 7.10. By Lemma C.4, $I \in C^2([0, 1])$ and $I'' = \int \int p (\dot{U}_y^2 + K' \dot{U}_{yy})$, since $K' = \dot{\sigma}$. Write $\mathcal{D} = \partial_l - \mathcal{G}$. The commutator $\partial_y(\mathcal{G}f) = \mathcal{G}f_y - cf_y$ gives $\mathcal{D}(f_y) = \partial_y(\mathcal{D}f) - cf_y$. With $\mathcal{D}c = -c^2$ and $\mathcal{D}\mathcal{Z} = -\frac{1}{2}Kc^2$ from (7.18), this gives

$$\mathcal{D}c_{yy} = -3c_y^2 - 4cc_{yy}, \quad \mathcal{D}\mathcal{Z}_{yy} = -K(c_y^2 + cc_{yy}) - c_y\mathcal{Z}_y - 2c\mathcal{Z}_{yy}.$$

Put $\Omega = K\mathcal{Z}_{yy} - \frac{1}{4}K^2c_{yy}$. Since K depends on l only, $\mathcal{D}(Kf) = K'f + K\mathcal{D}f$, and therefore

$$\mathcal{D}\Omega = K'\mathcal{Z}_{yy} - Kc_y\mathcal{Z}_y - 2Kc\mathcal{Z}_{yy} - \frac{1}{2}KK'c_{yy} - \frac{1}{4}K^2c_y^2.$$

Substituting $\dot{U} = \mathcal{Z} - \frac{1}{2}Kc$, we obtain the pointwise identity

$$\dot{U}_y^2 + K' \dot{U}_{yy} = (\mathcal{Z}_y - \frac{1}{2}Kc_y)^2 + K'(\mathcal{Z}_{yy} - \frac{1}{2}Kc_{yy}) = \mathcal{Z}_y^2 + 2Kc\mathcal{Z}_{yy} + \frac{1}{2}K^2c_y^2 + \mathcal{D}\Omega.$$

Every term is integrable against $p dy dl$ on $[l_0, 1] \times \mathbb{R}$, by Lemma C.4(b), the boundedness of c, c_y, K, K' and $p \leq 1/l_0$; the same bounds verify the hypotheses of (C.3) for $g = \Omega$. By (C.3) with $l' = l_0$ and $l = 1$,

$$\int_{l_0}^1 \int_{\mathbb{R}} p \mathcal{D}\Omega = \int_{\mathbb{R}} p_1 \Omega(1, \cdot) - \int_{\mathbb{R}} p_{l_0} \Omega(l_0, \cdot) = 0.$$

Indeed, $K(l_0) = 0$; and $p_1 = 1$, while $\int \Omega(1, \cdot) = K(1) \int \mathcal{Z}_{yy}(1, \cdot) - \frac{1}{4}K(1)^2 \int c_{yy}(1, \cdot) = 0$ because $\mathcal{Z}_y(1, \cdot)$ and $c_y(1, \cdot)$ tend to zero at $\pm\infty$. Finally, integrating by parts in y ,

$$\int_{\mathbb{R}} p_l c \mathcal{Z}_{yy} dy = - \int_{\mathbb{R}} (p_l c)_y \mathcal{Z}_y dy.$$

The boundary terms vanish because $p_l c$ is bounded and $\mathcal{Z}_y$ tends to zero, and the right side converges because $p_l c$ is bounded and monotone (Proposition 7.9) and $\mathcal{Z}_y$ is bounded. This proves (7.19). $\square$

Proofs for the local analysis at the crossing

This appendix proves the lemmas of Chapter 8. §D.1 proves the signs of the flat constants (Lemma 8.2). §D.2 computes the transition kernel of the replica-symmetric top diffusion (Lemma D.1), uses it to prove Proposition 8.4, and proves Lemma 8.5. §D.3 proves the natural-scale continuity Lemma 8.6 by one compactness argument in rescaled variables. §D.4 proves the bottom comparison Lemma 8.7 by a weighted parabolic comparison. Throughout we use the Gaussian facts of Lemmas 2.1 and C.1; in particular $R' = -V$, $\bar{R}' = 1 - V$, and the identity (2.2) for V'' .

D.1 Flat constants

Proof of Lemma 8.2. Put $p_V = \int (1 - 2V)V'^2$. By (2.2) and $(2R + z)' = 1 - 2V$, multiplying (2.2) by V'' and integrating by parts gives

$$\int (2R + z)V'V'' = -\frac{1}{2} \int (1 - 2V)V'^2, \quad \int V(1 - V)V'' = -\int (1 - 2V)V'^2,$$

where the boundary terms $(2R + z)V'^2$ and $V(1 - V)V'$ vanish at $\pm\infty$ by Lemma C.1(a). Hence $\mathcal{I}_2 = \frac{1}{2}p_V + 2p_V = \frac{5}{2}p_V$.

Next let $F \in C^2([0, 1])$ with $F(0) = F(1) = 0$. By (2.2), the derivative of $F'(V)V' + (2R + z)F(V)$ is $F''(V)V'^2 - 2F'(V)V(1 - V) + (1 - 2V)F(V)$. This function tends to zero at $\pm\infty$: at $+\infty$ by the Gaussian decay of V and V' , and at $-\infty$ because $F(V) = O(1 - V) = O(z^{-2})$, $2R + z = O(|z|)$, and $V' \rightarrow 0$ (Lemma C.1(a)). Hence

$$\int F''(V)V'^2 = \int [2F'(V)V(1 - V) - (1 - 2V)F(V)].$$

The choice $F(v) = (v^2 - v)/2$ gives $A_V = \int V(1 - V)(V - \frac{1}{2}) = B_V - \frac{1}{2}c_0$, and the choice $F(v) = v^2(1 - v)/2$ gives $\int (1 - 3V)V'^2 = \frac{1}{2} \int V^2(1 - V)(3 - 4V)$. Since $\int VV'^2 = \frac{1}{2}(A_V - p_V)$, these identities give

$$\int (3V - 1)V'^2 = \frac{A_V - 3p_V}{2}, \quad \int V^2(1 - V)(1 - 3V) = \frac{9p_V - 3A_V - 5B_V}{4}.$$

With $p_V = \frac{2}{5}\mathcal{I}_2$ we obtain, for every $k \in \mathbb{R}$,

$$\begin{aligned} \int [V'^2 + 4k(3V - 1)V'^2 + 2k^2V^2(1 - V)(1 - 3V)] &= \frac{1 + (3k - 2)^2}{5} \mathcal{I}_2 \\ &+ \left(2k - \frac{3}{2}k^2\right)A_V - \frac{5}{2}k^2B_V. \end{aligned}$$

At $k = x_f$ we have $A_V = 2x_fB_V$ and $B_V(1 - 2x_f) = B_V - A_V = \frac{1}{2}c_0$, so the last two terms equal $\frac{3}{2}x_f^2(1 - 2x_f)B_V = \frac{3}{4}x_f^2c_0$. This is the formula for Σ_{flat} , which is positive because $\mathcal{I}_2 > 0$. Finally $x_f = A_V/(2B_V) = \frac{1}{2} - c_0/(4B_V)$ is positive because $A_V > 0$, and $x_f < \frac{1}{4}$ because $c_0 = \int V(1 - V) > \int V^2(1 - V) = B_V$, as $0 < V < 1$. $\square$

D.2 The replica-symmetric kernel, obstacle, and family

For $0 < \rho < 1$ and $z_0, z \in \mathbb{R}$ put

$$\Pi_\rho(z_0, z) = \sqrt{\rho} \phi_{1-\rho}(\sqrt{\rho}z - z_0) \frac{\Phi(z)}{\Phi(z_0)}, \quad v(z) = \Phi(z)^2 e^{z^2/2},$$

where ϕ_s is the $N(0, s)$ density. For $c \in \mathbb{R}$ let $g_c(z_0) = \Pi_\rho(z_0, (-\infty, c]) - \mathbf{1}_{\{z_0 \leq c\}}$.

Lemma D.1. (a) $\Pi_\rho(z_0, \cdot)$ is the law at time $\log(1/\rho)$ of the diffusion (7.20) started at z_0 ; in particular $\Pi_\rho \Pi_{\rho'} = \Pi_{\rho\rho'}$. The detailed balance identity $v(z_0)\Pi_\rho(z_0, z) = v(z)\Pi_\rho(z, z_0)$ holds, v is strictly increasing, $\int_{-\infty}^c v < \infty$ for every c , and $\Pi_\rho(z_0, \cdot)$ is strictly increasing in z_0 in the likelihood ratio order. Moreover $\Pi_\rho(z_0, \cdot)$ is the law of

$$\sqrt{\rho}z_0 + \frac{1-\rho}{\sqrt{\rho}}\mathbb{T} + \sqrt{1-\rho}N,$$

where $\mathbb{T}$ has the law of $N(z_0, 1)$ conditioned on $[0, \infty)$ and $N \sim N(0, 1)$ is independent of $\mathbb{T}$.

- (b) $\int v|g_c| \leq 2 \int_{-\infty}^c v$ and $\int v g_c = 0$ for every c .
- (c) Put $\varkappa_1 = \min_{[-1, 1]}(2R(z) + z)$ and $\varkappa_2 = \phi(2)\Phi(-4)$. Then $\varkappa_1 \geq \min\{2\phi(1), 4\phi_0 - 1\} > 0$, and $g_c(z_0) \geq \varkappa_2$ whenever $\rho \in [\frac{1}{4}, 1)$, $c \in [-1, 0]$ and $z_0 \in (c, c + \sqrt{1-\rho})$.
- (d) Let $p_0 = v f_0$ be a probability density with f_0 nonincreasing, let $Z_0 \sim p_0$, and let Z_ρ have the law $\int p_0(z_0)\Pi_\rho(z_0, \cdot) dz_0$. Then Z_ρ has the density $v f_\rho$ with $f_\rho(z) = \int \Pi_\rho(z, z_0)f_0(z_0) dz_0$ nonincreasing, and

$$\mathbb{P}(Z_\rho \leq c) - \mathbb{P}(Z_0 \leq c) = \int v(f_0 - f_0(c))g_c \leq 0 \quad (c \in \mathbb{R}).$$

If f_0 is strictly decreasing, then f_ρ is strictly decreasing, the inequality is strict for every c , and $\mathbb{E}h(Z_\rho) < \mathbb{E}h(Z_0)$ for every bounded, strictly decreasing $h \in C^1(\mathbb{R})$ with h' integrable.

Proof. (a) Consider the order parameter $\mathbf{1}_{[0,1]}$ at the margin 0. By (2.9), $u(q, x) = \log H(q, x)$ with $H(q, x) = \Phi(x/\sqrt{1-q})$ on $[0, 1]$, and $H_q + \frac{1}{2}H_{xx} = 0$. Hence the diffusion (2.14) with drift $b = H_x/H$, started at $(0, z_0)$, is the Doob transform of Brownian motion by H , with transition densities $\phi_{q_2-q_1}(x_2 - x_1)H(q_2, x_2)/H(q_1, x_1)$ by (2.19). Put $Z_\zeta = X_q/\sqrt{1-q}$ with $\zeta = \log(1/(1-q))$. Itô's formula gives $dZ = (R(Z) + \frac{1}{2}Z) d\zeta + (1-q)^{-1/2} dB_q$, and the martingale part is a Brownian motion in the time ζ ; so Z solves (7.20), whose drift has the bounded derivative $\frac{1}{2} - V$, so that its solution is unique. In the variables $x_i = \sqrt{1-q_i} z_i$, with $\rho = (1-q_2)/(1-q_1)$, the transition density of Z becomes $\sqrt{1-q_2} \phi_{q_2-q_1}(\sqrt{1-q_2} z_2 - \sqrt{1-q_1} z_1) \Phi(z_2)/\Phi(z_1) = \Pi_\rho(z_1, z_2)$. This proves the first assertion; the semigroup property follows from the Markov property and time homogeneity.

For detailed balance, $\nu(z_0)\Pi_\rho(z_0, z) = \Phi(z_0)\Phi(z)\sqrt{\rho}(2\pi(1-\rho))^{-1/2}e^E$ with $E = (2\sqrt{\rho}zz_0 - \rho z^2 - \rho z_0^2)/(2(1-\rho))$, which is symmetric in (z_0, z) . Next $(\log \nu)' = 2R + z > 0$, because $R(z) > z_-$ (Lemma 2.1(a)). For $z < -1$, $\Phi(z) \leq \phi(z)/|z|$ (Lemma 2.1(b)) gives $\nu(z) \leq e^{-z^2/2}/(2\pi z^2)$, so $\int_{-\infty}^c \nu < \infty$. The mixed derivative $\partial_z \partial_{z_0} \log \Pi_\rho(z_0, z) = \sqrt{\rho}/(1-\rho) > 0$ gives the strict likelihood ratio order. For the representation, let $Y' \sim N(z_0/\sqrt{\rho}, (1-\rho)/\rho)$ and $G' \sim N(0, 1)$ be independent. The density of Y' is $\sqrt{\rho} \phi_{1-\rho}(\sqrt{\rho}z - z_0)$, and $\mathbb{P}(G' \leq Y') = \Phi(z_0)$, so $\Pi_\rho(z_0, \cdot)$ is the law of Y' conditioned on $\{G' \leq Y'\}$. Put $W = G' - Y'$, so that $\text{Var } W = 1/\rho$ and $\text{Cov}(Y', W) = -(1-\rho)/\rho$. Gaussian regression gives $Y' = \sqrt{\rho}z_0 - (1-\rho)W + \sqrt{1-\rho}N$ with $N \sim N(0, 1)$ independent of W . Writing $-W = T/\sqrt{\rho}$ with $T \sim N(z_0, 1)$, the event $\{G' \leq Y'\}$ is $\{T \geq 0\}$, which proves the representation.

(b) Detailed balance and Tonelli's theorem give

$$\int \nu(z_0)\Pi_\rho(z_0, (-\infty, c]) dz_0 = \int_{-\infty}^c \nu(z) \left(\int \Pi_\rho(z, z_0) dz_0 \right) dz = \int_{-\infty}^c \nu < \infty.$$

Subtracting $\int_{-\infty}^c \nu$ gives $\int \nu g_c = 0$, and the sum of the two integrals bounds $\int \nu |g_c|$.

(c) On $[0, 1]$, $2R + z \geq 2R \geq 2\phi(1)$. On $[-1, 0]$, $R \geq R(0) = 2\phi_0$, so $2R + z \geq 4\phi_0 - 1 > 0$. Let $\rho \in [\frac{1}{4}, 1)$, $c \in [-1, 0]$ and $z_0 \in (c, c + \sqrt{1-\rho}) \subset (-1, 1)$. Then $g_c(z_0) = \Pi_\rho(z_0, (-\infty, c])$. On the event $\{T \leq 1, N \leq -4\}$ the variable of (a) is at most $\sqrt{\rho}z_0 + 2(1-\rho) - 4\sqrt{1-\rho}$, because $(1-\rho)/\sqrt{\rho} \leq 2(1-\rho)$. Moreover $c - \sqrt{\rho}z_0 \geq c(1-\sqrt{\rho}) - \sqrt{\rho}\sqrt{1-\rho} \geq -(1-\rho) - \sqrt{1-\rho}$, and $3(1-\rho) \leq 3\sqrt{1-\rho}$. Hence the variable is at most c on this event. Since $\mathbb{P}(T \leq 1) \geq \Phi(1-z_0) - \Phi(-z_0) \geq \phi(2)$ and T, N are independent, $g_c(z_0) \geq \kappa_2$.

(d) By detailed balance, $\int \nu(z_0)f_0(z_0)\Pi_\rho(z_0, z) dz_0 = \nu(z)f_\rho(z)$. By the likelihood ratio order (see [114]), $\Pi_\rho(z, \cdot)$ is stochastically increasing in z , strictly because it has a positive density; so f_ρ is nonincreasing, and strictly decreasing if f_0 is. Next, $\mathbb{P}(Z_\rho \leq c) - \mathbb{P}(Z_0 \leq c) = \int p_0 g_c = \int \nu(f_0 - f_0(c))g_c$, where we used $\int \nu g_c = 0$ and the integrability of νg_c from (b). The integrand is nonpositive: for $z_0 \leq c$, $g_c \leq 0$ and $f_0(z_0) \geq f_0(c)$; for $z_0 > c$, $g_c \geq 0$ and $f_0(z_0) \leq f_0(c)$. If f_0 is strictly decreasing, the integrand is negative on $(-\infty, c)$,

where $g_c = -\Pi_\rho(\cdot, (c, \infty)) < 0$. Finally, for h as stated, $h(z) = h(+\infty) + \int_z^\infty |h'(c)| dc$, and Tonelli's theorem gives $\mathbb{E}h(Z_\rho) - \mathbb{E}h(Z_0) = \int |h'(c)| [\mathbb{P}(Z_\rho \leq c) - \mathbb{P}(Z_0 \leq c)] dc < 0$. $\square$

We now describe the Parisi diffusion of γ_d at κ_d . Write $\beta = \beta_d = \sqrt{\delta_d/q_d}$ and $k = \kappa_d/\sqrt{q_d}$, and let $Z_d = (X_{q_d} - \kappa_d)/\sqrt{\delta_d}$. Since $X_{q_d} = B_{q_d}$, Z_d has the density $w_d(z) = \beta\phi(k + \beta z)$.

Above the atom. On $[q_d, 1)$, $\gamma_d = 1$ and $u(q, x) = \log \Phi((x - \kappa_d)/\sqrt{1 - q})$ by (2.9). The computation in the proof of Lemma D.1(a), with the margin κ_d and the initial level q_d , shows that $Z_\rho := (X_q - \kappa_d)/\sqrt{1 - q}$, with $\rho = \hat{s}(q) \in (0, 1]$, starts from $Z_1 = Z_d$ and has the transition kernels Π . By (2.10), $v(q, X_q) = V(Z_\rho)$, so $S_d(q) = \alpha \mathbb{E}V(Z_\rho)^2$.

Below the atom. On $[0, q_d)$, $\gamma_d = 0$, $X = B$, and $\lambda_d = \delta_d$. By Lemma 2.9(a), $\mathcal{L}c = \gamma c^2 = 0$ there, so $c(q, \cdot)$ is the heat flow of $c(q_d, \cdot)$. With $Z'_q = (X_q - \kappa_d)/\sqrt{\delta_d}$ and $t = \hat{s}(q) - 1 = (q_d - q)/\delta_d$, this gives $v(q, X_q) = \mathbb{E}[V(Z'_q + \sqrt{t}G) \mid Z'_q]$, where $G \sim N(0, 1)$ is independent of Z'_q , and $Z'_q + \sqrt{t}G$ has the law of Z_d .

We use the positive constants $\varkappa_1, \varkappa_2$ of Lemma D.1(c) and

$$\begin{aligned}\varkappa_3 &= (2\Phi(1) - 1) \min_{[-2, 2]} |V'|, \\ \varkappa_4 &= \frac{(V(-1)^2 - V(0)^2) v(-1) \phi(1) \varkappa_1 \varkappa_2}{4v(1)}, \\ \varkappa_5 &= 2\phi(1)\varkappa_3^2;\end{aligned}$$

they are positive because $V' < 0$ (Lemma 2.1(a)).

Proof of Proposition 8.4. By Proposition 3.8(iii) and $|\kappa_d| \leq e$, we may choose e_0 so small that for $0 < e < e_0$:

$$\begin{aligned}\beta^2 &\leq \frac{1}{2}, \quad \beta|k| \leq \phi_0, \quad \beta + |k| \leq \frac{1}{\sqrt{3}}, \quad \beta|k| + \beta^2 \leq \frac{1}{2}\varkappa_1, \\ |\kappa_d| + \sqrt{\delta_d} &\leq \sqrt{q_d - \delta_d}, \quad c_\beta e \leq \beta \leq C_\beta e,\end{aligned}$$

with numerical $c_\beta, C_\beta > 0$.

(a) *Below q_d .* By (8.4) with $\gamma_d = 0$, $S'_d = \alpha\mathcal{N} > 0$ on $[0, q_d)$, since $v_x < 0$; so S_d is strictly increasing on $[0, q_d]$.

Above q_d . Let $f_0 = w_d/v$. Then $(\log f_0)'(z) = -\beta(k + \beta z) - (2R(z) + z)$. For $z \in [0, 1]$, $2R(z) + z \geq 2\phi(1) > \phi_0$; for $z > 1$, $2R + z > 1 > \phi_0$; in both cases $-\beta(k + \beta z) \leq \beta|k|$. For $z = -t < 0$, $-\beta(k + \beta z) \leq \beta|k| + \beta^2 t \leq \beta|k| + \frac{1}{2}t$, and

$$2R(-t) - \frac{3}{2}t > \max\left\{\frac{1}{2}t, 4\phi_0 - \frac{3}{2}t\right\} \geq \phi_0,$$

because $R(-t) > t$ (Lemma 2.1(b)) and $R(-t) > R(0) = 2\phi_0$, as R is decreasing. In all cases $(\log f_0)' < \beta|k| - \phi_0 \leq 0$, so f_0 is strictly decreasing. Let $0 < \rho' < \rho \leq 1$. By Lemma D.1(d) (with nothing to prove if $\rho = 1$), Z_ρ has a density νf_ρ with f_ρ strictly

decreasing. By the semigroup property, $Z_{\rho'}$ is obtained from Z_{ρ} by the kernel $\Pi_{\rho'/\rho}$. Applying Lemma D.1(d) again, with the initial density νf_{ρ} and the strictly decreasing function $h = V^2$ (for which $|h'| = 2V|V'|$ is integrable), gives $\mathbb{E}V(Z_{\rho'})^2 < \mathbb{E}V(Z_{\rho})^2$. Hence S_d is strictly decreasing on $[q_d, 1)$.

Since $S_d(q_d) = 1$ and $D_d(q_d) = 0$, the function $D_d(q) = \int_{q_d}^q (S_d - 1)\lambda_d^{-2}$ is positive on $[0, q_d)$ and negative on $(q_d, 1)$. As $\mathcal{H}'_d = -D_d/2$, the point q_d is the unique minimum point of $\mathcal{H}_d$ on $[0, 1)$.

(b) *Above q_d .* Let $\rho \in [\frac{1}{4}, 1)$ and let q be the level with $\hat{s}(q) = \rho$. Put $h = V^2$. Since h is decreasing with $h(+\infty) = 0$, Tonelli's theorem gives $\mathbb{E}h(Z) = \int |h'(c)|\mathbb{P}(Z \leq c) dc$ for every random variable Z . Since $1 - S_d(q) = \alpha(\mathbb{E}h(Z_1) - \mathbb{E}h(Z_{\rho}))$, and since the integrand in Lemma D.1(d) is nonpositive everywhere,

$$1 - S_d(q) = \alpha \int |h'(c)| [\mathbb{P}(Z_1 \leq c) - \mathbb{P}(Z_{\rho} \leq c)] dc,$$

$$\mathbb{P}(Z_1 \leq c) - \mathbb{P}(Z_{\rho} \leq c) = \int \nu |f_0 - f_0(c)| |g_c|.$$

Restrict to $c \in [-1, 0]$ and $z_0 \in (c, c + \sqrt{1-\rho}) \subset (-1, 1)$. There $g_c(z_0) \geq \kappa_2$ by Lemma D.1(c), and $\nu(z_0) \geq \nu(-1)$. On $[-1, 1]$ we have $w_d \geq \beta\phi(1)$, because $|k + \beta z| \leq 1$, and $\nu \leq \nu(1)$, so $f_0 \geq \beta\phi(1)/\nu(1)$; also $|(\log f_0)'| \geq \kappa_1 - \beta|k| - \beta^2 \geq \frac{1}{2}\kappa_1$. Hence $|f_0(z_0) - f_0(c)| \geq \beta\phi(1)\kappa_1(z_0 - c)/(2\nu(1))$, and integrating over z_0 and then over $c \in [-1, 0]$, where $\int_{-1}^0 |h'| = V(-1)^2 - V(0)^2$,

$$1 - S_d(q) \geq \alpha\beta\kappa_4(1 - \rho) \quad \left(\frac{1}{4} \leq \hat{s}(q) = \rho < 1\right).$$

Since $D_d(q_d) = 0$, (8.2) for γ_d gives $G_d(q) = \mathcal{A}_d(q_d, q)$. For $q \geq q_d$ we have $\lambda_d(r') = 1 - r' = \delta_d \hat{s}(r')$, and in the variable $\hat{s}' = \hat{s}(r')$,

$$G_d(q) = \frac{1}{2} \int_{\hat{s}(q)}^1 (\hat{s}' - \hat{s}(q)) (1 - S_d) \hat{s}'^{-2} d\hat{s}'$$

$$\geq \frac{\alpha\beta\kappa_4}{2} \int_{\hat{s}(q)}^1 (\hat{s}' - \hat{s}(q)) (1 - \hat{s}') d\hat{s}' = \frac{\alpha\beta\kappa_4}{12} (1 - \hat{s}(q))^3$$

for $\frac{1}{4} \leq \hat{s}(q) \leq 1$.

Below q_d . Let $q < q_d$ with $t = \hat{s}(q) - 1 \in (0, 1]$. Since $\alpha\mathbb{E}V(Z_d)^2 = S_d(q_d) = 1$, the description below the atom gives

$$1 - S_d(q) = \alpha\mathbb{E} \text{Var}(V(Z'_q + \sqrt{t}G) \mid Z'_q).$$

For a C^1 function F with bounded derivative, the Cauchy-Schwarz inequality and Gaussian integration by parts give $\text{Var} F(\sqrt{t}G) \geq (\mathbb{E}[F(\sqrt{t}G)G])^2 = t(\mathbb{E}F'(\sqrt{t}G))^2$, which is the lower bound of [26]. For $|z| \leq 1$ and $t \leq 1$, $|\mathbb{E}V'(z + \sqrt{t}G)| \geq$

$\mathbb{E}[|V'(z + \sqrt{t}G)|; |G| \leq 1] \geq \kappa_3$, because $V' < 0$. The density of Z'_q at $z \in [-1, 1]$ is $\sqrt{\delta_d/q} \phi((\kappa_d + \sqrt{\delta_d}z)/\sqrt{q}) \geq \beta \phi(1)$, because $q \leq q_d$ and $|\kappa_d| + \sqrt{\delta_d} \leq \sqrt{q_d - \delta_d} \leq \sqrt{q}$. Hence $1 - S_d(q) \geq \alpha\beta\kappa_5 t$ for $t \in (0, 1]$. Since $\lambda_d = \delta_d$ on $[0, q_d]$, in the variable $t' = \hat{s}(r') - 1$,

$$G_d(q) = \frac{1}{2} \int_0^t (t - t')(1 - S_d) dt' \geq \frac{\alpha\beta\kappa_5}{12} t^3.$$

Saturation. By (a), G_d decreases in $\hat{s}$ on $(0, 1]$ and increases on $[1, \infty)$. For $\hat{s} < \frac{1}{4}$ and $\hat{s} > 2$ we may therefore use the values at $\frac{1}{4}$ and at 2. With $\kappa_o = \min\{\kappa_4, \kappa_5\}$ this gives $G_d \geq \frac{1}{12}\alpha\beta\kappa_o \min\{|\hat{s} - 1|, \frac{3}{4}\}^3$, and $\alpha\beta \geq 2c_\beta e$.

(c) *Above q_d .* Let $\rho = \hat{s}(q) \in [\frac{1}{4}, 1]$ and $\varsigma = \log(1/\rho)$. By (7.25), the generator of (7.20) maps V^2 to $k_1 = V'^2 - 2V^2(1 - V)$, which is bounded and integrable. Dynkin's formula gives $S_d(q_d) - S_d(q) = -\alpha \int_0^\varsigma \mathbb{E}k_1(Z_{e^{-r}}) dr$. The drift of (7.20) has derivative $\frac{1}{2} - V \geq -\frac{1}{2}$. Hence two solutions driven by the same Brownian motion and started at $z_0 < z_1$ satisfy $Z_r^{z_1} - Z_r^{z_0} \geq e^{-r/2}(z_1 - z_0)$ by Gronwall's inequality, so the map $z_0 \mapsto Z_r^{z_0}$ is increasing and the preimage of an interval of length h is an interval of length at most $e^{r/2}h$. Since Z at time 0 has the density w_d and is independent of the driving Brownian motion, the law of Z at time r has a density at most $e^{r/2} \sup w_d \leq e^{r/2} \beta \phi_0$. Therefore

$$0 \leq 1 - S_d(q) \leq \alpha\beta\phi_0 \|k_1\|_{L^1} \int_0^\varsigma e^{r/2} dr \leq 2\alpha\beta\phi_0 \|k_1\|_{L^1} \log \frac{1}{\rho} \leq 8\alpha\beta\phi_0 \|k_1\|_{L^1} (1 - \rho),$$

where the lower bound is (a), and we used $\varsigma \leq \log 4$ and $\log(1/\rho) \leq (1 - \rho)/\rho$.

Below q_d . The Gaussian Poincaré inequality $\text{Var } F(\sqrt{t}G) \leq t\mathbb{E}F'(\sqrt{t}G)^2$ [41] (see also [21, Theorem 3.20]) gives $1 - S_d(q) \leq \alpha t \mathbb{E}V'(Z_d)^2 \leq \alpha t \beta \phi_0 A_V$.

In both cases $0 \leq 1 - S_d \leq C'_o e |\hat{s} - 1|$ for $|\hat{s} - 1| \leq \frac{3}{4}$, with a constant C'_o . Since $\hat{\lambda}_d = \lambda_d/\delta_d = \min\{\hat{s}, 1\} \geq \frac{1}{4}$ there, integration gives $|\delta_d D_d(q)| = |\int_{[1, \hat{s}(q)]} (1 - S_d) \hat{\lambda}_d^{-2} d\hat{s}'| \leq 16C'_o e |\hat{s} - 1|^2$ and $G_d(q) \leq 8C'_o e |\hat{s} - 1|^3$. With $C_o = 16C'_o$ this proves (c). $\square$

Proof of Lemma 8.5. Throughout, $\eta \leq \frac{1}{4}d_0$, $q \in \mathcal{W}_\eta$, and $|\kappa - \kappa_d| \leq \eta e^2$, so that $\delta_q \in [\frac{1}{4}d_0 e^2, 3d_0 e^2]$ and $q \geq \frac{1}{2}$ for small e .

Derivative formulas. By (3.2),

$$\partial_q D_{RS} = \frac{S_{RS} - 1}{\delta_q^2} - \frac{2\alpha}{\delta_q^2} \mathbb{E}[R^2(1 - V)](Z^q) + \frac{2D_{RS}}{\delta_q}. \quad (\text{D.1})$$

Next $S_{RS}(q) = \alpha \mathbb{E}p(q, B_q)$ with $p(q, x) = V((x - \kappa)/\sqrt{\delta_q})^2$, and Itô's formula, with $\partial_q \frac{x - \kappa}{\sqrt{\delta_q}} = \frac{x - \kappa}{2\delta_q^{3/2}}$ and (2.2), gives

$$\partial_q S_{RS} = \frac{\alpha}{\delta_q} \mathbb{E}[V'^2 + V(V'' + zV')](Z^q) = \frac{\alpha}{\delta_q} \mathbb{E}[V'^2 - 2RVV' - 2V^2(1 - V)](Z^q).$$

Differentiation under the expectation, justified by the boundedness of V, V' and $R(z) \leq R(0) + z_-$, gives

$$\partial_\kappa S_{\text{RS}} = -\frac{2\alpha}{\sqrt{\delta_q}} \mathbb{E}[VV'](Z^q), \quad \partial_\kappa D_{\text{RS}} = \frac{2\alpha}{\delta_q^{3/2}} \mathbb{E}[RV](Z^q).$$

Bounds. The density of Z^q is $\sqrt{\delta_q/q} \phi((\kappa + \sqrt{\delta_q}z)/\sqrt{q})$. It is at most $\sqrt{2\delta_q} \phi_0 \leq Ce$, and at least $\sqrt{\delta_q} \phi(1) \geq ce$ on $[-1, 1]$ for small e . The functions V'^2 , RVV' , $V^2(1-V)$, and VV' are integrable by Lemma C.1(a), and V'^2 is bounded below on $[-1, 1]$. Moreover $RV \leq R \leq R(0) + z_-$ and $\mathbb{E}|Z^q| \leq (1 + |\kappa|)/\sqrt{\delta_q}$. The function $R^2(1-V)$ is bounded, by Lemma 2.1(a),(b). For $z = -t \leq -2$, Lemma 2.1(a),(b) give $R(z) > t$ and $1 - V(z) \geq \frac{1}{3}\bar{R}(z)^2 \geq \frac{1}{3}(t^{-1} - 3t^{-3})^2 \geq \frac{1}{48}t^{-2}$, so $R^2(1-V) \geq \frac{1}{48}$ on $(-\infty, -2]$; and $\mathbb{P}(Z^q \leq -2) = \Phi((\kappa - 2\sqrt{\delta_q})/\sqrt{q}) \geq \frac{1}{4}$ for small e . Hence

$$\begin{aligned} |\partial_q S_{\text{RS}}| &\leq \frac{C}{e}, & |\partial_\kappa S_{\text{RS}}| &\leq C, & |\partial_\kappa D_{\text{RS}}| &\leq \frac{C}{e^4}, \\ ce \leq A_w(q) &\leq Ce, & \mathbb{E}[R^2(1-V)](Z^q) &\geq \frac{1}{192}. \end{aligned} \tag{D.2}$$

This proves (c) and the first sentence of (b).

(a) At (q_d, κ_d) we have $S_{\text{RS}} = 1$ and $D_{\text{RS}} = 0$. By (D.2), $|S_{\text{RS}} - 1| \leq C\eta e$ on $\mathcal{W}_\eta \times [\kappa_d - \eta e^2, \kappa_d + \eta e^2]$, and $|D_{\text{RS}}(q_d; \kappa)| \leq C\eta/e^2$. The identity (D.1) is a linear differential equation for D_{RS} in q , with coefficient $2/\delta_q \leq 8/(d_0 e^2)$ and with the other terms bounded by C/e^4 . Gronwall's inequality over a q -interval of length at most ηe^2 gives $|D_{\text{RS}}| \leq C\eta/e^2$ on $\mathcal{W}_\eta$. Inserting these bounds and $\alpha \geq 2$ into (D.1), and using $e^{-2} \leq 3d_0/\delta_q$,

$$\partial_q D_{\text{RS}} \leq \frac{C\eta e - \frac{1}{48} + C\eta}{\delta_q^2} \leq -\frac{1}{96\delta_q^2}$$

for $\eta = \eta_1$ small. Now let $|\kappa - \kappa_d| \leq \eta'_1 e^2$. By (D.2), $|D_{\text{RS}}(q_d; \kappa)| \leq C|\kappa - \kappa_d|/e^4$, while $\partial_q D_{\text{RS}} \leq -1/(864d_0^2 e^4)$ on $\mathcal{W}_{\eta_1}$, because $\delta_q \leq 3d_0 e^2$. Hence $D_{\text{RS}}(\cdot; \kappa)$ changes sign within distance $C_1|\kappa - \kappa_d|$ of q_d , with $C_1 = 864d_0^2 C$, and has a zero there if $C_1\eta'_1 < \eta_1$. This zero is RS-stationary at κ . Since $\kappa_{\text{RS}} - \kappa_d = \theta_d e^2 + o(e^2)$ by Proposition 3.8(iii), after decreasing η'_1 (so that $\eta'_1 < \frac{1}{2}\theta_d$) and e_r we have $\kappa \in [\kappa_{\text{RS}} - c_2, \kappa_{\text{RS}})$, and by the uniqueness in Proposition 3.8(i) the zero is $q^*(\kappa)$. This proves (a) with $c_r \leq \frac{1}{96}$ and $C_r \geq C_1$.

(b) The segments from (q_d, κ_d) to (q_d, κ) and from (q_d, κ) to $(q^*(\kappa), \kappa)$ lie in $\mathcal{W}_{\eta_1} \times [\kappa_d - \eta_1 e^2, \kappa_d + \eta_1 e^2]$. By (D.2) and (a),

$$|\tau(\kappa)| = |S_{\text{RS}}(q^*(\kappa); \kappa) - S_{\text{RS}}(q_d; \kappa_d)| \leq \frac{C}{e} C_1 |\kappa - \kappa_d| + C|\kappa - \kappa_d| \leq \frac{C_r |\kappa - \kappa_d|}{e}$$

after enlarging C_r . □

D.3 Natural-scale continuity

Proof of Lemma 8.6. We divide the proof into four steps. The first two hold for a fixed interval $[u_-, u_+]$ with $0 < u_- < 1 < u_+$; for part (b) we take $[u_-, u_+] = [\frac{1}{2}, 2]$.

STEP 1. Rescaling. Put $s_{\text{lo}} = \frac{1}{4}u_-d_0$ and $s_{\text{hi}} = 4u_+d_0$. Let $e < e_{\text{st}}$ be so small that $e^2s_{\text{hi}} \leq \frac{3}{4}$; then $d_d \in [\frac{1}{2}d_0, 2d_0]$ and $q_{\text{hi}} := 1 - e^2s_{\text{hi}} \geq \frac{1}{4}$. Every $\gamma \in \mathcal{U}$ with $\hat{s}(\text{supp } \mu_\gamma) \subset [u_-, u_+]$ has its support in the levels $s = (1 - q)/e^2 \in [\frac{1}{2}u_-d_0, 2u_+d_0] \subset (s_{\text{lo}}, s_{\text{hi}})$, so its rescaled profile ζ_γ (Definition 4.1) equals one on $[0, s_{\text{lo}}]$ and zero on (s_{hi}, ∞) ; the same holds for $\zeta_d = \mathbf{1}_{[0, d_d]}$. Thus all these profiles lie in $\mathcal{K}(s_{\text{lo}}, s_{\text{hi}})$. For such ζ let U_ζ solve (7.3), and put $v_\zeta = -\ell_\zeta U_{\zeta, yy} \in (0, 1)$. By Lemma C.2, U_ζ is an instance of the finite-cutoff problem at margin zero with cutoff $1 - s_{\text{lo}}/s_{\text{hi}}$, independent of e and κ , and $u_\gamma(q, \kappa + ey) = U_{\zeta_\gamma}(s, y)$ for $s = (1 - q)/e^2 \in [s_{\text{lo}}, s_{\text{hi}}]$. Since $\lambda = e^2\ell$ and $\partial_x = e^{-1}\partial_y$,

$$(v, e v_x, e^2 v_{xx})(q, \kappa + ey) = (v_\zeta, \partial_y v_\zeta, \partial_y^2 v_\zeta)(s, y).$$

Since $\gamma = 0$ below q_{hi} , $X_{q_{\text{hi}}} = B_{q_{\text{hi}}}$, and $(X_{q_{\text{hi}}} - \kappa)/e$ has the density $e\pi$ with

$$\pi(y) = \pi_{e, \kappa}(y) = \frac{1}{\sqrt{q_{\text{hi}}}} \phi\left(\frac{\kappa + ey}{\sqrt{q_{\text{hi}}}}\right), \quad 0 < \pi \leq 2\phi_0,$$

and $\pi \rightarrow \phi_0$ locally uniformly as $(e, \kappa) \rightarrow (0, 0)$. Let $Y_s^\zeta(y)$, $s \leq s_{\text{hi}}$, be the rescaled diffusion (C.1) of ζ started at level s_{hi} from y . By the Markov property at the level q_{hi} , for every polynomial P in three variables with $P(0, 0, 0) = 0$, and with

$$P_\zeta = P(v_\zeta, \partial_y v_\zeta, \partial_y^2 v_\zeta), \quad r_{P, \zeta}(s, z) = P_\zeta(s, z) - P(1, 0, 0)\mathbf{1}_{\{z < 0\}},$$

we obtain, for $s = (1 - q)/e^2 \in [s_{\text{lo}}, s_{\text{hi}}]$,

$$\begin{aligned} \frac{1}{e} \mathbb{E} P(v, e v_x, e^2 v_{xx})(q, X_q) &= P(1, 0, 0) \int \pi(y) \mathbb{P}(Y_s^\zeta(y) < 0) dy \\ &\quad + \int \pi(y) \mathbb{E} r_{P, \zeta}(s, Y_s^\zeta(y)) dy, \end{aligned} \tag{D.3}$$

with $\zeta = \zeta_\gamma$. Both integrals converge absolutely, by the majorants of Step 2 and $\int \pi = 1/e$.

STEP 2. Majorants. All constants in this step depend only on $(s_{\text{lo}}, s_{\text{hi}})$ and P ; they are uniform over $\zeta \in \mathcal{K}(s_{\text{lo}}, s_{\text{hi}})$, $s \in [s_{\text{lo}}, s_{\text{hi}}]$, e , and κ . By Lemma C.3(a),(b) for the class $\mathcal{K}(s_{\text{lo}}, s_{\text{hi}})$, applied from the level s_{hi} ,

$$|\mathbb{P}(Y_s^\zeta(y) < 0) - \mathbf{1}_{\{y < 0\}}| \leq C e^{-y^2/C}.$$

Indeed, for $y \leq -C$ the diffusion stays below $y/C < 0$ with probability at least $1 - 2e^{-y^2/C}$; for $y > 0$ it stays above $y/2 > 0$ with probability at least $1 - e^{-y^2/C}$; and

bounded $|y|$ is absorbed in the constant. By Lemma 7.2(b),(c), the functions v_ζ , $\partial_y v_\zeta$, and $\partial_y^2 v_\zeta$ are bounded, and

$$\begin{aligned} |1 - v_\zeta| + |\partial_y v_\zeta| + |\partial_y^2 v_\zeta| &\leq C(1 + z^2)^{-1} \quad (z < 0), \\ v_\zeta + |\partial_y v_\zeta| + |\partial_y^2 v_\zeta| &\leq Ce^{-z^2/C} \quad (z \geq 0), \end{aligned}$$

at (s, z) . Since P is Lipschitz on bounded sets and $P(0, 0, 0) = 0$, $|r_{P,\zeta}(s, z)|$ obeys the same two bounds. With Lemma C.3(a),(b) once more, $\mathbb{E}|r_{P,\zeta}(s, Y_s^\zeta(y))| \leq C(1 + y^2)^{-1}$.

STEP 3. *Convergence.* Let $e_n \rightarrow 0$ and $\kappa_n \rightarrow 0$, let $\zeta_n \in \mathcal{K}(s_{\text{lo}}, s_{\text{hi}})$ converge in L^1 to $\zeta^0 = \mathbf{1}_{[0, d_0]}$, let $s_n \rightarrow s_\infty \in (s_{\text{lo}}, s_{\text{hi}})$, and let $\pi_n = \pi_{e_n, \kappa_n}$. We claim that

$$\begin{aligned} \int \pi_n \mathbb{E} r_{P,\zeta_n}(s_n, Y_{s_n}^{\zeta_n}(y)) dy &\longrightarrow \phi_0 \int \mathbb{E} r_{P,\zeta^0}(s_\infty, Y_{s_\infty}^{\zeta^0}(y)) dy, \\ \int \pi_n [\mathbb{P}(Y_{s_n}^{\zeta_n}(y) < 0) - \mathbf{1}_{\{y < 0\}}] dy &\longrightarrow \phi_0 \int [\mathbb{P}(Y_{s_\infty}^{\zeta^0}(y) < 0) - \mathbf{1}_{\{y < 0\}}] dy. \end{aligned} \quad (\text{D.4})$$

By Lemma C.2, the dilated order parameters of ζ_n converge in L^1 to that of ζ^0 . Hence, by Lemma 2.4(f), $\partial_y^k U_{\zeta_n} \rightarrow \partial_y^k U_{\zeta^0}$ for $k \leq 4$, uniformly on $[s_{\text{lo}}, s_{\text{hi}}] \times [-M, M]$ for every M ; by Lemma 2.4(a) the limits are continuous; and $\ell_{\zeta_n} \rightarrow \ell_{\zeta^0}$ uniformly because $|\ell_{\zeta_n} - \ell_{\zeta^0}| \leq \|\zeta_n - \zeta^0\|_{L^1}$. Since $\partial_y^j v_\zeta = -\ell_\zeta \partial_y^{j+2} U_\zeta$, it follows that $P_{\zeta_n}(s_n, z_n) \rightarrow P_{\zeta^0}(s_\infty, z)$ whenever $z_n \rightarrow z$.

Fix a Brownian path. The difference $Y^{\zeta_n} - Y^{\zeta^0}$ of two solutions of (C.1) started from y solves an ordinary differential equation in the time $\varsigma = s_{\text{hi}} - s$, whose right side is $\zeta_n U_{\zeta_n, y}(Y^{\zeta_n}) - \zeta^0 U_{\zeta^0, y}(Y^{\zeta^0})$. The function $U_{\zeta_n, y}$ is Lipschitz in y with constant $1/s_{\text{lo}}$, because $|U_{\zeta, yy}| = v_\zeta / \ell_\zeta \leq 1/s_{\text{lo}}$ on $[s_{\text{lo}}, s_{\text{hi}}]$. The remaining terms $\zeta_n |U_{\zeta_n, y} - U_{\zeta^0, y}|(Y^{\zeta^0})$ and $|\zeta_n - \zeta^0| U_{\zeta^0, y}(Y^{\zeta^0})$ have integrals over $\varsigma \in [0, s_{\text{hi}} - s_{\text{lo}}]$ that tend to zero, because the path of Y^{ζ^0} is bounded and $\|\zeta_n - \zeta^0\|_{L^1} \rightarrow 0$. Gronwall's inequality gives $\sup_s |Y_{s_n}^{\zeta_n}(y) - Y_s^{\zeta^0}(y)| \rightarrow 0$, and continuity of the path gives $Y_{s_n}^{\zeta_n}(y) \rightarrow Y_{s_\infty}^{\zeta^0}(y)$.

The law of $Y_{s_\infty}^{\zeta^0}(y)$ has a density, because $s_\infty < s_{\text{hi}}$: in the dilated variables of Lemma C.2 it is the Parisi diffusion at a positive level, started from $y/\sqrt{s_{\text{hi}}}$ at level zero, which by Lemma 2.4(g) is a translate of the diffusion started from zero at the margin $-y/\sqrt{s_{\text{hi}}}$, and Lemma 2.7 applies. In particular the law does not charge 0, so $\mathbf{1}_{\{Y_{s_n}^{\zeta_n}(y) < 0\}} \rightarrow \mathbf{1}_{\{Y_{s_\infty}^{\zeta^0}(y) < 0\}}$ almost surely. Hence, for every y , $r_{P,\zeta_n}(s_n, Y_{s_n}^{\zeta_n}(y)) \rightarrow r_{P,\zeta^0}(s_\infty, Y_{s_\infty}^{\zeta^0}(y))$ almost surely. These functions are bounded uniformly in n , so bounded convergence gives convergence of the inner expectations and probabilities for every y . Since $\pi_n \rightarrow \phi_0$ pointwise and $\pi_n \leq 2\phi_0$, the majorants of Step 2 and dominated convergence in y give (D.4).

STEP 4. *Proof of (a) and (b).* (a) Suppose that the claim fails for some $\omega_0 > 0$. Then there are $e_n \rightarrow 0$, order parameters γ_n with $\hat{s}(\text{supp } \mu_{\gamma_n}) \subset [u_-, u_+]$ and $\|\gamma_n - \gamma_d\|_{L^1} / \delta_d \rightarrow$

0, and levels q_n with $\hat{s}(q_n) \in [u_-, u_+]$ and $|S_{\gamma_n, \kappa_d}(q_n) - S_d(q_n)| \geq \omega_0 e_n$, where $\kappa_d = \kappa_d(2 + e_n)$. Then $s_n = (1 - q_n)/e_n^2 = d_d \hat{s}(q_n) \in [\frac{1}{2}u_- d_0, 2u_+ d_0]$; pass to a subsequence with $s_n \rightarrow s_\infty \in (s_{\text{lo}}, s_{\text{hi}})$. The change of variables $q = 1 - e_n^2 s$ gives $\|\zeta_{\gamma_n} - \zeta_d\|_{L^1} = e_n^{-2} \|\gamma_n - \gamma_d\|_{L^1} \rightarrow 0$, and $\zeta_d = \mathbf{1}_{[0, d_d]} \rightarrow \zeta^0$ because $d_d \rightarrow d_0$. So both ζ_{γ_n} and ζ_d converge to ζ^0 . Both profiles are evaluated at the margin $\kappa_d \rightarrow 0$, hence with the same π_n . Apply (D.3) with $P(a_1, a_2, a_3) = a_1^2$, for which $\alpha e \cdot$ (left side) $= S$, to γ_n and to γ_d , writing $\mathbb{P}(Y < 0) = \mathbf{1}_{\{y < 0\}} + [\mathbb{P}(Y < 0) - \mathbf{1}_{\{y < 0\}}]$, and subtract. The terms $\int \pi_n \mathbf{1}_{\{y < 0\}}$ cancel, and by (D.4) for both sequences the remaining terms have the same limit. Hence $(S_{\gamma_n, \kappa_d}(q_n) - S_d(q_n))/(\alpha e_n) \rightarrow 0$, a contradiction.

(b) Let $\eta \leq \frac{1}{4}d_0$ and $e < e_{\text{st}}$, let $|\kappa - \kappa_d| \leq \eta e^2$, and let $\text{supp } \mu_\gamma \subset \mathcal{W}_\eta$. Then $\hat{s}(\text{supp } \mu_\gamma) \subset [1 - \eta/d_d, 1 + \eta/d_d] \subset [\frac{1}{2}, 2]$, and $\|\zeta_\gamma - \zeta_d\|_{L^1} = e^{-2} \int |t - q_d| d\mu_\gamma(t) \leq \eta$. A level $q \in [q_d - \eta e^2, Q]$ has $s = (1 - q)/e^2$ with $|s - d_d| \leq \eta$, because $1 - Q \geq \delta_d - \eta e^2$. Suppose that the convergence in (b) is not uniform. Then there are $\eta_n, e_n \rightarrow 0$, margins κ_n with $|\kappa_n - \kappa_d(2 + e_n)| \leq \eta_n e_n^2$, so that $\kappa_n \rightarrow 0$, order parameters γ_n with $\text{supp } \mu_{\gamma_n} \subset \mathcal{W}_{\eta_n}$, and levels q_n as above, along which $\tilde{x}(q_n)$ stays away from x_f or $\tilde{\Sigma}(q_n)/\tilde{B}(q_n)$ stays away from σ_f . Here $\zeta_{\gamma_n} \rightarrow \zeta^0$ in L^1 and $s_n = (1 - q_n)/e_n^2 \rightarrow d_0 \in (s_{\text{lo}}, s_{\text{hi}})$. Each of the polynomials

$$a_1^2(1 - a_1), \quad a_2^2, \quad (3a_1 - 1)a_2^2, \quad a_3^2, \quad a_1^2(1 - a_1)(1 - 3a_1),$$

evaluated at $(a_1, a_2, a_3) = (v, ev_x, e^2 v_{xx})$, vanishes at $(1, 0, 0)$ and at $(0, 0, 0)$. For such P , $r_{P, \zeta} = P_\zeta$, and (D.3), (D.4) give

$$\frac{1}{e_n} \mathbb{E} P(v, e_n v_x, e_n^2 v_{xx})(q_n, X_{q_n}) \longrightarrow \phi_0 \int \mathbb{E} P_{\zeta^0}(d_0, Y_{d_0}^{\zeta^0}(y)) dy.$$

On $(d_0, s_{\text{hi}}]$, $\zeta^0 = 0$, so $Y_{d_0}^{\zeta^0}$ is a Brownian motion there and $Y_{d_0}^{\zeta^0}(y) = y + W_{s_{\text{hi}} - d_0}$. On $(0, d_0]$, $\zeta^0 = 1$ and $U_{\zeta^0}(s, y) = \log \Phi(y/\sqrt{s})$ solves (7.3); hence $\ell_{\zeta^0}(d_0) = d_0$ and $v_{\zeta^0}(d_0, y) = V(y/\sqrt{d_0})$. The integrand $P_{\zeta^0}(d_0, \cdot)$ is signed, but by Step 2 it is bounded in absolute value by an integrable function; by Tonelli's theorem $\int \mathbb{E} |P_{\zeta^0}(d_0, y + W_{s_{\text{hi}} - d_0})| dy = \int |P_{\zeta^0}(d_0, z)| dz < \infty$. Fubini's theorem and the translation invariance of Lebesgue measure therefore give

$$\int \mathbb{E} P_{\zeta^0}(d_0, y + W_{s_{\text{hi}} - d_0}) dy = \int P_{\zeta^0}(d_0, z) dz = \sqrt{d_0} \int P\left(V(w), \frac{V'(w)}{\sqrt{d_0}}, \frac{V''(w)}{d_0}\right) dw.$$

Moreover $\lambda(q_n)/e_n^2 = \ell_{\zeta_{\gamma_n}}(s_n) \rightarrow d_0$. With $\ell = \lambda/e^2$,

$$\tilde{x} = \frac{\ell \mathbb{E}(ev_x)^2}{2\mathbb{E}v^2(1 - v)}, \quad \tilde{\Sigma} = \mathbb{E}\left[\ell^2(e^2 v_{xx})^2 + 4\tilde{x}\ell(3v - 1)(ev_x)^2 + 2\tilde{x}^2 v^2(1 - v)(1 - 3v)\right].$$

Hence $e_n^{-1} \tilde{B}(q_n) \rightarrow \phi_0 \sqrt{d_0} B_V > 0$, $e_n^{-1} \mathbb{E}(e_n v_x)^2 \rightarrow \phi_0 \sqrt{d_0} A_V/d_0$, so $\tilde{x}(q_n) \rightarrow x_f$, and then $e_n^{-1} \tilde{\Sigma}(q_n) \rightarrow \phi_0 \sqrt{d_0} \Sigma_{\text{flat}}$. Thus $\tilde{\Sigma}(q_n)/\tilde{B}(q_n) \rightarrow \sigma_f$, a contradiction. The consequence (8.6) follows because $x_f \in (0, \frac{1}{4})$ and $\sigma_f > 0$ (Lemma 8.2). $\square$

D.4 The bottom comparison

We first record a comparison principle for the linear parabolic equations that arise, with weights adapted to the Gaussian tails. Its first step is the maximum principle for parabolic equations on unbounded domains, for functions of polynomial growth (see [61]). For $k \geq 0$ and $c > 0$ put

$$\mathfrak{d}_{k,c}(y) = (1 + |y|)^{-k} \quad (y \leq 0), \quad \mathfrak{d}_{k,c}(y) = e^{-cy^2} \quad (y > 0).$$

Then $\mathfrak{d}_{k,c} \leq \mathfrak{d}_{k',c'}$ if $k \geq k'$ and $c \geq c'$, and $\mathfrak{d}_{k,c}\mathfrak{d}_{k',c'} \leq \mathfrak{d}_{k+k',\min\{c,c'\}}$.

Lemma D.2. *Let $0 < T \leq 1$, $k \geq 0$, $c_H > 0$, and $C_0, C_H \geq 0$. Let $0 = t_0 < t_1 < \dots < t_n = T$, and let β_1, β_0 be functions on $[0, T] \times \mathbb{R}$ whose restrictions to each $(t_{j-1}, t_j) \times \mathbb{R}$ extend continuously to $[t_{j-1}, t_j] \times \mathbb{R}$, with*

$$0 \leq \beta_1(t, y) \leq C_0(1 + y_-), \quad \beta_0 \leq 0.$$

Put $\mathsf{L} = \frac{1}{2}\partial_{yy} + \beta_1\partial_y + \beta_0$. Let f be continuous on $[0, T] \times \mathbb{R}$, of polynomial growth in y uniformly in t , of class $C^{1,2}$ on each $[t_{j-1}, t_j] \times \mathbb{R}$ (with one-sided t -derivatives at the endpoints), with $f(0, \cdot) = 0$ and $\partial_t f = \mathsf{L}f + H$ on each block.

- (i) If $|H| \leq C_H$, then $|f(t, y)| \leq C_H t$.
- (ii) If $|H(t, y)| \leq C_H t \mathfrak{d}_{k,c_H}(y)$, then $|f(t, y)| \leq C t^2 \mathfrak{d}_{k,\hat{c}}(y)$ with $\hat{c} = c_H / (1 + 2c_H)$, where C depends only on k, c_H, C_0 , and C_H .

Proof. We divide the proof into three steps.

STEP 1. Maximum principle. Let w be continuous on $[0, T] \times \mathbb{R}$, of class $C^{1,2}$ on each block, with $|w| \leq C_w(1 + |y|)^{p_0}$, $w(0, \cdot) \leq 0$, and $(\partial_t - \mathsf{L})w \leq 0$ on each block. We show that $w \leq 0$. Fix $p > p_0/2$ and put $\Psi(t, y) = e^{Kt}(1 + y^2)^p$. Then $|\partial_{yy}\Psi| \leq C_p\Psi$, and $|\beta_1\partial_y\Psi| \leq C_0(1 + |y|)2p|y|(1 + y^2)^{p-1}e^{Kt} \leq 4pC_0\Psi$, because $(1 + |y|)|y| \leq 2(1 + y^2)$; also $-\beta_0\Psi \geq 0$. Hence $(\partial_t - \mathsf{L})\Psi \geq (K - \frac{1}{2}C_p - 4pC_0)\Psi \geq \Psi$ for $K = K(p, C_0)$ large. For $\varepsilon > 0$ the function $w_\varepsilon = w - \varepsilon\Psi$ satisfies $(\partial_t - \mathsf{L})w_\varepsilon \leq -\varepsilon\Psi < 0$ on each block, $w_\varepsilon(0, \cdot) < 0$, and $w_\varepsilon \rightarrow -\infty$ as $|y| \rightarrow \infty$, uniformly in $t \in [0, T]$. Suppose that $\sup w_\varepsilon > 0$. The supremum is then attained at some (t^*, y^*) with $t^* > 0$. Choose j with $t^* \in (t_{j-1}, t_j]$. At (t^*, y^*) we have $\partial_y w_\varepsilon = 0$ and $\partial_{yy} w_\varepsilon \leq 0$, and the left t -derivative on the block $[t_{j-1}, t_j]$ is nonnegative. Since $w_\varepsilon(t^*, y^*) > 0$ and $\beta_0 \leq 0$,

$$(\partial_t - \mathsf{L})w_\varepsilon(t^*, y^*) = \partial_t w_\varepsilon - \frac{1}{2}\partial_{yy} w_\varepsilon - \beta_1\partial_y w_\varepsilon - \beta_0 w_\varepsilon \geq 0,$$

a contradiction. Hence $w_\varepsilon \leq 0$, and letting $\varepsilon \rightarrow 0$ gives $w \leq 0$.

STEP 2. Bounded sources. Since $(\partial_t - \mathsf{L})(C_H t) = C_H - \beta_0 C_H t \geq C_H$, the functions $w = \pm f - C_H t$ satisfy $(\partial_t - \mathsf{L})w \leq \pm H - C_H \leq 0$ and $w(0, \cdot) = 0$. Step 1 gives (i).

STEP 3. *Weighted sources.* Let $\chi \in C^\infty(\mathbb{R})$ satisfy $0 \leq \chi \leq 1$, $\chi = 1$ on $(-\infty, -1]$ and $\chi = 0$ on $[1, \infty)$. Put $c(t) = c_H / (1 + 2c_H t)$, so that $\dot{c}(t) = -2c(t)^2$ and $c(t) \geq \hat{c}$ for $t \leq T$, and

$$\mathfrak{w}_0(t, y) = \chi(y)(1 + y^2)^{-k/2} + (1 - \chi(y))e^{-c(t)y^2}.$$

For $y \geq 1$, $\mathfrak{w}_0 = e^{-c(t)y^2}$ and

$$(\partial_t - \frac{1}{2}\partial_{yy})\mathfrak{w}_0 = (-\dot{c}(t)y^2 - 2c(t)^2y^2 + c(t))\mathfrak{w}_0 = c(t)\mathfrak{w}_0 \geq 0;$$

moreover $-\beta_1\partial_y\mathfrak{w}_0 = 2c(t)y\beta_1\mathfrak{w}_0 \geq 0$ and $-\beta_0\mathfrak{w}_0 \geq 0$, so $(\partial_t - L)\mathfrak{w}_0 \geq 0$. For $y \leq -1$, $\mathfrak{w}_0 = (1 + y^2)^{-k/2}$ does not depend on t , $|\partial_y\mathfrak{w}_0| = k|y|(1 + y^2)^{-1}\mathfrak{w}_0 \leq k|y|^{-1}\mathfrak{w}_0$, and $|\partial_{yy}\mathfrak{w}_0| \leq C_k\mathfrak{w}_0$; since $\beta_1 \leq 2C_0|y|$ there, $(\partial_t - L)\mathfrak{w}_0 \geq -(\frac{1}{2}C_k + 2kC_0)\mathfrak{w}_0$. On $[0, T] \times [-1, 1]$, $\mathfrak{w}_0$ is smooth, bounded below by a positive constant, and its derivatives are bounded, with constants depending only on k and c_H (because $c(t) \in [\hat{c}, c_H]$); also $\beta_1 \leq 2C_0$ there. Hence $(\partial_t - L)\mathfrak{w}_0 \geq -C_2\mathfrak{w}_0$ on $[0, T] \times \mathbb{R}$, with C_2 depending only on k, c_H, C_0 . The barrier $\mathfrak{w} = e^{(C_2+1)t}\mathfrak{w}_0$ therefore satisfies

$$(\partial_t - L)\mathfrak{w} \geq \mathfrak{w}.$$

Comparing the two pieces of $\mathfrak{w}_0$ with $\mathfrak{d}_k$, on $y \leq -1$, $y \geq 1$, and $|y| \leq 1$, and using $(1 + |y|)^2 \leq 2(1 + y^2)$, we find C_3, C_4 , depending only on k, c_H , and C_2 , such that

$$\mathfrak{d}_{k, c_H} \leq C_3\mathfrak{w}, \quad \mathfrak{w} \leq C_4\mathfrak{d}_{k, \hat{c}} \quad \text{on } [0, T] \times \mathbb{R}.$$

Put $A = \frac{1}{2}C_H C_3$. Then $(\partial_t - L)(t^2\mathfrak{w}) = 2t\mathfrak{w} + t^2(\partial_t - L)\mathfrak{w} \geq 2t\mathfrak{w} \geq 2t\mathfrak{d}_{k, c_H}/C_3$, so $w = \pm f - At^2\mathfrak{w}$ satisfies $(\partial_t - L)w \leq C_H t\mathfrak{d}_{k, c_H} - 2At\mathfrak{d}_{k, c_H}/C_3 = 0$ and $w(0, \cdot) = 0$. Step 1 gives $|f| \leq At^2\mathfrak{w} \leq AC_4 t^2\mathfrak{d}_{k, \hat{c}}$. This proves (ii). $\square$

We now expand the solution near the top of the support around the Gaussian reference with the current value of λ . Fix $Q \in [\frac{1}{2}, 1)$, put $\delta = 1 - Q$, and fix a margin κ . For $\gamma' \in \mathcal{U}$ with $\gamma' = 1$ on $[Q, 1]$, $t \in [0, \frac{1}{8}]$ and $y \in \mathbb{R}$ put

$$\begin{aligned} \bar{\gamma}'(t) &= \gamma'(Q - \delta t), \quad l(t) = \frac{\lambda_{\gamma'}(Q - \delta t)}{\delta}, \\ \hat{b}(t, y) &= \sqrt{\delta} b_{\gamma'}(Q - \delta t, \kappa + \sqrt{\delta}y), \quad \varepsilon(t) = t - l(t) + 1. \end{aligned}$$

Then $l(0) = 1$ and $l' = \bar{\gamma}' \in [0, 1]$, so $1 \leq l \leq 1 + t \leq \frac{9}{8}$, $0 \leq \varepsilon \leq t$, and $\varepsilon' = 1 - \bar{\gamma}'$. The Gaussian reference is

$$f(l, y) = l^{-1/2}R(y/\sqrt{l}) = \partial_y \log \Phi(y/\sqrt{l}), \quad g = \frac{1}{2}\partial_{yy}f.$$

Lemma D.3. *There are numerical constants $C, c > 0$ such that, in this setting, for $t \in [0, \frac{1}{8}]$ and $y \in \mathbb{R}$,*

$$\begin{aligned} \hat{b}(t, y) &= f(l(t), y) + \varepsilon(t)g(l(t), y) + \mathcal{R}(t, y), \\ |\mathcal{R}(t, y)| &\leq Ct^2\mathfrak{d}_{3, c}(y), \quad |\partial_y \mathcal{R}(t, y)| \leq Ct^2\mathfrak{d}_{4, c}(y). \end{aligned} \tag{D.5}$$

Proof. We first let γ' be a step profile, and write $\bar{\gamma} = \gamma'$. We divide the proof into four steps.

STEP 1. *Equations and a priori bounds.* By Lemma 2.9(a), $\mathcal{L}b = 0$, that is, $b_q = -\frac{1}{2}b_{xx} - \gamma'bb_x$. Since $\partial_t = -\delta\partial_q$ and $\partial_y = \sqrt{\delta}\partial_x$, on each block of constancy of $\bar{\gamma}$

$$\hat{b}_t = \frac{1}{2}\hat{b}_{yy} + \bar{\gamma}\hat{b}\hat{b}_y, \quad \hat{b}(0, \cdot) = f(1, \cdot),$$

the initial condition by (2.10). By Lemma 2.4(a),(b), $\hat{b}$ is of class $C^{1,\infty}$ on each closed block, and it and its y -derivatives have polynomial growth. By Lemma 2.4(b),(c), $\hat{b} > 0$ and $\hat{b}_y = -\delta c = -v/l \in [-1, 0]$. By Lemma 2.10(b), $\lambda b(q, \kappa) \leq \sqrt{1-q}$, so $\hat{b}(t, 0) \leq \sqrt{1+t/l} \leq 2$. Hence

$$0 < \hat{b}(t, y) \leq 2 + y_-.$$

Since $\Phi(y/\sqrt{l})$ solves $\partial_l = \frac{1}{2}\partial_{yy}$, the function $\log \Phi(y/\sqrt{l})$ solves $\partial_l = \frac{1}{2}(\partial_{yy} + (\partial_y \cdot)^2)$, and differentiation in y gives

$$f_l = \frac{1}{2}f_{yy} + ff_y, \quad g_l = \frac{1}{2}g_{yy} + fg_y + 3f_yg,$$

the second by differentiating the first twice and using $(ff_y)_{yy} = ff_{yyy} + 3f_yf_{yy}$. Moreover $f_y = -l^{-1}V(y/\sqrt{l}) \in [-1, 0]$ and $\partial_y^j f = -l^{-(j+1)/2}V^{(j-1)}(y/\sqrt{l})$ for $j \geq 1$. By Lemma C.1(a), $|V^{(i)}(z)| \leq C_i(1+|z|)^{-2-i}$ for $z \leq 0$ and $|V^{(i)}(z)| \leq C_ie^{-z^2/4}$ for $z \geq 0$, for $i \geq 1$. Since $1 \leq l \leq \frac{9}{8}$, with $c_* = \frac{2}{9}$,

$$|\partial_y^j f(l, y)| \leq C\mathfrak{d}_{j+1, c_*}(y) \quad (2 \leq j \leq 5).$$

In particular $|g| \leq C\mathfrak{d}_{3, c_*}$, $|g_y| \leq C\mathfrak{d}_{4, c_*}$, $|g_{yy}| \leq C\mathfrak{d}_{5, c_*}$ and $|g_{yyy}| \leq C\mathfrak{d}_{6, c_*}$.

STEP 2. *First order.* Put $\Delta = \hat{b} - f(l(t), \cdot)$. Using $\frac{d}{dt}f(l(t), y) = \bar{\gamma}f_l$ and $\hat{b}\hat{b}_y - ff_y = \hat{b}\Delta_y + f_y\Delta$, on each block

$$\Delta_t = L\Delta + (1 - \bar{\gamma})g, \quad L = \frac{1}{2}\partial_{yy} + \bar{\gamma}\hat{b}\partial_y + \bar{\gamma}f_y, \quad \Delta(0, \cdot) = 0.$$

The drift $\beta_1 = \bar{\gamma}\hat{b}$ satisfies $0 \leq \beta_1 \leq 2 + y_-$, and the reaction $\beta_0 = \bar{\gamma}f_y$ is nonpositive. The source is bounded, so Lemma D.2(i) gives $|\Delta| \leq Ct$. Differentiating in y ,

$$\partial_t\Delta_y = L'\Delta_y + \bar{\gamma}f_{yy}\Delta + (1 - \bar{\gamma})g_y, \quad L' = \frac{1}{2}\partial_{yy} + \bar{\gamma}\hat{b}\partial_y + \bar{\gamma}(\hat{b}_y + f_y),$$

with the same drift, the nonpositive reaction $\bar{\gamma}(\hat{b}_y + f_y)$, and a bounded source. Hence also $|\Delta_y| \leq Ct$.

STEP 3. *Second order.* Put $\mathcal{R} = \Delta - \varepsilon g(l(t), \cdot)$. Since $\varepsilon' = 1 - \bar{\gamma}$ and $\frac{d}{dt}g(l(t), y) = \bar{\gamma}g_l$, the equations of Steps 1 and 2 give exactly

$$\mathcal{R}_t = L\mathcal{R} + h, \quad h = \varepsilon \left[\frac{1}{2}(1 - \bar{\gamma})g_{yy} + \bar{\gamma}\Delta g_y - 2\bar{\gamma}f_yg \right], \quad \mathcal{R}(0, \cdot) = 0.$$

By Steps 1 and 2 and $0 \leq \varepsilon \leq t$,

$$|h| \leq Ct\mathfrak{d}_{3,c_*}, \quad |h_y| = \left| \varepsilon \left[\frac{1}{2}(1 - \bar{\gamma})g_{yyy} + \bar{\gamma}\Delta_y g_y + \bar{\gamma}\Delta g_{yy} - 2\bar{\gamma}f_{yy}g - 2\bar{\gamma}f_y g_y \right] \right| \leq Ct\mathfrak{d}_{4,c_*}.$$

Lemma D.2(ii) with $k = 3$ gives $|\mathcal{R}| \leq Ct^2\mathfrak{d}_{3,c_1}$ with $c_1 = c_*/(1 + 2c_*)$.

STEP 4. *The derivative, and general profiles.* Differentiating the equation for $\mathcal{R}$ in y ,

$$\partial_t \mathcal{R}_y = \mathcal{L}' \mathcal{R}_y + \bar{\gamma} f_{yy} \mathcal{R} + h_y.$$

Here $|\bar{\gamma} f_{yy} \mathcal{R}| \leq C\mathfrak{d}_{3,c_*} t^2 \mathfrak{d}_{3,c_1} \leq Ct\mathfrak{d}_{4,c_1}$ and $|h_y| \leq Ct\mathfrak{d}_{4,c_1}$, so Lemma D.2(ii) with $k = 4$ gives $|\mathcal{R}_y| \leq Ct^2\mathfrak{d}_{4,c}$ with $c = c_1/(1 + 2c_1)$. All constants are numerical. This proves (D.5) for step profiles.

For a general γ' , choose step profiles $\gamma'_n \in \mathcal{U}_Q$ with $\gamma'_n \rightarrow \gamma'$ in L^1 . For instance, for partitions $0 = r_0 < \dots < r_N = Q$ of mesh at most $1/n$ put $\gamma'_n = \gamma'(r_i)$ on $[r_i, r_{i+1})$ and $\gamma'_n = 1$ on $[Q, 1]$; then $0 \leq \gamma' - \gamma'_n$ and $\|\gamma'_n - \gamma'\|_{L^1} \leq 1/n$. If γ' has a positive bottom q_0 , one may instead take the profiles of Lemma 6.2, which moreover have the same q_0 , the same top, and the same mean. By Lemma 2.4(f), $\hat{b}_n \rightarrow \hat{b}$ and $\partial_y \hat{b}_n \rightarrow \partial_y \hat{b}$ pointwise, and $l_n \rightarrow l$, $\varepsilon_n \rightarrow \varepsilon$ uniformly because $|\lambda_{\gamma'_n} - \lambda_{\gamma'}| \leq \|\gamma'_n - \gamma'\|_{L^1}$. Since f , g , and their y -derivatives are continuous in l , $\mathcal{R}_n \rightarrow \mathcal{R}$ and $\partial_y \mathcal{R}_n \rightarrow \partial_y \mathcal{R}$ pointwise, and the bounds, whose constants do not depend on n , pass to the limit. $\square$

Proof of Lemma 8.7. Let $\eta_3 \leq \min\{\eta_1, \frac{1}{64}d_0\}$, let e be small, $|\kappa - \kappa_d| \leq \eta_3 e^2$, and let $\text{supp } \mu_\gamma \subset \mathcal{W}_{\eta_3}$. Then $\delta = 1 - Q$ and δ_m lie in $[\frac{1}{4}d_0 e^2, 3d_0 e^2]$, $q_0 \geq \frac{1}{2}$, and $t_0 = (Q - q_0)/\delta = e^2 \tilde{\ell}/\delta \leq 4\tilde{\ell}/d_0 \leq 8\eta_3/d_0 \leq \frac{1}{8}$. Also $m - q_0 \leq Q - q_0 = e^2 \tilde{\ell}$.

Comparison at the bottom. Both γ and $\gamma_m = \mathbf{1}_{[m,1]}$ equal one on $[Q, 1]$ and vanish on $[0, q_0)$. At $t = t_0$, which corresponds to the level q_0 , both have $l(t_0) = \lambda(q_0)/\delta = \delta_m/\delta$, and hence the same $\varepsilon(t_0)$. By Lemma D.3 applied to both profiles, the first two terms in (D.5) cancel in the difference. Since $v = -l\hat{b}_y$ and $l(t_0) \leq \frac{9}{8}$, at $x = \kappa + \sqrt{\delta}y$,

$$\sqrt{\delta} |b_\gamma - b_{\gamma_m}|(q_0, x) \leq 2Ct_0^2 \mathfrak{d}_{3,c}(y), \quad |v_\gamma - v_{\gamma_m}|(q_0, x) \leq 3Ct_0^2 \mathfrak{d}_{4,c}(y). \quad (\text{D.6})$$

Under both profiles $X_{q_0} = B_{q_0}$, and $y = (B_{q_0} - \kappa)/\sqrt{\delta}$ has the density $\sqrt{\delta} \phi_{q_0}(\kappa + \sqrt{\delta}y) \leq \sqrt{\delta}$, as $q_0 \geq \frac{1}{2}$. Since $0 < v < 1$, integrating (D.6) gives

$$|S_\gamma(q_0) - S_{\gamma_m}(q_0)| \leq 2\alpha \mathbb{E} |v_\gamma - v_{\gamma_m}|(q_0, X_{q_0}) \leq Ct_0^2 \sqrt{\delta} \leq Ce\tilde{\ell}^2.$$

Both profiles have $\lambda = \delta_m$ on $[0, q_0]$, so the entropy parts of $D_\gamma(q_0)$ and $D_{\gamma_m}(q_0)$ agree. The bound $0 < \hat{b} \leq 2 + y_-$ of Step 1 of the proof of Lemma D.3 uses only Lemma 2.4(b),(c) and Lemma 2.10(b), so it holds for every profile; hence $\sqrt{\delta}(b_\gamma + b_{\gamma_m}) \leq 2(2 + y_-)$, and $(2 + y_-)\mathfrak{d}_{3,c}(y)$ is integrable. Hence

$$|D_\gamma(q_0) - D_{\gamma_m}(q_0)| \leq \alpha \mathbb{E} |b_\gamma^2 - b_{\gamma_m}^2|(q_0, X_{q_0}) \leq \frac{Ct_0^2}{\delta} \sqrt{\delta} \leq \frac{C\tilde{\ell}^2}{e}.$$

The replica-symmetric order parameter between q_0 and m . On $[0, m]$, $\gamma_m = 0$ and $\lambda_{\gamma_m} = \delta_m$. By (8.4), $S'_{\gamma_m} = \alpha \mathcal{N}^m$ there, where $\mathcal{N}^m(r) = \mathbb{E} v_x(r, B_r)^2$ is computed for γ_m . At $r = m$, $v(m, x) = V((x - \kappa)/\sqrt{\delta_m})$ by (2.10), so $\mathcal{N}^m(m) = A_w(m)/\delta_m$ and $S_{\gamma_m}(m) = S_{\text{RS}}(m)$. By Lemma 2.9(c)(iv) with $\gamma_m = 0$, $(\mathcal{N}^m)' = \mathbb{E} v_{xx}^2 \geq 0$ on $[0, m]$. Also v_{xx} solves the backward heat equation there, because $\mathcal{L}v_{xx} = 0$ where the order parameter vanishes (Lemma 2.9(a)); by Jensen's inequality and the density bound $\sqrt{\delta_m}$ for $(B_m - \kappa)/\sqrt{\delta_m}$,

$$(\mathcal{N}^m)'(r) \leq \mathbb{E} v_{xx}(m, B_m)^2 = \delta_m^{-2} \mathbb{E} V''\left(\frac{B_m - \kappa}{\sqrt{\delta_m}}\right)^2 \leq \mathcal{I}_2 \delta_m^{-3/2}.$$

Hence $0 \leq \mathcal{N}^m(m) - \mathcal{N}^m(r) \leq \mathcal{I}_2 \delta_m^{-3/2} (m - r)$ for $r \in [q_0, m]$, and

$$\begin{aligned} \left| S_{\text{RS}}(m) - S_{\gamma_m}(q_0) - \alpha(m - q_0) \frac{A_w(m)}{\delta_m} \right| &= \alpha \int_{q_0}^m (\mathcal{N}^m(m) - \mathcal{N}^m(r)) \, dr \\ &\leq \frac{\alpha \mathcal{I}_2 (m - q_0)^2}{2 \delta_m^{3/2}} \leq C e \tilde{\ell}^2. \end{aligned}$$

Together with the bound on $|S_{\gamma}(q_0) - S_{\gamma_m}(q_0)|$ this proves the first estimate. Since $0 \leq \mathcal{N}^m \leq \mathcal{N}^m(m)$ on $[q_0, m]$, Lemma 8.5(c) gives

$$\sup_{r \in [q_0, m]} |S_{\gamma_m}(r) - S_{\text{RS}}(m)| \leq \alpha(m - q_0) \frac{A_w(m)}{\delta_m} \leq C e \tilde{\ell},$$

because $m \in \mathcal{W}_{\eta_1}$. Finally, integrating $D'_{\gamma_m} = (S_{\gamma_m} - 1)/\delta_m^2$ (Lemma 2.9(c)(i)) from q_0 to m and using $D_{\gamma_m}(m) = D_{\text{RS}}(m)$ gives the formula for $D_{\gamma_m}(q_0)$. $\square$

References

- [1] Michael Aizenman, Robert Sims, and Shannon L. Starr. Extended variational principle for the Sherrington–Kirkpatrick spin-glass model. *Phys. Rev. B*, 68(21):214403, 2003.
- [2] Shlomo Alexander. Amorphous solids: their structure, lattice dynamics and elasticity. *Phys. Rep.*, 296(2-4):65–236, 1998.
- [3] Brandon L. Annesi, Enrico M. Malatesta, and Francesco Zamponi. Exact full-RSB SAT/UNSAT transition in infinitely wide two-layer neural networks. *SciPost Phys.*, 18(4):118, 2025.
- [4] Louis-Pierre Arguin. Spin glass computations and Ruelle’s probability cascades. *J. Stat. Phys.*, 126(4-5):951–976, 2007.
- [5] P. M. Aronow and Patrick Lopatto. Replica symmetry breaking at arbitrary depth. Preprint, arXiv:2609.05317, 2026.
- [6] Antonio Auffinger and Wei-Kuo Chen. Free energy and complexity of spherical bipartite models. *J. Stat. Phys.*, 157(1):40–59, 2014.
- [7] Antonio Auffinger and Wei-Kuo Chen. The Parisi formula has a unique minimizer. *Comm. Math. Phys.*, 335(3):1429–1444, 2015.
- [8] Antonio Auffinger and Wei-Kuo Chen. On properties of Parisi measures. *Probab. Theory Related Fields*, 161(3-4):817–850, 2015.
- [9] Antonio Auffinger and Wei-Kuo Chen. Parisi formula for the ground state energy in the mixed p -spin model. *Ann. Probab.*, 45(6B):4617–4631, 2017.
- [10] Antonio Auffinger and Aukosh Jagannath. On spin distributions for generic p -spin models. *J. Stat. Phys.*, 174(2):316–332, 2019.
- [11] Antonio Auffinger, Wei-Kuo Chen, and Qiang Zeng. The SK model is infinite step replica symmetry breaking at zero temperature. *Comm. Pure Appl. Math.*, 73(5):921–943, 2020.
- [12] Richard Baraniuk, Mark Davenport, Ronald DeVore, and Michael Wakin. A simple proof of the restricted isometry property for random matrices. *Constr. Approx.*, 28(3):253–263, 2008.
- [13] Jean Barbier, Florent Krzakala, Nicolas Macris, Léo Miolane, and Lenka Zdeborová. Optimal errors and phase transitions in high-dimensional generalized linear models. *Proc. Natl. Acad. Sci. USA*, 116(12):5451–5460, 2019.
- [14] Adriano Barra, Pierluigi Contucci, Emanuele Mingione, and Daniele Tantari. Multi-species mean field spin glasses. Rigorous results. *Ann. Henri Poincaré*, 16(3):691–708, 2015.
- [15] Erik Bates and Youngtak Sohn. Crisanti–Sommers formula and simultaneous symmetry breaking in multi-species spherical spin glasses. *Comm. Math. Phys.*, 394(3):1101–1152, 2022.

- [16] Erik Bates and Youngtak Sohn. Free energy in multi-species mixed p -spin spherical models. *Electron. J. Probab.*, 27:Paper No. 52, 75 pp., 2022.
- [17] Erik Bates and Youngtak Sohn. Balanced multi-species spin glasses. Preprint, arXiv:2507.06522, 2025.
- [18] Dimitris Bertsimas and John N. Tsitsiklis. *Introduction to Linear Optimization*, volume 6 of *Athena Scientific Series in Optimization and Neural Computation*. Athena Scientific, Belmont, MA, 1997.
- [19] Erwin Bolthausen and Alain-Sol Sznitman. On Ruelle’s probability cascades and an abstract cavity method. *Comm. Math. Phys.*, 197(2):247–276, 1998.
- [20] Erwin Bolthausen, Shuta Nakajima, Nike Sun, and Changji Xu. Gardner formula for Ising perceptron models at small densities. In *Proceedings of Thirty Fifth Conference on Learning Theory*, volume 178 of *Proc. Mach. Learn. Res.*, pages 1787–1911. PMLR, 2022.
- [21] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press, Oxford, 2013.
- [22] Michelle Boué and Paul Dupuis. A variational representation for certain functionals of Brownian motion. *Ann. Probab.*, 26(4):1641–1659, 1998.
- [23] M. Bouten. Replica symmetry instability in perceptron models. *J. Phys. A*, 27:6021, 1994.
- [24] Anton Bovier and Irina Kurkova. Derrida’s generalised random energy models 1: models with finitely many hierarchies. *Ann. Inst. H. Poincaré Probab. Statist.*, 40(4):439–480, 2004.
- [25] Herm Jan Brascamp and Elliott H. Lieb. On extensions of the Brunn–Minkowski and Prékopa–Leindler theorems, including inequalities for log concave functions, and with an application to the diffusion equation. *J. Funct. Anal.*, 22(4):366–389, 1976.
- [26] Theophilos Cacoullos. On upper and lower bounds for the variance of a function of a random variable. *Ann. Probab.*, 10(3):799–809, 1982.
- [27] Luis A. Caffarelli. Monotonicity properties of optimal transportation and the FKG and related inequalities. *Comm. Math. Phys.*, 214:547–563, 2000. Erratum: *Comm. Math. Phys.* 225 (2002), 449–450.
- [28] Emmanuel J. Candès and Terence Tao. Decoding by linear programming. *IEEE Trans. Inform. Theory*, 51(12):4203–4215, 2005.
- [29] Patrick Charbonneau, Jorge Kurchan, Giorgio Parisi, Pierfrancesco Urbani, and Francesco Zamponi. Fractal free energy landscapes in structural glasses. *Nat. Commun.*, 5:3725, 2014.
- [30] Patrick Charbonneau, Jorge Kurchan, Giorgio Parisi, Pierfrancesco Urbani, and Francesco Zamponi. Exact theory of dense amorphous hard spheres in high dimension. III. The full replica symmetry breaking solution. *J. Stat. Mech. Theory Exp.*, 2014(10):P10009, 2014.
- [31] Patrick Charbonneau, Enzo Marinari, Marc Mézard, Giorgio Parisi, Federico Ricci-Tersenghi, Gabriele Sicuro, and Francesco Zamponi, editors. *Spin Glass Theory and Far Beyond: Replica Symmetry Breaking after 40 Years*. World Scientific, Singapore, 2023.
- [32] Hong-Bin Chen. FRSB in the SK spin glass: convergence to full-interval support at zero temperature. Preprint, arXiv:2607.18032, 2026.

- [33] Hong-Bin Chen. On free energy of non-convex multi-species spin glasses. *ALEA Lat. Am. J. Probab. Math. Stat.*, 23(1):429–473, 2026.
- [34] Hong-Bin Chen and Jean-Christophe Mourrat. On the free energy of vector spin glasses with nonconvex interactions. *Probab. Math. Phys.*, 6(1):1–80, 2025.
- [35] Hong-Bin Chen and Jean-Christophe Mourrat. Free energy of non-convex multi-species spherical spin glasses. Preprint, arXiv:2609.32710, 2026.
- [36] Hong-Bin Chen, Victor Issa, and Jean-Christophe Mourrat. Free energy of non-convex multi-species spin glasses with centered Ising spins. Preprint, arXiv:2606.16636, 2026.
- [37] Wei-Kuo Chen. The Aizenman–Sims–Starr scheme and Parisi formula for mixed p -spin spherical models. *Electron. J. Probab.*, 18:Paper No. 94, 14 pp., 2013.
- [38] Wei-Kuo Chen. Variational representations for the Parisi functional and the two-dimensional Guerra–Talagrand bound. *Ann. Probab.*, 45(6A):3929–3966, 2017.
- [39] Wei-Kuo Chen. On the Almeida–Thouless transition line in the Sherrington–Kirkpatrick model with centered Gaussian external field. *Electron. Commun. Probab.*, 26:1–9, 2021. doi: 10.1214/21-ECP439.
- [40] Wei-Kuo Chen and Arnab Sen. Parisi formula, disorder chaos and fluctuation for the ground state energy in the spherical mixed p -spin models. *Comm. Math. Phys.*, 350(1): 129–173, 2017.
- [41] Herman Chernoff. A note on an inequality involving the normal distribution. *Ann. Probab.*, 9(3):533–535, 1981.
- [42] Julian D. Cole. On a quasi-linear parabolic equation occurring in aerodynamics. *Quart. Appl. Math.*, 9:225–236, 1951.
- [43] Thomas M. Cover. Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition. *IEEE Trans. Electron. Comput.*, EC-14(3):326–334, 1965.
- [44] Andrea Crisanti and Hans-Jürgen Sommers. The spherical p -spin interaction spin glass model: the statics. *Z. Phys. B*, 87(3):341–354, 1992.
- [45] Jairo R. L. de Almeida and David J. Thouless. Stability of the Sherrington–Kirkpatrick solution of a spin glass model. *J. Phys. A*, 11(5):983–990, 1978.
- [46] Sudhakar Dharmadhikari and Kumar Joag-Dev. *Unimodality, Convexity, and Applications*. Probability and Mathematical Statistics. Academic Press, Boston, MA, 1988.
- [47] Persi Diaconis and David Freedman. A dozen de Finetti-style results in search of a theory. *Ann. Inst. H. Poincaré Probab. Statist.*, 23(2, suppl.):397–423, 1987.
- [48] Jian Ding and Nike Sun. Capacity lower bound for the Ising perceptron. *Probab. Theory Related Fields*, 193(3-4):627–715, 2025.
- [49] Tomas Dominguez and Jean-Christophe Mourrat. *Statistical Mechanics of Mean-Field Disordered Systems: A Hamilton–Jacobi Approach*. Zurich Lectures in Advanced Mathematics. EMS Press, Berlin, 2024. All numbered items refer to arXiv:2311.08976v2.
- [50] Joseph L. Doob. Conditional Brownian motion and the boundary limits of harmonic functions. *Bull. Soc. Math. France*, 85:431–458, 1957.

- [51] L. N. Dovbysh and Vladimir N. Sudakov. Gram–de Finetti matrices. *Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI)*, 119:77–86, 1982.
- [52] Bertrand Duplantier. Comment on Parisi’s equation for the SK model for spin glasses. *J. Phys. A*, 14(1):283, 1981.
- [53] Bradley Efron. Tweedie’s formula and selection bias. *J. Amer. Statist. Assoc.*, 106(496):1602–1614, 2011.
- [54] Ahmed El Alaoui and Mark Sellke. Algorithmic pure states for the negative spherical perceptron. *J. Stat. Phys.*, 189(2):Paper No. 27, 35 pp., 2022. All numbered items refer to arXiv:2010.15811v1.
- [55] Ahmed El Alaoui, Andrea Montanari, and Mark Sellke. Optimization of mean-field spin glasses. *Ann. Probab.*, 49(6):2922–2960, 2021.
- [56] Simon Foucart and Holger Rauhut. *A Mathematical Introduction to Compressive Sensing. Applied and Numerical Harmonic Analysis*. Birkhäuser/Springer, New York, 2013.
- [57] Silvio Franz and Michele Leone. Replica bounds for optimization problems and diluted spin systems. *J. Stat. Phys.*, 111(3-4):535–564, 2003.
- [58] Silvio Franz and Giorgio Parisi. The simplest model of jamming. *J. Phys. A*, 49(14):145001, 2016.
- [59] Silvio Franz, Giorgio Parisi, Maksim Sevelev, Pierfrancesco Urbani, and Francesco Zamponi. Universality of the SAT–UNSAT (jamming) threshold in non-convex continuous constraint satisfaction problems. *SciPost Phys.*, 2(3):019, 2017.
- [60] Silvio Franz, Sungmin Hwang, and Pierfrancesco Urbani. Jamming in multilayer supervised learning models. *Phys. Rev. Lett.*, 123(16):160602, 2019.
- [61] Avner Friedman. *Partial Differential Equations of Parabolic Type*. Prentice-Hall, Englewood Cliffs, NJ, 1964.
- [62] Elizabeth Gardner. The space of interactions in neural network models. *J. Phys. A*, 21(1):257–270, 1988.
- [63] Elizabeth Gardner and Bernard Derrida. Optimal storage properties of neural network models. *J. Phys. A*, 21(1):271–284, 1988.
- [64] Stefano Ghirlanda and Francesco Guerra. General properties of overlap probability distributions in disordered spin systems. Towards Parisi ultrametricity. *J. Phys. A*, 31(46):9149–9155, 1998.
- [65] Robert D. Gordon. Values of Mills’ ratio of area to bounding ordinate and of the normal probability integral for large values of the argument. *Ann. Math. Statist.*, 12(3):364–366, 1941.
- [66] Yehoram Gordon. Some inequalities for Gaussian processes and applications. *Israel J. Math.*, 50:265–289, 1985.
- [67] Francesco Guerra. Sum rules for the free energy in the mean field spin glass model. In Roberto Longo, editor, *Mathematical Physics in Mathematics and Physics: Quantum and Operator Algebraic Aspects (Siena, 2000)*, volume 30 of *Fields Inst. Commun.*, pages 161–170. Amer. Math. Soc., Providence, RI, 2001.

- [68] Francesco Guerra. Broken replica symmetry bounds in the mean field spin glass model. *Comm. Math. Phys.*, 233(1):1–12, 2003.
- [69] Fu-Hsuan Ho. Convergence of the free energy of vector spin glasses with non-convex interactions. Preprint, arXiv:2609.39307, 2026.
- [70] Alan J. Hoffman. On approximate solutions of systems of linear inequalities. *J. Research Nat. Bur. Standards*, 49(4):263–265, 1952.
- [71] Eberhard Hopf. The partial differential equation $u_t + uu_x = \mu u_{xx}$. *Comm. Pure Appl. Math.*, 3:201–230, 1950.
- [72] Brice Huang. Capacity threshold for the Ising perceptron. In *2024 IEEE 65th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1126–1136. IEEE Computer Soc., 2024.
- [73] Brice Huang, Mark Sellke, and Nike Sun. Algorithmic threshold for high-dimensional projection pursuit I: general theory. Preprint, arXiv:2608.29416, 2026.
- [74] Nobuyuki Ikeda and Shinzo Watanabe. A comparison theorem for solutions of stochastic differential equations and its applications. *Osaka J. Math.*, 14(3):619–633, 1977.
- [75] Aukosh Jagannath and Subhabrata Sen. On the unbalanced cut problem and the generalized Sherrington–Kirkpatrick model. *Ann. Inst. H. Poincaré D*, 8(1):35–88, 2021.
- [76] Aukosh Jagannath and Ian Tobasco. A dynamic programming approach to the Parisi functional. *Proc. Amer. Math. Soc.*, 144(7):3135–3150, 2016.
- [77] Aukosh Jagannath and Ian Tobasco. Some properties of the phase diagram for mixed p -spin glasses. *Probab. Theory Related Fields*, 167(3-4):615–672, 2017.
- [78] Aukosh Jagannath and Ian Tobasco. Low temperature asymptotics of spherical mean field spin glasses. *Comm. Math. Phys.*, 352(3):979–1017, 2017.
- [79] Ioannis Karatzas and Steven E. Shreve. *Brownian Motion and Stochastic Calculus*, volume 113 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1991.
- [80] Chinubhai G. Khatri. On certain inequalities for normal distributions and their applications to simultaneous confidence bounds. *Ann. Math. Statist.*, 38(6):1853–1867, 1967.
- [81] John F. C. Kingman. *Poisson Processes*, volume 3 of *Oxford Studies in Probability*. The Clarendon Press, Oxford University Press, New York, 1993.
- [82] Pax Kivimae. The ground state energy and concentration of complexity in spherical bipartite models. *Comm. Math. Phys.*, 403(1):37–81, 2023.
- [83] Werner Krauth and Marc Mézard. Storage capacity of memory networks with binary couplings. *J. Physique*, 50(20):3057–3066, 1989.
- [84] James Kuelbs, Wenbo V. Li, and Werner Linde. The Gaussian measure of shifted balls. *Probab. Theory Related Fields*, 98:143–162, 1994.
- [85] Wenbo V. Li and Qi-Man Shao. Gaussian processes: inequalities, small ball probabilities and applications. In D. N. Shanbhag and C. R. Rao, editors, *Stochastic Processes: Theory and Methods*, volume 19 of *Handbook of Statistics*, pages 533–597. North-Holland, Amsterdam, 2001.

- [86] Patrick Lopatto. Replica symmetry up to the de Almeida–Thouless line in the Sherrington–Kirkpatrick model. Preprint, arXiv:2604.11921, 2026.
- [87] Patrick Lopatto. Full replica symmetry breaking in the Sherrington–Kirkpatrick model. Preprint, arXiv:2607.11756. All numbered items refer to arXiv:2607.11756v3, 2026.
- [88] Marc Mézard, Giorgio Parisi, and Miguel Angel Virasoro. *Spin Glass Theory and Beyond*, volume 9 of *World Scientific Lecture Notes in Physics*. World Scientific, Singapore, 1987.
- [89] Andrea Montanari and Kangjie Zhou. Which exceptional low-dimensional projections of a Gaussian point cloud can be found in polynomial time? Preprint, arXiv:2406.02970. To appear in *Ann. Probab.*, 2024.
- [90] Andrea Montanari, Yiqiao Zhong, and Kangjie Zhou. Tractability from overparametrization: the example of the negative perceptron. *Probab. Theory Related Fields*, 188(3-4):805–910, 2024. All numbered items refer to arXiv:2110.15824v3.
- [91] Cristian F. Moukarzel. Isostatic phase transition and instability in stiff granular materials. *Phys. Rev. Lett.*, 81(8):1634–1637, 1998.
- [92] Jean-Christophe Mourrat. Nonconvex interactions in mean-field spin glasses. *Probab. Math. Phys.*, 2(2):281–339, 2021. All numbered items refer to arXiv:2004.01679v8.
- [93] Jean-Christophe Mourrat. The Parisi formula is a Hamilton–Jacobi equation in Wasserstein space. *Canad. J. Math.*, 74(3):607–629, 2022.
- [94] Jean-Christophe Mourrat. Free energy upper bound for mean-field vector spin glasses. *Ann. Inst. H. Poincaré Probab. Statist.*, 59(3):1143–1182, 2023.
- [95] Corey S. O’Hern, Leonardo E. Silbert, Andrea J. Liu, and Sidney R. Nagel. Jamming at zero temperature and zero applied stress: the epitome of disorder. *Phys. Rev. E*, 68(1):011306, 2003.
- [96] Dmitry Panchenko. On the Dovbysh–Sudakov representation result. *Electron. Commun. Probab.*, 15:330–338, 2010.
- [97] Dmitry Panchenko. The Parisi ultrametricity conjecture. *Ann. of Math. (2)*, 177(1):383–393, 2013.
- [98] Dmitry Panchenko. *The Sherrington–Kirkpatrick Model*. Springer Monographs in Mathematics. Springer, New York, 2013.
- [99] Dmitry Panchenko. The Parisi formula for mixed p -spin models. *Ann. Probab.*, 42(3):946–958, 2014.
- [100] Dmitry Panchenko. The free energy in a multi-species Sherrington–Kirkpatrick model. *Ann. Probab.*, 43(6):3494–3513, 2015.
- [101] Dmitry Panchenko. Free energy in the mixed p -spin models with vector spins. *Ann. Probab.*, 46(2):865–896, 2018.
- [102] Dmitry Panchenko and Michel Talagrand. Bounds for diluted mean-fields spin glass models. *Probab. Theory Related Fields*, 130(3):319–336, 2004.
- [103] Dmitry Panchenko and Michel Talagrand. Guerra’s interpolation using Derrida–Ruelle cascades. Preprint, arXiv:0708.3641, 2007.

- [104] Giorgio Parisi. Infinite number of order parameters for spin-glasses. *Phys. Rev. Lett.*, 43(23): 1754–1756, 1979.
- [105] Giorgio Parisi. A sequence of approximated solutions to the S–K model for spin glasses. *J. Phys. A*, 13(4):L115–L121, 1980.
- [106] Giorgio Parisi and Francesco Zamponi. A proof of an identity for the critical exponents of jamming. *J. Stat. Mech. Theory Exp.*, 2026(7):073301, 2026.
- [107] Giorgio Parisi, Pierfrancesco Urbani, and Francesco Zamponi. *Theory of Simple Glasses: Exact Solutions in Infinite Dimensions*. Cambridge University Press, Cambridge, 2020.
- [108] András Prékopa. On logarithmic concave measures and functions. *Acta Sci. Math. (Szeged)*, 34:335–343, 1973.
- [109] Yuri V. Prokhorov. Convergence of random processes and limit theorems in probability theory. *Theory Probab. Appl.*, 1(2):157–214, 1956.
- [110] Daniel Revuz and Marc Yor. *Continuous Martingales and Brownian Motion*, volume 293 of *Grundlehren der Mathematischen Wissenschaften*. Springer-Verlag, Berlin, third edition, 1999.
- [111] Herbert Robbins. An empirical Bayes approach to statistics. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, Vol. I*, pages 157–163. University of California Press, 1956.
- [112] David Ruelle. A mathematical reformulation of Derrida’s REM and GREM. *Comm. Math. Phys.*, 108(2):225–239, 1987.
- [113] Michael R. Sampford. Some inequalities on Mill’s ratio and related functions. *Ann. Math. Statist.*, 24(1):130–132, 1953.
- [114] Moshe Shaked and J. George Shanthikumar. *Stochastic Orders*. Springer Series in Statistics. Springer, New York, 2007.
- [115] Mariya Shcherbina and Brunello Tirozzi. Rigorous solution of the Gardner problem. *Comm. Math. Phys.*, 234(3):383–422, 2003.
- [116] Zbyněk Šidák. Rectangular confidence regions for the means of multivariate normal distributions. *J. Amer. Statist. Assoc.*, 62(318):626–633, 1967.
- [117] Mihailo Stojnic. Negative spherical perceptron. Preprint, arXiv:1306.3980, 2013.
- [118] Mihailo Stojnic. Another look at the Gardner problem. Preprint, arXiv:1306.3979, 2013.
- [119] Mihailo Stojnic. Fl RDT based ultimate lowering of the negative spherical perceptron capacity. Preprint, arXiv:2312.16531, 2023.
- [120] Mihailo Stojnic. Fully lifted interpolating comparisons of bilinearly indexed random processes. Preprint, arXiv:2311.18092, 2023.
- [121] Mihailo Stojnic. Bilinearly indexed random processes—stationarization of fully lifted interpolation. Preprint, arXiv:2311.18097, 2023.
- [122] Eliran Subag. TAP approach for multispecies spherical spin glasses II: The free energy of the pure models. *Ann. Probab.*, 51(3):1004–1024, 2023.
- [123] Michel Talagrand. Intersecting random half-spaces: toward the Gardner–Derrida formula. *Ann. Probab.*, 28(2):725–758, 2000.

- [124] Michel Talagrand. The Parisi formula. *Ann. of Math. (2)*, 163(1):221–263, 2006.
- [125] Michel Talagrand. Free energy of the spherical mean field model. *Probab. Theory Related Fields*, 134(3):339–382, 2006.
- [126] Michel Talagrand. Parisi measures. *J. Funct. Anal.*, 231(2):269–286, 2006.
- [127] 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, 2011.
- [128] Michel Talagrand. *Mean Field Models for Spin Glasses. Volume II: Advanced Replica-Symmetry and Low Temperature*, volume 55 of *Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge*. Springer, Berlin, 2011.
- [129] Fabio Lucio Toninelli. About the Almeida–Thouless transition line in the Sherrington–Kirkpatrick mean-field spin glass model. *Europhys. Lett.*, 60(5):764–767, 2002.
- [130] Roman Vershynin. Introduction to the non-asymptotic analysis of random matrices. In Yonina C. Eldar and Gitta Kutyniok, editors, *Compressed Sensing*, pages 210–268. Cambridge University Press, Cambridge, 2012.
- [131] James G. Wendel. A problem in geometric probability. *Math. Scand.*, 11:109–111, 1962.
- [132] Matthieu Wyart. Marginal stability constrains force and pair distributions at random close packing. *Phys. Rev. Lett.*, 109(12):125502, 2012.
- [133] Changji Xu. Sharp threshold for the Ising perceptron model. *Ann. Probab.*, 49(5):2399–2415, 2021.
- [134] Yuxin Zhou. Existence of full replica symmetry breaking for the Sherrington–Kirkpatrick model at low temperature. *Comm. Pure Appl. Math.*, 79(7):1746–1770, 2026.