

Dynamics of the spherical Sherrington–Kirkpatrick model

Part I: Foundations and tools

P. M. Aronow

Patrick Lopatto

AUTHORS' NOTE

With the exception of this note and small edits to the abstract, 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 remains lacking in exposition and overall coherence. We are nonetheless releasing this manuscript on Hexagon to disseminate the result as quickly as possible, because we believe it may be of interest to the mathematics and physics communities. We welcome suggestions concerning mathematical corrections, omitted citations, and attribution of ideas.

For publication, the manuscript is divided into two volumes: *Part I: Foundations and tools* (Sections 1–12) and *Part II: Mixing profiles and critical dynamics* (Sections 13–21). Section, theorem, equation, and citation numbers are shared across the volumes, and each volume includes the complete bibliography.

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

ABSTRACT

We study the Langevin dynamics of the spherical Sherrington–Kirkpatrick model, the diffusion on the sphere of radius $\sqrt{N}$ in $\mathbb{R}^N$ that is reversible with respect to the Gibbs measure with density proportional to $\exp(\beta x^\top W x / \sqrt{2})$, where W is an $N \times N$ matrix from the Gaussian orthogonal ensemble. The dynamics is started from the uniform distribution on the sphere, independently of W . We determine the asymptotic quenched total-variation distance to equilibrium as a function of time in the fixed-temperature, mesoscopic, critical-window, and infinite-temperature crossover regimes described below, with convergence in probability or in distribution with respect to W . Let $\beta_c = 1/\sqrt{2}$ be the critical inverse temperature. For $\beta < \beta_c$ there is cutoff with a Gaussian profile at fixed $\beta \in (0, \beta_c)$, along every mesoscopic approach of β to β_c from below, and in the infinite-temperature crossover when $N\beta_N^2 \rightarrow \infty$; when $N\beta_N^2$ stays bounded, the distance relaxes over bounded times and there is no cutoff. At fixed $\beta \in (0, \beta_c)$ the mixing time is $(\log N - (3/2) \log \log N + O(1)) / (4(1 - \sqrt{2}\beta)^2)$, and the correction $-(3/2) \log \log N$ comes from the edge of the spectrum of W . For $\beta > \beta_c$, at fixed β and along every mesoscopic approach to β_c from above, there is cutoff at a time set by the gap between the two largest eigenvalues of W , of order $N^{2/3} \log N$ at fixed β . The profile is the total-variation distance between a standard Gaussian variable and its sum with an independent Cauchy variable; the Cauchy variable is the ratio of the first two eigencoordinates at an early stage of the evolution. In the critical window $\sqrt{2}\beta = 1 + \sigma N^{-1/3}$, the distance at time $N^{2/3}s$ converges in distribution to a random profile determined by the Airy point process, and there is no cutoff. As $\sigma \rightarrow \pm\infty$, this profile, suitably recentered and rescaled, converges to the profiles of the two mesoscopic regimes.

Contents

1	Introduction	1
1.1	The model and its Langevin dynamics	1
1.2	Cutoff and limit profiles	5
1.3	Main results	7
1.4	Discussion	13
1.5	Related work	17
1.6	Ideas of the proofs	19
1.7	Organization of the paper.	20
2	Overview of the proofs	21
2.1	Common reductions	21
2.2	High temperature	24
2.3	Low temperature	30
2.4	The critical window.	34
2.5	The critical mixing profile far from the critical temperature	38
2.6	Dependencies	39
PART I	Preliminaries	41
3	The model, conventions, and the limiting profiles	42
3.1	The model	42
3.2	Equilibrium parameters and temperature regimes	47
3.3	Distances, mixing times, and profiles	49
3.4	Disorder quantifiers and constants	51
3.5	The limiting profiles	53
4	Random-matrix inputs and the equilibrium Lagrange multiplier	57
4.1	Published random-matrix inputs	58
4.2	Eigenvalue gaps at fixed confidence.	63
4.3	Edge counts at a varying scale	65
4.4	The Lagrange multiplier at high temperature	69
4.5	The critical window.	74

PART II	Tools	82
5	Gaussian laws conditioned on a sphere	83
5.1	Notation	83
5.2	Densities of weighted sums of centered squares	84
5.3	Conditioning on the sphere: finite sequences	88
5.4	Countable sequences and coordinate marginals	90
5.5	Moments of coordinate marginals.	94
5.6	Exponential tilts and quadratic statistics.	95
5.7	Continuity in the variance sequence.	98
6	Likelihood ratios on a common sphere	101
6.1	Total variation and likelihood ratios.	102
6.2	Integration by parts under the conditioned law	102
6.3	Normal approximation of quadratic statistics	105
6.4	Likelihood ratios	108
7	Deterministic covariance evolution and heat estimates	113
7.1	The heat matrix.	114
7.2	The covariance evolution and the Bingham comparison law	119
7.3	Estimates under two-sided precision-spacing bounds	124
7.4	Estimates in the critical window	131
7.5	Positive measures from real squared resolvents	133
8	Functional inequalities on the sphere	139
8.1	Conventions and standard tools.	140
8.2	Ambient calculus for the quadratic generator	143
8.3	Many precisions comparable to the smallest one.	147
8.4	Logarithmic Sobolev inequalities from a Gaussian extension.	151
8.5	A few precisions close to the smallest one	159
8.6	The critical window.	169
9	Comparison with Bingham laws through the forward equation	173
9.1	Notation	174
9.2	The forward-equation defect	176
9.3	Comparison through the defect	179
9.4	Reverse variance inequalities and exponential decay.	181
9.5	An inverse moment under conditioned Gaussian laws.	184
9.6	Comparison when Nd^2 is large	185
9.7	Comparison for nearly isotropic matrices	188
9.8	A Poisson equation for Bingham laws.	189
10	The unnormalized process and its moments	195
10.1	The unnormalized process	196

10.2	Positive martingales	201
10.3	Moment bounds with backward semigroup weights.	203
10.4	The critical window.	208
10.5	Above the critical temperature: moments and occupation	212
10.6	Initial laws equal to Bingham comparison laws	217
10.7	The maximal inverse of the squared norm.	221
10.8	A Feynman–Kac bound for a weighted gradient.	226
11	Score bounds, smoothing, and conditional relative entropy	228
11.1	Setting, weak scores, and weighted endpoint measures	229
11.2	A Bismut-type formula for the score over a final time interval	232
11.3	Weak convergence to total variation convergence by Gaussian smoothing	237
11.4	Fisher information of finitely many coordinates	240
11.5	Conditional relative entropy of the remaining coordinates	244
11.6	Combining the coordinates	259
12	Density bounds relative to time-dependent Bingham laws	260
12.1	Setting and notation	261
12.2	The coefficients of the comparison law	264
12.3	Exponential moments of quadratic functions	268
12.4	A differential inequality for Rényi divergences	269
12.5	Density bounds on the time interval $[0, S]$	273
12.6	The critical window.	275
	References	279

Introduction

The spherical Sherrington–Kirkpatrick model is the simplest mean-field spin glass. Its configurations are points x of the sphere $S_N = \{x \in \mathbb{R}^N : |x|^2 = N\}$, and its energy is a quadratic form in x whose coefficients form a random symmetric matrix, so that its equilibrium can be analyzed exactly through the spectrum of that matrix. The Langevin diffusion on S_N that is reversible with respect to the Gibbs measure has been studied in physics for more than forty years, mostly through the evolution of observables such as the energy and two-time correlation functions in the limit of infinite system size.

The goal of this paper is to describe how the law of this diffusion approaches equilibrium when the dimension N is large and the disorder is fixed. We start the dynamics from the uniform distribution on S_N , independently of the disorder, and measure its distance to the Gibbs measure in total variation. We determine the asymptotic behavior of this distance as a function of time at fixed temperatures above and below the critical one, in the crossover to infinite temperature, along every mesoscopic approach to the critical temperature from either side, and in the critical window. In the regimes treated here outside the critical window we find cutoff, except in the part of the crossover to infinite temperature where $N\beta^2$ stays bounded; there the distance relaxes over bounded times. In each case we identify the center and the window of the transition and its limiting shape, which we call the mixing profile. In the critical window the mixing profile is random and determined by the edge of the spectrum of the disorder, and there is no cutoff. We also show that the critical mixing profiles match those of the two mesoscopic regimes far from the critical temperature.

1.1 The model and its Langevin dynamics

Spherical models. Berlin and Kac introduced the spherical model of a ferromagnet in [28], in which the Ising spins of a lattice model are replaced by real variables subject to a single constraint on the sum of their squares. Lewis and Wannier considered the mean spherical model, in which this constraint holds only in expectation [107, (1)–(2), p. 683], and Stanley identified the spherical model with the limit of classical vector spin models as the number of spin components tends to infinity [140]. Sherrington and Kirkpatrick introduced the mean-field Ising spin glass with independent Gaussian couplings [132]. Kosterlitz, Thouless, and Jones introduced its spherical version, whose couplings form a

random symmetric matrix, and solved its thermodynamics through the spectrum of that matrix [97].

We use the following normalization. Let W_N be an $N \times N$ matrix from the Gaussian orthogonal ensemble (GOE), that is, a real symmetric matrix whose entries on and above the diagonal are independent with $W_{ij} \sim \mathcal{N}(0, (1 + \mathbf{1}_{\{i=j\}})/N)$ for $i \leq j$. Let $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_N$ be its eigenvalues, and let ν_N be the uniform probability measure on S_N . For an inverse temperature $\beta \geq 0$ we define the Hamiltonian and the Gibbs measure by

$$H_N(x) = \frac{x^\top W_N x}{\sqrt{2}}, \quad \pi_{N,c}(dx) = \frac{1}{Z_{N,c}} e^{\beta H_N(x)} \nu_N(dx) = \frac{1}{Z_{N,c}} e^{(c/2)x^\top W_N x} \nu_N(dx), \quad (1.1)$$

where $Z_{N,c}$ is the normalizing constant and $c = \sqrt{2}\beta$. Throughout the paper we parameterize the temperature by the rescaled inverse temperature c , so that increasing c lowers the temperature. The critical value is $c = 1$, that is, $\beta_c = 1/\sqrt{2}$. Since its density with respect to ν_N is the exponential of a quadratic form, $\pi_{N,c}$ is a Bingham distribution [31]. Because $-W_N$ has the same law as W_N , negative inverse temperatures lead to the same results with the roles of the top and the bottom of the spectrum exchanged, and it suffices to consider $\beta \geq 0$.

We work in an orthonormal eigenbasis of W_N and write $x_1, \dots, x_N$ for the coordinates in this basis, the eigencoordinates. The Gibbs measure has the following Gaussian representation. Let $z > c\lambda_1$ be the unique solution of

$$\frac{1}{N} \sum_{i=1}^N \frac{1}{z - c\lambda_i} = 1, \quad (1.2)$$

and set $p_i = z - c\lambda_i$ and $q_i = 1/p_i$. Since $-\sum_i p_i x_i^2/2 = -zN/2 + (c/2)\sum_i \lambda_i x_i^2$ on S_N , the density of a centered Gaussian vector with independent coordinates of variances $q_1, \dots, q_N$ is proportional to the Gibbs density on S_N . The gradient of $|x|^2$ has constant length on S_N , and it follows that $\pi_{N,c}$ is the law of this Gaussian vector conditioned on $|x|^2 = N$, where the choice of z in (1.2) makes the expected squared norm of the unconditioned vector equal to N . The numbers z , p_i , and q_i are the Lagrange multiplier, the precisions, and the variances of Definition 3.3. Equation (1.2) is the saddle-point equation in the solution of Kosterlitz, Thouless, and Jones [97]; see [6, Lemma 1.3] for a modern account of their contour-integral formula for $Z_{N,c}$. The representation of Bingham distributions by conditioned Gaussian vectors is classical in directional statistics [98], and for spherical models it appears in [5, Proposition 2.7] and, in the critical regime of the present model, in [62, Section 1.2].

As $N \rightarrow \infty$, the empirical spectral measure of W_N converges to the semicircle law $\rho_{\text{sc}}(\lambda) d\lambda = (2\pi)^{-1} \sqrt{4 - \lambda^2} d\lambda$ on $[-2, 2]$ [152, 153]; for convergence in probability in the present normalization, see [4, Theorem 2.1.1, p. 7]. Also, λ_1 converges to 2 by the edge

estimates recalled in Lemma 4.2. For $c < 1$ we have $z \rightarrow 1 + c^2$ and the smallest precision $d = p_1 = z - c\lambda_1$ converges to $(1 - c)^2$, so that all variances q_i remain bounded. For $c > 1$ we have $d \rightarrow 0$, and the first eigencoordinate carries a macroscopic fraction of the squared norm under $\pi_{N,c}$, in the sense that $x_1^2/N \rightarrow 1 - 1/c$ in probability. This is the condensation of the spherical model onto the top eigenvector [97], [51, Section III]. At low temperature the Gibbs measure is therefore concentrated on two antipodal regions in which x_1 is close to $\sqrt{(1 - 1/c)N}$ or to $-\sqrt{(1 - 1/c)N}$. The free energy has a phase transition at $c = 1$ [97]; we refer to [47] for the replica analysis of general spherical p -spin models and to [143] for a proof of the free-energy formula for spherical models with even interactions, which include the present one.

Langevin dynamics. The Langevin dynamics of the model is the diffusion on S_N with generator

$$L = \Delta_{S_N} + \beta \nabla_{S_N} H_N \cdot \nabla_{S_N}, \quad (1.3)$$

where Δ_{S_N} and ∇_{S_N} are the Laplace–Beltrami operator and the gradient on S_N . Its noise coefficient is $\sqrt{2}$, and it is reversible with respect to $\pi_{N,c}$. We write $(P_t)_{t \geq 0}$ for its semigroup, so that $\nu_N P_t$ is the law at time t of the diffusion started from the uniform distribution. In the eigenbasis the drift is linear apart from the term that keeps the process on the sphere. However, the noise is projected onto the tangent space of S_N , and for finite N the eigencoordinates are coupled and are not independent Ornstein–Uhlenbeck processes.

This dynamics has a long history in physics. Shukla and Singh studied relaxation in the spherical spin-glass model [133]. Ciuchi and de Pasquale studied nonlinear relaxation and ergodicity breaking in random anisotropy spin glasses with many spin components [44]. Cugliandolo and Dean solved the dynamics completely in the thermodynamic limit, for quenches from high temperature and for more general temperature protocols [51]. They diagonalized the coupling matrix and reduced the spherical constraint to a scalar Volterra equation for a time-dependent Lagrange multiplier [51, (II.2)–(II.7)], and from its solution they computed correlation and response functions and showed that the dynamics ages at low temperature. Zippold, Kühn, and Horner refined the time scales that separate stationary relaxation from aging and compared the model with spherical ferromagnets [156, Sections III–IV]. Cugliandolo and Kurchan developed a microscopic aging analysis for spherical $p > 2$ spin glasses after taking the thermodynamic limit [50, Introduction], and the quadratic model behaves like a coarsening system [52, Section IV.C]. Caltagirone, Parisi, and Rizzo computed the off-equilibrium correlations at the critical temperature [38, Section III]. On the mathematical side, Ben Arous, Dembo, and Guionnet proved aging for the Langevin dynamics of the spherical Sherrington–Kirkpatrick model with a soft spherical constraint [26]. The same authors derived rigorously the Cugliandolo–Kurchan equations for the correlation and response functions of spherical spin glasses with a soft

spherical constraint [23, Theorem 1.2]. Dembo, Guionnet, and Mazza obtained the limit of these equations under the hard spherical constraint [55, Proposition 1.1], and proved that at sufficiently high temperature the solutions converge to a stationary pair that satisfies the fluctuation-dissipation relation and decays exponentially [55, Theorem 1.3]. These analyses take the thermodynamic limit first and then study the resulting correlation and response functions, including their long-time behavior.

Finite-size effects enter through the edge of the spectrum. Rodgers and Moore identified the first-gap barrier between the two antipodal low-temperature regions [124, Section 2, (25)–(26)]; see also [73, Section 4]. Fyodorov, Perret, and Schehr showed that at zero temperature the relaxation crosses over at times of order $N^{2/3}$ from the algebraic decay of the thermodynamic limit to an exponential decay governed by the random gap $\lambda_1 - \lambda_2$, in a regime where observables are no longer self-averaging [73, Section 3.1]. Barbier, de Freitas Pimenta, Cugliandolo, and Stariolo studied these finite-size effects further, including the escape from initial conditions concentrated near saddle points [13]. For gradient descent on the sphere started from the uniform distribution, which is the dynamics at zero temperature, Yu showed that the time needed to reach a fixed overlap with an extreme eigenvector lies between orders $N^{2/3}$ and $N^{2/3} \log N$ with probability at least $1 - \delta$, for each fixed $\delta > 0$ [155, Section 3, Theorem 3.3]. That result concerns a hitting time, while our results concern the law of the diffusion at positive temperature in total variation. In an approximation that neglects finite-size fluctuations of the spectral density, Cugliandolo, Dean, and Yoshino found a relaxation time of order $N(\beta - \beta_c)$ for the linear susceptibility at low temperature [53, Sections 1.1 and 4.3].

The question. In this paper we study the law of the diffusion itself at finite N . We start the diffusion from ν_N , independently of W_N , and measure its distance to equilibrium by

$$D_N(t) = \|\nu_N P_t - \pi_{N,c}\|_{\text{TV}}, \quad \text{where } \|\mu - \eta\|_{\text{TV}} = \sup_A |\mu(A) - \eta(A)| \quad (1.4)$$

and the supremum is over Borel sets $A \subset S_N$. We ask how $D_N(t)$ behaves as a function of t when N is large, and how the answer depends on the temperature.

We comment on three features of this formulation. First, total variation controls the law of the entire configuration, so that once $D_N(t)$ is small the law of every measurable observable is close to its equilibrium law, including observables that depend on all N coordinates; it is also the standard distance in the theory of mixing times [105]. Second, the uniform distribution is the Gibbs measure at $\beta = 0$, and the dynamics started from ν_N describes a quench from infinite temperature. In the eigenbasis a uniform point has coordinates $\sqrt{N} g_i / |g|$, where $g_1, \dots, g_N$ are independent standard Gaussian variables. Several of the computations cited above instead model the quench by giving every eigenmode the same deterministic initial amplitude [51, (III.1)], [73, (6)], [13, (12)]. For a fixed matrix this is a different initial condition, and Section 1.4 explains why the

difference matters at low temperature. Third, the law $\nu_N P_t$ is averaged over the initial point and the Brownian motion but not over the disorder, so that $D_N(t)$ is a random variable that depends on W_N only, and our results are statements in probability or in distribution with respect to W_N .

1.2 Cutoff and limit profiles

Diaconis and Shahshahani [58] and Aldous [1] found the cutoff phenomenon in their analysis of card shuffling and random walks on groups, and Aldous and Diaconis named it [2]; we refer to [56] and [105, Chapter 18] for surveys. Given a sequence of Markov processes, let $t_N(\varepsilon)$ be the first time at which the distance to equilibrium is at most ε . Ratios of mixing times are set to 1 when their denominator is zero. In the deterministic setting, there is cutoff if $t_N(1 - \varepsilon) > 0$ for all sufficiently large N and $t_N(\varepsilon)/t_N(1 - \varepsilon) \rightarrow 1$ for every $\varepsilon \in (0, 1/2)$, which means that the distance drops from close to one to close to zero over a window of time that is much shorter than the mixing time. A finer question asks for a limit profile, which consists of a center T_N , a window w_N with $w_N/T_N \rightarrow 0$, and a function P such that the distance at time $T_N + w_N u$ converges to $P(u)$ for every $u \in \mathbb{R}$.

Limit profiles are known for few processes. Diaconis, Graham, and Morrison studied the random walk on the hypercube $\{0, 1\}^n$ that at each step stays put or flips one of the n coordinates, each with probability $1/(n + 1)$. They proved in [59, Theorem 1] that its distance to the uniform distribution at time $(n \log n)/4 + un$ converges to

$$\Psi(u) = 2\Phi(e^{-2u}/2) - 1, \quad (1.5)$$

where Φ is the standard normal distribution function. The function $\Psi(u)$ is the total-variation distance between $\mathcal{N}(e^{-2u}, 1)$ and $\mathcal{N}(0, 1)$, and it is also the profile in our results at high temperature. Bayer and Diaconis found the profile of the riffle shuffle [21, Theorem 4], and Lacoïn found that of the simple exclusion process on the circle [100, Theorem 2.1]. Teyssier showed that the profile of random transpositions is a total-variation distance between two Poisson laws [144, Theorem 1.1], and Nestoridi and Olesker-Taylor developed a general method for the profiles of reversible Markov chains [114].

Cutoff is also known for many diffusions. Saloff-Coste [128, Theorems 2.1 and 4.2] and Méliot [113, Theorem 6] studied Brownian motion on spheres and other compact symmetric spaces, and Lachaud proved cutoff for samples of independent Ornstein–Uhlenbeck processes [99, Propositions 1–2]. Boursier, Chafaï, and Labbé proved it for the Dyson Ornstein–Uhlenbeck process [35] and computed exact profiles for Ornstein–Uhlenbeck processes [35, Lemma 3.4]. Barrera and Jara [14, 15] studied total-variation cutoff for small Gaussian perturbations of attracting dynamical systems, where the

profile is obtained through a Gaussian approximation [15, Theorem 2.2]. For small Lévy perturbations, Barrera, Högele, and Pardo treated Ornstein–Uhlenbeck systems in Wasserstein distance [16, Section 3] and nonlinear Langevin systems with layered stable noise in total variation [17, Section 1]. Chen and Saloff-Coste gave cutoff criteria for maximum L^p distances with $1 < p < \infty$ [41, Theorem 4.2]. Salez treated nonnegatively curved Markov chains [126, Theorem 1] and obtained a total-variation criterion of product-condition type for nonnegatively curved diffusions started from deterministic points [127, Theorem 1 and Corollary 1]. Chafaï and Fathi proved cutoff under a rigidity condition relating curvature and spectral gap [40, Corollaries 1.3–1.4], and Pedrotti and Salez bounded the total-variation window through Poincaré constants of the evolving laws, which allows arbitrary initial distributions [119, Theorem 1]. These criteria do not determine the profiles studied here.

Mixing across a phase transition is best understood for the Curie–Weiss model. Levin, Luczak, and Peres proved that the Glauber dynamics of the mean-field Ising model has cutoff at high temperature and mixes in a time of order $n^{3/2}$ at the critical temperature, while at low temperature it mixes exponentially slowly and a dynamics restricted to configurations of nonnegative magnetization mixes rapidly [106, Theorems 1–3 and Section 5]. Ding, Lubetzky, and Peres determined the mixing time through the critical window and along both mesoscopic approaches to it, which are governed by the parameter $n\delta^2$ where δ is the distance to the critical temperature [60, Theorems 1–3]. In [61, Theorem 1] they showed that at low temperature a censored dynamics, which removes the bottleneck created by the sign of the magnetization, has cutoff. Cuff, Ding, Luidor, Lubetzky, Peres, and Sly analyzed the mean-field Potts model, whose dynamical transition differs from its equilibrium transition [49]. In this paper we obtain a picture of the same kind for a disordered model, with two differences: the mesoscopic parameter is $N\delta^3$ instead of $n\delta^2$, and in the critical window the profile is random, since it depends on the edge of the spectrum of W_N .

The mixing time also depends on the initial distribution. Lubetzky and Sly showed that the one-dimensional stochastic Ising model started from the uniform distribution mixes in half the time needed from almost every fixed initial configuration [108, Theorem 2]; see also [83] for cutoff from given initial distributions. The quantity $D_N(t)$ concerns the law obtained by averaging over the initial point, which can be much closer to equilibrium than the law started from a typical fixed point. Our results therefore do not describe worst-case mixing (see Section 1.4).

1.3 Main results

Definitions

Throughout, W_N is a GOE matrix and the initial distribution is ν_N , independent of W_N . The rescaled inverse temperature may depend on N , in which case we write c_N and evaluate D_N , π_{N,c_N} , and the quantities below at $c = c_N$; the temperature sequences need not be monotone. Convergence in probability and in distribution refers to the randomness of W_N , and $X_N = o_{\mathbb{P}}(1)$ means that $X_N \rightarrow 0$ in probability. By the contraction property of Markov semigroups, $D_N(t)$ is nonincreasing in t . The mixing time at accuracy $\varepsilon \in (0, 1)$ is

$$t_N(\varepsilon) = \inf\{t \geq 0 : D_N(t) \leq \varepsilon\}.$$

We say that D_N has cutoff profile P with center T_N and window w_N , which may be random, if $w_N/T_N \rightarrow 0$ in probability and

$$\sup_{u \in K} |D_N(T_N + w_N u) - P(u)| \rightarrow 0$$

in probability for every compact set $K \subset \mathbb{R}$ (Definition 3.7). In this random setting, cutoff also requires $\mathbb{P}(t_N(1 - \varepsilon) > 0) \rightarrow 1$ for every $\varepsilon \in (0, 1/2)$. In every supremum of this kind we restrict to offsets u for which the evaluation time $T_N + w_N u$ is nonnegative, as in Definition 3.7; equivalently, we set $D_N(t) = D_N(0)$ for $t < 0$. Since $w_N/T_N \rightarrow 0$ in probability, either convention changes nothing in the limit. If P is continuous and decreases strictly from 1 to 0, then Lemma 3.11 implies that $(t_N(\varepsilon) - T_N)/w_N \rightarrow P^{-1}(\varepsilon)$ in probability for every $\varepsilon \in (0, 1)$, and in particular that there is cutoff. More generally, a mixing profile is a limit of the functions $u \mapsto D_N(T_N + w_N u)$, or of $s \mapsto D_N(b_N s)$ for a time scale b_N . The critical mixing profiles below are random, and they are not cutoff profiles.

Besides the function Ψ in (1.5), we need the profiles

$$\begin{aligned} \mathcal{V}(u) &= \|\text{Law}(G + e^{-2u}C) - \text{Law}(G)\|_{\text{TV}}, \\ \mathcal{V}_\theta(u) &= \|\text{Law}(G + \sqrt{2\theta}e^{-2u}C) - \text{Law}(G)\|_{\text{TV}}, \end{aligned} \tag{1.6}$$

where G is a standard Gaussian variable, C is an independent standard Cauchy variable, and $\theta > 0$. The law of $G + aC$ is the convolution of a Gaussian and a Cauchy law, and its density is a Voigt profile [149], [117, Section 7.19]. Both Ψ and $\mathcal{V}$ are continuous and decrease strictly from 1 at $-\infty$ to 0 at $+\infty$ (Lemma 3.9), and $\mathcal{V}_\theta(u) = \mathcal{V}(u - \log(2\theta)/4)$. Solving $\Psi(u) = \varepsilon$, we find $\Psi^{-1}(\varepsilon) = -\log(2\Phi^{-1}((1 + \varepsilon)/2))/2$.

At low temperature we use the scaled eigenvalue gaps a_i and the top fraction θ_N of Section 3.2, given by

$$a_i = \frac{c}{2}(\lambda_1 - \lambda_i), \quad \theta_N = 1 - \frac{1}{N} \sum_{i=2}^N \frac{1}{2a_i}. \tag{1.7}$$

Since $c\lambda_1|x|^2/2$ is constant on S_N , we have $\pi_{N,c}(dx) \propto \exp(-\sum_{i \geq 2} a_i x_i^2) \nu_N(dx)$. A centered Gaussian vector $(x_2, \dots, x_N)$ with variances $(2a_i)^{-1}$ has expected squared norm $(1 - \theta_N)N$, which is why θ_N approximates the squared projection x_1^2/N of $x/\sqrt{N}$ onto the top eigenvector under $\pi_{N,c}$. For fixed $c > 1$ we have $\theta_N \rightarrow 1 - 1/c$ in probability, and for $c_N = 1 + \epsilon_N$ with $\epsilon_N \rightarrow 0$ and $N\epsilon_N^3 \rightarrow \infty$ we have $\theta_N/\epsilon_N \rightarrow 1$ in probability (Section 4). The convergence (1.11) of the top eigenvalues of W_N to the Airy point process, stated below, implies that $N^{2/3}a_2$ converges in distribution to a positive random variable.

At high temperature we use a deterministic evolution of variances, defined for a fixed matrix W_N and $c \geq 0$. Let $v_1(t), \dots, v_N(t)$ solve

$$v_i'(t) = 2 + 2(c\lambda_i - z(t))v_i(t), \quad v_i(0) = 1, \quad z(t) = 1 + \frac{1}{N} \sum_{j=1}^N c\lambda_j v_j(t). \quad (1.8)$$

These are the variances at time t of the linear system $dx_i = (c\lambda_i - z(t))x_i dt + \sqrt{2} dB_i$ with independent standard Gaussian initial values, in which the spherical constraint holds only in expectation because the choice of $z(t)$ keeps $\sum_i v_i(t) = N$. Cugliandolo and Dean solved this system in the thermodynamic limit through the scalar Volterra equation mentioned above [51, (II.4)–(II.7)]. As $t \rightarrow \infty$, $z(t)$ converges to the Lagrange multiplier z and $v_i(t)$ converges to q_i . The Gaussian law with variances $v_i(t)$ conditioned on S_N is the Bingham comparison law μ_t at time t , defined in Definition 7.5. Its distance to $\pi_{N,c}$ is governed by the relative errors and by the information of $v(t)$ relative to q (Definition 3.5),

$$e_i(t) = \frac{v_i(t)}{q_i} - 1, \quad \mathcal{I}_N(t) = \frac{1}{2} \sum_{i=1}^N e_i(t)^2. \quad (1.9)$$

Away from the critical window

Our first result gives the profiles in the regimes treated here outside the critical window. Each of its five cases has its own section, and Table 1.1 below summarizes the results.

Theorem 1.1 (Cutoff profiles away from the critical window). *Suppose that the rescaled inverse temperatures c_N satisfy one of the conditions (1)–(5) below, and let P , T_N , and w_N be the profile, the center, and the window given there. Then $w_N/T_N \rightarrow 0$ in probability, and for every compact set $K \subset \mathbb{R}$ and every $\varepsilon \in (0, 1)$,*

$$\sup_{u \in K} |D_N(T_N + w_N u) - P(u)| \rightarrow 0, \quad \frac{t_N(\varepsilon) - T_N}{w_N} \rightarrow P^{-1}(\varepsilon) \quad (1.10)$$

in probability. In particular, there is cutoff.

- (1) Infinite-temperature crossover. Let $c_N \rightarrow 0$ and $J_N = Nc_N^2 \rightarrow \infty$. Then $P = \Psi$, $w_N = 1$, and $T_N = \tau_N$, where $\tau_N = \inf\{t \geq 0 : \mathcal{I}_N(t) \leq 1\}$ if $\mathcal{I}_N(0) > 1$ and $\tau_N = 1$ otherwise;

with probability tending to one, $\mathcal{I}_N(0) > 1$ and τ_N is the only solution of $\mathcal{I}_N(t) = 1$ in $[0, \infty)$. Here $\mathcal{I}_N$ is the information (1.9) at $c = c_N$. Further, let $h_N = (c_N/4) \log J_N$ and

$$\mathcal{B}(h) = \frac{1}{2} \int_{-2}^2 e^{4h\lambda} \rho_{\text{sc}}(\lambda) d\lambda = \frac{I_1(8h)}{8h},$$

where I_1 is the modified Bessel function of the first kind, and the quotient at $h = 0$ is defined by continuous extension, with $\mathcal{B}(0) = 1/2$. If (h_N) is bounded, then $\tau_N = (\log J_N + \log \mathcal{B}(h_N))/4 + o_{\mathbb{P}}(1)$. If $h_N \rightarrow \infty$, then $\tau_N - T'_N \rightarrow 0$ in probability, where T'_N is the center in (2) evaluated at $c = c_N$, and (1.10) also holds with $T_N = T'_N$ and $w_N = (1 - c_N)^{-2}$.

- (2) Fixed high temperature. Let $c_N = c \in (0, 1)$. Then $P = \Psi$, $w_N = 1/g$ with $g = (1 - c)^2$, and

$$T_N = \frac{\log N - \frac{3}{2} \log \log N + \log A_c}{4g}, \quad A_c = \frac{\sqrt{c} (1 - c)^3}{4\sqrt{\pi} (1 + c)^2}.$$

- (3) Approach to the critical temperature from above. Let $c_N = 1 - \delta_N$, where $\delta_N \rightarrow 0$ and $M_N = N\delta_N^3 \rightarrow \infty$. Then $P = \Psi$, $w_N = \delta_N^{-2}$, and

$$T_N = \frac{\log M_N - \frac{3}{2} \log \log M_N + \log B_{c_N}}{4\delta_N^2}, \quad B_c = \frac{\sqrt{c}}{4\sqrt{\pi} (1 + c)^2}.$$

- (4) Approach to the critical temperature from below. Let $c_N = 1 + \epsilon_N$, where $\epsilon_N > 0$, $\epsilon_N \rightarrow 0$, and $N\epsilon_N^3 \rightarrow \infty$. Then $P = \mathcal{V}$, $w_N = a_2^{-1}$, and

$$T_N = \begin{cases} \frac{\log(2\theta_N N a_2)}{4a_2}, & \theta_N > 0 \text{ and } 2\theta_N N a_2 > 1, \\ 1, & \text{otherwise.} \end{cases}$$

- (5) Fixed low temperature. Let $c_N = c > 1$. Then P , w_N , and T_N are as in (4). Equivalently, with $\theta = 1 - 1/c$, (1.10) holds with $\mathcal{V}_\theta$ in place of P and with T''_N in place of T_N , where $T''_N = \log(Na_2)/(4a_2)$ if $Na_2 > 1$ and $T''_N = 1$ otherwise. Further, $a_2 t_N(\epsilon)/\log N \rightarrow 1/12$ in probability.

For finitely many fixed temperatures in (2) and (5), the conclusions hold simultaneously for the same matrices W_N .

The centers τ_N in case (1) and T_N, T''_N in cases (4) and (5) are random. In case (1), τ_N is the first time at which $\mathcal{I}_N(t) \leq 1$ on the event $\{\mathcal{I}_N(0) > 1\}$, and $\tau_N = 1$ elsewhere (see (15.7)). On $\{\mathcal{I}_N(0) > 1\}$, we have $\mathcal{I}_N(\tau_N) = 1$, and this solution is unique with probability tending to one. In cases (4) and (5), the events on which the centers are set equal to 1 have probability tending to zero, so that $T_N = \log(2\theta_N N a_2)/(4a_2)$ and $T''_N = \log(Na_2)/(4a_2)$ with probability tending to one. We prove case (1) in Section 15 (Theorems 15.2 and 15.3), case (2) in Section 13 (Theorem 13.1), and case (3) in Section 14 (Theorem 14.1). Case

(4) is proved in Section 17 (Theorems 17.3 and 17.5 and Corollary 17.6), and case (5) in Section 16 (Theorem 16.1).

When Nc_N^2 stays bounded, the uniform distribution and the Gibbs measure are not asymptotically singular, and the distance relaxes over a bounded time.

Theorem 1.2 (Bounded-time relaxation near infinite temperature). *Suppose that $c_N \rightarrow 0$ and $\sqrt{N}c_N \rightarrow a \in [0, \infty)$. Then*

$$\sup_{t \geq 0} \left| D_N(t) - \left(2\Phi\left(\frac{ae^{-2t}}{2\sqrt{2}}\right) - 1 \right) \right| \rightarrow 0$$

in probability, and for every $\varepsilon \in (0, 1)$,

$$t_N(\varepsilon) \rightarrow \max \left\{ \frac{1}{2} \log \frac{a}{2\sqrt{2}\Phi^{-1}((1+\varepsilon)/2)}, 0 \right\}$$

in probability, where the limit is zero if $a = 0$. If $c_N = 0$, then $\nu_N = \pi_{N,0}$ and $D_N(t) = 0$ for all $t \geq 0$.

Theorem 1.2 is part of Theorem 15.2. Its limit is $2\Phi(\sqrt{I(t)}/2) - 1$ with $I(t) = a^2e^{-4t}/2$, and $I(t)$ is the limit of the information $\mathcal{I}_N(t)$ in (1.9), because the initial information satisfies $\mathcal{I}_N(0) = (Nc_N^2/2)(1 + o_{\mathbb{P}}(1))$ and decays at rate $4 + O(c_N)$ (Section 15). There is no cutoff in this regime.

Rescaled inverse temperature	Center T_N (leading order)	Window w_N	Profile
$c_N \rightarrow 0, Nc_N^2 \rightarrow a^2 < \infty$	bounded relaxation time, no cutoff		Theorem 1.2
$c_N \rightarrow 0, Nc_N^2 \rightarrow \infty$	$\log(Nc_N^2)/4$	1	Ψ
$c \in (0, 1)$ fixed	$\log N / (4(1-c)^2)$	$(1-c)^{-2}$	Ψ
$c_N = 1 - \delta_N, N\delta_N^3 \rightarrow \infty$	$\log(N\delta_N^3) / (4\delta_N^2)$	δ_N^{-2}	Ψ
$c_N = 1 + \sigma N^{-1/3}$	order $N^{2/3}$, no cutoff	—	F_σ (random)
$c_N = 1 + \epsilon_N, N\epsilon_N^3 \rightarrow \infty$	$\log(N\epsilon_N^3) / (12a_2)$	a_2^{-1}	$\mathcal{V}$
$c > 1$ fixed	$\log N / (12a_2)$	a_2^{-1}	$\mathcal{V}$

Table 1.1. The regimes of Theorems 1.1 to 1.3, from high to low temperature. The centers are given to leading order; the exact centers are in Theorem 1.1. At low temperature $a_2 = c(\lambda_1 - \lambda_2)/2$ is of order $N^{-2/3}$, and the center and the window are random at leading order, since $N^{2/3}a_2$ has a nondegenerate limit in distribution.

The critical window

In the critical window $c_N = 1 + \sigma N^{-1/3}$ we express our results through the Airy point process. Let $\chi_1 < \chi_2 < \dots$ be the reflected GOE Airy point process, given by the eigenvalues of the stochastic Airy operator $-d^2/dx^2 + x + 2b'(x)$ on $L^2(0, \infty)$ with Dirichlet boundary condition at 0, where b' is white noise [121, Section 2 and Theorem 1.1]. This is one random sequence, and for every fixed k we have

$$(N^{2/3}(2 - \lambda_i))_{i=1}^k \implies (\chi_i)_{i=1}^k \quad (1.11)$$

in distribution [121, Theorem 1.1]. The points χ_i are almost surely distinct and tend to infinity [121, Proposition 3.5]. We write TW_1 for the law of $-\chi_1$. By (1.11) with $k = 1$, $N^{2/3}(\lambda_1 - 2)$ converges in distribution to TW_1 , the GOE Tracy–Widom distribution [145, Section I and (53) in Section VII]. We set $\Delta_i = \chi_i - \chi_1$ for $i \geq 2$. Landon and Sosoie proved that the limit

$$\Xi = \lim_{k \rightarrow \infty} \left(\sum_{i=2}^k \frac{1}{\Delta_i} - \frac{2}{\pi} \left(\frac{3\pi k}{2} \right)^{1/3} \right)$$

exists almost surely and that $N^{1/3}(N^{-1} \sum_{i \geq 2} (\lambda_1 - \lambda_i)^{-1} - 1)$ converges in distribution to Ξ [102, Theorem 6.1]. This is the regularized inverse-gap statistic of Lemma 4.5. Du and Huang showed that the function

$$\mathcal{S}(x) = \Xi + \frac{1}{x} + \sum_{i \geq 2} \left(\frac{1}{x + \Delta_i} - \frac{1}{\Delta_i} \right), \quad x > 0, \quad (1.12)$$

is almost surely strictly decreasing, with limits $+\infty$ at 0 and $-\infty$ at $+\infty$ [62, Proposition 1.4]. For every $\sigma \in \mathbb{R}$ we let $q_\sigma > 0$ be the unique solution of $\mathcal{S}(q_\sigma) = \sigma$, and we set

$$R_1^\sigma = \frac{q_\sigma}{2}, \quad R_i^\sigma = \frac{\Delta_i + q_\sigma}{2} \quad (i \geq 2).$$

These numbers are the limiting scaled half-precisions. If $c_N = 1 + \sigma N^{-1/3}$, then the equation (1.2) for the Lagrange multiplier becomes $\mathcal{S}(q_\sigma) = \sigma$ in the limit, and $N^{2/3}p_i/2$ converges in distribution to R_i^σ , jointly for finitely many i (Lemma 4.17).

We let ρ_σ be the law of a sequence $(\xi_i)_{i \geq 1}$ of independent centered Gaussian variables with variances $(2R_i^\sigma)^{-1}$, conditioned on

$$\sum_{i \geq 1} \left(\xi_i^2 - \frac{1}{2R_i^\sigma} \right) = 0.$$

The series converges almost surely because $\sum_i (R_i^\sigma)^{-2} < \infty$, and since the event that it vanishes has probability zero, we define ρ_σ through densities (Sections 5 and 18). Note that ρ_σ describes Gibbs measures in the critical window, in the sense that the first k eigencoordinates under π_{N, c_N} , multiplied by $N^{-1/3}$, converge to the corresponding

marginal of ρ_σ (see Corollary 18.12 for a precise statement). Du and Huang introduced this conditioned Gaussian sequence and the root ϱ_0 at the critical temperature $\sigma = 0$, in their study of the overlap distribution [62, Sections 1.2 and 2.2].

We define the critical mixing profile F_σ in Section 19 through the limit of the dynamics constructed in Section 18, as follows. Let $c_N = 1 + \sigma N^{-1/3}$, and let $X(t)$ be the diffusion started from ν_N . We show that $N^{-1/3} X_i(N^{2/3}s)$ converges in distribution to $U_i(s) / \sqrt{Z(s)}$, jointly for finitely many i , where the processes U_i solve

$$dU_i(s) = -2R_i^\sigma U_i(s) ds + \sqrt{2Z(s)} dB_i(s), \quad U_i(0) = 0, \quad (1.13)$$

for independent Brownian motions $B_1, B_2, \dots$ and Z is a positive process that solves a stochastic Volterra equation driven by the same Brownian motions. We show that Z is the limit of the rescaled squared norm of the unnormalized process of Section 1.6. It tends to infinity as $s \downarrow 0$, and this singular behavior at time zero is how the uniform initial distribution enters the limit. Conditionally on the Airy point process, we define

$$F_\sigma(s) = \left\| \text{Law}\left((U_i(s) / \sqrt{Z(s)})_{i \geq 1}\right) - \rho_\sigma \right\|_{\text{TV}}, \quad s > 0. \quad (1.14)$$

Note that F_σ is a random function that is determined by the Airy point process.

Theorem 1.3 (The critical window). *Fix $\sigma_1, \dots, \sigma_m \in \mathbb{R}$ and let $c_{N,j} = 1 + \sigma_j N^{-1/3}$. Let $D_N^{(j)}$ and $t_N^{(j)}$ be the distance and the mixing time at $c = c_{N,j}$, for the same matrix W_N . Then*

$$\left(s \mapsto D_N^{(j)}(N^{2/3}s), \varepsilon \mapsto N^{-2/3} t_N^{(j)}(\varepsilon) \right)_{j=1}^m \Longrightarrow \left(F_{\sigma_j}, F_{\sigma_j}^{-1} \right)_{j=1}^m$$

in distribution, where all limits are defined from the same Airy point process. The topologies are those of uniform convergence on compact subsets of $(0, \infty)$ and of $(0, 1)$, respectively. Almost surely, for each j , the function F_{σ_j} is continuous and strictly decreasing on $(0, \infty)$, $F_{\sigma_j}(0+) = 1$, and $F_{\sigma_j}(s) \rightarrow 0$ as $s \rightarrow \infty$. Moreover, there exist $c_1, C_1 > 0$, depending on the Airy point process and on σ_j , such that $F_{\sigma_j}(s) \geq c_1 e^{-C_1 s}$ for all $s > 0$. For every $\varepsilon \in (0, 1/2)$ and every j , with the finite- N quotient defined to be 1 on the event $\{t_N^{(j)}(1 - \varepsilon) = 0\}$,

$$\frac{t_N^{(j)}(\varepsilon)}{t_N^{(j)}(1 - \varepsilon)} \Longrightarrow \frac{F_{\sigma_j}^{-1}(\varepsilon)}{F_{\sigma_j}^{-1}(1 - \varepsilon)},$$

and the limit is almost surely greater than one. Finally, for every $K > 0$ and $\eta > 0$ there exist $C > 0$ and events $\mathcal{A}_N$ with $\liminf_{N \rightarrow \infty} \mathbb{P}(\mathcal{A}_N) \geq 1 - \eta$ with the following property. On $\mathcal{A}_N$, for every $\sigma \in [-K, K]$ and every $\varepsilon \in (0, 1/2)$, the mixing time $t_N(\varepsilon)$ at $c = 1 + \sigma N^{-1/3}$ satisfies

$$C^{-1} N^{2/3} \log(1/\varepsilon) \leq t_N(\varepsilon) \leq C N^{2/3} \log(1/\varepsilon).$$

In particular, Theorem 1.3 implies that there is no cutoff in the critical window, since $t_N(\varepsilon)/t_N(1 - \varepsilon)$ converges in distribution to a limit that is almost surely greater than one. The convergence is stated in distribution: the theorem is formulated through the GOE edge limit without specifying a joint coupling across matrix sizes with the limiting Airy point process. We prove Theorem 1.3 in Sections 18 and 19. We prove convergence for deterministic spectra satisfying the hypotheses of Section 4 in Theorem 19.2 and deduce the statement for the GOE in Theorem 19.4. We prove the lower bound on F_σ in Corollary 19.5 and the bounds on $t_N(\varepsilon)$ in Corollary 19.11. These bounds allow accuracies that depend on N , and for instance they show that the mixing time at accuracy e^{-bN} is of order $N^{5/3}$.

Our last main result describes critical mixing profiles far from the center of the window, where they converge to the profiles of the mesoscopic regimes on the same side of the critical temperature.

Theorem 1.4 (Matching with the mesoscopic regimes). *Almost surely, for every compact set $K \subset \mathbb{R}$, as $S \rightarrow \infty$,*

$$\sup_{u \in K} \left| F_{-S} \left(\tau_S^h + \frac{u}{S^2} \right) - \Psi(u) \right| \rightarrow 0, \quad \tau_S^h = \frac{\log S^3 - \frac{3}{2} \log \log S^3 - \log(16\sqrt{\pi})}{4S^2},$$

and

$$\sup_{u \in K} \left| F_S \left(\tau_S^c + \frac{2u}{\Delta_2} \right) - \mathcal{V}(u) \right| \rightarrow 0, \quad \tau_S^c = \frac{\log((S - \Xi)\Delta_2)}{2\Delta_2}.$$

Further, $S^2(F_{-S}^{-1}(\varepsilon) - \tau_S^h) \rightarrow \Psi^{-1}(\varepsilon)$ and $(\Delta_2/2)(F_S^{-1}(\varepsilon) - \tau_S^c) \rightarrow \mathcal{V}^{-1}(\varepsilon)$, locally uniformly in $\varepsilon \in (0, 1)$.

We prove Theorem 1.4 in Section 20 (Theorem 20.4) and Section 21 (Theorem 21.2). The proof approximates each critical profile by finite-dimensional diffusions, taking $N \rightarrow \infty$ at fixed $\sigma = \mp S$ before letting $S \rightarrow \infty$.

1.4 Discussion

Three mechanisms. The three profiles Ψ , $\mathcal{V}$, and F_σ come from three different mechanisms.

At high temperature the law of the diffusion at time t is close in total variation to the Bingham comparison law, a Gaussian law with variances $v_i(t)$ conditioned on S_N , while $\pi_{N,c}$ is a Gaussian law with variances q_i conditioned on S_N . The log-likelihood ratio of the Bingham comparison law with respect to $\pi_{N,c}$ is $\sum_i e_i x_i^2 / (2v_i)$ up to an additive constant, and when the relative errors e_i are small and many coordinates contribute, it is approximately Gaussian under $\pi_{N,c}$ with variance $\mathcal{I}_N$ and mean $-\mathcal{I}_N/2$. This is Le Cam's Gaussian experiment [104], [146, Chapters 6–7], in which the total-variation

distance is $2\Phi(\sqrt{\mathcal{I}_N}/2) - 1$. The information decays over time, and for large t its decay is governed by the edge of the spectrum. Indeed, each relative error $e_i(t)$ decays roughly like $e^{-2p_i t}$, precisions near the edge satisfy $p_i - d = c(\lambda_1 - \lambda_i)$, and the semicircle density vanishes like a square root at the edge. A Laplace-type computation then suggests that

$$\mathcal{I}_N(t) \approx A_c N (4gt)^{-3/2} e^{-4gt}$$

at fixed $c \in (0, 1)$, with A_c and g as in Theorem 1.1. The center T_N in case (2) agrees, up to $o(g^{-1})$, with the time at which the right side equals one, so that $\mathcal{I}_N(T_N + u/g) \rightarrow e^{-4u}$ and the profile is $2\Phi(e^{-2u}/2) - 1 = \Psi(u)$. In particular, the correction $-(3/2) \log \log N$ comes from the factor $t^{-3/2}$, which reflects the square-root vanishing of the semicircle density. Factors $t^{-3/2}$ of this kind, which come from Laplace transforms of the spectral measure near the edge, appear in the physics literature on this model [51, Appendix A], [156, (14), (17)–(18)], [73, (27)–(28)].

At low temperature the mechanism is different. We describe it through the unnormalized process of Section 1.6, whose coordinates solve linear equations in which the first coordinate has no drift and coordinate i has drift $-2a_i Y_i$. During an initial period the first two coordinates acquire values that, after a common normalization, are asymptotically independent standard Gaussian variables H_1 and H_2 ; these values are built from the initial coordinates and the early noise. Afterwards the first eigencoordinate of the diffusion grows until it carries a fraction θ_N of the squared norm, while the ratio of the second eigencoordinate to the first decays at rate $2a_2$. At time $T_N + u/a_2$ the second eigencoordinate, divided by its equilibrium standard deviation $(2a_2)^{-1/2}$, is approximately $G + e^{-2u} H_2 / |H_1|$, where G is a fresh standard Gaussian variable independent of (H_1, H_2) . Since $H_2 / |H_1|$ has a standard Cauchy law, this leads to the profile $\mathcal{V}$. The center T_N solves $\sqrt{2\theta_N N a_2} e^{-2a_2 T_N} = 1$, where $\sqrt{\theta_N N}$ is the size of the first eigencoordinate, and all other eigencoordinates relax faster and do not contribute. The appearance of the Cauchy law depends on the initial distribution, since we need H_1 and H_2 to be asymptotically independent centered Gaussian variables with the same variance. A ratio of two such variables is a classical construction of the Cauchy law [110], and what is specific to this model is that the ratio survives in the total-variation limit, independently of late Gaussian noise.

In the critical window the relevant precisions are of order $N^{-2/3}$ and the time scale is $N^{2/3}$. On this scale eigencoordinates are of order $N^{1/3}$ and initial values of individual coordinates are negligible, while the limit (1.13) retains infinitely many random rates R_i^σ . The profile F_σ is determined by these random rates.

Matching across regimes. The constants in Theorem 1.1 are consistent with each other. Since $A_c = (1 - c)^3 B_c$, the center in case (2) equals

$$\frac{\log(N(1 - c)^3) - \frac{3}{2} \log \log N + \log B_c}{4(1 - c)^2}.$$

This has the form of the center in case (3) with $\delta = 1 - c$, except that $\log \log N$ replaces $\log \log M_N$. If we replace c by $c_N = 1 - \delta_N$ then the difference between the two centers is $-(3/8)\delta_N^{-2} \log(\log N / \log M_N)$, which is of the order of the window δ_N^{-2} or larger unless $\log M_N / \log N \rightarrow 1$. For instance, it equals $(3/8)\delta_N^{-2} \log(1 - 3\alpha)$ for $\delta_N = N^{-\alpha}$ with $0 < \alpha < 1/3$. Thus the formula for fixed temperature does not extend to temperatures approaching the critical one, and the logarithmic correction must be computed with M_N . Case (1) connects to case (2) in a similar way. When $h_N \rightarrow \infty$ the center τ_N is given by the formula of case (2) with $c = c_N$, and when h_N is bounded the function $\mathcal{B}$ describes the transition from relaxation governed by the bulk of the spectrum to relaxation governed by its edge. Below the critical temperature, cases (4) and (5) have the same formula.

Theorem 1.4 matches the critical profiles with cases (3) and (4). If we set $\delta_N = SN^{-1/3}$ in case (3), then $M_N = S^3$ and $B_{c_N} \rightarrow 1/(16\sqrt{\pi})$, so that the center and the window of case (3), divided by $N^{2/3}$, converge to τ_S^h and S^{-2} . If we set $\epsilon_N = SN^{-1/3}$ formally in case (4), then under a coupling realizing the convergence of the edge spectral data we have $N^{1/3}\theta_N \rightarrow S - \Xi$ and $N^{2/3}a_2 \rightarrow \Delta_2/2$. On the event $\{(S - \Xi)\Delta_2 > 1\}$, the center and the window divided by $N^{2/3}$ then converge to τ_S^c and $2/\Delta_2$. This event holds almost surely for all sufficiently large S . These substitutions are only formal, since $N\delta_N^3$ and $N\epsilon_N^3$ stay bounded. In case (4) itself we have $\epsilon_N N^{1/3} \rightarrow \infty$ and $\theta_N/\epsilon_N \rightarrow 1$, and the statistic Ξ appears only in the iterated limit, through τ_S^c . The mesoscopic theorems do not imply Theorem 1.4, which concerns the order of limits in which $N \rightarrow \infty$ first, and its proof uses estimates that are uniform in the effective parameters $N\delta_N^3$ and $\theta_N Na_2$ before the limit is taken.

Time scales and the width of the critical window. The mixing time increases from order $\log N$ at fixed high temperature to order $\delta_N^{-2} \log(N\delta_N^3)$ in the approach from above and to order $N^{2/3}$ in the critical window, and it is of order $N^{2/3} \log(N\epsilon_N^3)$ in the approach from below and of order $N^{2/3} \log N$ at fixed low temperature. The window grows from order one to order $N^{2/3}$ and remains of order $N^{2/3}$ below the critical temperature, where the center and the window are random at leading order. For instance, in case (5), (1.11) and $a_2 t_N(\epsilon) / \log N \rightarrow 1/12$ imply that $t_N(\epsilon) / (N^{2/3} \log N)$ converges in distribution to $1/(6c\Delta_2)$.

The critical window has width $N^{-1/3}$ in c for the following reason. At high temperature the smallest precision is $d \approx (1 - c)^2$, and the number of precisions in $[d, 2d]$ is of order $N(1 - c)^3$, since this is the number of eigenvalues within distance $(1 - c)^2$ of the edge. When $M_N = N\delta_N^3 \rightarrow \infty$, many coordinates contribute comparably to $\mathcal{I}_N$ and their sum produces the Gaussian profile. When $N\delta_N^3$ is bounded, the relevant precisions are comparable to the spacing $N^{-2/3}$ of the top eigenvalues, and the limit retains their randomness. Below the critical temperature the corresponding parameter is $\theta_N Na_2$, which is of order $(N\epsilon_N^3)^{1/3}$. The fluctuations of the free energy cross over on the wider scale $c = 1 + bN^{-1/3} \sqrt{\log N}$, as predicted by Baik and Lee (see [91, Section 1.2]).

Landon proved the critical Gaussian limit, tightness in the crossover, and the Gaussian and Tracy–Widom outer limits [101, Theorems 1.1–1.2]. Johnstone, Klochov, Onatski, and Pavlyshyn determined the full crossover law [91, Theorem 1.1, (1.7)]. Under their centering and normalization its GOE form is $Z + \sqrt{3/2} b_+ \text{TW}_1$, where $b_+ = \max\{b, 0\}$ and Z is an independent standard Gaussian variable, so that the Gaussian fluctuations from the bulk of the spectrum and the Tracy–Widom fluctuations from the edge are of the same order there. The dynamical window is determined by the edge alone, and the logarithm does not appear. A temperature scale $N^{-1/3}$ also appears in the static analysis of Wang and Tang, which is based on a maximum-likelihood approximation of the spectrum [150, Section 4.1.1].

Symmetry of the initial condition. The generator (1.3), the Gibbs measure, and ν_N are all invariant under each reflection $x_i \mapsto -x_i$ of an eigencoordinate, and therefore so is $\nu_N P_t$. This symmetry is important at low temperature. There the potential $\sum_{i \geq 2} a_i x_i^2$ vanishes at $\pm\sqrt{N}$ times the top eigenvector and is at least Na_2 on the equator $\{x_1 = 0\}$, where Na_2 is of order $N^{1/3}$. This is the antipodal barrier mechanism of Rodgers and Moore [124, Section 2, (25)–(26)], discussed further in [73, Section 4]. The sign of x_1 changes only when the process crosses the equator. Since $\nu_N P_t$ is symmetric under $x_1 \mapsto -x_1$, that sign does not affect $D_N(t)$, and the distance is the same as for the process folded onto the hemisphere $\{x_1 \geq 0\}$. From a starting point in one hemisphere the law at time t need not be symmetric, and Theorem 1.1 concerns only the uniform initial law. Gheissari, Grant, and Yu compare symmetric and worst-case initial conditions for spiked matrix models [78, Theorem 1.2], as we discuss in Section 1.5. For the Curie–Weiss model the censored dynamics of [61] removes the analogous bottleneck by changing the dynamics.

Which temperature sequences are covered. Theorems 1.1 to 1.3 cover fixed values $c \in [0, 1) \cup (1, \infty)$, every sequence $c_N \rightarrow 0$ along which Nc_N^2 converges in $[0, \infty]$, every sequence $c_N \rightarrow 1$ that stays on one side of 1 and satisfies $N^{1/3}|c_N - 1| \rightarrow \infty$, and $c_N = 1 + \sigma N^{-1/3}$ with σ fixed. Several cases are not covered. First, for fixed $c \notin \{0, 1\}$ we state the results at the fixed temperature c and not along sequences $c_N \rightarrow c$. The proofs suggest that the results extend to such sequences with centers evaluated at c_N , but a center evaluated at the limit c instead is shifted by an amount that does not vanish unless $(c_N - c) \log N \rightarrow 0$. Similarly, in the critical window we do not treat $c_N = 1 + \sigma_N N^{-1/3}$ with $\sigma_N \rightarrow \sigma$. Second, we do not cover the case $c_N \rightarrow \infty$, in which the temperature tends to zero, since our proofs at low temperature use a bounded range of the scaled eigenvalue gaps a_i and a limiting top fraction in $(0, 1)$ or tending to zero. Third, a sequence that oscillates between regimes is analyzed along subsequences. It need not have a single limiting profile, although subsequences with the same limiting shape can be combined using their respective centers and windows.

1.5 Related work

Equilibrium. The equilibrium theory of the model is well developed. Baik and Lee proved that the free energy has Gaussian fluctuations of order N^{-1} at high temperature and Tracy–Widom fluctuations of order $N^{-2/3}$ at low temperature [6, Theorems 1.1–1.2]; see [7–9] for related models with a ferromagnetic interaction or a bipartite structure. At the critical temperature, Landon proved a central limit theorem for the free energy on the scale $N^{-1}\sqrt{\log N}$ [101, Theorem 1.1], and Johnstone, Klochov, Onatski, and Pavlyshyn identified its limit law throughout the transition window [91, Theorem 1.1, (1.7)]. Nguyen and Sosoe proved a central limit theorem for the overlap of two replicas near the critical temperature on the high-temperature side [115, Theorem 1 and Section 5]. Landon and Sosoe identified the fluctuations of the moments of the overlap at low temperature and in the process constructed the statistic Ξ [102, Theorems 2.1–2.3 and 6.1]. Du and Huang determined the limiting overlap distribution at the critical temperature [62, Theorem 1.1], and our description of the Gibbs measure in the critical window follows theirs. Results with an external field are in [10, 45, 103], variational descriptions of spherical models of Thouless–Anderson–Palmer type are in [22, 142]. Chen extended the spherical free-energy formula to general mixtures, including odd interactions [43], while Panchenko and Talagrand studied multiple spherical systems with constrained overlaps [118, Theorem 1 and Section 3]. The relation between exact and mean spherical constraints was studied by Kac and Thompson [92], Crisanti, Sarracino, and Zannetti [48], and Aru and Korzhenkova [5], and Lukkarinen studied condensation in spherical models [109]. Genovese and Tantari studied the Legendre duality between spherical and Gaussian spin-glass ensembles [75]. These works concern the Gibbs measure, which in our setting is the end point of the dynamics.

Spherical spin glass dynamics and sampling. Gheissari and Jagannath proved that the spectral gap of the Langevin dynamics of spherical p -spin glasses with $p \geq 3$ is of order one at sufficiently high temperature and exponentially small at sufficiently low temperature [77, Theorem 1]; the quadratic case is excluded [77, Remark 5]. Ben Arous and Jagannath related spectral gaps to the free-energy landscape [24] and distinguished shattering from metastability [25, Theorems 2.5–2.6], and Ben Arous, Gheissari, and Jagannath bounded the evolution of observables along spherical dynamics [27]. Huang, Mohanty, Rajaraman, and Wu proved that simulated annealing based on Langevin dynamics samples from spherical spin glasses at high temperature in polynomial time, with small error in total variation [89, Theorem 7.2]. In [89, Section 1.5] they discuss the different question of sampling with Langevin dynamics at a fixed temperature from a uniform start. Huang, Montanari, and Pham obtained sampling in total variation for spherical spin glasses by algorithmic stochastic localization [88], an algorithm that differs from Langevin dynamics; see also [65]. For models with discrete spins, functional

inequalities and rapid mixing at high temperature were established in [3, 20, 66]. Ge, Lee, Lu, and Risteski gave an exact rejection sampler for Bingham distributions with expected runtime polynomial in the dimension and the spectral range, assuming an oracle for solving a univariate polynomial equation [74, Theorem 1.1]; it does not concern the Langevin dynamics. Other dynamics on the sphere have been analyzed as well; for instance, Kac's walk mixes in order $n \log n$ steps [120] and has cutoff in total variation from a coordinate vector [90, Theorem 1.1].

The closest work is that of Gheissari, Grant, and Yu [78], who study the Langevin dynamics on the sphere for spiked matrix models, in which a rank-one perturbation is added to a GOE matrix. For a sufficiently large fixed spike, they prove that at low temperature the worst-case mixing time is exponential, while the dynamics mixes in logarithmic time from initial distributions that are invariant under the reflection of the top eigencoordinate, such as the uniform distribution [78, Theorems 1.2–1.3]. They also prove cutoff in a high-temperature regime [78, Corollary 6.3]. Their results do not cover the model without a spike, in which the gap between the two largest eigenvalues is of order $N^{-2/3}$ instead of order one. They also do not identify a profile at low temperature, and the question of cutoff from symmetric initializations at low temperature is left open there [78, discussion following Theorem 1.2, p. 4]. Compared to their results, Theorem 1.1(4)–(5) concern the model without a spike and identify the profile, and Theorems 1.3 and 1.4 have no counterpart in [78].

The physics works discussed in Section 1.1 compute observables in the thermodynamic limit or at zero temperature, and several of them use a deterministic initial amplitude in each eigenmode. Our results concern the total-variation distance of the full law at positive temperature, from the uniform distribution and for fixed disorder. They are consistent with the physical picture, in which the relaxation rate $(1 - c)^2$ at high temperature corresponds to the classical saddle point, and in which the role of the edge of the spectrum and of the gap $\lambda_1 - \lambda_2$ at low temperature was identified in [13, 73]. We also note that the isolated (Hamiltonian) dynamics of the same model has additional conserved quantities that prevent relaxation to the Gibbs measure [54].

Tools. We use several classical tools and credit them where they are used. These include the representation of uniform and Bingham laws by conditioned Gaussian vectors [31, 98], with the Gaussian approximation of spherical coordinates going back to Borel [57, Section 6]. They also include Le Cam's theory of Gaussian likelihood ratios [104, 146], Stein's method for normal approximation [141], the scalar equation of Cugliandolo and Dean for the Lagrange multiplier [51], the Bakry–Émery calculus and logarithmic Sobolev inequalities [11, 12, 79], and stochastic calculus [123].

1.6 Ideas of the proofs

We describe the main ideas briefly here and refer to Section 2 for a detailed overview. The proofs use one exact representation of the dynamics throughout. Let $g = (g_1, \dots, g_N)$ be a standard Gaussian vector, independent of independent Brownian motions $B_1, \dots, B_N$, and let Y solve

$$dY_i = -p_i Y_i dt + \sqrt{2/N} |Y| dB_i, \quad Y(0) = g. \quad (1.15)$$

Then $\sqrt{N}Y/|Y|$ is the Langevin diffusion with generator (1.3) started from ν_N (Section 10). Shifting the precisions p_i by a common constant, for instance replacing them by $2a_i$, does not change $\sqrt{N}Y/|Y|$. The process Y is the unnormalized process of Section 10, with quadratic potential coefficients $p_i/2$. Its coordinates solve linear equations that are coupled only through the factor $|Y|$ in the noise.

At high temperature we compare $\nu_N P_t$ with the Bingham comparison law. We bound the total-variation distance between them by integrating the error that the comparison law makes in the forward equation of the diffusion (Section 9), which requires functional inequalities on the sphere (Section 8) and moment bounds for the unnormalized process (Section 10). The distance between the comparison law and $\pi_{N,c}$ is a distance between two Gaussian laws conditioned on the same sphere. We show that it equals $2\Phi(\sqrt{\mathcal{I}_N}/2) - 1$ up to an error that we control quantitatively (Sections 5 and 6). We compute the information from the covariance evolution and the edge of the spectrum (Section 7). At low temperature we follow the mechanism described in Section 1.4. We analyze the first coordinates of the unnormalized process directly and control the remaining coordinates through conditional relative entropy and logarithmic Sobolev inequalities, and smoothing estimates then upgrade weak convergence of the first coordinates to convergence in total variation (Sections 11 and 12). In the critical window we construct the limit process (1.13) and the limiting Gibbs law (Theorem 18.9 and Corollary 18.12). We also prove that at every time of order $N^{2/3}$ the law of the diffusion has a density with respect to the Gibbs measure that is bounded in L^2 uniformly in N (Theorem 12.13). With the smoothing of the finite-dimensional semigroup and a Poincaré inequality for the conditional law of the remaining coordinates, this bound controls the remaining coordinates at every fixed time (Lemma 19.15). Integration by parts under the Gibbs measure gives a uniform translation bound for the first-coordinate laws. Mollification upgrades their weak convergence to convergence in total variation at each fixed positive time (Lemma 19.17). Combining this with the bound on the remaining coordinates gives pointwise convergence of the whole total-variation distance (Proposition 19.19). Semigroup smoothing and weak convergence give continuity of the limiting profile; monotonicity then upgrades pointwise convergence to locally uniform convergence, and finite-dimensional spectral correlation bounds give strict decrease (Theorem 19.2). We prove the outer limits in Theorem 1.4 by applying

finite-dimensional estimates that are uniform in the effective parameters $N\delta_N^3$ and $\theta_N Na_2$ to finite approximations of the Airy point process (Sections 20 and 21).

Two principles apply throughout the paper. First, the object of study is always the law of the original diffusion on S_N started from ν_N . We use Gaussian and Bingham laws through explicit total-variation comparisons with the finite-dimensional diffusion. In the critical window, the countable limiting laws define the limiting distance, whose convergence follows from finite-coordinate convergence in total variation, estimates for the remaining coordinates, and continuity of the limiting profile. Second, random-matrix inputs enter only through deterministic hypotheses on the spectrum, which hold with probability at least $1 - \eta$ for every fixed $\eta > 0$ with constants depending on η . The dynamical estimates are deterministic statements about a matrix that satisfies these hypotheses. We state the published random-matrix results that we use, in our normalization, in Section 4.

1.7 Organization of the paper

For publication, the manuscript is divided into two volumes. The first, *Part I: Foundations and tools*, contains Sections 1–12: the main statements, an overview of the proofs, the preliminaries, and the general tools. The second, *Part II: Mixing profiles and critical dynamics*, contains Sections 13–21: the proofs in the temperature regimes, the construction and analysis of the critical dynamics, and the outer limits of the critical profiles.

Section 2 gives an overview of the proofs. Part I contains the preliminaries: in Section 3 we define the model, fix conventions, and collect the properties of the profiles, and in Section 4 we state the random-matrix inputs and analyze the Lagrange multiplier. Part II develops the tools, each in the generality needed later, namely Gaussian laws conditioned on a sphere (Section 5), their likelihood ratios (Section 6), the covariance evolution and its heat estimates (Section 7), functional inequalities on the sphere (Section 8), the comparison with Bingham laws (Section 9), the unnormalized process (Section 10), score bounds and conditional relative entropy (Section 11), and density bounds (Section 12). Part III proves Theorem 1.1(1)–(3) and Theorem 1.2 (Sections 13 to 15), and Part IV proves Theorem 1.1(4)–(5) (Sections 16 and 17). Part V constructs the critical limit process and proves Theorems 1.3 and 1.4 (Sections 18 to 21).

Overview of the proofs

In this section, we outline the proofs of Theorems 1.1 to 1.4; the section also serves as a map of the paper. For each regime, we explain where the mixing profile comes from, what the main obstacle is, and which results overcome it, with references to the precise statements in the later sections. Some computations in this overview are heuristic; precise statements and proofs appear in the sections cited.

Section 2.1 collects the reductions that are used in every regime. Section 2.2 treats the temperatures above the critical one, Section 2.3 the temperatures below it, and Section 2.4 the critical window. Section 2.5 explains how the critical mixing profiles are matched with the cutoff profiles of the mesoscopic regimes. Section 2.6 records which sections each proof uses.

2.1 Common reductions

Eigencoordinates and shifts. Since ν_N is invariant under orthogonal transformations, we may work in an orthonormal eigenbasis of W_N , and $D_N(t)$ depends on W_N only through its eigenvalues (Section 3.1). Most of our dynamical estimates concern a general real symmetric matrix A with eigenvalues $\alpha_1 \geq \dots \geq \alpha_N$, the generator L_A , and the Gibbs measure $\pi_A \propto e^{x^\top A x/2} \nu_N$; the model is the case $A = cW_N$. Because $|x|^2 = N$ on S_N , adding a multiple of the identity to A changes neither L_A nor π_A (see (3.7)). We use this freedom in two ways. At high temperature and in the critical window, we write the potential through the precisions $p_i = z - \alpha_i$ of Definition 3.3, which are positive and satisfy $\sum_i p_i^{-1} = N$, so that π_A is proportional to $e^{-\sum_i p_i x_i^2/2} \nu_N$. At low temperature, we use the scaled eigenvalue gaps $a_i = c(\lambda_1 - \lambda_i)/2$ of (3.8), for which $\pi_{N,c} \propto e^{-F} \nu_N$ with $F = \sum_{i \geq 2} a_i x_i^2$, and the first eigencoordinate carries no potential. Finally, we note that L_A , π_A , and ν_N are invariant under each reflection $x_i \mapsto -x_i$, and so is $\nu_N P_t^A$ (Lemma 3.1).

Gibbs measures as conditioned Gaussian laws. For positive numbers $v_1, \dots, v_N$ with $\sum_i v_i = N$, let μ_v be the law of a centered Gaussian vector with independent coordinates of variances v_i , conditioned on $|x|^2 = N$ (Definition 5.4). By Proposition 5.6, the Bingham law proportional to $\exp\{\frac{1}{2} \sum_i b_i x_i^2\} \nu_N$ with $b_i = z - v_i^{-1}$ is μ_v , for every $z \in \mathbb{R}$. In particular, we have $\nu_N = \mu_v$ with all $v_i = 1$, and $\pi_A = \mu_q$ with $q_i = 1/p_i$; this representation of Bingham laws is classical [31, 98]. Under μ_v , the coordinates behave like independent

Gaussian variables when no coordinate carries a macroscopic share of the variance of the radial constraint, and in Section 5 we measure this through

$$\eta_v = \frac{\max_i v_i}{(2 \sum_j v_j^2)^{1/2}}. \quad (2.1)$$

The main results of that section are a quantitative local limit theorem for the density of $\sum_i v_i (g_i^2 - 1)$, where the g_i are independent standard Gaussian variables (Theorem 5.3); an exact formula for the density of a set of coordinates with respect to independent Gaussian variables (Lemma 5.9), and a bound on this density when the set carries a small fraction of the variance (Theorem 5.11); moment bounds (Theorem 5.12); an exact formula for quadratic exponential moments (Lemma 5.13); and moment bounds for quadratic statistics (Theorem 5.15). For the model, η_q is of order $N^{-1/2}$ at a fixed temperature $c < 1$, and it is asymptotic to $M_N^{-1/2}$ along $c_N = 1 - \delta_N$ (Theorem 4.13). In the critical window and at low temperature, the first eigencoordinates carry a share of the radial variance that does not vanish, and η_q does not tend to zero. For the critical window, we also treat in Section 5 countable variance sequences, for which the radial constraint must be renormalized (Definition 5.8 and Proposition 5.17).

The unnormalized process. All our dynamical estimates use one exact representation of the diffusion. For quadratic potential coefficients $\rho_1, \dots, \rho_N \in \mathbb{R}$, let Y solve

$$dY_i = -2\rho_i Y_i dt + \sqrt{2/N} |Y| dB_i, \quad Y(0) = g, \quad (2.2)$$

where g is a standard Gaussian vector independent of the Brownian motions $B_1, \dots, B_N$. We show in Lemma 10.2 that Y never vanishes and that $X = \sqrt{N} Y / |Y|$ is the Langevin diffusion with potential $\sum_i \rho_i x_i^2$ started from ν_N , with the same time parameter. If we replace every ρ_i by $\rho_i + \tau$, then $Y(t)$ is multiplied by $e^{-2\tau t}$ and X does not change. We take $\rho_i = p_i/2$ at high temperature and in the critical window, which gives (1.15), and $\rho_i = a_i$ at low temperature, in which case Y_1 is a martingale, since $\mathbb{E}|Y(0)|^2 < \infty$ (Remark 10.3).

The coordinates of Y have linear drifts. By (10.5), the second moments $\mathbb{E}Y_i(t)^2$ solve the linear system with matrix $-4 \operatorname{diag}(\rho) + (2/N) \mathbf{1}\mathbf{1}^\top$, and for $\rho_i = p_i/2$ we find that

$$\mathbb{E}Y_i(t)^2 = h_i(t), \quad h(t) = e^{tH} \mathbf{1}, \quad H = -2 \operatorname{diag}(p) + \frac{2}{N} \mathbf{1}\mathbf{1}^\top, \quad (2.3)$$

where H is the heat matrix studied in Section 7. The only nonlinear feature of (2.2) is the random factor $|Y|$ in the noise. In Section 10, we bound $|Y(t)|^2$ from above and from below by solutions of the linear system, in moments, in inverse moments, and along whole paths. We also write $Y = \Phi X_0$ there, where Φ is a matrix-valued process independent of the initial point X_0 (Lemma 10.2), and this is how the initial distribution enters the proofs at low temperature.

Deterministic hypotheses. As explained in Section 1.6, the random-matrix results of Section 4 enter only through deterministic hypotheses on the spectrum. These hold on events of probability at least $1 - \eta$, for every fixed $\eta > 0$ and with constants depending on η (Section 3.4), and Lemma 3.8 removes the localization at the end of each proof. We list the hypotheses of each regime in Table 2.1. When a proof uses indexed bounds on the GOE edge gaps, Lemma 4.6 provides $\lambda_1 - \lambda_i \asymp ((i-1)/N)^{2/3}$ at every fixed confidence. The deterministic infinite-temperature comparison requires only the norm bound in Theorem 15.1.

Regime	Deterministic hypotheses	For the GOE
$c_N \rightarrow 0$	$A = cG$ with $\ G\ \leq M$ and $c \leq c_0(M)$ (Theorem 15.1)	Lemma 15.8; for the explicit centers, also Lemma 4.2
$c \in (0, 1)$ fixed	$\ A\ \leq \Lambda$ and $Nd^2 \geq C_w(\Lambda)$ (Theorem 9.13); eigenvalues within $o((\log N)^{-7/4})$ of the semicircle quantiles, uniformly (Lemma 13.11)	Lemma 4.2 and Theorem 4.13
$c_N = 1 - \delta_N$	$d \leq d_0$, $K^{-1}((i-1)/N)^{2/3} \leq p_i - d \leq K((i-1)/N)^{2/3}$, and $Nd^{3/2} \geq M_0$; see (14.33)	Lemma 4.6 and Theorem 4.13
$c_N = 1 + \sigma N^{-1/3}$	$\sum_i (2r_i)^{-1} = N$, $c_0(i/N)^{2/3} \leq r_i \leq C_0(i/N)^{2/3}$, and $r_2 - r_1 \geq \delta N^{-2/3}$ (Definition 4.16)	Lemma 4.17
$c_N = 1 + \epsilon_N$	$c_*(i-1)^{2/3}N^{-2/3} \leq a_i \leq C_*(i-1)^{2/3}N^{-2/3}$ and $a_3 \geq (1 + \epsilon_*)a_2$ (Definition 17.1); $\theta_N > 0$, $\theta_N \rightarrow 0$, and $m_N = \theta_N Na_2 \rightarrow \infty$ (Theorem 17.3)	Lemma 4.7
$c > 1$ fixed	$\theta_N \geq \theta_*$, $a_N \leq B$, $a_j \geq c_*(j-1)^{2/3}N^{-2/3}$, and $a_2 \leq C_*N^{-2/3}$ (Definition 16.3); $\theta_N \rightarrow \theta \in (0, 1)$ and $1 + \varpi \leq a_3/a_2 \leq D_3$ for fixed $\varpi > 0$, $D_3 < \infty$ (Theorem 16.5)	Lemmas 16.6 and 16.7

Table 2.1. The deterministic hypotheses on the spectrum in each regime, and the results that verify them for the GOE on events of probability close to one. Here $d = p_1$ is the smallest precision and $r_i = p_i/2$ are the half-precisions.

From pointwise limits to profiles. The distance D_N is nonincreasing by the contraction property of Markov semigroups (Lemma 3.6). Recall from Lemma 3.9 that the profiles Ψ , $\mathcal{V}$, and $\mathcal{V}_\theta$ are continuous and decrease strictly from 1 to 0. Suppose that $T_N/w_N \rightarrow \infty$ and that $D_N(T_N + w_N u) \rightarrow P(u)$ in probability for each fixed u , where only nonnegative times $T_N + w_N u$ are considered, as in Definition 3.7. Then Lemma 3.11 implies that the convergence is locally uniform in u , and that $(t_N(\epsilon) - T_N)/w_N \rightarrow P^{-1}(\epsilon)$ for every

$\varepsilon \in (0, 1)$. Outside the critical window, it therefore suffices to prove convergence at each fixed offset u . We also use repeatedly that the total variation distance between two laws is at least the distance between the laws of any function of the configuration, for instance of finitely many eigencoordinates; this gives lower bounds on D_N .

2.2 High temperature

We first describe, for a fixed matrix A , the comparison on which cases (1)–(3) of Theorem 1.1 and Theorem 1.2 rest. We then explain how each regime evaluates it.

The Bingham comparison law

Let A have precisions p_i , let H and $h(t)$ be as in (2.3), and let $\bar{h}(t) = N^{-1} \sum_i h_i(t)$. We define the covariance evolution of A by

$$v_i(t) = \frac{h_i(t)}{\bar{h}(t)}, \quad \text{so that} \quad \sum_{i=1}^N v_i(t) = N, \quad v_i(0) = 1, \quad \lim_{t \rightarrow \infty} v_i(t) = q_i. \quad (2.4)$$

These are the variances of the linear system (1.8), in which the spherical constraint holds only in expectation (Section 7.2), and the Bingham comparison law at time t is $\mu_t = \mu_{v(t)}$ (Definition 7.5). We have $\mu_0 = \nu_N$, and $\mu_t \rightarrow \pi_A$ as $t \rightarrow \infty$ (Lemma 7.6). The mean $\bar{h}$ solves the renewal equation $\bar{h} = k + 2k * \bar{h}$ with $k(t) = N^{-1} \sum_i e^{-2p_i t}$ (Lemma 7.1); this is the finite-dimensional counterpart of the scalar Volterra equation of Cugliandolo and Dean [51, (II.4)–(II.7)].

By the triangle inequality, we have

$$\left| D_N(t) - \|\mu_t - \pi_A\|_{\text{TV}} \right| \leq \|\nu_N P_t^A - \mu_t\|_{\text{TV}}. \quad (2.5)$$

The proof at high temperature consists of two independent estimates: the right side of (2.5) is small at the relevant times, and $\|\mu_t - \pi_A\|_{\text{TV}}$ converges to the profile. In general, μ_t need not coincide with the time- t law, and $v_i(t)$ need not equal the second moment of x_i under μ_t or under $\nu_N P_t^A$ (Remark 7.8). We compare the two laws only through the forward equation of the diffusion.

Likelihood ratios on a common sphere

Both μ_t and π_A are Gaussian laws conditioned on S_N , with variances $v_i(t)$ and q_i of equal sum. We write $v_i(t) = q_i(1 + e_i(t))$, so that $e_i(t)$ are the relative errors and $\mathcal{I}_N(t) = \mathcal{I}(q, v(t)) = \frac{1}{2} \sum_i e_i(t)^2$ is the information of Definition 3.5. Since the two Gaussian densities differ by the exponential of a quadratic form, we have on S_N

$$\ell = \log \frac{d\mu_t}{d\pi_A} = \frac{1}{2} \sum_{i=1}^N \frac{e_i(t)}{1 + e_i(t)} Z_i^2 + \text{const}, \quad Z_i = \frac{x_i}{\sqrt{q_i}}. \quad (2.6)$$

We now explain heuristically why ℓ is approximately Gaussian under π_A . When η_q is small, the variables Z_i behave under $\pi_A = \mu_q$ like independent standard Gaussian variables, subject to the constraint $\sum_i q_i(Z_i^2 - 1) = 0$. When the relative errors are small, ℓ is close to $\frac{1}{2} \sum_i e_i(Z_i^2 - 1)$ up to a constant. For independent standard Gaussian variables, this sum has variance $\frac{1}{2} \sum_i e_i^2 = \mathcal{I}_N$, and its covariance with the constraint is $\sum_i q_i e_i = \sum_i (v_i - q_i) = 0$, so that conditioning on the constraint does not change its variance to leading order. As $\int e^\ell d\pi_A = 1$, the mean of ℓ is then close to $-\mathcal{I}_N/2$. In Le Cam's Gaussian experiment [104], [146, Chapters 6–7], if ℓ has law $\mathcal{N}(-I/2, I)$ under π_A , then it has law $\mathcal{N}(I/2, I)$ under μ_t . In that case, we find that

$$\|\mu_t - \pi_A\|_{\text{TV}} = \mu_t(\ell > 0) - \pi_A(\ell > 0) = 2\Phi(\sqrt{I}/2) - 1 = \mathcal{G}(I), \quad (2.7)$$

where $\mathcal{G}$ is the function of (13.2).

We make this quantitative in Theorem 6.6. Let v and $\tilde{v}$ be variance sequences with equal sums, write $\tilde{v}_i = v_i(1 + e_i)$, let $I = \mathcal{I}(v, \tilde{v}) = \frac{1}{2} \sum_i e_i^2$ be the information of $\tilde{v}$ relative to v , and let $\ell = \log(d\mu_{\tilde{v}}/d\mu_v)$. Suppose that $\max_i |e_i| \leq 1/4$, that η_v is smaller than a universal constant, and that $I \leq B$. Then the Wasserstein distance between the law of ℓ under μ_v and $\mathcal{N}(-I/2, I)$ is at most $C_B(\eta_v + \max_i |e_i|)$, and so is the difference between $\|\mu_{\tilde{v}} - \mu_v\|_{\text{TV}}$ and $\mathcal{G}(I)$. The proof combines an exact integration by parts formula for quadratic statistics under μ_v , which accounts for the constraint (Lemma 6.3), a bound on inverse moments (Lemma 6.2), and Stein's method [141] (Lemma 6.4 and Theorem 6.5). We deduce the corresponding limit statements in Theorem 6.8; in particular, the distance tends to one when the information diverges.

We conclude that the high-temperature profile reduces to the following problem: find a center T_N and a window w_N such that, at the times $t = T_N + w_N u$, we have $\mathcal{I}_N(t) \rightarrow e^{-4u}$, $\max_i |e_i(t)| \rightarrow 0$, and $\eta_q \rightarrow 0$. Then $\|\mu_t - \pi_A\|_{\text{TV}}$ tends to $\mathcal{G}(e^{-4u}) = \Psi(u)$, which explains why the same profile Ψ appears in cases (1)–(3).

The information at a fixed temperature

The information is determined by the mean of the heat vector (Lemma 7.3):

$$\mathcal{I}_N(t) = \frac{N \bar{h}''(2t)}{8 \bar{h}(t)^2}, \quad \bar{h}(t) = \vartheta + \int e^{-\lambda t} \rho(d\lambda), \quad (2.8)$$

where $\vartheta = N / \sum_i q_i^2$, and ρ is a positive measure on $[2d, 2s]$: together with the atom ϑ at zero, it forms the spectral measure of $-H$ at the unit vector $N^{-1/2} \mathbf{1}$ (Lemma 7.1). In particular, $\mathcal{I}_N(t)$ decays at least like e^{-4dt} , and at large times its behavior is governed by the precisions near the smallest one, that is, by the eigenvalues near λ_1 .

At a fixed temperature $c \in (0, 1)$, we compute $\mathcal{I}_N(t)$ in Section 13 by comparison with the same evolution for the semicircle law, which we solve exactly by Laplace transforms (Proposition 13.4), using the same scalar-resolvent method as [51, Appendix A], whose

appendix treats physical temperatures $T < T_c$. By Proposition 13.7, its relative errors satisfy $\eta_t(\lambda) \approx -c(1+c)^{-1}e^{-2p(\lambda)t}$ for λ within distance $O(1/t)$ of the edge, where $p(\lambda) = 1 + c^2 - c\lambda = g + c(2 - \lambda)$ and $g = (1 - c)^2$. Using that $\rho_{sc}(2 - y) \approx \pi^{-1}\sqrt{y}$ for small y , we find heuristically that

$$\mathcal{I}_N(t) \approx \frac{N}{2} \int \eta_t(\lambda)^2 \rho_{sc}(\lambda) d\lambda \approx \frac{N}{2} \left(\frac{c}{1+c} \right)^2 e^{-4gt} \int_0^\infty e^{-4cty} \frac{\sqrt{y}}{\pi} dy = \frac{N}{2} K_c t^{-3/2} e^{-4gt}, \quad (2.9)$$

with $K_c = \sqrt{c}/(16\sqrt{\pi}(1+c)^2)$ as in (13.1). As $A_c = 4K_c g^{3/2}$, this is the asymptotic stated in Section 1.4, and it tends to e^{-4u} at the times $T_N + u/g$. For the continuum evolution, the last asymptotic in (2.9) holds uniformly for $c \in (0, c_1]$ as $ct \rightarrow \infty$, by Proposition 13.7. In Lemma 13.11, we compare the finite and continuum evolutions. Let E be the supremum distance on $(0, 1)$ between the step function of the eigenvalues and the quantile function of the semicircle law. If E is small and $cEt \leq 1/4$, the relative errors of the two evolutions differ in $L^2(0, 1)$ by at most $CcE(1+t)e^{-2gt}$, where C depends only on c_1 , and we bound the largest relative error by a separate argument. At the times $T_N + u/g$, the factor $\sqrt{N}e^{-2gt}$ is of order $(\log N)^{3/4}$, and the resulting error in $\sqrt{2\mathcal{I}_N}$ is $O(E(\log N)^{7/4})$. To prove the bound in $L^2(0, 1)$, we write both evolutions through operators with a gap on the orthogonal complement of their stationary vectors (Lemma 13.9), rotate one stationary vector onto the other, compare the two evolutions on the common complement by variation of constants, and then normalize (Lemma 13.10). Because of the factor e^{-2gt} , we need the rigidity estimate of Lemma 4.2 only in a weak form. It suffices that $E(\log N)^{7/4} \rightarrow 0$. For each fixed confidence $1 - \eta$, Lemma 13.8 gives $E \leq (C_\eta + 7)N^{-5/12}$ on events of probability at least $1 - \eta$ for all sufficiently large N ; we let $\eta \downarrow 0$ after the deterministic estimates. The other conditions of the reduction above hold because η_q is of order $N^{-1/2}$ and $\max_i |e_i(t)| \rightarrow 0$. Combining these estimates with Lemma 13.2, a uniform form of Theorem 6.8, we find that $\|\mu_t - \pi_{N,c}\|_{TV} \rightarrow \Psi(u)$ at $t = T_N + u/g$. Finally, Theorem 9.13, which we describe next, implies that the right side of (2.5) is at most $Ce^{-gt/4}$ on events of probability tending to one, and Theorem 13.1 follows.

Comparison through the forward equation

The density of μ_t does not solve the forward (Fokker–Planck) equation of L_A , but it solves it up to an explicit error. In Definition 9.2, we define the forward-equation defect R_t by $\partial_t \mu_t = L_A^* \mu_t + R_t \mu_t$, and in Lemma 9.3 we compute it:

$$R_t = -\frac{Q_B(t)Q_C(t) - \mu_t(Q_B(t)Q_C(t))}{N}, \quad (2.10)$$

$$Q_B(t) = x^\top B_t x - N(z(t) - 1), \quad Q_C(t) = x^\top (A - B_t)x.$$

Here $\mu_t \propto e^{x^\top B_t x/2} \nu_N$ with $B_t = \text{diag}(z(t) - v_i(t)^{-1})$. Testing against the backward evolution of a bounded function f , we have the Duhamel identity

$$\mu_t(f) - \nu_N P_t^A(f) = \int_0^t \mu_u(R_u P_{t-u}^A f) du$$

of Theorem 9.5. In that theorem, we turn the identity into bounds on $\|\nu_N P_t^A - \mu_t\|_{\text{TV}}$ of three forms. In the first, we measure the defect in $L^2(\mu_u)$. The second also requires a Poincaré inequality for μ_u and a bound on $\mu_u \Gamma(P_{t-u}^A f)$. In the third, we write the defect as a divergence, $\mu_u(R_u g) = \mu_u \langle j_u, \nabla_{S_N} g \rangle$, and only the size $\mu_u |j_u|^2$ of the vector field j_u enters.

Two facts make the defect small at late times. First, the coefficients of $Q_C(t)$ decay at rate $2d$: if $\|A\| \leq M$, then $\|A - B_t\| \leq C_M d^{-1} e^{-2dt}$ by Lemma 7.6. Second, the gradient of $P_\tau^A f$ decays. With the weight $\varphi(x) = N^{-1} \sum_i q_i x_i^2$, which is built from the equilibrium variances, we derive from the weighted Bochner identity of Lemma 8.5 that

$$\varphi \Gamma(P_t^A f) \leq e^{-2r_0 t} P_t^A(\varphi \Gamma(f)), \quad r_0 = d - \frac{\kappa + \kappa^{-1}}{N}, \quad \kappa = \frac{s}{d} \quad (2.11)$$

(Theorem 8.6). This is an instance of the Bakry-Émery calculus [11, 12]. When Nd^2 is large, the rate $r = r_0 - 1/N$ is at least $d/2$, and by the second form, if $\|A\| \leq \Lambda$, then $\|\nu_N P_t^A - \mu_t\|_{\text{TV}} \leq C_\Lambda d^{-2} e^{-rt}$ (Theorem 9.13). At a fixed temperature $c < 1$, we have $d_N \rightarrow (1-c)^2$, and this bound suffices. For a matrix within ϵ of a multiple of the identity, the distance is at most $C_1 \epsilon^2 \min(t, 1)$ and at most $C_1 \epsilon^2 e^{-rt}$, uniformly in time (Proposition 9.14); we use these bounds at infinite temperature.

The approach from above

Along $c_N = 1 - \delta_N$ with $M_N = N\delta_N^3 \rightarrow \infty$, the smallest precision satisfies $d_N/\delta_N^2 \rightarrow 1$ (Theorem 4.13). The product $Nd_N^2 \approx M_N \delta_N$ may tend to zero, and the times $T_N + u\delta_N^{-2}$ are of order $\delta_N^{-2} \log M_N$. There are two new difficulties, one for each side of (2.5).

The information at a smaller scale. Let $t = t_N > 0$ be deterministic with $\delta_N^2 t_N \rightarrow \infty$ and $\delta_N^2 t_N = o(M_N^{2/3})$, as in (4.15). At such times, the relative error $e_i(t)$ is small unless $p_i - d_N$ is at most of order $1/t$, which is much smaller than δ_N^2 . The number of such indices is of order $Nt^{-3/2}$, and on this scale the information is governed by eigenvalues within distance $O(1/t_N)$ of the edge. In Section 14, we therefore work with the measure $\rho = \rho_N$ of (2.8) directly. We rescale it at the scale $1/t$, as in (14.16), and the rescaled measure converges vaguely in probability to $\pi^{-1} \sqrt{y} dy$, the image of the semicircle density near the edge (Theorem 14.2). The difficulty is that ρ_N is not a simple function of the eigenvalue counts. However, its real resolvent is an explicit rational function of the precisions (Lemma 7.1). We find the limits of these resolvents at real arguments from the edge counts at a varying scale (Theorem 4.8 and Corollary 4.9), together with

the asymptotics of the Lagrange multiplier (Theorem 4.13) and of the inverse gaps (Lemma 4.5). In Lemma 7.15, we prove that convergence of the real squared resolvents of positive measures implies their convergence against a large class of test functions; the proof uses the compact-interval moment method (see [151, Chapter III]), a finite-difference calculation to recover the moments, and approximation by Bernstein polynomials [29, pp. 1–2]. The result is Corollary 14.3:

$$\mathcal{I}_N(t) = \frac{Nd_N^{3/2}}{128\sqrt{\pi}} w^{-3/2} e^{-4w} (1 + o_{\mathbb{P}}(1)), \quad w = d_N t. \quad (2.12)$$

This is the right side of (2.9) with $c = 1$ and g replaced by d_N . If we set the right side of (2.12) equal to e^{-4u} and solve for t , we recover the center T_N of case (3), with its correction $-\frac{3}{2} \log \log M_N$, up to an error that is small compared with the window δ_N^{-2} . To replace d_N by δ_N^2 in the center, we need that $\log(2 + M_N)|d_N/\delta_N^2 - 1| \rightarrow 0$ in probability, which follows from Theorem 4.13. As $\eta_q = O_{\mathbb{P}}(M_N^{-1/2})$, Theorem 6.6 applies, and $\|\mu_t - \pi_{N,c_N}\|_{\text{TV}}$ converges to $\Psi(u)$ at the times $T_N + u/\delta_N^2$ (Theorem 14.4).

The comparison with no condition on Nd^2 . The bound $C_{\Lambda} d^{-2} e^{-rt}$ of Theorem 9.13 requires Nd^2 to be large, and at the times $T_N + u\delta_N^{-2}$ it is about $\delta_N^{-4} M_N^{-1/4}$, which need not be small. The obstruction is the rate r_0 in (2.11), which is positive only when Nd^2 is large. To overcome this, we prove in Theorem 14.10 that for precisions that satisfy (14.33) with $M = Nd^{3/2} \geq M_0$,

$$\|\nu_N P_t^A - \mu_t\|_{\text{TV}} \leq C e^{-(d-2/N)t}, \quad 0 \leq t \leq C_T d^{-1} \log M, \quad (2.13)$$

where C and M_0 depend only on the constants in (14.33) and on C_T . At these times, the right side is $M^{-1/4}(\log M)^{O(1)}$, which tends to zero however slowly M grows. The proof has three ingredients.

The first is the third form of Theorem 9.5. We solve the Poisson equation $-L_{B_u} \psi_u = R_u$ for the generator whose Gibbs measure is μ_u and take $j_u = \nabla_{S_N} \psi_u$ (Lemma 9.6 and Corollary 9.7). In Theorem 9.16, we bound the energy $\mu_u \Gamma(\psi_u)$ by C/M times explicit sums of the coefficients of $Q_C(u)$, under conditions on the variances at a scale δ (Definition 9.15). The proof uses the lower bound on the spectral gap of Theorem 8.8, which is sharp when many precisions are comparable to the smallest one. Along the covariance evolution, we find that the energy is at most $C(1+u)^{-1}$ for $u \leq d^{-1}$ and at most $Cd(1+ds)^2 e^{-4ds}$ at $u = d^{-1} + s$ (Lemma 14.7).

The second ingredient is a bound on the gradient of $P_{\tau}^A f$ under μ_u with the rate $d - 2/N$ in place of r_0 . To obtain it, we keep the variable term of the weighted Bochner identity, which leads to the Feynman–Kac bound of Lemma 10.21:

$$\varphi(x) \Gamma(P_t^A f)(x) \leq e^{-2(d-2/N)t} \mathbb{E}_x \left[\exp \left(2 \int_0^t \Gamma(\log \varphi)(X_s) ds \right) \varphi(X_t) \Gamma(f)(X_t) \right]. \quad (2.14)$$

By Lemma 8.5, the function $\Gamma(\log \varphi)$ is at most $(\kappa + \kappa^{-1} - 2)/N$, which is not small compared with d when Nd^2 is small. However, its time integral along the path is small: by Theorem 10.15, its exponential moments are at most $1 + C\lambda\sqrt{\log M/M}$, for every initial law. Together with a reverse variance argument along the path, this implies the bound of Lemma 14.9 on $\mu_u \Gamma(P_\tau^A f)$ for $0 \leq f \leq 1$.

The third ingredient consists of the path estimates for the unnormalized process behind the second one. Under the initial law μ_u weighted by $1/\varphi$, we need moments of $\sup_t \varphi(X_t)$ along the whole time interval, and exponential moments of the occupation integral above (Lemma 14.8). We derive them from inverse moments of $|Y(t)|^2$ relative to $N\bar{h}(u+t)$, at fixed times (Lemma 10.14) and along the whole interval (Theorem 10.18), and from a comparison of the initial state with all future solutions of the linear system at once (Lemma 10.16). The inverse bounds use Lemma 10.4, which requires the conditional relative variances to be bounded uniformly by a small constant; how small depends on the order of the inverse moment or on the exponential parameter. Its occupation bound additionally uses a deterministic bound on the quadratic variation of the logarithm. Inserting the first two ingredients into the Duhamel integral, we obtain (2.13). The estimate holds for every N and every matrix with $M \geq M_0$, with the same constants. We use it again in Section 20, with M fixed while $N \rightarrow \infty$.

The infinite-temperature crossover

When $A = cG$ with $\|G\| \leq M$ and c small, the distance is asymptotically determined by the information alone, uniformly over all times (Theorem 15.1). Indeed, Proposition 9.14 implies that $\|\nu_N P_t^A - \mu_t\|_{\text{TV}} \leq Cc^2 \min(t, 1)$ for all t . The relative errors satisfy $\max_i |e_i(t)| \leq Cce^{-t}$, and if $\mathcal{I}_N(0) > 0$, then the logarithmic derivative of $\mathcal{I}_N(t)$ lies in $[-4 - Cc, -4 + Cc]$. Using Lemma 13.2, we find that $\sup_{t \geq 0} |D_N(t) - \mathcal{G}(\mathcal{I}_N(t))| \rightarrow 0$ as $c \rightarrow 0$ and $N \rightarrow \infty$. If $\mathcal{I}_N(0) \rightarrow \infty$, it follows that there is cutoff with a window of order one, at the time τ_N defined by $\mathcal{I}_N(\tau_N) = 1$.

For the GOE, we have $\mathcal{I}_N(0) = (J_N/2)(1 + o_{\mathbb{P}}(1))$ with $J_N = Nc_N^2$ (Theorem 15.2). Here $e_i(0) = p_i - 1$ is close to $-c(\lambda_i - N^{-1} \sum_j \lambda_j)$, and the centered second moment $N^{-1} \text{Tr}(W_N^2) - (N^{-1} \text{Tr} W_N)^2$ tends to 1 in probability. The proof of the classification uses only the norm of W_N and the traces of W_N and W_N^2 , and not the rigidity estimate (Lemma 15.8). If $\sqrt{N}c_N \rightarrow a$, the information is asymptotic to $a^2 e^{-4t}/2$, and $\mathcal{G}(a^2 e^{-4t}/2)$ is the limit in Theorem 1.2. If $J_N \rightarrow \infty$, there is cutoff at the first time τ_N at which $\mathcal{I}_N \leq 1$. With probability tending to one, $\mathcal{I}_N(0) > 1$ and τ_N is the only solution of $\mathcal{I}_N(t) = 1$.

In Theorem 15.3, we determine τ_N up to $o(1)$. Here we need the rigidity estimate, through the approximation of the eigenvalues by the semicircle quantiles in Lemma 13.8. On the time scale $\tau_N \approx \frac{1}{4} \log J_N$, the product $c_N \tau_N$ is close to $h_N = (c_N/4) \log J_N$. When h_N is bounded, the rescaled relative errors $c^{-1} e^{2t} e(t)$ are close to $-e^{2ct} \mathcal{T} \lambda$, where $\mathcal{T}$ is the multiplication by the eigenvalues compressed to vectors of mean zero, and λ is the

vector of eigenvalues of G , which we may take to have mean zero (Lemma 15.9). For the semicircle law, we compute this operator with Chebyshev polynomials (Lemma 15.10), which gives $\mathcal{I}_N(t) \approx J_N e^{-4t} \mathcal{B}(c_N t)$ with $\mathcal{B}(h) = I_1(8h)/(8h)$ (Lemma 15.11). Solving $\mathcal{I}_N(\tau_N) = 1$, we have $\tau_N = \frac{1}{4}[\log J_N + \log \mathcal{B}(h_N)] + o_{\mathbb{P}}(1)$. When $h_N \rightarrow \infty$, the relative errors concentrate near the edge of the spectrum. The computation of Section 13 holds uniformly for $c \in (0, c_1]$, and it yields the center of case (2) evaluated at $c = c_N$.

2.3 Low temperature

Sections 16 and 17 prove cases (5) and (4) of Theorem 1.1, at fixed $c > 1$ and along $c_N = 1 + \epsilon_N$. We use the scaled eigenvalue gaps a_i and the top fraction θ_N of Section 3.2, with $R_N = Na_2$ and $m_N = \theta_N R_N$. At fixed $c > 1$, a_2 is of order $N^{-2/3}$ and R_N of order $N^{1/3}$. Along $c_N = 1 + \epsilon_N$, we have $\theta_N/\epsilon_N \rightarrow 1$, and m_N is of order $(N\epsilon_N^3)^{1/3}$ (Lemmas 4.7 and 17.2). Throughout, the dynamics is represented by the unnormalized process (2.2) with $\rho_i = a_i$, whose first coordinate Y_1 is a martingale while Y_i is pulled toward zero at rate $2a_i$ for $i \geq 2$. At a fixed temperature, we take as its initial value a point X_0 with law ν_N , which also starts the diffusion from ν_N (Lemma 10.2).

The mechanism

We describe the mechanism at a fixed temperature, with $\theta_N \rightarrow \theta \in (0, 1)$, following the outline in Section 16.1. It has three steps.

Gaussian approximation at time t_p . At the time $t_p = (\log N)/100$, the pair $(Y_1(t_p), Y_2(t_p))$ is close in total variation to $\theta^{-1/2}(H_1, H_2)$, where H_1, H_2 are independent standard Gaussian variables (Proposition 16.18). These variables are built from the initial point and from the noise up to time t_p . To see why, we write $Y(t) = \Phi(t)X_0$ as in Lemma 10.2, with Φ independent of X_0 . Conditionally on Φ , finitely many coordinates of $Y(t_p)$ are linear functions of a uniform point of S_N . By the rotation invariance of ν_N , they are close to a centered Gaussian vector whose covariance is the corresponding block of $\Phi\Phi^T$; this is the classical Gaussian approximation of the coordinates of a uniform point [57, (1)]. The block concentrates around θ^{-1} times the identity. Its diagonal entries have means $h_i(t_p) = \mathbb{E}Y_i(t_p)^2$, which are close to θ_N^{-1} . Indeed, the mean of $|Y|^2$ solves a renewal equation whose kernel $(2/N) \sum_{i \geq 2} e^{-4a_i t}$ is a defective probability density with total mass $1 - \theta_N$, the mean square of Y_1 satisfies $|h_1(t) - \theta_N^{-1}| \leq C[(1+t)^{-1/2} + t/N]$ for $1 \leq t \leq N$, so that $h_1(t_p) = \theta_N^{-1} + o(1)$, and $h_i(t_p)$ is close to $h_1(t_p)$ for each fixed i , since $a_i t_p \rightarrow 0$ (Lemma 16.12).

Growth of the top coordinate. After t_p , the martingale Y_1 barely moves, by the second part of Proposition 16.18, while $|Y(t)|^2$ relaxes to $Y_1(t_p)^2/\theta_N$ (Proposition 16.19). In

particular, at the times $t_u = \log R_N / (4a_2) + u/a_2$ we have $X_1(t_u)^2/N \rightarrow \theta$: the first eigencoordinate grows from order one to $\sqrt{\theta N}$.

The ratio of the first two coordinates. By Duhamel's formula, we have for $t \geq t_p$

$$Y_2(t) = e^{-2a_2(t-t_p)} Y_2(t_p) + \sqrt{2/N} \int_{t_p}^t e^{-2a_2(t-s)} |Y(s)| dB_2(s).$$

We now argue heuristically. During the last part of $[t_p, t_u]$, of length of order a_2^{-1} , we may replace $|Y(s)|$ by $|Y_1(t_p)|/\sqrt{\theta N}$. Conditionally on $(Y_1(t_p), Y_2(t_p))$, the stochastic integral is approximately centered Gaussian with variance $Y_1(t_p)^2/(2\theta_N N a_2)$. After division by $|Y_1(t_p)|$, its Gaussian approximation is independent of this early pair. Recall that $X_2 = \sqrt{N} Y_2/|Y|$, and note that $a_2 t_p \rightarrow 0$ and $\sqrt{2a_2 N \theta_N} e^{-2a_2 t_u} = \sqrt{2\theta_N} e^{-2u}$, since $N a_2 = R_N$. We find that

$$\sqrt{2a_2} X_2(t_u) \approx \sqrt{2\theta_N} e^{-2u} \frac{Y_2(t_p)}{|Y_1(t_p)|} + G, \quad (2.15)$$

where G is a standard Gaussian variable independent of $(Y_1(t_p), Y_2(t_p))$. By the first step, the ratio is close in law to $H_2/|H_1|$, which has the standard Cauchy law [110]. Under $\pi_{N,c}$, the variable $\sqrt{2a_2} x_2$ is approximately standard Gaussian. As $\mathcal{V}_\theta(u)$ is the total variation distance between the laws of $G + \sqrt{2\theta} e^{-2u} C$ and G , this identifies the profile, provided that the other coordinates do not contribute.

More precisely, for every fixed k , the vector $(\sqrt{2a_j} X_j(t_u))_{j=2}^k$ converges in distribution to $(G_2 + \sqrt{2\theta} e^{-2u} C, G_3, \dots, G_k)$ by Theorem 16.20, under the hypothesis $H_k(\omega, D_k)$ of Definition 16.3 when $k \geq 3$. In particular, $a_3 \geq (1 + \omega)a_2$ for a constant $\omega > 0$. The coordinates $j \geq 3$ forget their values at time t_p , since this separation of the gaps makes their inherited part negligible. As the total variation distance decreases under projections, we conclude that $\liminf_N D_N(t_u) \geq \mathcal{V}_\theta(u)$.

The upper bound

The main difficulty is the upper bound, since D_N involves all N coordinates and weak convergence of finitely many of them does not control it. We first reduce to a hemisphere: by Lemma 3.1, $D_N(t)$ equals the distance between the laws folded onto $\{x_1 > 0\}$. There we use the graph coordinates $y = (x_2, \dots, x_N)$, with $x_1 = (N - |y|^2)^{1/2}$. By Proposition 16.9, the Gibbs measure in these coordinates is close to the Gaussian product law $G = \bigotimes_{j \geq 2} \mathcal{N}(0, (2a_j)^{-1})$, with an error at most $C/m_N + Ce^{-c_0 m_N}$ that includes the factor $(1 - |y|^2/N)^{-1/2}$ of the parametrization. The bound is uniform in θ_N , which we need in Section 17. We then split y into its first coordinates y_z (the retained coordinates $x_2, \dots, x_{J_N}$ at a fixed temperature, described below, and $x_2, \dots, x_J$ in the approach from below) and the remaining coordinates y_v . By Lemma 11.21, we have

$$\left| \|\rho - \pi\|_{\text{TV}} - \|\mu - G_z\|_{\text{TV}} \right| \leq 2(1 - Z) + \sqrt{H_v/2} + \|\rho_z - \mu\|_{\text{TV}} + \|\pi - G\|_{\text{TV}}. \quad (2.16)$$

Here ρ is the time- t law, and $\sigma \leq \rho$ is the endpoint subprobability measure weighted by a path weight with values in $[0, 1]$ and by a cutoff function, with mass $Z > 0$. The measure σ/Z is a normalized cutoff-weighted law in the sense of Section 11, with normalizing mass Z . Further, H_v is the relative entropy of σ/Z with respect to the product of its marginal on y_z and G_v , and μ is the limiting law of y_z . We now explain how we bound the four terms on the right side.

Weighting at the entrance into a cap. At a fixed temperature, we weight the path law by a smooth function of the position at the time $T_e = \chi_N/a_2$, with $\chi_N \rightarrow \infty$ slowly. This function takes values in $[0, 1]$, equals one on a cap $\{x_1^2 \geq \sigma N\}$, and vanishes outside a larger cap. By Proposition 16.13, the mean of the weight is at least $1 - \varepsilon$ for large N , where $\varepsilon > 0$ is small and fixed. By Proposition 16.14, the path then stays in a slightly larger cap up to times of order $\log R_N/a_2$, outside an event whose probability is exponentially small in $R_N/\log R_N$. We also multiply the endpoint measure by a cutoff function that depends only on x_1^2 , equals one on $\{x_1^2 \geq rN\}$, and vanishes on $\{x_1^2 \leq rN/2\}$, for a fixed $r > 0$. We bound the derivatives of the cutoff and the mass of the region outside which it equals one (Lemmas 16.22 and 16.23).

The conditional relative entropy of the remaining coordinates. In Section 11, we compute the time derivative of the conditional relative entropy of the remaining coordinates given the first ones exactly (Proposition 11.18) and prove a logarithmic Sobolev inequality for this conditional relative entropy (Proposition 11.16). It follows from the Gaussian logarithmic Sobolev inequality [79] after a comparison of the metric of the remaining coordinates (Lemma 11.15). The entropy therefore decays at a rate close to $4b$, where b is the smallest of the gaps a_j of the remaining coordinates, up to errors from the cutoff function. At a fixed temperature, the upper bound on the density of the weighted law gives an initial value of order N at the time $T_e + a_2^{-1}$ (Lemma 16.27). Heuristically, if only x_2 is retained, decay at rate $4a_3$ during a time close to $\log R_N/(4a_2)$ multiplies the entropy by about $R_N^{-a_3/a_2}$. Since N is of order R_N^3 , this does not beat the initial value unless $a_3 > 3a_2$. To overcome this, we retain all coordinates with $a_j \leq 16a_2/r$, and their number is bounded (Lemma 16.16). On the support of the cutoff function, the metric of the sphere in graph coordinates is at least $r/2$ times the Euclidean metric. The Gaussian logarithmic Sobolev inequality for the remaining coordinates then bounds the entropy by $(32a_2)^{-1}$ times its dissipation (Lemma 16.17). After the error terms are absorbed, the entropy decays at rate $16a_2$, which multiplies it by about R_N^{-4} during this time. This beats the initial value, and the entropy tends to zero at the times t_u (Proposition 16.28).

The first coordinates in total variation. We upgrade weak convergence to convergence in total variation by Gaussian smoothing (Theorem 11.8). For this, we need that the limit μ has finite Fisher information, and that the laws ρ_z are close in total variation to laws with bounded Fisher information. The limit is a Gaussian convolution, whose

Fisher information is at most the dimension; no moment of the Cauchy law is needed (Lemma 11.7). For the time- t law, we bound the score with a Bismut-type formula over the final time interval of length a_2^{-1} (Theorem 11.5 and Corollary 11.6). This formula controls the derivative, along a rotation in the plane of two eigencoordinates, of the logarithm of the density of an endpoint law weighted by a functional of the path. The bound uses only the final interval of the original uniformly initialized diffusion, with no score bound before that interval; its left-endpoint law may be evolved and prefix-weighted. Such formulas go back to Bismut [32] and to Elworthy and Li [67, Theorems 2.1 and 3.1]. With the Fisher information of the coordinates in the graph chart (Lemma 11.11), this leads to Corollary 11.12, which we use in the approach from below. At a fixed temperature, we apply the formula over final intervals of length $a_j^{-1} \leq a_2^{-1}$, bound the Fisher information of finitely many coordinates directly in the graph chart, and apply Theorem 11.8. This gives Proposition 16.29, and Theorem 16.30 is its case of the second coordinate.

The Gibbs measure. The last term in (2.16) is small by Proposition 16.9. The parameters of the proof are chosen in the order of Table 16.1. Since J_N is bounded, Section 16.11 combines the estimates along subsequences on which J_N is constant.

The approach from below

Along $c_N = 1 + \epsilon_N$, the top fraction $\theta = \theta_N$ tends to zero, and $m = \theta N a_2$ may tend to infinity arbitrarily slowly. The main difficulty compared with a fixed temperature is uniformity: the time scales θ^{-2} and a_2^{-1} and the size $\theta^{-1/2}$ of the first unnormalized coordinate all depend on θ , and every error must be bounded by a negative power of m . As we describe in Section 17.1, the mean $\mathbb{E}|Y(t)|^2$ decreases like $N(1+t)^{-1/2}$ up to the time θ^{-2} and like $N\theta^{-2}t^{-3/2}$ afterwards, until a time of order a_2^{-1} ; after that time, it is of order θ^{-2} , with its radial part determined by the first coordinate through $|Y(t)|^2 \approx Y_1(t)^2/\theta$. In Lemma 17.8 and Proposition 17.11, we prove renewal estimates and moment bounds of all fixed orders with constants that do not depend on θ . Let $\tau = m^\alpha \theta^{-2}$ with $\alpha = 1/10$, let $H_i = \sqrt{\theta} Y_i(\tau)$, and let $t_m(u) = \log(2m)/(4a_2) + u/a_2$. Then the following hold.

- The vector $(H_1, \dots, H_J)$ is close to a standard Gaussian vector (Lemma 17.16).
- The process $\theta^2 |Y(t)|^2$ stays close to H_1^2 along a time interval of length of order $a_2^{-1} \log m$ that contains the times $t_m(u)$ for u in a compact set, with an error bounded by a negative power of m in every fixed L^p norm (Theorem 17.14).
- Let $E = \{\ell_- \leq |H_1| \leq \ell_+, |H_i| \leq \ell_+ \ (2 \leq i \leq J)\}$. On E , outside an event of probability at most Cm^{-B_0} for every fixed $B_0 > 0$, the path does not leave the domain where $X_1^2/(\theta N) \in [1/4, 4]$ during the time interval of the previous item (Corollary 17.17). The losses from killing the path on leaving this domain and from

the cutoff function are controlled by Corollary 17.18. We choose the bounds ℓ_-, ℓ_+ so that $\mathbb{P}(E)$ is arbitrarily close to one and condition the path law on E .

At the times $t_m(u)$, the analog of (2.15) is $\sqrt{2a_2}X_2 \approx e^{-2u}H_2/|H_1| + G_2$, which gives the profile $\mathcal{V}$ and the lower bound (Theorem 17.19).

For the upper bound, the initial value of the conditional relative entropy comes from a different argument. Theorem 17.20 rests on the density bounds of Section 12, which we describe in Section 2.4: at times $t \geq \theta^{-2}$, every moment of the density of the time- t law with respect to the Gibbs measure is at most $e^{O(N\theta^3)} = e^{O(m^3)}$. The initial conditional relative entropy is then of order m^3 . Heuristically, decay at rate $4a_3$ over an interval of length $(1/4 - 1/100)a_2^{-1} \log m$ would multiply the entropy by $m^{-(24/25)(a_3/a_2)}$. If a_3/a_2 is close to one, this does not beat the initial factor m^3 . To overcome this, we keep the first J coordinates, with J so large that $b = a_{J+1} \geq 6a_2$. Then the conditional relative entropy of the remaining coordinates decays at a rate at least $4 \cdot \frac{7}{8} \cdot b$, up to errors of order m^{-1} (Theorem 11.19). Over the interval from $b a_2^{-1} \log(em)$ to $t_m(u)$, of length about $(\frac{1}{4} - \frac{1}{100})a_2^{-1} \log m$, the entropy is multiplied by at most m^{-5} , which beats m^3 (Corollary 17.23). We choose the parameters in the order of Table 17.1.

Finally, the convergence is uniform over all admissible gap vectors in $\mathcal{A}_N$ (Definition 17.1) with $\theta \leq 1/M$ and $m \geq M$, in the iterated limit $\lim_{M \rightarrow \infty} \limsup_{N \rightarrow \infty}$ (Corollary 17.4). We deduce it from the statement along sequences by a diagonal argument (Lemma 17.24), and this is the form that we use in Section 21.

2.4 The critical window

In the critical window $c_N = 1 + \sigma N^{-1/3}$, the half-precisions $r_i = p_i/2$ are of order $(i/N)^{2/3}$ for every i , including $i = 1$ (Lemma 4.17). For every fixed i , the inverse drift rate $(2r_i)^{-1}$ is of order $N^{2/3}$, and the limiting dynamics retains the individual scaled rates. To prove Theorem 1.3, we construct a countable limit of the dynamics in Section 18, and in Section 19 we show that the distance converges to the distance between the normalized limit and the limiting Gibbs law. We write $b = N^{2/3}$.

The critical limit process

Let $R_{i,N} = b r_{i,N}$ be the scaled half-precisions. We rescale the unnormalized process (2.2) with $\rho_i = r_{i,N}$ by

$$U_{i,N}(s) = N^{-1/6} Y_i(bs), \quad Z_N(s) = b^{-1} |Y(bs)|^2, \quad N^{-1/3} X_i(bs) = \frac{U_{i,N}(s)}{\sqrt{Z_N(s)}}, \quad (2.17)$$

so that $dU_{i,N} = -2R_{i,N}U_{i,N}ds + \sqrt{2Z_N}dB_i^{(N)}$ for independent Brownian motions $B_i^{(N)}$; see (18.7). Each coordinate starts at $U_{i,N}(0) = N^{-1/6}g_i$, which tends to zero. The rescaled

squared norm $Z_N = N^{-1/3} \sum_i U_{i,N}^2$ is a sum over all coordinates, and the corresponding series for the limiting coordinates diverges. We describe the limit of Z_N instead through its mean and its fluctuation. Let $\psi_N(s) = \mathbb{E} Z_N(s) = N^{1/3} \bar{h}(bs)$. Then ψ_N converges to the mean function ψ_R (Lemma 18.5), which is determined by its Laplace transform for $p > 0$:

$$\int_0^\infty e^{-ps} \psi_R(s) ds = \frac{1}{2\Lambda_R(p)}, \quad \Lambda_R(p) = \frac{1}{2} \sum_{i \geq 1} \left(\frac{1}{R_i} - \frac{1}{R_i + p/4} \right).$$

The mean function solves a Volterra equation that is the rescaled limit of the renewal equation for $\bar{h}$ (Section 2.2), and in this sense it is the rescaled limit of the scalar equation of Cugliandolo and Dean. Note that $\sqrt{s} \psi_R(s) \rightarrow \psi_* \in (0, \infty)$ as $s \downarrow 0$. The fluctuation $Z_N - \psi_N$ converges as well, and the limit (U, Z) solves the countable system (18.51):

$$\begin{aligned} U_i(s) &= \sqrt{2} \int_0^s e^{-2R_i(s-v)} \sqrt{Z(v)} dB_i(v) \quad (i \geq 1), \quad Z = \psi_R + V, \\ V(s) &= \sqrt{8} \sum_{i \geq 1} \int_0^s k_i(s-v) \sqrt{Z(v)} U_i(v) dB_i(v), \end{aligned}$$

with the explicit kernels k_i of (18.28). This is the critical limit process of Theorem 18.9. We show there that its law is unique and that finitely many rescaled eigencoordinates converge to the coordinates $\xi_i = U_i/\sqrt{Z}$ of the normalized critical limit process; see (18.53). Uniqueness comes from a Picard iteration for a system in which the noise amplitude $\sqrt{2Z}$ and a weighted norm of the coordinates are truncated (Lemma 18.8), and the moment bounds that we need for tightness come from Lemmas 7.14 and 10.10. A stochastic integral is not a continuous function of its integrand and its driving path. We identify the stochastic integrals of a limit point through the joint laws of the rescaled processes, by approximating them with Riemann sums (Lemma 18.11).

The start is singular, and this is how the uniform initial distribution enters the limit. Every coordinate starts at zero, while $Z(s) \sim \psi_* s^{-1/2}$ and the weighted norm $\sum_i U_i(s)^2/R_i$ tends to 2 as $s \downarrow 0$ (Remark 18.10). In finite dimension, this weighted norm is a martingale whose initial value $N^{-1/3} \sum_i g_i^2/R_{i,N}$ is spread over all coordinates, as it is for a uniform point of S_N .

The rescaled Gibbs measures converge as well. By Corollary 18.12, the first coordinates of $N^{-1/3}x$ under π_{N,c_N} converge in total variation to those of ρ_R , the Gaussian sequence with variances $(2R_i)^{-1}$ conditioned on the renormalized radial constraint of Definition 5.8. At the critical temperature, this law is due to Du and Huang [62, Section 2.2]. The proof of Theorem 19.2 compares the normalized critical entrance process with this limiting Gibbs law.

Density bounds

The second ingredient of the critical window, which we also use along $c_N = 1 + \epsilon_N$, bounds the density of the time- t law with respect to the Gibbs measure in every space L^p ,

at times when the distance is still close to one. By Theorem 12.8, under the hypotheses of Definition 12.2 on a time interval $[0, S]$, we have for every fixed $p > 1$

$$\log \int_{S_N} \left(\frac{d(\nu_N P_t^r)}{d\pi_r} \right)^p d\pi_r \leq C_p N(1+t)^{-3/2}, \quad 0 \leq t \leq S. \quad (2.18)$$

Heuristically, the right side counts the coordinates that have not equilibrated at time t , namely those with $r_{it} \lesssim 1$, and each of them contributes a bounded amount. In the critical window, the hypotheses hold on $[0, KN^{2/3}]$, and the right side is bounded at times of order $N^{2/3}$ (Theorem 12.13). With the Poincaré constant of order $N^{2/3}$ of Proposition 8.21, we deduce from the L^2 bound at time b that $D_N(b+u) \leq Ce^{-u/(K_P b)}$, where $K_P b$ is an upper bound on the Poincaré constant (see (19.28)). This bound gives the behavior of the critical profile at infinity (Proposition 12.14).

In the proof, we differentiate the Rényi divergence of the time- t law with respect to a moving reference measure, the Bingham comparison law $\mu_{\gamma t}$ at an earlier time γt (Lemma 12.7). Rényi divergences were introduced in [122], and Vempala and Wibisono computed their derivative along the Langevin dynamics when the reference measure is stationary [148, Lemmas 5 and 6]; here the reference measure moves. A logarithmic Sobolev inequality for $\mu_{\gamma t}$ with constant of order $1+t$ (Proposition 8.21) bounds the dissipation from below. The error terms are quadratic forms in the coordinates with small coefficients, because a term common to all coordinates cancels exactly (Lemma 12.4). We compute their exponential moments under $\mu_{\gamma t}$ exactly from Lemma 5.13 (Lemma 12.5). A scalar differential inequality closes the argument (Lemma 12.6).

The logarithmic Sobolev inequalities come from Section 8. Extending functions on the sphere to Gaussian space, we find logarithmic Sobolev inequalities with an additive error term (Theorem 8.11). Combined with Rothaus's lemma [125] and a Poincaré inequality, they imply a logarithmic Sobolev inequality as soon as three variances are comparable to the largest one (Corollary 8.13). The Poincaré inequalities are proved separately when many precisions are comparable to the smallest one (Corollary 8.9), and when only a few precisions are close to the smallest one, which is the case of the Gibbs measure in the critical window (Lemmas 8.18 and 8.19).

Convergence of the distance

Let μ_s^∞ be the law of $\xi(s)$ on $\mathbb{R}^N$, and let $F_R(s) = \|\mu_s^\infty - \rho_R\|_{TV}$. Let $(r_{\cdot, N})$ be a deterministic sequence of half-precisions that satisfies the critical spectral conditions and converges to R . We prove in Theorem 19.2 that $D_N(bs) \rightarrow F_R(s)$ uniformly on compact subsets of $(0, \infty)$, and that F_R is continuous and strictly decreasing, with $F_R(0+) = 1$ and $F_R(\infty) = 0$. The obstacle is again that D_N involves all coordinates: the weak convergence of finitely many coordinates of Theorem 18.9 controls neither the remaining coordinates nor the total variation distance of the first ones. We first prove convergence at every fixed positive time (Proposition 19.19), in two steps.

The remaining coordinates. By Lemma 19.15, for every $s_0 > 0$ we have $0 \leq D_N(bs) - D_N^{\leq J}(bs) \leq CJ^{-1/3}$ for all $s \geq s_0$, where $D_N^{\leq J}$ is the distance between the laws of the first J coordinates and C depends on s_0 . The argument is as follows. The density of the time- t law with respect to the Gibbs measure is bounded in L^2 at $t = s_0b/2$ (Theorem 12.13). The smoothing of the finite-dimensional semigroup then bounds its energy at every time bs with $s \geq s_0$ (Lemma 19.12). Next, the conditional law of the remaining coordinates given the first J satisfies a logarithmic Sobolev inequality with constant $Cb/(\gamma J^{2/3})$, uniformly over all conditioning values (Lemma 19.14); we deduce it from Corollaries 8.9 and 8.13 on spheres of arbitrary radius (Lemma 19.13). The resulting Poincaré inequality bounds the conditional variance of the density by a multiple of $J^{-2/3}$ times its energy, and we obtain the bound $J^{-1/3}$ by comparing the L^1 and L^2 norms. Here J is fixed before $N \rightarrow \infty$.

The first coordinates. For fixed J and $s > 0$, the total variation error between the first J rescaled coordinates under $\nu_N P_{bs}$ and their limiting law tends to zero (Lemma 19.17). Weak convergence holds by Theorem 18.9. To upgrade it, we integrate by parts under the Gibbs measure along vector fields that move the first J coordinates and compensate in the remaining ones. The divergences of these fields are bounded in $L^2(\pi_N)$ through inverse moments of a block of coordinates, which follow from the Gaussian domination of the Gibbs measure (Corollary 18.12 and Lemma 19.16). Together with the density and energy bounds above, this gives a bound on translations of the first-coordinate law that is uniform in N and in $s \geq s_0 > 0$. Mollification and weak convergence then give convergence in total variation at each fixed s . Since total variation on $\mathbb{R}^N$ is the limit of the distances between the first J coordinates (Lemma 19.7), we combine this convergence with the bound on the remaining coordinates, let $N \rightarrow \infty$, and then let $J \rightarrow \infty$. This proves $D_N(bs) \rightarrow F_R(s)$ for every $s > 0$ (Proposition 19.19).

Continuity, positivity, and strict decrease. We prove these properties of F_R with estimates for the finite-dimensional semigroups that are uniform in N , and we do not need a semigroup for the limiting process. By the spectral theorem, $\|(P_{bh} - I)P_{bt}u\|_{L^2(\pi_N)} \leq h\|u\|_{L^2(\pi_N)}/(et)$ for $t > 0$ and $h \geq 0$ (Lemma 19.12). We apply this bound to the density at time $s_0b/2$, and we find that $\|\nu_N P_{bt} - \nu_N P_{bs}\|_{TV} \leq L|t - s|$ for $s, t \geq s_0$, where L does not depend on N . This bound passes to the limit through the weak convergence of the first J coordinates and Lemma 19.7, and F_R is locally Lipschitz (Lemma 19.18). The exponential lower bound on D_N of Section 2.4 passes to F_R by the convergence at every fixed time, and F_R is positive. For strict decrease, suppose that $F_R(s_1) = F_R(s_2)$ for some $s_1 < s_2$, and write $f - 1 = g_+ - g_-$, where f is the density of the law at time bs_1 and $g_{\pm} \geq 0$. Then the correlations $\langle g_+, P_{bt}g_- \rangle_{L^2(\pi_N)}$ tend to zero for $t \in [s_2 - s_1, 2(s_2 - s_1)]$, while the constant terms of their spectral expansions tend to $F_R(s_1)^2 > 0$ (Lemma 19.20). On the time scale b , the spectral gap of Proposition 8.21 is bounded below uniformly in N , and the sums of the absolute values of the other spectral coefficients are bounded.

An argument with Chebyshev polynomials then forces $F_R(s_1) = 0$, which contradicts positivity (Lemma 19.9). Since $D_N(bs)$ is nonincreasing in s and converges at every s to the continuous function F_R , a finite grid of times gives the uniform convergence on compact sets (Theorem 19.2). We record the order in which the parameters are chosen in Table 19.2.

Lower bounds and the GOE

The lower bounds on the critical time scale do not use the profile. A finite-dimensional estimate at a short critical time shows that $D_N(s_*b) \geq 3/4$ (Lemma 19.10). For a reversible semigroup, a lower bound at time 2ℓ and a bound on the L^4 norm of the density at time ℓ imply an exponential lower bound at all times (Lemma 19.6). With Theorem 12.13, we find that $D_N(t) \geq c_1 e^{-C_1 t/b}$ for all $t \geq 0$. Together with the upper bound (19.28), this gives $c_2 b \log(1/\varepsilon) \leq t_N(\varepsilon) \leq C_2 b \log(1/\varepsilon)$ for every $\varepsilon \in (0, 1/2)$, including accuracies that depend on N (Corollary 19.11). Combining the lower bound with the pointwise convergence (19.51) proved in Theorem 19.2, we have $F_R(s) \geq c_1 e^{-C_1 s}$ (Corollary 19.5).

For the GOE, the half-precisions satisfy the critical spectral conditions on events of probability at least $1 - \eta$, uniformly for σ in compact sets, and the spectral data converge jointly in distribution to their limits, which are functions of the reflected GOE Airy point process (χ_i) , the limit of $N^{2/3}(2 - \lambda_i)$, for finitely many values of σ at once (Lemma 4.17). The finite and limiting profiles and mixing times are Borel functions of the spectral data (Lemmas 19.21 and 19.22). The maps from the spectral data to the finite profiles depend on N . We control them only along convergent sequences of spectral data that satisfy the critical spectral conditions with fixed constants (Theorem 19.2 and Corollary 19.3). A localized form of the continuous mapping theorem (Lemma 19.23), applied on the events where these conditions hold, then gives Theorem 19.4, jointly with the spectral data and for finitely many temperatures at once. The limit is a function of the Airy point process arising from the distributional edge limit, and the critical mixing profile is random. Since $F_\sigma^{-1}(\varepsilon)/F_\sigma^{-1}(1 - \varepsilon) > 1$ almost surely, there is no cutoff.

2.5 The critical mixing profile far from the critical temperature

Theorem 1.4 concerns the iterated limit in which $N \rightarrow \infty$ first, at fixed $\sigma = \mp S$, and then $S \rightarrow \infty$. The mesoscopic theorems give no information here, since they assume $N\delta_N^3 \rightarrow \infty$ or $m_N \rightarrow \infty$, and the critical theorem holds at fixed σ . The link between the regimes is that several estimates of Parts II to IV hold at every finite N , with constants that depend only on the spacing of the precisions, and not on N or on the size of the effective parameter, once this parameter exceeds a fixed threshold. We use them for finite sequences that approximate the limiting rates of the critical window: we let $N \rightarrow \infty$ with the effective parameter fixed, and only afterwards let it tend to infinity.

Finite approximations. Fix a realization of the Airy point process outside a null set. For $\xi > 0$ and all sufficiently large N depending on ξ , we construct in Lemma 20.1 the precisions

$$p_{i,N} = N^{-2/3}(\Delta_i + \xi) \quad (i \leq N/2), \quad p_{i,N} = \gamma_N N^{-2/3}(\Delta_i + \xi) \quad (i > N/2),$$

with $\gamma_N \in (1, 3]$ chosen such that $\sum_i p_{i,N}^{-1} = N$. The smallest precision is $\xi N^{-2/3}$, so that $M = N p_{1,N}^{3/2} = \xi^{3/2}$ does not depend on N , and the spacing constants do not depend on N or ξ . The half-precisions $p_{i,N}/2$ satisfy the critical spectral conditions of Definition 4.16 and converge to $R = R^{\mathcal{S}(\xi)}$, and by Theorem 19.2 their distances converge to $F_{\mathcal{S}(\xi)}$.

Far above the critical temperature. In Section 20, we take $\xi = \alpha^2$ with α large. For these approximations, we bound the three errors of Section 2.2 at finite N , with constants that do not depend on N or α . At the time $t = w/d_N$, the comparison error is at most Ce^{-w} by (2.13), and the relative errors are at most Cwe^{-2w} by Lemma 7.12. Further, η_q is of order $\alpha^{-3/2}$, so that Theorem 6.6 applies. The information converges as $N \rightarrow \infty$: by (2.8) and Lemma 18.5, we have $\mathcal{I}_N(N^{2/3}s) \rightarrow \psi''(2s)/(8\psi(s)^2)$, where ψ is the mean function of the limit (Lemma 20.2). As $\alpha \rightarrow \infty$, the transient part of ψ is the Laplace transform of a positive measure with an explicit real resolvent, whose asymptotics follow from Lemma 7.15 (Lemma 20.3). At the times $(w_\alpha^* + u)/\alpha^2$ with $w_\alpha^* = \frac{1}{4}(\log \alpha^3 - \frac{3}{2} \log \log \alpha^3 - \log(16\sqrt{\pi}))$, the limiting information tends to e^{-4u} . We let $N \rightarrow \infty$ at fixed α , then $\alpha \rightarrow \infty$, and finally pass from α to S through $\alpha^2 = \varrho_{-S}$. This proves Theorem 20.4; the order of the choices is in Table 20.1.

Far below the critical temperature. In Section 21, we take $\xi = \varrho_S$, which tends to zero as $S \rightarrow \infty$. By Lemma 21.1, the scaled gaps $a_{i,N} = (p_{i,N} - p_{1,N})/2$ are admissible in the sense of Definition 17.1, with constants that do not depend on S , while $N^{2/3}a_{2,N} = \Delta_2/2$, $N^{1/3}\theta_N \rightarrow S - \Xi$, and $m_N \rightarrow (S - \Xi)\Delta_2/2$. At fixed S , the parameter m_N has a finite limit, and Theorem 17.3 does not apply, but the uniform statement Corollary 17.4 does. For every sufficiently large L and all S with $(S - \Xi)\Delta_2/2 > L$, it implies that the error at fixed S is at most $E_I(L)$ in the limit $N \rightarrow \infty$, where $E_I(L) \rightarrow 0$ as $L \rightarrow \infty$. The critical theorem identifies the limit of the distance at the corresponding times with $F_S(\tau_\xi^c + 2u/\Delta_2)$. Letting $S \rightarrow \infty$ and then $L \rightarrow \infty$ proves Theorem 21.2.

2.6 Dependencies

In Table 2.2, we list for each section its main results and the earlier sections whose results it uses. Every section uses the conventions of Section 3, which we do not list.

Section	Main results	Uses Sections
4	random-matrix inputs; the Lagrange multiplier	—
5	Gaussian laws conditioned on a sphere	—
6	Theorems 6.6 and 6.8	5
7	covariance evolution; Lemmas 7.3, 7.12, and 7.14	5
8	Theorems 8.6 and 8.8 and Proposition 8.21	4, 5, 7
9	Theorems 9.5, 9.13, and 9.16	5, 7, 8
10	Lemmas 10.2 and 10.10 and Theorems 10.15 and 10.18	5, 7, 8
11	Theorems 11.5, 11.8, and 11.19	—
12	Theorems 12.8 and 12.13	5, 7, 8
13	Theorem 1.1(2)	4, 6, 7, 9
14	Theorem 1.1(3)	4–7, 9, 10
15	Theorem 1.1(1), Theorem 1.2	4, 7, 9, 13
16	Theorem 1.1(5)	4, 10, 11
17	Theorem 1.1(4)	4, 8, 10–12, 16
18	the critical limit process; Corollary 18.12	4, 5, 7, 10
19	Theorem 1.3	4, 5, 7, 8, 10, 12, 18
20	Theorem 1.4, first part	4, 6, 7, 14, 18, 19
21	Theorem 1.4, second part	4, 5, 17, 19, 20

Table 2.2. Main results of each section and the earlier sections that it uses, besides Section 3. Sections 5–12 form Part II.

PART I

Preliminaries

The model, conventions, and the limiting profiles

This section fixes the model and the conventions used throughout the paper. In Section 3.1, we define the spherical Langevin dynamics and its Gibbs measure, and record three exact reductions: diagonalization of the disorder, invariance under adding a multiple of the identity to the quadratic form, and invariance under reflections of eigencoordinates (Lemma 3.1). We also name the symmetry sectors of L^2 (Definition 3.2). Section 3.2 introduces the Lagrange multiplier and the precisions of the Gaussian representation of the Gibbs measure, the information of one variance sequence relative to another (Definition 3.5), and the parameters attached to each temperature regime. Section 3.3 defines the total variation distance, the mixing times, and what we mean by a mixing profile, by a cutoff profile, and by cutoff (Definition 3.7), and recalls standard facts about total variation and relative entropy (Lemma 3.6). Section 3.4 states once how probabilities over the disorder and the constants in our estimates are quantified. In Section 3.5, we prove the properties of the three cutoff profiles Ψ , $\mathcal{V}$, and $\mathcal{V}_\theta$ (Lemma 3.9), and show that pointwise convergence of a nonincreasing distance to a profile implies locally uniform convergence and convergence of the mixing times (Lemma 3.11).

The section uses no earlier results. Its definitions are used everywhere. Lemmas 3.9 and 3.11 are used in each section that identifies a profile, namely Sections 13 to 17, 20, and 21.

3.1 The model

Disorder. Fix $N \geq 2$. Let $W_N = (W_{ij})_{i,j=1}^N$ be a matrix from the Gaussian orthogonal ensemble (GOE): a real symmetric random matrix whose entries W_{ij} , $1 \leq i \leq j \leq N$, are independent centered Gaussian random variables with $\mathbb{E}W_{ij}^2 = 1/N$ for $i < j$ and $\mathbb{E}W_{ii}^2 = 2/N$. We denote its eigenvalues by $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_N$. We write $\lambda_{i,N}$ when the dependence on N matters. Almost surely the eigenvalues are distinct. Indeed, the discriminant of the characteristic polynomial of a symmetric matrix is a polynomial in its entries which does not vanish at a diagonal matrix with distinct diagonal entries. Its zero set has Lebesgue measure zero, while the law of W_N has a density with respect to Lebesgue measure on the space of real symmetric matrices. We write

$$\rho_{\text{sc}}(x) = \frac{\sqrt{\max\{4 - x^2, 0\}}}{2\pi}$$

for the semicircle density.

The dynamics. Let $S_N = \{x \in \mathbb{R}^N : |x|^2 = N\}$, where $|\cdot|$ is the Euclidean norm, and let ν_N be the uniform probability measure on S_N . We write ∇_{S_N} and Δ_{S_N} for the gradient and the Laplace–Beltrami operator of S_N with the metric induced from $\mathbb{R}^N$, and $\langle \cdot, \cdot \rangle$ for the Euclidean inner product. For an inverse temperature $\beta \geq 0$, we set $c = \sqrt{2}\beta$ and

$$H_N(x) = \frac{x^\top W_N x}{\sqrt{2}}, \quad L = \Delta_{S_N} + \beta \nabla_{S_N} H_N \cdot \nabla_{S_N}, \quad \pi_{N,c}(dx) = \frac{e^{\beta H_N(x)} \nu_N(dx)}{\int_{S_N} e^{\beta H_N} d\nu_N}. \quad (3.1)$$

The inverse temperature multiplies H_N and is not part of it: the Gibbs weight is $e^{+\beta H_N}$, so the physical energy is $-H_N$. Note that $\beta H_N(x) = (c/2) x^\top W_N x$. We call $c = \sqrt{2}\beta$ the rescaled inverse temperature, and use it in place of β throughout. Increasing c lowers the temperature. The critical value is $c = 1$, that is, $\beta_c = 1/\sqrt{2}$, and high and low temperature mean $c < 1$ and $c > 1$, respectively. The measure $\pi_{N,c}$ is the Gibbs measure of the quadratic spherical spin glass (the 2-spin spherical Sherrington–Kirkpatrick model) introduced by Kosterlitz, Thouless, and Jones [97]. As a probability measure on a sphere with density proportional to the exponential of a quadratic form, it is a Bingham distribution [31]. We write $P_t = e^{tL}$ for the Markov semigroup generated by L , and, for a probability measure μ on S_N , we write μP_t for the law at time t of the diffusion with initial law μ . We call μP_t the time- t law.

Normalization. The GOE variances $1/N$ off the diagonal and $2/N$ on the diagonal agree with those of the completion matrix for the Sherrington–Kirkpatrick model. In the convention of the Sherrington–Kirkpatrick literature in which the Gibbs weight is $e^{(\beta/2)x^\top W_N x}$, our weight $e^{(c/2)x^\top W_N x}$ has the same form. Thus c plays the role of the SK inverse temperature, and $c = 1$ corresponds to $\beta_c^{\text{SK}} = 1$. Here the reference measure ν_N is the uniform probability measure on S_N , rather than counting measure on the spin space.

In ambient coordinates, the diffusion generated by L is the solution of

$$dX_t = \sqrt{2} \Pi(X_t) dB_t + \left(c \Pi(X_t) W_N X_t - \frac{N-1}{N} X_t \right) dt, \quad \Pi(x) = I - \frac{xx^\top}{N}, \quad (3.2)$$

where $X_0 \in S_N$ and B is a standard Brownian motion in $\mathbb{R}^N$, independent of (W_N, X_0) . For the model, X_0 has law ν_N independently of W_N . The noise coefficient is $\sqrt{2}$, so that the generator contains Δ_{S_N} and not $\Delta_{S_N}/2$. The coefficients of (3.2) are polynomial, hence locally Lipschitz. To see that the solution stays on S_N , let $U_t = 1 - |X_t|^2/N$. Since $x^\top \Pi(x) = (1 - |x|^2/N) x^\top$ and $\text{Tr}(\Pi(x)\Pi(x)^\top) = N - 2|x|^2/N + |x|^4/N^2$, Itô's formula gives

$$dU_t = -\frac{2U_t}{N} \left(\sqrt{2} X_t^\top dB_t + (c X_t^\top W_N X_t + N - 1 + U_t) dt \right),$$

up to the exit time of X from any ball. This is a linear equation for U with locally bounded adapted coefficients, and its unique solution with $U_0 = 0$ is $U \equiv 0$. Hence

$|X_t|^2 = N$ for all $t \geq 0$, and in particular the solution exists for all times. For a function f that is smooth near S_N , the identities $\beta \nabla_{S_N} H_N(x) = c \Pi(x) W_N x$ and $\Delta_{S_N} f(x) = \text{Tr}(\Pi(x) \nabla^2 f(x)) - (N-1)N^{-1} \langle x, \nabla f(x) \rangle$ for $x \in S_N$ show that the generator of (3.2) is L .

General quadratic potentials. Many tools in this paper hold for an arbitrary deterministic real symmetric matrix A with eigenvalues $\alpha_1 \geq \dots \geq \alpha_N$. For such A , we write

$$L_A = \Delta_{S_N} + \nabla_{S_N} \left(\frac{1}{2} x^\top A x \right) \cdot \nabla_{S_N}, \quad P_t^A = e^{tL_A}, \quad \pi_A(dx) \propto e^{x^\top A x / 2} \nu_N(dx). \quad (3.3)$$

The model (3.1) is the case $A = cW_N$, for which $\alpha_i = c\lambda_i$. Let $V(x) = x^\top A x / 2$. Then $L_A f = e^{-V} \text{div}_{S_N}(e^V \nabla_{S_N} f)$, and the divergence theorem on the closed manifold S_N gives, for smooth f, g ,

$$- \int_{S_N} f L_A g d\pi_A = \int_{S_N} \langle \nabla_{S_N} f, \nabla_{S_N} g \rangle d\pi_A. \quad (3.4)$$

In particular, L_A is symmetric and nonpositive on $L^2(\pi_A)$, and π_A is reversible and invariant for P_t^A . The Dirichlet form of L_A is $f \mapsto \int |\nabla_{S_N} f|^2 d\pi_A$, built from the tangential gradient. All Poincaré and logarithmic Sobolev inequalities in this paper refer to this Dirichlet form. For a probability measure π on S_N and a nonnegative measurable function g with $\pi(g) < \infty$, let $\text{Ent}_\pi(g) = \pi(g \log g) - \pi(g) \log \pi(g) \in [0, \infty]$, with $0 \log 0 = 0$. We say that π satisfies a Poincaré inequality with constant C_P , and a logarithmic Sobolev inequality with constant C_{LS} , if for all smooth f ,

$$\text{Var}_\pi(f) \leq C_P \pi(|\nabla_{S_N} f|^2), \quad \text{Ent}_\pi(f^2) \leq C_{LS} \pi(|\nabla_{S_N} f|^2). \quad (3.5)$$

The spectral gap of L_A is

$$\lambda_{\text{gap}}(L_A) = \inf \left\{ \frac{\pi_A(|\nabla_{S_N} f|^2)}{\text{Var}_{\pi_A}(f)} : f \text{ smooth and nonconstant} \right\},$$

the smallest nonzero eigenvalue of $-L_A$, by compact weighted elliptic spectral theory [130, Section 2, p. 30]. The optimal constant in the Poincaré inequality is $1/\lambda_{\text{gap}}(L_A)$. Restricting the infimum to the symmetry sectors of Definition 3.2 below, we obtain the spectral gaps of these sectors.

Eigencoordinates. Conditional on W_N , fix an orthonormal basis $u_1, \dots, u_N$ of eigenvectors of W_N , chosen measurably, with $W_N u_i = \lambda_i u_i$, and set $x_i = \langle x, u_i \rangle$. The map $x \mapsto (x_1, \dots, x_N)$ is an orthogonal transformation of $\mathbb{R}^N$. It maps S_N to itself, preserves ν_N , and commutes with ∇_{S_N} and Δ_{S_N} . In these coordinates,

$$\beta H_N(x) = \frac{c}{2} \sum_{i=1}^N \lambda_i x_i^2. \quad (3.6)$$

The distance between the time- t law and the Gibbs measure does not depend on the choice of eigenbasis, because ν_N is invariant under rotations and the initial point is independent of W_N . When we study the dynamics for a fixed realization of W_N , we may therefore assume that $W_N = \text{diag}(\lambda_1, \dots, \lambda_N)$. We call $x_1, \dots, x_N$ the eigencoordinates. No property of the eigenvectors of W_N is used.

Adding a multiple of the identity. For every $\tau \in \mathbb{R}$ and $x \in S_N$, we have $x^\top(A + \tau I)x/2 = x^\top Ax/2 + \tau N/2$. Since the added term is constant on S_N , we obtain

$$L_{A+\tau I} = L_A, \quad \pi_{A+\tau I} = \pi_A, \quad P_t^{A+\tau I} = P_t^A. \quad (3.7)$$

Two choices of τ are used for the model. Taking $A = cW_N$ and $\tau = -c\lambda_1$, we find the form used at low temperature:

$$\begin{aligned} L &= \Delta_{S_N} - \nabla_{S_N} F \cdot \nabla_{S_N}, & \pi_{N,c}(dx) &\propto e^{-F(x)} \nu_N(dx), \\ F(x) &= \sum_{i=2}^N a_i x_i^2, & a_i &= \frac{c}{2}(\lambda_1 - \lambda_i). \end{aligned} \quad (3.8)$$

Here $a_1 = 0 \leq a_2 \leq \dots \leq a_N$, and we call the a_i the scaled eigenvalue gaps. Similarly, for any $z > c\lambda_1$, the choice $\tau = -z$ yields

$$\begin{aligned} L &= \Delta_{S_N} - \nabla_{S_N} \left(\sum_{i=1}^N r_i x_i^2 \right) \cdot \nabla_{S_N}, & \pi_{N,c}(dx) &\propto \exp \left(- \sum_{i=1}^N r_i x_i^2 \right) \nu_N(dx), \\ r_i &= \frac{z - c\lambda_i}{2} > 0. \end{aligned} \quad (3.9)$$

The particular choice of z that is used for the model is made in Section 3.2.

Reflections. For $1 \leq j \leq N$, let R_j be the reflection of the j th eigencoordinate, $R_j(x) = (x_1, \dots, -x_j, \dots, x_N)$, and let $\mathcal{G}$ be the group generated by $R_1, \dots, R_N$. It contains the antipodal map $x \mapsto -x$. For $J \subset \{1, \dots, N\}$, let Q_J be the map that replaces x_j by $|x_j|$ for all $j \in J$.

Lemma 3.1 (Reflections of eigencoordinates). *Let $A = \text{diag}(\alpha_1, \dots, \alpha_N)$.*

1. *For every $g \in \mathcal{G}$, every bounded measurable f on S_N , and every $t \geq 0$, we have $P_t^A(f \circ g) = (P_t^A f) \circ g$. The measures ν_N and π_A are invariant under $\mathcal{G}$, and so is $\nu_N P_t^A$ for every $t \geq 0$.*
2. *Let $\mathcal{H}$ be a subgroup of $\mathcal{G}$, and let μ, μ' be $\mathcal{H}$ -invariant probability measures on S_N . Then*

$$\|\mu - \mu'\|_{\text{TV}} = \sup \{ |\mu(B) - \mu'(B)| : B \subset S_N \text{ Borel and } \mathcal{H}\text{-invariant} \}.$$

In particular, if $\mathcal{H}$ is generated by $(R_j)_{j \in J}$, then $\|Q_{J\#}\mu - Q_{J\#}\mu'\|_{\text{TV}} = \|\mu - \mu'\|_{\text{TV}}$.

Proof. Each $g \in \mathcal{G}$ is a diagonal orthogonal matrix. It commutes with A , it satisfies $g\Pi(x) = \Pi(gx)g$, and gB is a standard Brownian motion. If X solves (3.2) with cW_N replaced by A and $X_0 = x$, then multiplying the equation by g shows that gX solves the same equation with Brownian motion gB and initial point gx . By uniqueness in law, we have $P_t^A(f \circ g)(x) = \mathbb{E}f(gX_t) = (P_t^A f)(gx)$. The measures ν_N and π_A are invariant because g is an isometry of S_N and $x^T A x = (gx)^T A (gx)$. For every bounded measurable f , we then have

$$(\nu_N P_t^A)(f \circ g) = \nu_N((P_t^A f) \circ g) = \nu_N(P_t^A f) = (\nu_N P_t^A)(f),$$

which proves the first part.

For the second part, the inequality $\geq$ is immediate. Let $\sigma = \mu - \mu'$, and let $P \subset S_N$ be a positive set of a Hahn decomposition of σ , so that $\sigma(P) = \|\mu - \mu'\|_{\text{TV}}$ and $\sigma(D) \geq 0$ for all Borel sets $D \subset P$. Each set hP , $h \in \mathcal{H}$, is also a positive set, because σ is $\mathcal{H}$ -invariant. The set $B = \bigcup_{h \in \mathcal{H}} hP$ is $\mathcal{H}$ -invariant, and it is a positive set because every Borel subset of B is a finite disjoint union of Borel subsets of the sets hP . We have $\sigma(B) = \sigma(P) + \sigma(B \setminus P) \geq \sigma(P)$ because $P \subset B$, and the equality follows. If $\mathcal{H}$ is generated by $(R_j)_{j \in J}$, every $\mathcal{H}$ -invariant Borel set B satisfies $B = Q_J^{-1}(B')$ with $B' = \{x \in B : x_j \geq 0 \text{ for } j \in J\}$. For such B , we have $|\mu(B) - \mu'(B)| = |Q_{J\#}\mu(B') - Q_{J\#}\mu'(B')| \leq \|Q_{J\#}\mu - Q_{J\#}\mu'\|_{\text{TV}}$. The reverse inequality holds because $Q_{J\#}\mu(B'') - Q_{J\#}\mu'(B'') = \mu(Q_J^{-1}B'') - \mu'(Q_J^{-1}B'')$ for all Borel sets B'' . $\square$

It follows from the first part of Lemma 3.1 that the density of $\nu_N P_t$ with respect to $\pi_{N,c}$ is invariant under every reflection of an eigencoordinate. It is therefore orthogonal in $L^2(\pi_{N,c})$ to every function that is odd in some eigencoordinate, and such eigenfunctions of L do not affect the distance of the time- t law from equilibrium. An analogous observation for symmetric initial laws, in a spherical quadratic model with a large spike, was made by Gheissari, Grant, and Yu [78, Definition 1.1 and Theorem 1.2(2)].

We use three classes of symmetric functions, whose spectral gaps can differ.

Definition 3.2 (Symmetry sectors). Let $A = \text{diag}(\alpha_1, \dots, \alpha_N)$, and let μ be either ν_N or π_A . The *coordinate-even sector*, the *antipodal-even sector*, and the *sector even in x_1* are the closed subspaces of $L^2(\mu)$ that consist of the functions f with $f \circ R_j = f$ for all j , with $f(-x) = f(x)$, and with $f \circ R_1 = f$, respectively. The spectral gap of L_A in one of these sectors is the infimum in the definition of $\lambda_{\text{gap}}(L_A)$, restricted to nonconstant smooth functions in that sector; both the variance and the Dirichlet energy are taken under π_A . We speak of the coordinate-even and the antipodal-even spectral gaps, and of the spectral gap of the sector even in x_1 .

The measures ν_N and π_A have equivalent L^2 norms, since $d\pi_A/d\nu_N$ is continuous and strictly positive on the compact sphere, so these sectors have the same functions

and null-set convention for either choice of μ . The coordinate-even sector is contained in the other two. By the first part of Lemma 3.1, the semigroup P_t^A maps each sector into itself. A function in the sector even in x_1 is determined by its restriction to the hemisphere $\{x_1 \geq 0\}$, and its variance and Dirichlet energy under π_A equal those of this restriction under $\pi_A(\cdot \mid x_1 \geq 0)$, with the Dirichlet form of the hemisphere that imposes no boundary condition at the equator $\{x_1 = 0\}$ (see the proof of Lemma 8.1). In this sense, the sector even in x_1 describes the dynamics folded onto the hemisphere. Similarly, the coordinate-even sector describes the dynamics folded onto the orthant $\{x_i \geq 0 \text{ for all } i\}$.

3.2 Equilibrium parameters and temperature regimes

Definition 3.3 (Lagrange multiplier and precisions). Let A be a real symmetric matrix with eigenvalues $\alpha_1 \geq \dots \geq \alpha_N$. The function $z \mapsto N^{-1} \sum_i (z - \alpha_i)^{-1}$ is continuous and strictly decreasing on (α_1, ∞) , and it tends to $+\infty$ as $z \downarrow \alpha_1$ and to 0 as $z \rightarrow \infty$. By the intermediate value theorem, there exists a unique $z > \alpha_1$ such that

$$\frac{1}{N} \sum_{i=1}^N \frac{1}{z - \alpha_i} = 1. \quad (3.10)$$

We call z the Lagrange multiplier of A . We define the precisions p_i , the variances q_i , the matrix Q , the smallest precision d , the largest precision s , the ratio κ , and the half-precisions r_i by

$$p_i = z - \alpha_i > 0, \quad q_i = \frac{1}{p_i}, \quad Q = (zI - A)^{-1}, \quad d = p_1, \quad s = p_N, \quad \kappa = \frac{s}{d}, \quad r_i = \frac{p_i}{2}. \quad (3.11)$$

By (3.10), $\sum_i q_i = \text{Tr } Q = N$. We call this identity the Gaussian covariance trace constraint, or briefly the trace condition.

For the model, $A = cW_N$ and $c \geq 0$, we write $z_N = z_N(c)$ and d_N, s_N for the corresponding quantities, so that

$$\begin{aligned} \frac{1}{N} \sum_{i=1}^N \frac{1}{z_N - c\lambda_i} &= 1, & p_i &= z_N - c\lambda_i, & q_i &= \frac{1}{z_N - c\lambda_i}, \\ d_N &= z_N - c\lambda_1, & s_N &= z_N - c\lambda_N. \end{aligned} \quad (3.12)$$

At $c = 0$, we have $z_N = p_i = q_i = d_N = s_N = 1$. With $z = z_N$ in (3.9), the half-precisions $r_i = p_i/2$ are the coefficients of the quadratic potential. These names are used throughout the paper: the a_i are the scaled eigenvalue gaps, the $r_i = p_i/2$ are the half-precisions, and in the critical window the limits of $N^{2/3}r_i$ are the limiting scaled half-precisions R_i of Section 4.5. The unnormalized process of Section 10 is built from a potential $\sum_i \rho_i x_i^2$, where the quadratic potential coefficients ρ_i are either r_i or a_i . Its coordinate Y_i then has drift rate $2\rho_i$, that is, $2r_i = p_i$ or $2a_i$.

Remark 3.4 (Gaussian representation). The names in Definition 3.3 come from the following representation. For any $z' > \alpha_1$, let Y be a centered Gaussian vector whose coordinates in an eigenbasis of A are independent with variances $(z' - \alpha_i)^{-1}$. Its density is proportional to $\exp(-z'|y|^2/2 + y^T A y/2)$. Here conditioning on $|Y|^2 = N$ means the weakly continuous version given by the coarea formula: its density with respect to the surface measure of S_N is proportional to the density of Y restricted to S_N , since the coarea factor $|\nabla|y|^2|^{-1} = (2\sqrt{N})^{-1}$ is constant there. The term $-z'|y|^2/2$ is also constant on S_N , and this conditional law is π_A ; see also Section 5. The choice $z' = z$ in (3.10) makes the unconditioned mean squared norm $\mathbb{E}|Y|^2 = \sum_i q_i$ equal to N , and z plays the role of a Lagrange multiplier for this constraint. This is the device used by Kosterlitz, Thouless, and Jones [97] to solve the statics of the model. At the critical temperature $c = 1$, this representation is used by Du and Huang [62, Section 1.2, (1.7)–(1.8)]. Two cautions apply. The coordinates of a sample from π_A are not independent, and in general their variances are not exactly q_i . Also, conditioning Y on $|Y|^2 = N$ is different from normalizing it radially: the law of $\sqrt{N}Y/|Y|$ has density proportional to $(\sum_i x_i^2/q_i)^{-N/2}$ with respect to ν_N , which is not π_A unless all q_i are equal. In the isotropic case both constructions produce ν_N . In particular, if g is a standard Gaussian vector in $\mathbb{R}^N$, then $\sqrt{N}g/|g|$ has law ν_N .

The distance between two Gaussian laws conditioned on S_N is governed by the following quantity, which we use for the high-temperature profiles.

Definition 3.5 (Information). Let $q = (q_i)_{i=1}^n$ and $v = (v_i)_{i=1}^n$ be sequences of positive numbers with $\sum_i q_i = \sum_i v_i$, that is, the diagonals of two diagonal covariance matrices with the same trace. We call

$$\mathcal{I}(q, v) = \frac{1}{2} \sum_{i=1}^n \left(\frac{v_i}{q_i} - 1 \right)^2 \quad (3.13)$$

the *information of v relative to q* . For a real symmetric matrix A with variances q_i as in Definition 3.3, and for $t \geq 0$, we set $\mathcal{I}_N(t) = \mathcal{I}(q, v(t))$, where $v(t)$ is the covariance evolution of A , defined in Definition 7.5. For the model, $A = cW_N$.

In terms of the relative errors $e_i = v_i/q_i - 1$, we have $\mathcal{I}(q, v) = \frac{1}{2} \sum_i e_i^2$. The information is not symmetric in q and v , and it is not a Fisher information; in this paper, Fisher information and relative Fisher information always refer to integrated squared logarithmic derivatives of densities (Section 11). The name comes from Le Cam's theory: when the relative errors are small and no coordinate carries a macroscopic share of the variance, the log-likelihood ratio between the conditioned Gaussian laws with variances v and q is approximately Gaussian with variance $\mathcal{I}(q, v)$ (Section 6), as in a Gaussian shift experiment with information $\mathcal{I}(q, v)$ (Remark 3.10).

Temperature regimes. The main results concern the following regimes and parameters. Temperature sequences need not be monotone.

- *Infinite-temperature crossover:* $c_N \rightarrow 0$, with $J_N = Nc_N^2$.
- *Fixed high temperature:* $0 < c < 1$ fixed. The smallest precision satisfies $d_N \rightarrow (1 - c)^2$ (Theorem 4.13).
- *Approach to the critical temperature from above:* $c_N = 1 - \delta_N$ with $\delta_N \in (0, 1)$, $\delta_N \rightarrow 0$, and $M_N = N\delta_N^3 \rightarrow \infty$. We write $\ell_N = \log(2 + M_N)$.
- *Critical window:* $c_N = 1 + \sigma N^{-1/3}$ with $\sigma \in \mathbb{R}$ fixed, and times $t = N^{2/3}s$ with $s > 0$ fixed. The half-precisions $r_i = p_i/2$ are used, and the random limiting objects are defined in Section 4.5.
- *Approach to the critical temperature from below:* $c_N = 1 + \epsilon_N$ with $\epsilon_N > 0$, $\epsilon_N \rightarrow 0$, and $N\epsilon_N^3 \rightarrow \infty$.
- *Fixed low temperature:* $c > 1$ fixed.

In the two low-temperature regimes we use the form (3.8) and the parameters

$$\theta_N = 1 - \frac{1}{N} \sum_{i=2}^N \frac{1}{2a_i}, \quad R_N = Na_2, \quad m_N = \theta_N R_N. \quad (3.14)$$

The quantity $N(1 - \theta_N)$ is the mean squared norm of a centered Gaussian vector $(Z_i)_{i \geq 2}$ with independent coordinates of variances $(2a_i)^{-1}$, and $N\theta_N$ is the deficit of this mean with respect to N . We call θ_N the top fraction. It is a Gaussian variance deficit, defined through an unconditioned Gaussian mean. For general positive $a_2, \dots, a_N$, it satisfies $\theta_N < 1$ but may be negative, and at finite N it need not equal $\pi_{N,c}(x_1^2)/N$. Only asymptotically does it describe the squared projection x_1^2/N of $x/\sqrt{N}$ onto the top eigenvector under the Gibbs measure: in either low-temperature regime, Corollary 16.10 shows that for every fixed $\gamma > 0$,

$$\pi_{N,c_N} \left(\left| \frac{x_1^2}{N\theta_N} - 1 \right| > \gamma \right) \rightarrow 0 \quad \text{in probability,}$$

where the left side is set to 0 on the event $\{\theta_N \leq 0\}$, whose probability tends to zero. At fixed $c > 1$, we have $\theta_N \rightarrow \theta = 1 - 1/c > 0$ in probability, while along the approach from below $\theta_N/\epsilon_N \rightarrow 1$, so that $\theta_N \rightarrow 0$ (Lemma 4.7). On the events of that lemma, $R_N \asymp N^{1/3}$, and therefore $m_N^3 \asymp N\theta_N^3$; along an approach from below, $N\theta_N^3$ plays the role of M_N . In the low-temperature estimates, we abbreviate $m_N = \theta_N R_N$ by m .

3.3 Distances, mixing times, and profiles

For probability measures μ, η on a measurable space, the total variation distance is

$$\|\mu - \eta\|_{\text{TV}} = \sup_B |\mu(B) - \eta(B)|,$$

the supremum over measurable sets. If $\mu \ll \eta$, the relative entropy is $\text{KL}(\mu\|\eta) = \int \log(d\mu/d\eta) d\mu \in [0, \infty]$; otherwise $\text{KL}(\mu\|\eta) = \infty$. For the model, we set

$$D_N(t) = \|\nu_N P_t - \pi_{N,c}\|_{\text{TV}}, \quad t_N(\varepsilon) = \inf\{t \geq 0 : D_N(t) \leq \varepsilon\} \quad (0 < \varepsilon < 1). \quad (3.15)$$

We call $t_N(\varepsilon)$ the mixing time at accuracy ε . Ratios of mixing times are set to 1 when their denominator is zero. Here the initial point has law ν_N and is independent of W_N , and both the initial point and the Brownian motion are integrated out, so that $D_N(t)$ and $t_N(\varepsilon)$ are random only through W_N . When several temperatures are considered at once, they share the same matrix W_N . We never couple the diffusions at different temperatures.

The initial point is integrated out before the distance is taken: $D_N(t)$ is the distance of the averaged law $\nu_N P_t$ from equilibrium. A formulation in which an initial point is sampled once from the Gibbs measure and held fixed would instead concern the Gibbs mass of initial points x for which $\|\delta_x P_t - \pi_{N,c}\|_{\text{TV}}$ exceeds a threshold. Our results do not address that question; see the discussion in the introduction.

We collect standard facts about these distances.

Lemma 3.6. *Let μ, η be probability measures on a measurable space E .*

1. *If μ and η have densities f, g with respect to a σ -finite measure λ , then $\|\mu - \eta\|_{\text{TV}} = \int (f - g)_+ d\lambda$, which is half of $\int |f - g| d\lambda$. Moreover $\|\mu - \eta\|_{\text{TV}} = \sup\{|\mu(\varphi) - \eta(\varphi)| : 0 \leq \varphi \leq 1\}$.*
2. *If K is a Markov kernel from E to another measurable space, then $\|\mu K - \eta K\|_{\text{TV}} \leq \|\mu - \eta\|_{\text{TV}}$ and $\text{KL}(\mu K\|\eta K) \leq \text{KL}(\mu\|\eta)$. In particular, both inequalities hold for the push-forwards under a measurable map.*
3. *For the model, $t \mapsto D_N(t)$ is nonincreasing.*
4. (Pinsker's inequality) $\|\mu - \eta\|_{\text{TV}} \leq \sqrt{\text{KL}(\mu\|\eta)/2}$.
5. *If $\mu = \mu_1 \otimes \mu_2$ and $\eta = \eta_1 \otimes \eta_2$, then $\|\mu - \eta\|_{\text{TV}} \leq \|\mu_1 - \eta_1\|_{\text{TV}} + \|\mu_2 - \eta_2\|_{\text{TV}}$.*
6. *For $m \in \mathbb{R}$, $\|\mathcal{N}(m, 1) - \mathcal{N}(0, 1)\|_{\text{TV}} = 2\Phi(|m|/2) - 1$, where Φ is the standard normal distribution function.*

Proof. For the first part, let $B = \{f > g\}$. For every measurable B' , we have $\mu(B') - \eta(B') \leq \int_{B' \cap B} (f - g) d\lambda \leq \mu(B) - \eta(B)$, and similarly $\eta(B') - \mu(B') \leq \eta(B^c) - \mu(B^c) = \mu(B) - \eta(B)$. We conclude that $\|\mu - \eta\|_{\text{TV}} = \int (f - g)_+ d\lambda$, which is half of $\int |f - g| d\lambda$ because $\int (f - g) d\lambda = 0$. The same argument bounds $\int \varphi(f - g) d\lambda$ for $0 \leq \varphi \leq 1$, which proves the last claim. For the second part, observe that for $0 \leq \varphi \leq 1$, the function $K\varphi$ also takes values in $[0, 1]$, and $(\mu K)(\varphi) - (\eta K)(\varphi) = \mu(K\varphi) - \eta(K\varphi)$. The claim follows from the last statement of the first part, applied to μ, η and $\lambda = \mu + \eta$. The inequality for relative entropy is the data processing inequality [147, Theorem 9], with order one. The third part follows from the second. Indeed, $\pi_{N,c} P_s = \pi_{N,c}$, and the second part implies that $D_N(t + s) = \|(\nu_N P_t) P_s - \pi_{N,c} P_s\|_{\text{TV}} \leq D_N(t)$. The fourth part is Pinsker's inequality [147,

Theorem 31], whose variational distance is twice ours. For the fifth part, by the triangle inequality, we have $\|\mu - \eta\|_{\text{TV}} \leq \|\mu_1 \otimes \mu_2 - \eta_1 \otimes \mu_2\|_{\text{TV}} + \|\eta_1 \otimes \mu_2 - \eta_1 \otimes \eta_2\|_{\text{TV}}$, and the second part, applied to the kernels $x \mapsto \delta_x \otimes \mu_2$ and $y \mapsto \eta_1 \otimes \delta_y$, bounds each term. For the sixth part, we may take $m > 0$. The densities of $\mathcal{N}(m, 1)$ and $\mathcal{N}(0, 1)$ cross only at $m/2$. By the first part, the distance is then $\Phi(m/2) - \Phi(-m/2) = 2\Phi(m/2) - 1$. $\square$

We now define what we mean by a mixing profile, by a cutoff profile, and by cutoff. The centers, windows, and time scales below depend on the realized spectrum, and they are random variables.

Definition 3.7 (Mixing profiles, cutoff profiles, and cutoff). A *mixing profile* of D_N is a limit of the functions $u \mapsto D_N(T_N + w_N u)$, for some centers $T_N \geq 0$ and windows $w_N > 0$, or of the functions $s \mapsto D_N(b_N s)$, for some time scale $b_N > 0$, in probability or in distribution, uniformly on compact subsets of the domain of the limit. A mixing profile may be random.

Let $P : \mathbb{R} \rightarrow (0, 1)$ be continuous and strictly decreasing, with $P(u) \rightarrow 1$ as $u \rightarrow -\infty$ and $P(u) \rightarrow 0$ as $u \rightarrow \infty$. Let $T_N, w_N > 0$ be random variables. We say that D_N has *cutoff profile* P with center T_N and window w_N if $w_N/T_N \rightarrow 0$ in probability and, for every compact $K \subset \mathbb{R}$,

$$\sup_{u \in K} |D_N(T_N + w_N u) - P(u)| \rightarrow 0 \quad \text{in probability,}$$

where the supremum is restricted to u with $T_N + w_N u \geq 0$. We say that there is *cutoff* if, for every $\varepsilon \in (0, 1/2)$, we have $t_N(1 - \varepsilon) > 0$ with probability tending to one and $t_N(\varepsilon)/t_N(1 - \varepsilon) \rightarrow 1$ in probability.

A cutoff profile is a mixing profile. The critical mixing profiles F_σ of Section 19 are limits in distribution of $s \mapsto D_N(N^{2/3}s)$, uniformly on compact subsets of $(0, \infty)$. They are random, and they are not cutoff profiles. This definition of cutoff refers to the fixed initial law ν_N and not to the worst initial point. For background on cutoff and profiles, see [105, Chapter 18] and [56]. A cutoff profile implies cutoff by Corollary 3.12 below.

3.4 Disorder quantifiers and constants

Our estimates for the dynamics are deterministic statements about a fixed matrix whose spectrum satisfies explicit hypotheses. The random matrix enters only through the probability that these hypotheses hold. We use the following convention throughout.

Eventual statements. A statement holds *for all sufficiently large N* , or *eventually*, if there exists N_0 such that it holds for all $N \geq N_0$. The threshold N_0 may depend on c (or on the sequence (c_N)), on the confidence η , and on the fixed parameters named in the

statement, including fixed time anchors. It never depends on W_N , on the initial point, or on a varying time over which uniformity is claimed.

Fixed-confidence events. A statement holds *at every fixed confidence* if, for every $\eta \in (0, 1)$, there exist constants depending only on η and on the fixed parameters of the statement, and events $\mathcal{G}_N^\eta$ with $\liminf_{N \rightarrow \infty} \mathbb{P}(\mathcal{G}_N^\eta) \geq 1 - \eta$, such that the statement holds on $\mathcal{G}_N^\eta$ with these constants for all sufficiently large N . The constants are chosen before $N \rightarrow \infty$, and the confidence η is sent to zero last. Such estimates say nothing about exponentially small failure probabilities, and the events $\mathcal{G}_N^\eta$ are not a single family of events of probability tending to one with fixed constants.

Probability modes. Unless said otherwise, “in probability” refers to the law of W_N , and so do $X_N = O_{\mathbb{P}}(a_N)$ (that is, (X_N/a_N) is tight) and $X_N = o_{\mathbb{P}}(a_N)$ (that is, $X_N/a_N \rightarrow 0$ in probability). Probabilities over the initial point and the noise are written explicitly when they occur. For fixed W_N , they are deterministic quantities. Convergence in distribution, written $\Rightarrow$, is used only where a random limit appears, as in the critical window. “With probability tending to one” means with probability over W_N tending to one as $N \rightarrow \infty$.

Power exponents. For a positive sequence (a_N) and $r \in \mathbb{R}$, we write $a_N = N^{r+o(1)}$ if, for every fixed $\eta > 0$, we have $N^{r-\eta} \leq a_N \leq N^{r+\eta}$ for all sufficiently large N , with a threshold that may depend on η . For nonnegative random variables U_N , we write $U_N = N^{r+o_{\mathbb{P}}(1)}$ if $\mathbb{P}\{N^{r-\eta} \leq U_N \leq N^{r+\eta}\} \rightarrow 1$ for every fixed $\eta > 0$.

Comparisons. We write $X \lesssim Y$ if $X \leq CY$ for a positive constant C , and $X \asymp Y$ if $X \lesssim Y \lesssim X$, with constants as below. Subscripts indicate extra dependence, as in $\asymp_K$.

The next lemma records how a fixed-confidence statement is converted into convergence in probability. It is used in every application to the GOE without further comment.

Lemma 3.8 (Convergence in probability from fixed-confidence estimates). *Let X_N be real random variables. Suppose that for every $\eta \in (0, 1)$ there exist events $\mathcal{G}_N^\eta$ with $\liminf_N \mathbb{P}(\mathcal{G}_N^\eta) \geq 1 - \eta$ such that $\mathbb{P}(\mathcal{G}_N^\eta \cap \{|X_N| > \varepsilon\}) \rightarrow 0$ for every $\varepsilon > 0$. Then $X_N \rightarrow 0$ in probability. If instead, for every η , there exists a constant C_η with $|X_N| \leq C_\eta$ on $\mathcal{G}_N^\eta$ for all large N , then $X_N = O_{\mathbb{P}}(1)$.*

Proof. For every $\varepsilon > 0$ and $\eta \in (0, 1)$, we have $\limsup_N \mathbb{P}(|X_N| > \varepsilon) \leq \limsup_N \mathbb{P}(\mathcal{G}_N^\eta \cap \{|X_N| > \varepsilon\}) + \limsup_N \mathbb{P}((\mathcal{G}_N^\eta)^c) \leq \eta$. Let $\eta \rightarrow 0$. The second claim is proved in the same way. $\square$

Constants. We write $C, C', C_0, c_0, c_*, \dots$ for positive constants whose dependence is stated each time. Their values may change from line to line. The letter c alone is reserved for the rescaled inverse temperature $c = \sqrt{2}\beta$. Unless stated otherwise, constants do not depend on N or on W_N ; any dependence on the realized spectrum or on the Airy point process, as in the critical-window theorem, is stated explicitly.

3.5 The limiting profiles

Let Φ and ϕ be the standard normal distribution function and density. Throughout, G, G' denote standard Gaussian random variables and C a standard Cauchy random variable, all independent. We define

$$\Psi(u) = 2\Phi\left(\frac{e^{-2u}}{2}\right) - 1, \quad u_\varepsilon = -\frac{1}{2}\log\left[2\Phi^{-1}\left(\frac{1+\varepsilon}{2}\right)\right] \quad (0 < \varepsilon < 1), \quad (3.16)$$

and, for $u \in \mathbb{R}$ and $\theta > 0$,

$$\begin{aligned} \mathcal{V}(u) &= \|\text{Law}(G + e^{-2u}C) - \text{Law}(G)\|_{\text{TV}}, \\ \mathcal{V}_\theta(u) &= \|\text{Law}(G + \sqrt{2\theta}e^{-2u}C) - \text{Law}(G)\|_{\text{TV}}, \end{aligned} \quad (3.17)$$

where $\text{Law}(\cdot)$ denotes the law of a random variable. For $h > 0$, let K_h be the Ornstein–Uhlenbeck transition kernel

$$K_h(x, \cdot) = \text{Law}(e^{-2h}x + \sqrt{1 - e^{-4h}}G'), \quad (3.18)$$

which is the transition kernel at time h of $dZ_h = -2Z_h dh + 2dB_h$. Its density $y \mapsto (1 - e^{-4h})^{-1/2}\phi((y - e^{-2h}x)(1 - e^{-4h})^{-1/2})$ is strictly positive, and $\mathcal{N}(0, 1)K_h = \mathcal{N}(0, 1)$.

Lemma 3.9 (Properties of the limiting profiles). *1. We have $\Psi(u) = \|\mathcal{N}(e^{-2u}, 1) - \mathcal{N}(0, 1)\|_{\text{TV}}$. The function Ψ is continuous and strictly decreasing on $\mathbb{R}$, with $\Psi(u) \rightarrow 1$ as $u \rightarrow -\infty$ and $\Psi(u) \rightarrow 0$ as $u \rightarrow \infty$, and $\Psi^{-1}(\varepsilon) = u_\varepsilon$ for $0 < \varepsilon < 1$.*

2. The function $\mathcal{V}$ is continuous and strictly decreasing on $\mathbb{R}$, with $\mathcal{V}(u) \rightarrow 1$ as $u \rightarrow -\infty$ and $\mathcal{V}(u) \rightarrow 0$ as $u \rightarrow \infty$. Every $\varepsilon \in (0, 1)$ has a unique preimage $\mathcal{V}^{-1}(\varepsilon)$, which depends continuously on ε . Moreover, for $z > 0$ and $h > 0$, we have $\text{Law}(G + zC)K_h = \text{Law}(G + e^{-2h}zC)$.

3. For every $\theta > 0$ and $u \in \mathbb{R}$, we have $\mathcal{V}_\theta(u) = \mathcal{V}(u - \log(2\theta)/4)$. In particular, $\mathcal{V}_\theta$ has the properties of the second part, and $\mathcal{V}_\theta^{-1}(\varepsilon) = \mathcal{V}^{-1}(\varepsilon) + \log(2\theta)/4$.

Proof. The first part follows from the last part of Lemma 3.6 with $m = e^{-2u}$, since $m \mapsto 2\Phi(m/2) - 1$ is continuous and strictly increasing from 0 to 1 on $[0, \infty)$ and $u \mapsto e^{-2u}$ is a decreasing bijection of $\mathbb{R}$ onto $(0, \infty)$. Solving $2\Phi(e^{-2u}/2) - 1 = \varepsilon$ for u yields $u = u_\varepsilon$.

For the second part, write $T(z) = \|\text{Law}(G + zC) - \text{Law}(G)\|_{\text{TV}}$ for $z \geq 0$, so that $\mathcal{V}(u) = T(e^{-2u})$. The law of $G + zC$ has the density $f_z(y) = \mathbb{E}\phi(y - zC)$, which is continuous and strictly positive. If $z_n \rightarrow z \geq 0$, then $f_{z_n}(y) \rightarrow f_z(y)$ for all y by bounded convergence, with $f_0 = \phi$, and Scheffé's lemma [129, pp. 435–437] implies that $f_{z_n} \rightarrow f_z$ in L^1 ; equivalently, apply dominated convergence to $\min\{f_{z_n}, f_z\}$ and use that both densities integrate to one. By the first part of Lemma 3.6, T is continuous on $[0, \infty)$ and $T(0) = 0$. This proves that $\mathcal{V}$ is continuous and that $\mathcal{V}(u) \rightarrow 0$ as $u \rightarrow \infty$. For the other limit, fix

$\ell > 0$ and the interval $I = [-\ell, \ell]$. Conditioning on G and using that the density of zC is bounded by $1/(\pi z)$, we obtain $\mathbb{P}(G + zC \in I) \leq 2\ell/(\pi z)$. Then

$$\mathsf{T}(z) \geq \mathbb{P}(G \in I) - \mathbb{P}(G + zC \in I) \geq 1 - \mathbb{P}(|G| > \ell) - \frac{2\ell}{\pi z}.$$

Letting $z \rightarrow \infty$ and then $\ell \rightarrow \infty$, we obtain $\mathsf{T}(z) \rightarrow 1$, that is, $\mathcal{V}(u) \rightarrow 1$ as $u \rightarrow -\infty$.

To prove strict monotonicity, we first note that $e^{-2h}G + \sqrt{1 - e^{-4h}}G'$ is a standard Gaussian independent of C . Then

$$\text{Law}(G + zC)K_h = \text{Law}(e^{-2h}G + \sqrt{1 - e^{-4h}}G' + e^{-2h}zC) = \text{Law}(G + e^{-2h}zC),$$

and $\text{Law}(G)K_h = \text{Law}(G)$. Fix $z > 0$ and $h > 0$, and let $\sigma = \text{Law}(G + zC) - \text{Law}(G)$. We have $\sigma \neq 0$, because the characteristic function of $G + zC$ is $e^{-\xi^2/2 - z|\xi|}$. Let $\sigma = \sigma^+ - \sigma^-$ be its Jordan decomposition. Since $\sigma(\mathbb{R}) = 0$, both σ^+ and σ^- are nonzero, with total mass $\mathsf{T}(z)$. The measures $\sigma^\pm K_h$ have densities $a^\pm(y) = \int k_h(x, y) \sigma^\pm(dx)$, where k_h is the strictly positive density of $K_h(x, \cdot)$. In particular, $a^+(y) > 0$ and $a^-(y) > 0$ for all y . By the first part of Lemma 3.6,

$$\mathsf{T}(e^{-2h}z) = \|\sigma K_h\|_{\text{TV}} = \frac{1}{2} \int |a^+ - a^-| dy = \frac{1}{2} \int (a^+ + a^-) dy - \int \min(a^+, a^-) dy < \mathsf{T}(z).$$

In the last step, we used $\int (a^+ + a^-) dy = 2\mathsf{T}(z)$ and $\int \min(a^+, a^-) dy > 0$. We conclude that T is strictly increasing on $(0, \infty)$, and that $\mathcal{V}$ is strictly decreasing. A continuous strictly decreasing bijection of $\mathbb{R}$ onto $(0, 1)$ has a continuous inverse, which proves the second part.

For the third part, note that $\sqrt{2\theta}e^{-2u} = \exp(-2(u - \log(2\theta)/4))$, so that $\mathcal{V}_\theta(u) = \mathcal{V}(u - \log(2\theta)/4)$. The remaining claims follow from the second part. $\square$

Remark 3.10 (The Gaussian and Gaussian–Cauchy profiles). The function $\Psi(u) = 2\Phi(\sqrt{I}/2) - 1$ with $I = e^{-4u}$ is the total variation distance in a Gaussian shift experiment with information I , the limit experiment in Le Cam’s theory of local asymptotic normality [146, Chapters 6–7]. As a cutoff profile, Ψ appears in the work of Diaconis, Graham, and Morrison on the random walk on the hypercube [59, Theorem 1]: at time $(n \log n)/4 + un$ the distance tends to $\text{erf}(e^{-2u}/\sqrt{8}) = \Psi(u)$. Gaussian profiles of this kind also arise for Ornstein–Uhlenbeck processes [35, Lemma 3.4]. The law of $G + zC$ is the convolution of a Gaussian and a Cauchy law, and its density is a Voigt profile [149], [117, §7.19]. We use the total variation representation (3.17) for $\mathcal{V}$. In the low-temperature sections, the variable C arises as the ratio of two independent centered Gaussian variables.

The next lemma converts pointwise convergence of a nonincreasing distance to a profile into locally uniform convergence and convergence of the mixing times.

Lemma 3.11 (Monotone profiles and mixing times). *Let $D_n : \Omega \times [0, \infty) \rightarrow [0, 1]$ be jointly measurable, with nonincreasing sample functions $t \mapsto D_n(t)$ satisfying $\lim_{t \rightarrow \infty} D_n(t) = 0$ almost surely, and let $b_n, w_n > 0$ be random variables with $b_n/w_n \rightarrow \infty$ in probability. Let $V : \mathbb{R} \rightarrow (0, 1)$ be continuous and strictly decreasing, with $V(u) \rightarrow 1$ as $u \rightarrow -\infty$ and $V(u) \rightarrow 0$ as $u \rightarrow \infty$. Suppose that for each fixed $u \in \mathbb{R}$ and $\varepsilon > 0$,*

$$\mathbb{P}(b_n + w_n u \geq 0, |D_n(b_n + w_n u) - V(u)| > \varepsilon) \rightarrow 0. \quad (3.19)$$

Then the following hold, with all limits in probability.

1. *For each compact interval $K \subset \mathbb{R}$, we have $\sup_{u \in K} |D_n(b_n + w_n u) - V(u)| \rightarrow 0$, where the supremum is over $u \in K$ with $b_n + w_n u \geq 0$; with probability tending to one, this is all of K .*
2. *Let $t_n(\varepsilon) = \inf\{t \geq 0 : D_n(t) \leq \varepsilon\}$. For each compact interval $I \subset (0, 1)$,*

$$\sup_{\varepsilon \in I} \left| \frac{t_n(\varepsilon) - b_n}{w_n} - V^{-1}(\varepsilon) \right| \rightarrow 0.$$

3. *For every sequence of random variables $\omega_n \rightarrow \infty$ in probability, $\inf\{D_n(t) : 0 \leq t \leq b_n - w_n \omega_n\} \rightarrow 1$, where the infimum over the empty set is 1.*

If the hypotheses hold for deterministic D_n, b_n, w_n , with ordinary convergence in place of convergence in probability, then the conclusions hold with ordinary convergence.

Proof. For an empty K , the error supremum in the first part is zero. In its proof assume that K is nonempty. For each fixed u , the event $\{b_n + w_n u \geq 0\}$ has probability tending to one, because $b_n/w_n \rightarrow \infty$ in probability. We work on these events below. The restricted suprema in the first part are measurable. To see this, write $K = [a, b]$ and $\ell_n = \max\{a, -b_n/w_n\}$. If $\ell_n > b$, set the supremum over the empty interval to 0. Otherwise, use a fixed countable dense subset of K containing its endpoints, retain the points u with $b_n + w_n u \geq 0$, and also include the random left endpoint ℓ_n . At every interior point, monotonicity places $D_n(b_n + w_n u)$ between its one-sided limits; continuity of V and convexity of $x \mapsto |x - V(u)|$ therefore show that these values have the same supremum as the whole restricted interval. The countable supremum and the additional value $|D_n(b_n + w_n \ell_n) - V(\ell_n)|$ are measurable by joint measurability.

For the first part, fix $\delta > 0$ and let $u_0 < u_1 < \dots < u_k$ be a mesh with $K \subset [u_0, u_k]$ and $V(u_j) - V(u_{j+1}) \leq \delta$ for all j , which is possible because V is uniformly continuous on $[u_0, u_k]$. Let $e_n = \max_j |D_n(b_n + w_n u_j) - V(u_j)|$. Using (3.19) and a union bound over the finitely many mesh points, we find that $e_n \rightarrow 0$ in probability. For $u \in [u_j, u_{j+1}]$, since $w_n > 0$ and D_n is nonincreasing, we have

$$\begin{aligned} V(u) - \delta - e_n &\leq V(u_{j+1}) - e_n \leq D_n(b_n + w_n u_{j+1}) \leq D_n(b_n + w_n u) \\ &\leq D_n(b_n + w_n u_j) \leq V(u_j) + e_n \leq V(u) + \delta + e_n. \end{aligned}$$

Taking the supremum over $u \in K$, we find $\sup_{u \in K} |D_n(b_n + w_n u) - V(u)| \leq \delta + e_n$, and δ is arbitrary.

For the second part, the conclusion for $I = \emptyset$ is immediate, with the error supremum over the empty set equal to zero. Assume henceforth that I is nonempty. Write $v = V^{-1}$, which is continuous on $(0, 1)$. Fix $h > 0$. The functions $\varepsilon \mapsto V(v(\varepsilon) - h) - \varepsilon$ and $\varepsilon \mapsto \varepsilon - V(v(\varepsilon) + h)$ are continuous and positive on the compact set I . Let $\gamma_h > 0$ be a common lower bound for them on I . Let $K_h = [\min_I v - h, \max_I v + h]$. By the first part, with probability tending to one we have $\sup_{u \in K_h} |D_n(b_n + w_n u) - V(u)| < \gamma_h$, and then, simultaneously for all $\varepsilon \in I$,

$$D_n(b_n + w_n(v(\varepsilon) - h)) > \varepsilon, \quad D_n(b_n + w_n(v(\varepsilon) + h)) < \varepsilon.$$

The second inequality gives $t_n(\varepsilon) \leq b_n + w_n(v(\varepsilon) + h)$. Because D_n is nonincreasing, the first inequality implies that $D_n(t) > \varepsilon$ for all $t \leq b_n + w_n(v(\varepsilon) - h)$, and then $t_n(\varepsilon) \geq b_n + w_n(v(\varepsilon) - h)$. We have shown that $|(t_n(\varepsilon) - b_n)/w_n - v(\varepsilon)| \leq h$ for all $\varepsilon \in I$, and h is arbitrary. No continuity of D_n is needed.

For the third part, fix $\varepsilon > 0$ and let $A > 0$ be such that $V(-A) > 1 - \varepsilon/2$. On the event $\{\omega_n \geq A\}$, which has probability tending to one, the infimum is either 1 or equal to $D_n(b_n - w_n \omega_n) \geq D_n(b_n - w_n A)$, since D_n is nonincreasing. By (3.19), $D_n(b_n - w_n A) > 1 - \varepsilon$ with probability tending to one.

The deterministic version follows from the same inequalities. $\square$

Corollary 3.12 (Cutoff profiles imply cutoff). *If D_N has cutoff profile P with center T_N and window w_N in the sense of Definition 3.7, then $(t_N(\varepsilon) - T_N)/w_N \rightarrow P^{-1}(\varepsilon)$ in probability, uniformly for ε in compact subsets of $(0, 1)$, and there is cutoff.*

Proof. The definition of a cutoff profile includes the hypotheses of Lemma 3.11 with $b_n = T_N$ and $V = P$, since D_N is nonincreasing by Lemma 3.6. It is also jointly measurable in (W_N, t) , because it is measurable in W_N for each t and continuous in t : the density $P_t h$ below is continuous in t in $L^2(\pi_{N,c})$. Also $D_N(t) \rightarrow 0$ as $t \rightarrow \infty$. Indeed, ν_N has a bounded density h with respect to $\pi_{N,c}$, the density of $\nu_N P_t$ is $P_t h$ by (3.4), and $\|P_t h - 1\|_{L^2(\pi_{N,c})} \rightarrow 0$ because the spectrum of $-L$ on the compact connected manifold S_N is discrete with kernel spanned by the constants. Then $2D_N(t) \leq \|P_t h - 1\|_{L^2(\pi_{N,c})}$ by the Cauchy–Schwarz inequality. The first claim then follows from its second part. Fix $\varepsilon \in (0, 1/2)$ and write $t_N(\varepsilon) = T_N(1 + (w_N/T_N)(P^{-1}(\varepsilon) + o_{\mathbb{P}}(1)))$, and similarly for $1 - \varepsilon$. Since $w_N/T_N \rightarrow 0$ in probability, we have $t_N(1 - \varepsilon) > 0$ with probability tending to one and $t_N(\varepsilon)/t_N(1 - \varepsilon) \rightarrow 1$ in probability. $\square$

Random-matrix inputs and the equilibrium Lagrange multiplier

This section collects the properties of the GOE spectrum used in the paper, and it determines the Lagrange multiplier of Definition 3.3 in each temperature regime. Section 4.1 states the published random-matrix results that we use, in our normalization. These are rigidity and convergence at the spectral edge (Lemma 4.2), simultaneous control of the eigenvalue locations (Lemma 4.3), a localized estimate for the Stieltjes transform (Lemma 4.4), and the regularized inverse-gap statistic (Lemma 4.5). The proofs below give the precise published estimates used to derive the stated bounds. Section 4.2 proves two-sided bounds on all eigenvalue gaps below the top eigenvalue at every fixed confidence (Lemma 4.6), and derives from them the spectral hypotheses used at low temperature (Lemma 4.7). Section 4.3 proves a law of large numbers for the eigenvalues near the edge at every scale between $N^{-2/3}$ and 1 (Theorem 4.8), together with its version for the gaps below the top eigenvalue (Corollary 4.9). Section 4.4 determines the Lagrange multiplier at fixed high temperature and along every approach $c_N = 1 - \delta_N$ with $N\delta_N^3 \rightarrow \infty$ (Proposition 4.11 and Theorem 4.13). Finally, Section 4.5 treats the critical window. It proves almost-sure properties of the reflected GOE Airy point process and of the limiting equation for the Lagrange multiplier (Theorem 4.15), and the convergence of the scaled half-precisions together with uniform bounds on them (Lemma 4.17).

This section uses the definitions of Section 3. Theorems 4.8 and 4.13 are used at high temperature, in Sections 13 to 15, and Corollary 4.9 is used in Section 14. Lemma 4.7 supplies the spectral hypotheses in the approach to the critical temperature from below (Section 17). At a fixed low temperature, the proof of the main result of Section 16 uses instead Lemmas 4.2 and 4.6, through the law of large numbers for the inverse gaps of Lemma 16.6. The critical window uses Lemma 4.17 and the conditions (4.36), in Sections 8, 12, 18, and 19. Theorem 4.15 is used for the limits of the critical profile far from the critical temperature (Sections 20 and 21).

Notation. Throughout this section, $\lambda_1 \geq \dots \geq \lambda_N$ are the eigenvalues of W_N . For $E > \lambda_1$ and $E > 2$, respectively, we write

$$\mathfrak{m}_N(E) = \frac{1}{N} \sum_{i=1}^N \frac{1}{E - \lambda_i}, \quad \mathfrak{m}_{\text{sc}}(E) = \int \frac{\rho_{\text{sc}}(x)}{E - x} dx = \frac{E - \sqrt{E^2 - 4}}{2} \quad (4.1)$$

for the Stieltjes transforms of the empirical eigenvalue distribution and of the semicircle law. With this sign convention both are positive. The papers [69, 101, 102] use the opposite sign, which does not affect absolute differences. We let $\gamma_1 \geq \gamma_2 \geq \dots \geq \gamma_N$ be the semicircle quantiles, defined by $\int_{\gamma_i}^2 \rho_{\text{sc}}(x) dx = i/N$. We write

$$C_N(t) = \#\{i : \lambda_i \geq 2 - tN^{-2/3}\}, \quad \Delta_{i,N} = N^{2/3}(\lambda_1 - \lambda_i) \quad (1 \leq i \leq N),$$

for the number of eigenvalues within distance $tN^{-2/3}$ of the edge and for the scaled gaps below the top eigenvalue. The following elementary bounds on the quantiles are used several times.

Lemma 4.1 (Semicircle quantiles). *For $1 \leq i \leq N$, we have*

$$\left(\frac{3\pi i}{2N}\right)^{2/3} \leq 2 - \gamma_i \leq 7\left(\frac{i}{N}\right)^{2/3}.$$

Moreover, $2 - \gamma_i = (3\pi i/(2N))^{2/3}(1 + O((i/N)^{2/3}))$ uniformly for $1 \leq i \leq N/2$.

Proof. For $0 \leq v \leq 4$, we have $\rho_{\text{sc}}(2 - v) = (2\pi)^{-1}\sqrt{v(4 - v)}$. Using $\sqrt{4 - v} \leq 2$, we find that $i/N = \int_0^{2-\gamma_i} \rho_{\text{sc}}(2 - v) dv \leq 2(2 - \gamma_i)^{3/2}/(3\pi)$, which is the lower bound. For $i \leq N/2$, we have $\gamma_i \geq 0$ by symmetry of ρ_{sc} . Since $\sqrt{4 - v} \geq \sqrt{2}$ for $v \leq 2$, we then have $i/N \geq \sqrt{2}(2 - \gamma_i)^{3/2}/(3\pi)$ and $2 - \gamma_i \leq (3\pi/\sqrt{2})^{2/3}(i/N)^{2/3} \leq 7(i/N)^{2/3}$. For $i > N/2$, we use $2 - \gamma_i \leq 4 \leq 7(i/N)^{2/3}$. Finally, since $\sqrt{4 - v} = 2(1 + O(v))$, we have $\int_0^u \rho_{\text{sc}}(2 - v) dv = 2u^{3/2}(1 + O(u))/(3\pi)$ for $u \leq 2$. Setting $u = 2 - \gamma_i \leq 7(i/N)^{2/3}$ and solving for u , we obtain the last claim. $\square$

4.1 Published random-matrix inputs

Imported results and their normalization. Apart from standard tools of analysis and probability, the following are the only random-matrix inputs to the proofs of Theorems 1.1–1.4: eigenvalue rigidity; convergence at the edge to the reflected GOE Airy point process, including the GOE Tracy–Widom law; published counting estimates, used with rigidity to obtain simultaneous control of the eigenvalue locations; the localized Stieltjes-transform estimate; and the regularized inverse-gap statistic, including Du and Huang’s joint inverse-gap convergence in $\mathbb{R} \times \ell^2$. The versions collected in this subsection suffice for the downstream profile estimates; proofs also cite the sharper published estimates from which they are obtained. Du and Huang’s limiting equation $\mathcal{S}$ and their critical half-precision result at $\sigma = 0$ are identified where they occur in the critical-window statements below; the quantitative bounds and the extensions in σ are proved here. All statements use the GOE variances $1/N$ off the diagonal and $2/N$ on the diagonal, the reflected edge coordinates $N^{2/3}(2 - \lambda_i)$, and the Stieltjes-transform sign convention fixed above.

We record the published results that we use, in the normalization of Section 3. The random-matrix ensemble parameter of Ramírez, Rider, and Virág [121] equals 1 for the GOE and is unrelated to the inverse temperature β .

Lemma 4.2 (Rigidity and convergence at the edge [101, Theorem 2.1, (2.16)], [69, Theorem 2.2, (2.25)], [121, Theorem 1.1, (1.7), and Proposition 3.5]).

1. For every $\eta \in (0, 1)$ there exist $C > 0$ and N_0 such that, for $N \geq N_0$, the event

$$F_C = \left\{ \max_{1 \leq i \leq N} |\lambda_i - \gamma_i| \leq CN^{-5/12} \right\} \quad (4.2)$$

has probability at least $1 - \eta$.

2. There exists a random sequence $\chi_1 < \chi_2 < \dots$, the reflected GOE Airy point process, which almost surely is strictly increasing, has no finite accumulation point, and satisfies $\chi_i \rightarrow \infty$, such that for every fixed k ,

$$(N^{2/3}(2 - \lambda_i))_{i=1}^k \Longrightarrow (\chi_i)_{i=1}^k.$$

The law of $-\chi_1$ is the GOE Tracy–Widom law TW_1 . In particular, $N^{2/3}(\lambda_1 - 2)$ is tight, and for every fixed k , $(\Delta_{i,N})_{i=2}^k \Rightarrow (\Delta_i)_{i=2}^k$, where $\Delta_i = \chi_i - \chi_1 \in (0, \infty)$ almost surely.

3. Since $-W_N$ has the law of W_N , the same statements hold at the lower edge. In particular, $\lambda_N \rightarrow -2$ in probability, and $\max_i |\lambda_i| \leq 2 + CN^{-5/12}$ on F_C .

Used in: The high-temperature, low-temperature, and critical-window analyses, through the consequences below.

Proof. The first part is a consequence of the sharper published rigidity estimate [101, Theorem 2.1], based on [69, Theorem 2.2]. Choosing its exponent smaller than $1/4$ gives the stated fixed-confidence bound. The variance normalization in the latter reference applies to $\sqrt{N/(N+1)} W_N$; undoing this scaling changes eigenvalues by $O(N^{-1})$ on a bounded-spectrum event and is absorbed by $N^{-5/12}$. We use the sharper published estimates in a few proofs below, with explicit citations, but state only the bounds needed here. For the second part, the eigenvalues of $\sqrt{N} W_N$ have the law of the Hermite β -ensemble with $\beta = 1$, whose density is proportional to $\prod_{i < j} |\lambda_i - \lambda_j| \exp(-\sum_i \lambda_i^2/4)$. By [121, Theorem 1.1, (1.7)], the vector $(N^{1/6}(2\sqrt{N} - \sqrt{N}\lambda_i))_{i \leq k} = (N^{2/3}(2 - \lambda_i))_{i \leq k}$ converges in distribution to the k smallest eigenvalues $\Lambda_0 < \dots < \Lambda_{k-1}$ of the stochastic Airy operator $-d^2/dx^2 + x + 2b'(x)$ on $L^2(0, \infty)$ with Dirichlet boundary condition at 0. Here b is a standard Brownian motion, and b' is its derivative in the sense of distributions, that is, white noise. Since b is not differentiable, this is not an ordinary differential operator, and we use the definition of [121, Section 2]. There, the product fb' is defined through integration by parts, and λ is an eigenvalue if there exists a nonzero f with $f(0) = 0$ and $\int_0^\infty ((f')^2 + (1+x)f^2) dx < \infty$ such that $-f'' + xf + 2fb' = \lambda f$ in the

sense of distributions [121, Definition 2.1]. We set $\chi_i = \Lambda_{i-1}$. These eigenvalues are almost surely distinct, have no finite accumulation point, and tend to infinity [121, Proposition 3.5]. The stochastic Airy operator was proposed by Edelman and Sutton [64] on the basis of the tridiagonal models of Dumitriu and Edelman [63]. We note that the points χ_i form one random sequence, and that for every k the limit in the second part is the law of its first k coordinates. We write TW_1 for the law of $-\chi_1$. After a change of sign, the case $k = 1$ of the second part reads $N^{2/3}(\lambda_1 - 2) \Rightarrow -\chi_1$. When TW_1 appears as a random variable in a joint limit with other functions of the χ_i , it denotes $-\chi_1$ for this same sequence. The law TW_1 is the GOE Tracy–Widom distribution, whose distribution function is F_1 in [145, Section I and (53) in Section VII]. Indeed, F_1 is defined there as the limiting distribution function of the largest eigenvalue in edge scaling. Tracy and Widom normalize the GOE so that its off-diagonal entries have variance $1/2$, which in our notation is the matrix $\sqrt{N/2} W_N$. Their edge coordinate for its largest eigenvalue is $\sqrt{2} N^{1/6}(\sqrt{N/2} \lambda_1 - \sqrt{2N}) = N^{2/3}(\lambda_1 - 2)$, which converges in distribution to $-\chi_1$. The universality of this law for Wigner matrices is due to Soshnikov [136]. The statements about $\Delta_{i,N}$ follow from the continuous mapping theorem. The third part follows from the first two parts. $\square$

Lemma 4.3 (Simultaneous control of the eigenvalue locations). *For every $\eta \in (0, 1)$, there exist $C_\eta > 0$, an integer N_0 , and measurable events $\mathcal{H}_N^\eta$ with $\mathbb{P}(\mathcal{H}_N^\eta) \geq 1 - \eta$ for $N \geq N_0$, on which*

$$N^{2/3}|\lambda_i - \gamma_i| \leq C_\eta(1 + i)^{1/4} \quad (1 \leq i \leq N). \quad (4.3)$$

Used in: The fixed-confidence gap bounds, mesoscopic edge counts, and almost-sure Airy geometry below.

Proof. Write $q_{i,N} = N^{2/3}(2 - \gamma_i)$ and $a = 2/(3\pi)$. Choose a fixed $b > 0$ so small that $7b^{2/3} < 1/2$. By Lemma 4.1, uniformly for $i \leq bN^{2/5}$,

$$q_{i,N} \asymp i^{2/3}, \quad aq_{i,N}^{3/2} = i + O(i^{5/3}N^{-2/3}) = i + O(1).$$

Consequently, for all sufficiently large fixed K and all large N , the thresholds $s_\pm = q_{i,N} \pm i^{1/4}$, $K \leq i \leq bN^{2/5}$, lie in $[t_0, N^{4/15}]$ and satisfy

$$as_+^{3/2} - i \geq ci^{7/12}, \quad i - as_-^{3/2} \geq ci^{7/12},$$

where t_0 is any fixed sufficiently large constant. Here the exponent $7/12$ is the sum of $1/3$, from the derivative of $s^{3/2}$ at $s \asymp i^{2/3}$, and $1/4$. The published bounds [102, Proposition 6.5, (6.11)–(6.12)] give $|\text{EC}_N(s) - as^{3/2}| \leq C$ and $\text{Var}(C_N(s)) \leq C \log(2 + s)$ throughout this range. These counting estimates follow there from the Airy estimates of Soshnikov [137], the GUE estimates of Gustavsson [80], and the superposition identity

of Forrester and Rains [72]. If $N^{2/3}(2 - \lambda_i) < s_-$, then $C_N(s_-) \geq i$; if $N^{2/3}(2 - \lambda_i) > s_+$, then $C_N(s_+) < i$. Chebyshev's inequality therefore gives

$$\mathbb{P}(N^{2/3}|\lambda_i - \gamma_i| > i^{1/4}) \leq C \log(2+i)i^{-7/6} \quad (K \leq i \leq bN^{2/5}).$$

The series on the right is summable. Choose K , depending only on η , so that its sum over $i \geq K$ is at most $\eta/4$.

For the remaining indices $i > bN^{2/5}$, use the sharper rigidity estimate [101, Theorem 2.1], with exponent $1/10$, on an event whose probability tends to one. If $i \leq N/2$, it gives

$$N^{2/3}|\lambda_i - \gamma_i| \leq N^{1/10}i^{-1/3} \leq C_b i^{1/4},$$

since $i > bN^{2/5}$; if $i > N/2$, it gives the upper bound $N^{1/10} \leq C i^{1/4}$. Finally, the first $K-1$ scaled eigenvalues are tight by Lemma 4.2, and $q_{i,N} \rightarrow (3\pi i/2)^{2/3}$ for each fixed i . We can therefore choose C_η so that (4.3) holds simultaneously for these first indices with limiting failure probability at most $\eta/4$. Enlarge C_η to dominate the constants for the other indices. Intersecting the three events gives failure probability at most η for all sufficiently large N . $\square$

Lemma 4.4 (Localized Stieltjes transform estimate). *For every $\eta \in (0, 1)$ there are $U_\eta, C_\eta > 0$, an integer N_0 , and measurable events $\mathcal{R}_N^\eta$ of probability at least $1 - \eta$ for $N \geq N_0$, such that $\lambda_1 < 2 + U_\eta N^{-2/3}$ on these events and, for every deterministic $u \geq U_\eta$, with $E = 2 + uN^{-2/3}$,*

$$\mathbb{E} \left[\mathbf{1}_{\mathcal{R}_N^\eta} |\mathfrak{m}_N(E) - \mathfrak{m}_{\text{sc}}(E)| \right] \leq C_\eta N^{-1/3} \left(u^{-1/8} + N^{-1/3} u^{-1/2} + N^{-1/6} + u^{-1} \right). \quad (4.4)$$

The random variable inside the expectation is defined to be zero off $\mathcal{R}_N^\eta$; on this event $E > \lambda_1$, so the resolvent is defined.

Used in: The high-temperature multiplier estimates below.

Proof. Fix $\xi = 1/200$. The localized estimate of [101, Lemmas 2.2 and 2.6] supplies, on a common event of confidence $1 - \eta$, the bound $C_\eta \{N^{-1/3}(1 + \log^2 u)/u + N^{-1+\xi}\}$ for $U_\eta \leq u < N^{1/15-\xi}$, with U_η enlarged to put E strictly above the top eigenvalue. Since $(1 + \log^2 u)/u \leq Cu^{-1/8}$ and $N^{-1+\xi} \leq N^{-1/2}$, this is bounded by the right side of (4.4). Intersect this event with the sharper rigidity event of [101, Theorem 2.1], choosing both failure probabilities at most $\eta/2$. For $u \geq N^{1/15-\xi}$, that rigidity estimate and the semicircle-quantile quadrature give, on this same event,

$$|\mathfrak{m}_N(2 + uN^{-2/3}) - \mathfrak{m}_{\text{sc}}(2 + uN^{-2/3})| \leq CN^{-1/3} \frac{N^\xi}{u}.$$

To obtain this inequality, split the resolvent comparison at $i \asymp u^{3/2}$, use $2 - \gamma_i \asymp (i/N)^{2/3}$ below the midpoint and a fixed denominator above it, and add the quadrature error

$(N(E-2))^{-1}$; when $E \geq 3$, the fixed denominators only improve the bound. Here $N^\xi/u \leq u^{-1/8}$ for all large N , since $\xi < (7/8)(1/15 - \xi)$. Integration over the common event proves (4.4) on the remaining range. All event and threshold choices precede the choice of u . $\square$

Lemma 4.5 (Regularized inverse gaps [102, Theorem 6.1, (6.3)–(6.4)], [62, Propositions 1.3 and 4.1, (4.1)]). *Let $t_k = (3\pi k/2)^{2/3}$.*

1. *Almost surely, the limit*

$$\Xi = \lim_{k \rightarrow \infty} \left(\sum_{i=2}^k \frac{1}{\Delta_i} - \frac{2}{\pi} \sqrt{t_k} \right) \quad (4.5)$$

exists and is finite. We call Ξ the regularized inverse-gap statistic.

2. *Let $\Xi_N = \sum_{i=2}^N \Delta_{i,N}^{-1} - N^{1/3}$, and let $\mathbf{b}_N = (\Delta_{2,N}^{-1}, \dots, \Delta_{N,N}^{-1}, 0, 0, \dots) \in \ell^2$. Then*

$$(\Xi_N, \mathbf{b}_N) \implies (\Xi, (\Delta_i^{-1})_{i \geq 2}) \quad \text{in } \mathbb{R} \times \ell^2.$$

In particular, Ξ_N is tight.

3. *For every $c > 0$,*

$$\frac{1}{N} \sum_{i=2}^N \frac{1}{c(\lambda_1 - \lambda_i)} = \frac{1}{c} + \frac{\Xi_N}{cN^{1/3}}.$$

4. *If $c_N = 1 - \delta_N$ with $\delta_N \rightarrow 0$ and $M_N = N\delta_N^3 \rightarrow \infty$, then*

$$\frac{1}{\delta_N} \left(\frac{1}{N} \sum_{i=2}^N \frac{1}{c_N(\lambda_1 - \lambda_i)} - 1 \right) \longrightarrow 1 \quad \text{in probability.}$$

Used in: The high-temperature multiplier estimates, the low-temperature and varying low-temperature analyses, the critical process and profile, through the consequences below.

Proof. The first part, and the convergence in distribution of Ξ_N to Ξ , are due to Landon and Sosoë [102, Theorem 6.1, (6.3)–(6.4)]. Note that $2\sqrt{t_k}/\pi = \pi^{-1} \int_0^{t_k} x^{-1/2} dx$ and that the random variable in (6.3) there equals Ξ_N . The joint convergence in $\mathbb{R} \times \ell^2$ is [62, Proposition 4.1, (4.1)], whose inverse-gap vector has an additional first coordinate equal to 0. See also [62, Proposition 1.3], which restates the result of Landon and Sosoë. For the third part, recall that $\sum_{i \geq 2} \Delta_{i,N}^{-1} = \Xi_N + N^{1/3}$ by the definition of Ξ_N . Then

$$\frac{1}{N} \sum_{i=2}^N \frac{1}{c(\lambda_1 - \lambda_i)} = \frac{N^{2/3}}{cN} (\Xi_N + N^{1/3}) = \frac{1}{c} + \frac{\Xi_N}{cN^{1/3}}.$$

For the fourth part, the third part implies that

$$\frac{1}{\delta_N} \left(\frac{1}{N} \sum_{i=2}^N \frac{1}{c_N(\lambda_1 - \lambda_i)} - 1 \right) = \frac{1}{c_N} + \frac{\Xi_N}{c_N \delta_N N^{1/3}} = \frac{1}{c_N} + \frac{\Xi_N}{c_N M_N^{1/3}},$$

where we used $1/c_N - 1 = \delta_N/c_N$. The right side tends to 1 in probability, because $c_N \rightarrow 1$, $M_N \rightarrow \infty$, and Ξ_N is tight. The error is $O(\delta_N) + O_{\mathbb{P}}(M_N^{-1/3})$. $\square$

4.2 Eigenvalue gaps at fixed confidence

The published results determine the locations of any fixed number of edge eigenvalues and give quantitative control away from the edge. The next lemma provides two-sided bounds on all gaps $\lambda_1 - \lambda_i$ with constants independent of N , at every fixed confidence in the sense of Section 3.4. The proof uses the simultaneous location bound of Lemma 4.3, and treats the first few gaps through edge convergence.

Lemma 4.6 (Eigenvalue gaps at fixed confidence). *For every $\eta \in (0, 1)$, there exist constants $A, K, C, c_\eta, C_\eta > 0$, depending only on η , and events $\mathcal{G}_N^\eta$ with $\liminf_N \mathbb{P}(\mathcal{G}_N^\eta) \geq 1 - \eta$, on which, for all sufficiently large N ,*

$$\begin{aligned} 2 - \lambda_i &\geq -AN^{-2/3} \quad (1 \leq i \leq N), \quad \#\{i : \lambda_i > 2\} \leq K, \\ \#\{i : 2 - \lambda_i \leq u\} &\leq C(1 + Nu^{3/2}) \quad (u \geq 0), \end{aligned} \quad (4.6)$$

and

$$c_\eta(i-1)^{2/3} \leq \Delta_{i,N} = N^{2/3}(\lambda_1 - \lambda_i) \leq C_\eta(i-1)^{2/3} \quad (2 \leq i \leq N). \quad (4.7)$$

Proof. Fix $\eta \in (0, 1)$ and use Lemma 4.3 with confidence $\eta/2$. Write $d_{i,N} = N^{2/3}(2 - \lambda_i)$ and $q_{i,N} = N^{2/3}(2 - \gamma_i)$. On its event, $|d_{i,N} - q_{i,N}| \leq C_{\eta/2}(1+i)^{1/4}$ simultaneously for every i . Since Lemma 4.1 gives $(3\pi i/2)^{2/3} \leq q_{i,N} \leq 7i^{2/3}$, there exist $K \geq 2$ and $0 < c_2 \leq C_2$, depending only on η , such that on this event

$$c_2(i/N)^{2/3} \leq 2 - \lambda_i \leq C_2(i/N)^{2/3} \quad (K \leq i \leq N). \quad (4.8)$$

The location bound at $i = 1$ also gives $|d_{1,N}| \leq A$ for a constant A depending only on η . Ordering implies $d_{i,N} \geq -A$ for all i . If $i \geq K$ and $2 - \lambda_i \leq u$ with $u \geq 0$, then (4.8) implies $i \leq N(u/c_2)^{3/2}$. Counting the first $K - 1$ indices separately gives

$$\#\{i : 2 - \lambda_i \leq u\} \leq K + c_2^{-3/2}Nu^{3/2}, \quad \#\{i : \lambda_i > 2\} \leq K.$$

This proves (4.6) after increasing its constant C .

Choose $i_0 \geq K$ so large that $c_2 i_0^{2/3} \geq 2A$ and $C_2 i_0^{2/3} \geq A$. For $i \geq i_0$, subtracting $d_{1,N}$ from (4.8) gives

$$\frac{c_2}{2}i^{2/3} \leq \Delta_{i,N} = d_{i,N} - d_{1,N} \leq 2C_2 i^{2/3}.$$

For the first few gaps, Lemma 4.2 gives $(\Delta_{2,N}, \Delta_{i_0,N}) \Rightarrow (\Delta_2, \Delta_{i_0})$, where $0 < \Delta_2 \leq \Delta_{i_0} < \infty$ almost surely. Choose $0 < \delta_* < D_*$ so that $\mathbb{P}(\Delta_2 \leq \delta_*) + \mathbb{P}(\Delta_{i_0} \geq D_*) \leq \eta/2$. By

the closed-set portmanteau bound, the event $\{\delta_* < \Delta_{2,N} \leq \Delta_{i_0,N} < D_*\}$ has limiting probability at least $1 - \eta/2$. On it, for $2 \leq i < i_0$, ordering gives

$$\delta_* i_0^{-2/3} (i-1)^{2/3} \leq \Delta_{i,N} \leq D_* (i-1)^{2/3}.$$

Intersect this event with $\mathcal{H}_N^{\eta/2}$. Its limiting probability is at least $1 - \eta$. Since $(i-1)^{2/3} \leq i^{2/3} \leq 2^{2/3} (i-1)^{2/3}$ for $i \geq 2$, the last two displays prove (4.7) on this intersection, with $c_\eta = \min(c_2/2, \delta_* i_0^{-2/3})$ and $C_\eta = \max(2^{5/3} C_2, D_*)$. $\square$

We now derive the spectral hypotheses used at low temperature. Recall from (3.8) and (3.14) the scaled eigenvalue gaps $a_i = c(\lambda_1 - \lambda_i)/2$, the top fraction θ_N , $R_N = Na_2$, and $m_N = \theta_N R_N$. The lemma covers both a fixed temperature $c > 1$ and an approach $c_N = 1 + \epsilon_N$ to the critical temperature from below. The low-temperature sections use its conclusions as follows. The bounds of the second part are the deterministic spectral hypotheses there, and the last part describes the Gaussian vector of Section 3.2.

Lemma 4.7 (Eigenvalue gaps and the top fraction at low temperature). *We consider the following two cases, with $c_\infty = \lim_N c_N$:*

- (F) $c_N = c > 1$ is fixed;
 - (M) $c_N = 1 + \epsilon_N$ with $\epsilon_N > 0$, $\epsilon_N \rightarrow 0$, and $N\epsilon_N^3 \rightarrow \infty$.
- Let $a_i = c_N(\lambda_1 - \lambda_i)/2$, and let θ_N , R_N , m_N be as in (3.14) with $c = c_N$.

1. We have the identity

$$\theta_N = 1 - \frac{1}{c_N} - \frac{\Xi_N}{c_N N^{1/3}}. \quad (4.9)$$

In case (F), $\theta_N \rightarrow \theta = 1 - 1/c$ in probability. In case (M), $\theta_N/\epsilon_N \rightarrow 1$ in probability, and in particular $\theta_N > 0$ with probability tending to one. In both cases, $N^{2/3}a_2 \Rightarrow c_\infty \Delta_2/2$, and $m_N \rightarrow \infty$ in probability. In case (F), moreover, $\log R_N / \log N \rightarrow 1/3$ and $R_N / \log^2 N \rightarrow \infty$ in probability.

2. For every $\eta \in (0, 1)$, there exist constants $B, c_*, C_*, \epsilon_* > 0$, depending only on η (and on c in case (F)), and events with limiting probability at least $1 - \eta$, on which for all sufficiently large N :

$$\begin{aligned} c_*(i-1)^{2/3} N^{-2/3} &\leq a_i \leq C_*(i-1)^{2/3} N^{-2/3} \quad (2 \leq i \leq N), \\ a_N &\leq B, \quad \frac{a_3}{a_2} \geq 1 + \epsilon_*. \end{aligned} \quad (4.10)$$

On these events, for every integer $2 \leq k \leq N$ and every integer $m' \geq 2$,

$$\frac{a_k}{a_2} \leq \frac{C_*}{c_*} (k-1)^{2/3}, \quad a_2^2 \sum_{i=2}^N a_i^{-2} \leq 4 \left(\frac{C_*}{c_*} \right)^2, \quad \sum_{i>m'} \left(\frac{a_2}{a_i} \right)^2 \leq 4 \left(\frac{C_*}{c_*} \right)^2 m'^{-1/3}, \quad (4.11)$$

and $c_* N^{1/3} \leq R_N \leq C_* N^{1/3}$. In case (F), for any fixed $\theta_0 \in (0, 1 - 1/c)$, these events may also be chosen so that $\theta_N \geq \theta_0$.

3. Conditional on W_N , let $(Z_i)_{i \geq 2}$ be independent centered Gaussian random variables with variances $(2a_i)^{-1}$. Then $\mathbb{E}[\sum_{i \geq 2} Z_i^2 \mid W_N] = N(1 - \theta_N)$, and for every $\varepsilon > 0$,

$$\mathbb{P}\left(\left|\sum_{i=2}^N Z_i^2 - N(1 - \theta_N)\right| > \varepsilon N \mid W_N\right) \longrightarrow 0 \quad \text{in probability.}$$

More precisely, on the events of the second part, this conditional probability is at most $C\varepsilon^{-2}N^{-2/3}$, with C depending only on c_* .

Proof. The identity (4.9) is the third part of Lemma 4.5, since $(2a_i)^{-1} = (c_N(\lambda_1 - \lambda_i))^{-1}$. In case (F), we have $\theta_N \rightarrow 1 - 1/c$, because Ξ_N is tight. In case (M), using $1 - 1/c_N = \epsilon_N/(1 + \epsilon_N)$, we find

$$\frac{\theta_N}{\epsilon_N} = \frac{1}{1 + \epsilon_N} - \frac{\Xi_N}{(1 + \epsilon_N)\epsilon_N N^{1/3}} \longrightarrow 1$$

in probability, because $\epsilon_N N^{1/3} = (N\epsilon_N^3)^{1/3} \rightarrow \infty$. Next, $N^{2/3}a_2 = c_N\Delta_{2,N}/2 \Rightarrow c_\infty\Delta_2/2$ by Lemma 4.2, and the limit is almost surely positive and finite. We have $m_N = \theta_N N^{1/3} \cdot N^{2/3}a_2$. In both cases $\theta_N N^{1/3}$ tends to infinity in probability, and therefore $m_N \rightarrow \infty$ in probability. In case (F), $\log R_N = (\log N)/3 + \log(N^{2/3}a_2)$ with $\log(N^{2/3}a_2) = O_{\mathbb{P}}(1)$, which implies the remaining two claims of the first part.

For the second part, fix η and work on the events $\mathcal{G}_N^{\eta/3}$ of Lemma 4.6. Using (4.7) and $1 \leq c_N \leq C$, we obtain the two-sided bound on a_i in (4.10), with $c_* = c_{\eta/3}/2$ and $C_* = CC_{\eta/3}/2$. Taking $i = N$, we also have $a_N \leq C_*$. For the third gap, $a_3/a_2 = \Delta_{3,N}/\Delta_{2,N} \Rightarrow \Delta_3/\Delta_2$, which is almost surely larger than 1 because the GOE Airy point process is simple (Lemma 4.2). Let $\varepsilon_* > 0$ be such that $\mathbb{P}(\Delta_3/\Delta_2 \leq 1 + \varepsilon_*) \leq \eta/3$. The portmanteau theorem gives $\limsup_N \mathbb{P}(a_3/a_2 \leq 1 + \varepsilon_*) \leq \eta/3$. In case (F), the event $\{\theta_N \geq \theta_0\}$ has probability tending to one by the first part. The intersection of these events has the properties stated in the second part. The bounds (4.11) follow from (4.10), the bounds $\sum_{j \geq 1} j^{-4/3} \leq 4$ and $\sum_{j \geq m'} j^{-4/3} \leq m'^{-4/3} + \int_{m'}^\infty x^{-4/3} dx \leq 4m'^{-1/3}$, and $R_N = N^{1/3} \cdot N^{2/3}a_2$.

For the third part, the conditional mean is $\sum_{i \geq 2} (2a_i)^{-1} = N(1 - \theta_N)$ by the definition of θ_N , and the conditional variance is $2 \sum_{i \geq 2} (2a_i)^{-2}$. On the events of the second part, it is at most $c_*^{-2}N^{4/3} \sum_{j \geq 1} j^{-4/3}/2 \leq 2c_*^{-2}N^{4/3}$. Chebyshev's inequality bounds the conditional probability by $2c_*^{-2}\varepsilon^{-2}N^{-2/3}$ on these events, and the convergence in probability follows from Lemma 3.8. $\square$

4.3 Edge counts at a varying scale

We now prove a law of large numbers for the eigenvalues within distance y/h_N of the edge, at each scale h_N^{-1} between $N^{-2/3}$ and 1. The theorem is used at the scale $h_N \asymp \delta_N^{-2}$

in Theorem 4.13, and at the smaller scales $h_N = t$ of the times t in the high-temperature profile windows. To cover the second use, we allow a random shift of the edge and a factor $\alpha_N \rightarrow 1$. For counts, the restriction $h_N = o(N^{2/3})$ is what allows the number of eigenvalues at the scale h_N^{-1} to grow arbitrarily slowly.

Theorem 4.8 (Edge counts at a varying scale). *Let $h_N > 0$ be deterministic with $h_N \rightarrow \infty$ and $h_N = o(N^{2/3})$, and set $K_N = Nh_N^{-3/2}$. Let $\alpha_N > 0$ be deterministic with $\alpha_N \rightarrow 1$, and let ξ_N be real random variables with $h_N \xi_N \rightarrow 0$ in probability. We define the measure*

$$\omega_N = \frac{1}{K_N} \sum_{i=1}^N \delta_{\alpha_N h_N (2 - \lambda_i - \xi_N)}. \quad (4.12)$$

1. *For every compact $Y \subset \mathbb{R}$, $\sup_{y \in Y} |\omega_N((-\infty, y]) - 2y_+^{3/2}/(3\pi)| \rightarrow 0$ in probability. In particular, for $\alpha_N = 1$ and $\xi_N = 0$, the first limit in (4.14) below holds.*
2. *With probability tending to one, every atom of ω_N lies in $(-1/2, \infty)$. For every $\eta \in (0, 1)$, there exist C_η and events with limiting probability at least $1 - \eta$ on which $\omega_N((-\infty, y]) \leq C_\eta(K_N^{-1} + y^{3/2})$ for all $y \geq 1$.*
3. *Let $k > 3/2$, and let $\varphi : [-1/2, \infty) \rightarrow \mathbb{R}$ be continuous with $\sup_{y \geq 0} (1 + y)^k |\varphi(y)| < \infty$. Then*

$$\int \varphi d\omega_N \rightarrow \frac{1}{\pi} \int_0^\infty \varphi(y) \sqrt{y} dy \quad \text{in probability.} \quad (4.13)$$

The same holds for deterministic continuous φ_N that converge to φ uniformly on compact subsets of $[-1/2, \infty)$ and satisfy $\sup_N (\sup_{[-1/2, 0]} |\varphi_N| + \sup_{y \geq 0} (1 + y)^k |\varphi_N(y)|) < \infty$. In particular, ω_N converges vaguely in probability to $\pi^{-1} \sqrt{y} \mathbf{1}_{\{y \geq 0\}} dy$.

4. *For $\alpha_N = 1$ and $\xi_N = 0$, we have, for each $y > 0$ and each fixed $k > 3/2$,*

$$\begin{aligned} \frac{1}{K_N} \# \left\{ i : 2 - \lambda_i \leq \frac{y}{h_N} \right\} &\rightarrow \frac{2}{3\pi} y^{3/2} \quad \text{in probability,} \\ \sum_{i=1}^N e^{-h_N(2 - \lambda_i)} &= \frac{N}{2\sqrt{\pi} h_N^{3/2}} (1 + o_{\mathbb{P}}(1)), \\ \sum_{i=1}^N (1 + h_N(2 - \lambda_i))^{-k} &= K_N C_k (1 + o_{\mathbb{P}}(1)), \end{aligned} \quad (4.14)$$

where $C_k = \pi^{-1} B(3/2, k - 3/2)$. In the last sum, the finitely many terms with $1 + h_N(2 - \lambda_i) \leq 1/2$ are excluded, and with probability tending to one there are none.

In the third part, the integral is taken on the event of the second part that all atoms lie in $(-1/2, \infty)$; on the complementary event, we define it to be 0. Any other measurable finite choice gives the same convergence in probability, and the same convention applies to φ_N .

Proof. The proof has four steps. We first prove the count asymptotic for $\alpha_N = 1$, $\xi_N = 0$, at a fixed $y > 0$, using the simultaneous location bound. We then use the uniform upper

bound (4.6) on the counting function to bound the counts uniformly in y , add the shift and the factor α_N , and integrate.

Step 1: counts without shift. Let $F_N^0(y) = K_N^{-1} \#\{i : h_N(2 - \lambda_i) \leq y\}$, and fix $y > 0$. We have $K_N \rightarrow \infty$ and $K_N = o(N)$. Fix $r \in (0, 2y^{3/2}/(3\pi))$ and let $i_{\pm} = \lfloor (2y^{3/2}/(3\pi) \pm r)K_N \rfloor$. On the events of Lemma 4.3, for every fixed confidence $\eta > 0$,

$$h_N |\lambda_{i_{\pm}} - \gamma_{i_{\pm}}| \leq C_{\eta} h_N N^{-2/3} (1 + i_{\pm})^{1/4} \leq C_{\eta, y, r} K_N^{-5/12} \rightarrow 0,$$

because $h_N N^{-2/3} = K_N^{-2/3}$. By Lemma 4.1, $h_N(2 - \gamma_{i_{\pm}}) \rightarrow (y^{3/2} \pm 3\pi r/2)^{2/3}$. Thus, for all large N on these events, ordering brackets $K_N F_N^0(y)$ between i_- and i_+ . Letting $r \downarrow 0$ and then $\eta \downarrow 0$ proves $F_N^0(y) \rightarrow 2y^{3/2}/(3\pi)$ in probability, however slowly K_N tends to infinity.

For $y \leq 0$, we have $F_N^0(y) \leq K_N^{-1} \#\{i : \lambda_i \geq 2\}$, which tends to zero in probability by (4.6) and Lemma 3.8. Altogether, $F_N^0(y) \rightarrow F(y) = 2y_+^{3/2}/(3\pi)$ for every $y \in \mathbb{R}$. Each F_N^0 is nondecreasing and F is continuous. Convergence at the points of a finite mesh with spacing ϱ of a compact interval Y' , together with monotonicity, then gives $\sup_{Y'} |F_N^0 - F| \leq o_{\mathbb{P}}(1) + \sup\{|F(y) - F(y')| : y, y' \in Y', |y - y'| \leq \varrho\}$. Letting $\varrho \rightarrow 0$, we obtain $\sup_{Y'} |F_N^0 - F| \rightarrow 0$ in probability.

Step 2: a uniform bound on the counts. On the events of Lemma 4.6 with confidence $\eta/2$, the last bound of (4.6) with $u = y/h_N$ shows that $F_N^0(y) \leq C(K_N^{-1} + y^{3/2})$ for all $y \geq 0$, and by the first bound we have $h_N(2 - \lambda_i) \geq -Ah_N N^{-2/3}$, which tends to zero. Let $F_N(y) = \omega_N((-\infty, y])$. On the further event $\{|h_N \xi_N| \leq 1/8, \alpha_N \in [1/2, 2]\}$, whose probability tends to one, every atom $\alpha_N(h_N(2 - \lambda_i) - h_N \xi_N)$ of ω_N is larger than $-1/2$ for N large, and for $y \geq 1$,

$$F_N(y) = F_N^0(y/\alpha_N + h_N \xi_N) \leq F_N^0(2y + 1) \leq C(K_N^{-1} + (3y)^{3/2}).$$

This proves the second part.

Step 3: shift and factor. Fix a compact interval Y and $\varrho \in (0, 1)$. With probability tending to one, $|y/\alpha_N + h_N \xi_N - y| \leq \varrho$ for all $y \in Y$, and then $F_N^0(y - \varrho) \leq F_N(y) \leq F_N^0(y + \varrho)$ by monotonicity. By Step 1 on the interval $Y' = Y + [-1, 1]$ and the uniform continuity of F , we obtain $\sup_Y |F_N - F| \leq o_{\mathbb{P}}(1) + \sup_{y \in Y} |F(y \pm \varrho) - F(y)|$. Letting $\varrho \rightarrow 0$ proves the first part.

Step 4: integration. Let φ be as in the third part, with $\sup_{[-1/2, 0]} |\varphi| + \sup_{y \geq 0} (1 + y)^k |\varphi(y)| \leq \Lambda$. Fix $\eta, \varepsilon > 0$ and work on the events of the second part. For $R \geq 1$, we decompose (R, ∞) into the intervals $(2^j R, 2^{j+1} R]$ and find

$$\int_{(R, \infty)} |\varphi| d\omega_N \leq \Lambda \sum_{j \geq 0} (2^j R)^{-k} F_N(2^{j+1} R) \leq C_{\eta, k} \Lambda (R^{3/2-k} + K_N^{-1} R^{-k}),$$

and also $\pi^{-1} \int_R^\infty |\varphi(y)| \sqrt{y} dy \leq C_k \Lambda R^{3/2-k}$. Since $k > 3/2$, we may choose R , depending on $\eta, \varepsilon, k, \Lambda$, such that both tails are at most ε for all large N . On $[-1/2, R]$, we fix a mesh $-1/2 = y_0 < y_1 < \dots < y_n = R$ on which φ oscillates by at most $\varepsilon/(1 + F(R))$ on each interval $[y_{j-1}, y_j]$. Then

$$\left| \int_{(-1/2, R]} \varphi d\omega_N - \sum_{j=1}^n \varphi(y_j) (F_N(y_j) - F_N(y_{j-1})) \right| \leq \frac{\varepsilon F_N(R)}{1 + F(R)},$$

and the same holds with F in place of F_N and $\pi^{-1} \sqrt{y} \mathbf{1}_{\{y \geq 0\}} dy$ in place of ω_N . By the first part, the Riemann–Stieltjes sums converge in probability, and $F_N(R) \rightarrow F(R)$. Combining the three estimates, we find that $|\int \varphi d\omega_N - \pi^{-1} \int_0^\infty \varphi \sqrt{y} dy| \leq 4\varepsilon$ with probability at least $1 - \eta - o(1)$. The parameters η and ε are arbitrary, and (4.13) follows. For the sequence φ_N , the tail bound is uniform in N , and on $[-1/2, R]$ we add the error $\sup_{[-1/2, R]} |\varphi_N - \varphi| \cdot F_N(R)$, which tends to zero in probability. Vague convergence is the case of compactly supported φ .

For the fourth part, the first limit in (4.14) is the first part with $\alpha_N = 1$ and $\xi_N = 0$. For the other two, take $\varphi(y) = e^{-y}$ and $\varphi(y) = (1 + y)^{-k}$ in (4.13). We have $\pi^{-1} \int_0^\infty e^{-y} \sqrt{y} dy = \Gamma(3/2)/\pi = 1/(2\sqrt{\pi})$ and $\pi^{-1} \int_0^\infty \sqrt{y} (1 + y)^{-k} dy = \pi^{-1} B(3/2, k - 3/2) = C_k$, and $K_N/(2\sqrt{\pi}) = N/(2\sqrt{\pi} h_N^{3/2})$. $\square$

The theorem applies in particular to the gaps below the top eigenvalue at the scale of the times in the window of the high-temperature profile. Let $c_N = 1 - \delta_N$, $M_N = N\delta_N^3$ and d_N be as in Section 3.2, and let $t_N > 0$ be deterministic with

$$\delta_N^2 t_N \rightarrow \infty, \quad \delta_N^2 t_N = o(M_N^{2/3}). \quad (4.15)$$

The second condition is equivalent to $t_N = o(N^{2/3})$, because $M_N^{2/3} = N^{2/3} \delta_N^2$. Note that $c_N(\lambda_1 - \lambda_i) = p_i - d_N$ in terms of the precisions (3.12).

Corollary 4.9 (Gaps below the top eigenvalue at a smaller scale). *Assume $\delta_N \rightarrow 0$, $M_N \rightarrow \infty$, and (4.15). Let*

$$\omega_N^{\text{gap}} = \frac{t_N^{3/2}}{N} \sum_{i=1}^N \delta_{t_N c_N(\lambda_1 - \lambda_i)}.$$

Then ω_N^{gap} converges vaguely in probability to $\pi^{-1} \sqrt{y} dy$ on $[0, \infty)$. If $\varphi : [0, \infty) \rightarrow \mathbb{R}$ is continuous with $\sup_{y \geq 0} (1 + y)^2 |\varphi(y)| < \infty$, then

$$\int \varphi d\omega_N^{\text{gap}} \longrightarrow \frac{1}{\pi} \int_0^\infty \varphi(y) \sqrt{y} dy \quad \text{in probability.}$$

The same conclusion holds for deterministic continuous φ_N that converge to φ uniformly on compact subsets of $[0, \infty)$ and satisfy $\sup_N \sup_{y \geq 0} (1 + y)^2 |\varphi_N(y)| < \infty$. The atom of ω_N^{gap} at 0 has mass $t_N^{3/2}/N \rightarrow 0$.

Proof. We use Theorem 4.8 with $h_N = t_N$, $\alpha_N = c_N$, and $\xi_N = 2 - \lambda_1$. Then $K_N = Nt_N^{-3/2}$ and $\omega_N = \omega_N^{\text{gap}}$. Its hypotheses hold. Indeed, $t_N \rightarrow \infty$ because $\delta_N \rightarrow 0$ and $\delta_N^2 t_N \rightarrow \infty$, we have $t_N = o(N^{2/3})$ by (4.15), and $h_N \xi_N = t_N N^{-2/3} \cdot N^{2/3}(2 - \lambda_1) \rightarrow 0$ in probability, since $N^{2/3}(2 - \lambda_1)$ is tight by Lemma 4.2. All atoms are nonnegative. We may therefore extend φ and φ_N to $[-1/2, 0)$ by their values at 0, and the claims follow from the third part of Theorem 4.8 with $k = 2$. $\square$

Only $t_N \rightarrow \infty$, $t_N = o(N^{2/3})$ and $c_N \rightarrow 1$ were used in this proof.

4.4 The Lagrange multiplier at high temperature

At high temperature, $0 \leq c < 1$, the Lagrange multiplier is determined by the Stieltjes transform of the empirical eigenvalue distribution. Indeed, for $c > 0$, set $E_N = z_N/c$. Then $E_N > \lambda_1$, and (3.12) is equivalent to

$$m_N(E_N) = c. \quad (4.16)$$

For $0 < c < 1$, the semicircle equation $m_{\text{sc}}(E) = c$ has the unique solution $E_* = c + c^{-1}$, since $\sqrt{E_*^2 - 4} = c^{-1} - c$. Correspondingly, $cE_* = 1 + c^2$. Equation (4.16) is the saddle-point equation of the contour-integral representation of the partition function in [6, Lemma 1.3], which goes back to [97], as noted after that lemma. Since m_N is random, we compare it with m_{sc} only at finitely many deterministic points and transfer the comparison to the root by monotonicity. The following deterministic lemma does this.

Lemma 4.10 (Comparison of roots and derivatives). *Let $\lambda_1 \geq \dots \geq \lambda_n$ be real numbers and $w_1, \dots, w_n > 0$, and let $f(x) = \sum_{i=1}^n w_i(x - \lambda_i)^{-1}$ for $x > \lambda_1$. Let $x_* \in \mathbb{R}$ and $h > 0$ with $x_* - h > \lambda_1$, let $I = [x_* - h, x_* + h]$, and let $\bar{f} \in C^1(I)$ with $\Lambda = -\bar{f}'(x_*) > 0$ and*

$$(1 - \alpha)\Lambda \leq -\bar{f}'(x) \leq (1 + \alpha)\Lambda \quad (x \in I)$$

for some $\alpha \in [0, 1)$.

1. *If $|f(x_* \pm h/2) - \bar{f}(x_* \pm h/2)| \leq e$ for both signs, where $e < (1 - \alpha)\Lambda h/2$, then $\bar{f}(x_*) > 0$, the equation $f(x) = \bar{f}(x_*)$ has a unique solution $x_f \in (\lambda_1, \infty)$, and $|x_f - x_*| < h/2$.*
2. *If, in addition, $|f(x_* \pm h) - \bar{f}(x_* \pm h)| \leq e$ for both signs, then*

$$\left| \frac{-f'(x_f)}{\Lambda} - 1 \right| \leq \alpha + \frac{4e}{h\Lambda}. \quad (4.17)$$

3. *Let $a \in (0, h]$. If $|f(x_* \pm a) - \bar{f}(x_* \pm a)| \leq e_a$ for both signs, where $e_a < (1 - \alpha)\Lambda a$, then the equation $f(x) = \bar{f}(x_*)$ has a unique solution $x_f \in (\lambda_1, \infty)$, and $|x_f - x_*| < a$.*

Proof. We prove the third part first. The mean value theorem and the bound on $\bar{f}'$ give $\bar{f}(x_* - a) \geq \bar{f}(x_*) + (1 - \alpha)\Lambda a$ and $\bar{f}(x_* + a) \leq \bar{f}(x_*) - (1 - \alpha)\Lambda a$. Combining these bounds with the hypothesis on the errors, we obtain

$$f(x_* - a) \geq \bar{f}(x_* - a) - e_a > \bar{f}(x_*), \quad \bar{f}(x_*) > \bar{f}(x_* + a) + e_a \geq f(x_* + a) > 0.$$

The function f is continuous and strictly decreasing on (λ_1, ∞) , with limits $+\infty$ and 0. As the level $\bar{f}(x_*)$ is positive, it is attained exactly once, at a point of $(x_* - a, x_* + a)$. The first part is the case $a = h/2$.

For the second part, the function $q = -f' = \sum_i w_i(x - \lambda_i)^{-2}$ is positive and decreasing on (λ_1, ∞) . We have $x_* - h < x_* - h/2 < x_f < x_* + h/2 < x_* + h$, and by the mean value theorem,

$$\frac{f(x_* + h/2) - f(x_* + h)}{h/2} \leq q(x_f) \leq \frac{f(x_* - h) - f(x_* - h/2)}{h/2}.$$

By the mean value theorem, the corresponding difference quotients of $\bar{f}$ lie in $[(1 - \alpha)\Lambda, (1 + \alpha)\Lambda]$. Replacing $\bar{f}$ by f at the two endpoints changes each quotient by at most $2e/(h/2) = 4e/h$. We conclude that $(1 - \alpha)\Lambda - 4e/h \leq q(x_f) \leq (1 + \alpha)\Lambda + 4e/h$, which is (4.17). $\square$

We now determine the Lagrange multiplier along an approach to the critical temperature from above. The error e_N below comes from the common-event localized estimate (4.4). No lower bound on the growth of M_N is assumed.

Proposition 4.11 (Quantitative Lagrange multiplier near the critical temperature). *Let $c_N = 1 - \delta_N$ with $\delta_N \in (0, 1)$, $\delta_N \rightarrow 0$, and $M_N = N\delta_N^3 \rightarrow \infty$. We write*

$$\ell_N = \log(2 + M_N), \quad v_N = \frac{\delta_N^2}{c_N}, \quad E_* = c_N + c_N^{-1} = 2 + v_N,$$

$$e_N = M_N^{-5/12} + N^{-1/3}M_N^{-2/3} + N^{-1/6}M_N^{-1/3} + M_N^{-1}.$$

Let z_N be the Lagrange multiplier of $c_N W_N$, let $E_N = z_N/c_N$, so that $\mathfrak{m}_N(E_N) = c_N$, and let d_N and q_i be as in (3.12). Then

$$|E_N - E_*| = O_{\mathbb{P}}(v_N e_N), \tag{4.18}$$

$$\left| \frac{1 - c_N^2}{N} \sum_{i=1}^N q_i^2 - 1 \right| = O_{\mathbb{P}}(\sqrt{e_N}). \tag{4.19}$$

Moreover, in probability,

$$\ell_N^2 \frac{|z_N - (1 + c_N^2)|}{\delta_N^2} \longrightarrow 0, \quad \ell_N \left| \frac{1 - c_N^2}{N} \sum_{i=1}^N q_i^2 - 1 \right| \longrightarrow 0, \tag{4.20}$$

$$\ell_N \left| \frac{d_N}{\delta_N^2} - 1 \right| \longrightarrow 0. \tag{4.21}$$

Proof. We write c, δ, M, ℓ, v, e for $c_N, \delta_N, M_N, \ell_N, v_N, e_N$. Note that $\delta = (M/N)^{1/3}$ and $N^{2/3}v = M^{2/3}/c$, and that $e \rightarrow 0$. The proof has four steps. In Step 1, we record properties of m_{sc} near E_* , and in Step 2, we bound $m_N - m_{sc}$ at deterministic points. Step 3 proves the estimate (4.18) for the root, and Step 4 proves (4.19) and the consequences.

Step 1: the semicircle transform. Differentiating (4.1), we have $-m'_{sc}(E) = m_{sc}(E)/\sqrt{E^2 - 4}$ and

$$\frac{d}{dE} \log(-m'_{sc}(E)) = -\frac{E + \sqrt{E^2 - 4}}{E^2 - 4}.$$

At E_* , using $\sqrt{E_*^2 - 4} = c^{-1} - c$, we find

$$\Lambda_N = -m'_{sc}(E_*) = \frac{c^2}{1 - c^2}, \quad \left| \frac{m'_{sc}(E)}{m'_{sc}(E_*)} - 1 \right| \leq C_0 \frac{|E - E_*|}{v} \quad \left(|E - E_*| \leq \frac{v}{2} \right), \quad (4.22)$$

with a universal constant C_0 . For the second bound, note that $E - 2 \geq v/2$ and $E \leq 3$ on this interval for N large. The logarithmic derivative is then bounded by $C/(E - 2) \leq 2C/v$. Integrating it from E_* to E , we bound the absolute value of the logarithm of the ratio by $s_0 = 2C|E - E_*|/v \leq C$, and the claim follows from $|e^s - 1| \leq e^C|s|$ for $|s| \leq C$. Note also that $\Lambda_N v = c\delta/(1 + c) \geq \delta/4$ for $c \geq 1/2$. Let $h_0 = \min(1/2, 1/(2C_0))$. On $I_0 = [E_* - h_0 v, E_* + h_0 v]$, (4.22) shows that the hypothesis of Lemma 4.10 holds with $\tilde{f} = m_{sc}$, $x_* = E_*$, $\Lambda = \Lambda_N$, and $\alpha = 1/2$, and on a subinterval of half-length $h \leq h_0 v$ it holds with $\alpha = C_0 h/v$. The interval I_0 lies in (λ_1, ∞) with probability tending to one, because $N^{2/3}v \asymp M^{2/3} \rightarrow \infty$ and $N^{2/3}(\lambda_1 - 2)$ is tight.

Step 2: the transform at deterministic points. We show that for every $\varepsilon > 0$ there exist L, N_0 such that for all $N \geq N_0$ and all deterministic E with $|E - E_*| \leq v/2$,

$$\mathbb{P}\left(E \leq \lambda_1 \text{ or } [E > \lambda_1 \text{ and } |m_N(E) - m_{sc}(E)| > L\delta e]\right) \leq \varepsilon. \quad (4.23)$$

Here $m_N(E)$ is considered only on the event $E > \lambda_1$, where it is defined. This is a bound for each deterministic point, uniform in its choice, and not a bound for the supremum over a continuum of points. We apply it below at four points.

Let $u = N^{2/3}(E - 2)$, so that $u \asymp M^{2/3}$ uniformly over these deterministic points. For large N it exceeds the fixed threshold in Lemma 4.4. On its common confidence event $\mathcal{R}_N$, $E > \lambda_1$, and (4.4), divided by $\delta = M^{1/3}N^{-1/3}$, gives

$$\begin{aligned} \frac{1}{\delta} \mathbb{E}[\mathbf{1}_{\mathcal{R}_N} |m_N(E) - m_{sc}(E)|] &\leq C_\varepsilon M^{-1/3} \left(M^{-1/12} + N^{-1/3} M^{-1/3} + N^{-1/6} + M^{-2/3} \right) \\ &= C_\varepsilon e. \end{aligned}$$

Choose the event with failure probability at most $\varepsilon/2$, then apply Markov's inequality with a sufficiently large L . This proves (4.23); the constants are uniform in the deterministic choice of E .

Step 3: the root. Fix $\varepsilon > 0$, let L be as in (4.23), and set $a = 16Lve$, which is at most h_0v for N large. By (4.23) at the two points $E_* \pm a$, with probability at least $1 - 2\varepsilon - o(1)$ both errors are at most $L\delta e$, and $I_0 \subset (\lambda_1, \infty)$. The threshold in the third part of Lemma 4.10 with $\alpha = 1/2$ is $\Lambda_N a/2 \geq 2L\delta e$, by Step 1. Recall that $m_N(E_N) = c = m_{sc}(E_*)$. The lemma implies that $|E_N - E_*| < 16Lve$, and since ε is arbitrary, this proves (4.18).

Step 4: the derivative and the consequences. Fix $\varepsilon > 0$ and L as before, and apply the first two parts of Lemma 4.10 with $h = v\sqrt{e} \leq h_0v$ and $\alpha = C_0\sqrt{e}$. With probability at least $1 - 4\varepsilon - o(1)$, the errors at the four points $E_* \pm h/2, E_* \pm h$ are at most $L\delta e$. The condition $L\delta e < (1 - \alpha)\Lambda_N h/2$ holds for N large, since $\Lambda_N h \geq \delta\sqrt{e}/4$ and $e \rightarrow 0$. By (4.17),

$$\left| \frac{-m'_N(E_N)}{\Lambda_N} - 1 \right| \leq C_0\sqrt{e} + \frac{4L\delta e}{\Lambda_N v\sqrt{e}} \leq (C_0 + 16L)\sqrt{e}.$$

We have $N^{-1} \sum_i q_i^2 = c^{-2} N^{-1} \sum_i (E_N - \lambda_i)^{-2} = c^{-2} (-m'_N(E_N))$ and $c^{-2} \Lambda_N = (1 - c^2)^{-1}$, which gives (4.19).

For the consequences, $1 \leq M \leq N$ eventually, so $e \leq 4M^{-5/12}$. Thus $\ell^2 e \rightarrow 0$ and $\ell\sqrt{e} \rightarrow 0$. Since $z_N - (1 + c^2) = c(E_N - E_*) = O_{\mathbb{P}}(\delta^2 e)$, the first limit in (4.20) follows. The second follows from (4.19). Finally, since $1 + c^2 - \delta^2 = 2c$,

$$d_N - \delta^2 = (z_N - (1 + c^2)) + c(2 - \lambda_1). \quad (4.24)$$

The last term is $O_{\mathbb{P}}(N^{-2/3})$ by Lemma 4.2, and $N^{-2/3}/\delta^2 = M^{-2/3}$. Dividing (4.24) by δ^2 , we find $\ell|d_N/\delta^2 - 1| = O_{\mathbb{P}}(\ell^2 e + \ell M^{-2/3}) \rightarrow 0$. $\square$

At a fixed temperature the same comparison is simpler, because the edge stays at a fixed distance from E_* .

Lemma 4.12 (The Lagrange multiplier at fixed high temperature). *Fix $0 < c < 1$. Then $z_N - (1 + c^2) = O_{\mathbb{P}}(N^{-5/12})$ and $N^{-1} \sum_i q_i^2 \rightarrow (1 - c^2)^{-1}$ in probability.*

Proof. Let $E_* = c + c^{-1}$, choose a fixed compact interval $I \subset (2, \infty)$ around E_* , and set $\Lambda = -m'_{sc}(E_*) = c^2/(1 - c^2) > 0$. On F_C of Lemma 4.2, all denominators on I stay uniformly positive for large N . Comparing eigenvalues with quantiles, and then using monotone semicircle quadrature, gives

$$\sup_{E \in I} |m_N(E) - m_{sc}(E)| \leq C_c(CN^{-5/12} + N^{-1}).$$

The root bracket Lemma 4.10 gives $|E_N - E_*| \leq C_{c,C} N^{-5/12}$ on F_C . Taking confidence arbitrarily close to one proves the stated multiplier rate. For the derivative bracket take $h = N^{-5/24}$; the relative derivative error is bounded by $C_{c,C}(h + N^{-5/12}/h) \rightarrow 0$ on F_C . Since $N^{-1} \sum_i q_i^2 = c^{-2} (-m'_N(E_N))$ and $c^{-2} \Lambda = (1 - c^2)^{-1}$, this gives the trace limit. $\square$

We collect the consequences used in the high-temperature sections.

Theorem 4.13 (The Lagrange multiplier above the critical temperature).

1. For fixed $0 < c < 1$, in probability,

$$z_N - (1 + c^2) = O_{\mathbb{P}}(N^{-5/12}), \quad d_N \rightarrow (1 - c)^2, \quad s_N \rightarrow (1 + c)^2, \quad \frac{1}{N} \sum_{i=1}^N q_i^2 \rightarrow \frac{1}{1 - c^2}.$$

2. Let $c_N = 1 - \delta_N$ with $\delta_N \in (0, 1)$, $\delta_N \rightarrow 0$, $M_N = N\delta_N^3 \rightarrow \infty$, and $\ell_N = \log(2 + M_N)$. Then, in probability,

$$\begin{aligned} \ell_N^2 \frac{|z_N - (1 + c_N^2)|}{\delta_N^2} &\rightarrow 0, \quad \ell_N \left| \frac{d_N}{\delta_N^2} - 1 \right| \rightarrow 0, \\ s_N &\rightarrow 4, \quad \ell_N \left| \frac{1 - c_N^2}{N} \sum_{i=1}^N q_i^2 - 1 \right| \rightarrow 0. \end{aligned} \quad (4.25)$$

3. In the setting of the second part, for every fixed $k > 3/2$,

$$\begin{aligned} \sum_{i=1}^N q_i^k &= N c_N^{-3/2} \delta_N^{3-2k} C_k (1 + o_{\mathbb{P}}(1)), \quad C_k = \frac{1}{\pi} B\left(\frac{3}{2}, k - \frac{3}{2}\right), \\ \frac{\max_i q_i}{\sqrt{2 \sum_i q_i^2}} &= M_N^{-1/2} (1 + o_{\mathbb{P}}(1)). \end{aligned} \quad (4.26)$$

4. In the setting of the second part, fix $A > 0$, and let J_N be the number of indices i with $2 - \lambda_i \leq A\delta_N^2/\ell_N$. Then

$$J_N = \frac{2A^{3/2}}{3\pi} \frac{M_N}{\ell_N^{3/2}} (1 + o_{\mathbb{P}}(1)), \quad \frac{\sum_{i \leq J_N} q_i^2}{\sum_{i=1}^N q_i^2} = \frac{4A^{3/2}}{3\pi} \ell_N^{-3/2} (1 + o_{\mathbb{P}}(1)). \quad (4.27)$$

Proof. In the first part, the statements on z_N and on $\sum_i q_i^2$ are contained in Lemma 4.12. The limits of d_N and s_N follow from the identities $d_N = (z_N - (1 + c^2)) + (1 - c)^2 + c(2 - \lambda_1)$ and $s_N = (z_N - (1 + c^2)) + (1 + c)^2 - c(2 + \lambda_N)$, together with $\lambda_1 \rightarrow 2$ and $\lambda_N \rightarrow -2$ in probability (Lemma 4.2). In the second part, all statements except $s_N \rightarrow 4$ are (4.20) and (4.21), and $s_N \rightarrow 4$ follows from the same identity for s_N .

For the third part, write $v_N = \delta_N^2/c_N$ and $E_N = z_N/c_N$ as in Proposition 4.11, and set $y_i = (2 - \lambda_i)/v_N$ and $\vartheta_N = (E_N - 2)/v_N$. Then $q_i = c_N^{-1}(E_N - \lambda_i)^{-1} = c_N^{-1}v_N^{-1}(\vartheta_N + y_i)^{-1}$, and $\vartheta_N = 1 + O_{\mathbb{P}}(e_N) \rightarrow 1$ in probability by (4.18). We apply Theorem 4.8 with $h_N = v_N^{-1}$, $\alpha_N = 1$, and $\xi_N = 0$; this is admissible because $h_N \rightarrow \infty$ and $h_N/N^{2/3} = c_N/M_N^{2/3} \rightarrow 0$, and then $K_N = Nv_N^{3/2}$. With probability tending to one, $\min_i y_i \geq -1/4$ and $|\vartheta_N - 1| \leq 1/4$, and then, by the mean value theorem,

$$|(\vartheta_N + y_i)^{-k} - (1 + y_i)^{-k}| \leq k 2^{k+2} |\vartheta_N - 1| (1 + y_i)^{-k},$$

since every point between $\vartheta_N + y_i$ and $1 + y_i$ is at least $(1 + y_i)/2$, and $1 + y_i \geq 3/4$. The fourth part of Theorem 4.8 gives $K_N^{-1} \sum_i (1 + y_i)^{-k} \rightarrow C_k$ in probability. We conclude that $\sum_i (\vartheta_N + y_i)^{-k} = K_N C_k (1 + o_{\mathbb{P}}(1))$, and

$$\sum_{i=1}^N q_i^k = c_N^{-k} v_N^{-k} K_N C_k (1 + o_{\mathbb{P}}(1)) = N c_N^{-k} v_N^{3/2-k} C_k (1 + o_{\mathbb{P}}(1)),$$

and $c_N^{-k} v_N^{3/2-k} = c_N^{-3/2} \delta_N^{3-2k}$. For the ratio, we have $\max_i q_i = q_1 = d_N^{-1} = \delta_N^{-2} (1 + o_{\mathbb{P}}(1))$ by the second part, and $\sum_i q_i^2 = (N/(2\delta_N))(1 + o_{\mathbb{P}}(1))$ by the case $k = 2$, where $C_2 = 1/2$. These two asymptotics give $q_1 / \sqrt{2 \sum_i q_i^2} = \delta_N^{-2} (N/\delta_N)^{-1/2} (1 + o_{\mathbb{P}}(1)) = M_N^{-1/2} (1 + o_{\mathbb{P}}(1))$.

For the fourth part, apply Theorem 4.8 with $h_N = \ell_N / (A\delta_N^2)$, $\alpha_N = 1$, and $\xi_N = 0$. This is admissible because $h_N \rightarrow \infty$ and $h_N / N^{2/3} = \ell_N / (AM_N^{2/3}) \rightarrow 0$. Here $K_N = A^{3/2} M_N / \ell_N^{3/2}$, and the asymptotic for J_N is the count at $y = 1$. The eigenvalues are ordered, and the selected indices are $1, \dots, J_N$. For these indices, by (4.24),

$$p_i = z_N - c_N \lambda_i = \delta_N^2 + (z_N - (1 + c_N^2)) + c_N (2 - \lambda_i),$$

where $z_N - (1 + c_N^2) = o_{\mathbb{P}}(\delta_N^2)$, $0 \leq c_N (2 - \lambda_i) \leq A\delta_N^2 / \ell_N = o(\delta_N^2)$ when $\lambda_i \leq 2$, and $c_N (2 - \lambda_i) \geq -c_N (\lambda_1 - 2) = O_{\mathbb{P}}(N^{-2/3}) = o_{\mathbb{P}}(\delta_N^2)$ otherwise. Thus $q_i^2 = \delta_N^{-4} (1 + o_{\mathbb{P}}(1))$ uniformly for $i \leq J_N$. Dividing $J_N \delta_N^{-4} (1 + o_{\mathbb{P}}(1))$ by $\sum_i q_i^2 = (N/(2\delta_N))(1 + o_{\mathbb{P}}(1))$ we arrive at the second claim, since $2J_N / (N\delta_N^3) = 2J_N / M_N$. $\square$

Remark 4.14 (Scales of the precisions and counts when $t_N = o(N^{2/3})$). In the setting of the second part of Theorem 4.13, assume also (4.15). Then, in probability,

$$\frac{d_N t_N}{\delta_N^2 t_N} \rightarrow 1, \quad \frac{N d_N^{3/2}}{M_N} \rightarrow 1, \quad t_N = o(N^{2/3}), \quad N t_N^{-3/2} \rightarrow \infty, \quad \frac{t_N}{N \delta_N} = \frac{\delta_N^2 t_N}{M_N} \rightarrow 0. \quad (4.28)$$

4.5 The critical window

In the critical window, $c_N = 1 + \sigma N^{-1/3}$, the Lagrange multiplier is determined by the gaps at the top of the spectrum. This section first studies the GOE Airy point process and the limiting equation for the multiplier, and then the half-precisions at finite N .

The limiting equation. Recall the reflected GOE Airy point process $\chi_1 < \chi_2 < \dots$ of Lemma 4.2, the limit of $N^{2/3}(2 - \lambda_i)$, its gaps $\Delta_i = \chi_i - \chi_1$, the regularized inverse-gap statistic Ξ of Lemma 4.5, and $t_k = (3\pi k/2)^{2/3}$. By [62, Proposition 1.4], almost surely $\chi_k/t_k \rightarrow 1$, and the function

$$\mathcal{S}(x) = \Xi + \frac{1}{x} + \sum_{i=2}^{\infty} \left(\frac{1}{x + \Delta_i} - \frac{1}{\Delta_i} \right) = \lim_{k \rightarrow \infty} \left(\sum_{i=1}^k \frac{1}{x + \Delta_i} - \frac{2}{\pi} \sqrt{t_k} \right) \quad (x > 0) \quad (4.29)$$

is well defined, strictly decreasing on $(0, \infty)$, and tends to $+\infty$ as $x \downarrow 0$ and to $-\infty$ as $x \rightarrow \infty$. We prove these facts, with quantitative bounds, in Theorem 4.15 below. For $\sigma \in \mathbb{R}$, we let $\varrho_\sigma > 0$ be the unique solution of $\mathcal{S}(\varrho_\sigma) = \sigma$. We set

$$R_1^\sigma = \frac{\varrho_\sigma}{2}, \quad R_i^\sigma = \frac{\Delta_i + \varrho_\sigma}{2} \quad (i \geq 2). \quad (4.30)$$

We call the R_i^σ the limiting scaled half-precisions. In the third part of Lemma 4.17 below, we prove that the scaled half-precisions $N^{2/3}r_i$ converge to them in distribution, jointly for finitely many i . The map $\sigma \mapsto \varrho_\sigma$ is a continuous strictly decreasing bijection of $\mathbb{R}$ onto $(0, \infty)$. At $\sigma = 0$, the root ϱ_0 is the quantity $\Delta(\chi)$ of [62, Proposition 1.4(b)], which describes the Gibbs measure at the critical temperature.

Theorem 4.15 (The GOE Airy point process and the limiting equation (cf. [62, Proposition 1.4])). *Let $C_A(T) = \#\{i : \chi_i \leq T\}$ and $\mathcal{E}(T) = C_A(T) - 2T^{3/2}/(3\pi)$. There exists an event of probability one on which the following hold, with random constants C_ω (and $C_{\omega,k}$, depending also on k).*

1. $|\mathcal{E}(T)| \leq C_\omega(1+T)^{7/8}$ for all $T \geq 0$, and $|\chi_k - t_k| \leq C_\omega t_k^{3/8}$ for all $k \geq 1$. In particular, $C_\omega^{-1}(i-1)^{2/3} \leq \Delta_i \leq C_\omega(i-1)^{2/3}$ for $i \geq 2$, and $\sum_{i \geq 2} \Delta_i^{-2} < \infty$.
2. The series in (4.29) converges locally uniformly on $(0, \infty)$, together with its derivative, both expressions in (4.29) agree, and $\mathcal{S}$ is continuously differentiable and strictly decreasing, with $\mathcal{S}(x) \rightarrow +\infty$ as $x \downarrow 0$ and $\mathcal{S}(x) \rightarrow -\infty$ as $x \rightarrow \infty$. For $x > \max(0, \chi_1)$ and $t = x - \chi_1$,

$$\mathcal{S}(x) = -\sqrt{t} + \int_0^\infty \frac{\mathcal{E}(u)}{(t+u)^2} du + B_\omega(t), \quad B_\omega(t) = \sum_{i: \chi_i < 0} \left(\frac{1}{t + \chi_i} - \frac{1}{t} \right). \quad (4.31)$$

3. As $h \downarrow 0$ and $x \rightarrow \infty$, respectively, and for each fixed $k > 3/2$,

$$\begin{aligned} \sum_{i \geq 1} e^{-h\Delta_i} &= \frac{1 + O_\omega(h^{5/8})}{2\sqrt{\pi} h^{3/2}}, & -\mathcal{S}'(x) &= \frac{1 + O_\omega(x^{-5/8})}{2\sqrt{x}}, \\ \sum_{i \geq 1} \frac{1}{(x + \Delta_i)^k} &= C_k x^{3/2-k} (1 + O_{\omega,k}(x^{-5/8})), \end{aligned} \quad (4.32)$$

where $C_k = \pi^{-1}B(3/2, k - 3/2)$.

4. As $S \rightarrow \infty$, we have $\varrho_{-S} = S^2 + O_\omega(S^{3/4})$ on this event. Moreover, with probability over the GOE Airy point process,

$$\varrho_{-S} = S^2 + \chi_1 + O_{\mathbb{P}}(S^{3/4}) \quad (S \rightarrow \infty). \quad (4.33)$$

Used in: The high-temperature and low-temperature limits of the critical profile.

Proof. Let $a = 2/(3\pi)$. We first transfer the simultaneous location bound to the Airy limit and derive its counting asymptotics. We then use Stieltjes integration to study $\mathcal{S}$, the traces, and the root.

Step 1: the points and the counting function. Fix $\eta \in (0, 1)$ and use Lemma 4.3. For each fixed L , the vector $(N^{2/3}(2 - \lambda_i))_{i \leq L}$ converges in law to $(\chi_i)_{i \leq L}$, and $q_{i,N} = N^{2/3}(2 - \gamma_i) \rightarrow t_i$. Hence, for every $\varepsilon > 0$ and all large N ,

$$\mathbb{P}(|N^{2/3}(2 - \lambda_i) - t_i| \leq C_\eta(1 + i)^{1/4} + \varepsilon \text{ for } 1 \leq i \leq L) \geq 1 - \eta.$$

The set defined by these inequalities is closed. The portmanteau theorem, followed by $\varepsilon \downarrow 0$ and then $L \rightarrow \infty$, gives

$$\mathbb{P}(|\chi_i - t_i| \leq C_\eta(1 + i)^{1/4} \text{ for all } i \geq 1) \geq 1 - \eta.$$

Taking the union of these events for $\eta = 2^{-j}$, $j \geq 1$, gives an event of probability one on which $|\chi_i - t_i| \leq C_\omega t_i^{3/8}$ for all i . In particular, $\chi_i/t_i \rightarrow 1$. Intersect this event with the probability-one events of Lemmas 4.2 and 4.5 on which the χ_i are distinct, tend to infinity, and Ξ exists.

For large T , write $k = C_A(T)$, so that $\chi_k \leq T < \chi_{k+1}$. The point asymptotics first imply $t_k \asymp T \asymp t_{k+1}$, and the point bounds then give

$$t_k \leq T + C_\omega T^{3/8}, \quad t_{k+1} \geq T - C_\omega T^{3/8}.$$

Using $k = at_k^{3/2}$ and $k+1 = at_{k+1}^{3/2}$ in these inequalities proves $|k - aT^{3/2}| \leq C_\omega T^{7/8}$. The counting function is finite on bounded intervals, so enlarging C_ω proves $|\mathcal{E}(T)| \leq C_\omega(1 + T)^{7/8}$ for every $T \geq 0$.

Step 2: the gaps. The point bound gives $\Delta_i = \chi_i - \chi_1 \asymp_\omega i^{2/3}$ for large i . The finitely many remaining Δ_i , $i \geq 2$, are positive, so the two-sided bound in the first part follows, as does $\sum_{i \geq 2} \Delta_i^{-2} < \infty$. For all sufficiently large k , we also have

$$\sqrt{\chi_k} - \sqrt{t_k} = \frac{\chi_k - t_k}{\sqrt{\chi_k} + \sqrt{t_k}} = O_\omega(t_k^{-1/8}) \rightarrow 0. \quad (4.34)$$

Step 3: the function $\mathcal{S}$. For $x > 0$ and $i \geq 2$, we have $|(x + \Delta_i)^{-1} - \Delta_i^{-1}| \leq x\Delta_i^{-2}$ and $|\partial_x(x + \Delta_i)^{-1}| \leq \Delta_i^{-2}$. By the first part, the series in (4.29) and its termwise derivative converge uniformly on compact subsets of $(0, \infty)$, and $\mathcal{S}$ is continuously differentiable with

$$-\mathcal{S}'(x) = \sum_{i \geq 1} (x + \Delta_i)^{-2} > 0.$$

The two expressions in (4.29) agree because the partial sums of the second expression equal $x^{-1} + \sum_{i=2}^k ((x + \Delta_i)^{-1} - \Delta_i^{-1}) + (\sum_{i=2}^k \Delta_i^{-1} - 2\sqrt{t_k}/\pi)$, and the last bracket tends to Ξ by (4.5). As $x \downarrow 0$, we have $\mathcal{S}(x) \rightarrow +\infty$, because the term x^{-1} tends to infinity and the series stays bounded.

Now fix $x > \max(0, \chi_1)$ and let $t = x - \chi_1 > 0$, so that $x + \Delta_i = t + \chi_i > 0$ for all i . For a real cutoff $T > 0$, we have the exact identity

$$\sum_{i: \chi_i \leq T} \frac{1}{t + \chi_i} = \frac{C_A(T)}{t + T} + \int_0^T \frac{C_A(u)}{(t + u)^2} du + B_\omega(t). \quad (4.35)$$

Indeed, a point $\chi_i \in [0, T]$ contributes $(t + T)^{-1} + \int_{\chi_i}^T (t + u)^{-2} du = (t + \chi_i)^{-1}$ to the first two terms on the right, and a point $\chi_i < 0$ contributes $(t + T)^{-1} + \int_0^T (t + u)^{-2} du = t^{-1}$, which $B_\omega(t)$ corrects to $(t + \chi_i)^{-1}$. Inserting $C_A(u) = au^{3/2} + \mathcal{E}(u)$ and integrating by parts in the continuum term, we obtain

$$\frac{aT^{3/2}}{t + T} + \int_0^T \frac{au^{3/2}}{(t + u)^2} du = \frac{1}{\pi} \int_0^T \frac{\sqrt{u}}{t + u} du = \frac{2}{\pi} \left(\sqrt{T} - \sqrt{t} \arctan \sqrt{T/t} \right),$$

where we used $3a/2 = 1/\pi$. We obtain

$$\sum_{i: \chi_i \leq T} \frac{1}{t + \chi_i} - \frac{2}{\pi} \sqrt{T} = -\frac{2}{\pi} \sqrt{t} \arctan \sqrt{T/t} + \frac{\mathcal{E}(T)}{t + T} + \int_0^T \frac{\mathcal{E}(u)}{(t + u)^2} du + B_\omega(t).$$

As $T \rightarrow \infty$, the right side tends to the right side of (4.31), since $\mathcal{E}(T)/(t + T) = O_\omega(T^{-1/8})$ and the integral converges absolutely by Step 1. On the left side, for $\chi_k \leq T < \chi_{k+1}$, the sum equals $\sum_{i \leq k} (x + \Delta_i)^{-1}$, and $2(\sqrt{T} - \sqrt{t_k})/\pi \rightarrow 0$ by (4.34) applied at k and $k + 1$, together with $\sqrt{t_{k+1}} - \sqrt{t_k} \rightarrow 0$. By the second expression in (4.29), the left side tends to $\mathcal{S}(x)$, and (4.31) follows. For $t \geq 1$, Step 1 gives

$$\left| \int_0^\infty \frac{\mathcal{E}(u)}{(t + u)^2} du \right| \leq C_\omega \int_0^\infty \frac{(1 + u)^{7/8}}{(t + u)^2} du \leq C_\omega t^{-1/8},$$

and $|B_\omega(t)| \leq C_\omega t^{-2}$ for $t \geq 2|\chi_1|$, since only finitely many points are negative and all are at least χ_1 . Combining these bounds with (4.31), we find $\mathcal{S}(x) = -\sqrt{x - \chi_1} + O_\omega((x - \chi_1)^{-1/8}) \rightarrow -\infty$. This completes the proof of the second part.

Step 4: traces. For $h \in (0, 1]$, the same bookkeeping as in (4.35) shows that

$$\sum_{i \geq 1} e^{-h\chi_i} = h \int_0^\infty e^{-hu} C_A(u) du + \sum_{i: \chi_i < 0} (e^{-h\chi_i} - 1).$$

The leading part is $ah \int_0^\infty e^{-hu} u^{3/2} du = a\Gamma(5/2)h^{-3/2} = (2\sqrt{\pi})^{-1}h^{-3/2}$, the part coming from $\mathcal{E}$ is at most $C_\omega h \int_0^\infty e^{-hu} (1 + u)^{7/8} du \leq C_\omega h^{-7/8}$, and the finite sum is $O_\omega(h)$. Multiplying by $e^{h\chi_1} = 1 + O_\omega(h)$, we obtain the first formula in (4.32). For $k > 3/2$ and $t = x - \chi_1 \geq 1$, similarly,

$$\sum_{i \geq 1} \frac{1}{(t + \chi_i)^k} = \int_0^\infty \frac{k C_A(u)}{(t + u)^{k+1}} du + \sum_{i: \chi_i < 0} \left(\frac{1}{(t + \chi_i)^k} - \frac{1}{t^k} \right).$$

The leading part is $\pi^{-1} \int_0^\infty \sqrt{u} (t+u)^{-k} du = \pi^{-1} t^{3/2-k} B(3/2, k-3/2) = C_k t^{3/2-k}$, the part coming from $\mathcal{E}$ is at most $C_{\omega,k} t^{7/8-k}$, and the finite sum is $O_{\omega,k}(t^{-k-1})$. The third formula follows, because $x + \Delta_i = t + \chi_i$ and $t^{3/2-k} = x^{3/2-k}(1 + O_{\omega,k}(x^{-1}))$. The second formula is the case $k = 2$, since $-\mathcal{S}'(x) = \sum_i (x + \Delta_i)^{-2}$ and $C_2 = \pi^{-1} B(3/2, 1/2) = 1/2$.

Step 5: the root. Let $x = \varrho_{-S}$ and $t = x - \chi_1$. We have $x \rightarrow \infty$ as $S \rightarrow \infty$, because $\mathcal{S}$ decreases from $+\infty$ to $-\infty$. From the end of Step 3, we have $\sqrt{t} = S + O_\omega(t^{-1/8})$. Then $t/S^2 \rightarrow 1$ and $\sqrt{t} - S = O_\omega(S^{-1/4})$, and multiplying by $\sqrt{t} + S \leq 3S$, we obtain $t - S^2 = O_\omega(S^{3/4})$, that is, $\varrho_{-S} = S^2 + \chi_1 + O_\omega(S^{3/4}) = S^2 + O_\omega(S^{3/4})$.

For (4.33), the pathwise bound just proved and the finiteness of χ_1 imply that $S^{-3/4}(\varrho_{-S} - S^2 - \chi_1)$ is bounded eventually on each sample path. The root is continuous in S , so its eventual bounds can be tested on rational S and then localized by measurable events at any fixed confidence. This gives the asserted $O_P(S^{3/4})$ estimate. $\square$

The half-precisions at finite N . The critical sections use the following deterministic conditions on the half-precisions. Recall from (3.9) that the Gibbs measure is $\pi \propto \exp(-\sum_i r_i x_i^2) \nu_N$, where $r_i = p_i/2$.

Definition 4.16 (Critical spectral conditions). Let $0 < c_0 \leq C_0$ and $\delta > 0$. A sequence $0 < r_1 \leq r_2 \leq \dots \leq r_N$ satisfies the critical spectral conditions with constants (c_0, C_0, δ) if

$$\sum_{i=1}^N \frac{1}{2r_i} = N, \quad c_0 \left(\frac{i}{N}\right)^{2/3} \leq r_i \leq C_0 \left(\frac{i}{N}\right)^{2/3} \quad (1 \leq i \leq N), \quad r_2 - r_1 \geq \frac{\delta}{N^{2/3}}. \quad (4.36)$$

For $M > 0$, $N > (2M)^3$, and $|\sigma| \leq M$, let $c_{N,\sigma} = 1 + \sigma N^{-1/3} \in [1/2, 3/2]$. Let $z_{N,\sigma}$ be the Lagrange multiplier of $c_{N,\sigma} W_N$, let $d_{N,\sigma} = z_{N,\sigma} - c_{N,\sigma} \lambda_1$ be the smallest precision, and let

$$r_{i,N,\sigma} = \frac{z_{N,\sigma} - c_{N,\sigma} \lambda_i}{2}, \quad \varrho_{N,\sigma} = \frac{N^{2/3} d_{N,\sigma}}{c_{N,\sigma}}.$$

Then $r_{1,N,\sigma} = d_{N,\sigma}/2$ and $N^{2/3} r_{i,N,\sigma} = c_{N,\sigma} (\varrho_{N,\sigma} + \Delta_{i,N})/2$, where $\Delta_{1,N} = 0$. The condition $N > (2M)^3$ ensures $c_{N,\sigma} > 0$. Without such a condition the analysis below fails at small N : if $c_{N,\sigma} < 0$, then $c_{N,\sigma} \lambda_1$ is the smallest of the numbers $c_{N,\sigma} \lambda_i$, and the half-precisions are no longer ordered as above.

Lemma 4.17 (Half-precisions in the critical window (cf. [62, Proposition 4.1, (4.2)])). Fix $M > 0$, and let $N > (2M)^3$.

1. For $|\sigma| \leq M$, the root $\varrho_{N,\sigma}$ is the unique positive solution of $\mathcal{S}_N(\varrho_{N,\sigma}) = \sigma$, where

$$\mathcal{S}_N(x) = \Xi_N + \frac{1}{x} + \sum_{i=2}^N \left(\frac{1}{x + \Delta_{i,N}} - \frac{1}{\Delta_{i,N}} \right), \quad (4.37)$$

and Ξ_N is as in Lemma 4.5. The map $\sigma \mapsto \varrho_{N,\sigma}$ is strictly decreasing.

2. For every $\eta \in (0, 1)$, there exist $0 < c_0 \leq C_0$ and $\delta > 0$, depending only on η and M , and events $\mathcal{A}_N$ with $\liminf_N \mathbb{P}(\mathcal{A}_N) \geq 1 - \eta$, on which, for every $\sigma \in [-M, M]$ simultaneously, the sequence $(r_{i,N,\sigma})_{i=1}^N$ satisfies the critical spectral conditions (4.36) with constants (c_0, C_0, δ) .
3. Let $\sigma_1, \dots, \sigma_m \in [-M, M]$. Let $\mathbf{v}_{N,\sigma} = ((N^{2/3}r_{i,N,\sigma})^{-1})_{i \geq 1}$, padded with zeros for $i > N$, and $\mathbf{v}_\sigma = ((R_i^\sigma)^{-1})_{i \geq 1}$, with R_i^σ as in (4.30), where q_σ solves

$$\Xi + \frac{1}{x} - \sum_{i \geq 2} \frac{x \Delta_i^{-2}}{1 + x \Delta_i^{-1}} = \sigma. \quad (4.38)$$

Then, jointly in distribution,

$$(\Xi_N, \mathbf{b}_N, (q_{N,\sigma_j})_{j \leq m}, (\mathbf{v}_{N,\sigma_j})_{j \leq m}) \implies (\Xi, (\Delta_i^{-1})_{i \geq 2}, (q_{\sigma_j})_{j \leq m}, (\mathbf{v}_{\sigma_j})_{j \leq m})$$

in $\mathbb{R} \times \ell^2 \times \mathbb{R}^m \times (\ell^2)^m$, with $\mathbf{b}_N$ as in Lemma 4.5. All limits are functions of the same GOE Airy point process.

Used in: The functional inequalities, density estimates, critical process, critical profile.

The left side of (4.38) is $\mathcal{S}(x)$, since $(x + \Delta_i)^{-1} - \Delta_i^{-1} = -x\Delta_i^{-2}/(1 + x\Delta_i^{-1})$; thus q_σ is the root defined before Theorem 4.15. Similarly, $\mathcal{S}_N$ is $\mathcal{S}$ with the GOE Airy point process replaced by the scaled eigenvalues.

Proof. We write $c = c_{N,\sigma}$ and $q_N = q_{N,\sigma}$ when σ is fixed. The proof has three steps: the finite equation, convergence through a deterministic continuous map, and the bounds at fixed confidence.

Step 1: the finite equation. We have $z_{N,\sigma} - c\lambda_i = cN^{-2/3}(q_N + \Delta_{i,N})$, and the equation (3.12) reads $\sum_{i=1}^N (q_N + \Delta_{i,N})^{-1} = cN^{1/3} = N^{1/3} + \sigma$. Subtracting $\sum_{i \geq 2} \Delta_{i,N}^{-1} = \Xi_N + N^{1/3}$, which is finite because the eigenvalues are almost surely distinct, we obtain $\mathcal{S}_N(q_N) = \sigma$. The function $\mathcal{S}_N$ is continuous and strictly decreasing on $(0, \infty)$, with limit $+\infty$ at 0 and limit $\Xi_N - \sum_{i \geq 2} \Delta_{i,N}^{-1} = -N^{1/3}$ at ∞ . Since $\sigma \geq -M > -N^{1/3}$, the level σ is attained exactly once, and the root is decreasing in σ . The condition $z_{N,\sigma} > c\lambda_1$ of Definition 3.3 is equivalent to $q_N > 0$.

Step 2: convergence. For $\zeta \in \mathbb{R}$, a sequence $u = (u_i)_{i \geq 2} \in \ell^2$ with $u_i \geq 0$, and $x > 0$, let

$$\Theta(\zeta, u; x) = \zeta + \frac{1}{x} - \sum_{i \geq 2} \frac{xu_i^2}{1 + xu_i},$$

so that $\mathcal{S}_N(x) = \Theta(\Xi_N, \mathbf{b}_N; x)$ and $\mathcal{S}(x) = \Theta(\Xi, (\Delta_i^{-1})_{i \geq 2}; x)$. For $0 < x \leq L$ and $u_i, w_i \geq 0$, the derivative of $v \mapsto xv^2/(1 + xv)$ is $xv(2 + xv)/(1 + xv)^2 \leq 2xv$. Integrating this bound between w_i and u_i , we obtain

$$\left| \frac{xu_i^2}{1 + xu_i} - \frac{xw_i^2}{1 + xw_i} \right| \leq 2L(u_i + w_i)|u_i - w_i|.$$

Summing and using the Cauchy–Schwarz inequality, we obtain

$$\sup_{0 < x \leq L} |\Theta(\xi, u; x) - \Theta(\xi', w; x)| \leq |\xi - \xi'| + 2L(\|u\|_2 + \|w\|_2)\|u - w\|_2.$$

By Lemma 4.5 and the Skorokhod representation theorem [30, Theorem 6.7], we may assume that $(\Xi_N, \mathbf{b}_N) \rightarrow (\Xi, (\Delta_i^{-1})_{i \geq 2})$ almost surely in $\mathbb{R} \times \ell^2$, where the limit is built from a GOE Airy point process. Then $\mathcal{S}_N \rightarrow \mathcal{S}$ uniformly on compact subsets of $(0, \infty)$, almost surely. Let $x_- = \varrho_M/2$ and $x_+ = 2\varrho_{-M}$, so that $\mathcal{S}(x_-) > M$ and $\mathcal{S}(x_+) < -M$. For N large, we have $\mathcal{S}_N(x_-) > M$ and $\mathcal{S}_N(x_+) < -M$, and by monotonicity all roots $\varrho_{N,\sigma}$, $|\sigma| \leq M$, lie in (x_-, x_+) . For $\varepsilon \in (0, x_-/2)$, the quantity

$$\gamma_\varepsilon = \min_{|\sigma| \leq M} \min \left(\mathcal{S}(\varrho_\sigma - \varepsilon) - \sigma, \sigma - \mathcal{S}(\varrho_\sigma + \varepsilon) \right)$$

is positive, by continuity and compactness. Once $\sup_{[x_-/2, 2x_+]} |\mathcal{S}_N - \mathcal{S}| < \gamma_\varepsilon$, we have $|\varrho_{N,\sigma} - \varrho_\sigma| < \varepsilon$ for all $|\sigma| \leq M$, by monotonicity. We conclude that $\varrho_{N,\sigma} \rightarrow \varrho_\sigma$ almost surely, uniformly for $|\sigma| \leq M$. Finally,

$$(N^{2/3}r_{1,N,\sigma})^{-1} = \frac{2}{c \varrho_{N,\sigma}}, \quad (N^{2/3}r_{i,N,\sigma})^{-1} = \frac{2}{c} \frac{\Delta_{i,N}^{-1}}{1 + \varrho_{N,\sigma} \Delta_{i,N}^{-1}} \quad (i \geq 2).$$

The map $v \mapsto v/(1+xv)$ is 1-Lipschitz on $[0, \infty)$, and $|v/(1+xv) - v/(1+x'v)| \leq v^2|x - x'|$. These two bounds show that the map $(x, u) \mapsto (u_i/(1+xu_i))_{i \geq 2}$ from $(0, \infty) \times \ell_+^2$ to ℓ^2 is continuous, with modulus bounded by $\|u - w\|_2 + \|u\|_4^2|x - x'|$. Since $c_{N,\sigma} \rightarrow 1$, we conclude that $\mathbf{v}_{N,\sigma} \rightarrow \mathbf{v}_\sigma$ in ℓ^2 almost surely, for each σ . This proves the third part.

Step 3: bounds at fixed confidence. Fix $\eta \in (0, 1)$. By Lemma 4.6 with confidence $\eta/2$, there exist $0 < c_1 \leq C_1$ and events of limiting probability at least $1 - \eta/2$ on which

$$c_1(i-1)^{2/3} \leq \Delta_{i,N} \leq C_1(i-1)^{2/3} \quad (2 \leq i \leq N). \quad (4.39)$$

These events and constants do not depend on σ . By Step 2, $(\varrho_{N,M}, \varrho_{N,-M}) \Rightarrow (\varrho_M, \varrho_{-M})$, and the limit is almost surely in $(0, \infty)^2$. Let $0 < \varrho_{\min} < \varrho_{\max}$ be such that $\mathbb{P}(\varrho_M \leq \varrho_{\min}) + \mathbb{P}(\varrho_{-M} \geq \varrho_{\max}) \leq \eta/2$. By the portmanteau theorem, $\limsup_N \mathbb{P}(\varrho_{N,M} \leq \varrho_{\min} \text{ or } \varrho_{N,-M} \geq \varrho_{\max}) \leq \eta/2$. Let $\mathcal{A}_N$ be the intersection of the complement of this event with the events above. On $\mathcal{A}_N$, by the monotonicity in Step 1, $\varrho_{\min} \leq \varrho_{N,\sigma} \leq \varrho_{\max}$ for all $|\sigma| \leq M$, and since $1/2 \leq c_{N,\sigma} \leq 3/2$,

$$\begin{aligned} \frac{\varrho_{\min}}{4} &\leq N^{2/3}r_{1,N,\sigma} \leq \frac{3\varrho_{\max}}{4}, \\ \frac{\varrho_{\min} + c_1(i-1)^{2/3}}{4} &\leq N^{2/3}r_{i,N,\sigma} \leq \frac{3(\varrho_{\max} + C_1(i-1)^{2/3})}{4} \quad (i \geq 2). \end{aligned}$$

Using $2^{-2/3}i^{2/3} \leq (i-1)^{2/3} \leq i^{2/3}$ for $i \geq 2$, we deduce from these bounds that $c_0(i/N)^{2/3} \leq r_{i,N,\sigma} \leq C_0(i/N)^{2/3}$ for all i , with $c_0 = \min(\varrho_{\min}, c_1)/8$ and $C_0 = 3(\varrho_{\max} + C_1)/4$. For the first gap, $N^{2/3}(r_{2,N,\sigma} - r_{1,N,\sigma}) = c_{N,\sigma}\Delta_{2,N}/2 \geq c_1/4$, which allows $\delta = c_1/4$. The half-precisions are nondecreasing in i because the eigenvalues are, and $\sum_i (2r_{i,N,\sigma})^{-1} = N$ holds by the definition of the Lagrange multiplier. This proves the second part. All constants are chosen before $N \rightarrow \infty$, and they do not depend on $\sigma \in [-M, M]$. $\square$

At $\sigma = 0$, the convergence in distribution of $\mathbf{v}_{N,0}$ is equivalent to [62, Proposition 4.1, (4.2)], whose coefficients $(N^{2/3}p_i)^{-1}$ are half of ours. There it is used to describe the Gibbs measure at the critical temperature as a Gaussian sequence conditioned on a renormalized radial constraint. The extension to bounded σ and the uniform bounds of the second part are the deterministic arguments above.

PART II

Tools

Gaussian laws conditioned on a sphere

This section and the next collect deterministic estimates for a centered Gaussian vector with independent coordinates, conditioned on the event that its squared norm equals its expectation. We show in Proposition 5.6 that the Gibbs measures π_A of Section 3 are laws of this kind, and so are the Bingham comparison laws μ_t of Section 7. The main results are a quantitative local limit theorem for weighted sums of centered squares of Gaussian variables (Theorem 5.3); a formula for the density of a set of coordinates under the conditioned law, relative to independent Gaussian variables, together with a comparison when these coordinates carry a small fraction of the variance (Lemma 5.9 and Theorem 5.11); moment bounds (Theorem 5.12); an exact formula for quadratic exponential moments (Lemma 5.13); and moment bounds for quadratic statistics (Theorem 5.15 and Corollary 5.16). We treat finite and countably infinite variance sequences together. The countable case is needed in the critical window, where the limiting Gibbs law is a countable Gaussian sequence conditioned on a renormalized radial constraint. We construct this law in Lemma 5.7, identify its finite-coordinate marginals in Lemma 5.9, and prove in Proposition 5.17 that it depends continuously on the variance sequence, with moment bounds that hold uniformly along a converging sequence of finite approximations; Section 18 specializes these results to the critical variances.

The proofs use only the notation of Sections 3 and 4. The results are used in Section 6, in the comparison and moment bounds of Sections 9, 10, and 12, in Part III, and in Section 18. All statements are deterministic. A constant is called universal if it does not depend on the number of coordinates, on the size of the smallest variance, or on any other feature of the variance sequence.

5.1 Notation

A *variance sequence* is a family $v = (v_i)_{i \in I}$ of positive real numbers, indexed by a nonempty finite or countably infinite set I , such that $\sum_{i \in I} v_i^2 < \infty$. We call v finite or countable according to whether I is finite or countably infinite. Let $g = (g_i)_{i \in I}$ be independent standard Gaussian random variables, and define

$$T_v = \sum_{i \in I} v_i(g_i^2 - 1), \quad S_v = \left(2 \sum_{i \in I} v_i^2\right)^{1/2}, \quad \hat{v}_i = \frac{v_i}{S_v}, \quad \eta_v = \max_{i \in I} \hat{v}_i. \quad (5.1)$$

For countable I , the summands of T_v are independent and centered, with variances $2v_i^2$ whose sum is finite. The series converges in L^2 to a limit that does not depend on the order of summation. For any enumeration of I , its partial sums form a martingale bounded in L^2 , which converges almost surely by the martingale convergence theorem. We have $\text{Var } T_v = S_v^2$ and $\sum_i \hat{v}_i^2 = 1/2$, so that $\eta_v \leq 2^{-1/2}$; the maximum in (5.1) is attained because $\hat{v}_i \rightarrow 0$ along any enumeration of a countable I . The quantity η_v is small exactly when no single coordinate carries a macroscopic part of the variance of T_v . For instance, if $v_i = 1$ for $1 \leq i \leq n$, then $S_v = (2n)^{1/2}$ and $\eta_v = (2n)^{-1/2}$.

When T_v has a continuous density, we denote it by p_v , and write $f_v(x) = S_v p_v(S_v x)$ for the density of T_v/S_v . We write $\hat{f}_v(t) = \mathbb{E} e^{itT_v/S_v}$ for the characteristic function of T_v/S_v . The standard Gaussian density and distribution function are ϕ and Φ , and G denotes a standard Gaussian random variable.

For $A \subset I$, we write $v|_A$ for the restriction of v to A , and $T_{v,A} = \sum_{i \in A} v_i(g_i^2 - 1)$ and $p_{v,A}$ for the corresponding sum and its density. We let γ_A be the standard Gaussian law on $\mathbb{R}^A$, the product of standard Gaussian laws on the coordinates, and, for finite A , $\phi_A(z) = \prod_{i \in A} \phi(z_i)$ its density. For $z \in \mathbb{R}^A$ we set

$$U_A(z) = \sum_{i \in A} v_i(z_i^2 - 1), \quad h_A = \frac{\sum_{i \in A} v_i^2}{\sum_{i \in I} v_i^2}, \quad S_{v|_{I \setminus A}} = S_v(1 - h_A)^{1/2}. \quad (5.2)$$

For infinite A , the series $U_A(z)$ is defined for γ_A -almost every z , as the almost sure and $L^2(\gamma_A)$ limit of its partial sums. Note that $T_{v,A} = U_A(g_A)$, where $g_A = (g_i)_{i \in A}$, and h_A is the fraction of the variance of T_v contributed by the coordinates in A .

Under the conditioned laws defined below, the random point is denoted by $Y = (Y_i)_{i \in I}$, and its standardized coordinates are $Z_i = Y_i / \sqrt{v_i}$. For a probability measure μ and a coordinate set A , we write μ^{Z_A} for the law of $Z_A = (Z_i)_{i \in A}$ under μ . For $r > 0$, let $\mathcal{S}_r = \{y \in \mathbb{R}^I : |y|^2 = r\}$ when I is finite, with $(|I| - 1)$ -dimensional surface measure $\mathcal{H}^{|I|-1}$; in particular $\mathcal{S}_N = S_N$ when $I = \{1, \dots, N\}$.

5.2 Densities of weighted sums of centered squares

The following lemma is the basis of all density estimates in this section.

Lemma 5.1 (A differential equation for the characteristic function). *Let v be a variance sequence. The characteristic function $\hat{f}_v$ is continuously differentiable and satisfies*

$$\hat{f}'_v(t) = \omega_v(t) \hat{f}_v(t), \quad \omega_v(t) = -2t \sum_{i \in I} \frac{\hat{v}_i^2}{1 - 2it\hat{v}_i}, \quad \hat{f}_v(0) = 1, \quad (5.3)$$

where the series defining ω_v converges uniformly on bounded intervals. Further,

$$|\hat{f}_v(t)| = \prod_{i \in I} (1 + 4t^2 \hat{v}_i^2)^{-1/4} \leq (1 + 4\eta_v^2 t^2)^{-1/(8\eta_v^2)}, \quad (5.4)$$

and, if $|t|\eta_v \leq 1/4$, then

$$|\hat{f}_v(t) - e^{-t^2/2}| \leq \eta_v |t|^3 e^{-2t^2/5}. \quad (5.5)$$

Proof. We begin with one coordinate. For $b > 0$, let $k_b(t) = \mathbb{E}e^{itb(G^2-1)}$. We may differentiate under the expectation, since $\mathbb{E}G^2 < \infty$. Applying Gaussian integration by parts, $\mathbb{E}[G\psi(G)] = \mathbb{E}[\psi'(G)]$, to $\psi(x) = xe^{itbx^2}$, we find that

$$\mathbb{E}[G^2 e^{itbG^2}] = \mathbb{E}[e^{itbG^2}] + 2itb \mathbb{E}[G^2 e^{itbG^2}], \quad \text{that is,} \quad (1 - 2itb) \mathbb{E}[G^2 e^{itbG^2}] = \mathbb{E}[e^{itbG^2}].$$

Inserting this into $k'_b(t) = ib e^{-itb} (\mathbb{E}[G^2 e^{itbG^2}] - \mathbb{E}[e^{itbG^2}])$, we obtain

$$k'_b(t) = -\frac{2tb^2}{1 - 2itb} k_b(t).$$

Taking the real part of $2\overline{k_b}k'_b$, we find that $|k_b|^2$ solves the real equation

$$\frac{d}{dt}|k_b(t)|^2 = -\frac{4tb^2}{1 + 4t^2b^2}|k_b(t)|^2,$$

and solving this equation with $|k_b(0)| = 1$ gives $|k_b(t)| = (1 + 4t^2b^2)^{-1/4}$. For finite I , the variable T_v/S_v is the sum of the independent variables $\hat{v}_i(g_i^2 - 1)$, and $\hat{f}_v = \prod_i k_{\hat{v}_i}$. Differentiating this finite product yields (5.3) and the identity in (5.4).

Let I be countable, fix an enumeration of I , and let X_n be the n -th partial sum of the series $X = T_v/S_v$. The characteristic functions $\hat{f}_n$ of X_n converge to $\hat{f}_v$ uniformly on bounded intervals, because $X_n \rightarrow X$ in L^2 . Their derivatives $\hat{f}'_n(t) = i\mathbb{E}[X_n e^{itX_n}]$ also converge uniformly for $|t| \leq T$, since

$$\begin{aligned} |\mathbb{E}[X_n e^{itX_n}] - \mathbb{E}[X e^{itX}]| &\leq \mathbb{E}|X_n - X| + \mathbb{E}[|X| |e^{itX_n} - e^{itX}|] \\ &\leq \|X_n - X\|_2 + T\|X\|_2\|X_n - X\|_2. \end{aligned}$$

The summands of ω_v have modulus at most $2T\hat{v}_i^2$ for $|t| \leq T$, and the coefficients of the finite equations converge to ω_v uniformly for $|t| \leq T$. Passing to the limit in the finite equations, we find (5.3); the finite products of moduli converge to the infinite product in (5.4).

For the inequality in (5.4), we use that $a \mapsto a^{-1} \log(1 + ax)$ is nonincreasing on $a > 0$ for every $x \geq 0$, since its derivative has the sign of $ax/(1 + ax) - \log(1 + ax) \leq 0$. Using $\hat{v}_i^2 \leq \eta_v^2$, we find

$$\sum_{i \in I} \log(1 + 4t^2\hat{v}_i^2) \geq \sum_{i \in I} \frac{\hat{v}_i^2}{\eta_v^2} \log(1 + 4\eta_v^2 t^2) = \frac{1}{2\eta_v^2} \log(1 + 4\eta_v^2 t^2),$$

which is the inequality in (5.4). Next, using $\log(1 + x) \geq x/(1 + x)$, we have

$$|\hat{f}_v(t)| \leq \exp \left\{ -\sum_{i \in I} \frac{t^2\hat{v}_i^2}{1 + 4t^2\hat{v}_i^2} \right\} \leq \exp \left\{ -\frac{t^2}{2(1 + 4\eta_v^2 t^2)} \right\} \leq e^{-2t^2/5}, \quad |t|\eta_v \leq 1/4.$$

Since $\sum_i \hat{v}_i^2 = 1/2$, we have

$$|\omega_v(t) + t| = \left| 2t \sum_{i \in I} \hat{v}_i^2 \frac{2it\hat{v}_i}{1 - 2it\hat{v}_i} \right| \leq 4t^2 \sum_{i \in I} \hat{v}_i^3 \leq 2\eta_v t^2.$$

By variation of constants in (5.3), we have, for $t \geq 0$,

$$\hat{f}_v(t) - e^{-t^2/2} = \int_0^t e^{-(t^2-s^2)/2} (\omega_v(s) + s) \hat{f}_v(s) ds.$$

For $0 \leq s \leq t \leq (4\eta_v)^{-1}$, the integrand has modulus at most $2\eta_v s^2 e^{-t^2/2+s^2/2-2s^2/5} \leq 2\eta_v s^2 e^{-2t^2/5}$. Integrating in s gives (5.5) for $t \geq 0$. Negative t follow by complex conjugation. $\square$

Lemma 5.2 (Bounded densities from three coordinates). *Let v be a variance sequence such that I has at least three elements.*

1. *The variable T_v has a bounded continuous density p_v . For any three distinct indices $j_1, j_2, j_3 \in I$,*

$$\|p_v\|_\infty \leq \frac{c_3}{\min\{v_{j_1}, v_{j_2}, v_{j_3}\}}, \quad c_3 = \frac{1}{4\pi} \int_{\mathbb{R}} (1+s^2)^{-3/4} ds. \quad (5.6)$$

2. *We have $p_v(t) > 0$ for every $t > -\sum_i v_i$, where the sum may be infinite. If $\sum_i v_i < \infty$, then $p_v(t) = 0$ for $t \leq -\sum_i v_i$.*
3. *If $A \subset I$ and $I \setminus A$ has at least three elements, then*

$$p_v(t) = \mathbb{E} p_{v, I \setminus A}(t - T_{v,A}) \quad \text{for every } t \in \mathbb{R}. \quad (5.7)$$

Proof. By (5.4) with $t = uS_v$, we have $|\mathbb{E} e^{iuT_v}| = \prod_i (1 + 4u^2 v_i^2)^{-1/4}$. Every factor is at most one. Keeping only the three factors with indices j_1, j_2, j_3 , we have $|\mathbb{E} e^{iuT_v}| \leq (1 + 4u^2 m^2)^{-3/4}$, where $m = \min\{v_{j_1}, v_{j_2}, v_{j_3}\}$. This bound is integrable in u , and Fourier inversion yields a bounded continuous density p_v with $\|p_v\|_\infty \leq (2\pi)^{-1} \int (1 + 4u^2 m^2)^{-3/4} du = c_3/m$, which proves the first part.

For the third part, $T_v = T_{v,A} + T_{v, I \setminus A}$ with independent summands. By Fubini's theorem, the right side of (5.7) is a density of T_v , and it is continuous in t by dominated convergence, since $p_{v, I \setminus A}$ is bounded and continuous. Two continuous densities of the same law agree everywhere, and (5.7) follows.

For the second part, if $\sum_i v_i < \infty$, then $T_v \geq -\sum_i v_i$ and p_v vanishes on $(-\infty, -\sum_i v_i)$, and by continuity also at $-\sum_i v_i$. Now let $t > -\sum_i v_i$, and choose a finite set $B \subset I$ with at least three elements and $\sum_{i \in B} v_i > -t$. By [62, Lemma 2.2], the variable $\sum_{i \in B} v_i g_i^2$ has a continuous density that is positive on $(0, \infty)$. Equivalently, $p_{v,B}$ is positive on $(-\sum_{i \in B} v_i, \infty)$. By (5.7) with $A = I \setminus B$, we have

$$p_v(t) = \mathbb{E} p_{v,B}(t - T_{v, I \setminus B}) \geq \mathbb{E} [p_{v,B}(t - T_{v, I \setminus B}) \mathbf{1}\{T_{v, I \setminus B} \leq 0\}] > 0.$$

In the last step, we used that $t - T_{v,I \setminus B} \geq t > -\sum_{i \in B} v_i$ on the event $\{T_{v,I \setminus B} \leq 0\}$, and that this event has positive probability because $T_{v,I \setminus B}$ has mean zero (it is identically zero if $B = I$). $\square$

Theorem 5.3 (A quantitative local limit theorem). *For each integer $j \geq 0$, there exist universal constants $\eta_j > 0$ and $C_j < \infty$ with the following property. If v is a variance sequence with $\eta_v \leq \eta_j$, then f_v has j continuous derivatives and*

$$\sup_{x \in \mathbb{R}} |f_v^{(j)}(x) - \phi^{(j)}(x)| \leq C_j \eta_v. \quad (5.8)$$

Further, there exist universal constants $\eta_{\text{loc}}, c_0 > 0$ and $C < \infty$ such that $\eta_v \leq \eta_{\text{loc}}$ implies

$$f_v(0) \geq c_0, \quad \|f_v\|_\infty + \|f_v''\|_\infty \leq C, \quad |f_v'(0)| \leq C \eta_v, \quad \int_{\mathbb{R}} t^2 |\hat{f}_v(t)| dt \leq C. \quad (5.9)$$

The constants do not depend on the number of coordinates, the smallest variance, or the relative sizes of the variances, and the theorem applies to finite and countable sequences alike. Note that $\eta_v \leq \eta_j$ forces I to have at least $(2\eta_j^2)^{-1}$ elements.

Proof. Fix $j \geq 0$ and let $\eta_j^2 \leq a_j$, where $a_j = 1/(8(j+4))$. The function $a \mapsto (1 + 4at^2)^{-1/(8a)}$ is nondecreasing on $a > 0$, by the monotonicity used in the proof of Lemma 5.1. Then (5.4) implies that, for $\eta_v \leq \eta_j$,

$$|\hat{f}_v(t)| \leq (1 + 4a_j t^2)^{-1/(8a_j)} = (1 + 4a_j t^2)^{-(j+4)},$$

and $|t|^j |\hat{f}_v(t)|$ is integrable. By Fourier inversion, f_v exists, is j times continuously differentiable, and

$$f_v^{(j)}(x) - \phi^{(j)}(x) = \frac{1}{2\pi} \int_{\mathbb{R}} (-it)^j e^{-itx} (\hat{f}_v(t) - e^{-t^2/2}) dt.$$

We bound the integral of $|t|^j |\hat{f}_v(t) - e^{-t^2/2}|$ over two regions. On $|t| \leq (4\eta_v)^{-1}$, (5.5) bounds it by $\eta_v \int |t|^{j+3} e^{-2t^2/5} dt = C_j \eta_v$. On $|t| > (4\eta_v)^{-1}$, we set $u = 2\eta_v |t| \geq 1/2$ and split the exponent in (5.4) into two equal halves:

$$|\hat{f}_v(t)| \leq (1 + u^2)^{-1/(16\eta_v^2)} (1 + u^2)^{-1/(16\eta_v^2)} \leq e^{-\log(5/4)/(16\eta_v^2)} (1 + u^2)^{-(j+4)/2},$$

where we used $u^2 \geq 1/4$ in the first factor and $1/(16\eta_v^2) \geq (j+4)/2$ in the second. Changing variables, we find

$$\int_{|t| > (4\eta_v)^{-1}} |t|^j |\hat{f}_v(t)| dt \leq e^{-\log(5/4)/(16\eta_v^2)} (2\eta_v)^{-j-1} \int_{\mathbb{R}} |u|^j (1 + u^2)^{-(j+4)/2} du \leq C_j \eta_v,$$

since $x^{-j-2} e^{-\log(5/4)/(16x^2)}$ is bounded on $x > 0$. The Gaussian tail $\int_{|t| > (4\eta_v)^{-1}} |t|^j e^{-t^2/2} dt$ is also at most $C_j \eta_v$. This proves (5.8), and also $\int |t|^j |\hat{f}_v(t) - e^{-t^2/2}| dt \leq C_j \eta_v$.

For (5.9), let $\eta_{\text{loc}} \leq \min\{\eta_0, \eta_1, \eta_2\}$ be so small that $C_0\eta_{\text{loc}} \leq \phi(0)/2$. Using (5.8) with $j = 0, 1, 2$ and $\phi'(0) = 0$, we have $f_v(0) \geq \phi(0)/2$, the bounds on $\|f_v\|_\infty$ and $\|f_v''\|_\infty$, and $|f_v'(0)| \leq C_1\eta_v$. Finally, $\int t^2|\hat{f}_v| \leq \int t^2 e^{-t^2/2} dt + C_2\eta_v \leq C$ by the last bound of the previous paragraph. $\square$

5.3 Conditioning on the sphere: finite sequences

Definition 5.4 (Gaussian law conditioned on a sphere). Let v be a finite variance sequence and $R^2 = \sum_{i \in I} v_i$. The *Gaussian law* μ_v conditioned on the sphere is the probability measure on $\mathcal{S}_{R^2}$ given by

$$\mu_v(dy) = \frac{1}{\mathcal{Z}_v} \exp \left\{ -\frac{1}{2} \sum_{i \in I} \frac{y_i^2}{v_i} \right\} \mathcal{H}^{|I|-1}(dy), \quad y \in \mathcal{S}_{R^2}, \quad (5.10)$$

where $\mathcal{Z}_v$ is the normalizing constant.

This is a Bingham distribution [31]: a probability measure on a sphere whose density with respect to surface measure is the exponential of a quadratic form. The next lemma justifies the name: μ_v is the law of $Y = (\sqrt{v_i} g_i)_{i \in I}$ conditioned on $|Y|^2 = R^2$, that is, on $T_v = 0$. This description is classical; see [98, Abstract] for its use in the computation of Bingham normalizing constants, [62, Section 1.2, (1.7)–(1.8)] for the case of the spherical Sherrington–Kirkpatrick model at the critical temperature, and [5, Proposition 2.7] for a lattice spherical model.

Lemma 5.5 (Radial densities). Let v be a finite variance sequence, $n = |I|$, $R^2 = \sum_i v_i$, and $Y = (\sqrt{v_i} g_i)_{i \in I}$, with density $\gamma_v(y) = \prod_i (2\pi v_i)^{-1/2} e^{-y_i^2/(2v_i)}$. For a bounded nonnegative measurable function F , or a continuous real-valued function F , on $\mathbb{R}^I$, and for $t > -R^2$, let

$$p_v[F](t) = \frac{1}{2(R^2 + t)^{1/2}} \int_{\mathcal{S}_{R^2+t}} F(y) \gamma_v(y) \mathcal{H}^{n-1}(dy). \quad (5.11)$$

1. For bounded nonnegative measurable F on $\mathbb{R}^I$ and nonnegative measurable K on $\mathbb{R}$, we have $\mathbb{E}[F(Y)K(T_v)] = \int_{-R^2}^\infty K(t) p_v[F](t) dt$.
2. If F is continuous, then $p_v[F]$ is finite and continuous on $(-R^2, \infty)$. In particular, $p_v[1]$ is a continuous and positive version of the density of T_v on $(-R^2, \infty)$.
3. For every continuous F , we have $\mu_v(F) = p_v[F](0) / p_v[1](0)$. For every bounded continuous F , we have $\mathbb{E}[F(Y) \mid |T_v| \leq \varepsilon] \rightarrow \mu_v(F)$ as $\varepsilon \downarrow 0$.

Proof. In polar coordinates, $\int_{\mathbb{R}^n} H(y) dy = \int_0^\infty \int_{\mathcal{S}_{r^2}} H d\mathcal{H}^{n-1} dr$ for nonnegative measurable H . Substituting $s = r^2$, we find that $\int_{\mathbb{R}^n} H(y) dy = \int_0^\infty (2\sqrt{s})^{-1} \int_{\mathcal{S}_s} H d\mathcal{H}^{n-1} ds$. With

$H(y) = F(y)\gamma_v(y)K(|y|^2 - R^2)$ and $t = s - R^2$, this is the first part, since $T_v = |Y|^2 - R^2$. For the second part, by scaling the sphere, we have

$$\int_{\mathcal{S}_{R^2+t}} F\gamma_v \, d\mathcal{H}^{n-1} = (R^2 + t)^{(n-1)/2} \int_{\mathcal{S}_1} F((R^2 + t)^{1/2}\theta) \gamma_v((R^2 + t)^{1/2}\theta) \mathcal{H}^{n-1}(d\theta).$$

The integrand is jointly continuous in (t, θ) and the unit sphere is compact, so the integral is continuous in $t > -R^2$. Taking $F = 1$ in the first part, we see that $p_v[1]$ is a density of T_v ; it is positive because $\gamma_v > 0$.

For the third part, recall from (5.10) that the density of μ_v with respect to surface measure is proportional to γ_v on $\mathcal{S}_{R^2}$, which is the identity $\mu_v(F) = p_v[F](0)/p_v[1](0)$. Finally, for bounded continuous F , the second part implies that

$$\mathbb{E}[F(Y) \mid |T_v| \leq \varepsilon] = \frac{\int_{-\varepsilon}^{\varepsilon} p_v[F](t) \, dt}{\int_{-\varepsilon}^{\varepsilon} p_v[1](t) \, dt} \longrightarrow \frac{p_v[F](0)}{p_v[1](0)} = \mu_v(F),$$

as $\varepsilon \downarrow 0$. □

For a nonnegative continuous F that is unbounded on $\mathbb{R}^I$, bounded radial truncation justifies each later use of the coarea identity: on any compact interval $J \subset (-R^2, \infty)$, multiply F by a continuous compactly supported function of $|y|^2$ that equals one on the radii $R^2 + J$. The resulting test is bounded and agrees with F on the relevant spheres. Thus the first part applies on J ; increasing such intervals gives any nonnegative integral identity used below. The exact-level identity already applies directly to continuous polynomial and exponential tests.

For every finite variance sequence, we define $p_v(t) = p_v[1](t)$ for $t > -R^2$ and $p_v(t) = 0$ for $t \leq -R^2$; this is a density of T_v . When I has at least three elements, it agrees everywhere with the continuous density of Lemma 5.2, since both are densities of T_v that are continuous on $(-R^2, \infty)$ and vanish on $(-\infty, -R^2]$. When I has one or two elements, p_v is continuous only on $(-R^2, \infty)$: for one variance v_1 , it equals $(2\pi v_1(t + v_1))^{-1/2} e^{-(t+v_1)/(2v_1)}$ for $t > -v_1$, which is unbounded near $-v_1$. Taking $F = 1$ in (5.11) at $t = 0$ gives the normalization

$$\int_{\mathcal{S}_{R^2}} \gamma_v \, d\mathcal{H}^{|I|-1} = 2R p_v(0), \quad \frac{d\mu_v}{d\mathcal{H}^{|I|-1}}(y) = \frac{\gamma_v(y)}{2R p_v(0)}, \quad y \in \mathcal{S}_{R^2}. \quad (5.12)$$

Proposition 5.6 (Bingham laws as conditioned Gaussian laws). *Let $N \geq 2$.*

1. *The uniform measure ν_N on S_N is μ_v for $v = (1, \dots, 1) \in \mathbb{R}^N$. If g is a standard Gaussian vector in $\mathbb{R}^N$, then $g/|g|$ is independent of $|g|$, and $\sqrt{N} g/|g|$ has law ν_N .*
2. *Let v be a variance sequence on $\{1, \dots, N\}$ with $\sum_i v_i = N$, let $z \in \mathbb{R}$, and let $b_i = z - v_i^{-1}$. Then the probability measure on S_N proportional to $\exp\{\frac{1}{2} \sum_i b_i x_i^2\} \nu_N(dx)$ is μ_v .*

3. Let A be a real symmetric $N \times N$ matrix with eigenvalues $\alpha_1 \geq \dots \geq \alpha_N$, and let $z > \alpha_1$ be its Lagrange multiplier, the unique solution of $N^{-1} \sum_i (z - \alpha_i)^{-1} = 1$. In the eigencoordinates of A , the Gibbs measure $\pi_A \propto e^{x^\top A x / 2} \nu_N$ is μ_q , where $q_i = (z - \alpha_i)^{-1}$.
4. For every variance sequence v and $\lambda > 0$, if Y has law μ_v , then λY has law $\mu_{\lambda^2 v}$. Each μ_v is invariant under the reflections $y_i \mapsto -y_i$, $i \in I$.

The last part also holds for countable v , with μ_v as in Definition 5.8 below; see the remark after that definition.

Proof. For the first part, we note that the density in (5.10) with $v_i = 1$ equals $e^{-N/2}$ on S_N , and μ_v is normalized surface measure, that is, ν_N . In polar coordinates $g = r\theta$, the density of g is $(2\pi)^{-N/2} e^{-r^2/2} r^{N-1}$ times the surface measure $\mathcal{H}^{N-1}(d\theta)$ on the unit sphere. This density is a product of a function of r and a function of θ . It follows that $|g|$ and $g/|g|$ are independent, and that $g/|g|$ is uniformly distributed on the unit sphere. For the second part, on S_N we have $\frac{1}{2} \sum_i b_i x_i^2 = \frac{zN}{2} - \frac{1}{2} \sum_i x_i^2 / v_i$. The first term is constant, and μ_v lives on S_N because $\sum_i v_i = N$, so the two densities agree up to a constant factor. We obtain the third part as the case $v_i = q_i$ and $b_i = \alpha_i$ of the second, since $\sum_i q_i = N$ by the definition of z . The existence and uniqueness of z follow since $z \mapsto N^{-1} \sum_i (z - \alpha_i)^{-1}$ decreases continuously from $+\infty$ to 0 on (α_1, ∞) ; see Section 4. For the fourth part, the map $y \mapsto \lambda y$ carries $\mathcal{S}_{R^2}$ onto $\mathcal{S}_{\lambda^2 R^2}$ and multiplies surface measure by a constant, and $\sum_i (\lambda y_i)^2 / (\lambda^2 v_i) = \sum_i y_i^2 / v_i$. The density in (5.10) and the sphere are invariant under each reflection. $\square$

The spherical constraint goes back to the spherical model of [28, Abstract], and its spin-glass version was introduced in [97, Abstract]. In the present normalization, the Lagrange multiplier z is the unique critical point on (α_1, ∞) of

$$\mathcal{A}_N(u) = \frac{Nu}{2} - \frac{1}{2} \sum_i \log(u - \alpha_i),$$

since $\mathcal{A}'_N(u) = \frac{1}{2}[N - \sum_i (u - \alpha_i)^{-1}]$ is increasing and vanishes exactly at $u = z$.

5.4 Countable sequences and coordinate marginals

For countable v there is no sphere of finite radius when $\sum_i v_i = \infty$, which is the case in the critical window. We define the conditioned law through its finite-coordinate marginals, following [62, Section 2.2], where this is done for sequences with the decay of the inverse Airy gaps; the arguments apply to every variance sequence. For finite $A \subset I$ such that $I \setminus A$ has at least three elements, we define

$$r_{v,A}(z) = \frac{p_{v,I \setminus A}(-U_A(z))}{p_v(0)}, \quad z \in \mathbb{R}^A. \quad (5.13)$$

This is well defined by Lemma 5.2, since $p_v(0) > 0$.

Lemma 5.7 (Consistent marginals). *Let v be a countable variance sequence. For every finite $A \subset I$, $r_{v,A}$ is a probability density with respect to γ_A . If $A \subset A'$ are finite, then the image of $r_{v,A'} d\gamma_{A'}$ under the projection $\mathbb{R}^{A'} \rightarrow \mathbb{R}^A$ is $r_{v,A} d\gamma_A$. There exists a unique probability measure on $\mathbb{R}^I$ whose image under the projection to $\mathbb{R}^A$ is $r_{v,A} d\gamma_A$ for every finite $A \subset I$.*

Proof. By (5.7) at $t = 0$, we have $\int r_{v,A} d\gamma_A = \mathbb{E} p_{v,I \setminus A}(-T_{v,A}) / p_v(0) = 1$. It suffices to check consistency when $A' = A \cup \{j\}$. For fixed $z \in \mathbb{R}^A$, integrating out the coordinate j and applying (5.7) to the sequence $v|_{I \setminus A'}$, split into $\{j\}$ and $I \setminus A'$, we have

$$\int_{\mathbb{R}} p_{v,I \setminus A'}(-U_A(z) - v_j(x^2 - 1)) \phi(x) dx = p_{v,I \setminus A}(-U_A(z)).$$

Dividing by $p_v(0)$ proves the consistency. The last statement follows from Kolmogorov's extension theorem. $\square$

Definition 5.8 (Countable conditioned law). Let v be a countable variance sequence, and let $Z = (Z_i)_{i \in I}$ have the law constructed in Lemma 5.7. The *Gaussian law μ_v conditioned on $T_v = 0$* is the law of $Y = (\sqrt{v_i} Z_i)_{i \in I}$ on $\mathbb{R}^I$.

By (5.13), the laws $\mu_v^{Z_A}$ are invariant under each reflection $z_i \mapsto -z_i$, and the standardized coordinates of λY with respect to $\lambda^2 v$ are again Z . This proves the fourth part of Proposition 5.6 for countable v . As shown in Propositions 5.10 and 5.17 below, μ_v is supported on the set where $\sum_i (y_i^2 - v_i) = 0$, and, for any enumeration $(i_k)_{k \geq 1}$ of I , it is the limit of the finite laws $\mu_{v|_{\{i_1, \dots, i_n\}}}$ as $n \rightarrow \infty$, in total variation on every finite set of coordinates.

Lemma 5.9 (Marginal densities). *Let v be a variance sequence, and let $A \subset I$. Suppose that $I \setminus A$ has at least three elements, or that A is infinite and $T_{v,I \setminus A}$ has a bounded density $p_{v,I \setminus A}$. Then $\mu_v^{Z_A}$ has density*

$$r_{v,A}(z) = \frac{p_{v,I \setminus A}(-U_A(z))}{p_v(0)} \quad (5.14)$$

with respect to γ_A . If A is infinite, $U_A(z)$ is the series defined for γ_A -almost every z after (5.2), and the right side of (5.14) does not depend on the choice of a version of $p_{v,I \setminus A}$, since U_A has an absolutely continuous law under γ_A .

Proof. Let I be finite, and let F be a bounded continuous function of $z \in \mathbb{R}^A$. Define the bounded continuous physical-coordinate test $\tilde{F}(y) = F((y_i / \sqrt{v_i})_{i \in A})$ on $\mathbb{R}^I$. By independence of g_A and $g_{I \setminus A}$, for bounded nonnegative measurable K ,

$$\begin{aligned} \mathbb{E}[\tilde{F}(Y)K(T_v)] &= \int_{\mathbb{R}^A} F(z) \phi_A(z) \int_{\mathbb{R}} K(U_A(z) + x) p_{v,I \setminus A}(x) dx dz \\ &= \int_{\mathbb{R}} K(t) \left[\int_{\mathbb{R}^A} F \phi_A p_{v,I \setminus A}(t - U_A) dz \right] dt. \end{aligned}$$

The bracket is a continuous function of t , by dominated convergence, since $p_{v,I \setminus A}$ is bounded and continuous (Lemma 5.2). By Lemma 5.5, applied to the positive and negative parts of $\tilde{F}$, $p_v[\tilde{F}]$ is another density of the same measure, continuous on $(-R^2, \infty)$. The two densities agree at $t = 0$, by continuity. Dividing by $p_v[1](0) = p_v(0)$ and using the third part of Lemma 5.5, we obtain $\mu_v(\tilde{F}) = \int F r_{v,A} d\gamma_A$, that is, $\mathbb{E}_{\mu_v} F(Z_A) = \int F r_{v,A} d\gamma_A$. Since bounded continuous functions determine a probability measure on $\mathbb{R}^A$, this proves (5.14) for finite I .

Let I be countable. For finite A , (5.14) is the definition of μ_v . Let A be infinite, and let $\lambda_A(dz) = p_v(0)^{-1} p_{v,I \setminus A}(-U_A(z)) \gamma_A(dz)$, where $p_{v,I \setminus A}$ is a bounded density of $T_{v,I \setminus A}$; this holds in both cases of the hypothesis, by Lemma 5.2. For every infinite $A' \subset A$, the variable $T_{v,A'}$ has a bounded continuous density $p_{v,A'}$, by Lemma 5.2, and the convolution $t \mapsto \int p_{v,I \setminus A}(t-x) p_{v,A'}(x) dx$ of a bounded and an integrable function is continuous. It is a density of $T_{v,(I \setminus A) \cup A'}$, which has a continuous density by Lemma 5.2, and the two agree everywhere:

$$p_{v,(I \setminus A) \cup A'}(t) = \mathbb{E} p_{v,I \setminus A}(t - T_{v,A'}) \quad \text{for every } t \in \mathbb{R}. \quad (5.15)$$

With $A' = A$ and $t = 0$, this shows that λ_A is a probability measure. Let $A_0 \subset A$ be finite. Since $U_A = U_{A_0} + U_{A \setminus A_0}$ almost surely and the coordinates are independent under γ_A , integrating out the coordinates in $A \setminus A_0$ yields the density

$$\frac{\mathbb{E} p_{v,I \setminus A}(-U_{A_0}(z) - T_{v,A \setminus A_0})}{p_v(0)} = \frac{p_{v,I \setminus A_0}(-U_{A_0}(z))}{p_v(0)} = r_{v,A_0}(z), \quad z \in \mathbb{R}^{A_0},$$

where we used (5.15) with $A' = A \setminus A_0$. We conclude that λ_A and $\mu_v^{Z_A}$ have the same finite-dimensional marginals. Since two probability measures on $\mathbb{R}^A$ that agree on cylinder sets are equal, we have $\lambda_A = \mu_v^{Z_A}$. $\square$

For finite A , the formula (5.14) appears in [62, Lemma 4.6] for finite sequences and in [62, (2.14)] for countable ones.

Proposition 5.10 (The radial constraint). *Let v be a countable variance sequence and $(i_k)_{k \geq 1}$ an enumeration of I . Under μ_v , the partial sums $\sum_{k \leq n} (Y_{i_k}^2 - v_{i_k}) = \sum_{k \leq n} v_{i_k} (Z_{i_k}^2 - 1)$ converge to zero almost surely and in $L^2(\mu_v)$.*

Proof. This is proved in [62, Lemma 2.4] for sequences with the decay of the inverse Airy gaps; we give the argument for general v . Let $A_n = \{i_1, \dots, i_n\}$ and $Q_n = U_{A_n}(Z_{A_n})$. For $n \geq 3$, the density q_n of T_{v,A_n} under the Gaussian product law satisfies $\|q_n\|_\infty \leq \|p_{v,A_3}\|_\infty$, since T_{v,A_n} is the sum of T_{v,A_3} and an independent variable. By Lemma 5.9, we have

$$\mathbb{E}_{\mu_v} Q_n^2 = \frac{1}{p_v(0)} \int_{\mathbb{R}} x^2 q_n(x) p_{v,I \setminus A_n}(-x) dx \leq \frac{\|p_{v,A_3}\|_\infty}{p_v(0)} \mathbb{E} T_{v,I \setminus A_n}^2 = \frac{\|p_{v,A_3}\|_\infty}{p_v(0)} 2 \sum_{k > n} v_{i_k}^2,$$

which tends to zero. For almost sure convergence, let $B = A_3$. Applying Lemma 5.9 with $A = I \setminus B$, we find that the law of $Z_{I \setminus B}$ under μ_v is absolutely continuous with respect to $\gamma_{I \setminus B}$. Under $\gamma_{I \setminus B}$, the partial sums $\sum_{3 < k \leq n} v_{i_k} (z_{i_k}^2 - 1)$ converge almost surely, as noted after (5.1). It follows that Q_n converges μ_v -almost surely, and its limit is zero because $Q_n \rightarrow 0$ in $L^2(\mu_v)$. $\square$

The next theorem compares a set of coordinates of μ_v with independent Gaussian variables when the coordinates carry at most a fixed fraction of the variance of T_v .

Theorem 5.11 (Gaussian approximation of coordinate marginals). *There exist universal constants $\eta_* > 0$ and $C < \infty$ with the following property. Let v be a variance sequence with $\eta_v \leq \eta_*$, and let $A \subset I$ with $h_A \leq 1/2$. Then $I \setminus A$ has at least three elements, and*

$$\|\mu_v^{Z_A} - \gamma_A\|_{TV} \leq C(\eta_v + h_A), \quad 0 \leq r_{v,A} \leq C. \quad (5.16)$$

Further, for each $\varepsilon \in (0, 1)$ there exist $\eta(\varepsilon) > 0$ and $C_\varepsilon < \infty$ such that $\eta_v \leq \eta(\varepsilon)$ and $h_A \leq 1 - \varepsilon$ imply that $I \setminus A$ has at least three elements and $r_{v,A} \leq C_\varepsilon$.

For the uniform measure, Borel gave the Gaussian limit of one coordinate; see [57, Section 6.1]. Stam proved a bound in total variation for a number of coordinates that grows more slowly than $\sqrt{N}$ [139], and Diaconis and Freedman proved a bound of order k/N for k coordinates, which is small whenever $k = o(N)$ [57, (1)]; see [57, Sections 6.1 and 6.4] for the history. Theorem 5.11 allows arbitrary variances, with constants that do not depend on the smallest variance.

Proof. The standardized variances of $v|_{I \setminus A}$ are $v_i/S_{v|_{I \setminus A}} = \hat{v}_i(1 - h_A)^{-1/2}$, and in particular $\eta_{v|_{I \setminus A}} \leq \eta_v(1 - h_A)^{-1/2}$. Let $\eta_{\text{loc}}, c_0, C_0$ be as in Theorem 5.3. If $\eta_v(1 - h_A)^{-1/2} \leq \eta_{\text{loc}}$, then $I \setminus A$ has at least $(2\eta_{\text{loc}}^2)^{-1} \geq 3$ elements, and Lemma 5.9 applies. Writing $p_{v, I \setminus A}(x) = S_{v|_{I \setminus A}}^{-1} f_{v|_{I \setminus A}}(x/S_{v|_{I \setminus A}})$ and $p_v(0) = S_v^{-1} f_v(0)$ in (5.14), we obtain

$$r_{v,A}(z) = \frac{1}{(1 - h_A)^{1/2}} \frac{f_{v|_{I \setminus A}}(-\tau_A)}{f_v(0)}, \quad \tau_A = \frac{U_A(z)}{S_v(1 - h_A)^{1/2}}. \quad (5.17)$$

By Theorem 5.3 with $j = 0$, we have $f_{v|_{I \setminus A}} = \phi + e_1$ and $f_v(0) = \phi(0) + e_2$, where

$$\|e_1\|_\infty \leq C_0 \eta_v (1 - h_A)^{-1/2}, \quad |e_2| \leq C_0 \eta_v.$$

Let $h_A \leq 1/2$ and $\eta_* \leq \eta_{\text{loc}}/\sqrt{2}$, with $\sqrt{2}C_0\eta_* \leq \phi(0)/2$. Then $f_v(0) \geq \phi(0)/2$ and $f_{v|_{I \setminus A}} \leq \phi(0) + \sqrt{2}C_0\eta_v$, and (5.17) gives $r_{v,A} \leq C$. Further, we have $|(1 - h)^{-1/2} - 1| \leq h$ for $h \in [0, 1/2]$ and $|\phi(x) - \phi(0)| \leq \phi(0)x^2/2$. Inserting these bounds into the numerator of $r_{v,A} - 1 = [(1 - h_A)^{-1/2} f_{v|_{I \setminus A}}(-\tau_A) - f_v(0)]/f_v(0)$, we find that it is at most $C(\eta_v + h_A + \tau_A^2)$ in absolute value. We conclude that

$$|r_{v,A}(z) - 1| \leq C(\eta_v + h_A + \tau_A(z)^2).$$

Under γ_A , we have $\mathbb{E} U_A^2 = 2 \sum_{i \in A} v_i^2 = h_A S_v^2$, which gives $\int \tau_A^2 d\gamma_A = h_A / (1 - h_A) \leq 2h_A$. Recalling that $\|\mu_v^{Z_A} - \gamma_A\|_{TV} = \frac{1}{2} \int |r_{v,A} - 1| d\gamma_A$, we conclude that (5.16) holds.

Now let $h_A \leq 1 - \varepsilon$, and let $\eta(\varepsilon) = \min\{\eta_{\text{loc}}\varepsilon^{1/2}, \phi(0)/(2C_0)\}$. Then $\eta_{v|_{I \setminus A}} \leq \eta_v \varepsilon^{-1/2} \leq \eta_{\text{loc}}$ and $f_v(0) \geq \phi(0)/2$, so that (5.17) yields $r_{v,A} \leq \varepsilon^{-1/2}(\phi(0) + C_0 \eta_v \varepsilon^{-1/2}) / (\phi(0)/2) \leq C_\varepsilon$. $\square$

5.5 Moments of coordinate marginals

The density bounds of Theorem 5.11 transfer moments of Gaussian observables to the conditioned law.

Theorem 5.12 (Moments under the conditioned law). *Let $\eta_* \leq 1/2$ be as in Theorem 5.11, and let v be a variance sequence with $\eta_v \leq \eta_*$. If $A \subset I$ satisfies $h_A \leq 1/2$, $p > 1$, and $F \in L^p(\gamma_A)$, then $F(Z_A)$ is integrable under μ_v and*

$$|\mathbb{E}_{\mu_v} F(Z_A) - \mathbb{E}_{\gamma_A} F| \leq C_p \|F\|_{L^p(\gamma_A)} (\eta_v + h_A)^{1-1/p}. \quad (5.18)$$

In particular, uniformly in $i \in I$,

$$\begin{aligned} \mathbb{E}_{\mu_v} Y_i^2 &= v_i [1 + O(\eta_v^{1/2})], \\ \mathbb{E}_{\mu_v} Y_i^4 &= 3v_i^2 [1 + O(\eta_v^{1/2})], \\ \text{Var}_{\mu_v}(Y_i^2) &= 2v_i^2 [1 + O(\eta_v^{1/2})], \end{aligned} \quad (5.19)$$

where the implicit constants are universal. For each $\varepsilon \in (0, 1)$, let $\eta(\varepsilon)$ and C_ε be as in Theorem 5.11. If $\eta_v \leq \eta(\varepsilon)$ and $h_A \leq 1 - \varepsilon$, then for every $p > 1$, $F \in L^p(\gamma_A)$, and $L > 0$,

$$\mathbb{E}_{\mu_v} [|F(Z_A)| \mathbf{1}\{|F(Z_A)| > L\}] \leq C_\varepsilon L^{1-p} \|F\|_{L^p(\gamma_A)}^p. \quad (5.20)$$

In particular, every family of such observables $F(Z_A)$ with bounded $L^p(\gamma_A)$ norms is uniformly integrable under the corresponding conditioned laws.

Proof. For (5.18), let $q = p/(p-1)$. We have $\|r_{v,A} - 1\|_\infty \leq C$ and $\int |r_{v,A} - 1| d\gamma_A \leq C(\eta_v + h_A)$ by Theorem 5.11, so

$$\|r_{v,A} - 1\|_{L^q(\gamma_A)} \leq \|r_{v,A} - 1\|_\infty^{1-1/q} \|r_{v,A} - 1\|_{L^1(\gamma_A)}^{1/q} \leq C_p (\eta_v + h_A)^{1-1/p}.$$

The bound $r_{v,A} \leq C$ implies that $F(Z_A)$ is integrable under μ_v . By Lemma 5.9 and Hölder's inequality,

$$|\mathbb{E}_{\mu_v} F(Z_A) - \mathbb{E}_{\gamma_A} F| = \left| \int F(r_{v,A} - 1) d\gamma_A \right| \leq \|F\|_{L^p(\gamma_A)} \|r_{v,A} - 1\|_{L^q(\gamma_A)},$$

which proves (5.18). For $A = \{i\}$ we have $h_A = 2\hat{\sigma}_i^2 \leq 2\eta_v^2 \leq 1/2$. Taking $p = 2$ and $F(z) = z_i^2$ or $F(z) = z_i^4$, whose Gaussian L^2 norms are universal constants, gives

$$|\mathbb{E}_{\mu_v} Z_i^2 - 1| + |\mathbb{E}_{\mu_v} Z_i^4 - 3| \leq C(\eta_v + 2\eta_v^2)^{1/2} \leq C\eta_v^{1/2}.$$

Scaling by v_i and v_i^2 gives the first two estimates in (5.19); the third follows by subtracting $(\mathbb{E}_{\mu_v} Y_i^2)^2$ from $\mathbb{E}_{\mu_v} Y_i^4$. Finally, under the assumptions of the last statement, $r_{v,A} \leq C_\varepsilon$ by Theorem 5.11, and

$$\mathbb{E}_{\mu_v} [F(Z_A) | \mathbf{1}\{|F(Z_A)| > L\}] \leq C_\varepsilon \int |F| \mathbf{1}\{|F| > L\} d\gamma_A \leq C_\varepsilon L^{1-p} \int |F|^p d\gamma_A,$$

which is (5.20). □

5.6 Exponential tilts and quadratic statistics

Lemma 5.13 (Quadratic exponential moments). *Let v be a variance sequence, and let $w = (w_i)_{i \in I}$ be real numbers, with only finitely many nonzero when I is countable. Let $t \in \mathbb{R}$ satisfy $1 - tw_i > 0$ for every i , and let*

$$v_i(t) = \frac{v_i}{1 - tw_i}, \quad \Delta(t) = \sum_{i \in I} (v_i(t) - v_i).$$

Then $v(t)$ is a variance sequence, and

$$\mathbb{E}_{\mu_v} \exp \left\{ \frac{t}{2} \sum_{i \in I} w_i Z_i^2 \right\} = \prod_{i \in I} (1 - tw_i)^{-1/2} \frac{p_{v(t)}(-\Delta(t))}{p_v(0)}. \quad (5.21)$$

All the quantities on the right side are real, and no smallness of $\Delta(t)$ is assumed.

Proof. For one coordinate, multiplying the Gaussian density with variance v_i by $\exp\{tw_i y^2 / (2v_i)\}$ gives $(1 - tw_i)^{-1/2}$ times the Gaussian density with variance $v_i(t)$. Multiplying over i , we have, with γ_v as in Lemma 5.5,

$$\exp \left\{ \frac{t}{2} \sum_i w_i \frac{y_i^2}{v_i} \right\} \gamma_v(y) = \prod_i (1 - tw_i)^{-1/2} \gamma_{v(t)}(y). \quad (5.22)$$

Let I be finite, $R^2 = \sum_i v_i$, and let $W(y)$ be the exponential on the left of (5.22). Since $\sum_i v_i(t) = R^2 + \Delta(t)$, the definition (5.11) applied to $v(t)$ at the point $s - \Delta(t)$ integrates over the same sphere $\mathcal{S}_{R^2+s}$ as $p_v[W](s)$, and (5.22) implies that

$$p_v[W](s) = \prod_i (1 - tw_i)^{-1/2} p_{v(t)}(s - \Delta(t)), \quad s > -R^2.$$

Evaluating at $s = 0$ and dividing by $p_v(0)$, we obtain (5.21) from the third part of Lemma 5.5.

Let I be countable, and let $F \subset I$ be a finite set containing the indices with $w_i \neq 0$. The sequence $v(t)$ agrees with v off F . In particular, it is a variance sequence and $p_{v(t), I \setminus F} = p_{v, I \setminus F}$. By Lemma 5.9 and (5.22),

$$\mathbb{E}_{\mu_v} \exp \left\{ \frac{t}{2} \sum_{i \in F} w_i Z_i^2 \right\} = \prod_{i \in F} (1 - tw_i)^{-1/2} \frac{\mathbb{E} p_{v, I \setminus F}(-\sum_{i \in F} (v_i(t) g_i^2 - v_i))}{p_v(0)}.$$

Since $\sum_{i \in F} (v_i(t) g_i^2 - v_i) = T_{v(t), F} + \Delta(t)$, the expectation equals $p_{v(t)}(-\Delta(t))$ by (5.7) for $v(t)$. This proves (5.21). $\square$

For independent Gaussian variables, the analogous moment bound for quadratic statistics is classical; we record it with its short proof, since it is also used in Proposition 5.17.

Lemma 5.14 (Quadratic statistics of independent Gaussians). *For every $p \geq 1$ there exists $C_p < \infty$ such that, for every finite or countable index set J and every real $w \in \ell^2(J)$, the series $X = \sum_{i \in J} w_i (g_i^2 - 1)$ satisfies $\mathbb{E} e^{tX} \leq e^{2t^2 \|w\|_2^2}$ for $|t| \leq (4\|w\|_2)^{-1}$, and $\|X\|_{L^p} \leq C_p \|w\|_2$. Here $(4\|w\|_2)^{-1} = +\infty$ when $w = 0$.*

Proof. If $w = 0$, then $X = 0$ and both claims hold immediately. Fix a finite nonzero w and t with $|t| \leq (4\|w\|_2)^{-1}$, so that $|2tw_i| \leq 1/2$ for every i . For $|2u| \leq 1/2$, we have $-u - \frac{1}{2} \log(1 - 2u) = \sum_{k \geq 2} (2u)^k / (2k) \leq \frac{1}{4} (2u)^2 / (1 - |2u|) \leq 2u^2$. Since $\mathbb{E} e^{sw(g^2 - 1)} = e^{-sw} (1 - 2sw)^{-1/2}$ for $2sw < 1$, we have $\log \mathbb{E} e^{tX} = \sum_i [-tw_i - \frac{1}{2} \log(1 - 2tw_i)] \leq 2t^2 \|w\|_2^2$. For countable J , the partial sums converge almost surely along a subsequence, and the same bound follows from Fatou's lemma. At $t = \pm(4\|w\|_2)^{-1}$, Markov's inequality implies that $\mathbb{P}(|X| > s\|w\|_2) \leq 2e^{1/8} e^{-s/4}$ for $s \geq 0$. Integrating this tail against $ps^{p-1} ds$ yields $\mathbb{E}|X|^p \leq C_p^p \|w\|_2^p$. $\square$

Theorem 5.15 (Quadratic statistics under the conditioned law). *There exist universal constants $\eta_0, c_* > 0$ and $C < \infty$ such that the following holds for every variance sequence v with $\eta_v \leq \eta_0$. Let $w = (w_i)_{i \in I}$ be real numbers, with only finitely many nonzero, and let*

$$Q_w = \sum_{i \in I} w_i (Z_i^2 - 1) = \sum_{i \in I} w_i \left(\frac{Y_i^2}{v_i} - 1 \right), \quad \|w\|_2 = \left(\sum_i w_i^2 \right)^{1/2} > 0.$$

Then

$$\mathbb{E}_{\mu_v} e^{tQ_w} \leq C e^{Ct^2 \|w\|_2^2}, \quad |t| \leq c_* / \|w\|_2. \quad (5.23)$$

For every real $p \geq 1$, there exists $C_p < \infty$ such that

$$\|Q_w\|_{L^p(\mu_v)} \leq C_p \|w\|_2. \quad (5.24)$$

If I is countable and $w \in \ell^2(I)$, then the series defining Q_w converges in $L^p(\mu_v)$ for every $p \geq 1$, and its limit satisfies the same two bounds.

If $w = 0$, then $Q_w = 0$ and there is nothing to prove. The theorem bounds Q_w itself, which is centered at the mean under the independent Gaussian law; the mean $\mathbb{E}_{\mu_v} Q_w$ need not vanish.

Proof. We set $c_* = 1/8$ and fix t with $|t| \leq c_*/\|w\|_2$, so that $|2tw_i| \leq 1/4$ for every i . We apply Lemma 5.13 to the coefficients $2w_i$, with

$$v_i(t) = \frac{v_i}{1 - 2tw_i} \in \left[\frac{4}{5}v_i, \frac{4}{3}v_i \right], \quad \frac{4}{5}S_v \leq S_{v(t)} \leq \frac{4}{3}S_v, \quad \eta_{v(t)} \leq \frac{5}{3}\eta_v.$$

Since $tQ_w = t \sum_i w_i Z_i^2 - t \sum_i w_i$, we find from (5.21) that

$$\mathbb{E}_{\mu_v} e^{tQ_w} = \exp \left\{ \sum_{i \in I} \left[-tw_i - \frac{1}{2} \log(1 - 2tw_i) \right] \right\} \frac{p_{v(t)}(-\Delta(t))}{p_v(0)}.$$

Let $\eta_0 = \frac{3}{5}\eta_{\text{loc}}$, so that (5.9) applies to v and $v(t)$. Using the upper bound for $f_{v(t)}$ and the lower bound for $f_v(0)$, we have

$$\frac{p_{v(t)}(-\Delta(t))}{p_v(0)} \leq \frac{C/S_{v(t)}}{c_0/S_v} \leq C,$$

whatever the size of $\Delta(t)$. The elementary inequality in the proof of Lemma 5.14 bounds the exponent by $2t^2\|w\|_2^2$, which proves (5.23).

At $t = \pm c_*/\|w\|_2$, Markov's inequality and (5.23) give $\mu_v(|Q_w| > s\|w\|_2) \leq Ce^{-c_*s}$ for $s \geq 0$. Integrating this tail against $ps^{p-1}ds$ gives (5.24). For countable I and $w \in \ell^2(I)$, the bound (5.24) applied to differences of partial sums shows that the partial sums form a Cauchy sequence in every $L^p(\mu_v)$. The limits for different p agree almost surely, since they are also limits in probability. Along an almost surely convergent subsequence, Fatou's lemma yields (5.23) for the limit, since $c_*/\|w\|_2$ is at most the corresponding threshold for each partial sum; norm convergence yields (5.24). $\square$

The case $p = 2$ is used often enough to be stated separately, in the variables Y_i . Its left side is a second moment about the deterministic center $\sum_i w_i v_i$, and it bounds the variance of $\sum_i w_i Y_i^2$ under μ_v .

Corollary 5.16 (Quadratic second-moment bound). *Let η_0 be the constant of Theorem 5.15. There exists a universal constant C such that, for every variance sequence v with $\eta_v \leq \eta_0$ and all real w_i , only finitely many nonzero,*

$$\mathbb{E}_{\mu_v} \left[\sum_{i \in I} w_i (Y_i^2 - v_i) \right]^2 \leq C \sum_{i \in I} w_i^2 v_i^2. \quad (5.25)$$

For countable I and $\sum_i w_i^2 v_i^2 < \infty$, the series is the $L^2(\mu_v)$ limit of its partial sums, and (5.25) holds.

Proof. We have $\sum_i w_i (Y_i^2 - v_i) = Q_{w'}$ with $w'_i = w_i v_i$, and the claim is Theorem 5.15 with $p = 2$. $\square$

5.7 Continuity in the variance sequence

The estimates of the previous sections assume that η_v is small. The conditioned law in the critical window has a fixed positive η_v . Instead, we use the following domination by independent Gaussian variables, together with continuity of the law under convergence of the variance sequence in ℓ^2 . For finite $B \subset I$ such that B and $I \setminus B$ both have at least three elements, let

$$M(v, B) = \frac{\max\{\|p_{v,B}\|_\infty, \|p_{v,I \setminus B}\|_\infty\}}{p_v(0)}. \quad (5.26)$$

Proposition 5.17 (Continuity and domination). *Let v be a variance sequence.*

1. *Let $B \subset I$ be finite, and suppose that B and $I \setminus B$ both have at least three elements. Then for all nonnegative measurable functions H_1 on $\mathbb{R}^B$ and H_2 on $\mathbb{R}^{I \setminus B}$,*

$$\mathbb{E}_{\mu_v} H_1(Z_B) \leq M(v, B) \int H_1 d\gamma_B, \quad \mathbb{E}_{\mu_v} H_2(Z_{I \setminus B}) \leq M(v, B) \int H_2 d\gamma_{I \setminus B}. \quad (5.27)$$

In particular, for every $p \geq 1$ we have $\sup_{i \in I} \mathbb{E}_{\mu_v} |Z_i|^p \leq M(v, B) \mathbb{E}|G|^p$, and $\mathbb{E}_{\mu_v} \sum_i u_i Z_i^2 \leq M(v, B) \sum_i u_i$ for all $u_i \geq 0$. For every real $w \in \ell^2(I)$, the series $\sum_i w_i (Z_i^2 - 1)$ converges in $L^p(\mu_v)$, and

$$\left\| \sum_{i \in I} w_i (Z_i^2 - 1) \right\|_{L^p(\mu_v)} \leq 2C_p M(v, B)^{1/p} \|w\|_2, \quad (5.28)$$

where C_p is the constant of Lemma 5.14.

2. *Let $v^{(n)}$, $n \geq 1$, be variance sequences indexed by subsets $I_n \subset I$, extended by zero to I , such that $\sum_{i \in I} (v_i^{(n)} - v_i)^2 \rightarrow 0$. Suppose that I has at least three elements. Then $p_{v^{(n)}}(0) \rightarrow p_v(0)$. Let $A \subset I$ be finite such that $I \setminus A$ has at least three elements. Then $A \subset I_n$ and $I_n \setminus A$ has at least three elements for all large n , and*

$$\sup_{x \in \mathbb{R}} |p_{v^{(n)}, I_n \setminus A}(x) - p_{v, I \setminus A}(x)| \rightarrow 0, \\ \|\mu_{v^{(n)}}^{Z_A} - \mu_v^{Z_A}\|_{\text{TV}} \rightarrow 0, \quad \|\mu_{v^{(n)}}^{Y_A} - \mu_v^{Y_A}\|_{\text{TV}} \rightarrow 0,$$

where μ^{Y_A} is the law of $Y_A = (Y_i)_{i \in A}$ under μ . If B is as in the first part, then $M(v^{(n)}, B) \rightarrow M(v, B)$. In particular, there exists n_0 such that the bounds of the first part hold for $\mu_{v^{(n)}}$, $n \geq n_0$, with $M(v, B)$ replaced by $2M(v, B)$.

For variance sequences with the decay of the inverse Airy gaps, the convergence of finite-coordinate marginals in the second part, with uniform bounds on second moments, is proved in [62, Lemma 4.7]; see (4.24)–(4.27) there. The domination in the first part, from which we obtain all moments and the bound (5.28), is the additional input needed in Section 18.

Proof. For the first part, we use Lemma 5.9 with $A = B$ and with $A = I \setminus B$. The densities $r_{v,B} = p_{v,I \setminus B}(-U_B)/p_v(0)$ and $r_{v,I \setminus B} = p_{v,B}(-U_{I \setminus B})/p_v(0)$ are bounded by $M(v, B)$, which proves (5.27). We obtain the bound on $\mathbb{E}_{\mu_v} |Z_i|^p$ by applying (5.27) to $H(z) = |z|^p$ in whichever of the two sets contains i , and the bound on $\mathbb{E}_{\mu_v} \sum_i u_i Z_i^2$ follows by monotone convergence. For (5.28), the partial sums of $\sum_{i \in I \setminus B} w_i (Z_i^2 - 1)$, along any enumeration, form a Cauchy sequence in $L^p(\mu_v)$, since by (5.27) and Lemma 5.14,

$$\mathbb{E}_{\mu_v} \left| \sum_{i \in F} w_i (Z_i^2 - 1) \right|^p \leq M(v, B) C_p^p \left(\sum_{i \in F} w_i^2 \right)^{p/2}$$

for every finite $F \subset I \setminus B$. The same bound holds for the limit, with F replaced by $I \setminus B$, and for the sum over B . The triangle inequality and $\|w|_B\|_2 + \|w|_{I \setminus B}\|_2 \leq 2\|w\|_2$ prove (5.28).

For the second part, we take the same independent variables g for all n , so that $T_{v^{(n)}} = \sum_{i \in I} v_i^{(n)} (g_i^2 - 1)$. Then $\mathbb{E} |T_{v^{(n)}} - T_v|^2 = 2 \sum_i (v_i^{(n)} - v_i)^2 \rightarrow 0$, and the characteristic functions φ_n of $T_{v^{(n)}}$ converge pointwise to the characteristic function φ of T_v . Fix three indices $j_1, j_2, j_3 \in I$, and let $m = \frac{1}{2} \min_k v_{j_k}$. For large n , we have $v_{j_k}^{(n)} \geq m$, and the argument in the proof of Lemma 5.2 shows that $|\varphi_n(u)| + |\varphi(u)| \leq 2(1 + 4m^2 u^2)^{-3/4}$. By Fourier inversion and dominated convergence,

$$\sup_x |p_{v^{(n)}}(x) - p_v(x)| \leq \frac{1}{2\pi} \int_{\mathbb{R}} |\varphi_n(u) - \varphi(u)| du \rightarrow 0.$$

The same argument, with three indices of $I \setminus A$, proves the uniform convergence of $p_{v^{(n)}, I_n \setminus A}$. Every finite subset of I is contained in I_n for large n , since $v_i^{(n)} \rightarrow v_i > 0$ for each i ; this proves the claims about $A \subset I_n$. The laws $\mu_{v^{(n)}}^{Z_A}$ and $\mu_v^{Z_A}$ have the densities $r_{v^{(n)}, A}$ and $r_{v, A}$ with respect to γ_A , by Lemma 5.9. Using $U_A^{(n)}(z) = \sum_{i \in A} v_i^{(n)} (z_i^2 - 1) \rightarrow U_A(z)$, the uniform convergence of densities, the continuity of $p_{v, I \setminus A}$, and $p_{v^{(n)}}(0) \rightarrow p_v(0) > 0$, we find that $r_{v^{(n)}, A} \rightarrow r_{v, A}$ pointwise. By Scheffé's lemma, we have $\int |r_{v^{(n)}, A} - r_{v, A}| d\gamma_A \rightarrow 0$, which is the convergence of $\mu_{v^{(n)}}^{Z_A}$ in total variation.

For Y_A , let D_n and D be the diagonal matrices with entries $(v_i^{(n)})^{1/2}$ and $v_i^{1/2}$, $i \in A$. Total variation is invariant under the bijection D_n , and by the triangle inequality we have

$$\|\mu_{v^{(n)}}^{Y_A} - \mu_v^{Y_A}\|_{\text{TV}} \leq \|\mu_{v^{(n)}}^{Z_A} - \mu_v^{Z_A}\|_{\text{TV}} + \|\text{Law}(D_n D^{-1} Y_A) - \text{Law}(Y_A)\|_{\text{TV}},$$

where Y_A has law $\mu_v^{Y_A}$. The law of Y_A has a Lebesgue density $\rho \in L^1(\mathbb{R}^A)$, and the law of $D_n D^{-1} Y_A$ has the density $\rho(DD_n^{-1} y) \det(DD_n^{-1})$. Since $D_n D^{-1} \rightarrow \text{Id}$, this density converges to ρ in L^1 : this is clear for continuous ρ with compact support, by dominated convergence, and the general case follows by approximation, since the change of variables preserves the L^1 norm. This proves the convergence of $\mu_{v^{(n)}}^{Y_A}$. Finally, the same Fourier

argument shows that $p_{v^{(n)},B} \rightarrow p_{v,B}$ and $p_{v^{(n)},I_n \setminus B} \rightarrow p_{v,I \setminus B}$ uniformly, so that their suprema converge, and together with $p_{v^{(n)}}(0) \rightarrow p_v(0)$ this gives $M(v^{(n)}, B) \rightarrow M(v, B)$. $\square$

Likelihood ratios on a common sphere

This section compares two Gaussian laws conditioned on the same sphere, μ_v and $\mu_{\tilde{v}}$ with $\sum_i v_i = \sum_i \tilde{v}_i$. We write $\tilde{v}_i = v_i(1 + e_i)$, so that $I = \mathcal{I}(v, \tilde{v}) = \frac{1}{2} \sum_i e_i^2$ is the information of $\tilde{v}$ relative to v (Definition 3.5). The main result, Theorem 6.6, states that under μ_v the logarithm of the likelihood ratio $d\mu_{\tilde{v}}/d\mu_v$ is close in the Wasserstein distance W_1 to the Gaussian law $\mathcal{N}(-I/2, I)$, with an error $C_B(\eta_v + \max_i |e_i|)$ that is uniform over $I \leq B$, including $I = 0$. As a consequence, $\|\mu_{\tilde{v}} - \mu_v\|_{\text{TV}}$ is close to $2\Phi(\sqrt{I}/2) - 1$. With $I = e^{-4u}$, this is the Gaussian profile $\Psi(u)$ of Section 3, which is how Ψ arises at high temperature. In Theorem 6.8, we deduce the corresponding limit statements, including convergence of exponential moments and the case of diverging information.

The proof has four ingredients: an identity expressing total variation through the likelihood ratio (Lemma 6.1); a bound on inverse moments of a positive quadratic form under μ_v (Lemma 6.2); an exact integration by parts formula for quadratic statistics under μ_v (Lemma 6.3); and Stein's method, which turns this formula into a normal approximation for quadratic statistics (Lemma 6.4 and Theorem 6.5). The log likelihood ratio is a quadratic statistic plus a constant, and the information I is its asymptotic variance. We use Theorems 5.3, 5.11, and 5.15, Lemmas 5.9 and 5.13, and Corollary 5.16 from Section 5. The results are used in the high-temperature profiles of Sections 13 to 15 and in the outer limit of Section 20.

The Gaussian law $\mathcal{N}(-I/2, I)$ and the profile $2\Phi(\sqrt{I}/2) - 1$ are those of the Gaussian shift experiment in Le Cam's theory of local asymptotic normality [104, Section 2, pp. 40–43]; see [146, Chapters 6–7]. The results below concern the log likelihood ratio of two fixed laws, or of finitely many, and do not involve a parametrized family of laws. Normal approximation by an integration by parts identity is Stein's method [141]; see [116, Sections 3.1–3.5] and [42, Chapter 2] for modern accounts.

Conventions. Throughout this section, all variance sequences are finite, indexed by $\{1, \dots, n\}$, and μ_v is the law of Definition 5.4 on the sphere $\mathcal{S}_{R^2}$, $R^2 = \sum_i v_i$. We keep the notation S_v , $\hat{v}_i$, η_v of (5.1) and write $Z_i = Y_i / \sqrt{\hat{v}_i}$ under μ_v . Recall that $|Y|^2 = R^2$ under μ_v ; dividing $|Y|^2 - R^2 = \sum_i v_i(Z_i^2 - 1)$ by S_v , we obtain the identity

$$\sum_{i=1}^n \hat{v}_i (Z_i^2 - 1) = 0 \quad \text{everywhere on } \mathcal{S}_{R^2}. \quad (6.1)$$

For probability measures λ_1, λ_2 on $\mathbb{R}$ with finite first moments, $W_1(\lambda_1, \lambda_2) = \sup_h |\int h d\lambda_1 - \int h d\lambda_2|$, the supremum being over 1-Lipschitz functions h . We write $\text{Law}_\mu(X)$ for the law of X under μ , and $\mathcal{N}(m, 0)$ is the point mass at m . For $w \in \mathbb{R}^n$, $\|w\|_2$ and $\|w\|_\infty$ are the Euclidean and maximum norms, and $\langle \cdot, \cdot \rangle$ is the Euclidean inner product.

6.1 Total variation and likelihood ratios

Lemma 6.1 (Total variation through the likelihood ratio). *Let P, Q be probability measures on a common space, with Q absolutely continuous with respect to P , and let $\ell = \log(dQ/dP) \in [-\infty, \infty)$. Then*

$$\|P - Q\|_{\text{TV}} = \mathbb{E}_P(1 - e^\ell)_+. \quad (6.2)$$

Let $I \geq 0$ and let X have law $\mathcal{N}(-I/2, I)$. Then

$$\mathbb{E}(1 - e^X)_+ = 2\Phi(\sqrt{I}/2) - 1, \quad \mathbb{E}e^{tX} = e^{(t^2 - t)I/2} \quad (t \in \mathbb{R}). \quad (6.3)$$

Further, the function $x \mapsto (1 - e^x)_+$ is bounded by one and 1-Lipschitz on $\mathbb{R}$, which gives $|\mathbb{E}_P(1 - e^\ell)_+ - \mathbb{E}(1 - e^X)_+| \leq W_1(\text{Law}_P(\ell), \mathcal{N}(-I/2, I))$ whenever ℓ is real-valued and integrable under P .

Note that $\mathcal{N}(-I/2, I)$ is the law of $\log(d\mathcal{N}(\sqrt{I}, 1)/d\mathcal{N}(0, 1))$ under $\mathcal{N}(0, 1)$, so that (6.2) and (6.3) imply that $\|\mathcal{N}(\sqrt{I}, 1) - \mathcal{N}(0, 1)\|_{\text{TV}} = 2\Phi(\sqrt{I}/2) - 1$.

Proof. Let $L = dQ/dP = e^\ell$. Note that $\mathbb{E}_P(1 - L) = 0$, so that $\mathbb{E}_P(1 - L)_+ = \mathbb{E}_P(1 - L)_- = \frac{1}{2}\mathbb{E}_P|1 - L|$, and $\frac{1}{2}\mathbb{E}_P|1 - L| = \frac{1}{2}\int |dP - dQ| = \|P - Q\|_{\text{TV}}$. This proves (6.2). The first identity in (6.3) is immediate at $I = 0$, when $X = 0$; for its proof we may therefore assume $I > 0$. We write $X = -I/2 + \sqrt{I}G$ and $x = \sqrt{I}/2$, so that $X < 0$ exactly when $G < x$. Completing the square, we have $e^{-I/2 + \sqrt{I}g}\phi(g) = \phi(g - \sqrt{I})$, and

$$\mathbb{E}(1 - e^X)_+ = \Phi(x) - \int_{-\infty}^x \phi(g - \sqrt{I}) dg = \Phi(x) - \Phi(x - \sqrt{I}) = 2\Phi(x) - 1,$$

where we used $x - \sqrt{I} = -x$. The second identity in (6.3) is the Gaussian moment generating function. The function $(1 - e^x)_+$ vanishes for $x \geq 0$ and has derivative $-e^x \in [-1, 0)$ for $x < 0$, which implies the last statement. $\square$

6.2 Integration by parts under the conditioned law

Lemma 6.2 (Inverse moments of a quadratic form). *There exist universal constants $\eta_{\text{inv}} > 0$ and $C < \infty$ such that every finite variance sequence v with $\eta_v \leq \eta_{\text{inv}}$ satisfies*

$$\mathbb{E}_{\mu_v} V^{-p} \leq C, \quad p = 1, 2, \quad \text{where} \quad V = \sum_{i=1}^n \hat{\sigma}_i^2 Z_i^2. \quad (6.4)$$

The constant does not depend on the dimension or on the smallest variance. Note that $V > 0$ everywhere on $\mathcal{S}_{R^2}$.

Proof. We fix η_{inv} such that $\eta_{\text{inv}}^2 \leq 1/48$ and η_{inv} is at most the constant η_* of Theorem 5.11. We add indices $1, 2, \dots$ to a set J until $q_J = \sum_{i \in J} \hat{v}_i^2$ first reaches $1/6$. This happens, since $\sum_i \hat{v}_i^2 = 1/2$, and since the last index added contributes at most $\eta_v^2 \leq 1/48$, we have $1/6 \leq q_J \leq 3/16$. The fraction of the variance of T_v carried by J is $h_J = 2q_J \leq 3/8 \leq 1/2$, and Theorem 5.11 implies that $r_{v,J} \leq C$. Using $V \geq \sum_{i \in J} \hat{v}_i^2 Z_i^2$ and Lemma 5.9, we obtain

$$\mathbb{E}_{\mu_v} V^{-p} \leq \int \left(\sum_{i \in J} \hat{v}_i^2 Z_i^2 \right)^{-p} r_{v,J}(z) \gamma_J(\mathrm{d}z) \leq C \mathbb{E} \left(\sum_{i \in J} \hat{v}_i^2 g_i^2 \right)^{-p}.$$

Using $x^{-p} = \Gamma(p)^{-1} \int_0^\infty t^{p-1} e^{-tx} \mathrm{d}t$, Tonelli's theorem, and $\mathbb{E} e^{-s g^2} = (1 + 2s)^{-1/2}$, we have

$$\mathbb{E} \left(\sum_{i \in J} \hat{v}_i^2 g_i^2 \right)^{-p} = \frac{1}{\Gamma(p)} \int_0^\infty t^{p-1} \prod_{i \in J} (1 + 2t \hat{v}_i^2)^{-1/2} \mathrm{d}t.$$

Let $m_J = \max_{i \in J} \hat{v}_i^2 \leq \eta_v^2$. The function $x \mapsto \log(1 + 2tx)$ is concave and lies above its chord on $[0, m_J]$, that is, $\log(1 + 2t \hat{v}_i^2) \geq (\hat{v}_i^2 / m_J) \log(1 + 2tm_J)$ for $i \in J$. Then the product is at most $(1 + 2tm_J)^{-a}$ with $a = q_J / (2m_J)$, and with the substitution $s = 2tm_J$ and the beta integral we find

$$\frac{1}{\Gamma(p)} \int_0^\infty t^{p-1} (1 + 2tm_J)^{-a} \mathrm{d}t = \frac{\Gamma(a-p)}{(2m_J)^p \Gamma(a)} = \prod_{r=1}^p (q_J - 2rm_J)^{-1}, \quad p = 1, 2.$$

Since $2pm_J \leq 4\eta_v^2 \leq 1/12 \leq q_J/2$, each factor satisfies $q_J - 2rm_J \geq q_J/2 \geq 1/12$; in particular $a > p$, and the integral converges. This bounds the product by 12^p and proves (6.4). $\square$

Lemma 6.3 (Integration by parts for quadratic statistics). *Let v be a finite variance sequence, and let $w \in \mathbb{R}^n$ satisfy $\sum_i \hat{v}_i w_i = 0$. Define, on $\mathcal{S}_{R^2}$,*

$$\begin{aligned} Q_w &= \sum_i w_i (Z_i^2 - 1), & V &= \sum_i \hat{v}_i^2 Z_i^2, & K_w &= \sum_i \hat{v}_i w_i Z_i^2, \\ J_w &= \sum_i w_i^2 Z_i^2, & L_w &= \sum_i \hat{v}_i^2 w_i Z_i^2, & V_3 &= \sum_i \hat{v}_i^3 Z_i^2, \end{aligned}$$

and

$$A_w = 2J_w - \frac{2K_w^2}{V}, \quad R_w = \frac{2L_w}{V} - \frac{2K_w V_3}{V^2}. \quad (6.5)$$

For every continuously differentiable $f : \mathbb{R} \rightarrow \mathbb{R}$,

$$\mathbb{E}_{\mu_v} [(Q_w + R_w) f(Q_w)] = \mathbb{E}_{\mu_v} [A_w f'(Q_w)]. \quad (6.6)$$

Further, there exist universal constants $\eta_{\text{ibp}} > 0$ and $C < \infty$ such that, if $\eta_v \leq \eta_{\text{ibp}}$, then

$$\begin{aligned} \mathbb{E}_{\mu_v} |A_w - 2\|w\|_2^2| &\leq C(\|w\|_\infty \|w\|_2 + \eta_v^2 \|w\|_2^2), \\ \mathbb{E}_{\mu_v} |R_w| &\leq C\eta_v \|w\|_2, \quad \mathbb{E}_{\mu_v} |R_w Q_w| \leq C\eta_v \|w\|_2^2. \end{aligned} \quad (6.7)$$

The identity (6.6) is exact and holds without any smallness assumption. If μ_v were replaced by the law of independent Gaussian variables, the corresponding identity would hold with $R_w = 0$ and $A_w = 2J_w$; the term $-2K_w^2/V$ and the correction R_w come from the constraint (6.1).

Proof. We first prove the identity, using a vector field tangent to the spheres $\mathcal{S}_r$, and then bound the coefficients.

Step 1: a divergence identity on spheres. Let X be a continuously differentiable vector field on $\mathbb{R}^n \setminus \{0\}$ with $X(y) \cdot y = 0$ for all y . We claim that $\int_{\mathcal{S}_r} \operatorname{div} X \, d\mathcal{H}^{n-1} = 0$ for every $r > 0$. Indeed, for $0 < a < b$, the divergence theorem on the shell $\{a \leq |y| \leq b\}$ gives $\int_{a \leq |y| \leq b} \operatorname{div} X \, dy = 0$, since the outward normals at the two boundary spheres are $\pm y/|y|$ and $X \cdot y = 0$ there. In polar coordinates, this says that $\int_a^b (\int_{\mathcal{S}_s} \operatorname{div} X \, d\mathcal{H}^{n-1}) \, ds = 0$ for all $0 < a < b$. The inner integral is continuous in s , and it vanishes for every $s > 0$.

Step 2: the identity (6.6). We regard $z_i = y_i / \sqrt{v_i}$, and hence $Q_w, V, K_w, J_w, L_w, V_3, A_w, R_w$, as functions of $y \in \mathbb{R}^n$; they are smooth on $\mathbb{R}^n \setminus \{0\}$, where $V > 0$. Let $\alpha = K_w/V$ and define the vector field $\tilde{T}$ on $\mathbb{R}^n \setminus \{0\}$ by

$$\tilde{T}_i(y) = \sqrt{v_i} (w_i - \alpha \hat{v}_i) z_i.$$

Using $\sqrt{v_i} z_i = y_i$ and $v_i = S_v \hat{v}_i$, we have

$$\tilde{T}(y) \cdot y = S_v \sum_i \hat{v}_i (w_i - \alpha \hat{v}_i) z_i^2 = S_v (K_w - \alpha V) = 0.$$

Recalling that $\partial_{y_i} = v_i^{-1/2} \partial_{z_i}$, we find by direct computation that

$$\begin{aligned} \operatorname{div} \tilde{T} &= \sum_i (w_i - \alpha \hat{v}_i) - \sum_i \hat{v}_i z_i \partial_{z_i} \alpha, \quad \tilde{T} \cdot \nabla \log \rho = - \sum_i (w_i - \alpha \hat{v}_i) z_i^2, \\ \tilde{T} \cdot \nabla Q_w &= 2J_w - 2\alpha K_w = A_w, \end{aligned}$$

where $\rho(y) = \exp\{-\frac{1}{2} \sum_i y_i^2 / v_i\}$. Further, $\partial_{z_i} \alpha = 2\hat{v}_i w_i z_i / V - 2K_w \hat{v}_i^2 z_i / V^2$, which gives $\sum_i \hat{v}_i z_i \partial_{z_i} \alpha = R_w$. Adding the first two identities gives

$$\rho^{-1} \operatorname{div} (\rho \tilde{T}) = - \sum_i (w_i - \alpha \hat{v}_i) (z_i^2 - 1) - R_w = -Q_w + \alpha \sum_i \hat{v}_i (z_i^2 - 1) - R_w.$$

On $\mathcal{S}_{R^2}$, the middle term vanishes by (6.1). Now apply Step 1 to $X = \rho f(Q_w) \tilde{T}$, which is continuously differentiable on $\mathbb{R}^n \setminus \{0\}$ and satisfies $X \cdot y = 0$. Since $\operatorname{div} X = \rho[f(Q_w)\rho^{-1} \operatorname{div}(\rho \tilde{T}) + f'(Q_w) \tilde{T} \cdot \nabla Q_w]$, we obtain

$$\int_{\mathcal{S}_{R^2}} \rho[-(Q_w + R_w)f(Q_w) + A_w f'(Q_w)] d\mathcal{H}^{n-1} = 0.$$

The density of μ_v with respect to $\mathcal{H}^{n-1}$ on $\mathcal{S}_{R^2}$ is proportional to ρ , by (5.10), and this identity is (6.6).

Step 3: the error bounds. We take $\eta_{\text{ibp}} \leq \min\{\eta_0, \eta_{\text{inv}}\}$, where η_0 is the constant of Theorem 5.15 and η_{inv} that of Lemma 6.2. All norms below are in $L^p(\mu_v)$. We have $\|Q_w\|_4 \leq C\|w\|_2$ by Theorem 5.15. Since $\sum_i \hat{v}_i w_i = 0$, we have $K_w = \sum_i \hat{v}_i w_i (Z_i^2 - 1)$, and $\|K_w\|_4 \leq C\eta_v \|w\|_2$ by Theorem 5.15 with coefficients $\hat{v}_i w_i$. Similarly, using $J_w - \|w\|_2^2 = \sum_i w_i^2 (Z_i^2 - 1)$, we find that

$$\|J_w - \|w\|_2^2\|_1 \leq \|J_w - \|w\|_2^2\|_2 \leq C\left(\sum_i w_i^4\right)^{1/2} \leq C\|w\|_\infty \|w\|_2.$$

Next, $L_w = \sum_i \hat{v}_i^2 w_i + \sum_i \hat{v}_i^2 w_i (Z_i^2 - 1)$. By the Cauchy–Schwarz inequality, $|\sum_i \hat{v}_i^2 w_i| \leq (\sum_i \hat{v}_i^2)^{1/2} (\sum_i \hat{v}_i^2 w_i^2)^{1/2} \leq \eta_v \|w\|_2$, and Theorem 5.15 bounds the second sum in L^4 by $C(\sum_i \hat{v}_i^4 w_i^2)^{1/2} \leq C\eta_v^2 \|w\|_2$. We conclude that $\|L_w\|_4 \leq C\eta_v \|w\|_2$. Finally, $0 \leq V_3 \leq \eta_v V$ pointwise, so that $|R_w| \leq 2(|L_w| + \eta_v |K_w|)/V$, and $\|V^{-1}\|_2 \leq C$ by Lemma 6.2.

We now combine these bounds. By Hölder’s inequality,

$$\mathbb{E}|A_w - 2\|w\|_2^2| \leq 2\|J_w - \|w\|_2^2\|_1 + 2\|K_w\|_4^2 \|V^{-1}\|_2 \leq C(\|w\|_\infty \|w\|_2 + \eta_v^2 \|w\|_2^2).$$

Using the Cauchy–Schwarz inequality, we find

$$\mathbb{E}|R_w| \leq 2(\|L_w\|_2 + \eta_v \|K_w\|_2) \|V^{-1}\|_2 \leq C\eta_v \|w\|_2,$$

and by Hölder’s inequality with exponents 4, 4, and 2,

$$\mathbb{E}|R_w Q_w| \leq 2(\|L_w\|_4 + \eta_v \|K_w\|_4) \|Q_w\|_4 \|V^{-1}\|_2 \leq C\eta_v \|w\|_2^2.$$

This proves (6.7). □

6.3 Normal approximation of quadratic statistics

Lemma 6.4 (Solutions of the Stein equation). *There exists a universal constant C with the following property. Let $h : \mathbb{R} \rightarrow \mathbb{R}$ be smooth, and suppose that h is 1-Lipschitz or that $\|h\|_\infty \leq 1$. Then there exists a smooth bounded solution f of*

$$f'(x) - xf(x) = h(x) - \mathbb{E}h(G), \quad x \in \mathbb{R}, \quad (6.8)$$

with $\|f\|_\infty + \|f'\|_\infty \leq C$.

For background on Stein's equation and the generator approach, see [141] and [42, Chapter 2]. We prove the bounds here.

Proof. We set $\bar{h} = h - \mathbb{E}h(G)$ and define

$$f(x) = e^{x^2/2} \int_{-\infty}^x \bar{h}(u) e^{-u^2/2} du = -e^{x^2/2} \int_x^{\infty} \bar{h}(u) e^{-u^2/2} du,$$

where the two expressions agree since $\int \bar{h} \phi = 0$. Differentiation shows that f is smooth and solves (6.8). It is the only bounded solution, since the solutions of the homogeneous equation are the multiples of $e^{x^2/2}$.

Suppose $\|h\|_{\infty} \leq 1$, so that $|\bar{h}| \leq 2$. For $x > 0$, the substitution $u = x + s$ shows that

$$e^{x^2/2} \int_x^{\infty} e^{-u^2/2} du = \int_0^{\infty} e^{-xs - s^2/2} ds \leq \min \{ \sqrt{\pi/2}, 1/x \},$$

and the second expression for f gives $|f(x)| \leq 2\sqrt{\pi/2}$ and $|xf(x)| \leq 2$. The case $x < 0$ follows from the first expression in the same way, and the bounds at $x = 0$ follow by continuity. Finally, (6.8) gives $|f'| \leq |xf| + |\bar{h}| \leq 4$.

Suppose h is 1-Lipschitz, and assume first that all positive-order derivatives of h are bounded. Let $P_t h(x) = \mathbb{E}h(e^{-t}x + s_t G)$ with $s_t = (1 - e^{-2t})^{1/2}$ be the Ornstein–Uhlenbeck semigroup, and let

$$F(x) = - \int_0^{\infty} e^{-t} P_t(h')(x) dt.$$

Then $\|F\|_{\infty} \leq \|h'\|_{\infty} \leq 1$. For $t > 0$, Gaussian integration by parts gives $(P_t(h'))'(x) = e^{-t} \mathbb{E}[h''(e^{-t}x + s_t G)] = e^{-t} s_t^{-1} \mathbb{E}[G h'(e^{-t}x + s_t G)]$, and

$$|F'(x)| \leq \int_0^{\infty} \frac{e^{-2t}}{s_t} \mathbb{E}|G| dt = \mathbb{E}|G| = \sqrt{2/\pi},$$

where we used $\int_0^{\infty} e^{-2t} (1 - e^{-2t})^{-1/2} dt = 1$. Further, $(P_t h)' = e^{-t} P_t(h')$, and $\partial_t P_t h = (P_t h)'' - x(P_t h)'$ by direct differentiation of the Gaussian integral. We obtain

$$F'(x) - xF(x) = - \int_0^{\infty} [(P_t h)''(x) - x(P_t h)'(x)] dt = - \int_0^{\infty} \partial_t P_t h(x) dt = h(x) - \mathbb{E}h(G),$$

since $P_0 h = h$ and $P_t h(x) \rightarrow \mathbb{E}h(G)$ as $t \rightarrow \infty$. The bounds on the integrands justify the interchanges of derivatives and integrals. We conclude that F is a bounded solution of (6.8), and $F = f$. For a general smooth 1-Lipschitz h , let $h_{\varepsilon} = h * \rho_{\varepsilon}$ for a nonnegative smooth compactly supported mollifier of unit mass. Then h_{ε} remains 1-Lipschitz, has bounded derivatives of every positive order, and converges uniformly to h . The integral formula gives $f_{\varepsilon}(x) \rightarrow f(x)$, while $f'_{\varepsilon}(x) = x f_{\varepsilon}(x) + h_{\varepsilon}(x) - \mathbb{E}h_{\varepsilon}(G) \rightarrow f'(x)$ for each x . Thus $\|f\|_{\infty} \leq 1$ and $\|f'\|_{\infty} \leq \sqrt{2/\pi}$. $\square$

Theorem 6.5 (Normal approximation of quadratic statistics). *There exist universal constants $\eta_{\text{nor}} > 0$ and $C < \infty$ with the following property. Let v be a finite variance sequence with $\eta_v \leq \eta_{\text{nor}}$, and let $w \in \mathbb{R}^n$ be arbitrary. Let*

$$w^\perp = w - 2\langle \hat{v}, w \rangle \hat{v}, \quad \sigma_w^2 = 2\|w^\perp\|_2^2, \quad Q_w = \sum_i w_i(Z_i^2 - 1).$$

If $\sigma_w = 0$, then $Q_w = 0$ on $\mathcal{S}_{R^2}$. In all cases,

$$W_1(\text{Law}_{\mu_v}(Q_w), \mathcal{N}(0, \sigma_w^2)) \leq C(\eta_v \sigma_w + \|w^\perp\|_\infty). \quad (6.9)$$

If $\sigma_w > 0$, then also

$$\|\text{Law}_{\mu_v}(Q_w) - \mathcal{N}(0, \sigma_w^2)\|_{\text{TV}} \leq \min\{1, C(\eta_v + \|w^\perp\|_\infty / \sigma_w)\}. \quad (6.10)$$

Further,

$$|\mathbb{E}_{\mu_v} Q_w| \leq C(\eta_v \sigma_w + \|w^\perp\|_\infty), \quad (6.11)$$

$$|\mathbb{E}_{\mu_v} Q_w^2 - \sigma_w^2| \leq C(\eta_v \sigma_w^2 + \sigma_w \|w^\perp\|_\infty). \quad (6.12)$$

Since $\|\hat{v}\|_2^2 = 1/2$, the vector $w^\perp$ is the orthogonal projection of w onto the orthogonal complement of $\hat{v}$. The variance σ_w^2 is therefore no greater than the variance $2\|w\|_2^2$ of $\sum_i w_i(g_i^2 - 1)$ for independent Gaussian variables, by the part of w along the constrained direction. The constants do not depend on the dimension or on the smallest variance.

Proof. We set $\eta_{\text{nor}} = \eta_{\text{ibp}}$, the constant of Lemma 6.3. By (6.1), we have $Q_w - Q_{w^\perp} = 2\langle \hat{v}, w \rangle \sum_i \hat{v}_i(Z_i^2 - 1) = 0$, which means that $Q_w = Q_{w^\perp}$ on $\mathcal{S}_{R^2}$; also $\langle \hat{v}, w^\perp \rangle = 0$. If $\sigma_w = 0$, then $w^\perp = 0$ and $Q_w = 0$. Let $\sigma_w > 0$ and $u = w^\perp / \sigma_w$, so that $\langle \hat{v}, u \rangle = 0$, $2\|u\|_2^2 = 1$, and $Q_u = Q_w / \sigma_w$.

Let h be smooth, and either 1-Lipschitz with bounded derivatives of every positive order or bounded by one, and let f be the solution of (6.8) from Lemma 6.4. By (6.6) for u , we have $\mathbb{E}[Q_u f(Q_u)] = \mathbb{E}[A_u f'(Q_u)] - \mathbb{E}[R_u f(Q_u)]$, and

$$\mathbb{E}h(Q_u) - \mathbb{E}h(G) = \mathbb{E}[f'(Q_u) - Q_u f(Q_u)] = \mathbb{E}[(1 - A_u)f'(Q_u) + R_u f(Q_u)].$$

Since $1 = 2\|u\|_2^2$, the bounds (6.7) imply that

$$|\mathbb{E}h(Q_u) - \mathbb{E}h(G)| \leq C(\mathbb{E}|A_u - 2\|u\|_2^2| + \mathbb{E}|R_u|) \leq C(\eta_v + \|w^\perp\|_\infty / \sigma_w), \quad (6.13)$$

where we used $\|u\|_\infty \|u\|_2 \leq \|w^\perp\|_\infty / \sigma_w$ and $\eta_v^2 \leq \eta_v$.

Let h be 1-Lipschitz. The convolution h_δ of h with a smooth probability density supported in $[-\delta, \delta]$ is smooth, 1-Lipschitz, has bounded derivatives of every positive order, and satisfies $\|h_\delta - h\|_\infty \leq \delta$. Applying (6.13) to h_δ and letting $\delta \rightarrow 0$, we find that (6.13) holds for h . Taking the supremum over h , we obtain $W_1(\text{Law}(Q_u), \mathcal{N}(0, 1)) \leq$

$C(\eta_v + \|w^\perp\|_\infty/\sigma_w)$. Since $W_1(\text{Law}(\sigma X), \text{Law}(\sigma X')) = \sigma W_1(\text{Law}(X), \text{Law}(X'))$ for $\sigma > 0$, this is (6.9).

For (6.10), we claim that the total variation distance between two probability measures λ_1, λ_2 on $\mathbb{R}$ is the supremum of $|\int h d\lambda_1 - \int h d\lambda_2|$ over smooth h with $0 \leq h \leq 1$. Indeed, given a Borel set E and $\delta > 0$, by regularity of the finite measure $\lambda_1 + \lambda_2$ there exist a compact K and an open O with $K \subset E \subset O$ and $(\lambda_1 + \lambda_2)(O \setminus K) \leq \delta$. Covering K by finitely many open intervals whose closures lie in O , we find a smooth h with $\mathbf{1}_K \leq h \leq \mathbf{1}_O$, and then $|\lambda_j(E) - \int h d\lambda_j| \leq \delta$ for $j = 1, 2$. By the claim and (6.13) for smooth h with values in $[0, 1]$, we have $\|\text{Law}(Q_u) - \mathcal{N}(0, 1)\|_{\text{TV}} \leq C(\eta_v + \|w^\perp\|_\infty/\sigma_w)$. Total variation is invariant under multiplication of both variables by σ_w , and it is at most one, which proves (6.10).

For the moment bounds, we apply (6.6) to $w^\perp$ with $f(x) = 1$ and $f(x) = x$, which gives

$$\mathbb{E}Q_{w^\perp} = -\mathbb{E}R_{w^\perp}, \quad \mathbb{E}Q_{w^\perp}^2 = \mathbb{E}A_{w^\perp} - \mathbb{E}[R_{w^\perp}Q_{w^\perp}].$$

Using $Q_w = Q_{w^\perp}$, $2\|w^\perp\|_2^2 = \sigma_w^2$, and (6.7), we have $|\mathbb{E}Q_w| \leq C\eta_v\|w^\perp\|_2 \leq C\eta_v\sigma_w$, which implies (6.11), and

$$|\mathbb{E}Q_w^2 - \sigma_w^2| \leq \mathbb{E}|A_{w^\perp} - 2\|w^\perp\|_2^2| + \mathbb{E}|R_{w^\perp}Q_{w^\perp}| \leq C(\|w^\perp\|_\infty\sigma_w + \eta_v\sigma_w^2),$$

which is (6.12). □

6.4 Likelihood ratios

Let v and $\tilde{v}$ be finite variance sequences on $\{1, \dots, n\}$ with $\sum_i v_i = \sum_i \tilde{v}_i = R^2$. Then μ_v and $\mu_{\tilde{v}}$ are laws on the same sphere $\mathcal{S}_{R^2}$, with positive continuous densities, so they are mutually absolutely continuous. We write

$$\tilde{v}_i = v_i(1 + e_i), \quad \tilde{e}_i = \frac{e_i}{1 + e_i}, \quad \bar{e} = \max_i |e_i|, \quad I = \frac{1}{2} \sum_i e_i^2, \quad (6.14)$$

where $I = \mathcal{I}(v, \tilde{v})$ is the information of $\tilde{v}$ relative to v (Definition 3.5). The equality of the traces means that $\sum_i v_i e_i = 0$, that is,

$$\sum_i \hat{v}_i e_i = 0. \quad (6.15)$$

By (5.12), $d\mu_v/d\mathcal{H}^{n-1} = \gamma_v/(2Rp_v(0))$ on $\mathcal{S}_{R^2}$, and $\gamma_{\tilde{v}}(y)/\gamma_v(y) = \prod_i (1 + e_i)^{-1/2} \exp\{\frac{1}{2} \sum_i \tilde{e}_i y_i^2 / v_i\}$, since $v_i^{-1} - \tilde{v}_i^{-1} = \tilde{e}_i / v_i$. We conclude that the log likelihood ratio $\ell = \log(d\mu_{\tilde{v}}/d\mu_v)$ is, exactly,

$$\ell = k_0 + Q_w, \quad w = \frac{1}{2}\tilde{e}, \quad k_0 = \frac{1}{2} \sum_i [\tilde{e}_i - \log(1 + e_i)] + \log \frac{p_v(0)}{p_{\tilde{v}}(0)}, \quad (6.16)$$

where $Q_w = \sum_i w_i(Z_i^2 - 1)$ as in Theorem 6.5. Equivalently, with $\ell^{\text{prod}} = -\frac{1}{2} \sum_i \log(1 + e_i) + \frac{1}{2} \sum_i \tilde{e}_i Z_i^2$ the log likelihood ratio of the unconditioned Gaussian laws,

$$e^\ell = e^{\ell^{\text{prod}}} \frac{p_v(0)}{p_{\tilde{v}}(0)}. \quad (6.17)$$

Theorem 6.6 (Quantitative likelihood ratios on a common sphere). *There exists a universal constant $\eta_{\text{lik}} > 0$ such that, for every $B < \infty$, there exists $C_B < \infty$ with the following property. Let $v, \tilde{v}$ be finite variance sequences with $\sum_i v_i = \sum_i \tilde{v}_i$, and define $e_i, \bar{e}, I$ by (6.14). Suppose that*

$$\bar{e} \leq \frac{1}{4}, \quad \eta_v \leq \eta_{\text{lik}}, \quad I \leq B.$$

Then, with $\ell = \log(d\mu_{\tilde{v}}/d\mu_v)$,

$$W_1(\text{Law}_{\mu_v}(\ell), \mathcal{N}(-I/2, I)) \leq C_B(\eta_v + \bar{e}), \quad (6.18)$$

and

$$\left| \|\mu_{\tilde{v}} - \mu_v\|_{\text{TV}} - (2\Phi(\sqrt{I}/2) - 1) \right| \leq C_B(\eta_v + \bar{e}). \quad (6.19)$$

The threshold η_{lik} does not depend on B , and the bounds hold uniformly down to $I = 0$. The constants do not depend on the dimension or on the smallest variance.

Proof. If $I = 0$, then $\tilde{v} = v$, $\ell = 0$, and the left sides of (6.18) and (6.19) vanish, so both inequalities hold. Let $I > 0$. We use the exact decomposition (6.16): we show that the constant k_0 is close to $-I/2$, that the projected coefficient vector $w^\perp$ of Theorem 6.5 has small entries and norm close to $\sqrt{I/2}$, and then apply Theorem 6.5 and Lemma 6.1. We take $\eta_{\text{lik}} \leq \min\{\frac{3}{5}\eta_{\text{loc}}, \eta_{\text{nor}}\}$, where η_{loc} is the constant of Theorem 5.3 and η_{nor} that of Theorem 6.5. Note that $|\tilde{e}_i| \leq |e_i|/(1 - \bar{e}) \leq \frac{4}{3}|e_i|$.

Step 1: the constant k_0 . Recall that $S_{\tilde{v}}^2 = 2 \sum_i v_i^2(1 + e_i)^2$, so that $1 - \bar{e} \leq S_{\tilde{v}}/S_v \leq 1 + \bar{e}$ and $\eta_{\tilde{v}} \leq \frac{1+\bar{e}}{1-\bar{e}}\eta_v \leq \frac{5}{3}\eta_v \leq \eta_{\text{loc}}$. Writing $p_v(0) = f_v(0)/S_v$, and similarly for $\tilde{v}$, we have

$$\log \frac{p_v(0)}{p_{\tilde{v}}(0)} = \log \frac{f_v(0)}{f_{\tilde{v}}(0)} + \log \frac{S_{\tilde{v}}}{S_v}.$$

By Theorem 5.3, $f_v(0)$ and $f_{\tilde{v}}(0)$ are at least $\phi(0)/2$ and differ from $\phi(0)$ by at most $C\eta_v$, and the first logarithm is at most $C\eta_v$ in absolute value; the second is at most $\frac{4}{3}\bar{e}$. Thus

$$\left| \log \frac{p_v(0)}{p_{\tilde{v}}(0)} \right| \leq C(\eta_v + \bar{e}). \quad (6.20)$$

Next, the function $\kappa(e) = \frac{1}{2}[e/(1+e) - \log(1+e)] + e^2/4$ satisfies $\kappa(0) = \kappa'(0) = \kappa''(0) = 0$ and $\kappa'''(e) = (2-e)(1+e)^{-4}$, which is at most 8 for $|e| \leq 1/4$. Taylor's formula gives $|\kappa(e)| \leq \frac{4}{3}|e|^3$ for $|e| \leq 1/4$. Summing over i , we obtain

$$\left| \frac{1}{2} \sum_i [\tilde{e}_i - \log(1 + e_i)] + \frac{I}{2} \right| \leq \frac{4}{3} \sum_i |e_i|^3 \leq \frac{8}{3}\bar{e}I.$$

We conclude that

$$|k_0 + I/2| \leq C(\eta_v + \bar{e}) + 3\bar{e}I \leq C_B(\eta_v + \bar{e}). \quad (6.21)$$

Step 2: the projected coefficients. With $w = \frac{1}{2}\tilde{e}$, we have $2\langle \hat{v}, w \rangle = \langle \hat{v}, \tilde{e} \rangle$ and $w^\perp = \frac{1}{2}\tilde{e} - \langle \hat{v}, \tilde{e} \rangle \hat{v}$. By (6.15) and $\tilde{e}_i - e_i = -e_i^2/(1 + e_i)$,

$$|\langle \hat{v}, \tilde{e} \rangle| = \left| \sum_i \hat{v}_i \frac{e_i^2}{1 + e_i} \right| \leq \frac{4}{3} \eta_v \sum_i e_i^2 = \frac{8}{3} \eta_v I.$$

This cancellation, which comes from the equality of the traces, keeps the projection small, and it gives

$$\|w^\perp\|_\infty \leq \frac{2}{3}\bar{e} + \frac{8}{3}\eta_v^2 I.$$

Recall that $w^\perp$ is the orthogonal projection of w onto the complement of $\hat{v}$ and $\|\hat{v}\|_2^2 = 1/2$, so that

$$\sigma_w^2 = 2\|w^\perp\|_2^2 = \frac{1}{2}\|\tilde{e}\|_2^2 - \langle \hat{v}, \tilde{e} \rangle^2.$$

Using $|\tilde{e}_i^2 - e_i^2| = e_i^2|(1 + e_i)^{-2} - 1| \leq 4\bar{e}e_i^2$, we find

$$|\sigma_w^2 - I| \leq 4\bar{e}I + 8\eta_v^2 I^2, \quad \sigma_w^2 \leq \frac{1}{2}\|\tilde{e}\|_2^2 \leq 2I.$$

Dividing by $\sigma_w + \sqrt{I} \geq \sqrt{I}$, we find

$$|\sigma_w - \sqrt{I}| \leq 4\bar{e}\sqrt{I} + 8\eta_v^2 I^{3/2}. \quad (6.22)$$

Step 3: conclusion. By Theorem 6.5 and Step 2,

$$W_1(\text{Law}_{\mu_v}(Q_w), \mathcal{N}(0, \sigma_w^2)) \leq C(\eta_v \sigma_w + \|w^\perp\|_\infty) \leq C_B(\eta_v + \bar{e}).$$

Coupling $\mathcal{N}(0, \sigma_w^2)$ and $\mathcal{N}(0, I)$ as $\sigma_w G$ and $\sqrt{I}G$, and using (6.22), we have $W_1(\mathcal{N}(0, \sigma_w^2), \mathcal{N}(0, I)) \leq \mathbb{E}|G| |\sigma_w - \sqrt{I}| \leq C_B(\eta_v + \bar{e})$. Since $\ell = k_0 + Q_w$ and a deterministic shift changes W_1 by at most its size, (6.21) implies that

$$W_1(\text{Law}_{\mu_v}(\ell), \mathcal{N}(-I/2, I)) \leq |k_0 + I/2| + W_1(\text{Law}_{\mu_v}(Q_w), \mathcal{N}(0, I)) \leq C_B(\eta_v + \bar{e}),$$

which is (6.18). Finally, by Lemma 6.1 with $P = \mu_v$ and $Q = \mu_{\tilde{v}}$, we have $\|\mu_{\tilde{v}} - \mu_v\|_{\text{TV}} = \mathbb{E}_{\mu_v}(1 - e^\ell)_+$, and the last statement of Lemma 6.1, together with (6.3), proves (6.19). $\square$

Remark 6.7 (Disorder-dependent sequences). Theorem 6.6 is deterministic. In the applications, v and $\tilde{v}$ depend on the disorder and on time, and the hypotheses are verified on events of high probability. The bound B is then fixed before $N \rightarrow \infty$: if $\eta_v + \bar{e} \rightarrow 0$ in probability and the information is tight, then for each $\delta > 0$ we choose a deterministic B with $\limsup_N \mathbb{P}(I > B) \leq \delta$, apply (6.19) on the event $\{I \leq B, \bar{e} \leq 1/4, \eta_v \leq \eta_{\text{lik}}\}$, and then let $\delta \rightarrow 0$. We obtain $\|\mu_{\tilde{v}} - \mu_v\|_{\text{TV}} - (2\Phi(\sqrt{I}/2) - 1) \rightarrow 0$ in probability. The statement concerns a fixed pair of laws; uniformity over a family of pairs indexed by time requires the hypotheses to hold uniformly over the family.

Theorem 6.8 (Gaussian limits of likelihood ratios on a common sphere). *For each $N \geq 1$, let $v = v^{(N)}$ and $\tilde{v} = \tilde{v}^{(N)}$ be finite variance sequences with*

$$\sum_i v_i = \sum_i \tilde{v}_i, \quad \tilde{v}_i = v_i(1 + e_i), \quad \bar{e}_N = \max_i |e_i| \rightarrow 0, \quad \eta_v \rightarrow 0, \quad (6.23)$$

as $N \rightarrow \infty$, where the indices i and the coefficients e_i depend on N . Let $I_N = \frac{1}{2} \sum_i e_i^2$ and $\ell_N = \log(d\mu_{\tilde{v}}/d\mu_v)$. If $I_N \rightarrow I \in [0, \infty)$, then under μ_v ,

$$\ell_N \Rightarrow \mathcal{N}(-I/2, I), \quad \mathbb{E}_{\mu_v} e^{t\ell_N} \rightarrow e^{(t^2 - t)I/2} \quad \text{for each fixed } t \in \mathbb{R}, \quad (6.24)$$

and

$$\|\mu_{\tilde{v}} - \mu_v\|_{\text{TV}} \rightarrow 2\Phi(\sqrt{I}/2) - 1. \quad (6.25)$$

If $I_N \rightarrow \infty$, then $\|\mu_{\tilde{v}} - \mu_v\|_{\text{TV}} \rightarrow 1$.

The convergence of exponential moments is for each fixed t ; in its proof, the tilted variances are positive only for N large depending on t . In the case $I_N \rightarrow \infty$, no relation between η_v and I_N is assumed.

Proof. We prove the three assertions in turn.

Weak convergence and total variation. Let $I_N \rightarrow I < \infty$ and $B = I + 1$. For N large, $\bar{e}_N \leq 1/4$, $\eta_v \leq \eta_{\text{lik}}$, and $I_N \leq B$, and by Theorem 6.6 we have $W_1(\text{Law}_{\mu_v}(\ell_N), \mathcal{N}(-I_N/2, I_N)) \leq C_B(\eta_v + \bar{e}_N) \rightarrow 0$. Coupling with one standard Gaussian variable, we have $W_1(\mathcal{N}(-I_N/2, I_N), \mathcal{N}(-I/2, I)) \leq |I_N - I|/2 + \mathbb{E}|G| |\sqrt{I_N} - \sqrt{I}| \rightarrow 0$. Convergence in W_1 implies weak convergence, which proves the first part of (6.24), including the case $I = 0$. By (6.19) and continuity of $I \mapsto 2\Phi(\sqrt{I}/2) - 1$, we obtain (6.25).

Exponential moments. Fix $t \in \mathbb{R}$. Recall that $|\tilde{e}_i| \leq \frac{4}{3}\bar{e}_N \rightarrow 0$; we have $|t\tilde{e}_i| \leq 1/2$ and $|(1-t)e_i| \leq 1/2$ for all i when N is large. We assume this from now on. By (6.17),

$$\mathbb{E}_{\mu_v} e^{t\ell_N} = \left(\frac{p_v(0)}{p_{\tilde{v}}(0)} \right)^t \prod_i (1 + e_i)^{-t/2} \mathbb{E}_{\mu_v} \exp \left\{ \frac{t}{2} \sum_i \tilde{e}_i Z_i^2 \right\}.$$

We apply Lemma 5.13 with coefficients $\tilde{e}_i$. The tilted variances are

$$v_i(t) = \frac{v_i}{1 - t\tilde{e}_i} = \frac{v_i(1 + e_i)}{1 + (1 - t)e_i}, \quad K_N(t) = \prod_i (1 + e_i)^{-t/2} (1 - t\tilde{e}_i)^{-1/2}, \quad (6.26)$$

and we obtain

$$\mathbb{E}_{\mu_v} e^{t\ell_N} = \left(\frac{p_v(0)}{p_{\tilde{v}}(0)} \right)^t K_N(t) \frac{p_{v(t)}(-\Delta_N(t))}{p_v(0)}, \quad \Delta_N(t) = \sum_i (v_i(t) - v_i).$$

We treat the three factors separately. The first tends to one, by (6.20). For the third, using the identity $\sum_i v_i e_i = 0$, we find, exactly,

$$\Delta_N(t) = \sum_i \frac{t v_i e_i}{1 + (1-t)e_i} = -t(1-t) \sum_i \frac{v_i e_i^2}{1 + (1-t)e_i}, \quad |\Delta_N(t)| \leq 2|t(1-t)| \eta_v S_v \sum_i e_i^2. \quad (6.27)$$

The ratios $v_i(t)/v_i$ tend to one uniformly in i , so that $S_{v(t)}/S_v \rightarrow 1$ and $\eta_{v(t)} \leq \eta_v(1 + o(1)) \rightarrow 0$. As I_N is bounded, we deduce from (6.27) that $\Delta_N(t)/S_{v(t)} \rightarrow 0$. Using Theorem 5.3 with $j = 0$, we find that $f_{v(t)} \rightarrow \phi$ and $f_v \rightarrow \phi$ uniformly. Since ϕ is continuous, it follows that

$$\frac{p_{v(t)}(-\Delta_N(t))}{p_v(0)} = \frac{S_v}{S_{v(t)}} \frac{f_{v(t)}(-\Delta_N(t)/S_{v(t)})}{f_v(0)} \rightarrow 1. \quad (6.28)$$

For the second factor, $\log K_N(t) = \sum_i k_t(e_i)$ with

$$k_t(e) = \frac{1-t}{2} \log(1+e) - \frac{1}{2} \log(1+(1-t)e),$$

which follows from $1 - t\tilde{e}_i = (1 + (1-t)e_i)/(1 + e_i)$. We have $k_t(0) = k'_t(0) = 0$ and $k''_t(0) = (t^2 - t)/2$, and Taylor's formula gives $|k_t(e) - (t^2 - t)e^2/4| \leq C_t|e|^3$ for $|e|$ small, depending on t . We obtain

$$\log K_N(t) = \frac{t^2 - t}{4} \sum_i e_i^2 + O_t\left(\tilde{e}_N \sum_i e_i^2\right) \rightarrow \frac{(t^2 - t)I}{2}.$$

Combining the three factors, we obtain the second part of (6.24).

Diverging information. Let $I_N \rightarrow \infty$, and consider the single quadratic statistic

$$Q_N(y) = \sum_i e_i \left(\frac{y_i^2}{v_i} - 1 \right).$$

For N large, η_v and $\eta_{\tilde{v}} \leq \frac{5}{3}\eta_v$ are at most the constant η_0 of Theorem 5.15. Under μ_v , by Theorem 5.15 with $p = 2$, we have $\mathbb{E}_{\mu_v} Q_N^2 \leq C \sum_i e_i^2 = 2CI_N$. Under $\mu_{\tilde{v}}$, with $\tilde{Z}_i = Y_i/\sqrt{\tilde{v}_i}$, we have $y_i^2/v_i = (1 + e_i)\tilde{Z}_i^2$, and hence the exact identity

$$Q_N = 2I_N + \sum_i e_i(1 + e_i)(\tilde{Z}_i^2 - 1).$$

By Theorem 5.15 under $\mu_{\tilde{v}}$, we have $\mathbb{E}_{\mu_{\tilde{v}}}(Q_N - 2I_N)^2 \leq C \sum_i e_i^2(1 + e_i)^2 \leq CI_N$. By Chebyshev's inequality,

$$\mu_v(Q_N > I_N) \leq \frac{C}{I_N}, \quad \mu_{\tilde{v}}(Q_N \leq I_N) \leq \mu_{\tilde{v}}(|Q_N - 2I_N| \geq I_N) \leq \frac{C}{I_N}.$$

We conclude that $\|\mu_{\tilde{v}} - \mu_v\|_{TV} \geq \mu_{\tilde{v}}(Q_N > I_N) - \mu_v(Q_N > I_N) \geq 1 - 2C/I_N \rightarrow 1$, as desired. $\square$

Deterministic covariance evolution and heat estimates

This section collects deterministic facts about the exponential of a diagonal matrix perturbed by a matrix of rank one. For positive precisions $p_1, \dots, p_N$ with $\sum_i p_i^{-1} = N$, we study the heat matrix $H = -2 \operatorname{diag}(p) + (2/N) \mathbf{1}\mathbf{1}^\top$, the heat vector $h(t) = e^{tH} \mathbf{1}$, and its mean $\bar{h}(t)$. These objects describe the evolution of Gaussian variances that we use at high temperature to approximate the time- t law, and they also describe the second moments of the unnormalized process of Section 10. In Section 7.1, we prove the algebraic facts once (Lemma 7.1): positivity of e^{tH} , the kernel of H and the weight of its zero eigenvalue, a representation of $\bar{h}$ by a positive spectral measure, the coordinate Duhamel formula, and the resolvent of H . We also prove an exact formula for the information in terms of $\bar{h}$ (Lemma 7.3) and rescale the spectral measure at the scale of the smallest precision (Lemma 7.4). In Section 7.2, we define the covariance evolution and the Bingham comparison law μ_t of a symmetric matrix A , and prove their basic properties (Lemmas 7.6 and 7.7). The next two subsections prove estimates under two different hypotheses on the precisions. In Section 7.3, the smallest precision d is at most one and the other precisions exceed d by an amount of order $((i-1)/N)^{2/3}$, which is the situation at high temperature (Lemmas 7.10 to 7.13). In Section 7.4, the half-precisions $r_i = p_i/2$ are of order $(i/N)^{2/3}$ for all i , including $i = 1$, which is the situation in the critical window (Lemma 7.14). Finally, in Section 7.5, we show that convergence of the real squared resolvents of positive measures implies their convergence against a large class of test functions (Lemma 7.15).

We use only the definitions of Section 3, namely the operators (3.3) and the Lagrange multiplier of Definition 3.3, and the identification of Bingham laws with conditioned Gaussian laws in Proposition 5.6. All statements are deterministic, except for the random version of Lemma 7.15. The covariance evolution and the Bingham comparison law are used in Section 9 and in Part III. The algebra of Lemma 7.1 is used in Sections 10 and 14, and the estimates of Section 7.3 are used in Sections 10, 14, and 20. Lemma 7.14 is used in Sections 8, 10, 12, 18, and 19, and Lemmas 7.3 and 7.15 are used in Sections 14 and 20.

Notation. Throughout this section, $N \geq 2$ and $p = (p_1, \dots, p_N)$ is a sequence of positive numbers such that

$$\sum_{i=1}^N \frac{1}{p_i} = N. \quad (7.1)$$

Recall from Definition 3.3 that the precisions $p_i = z - \alpha_i$ of a real symmetric matrix A with eigenvalues $\alpha_1 \geq \dots \geq \alpha_N$ satisfy (7.1). Conversely, every sequence p satisfying (7.1) arises in this way up to a permutation of the eigencoordinates, since the Lagrange multiplier of $A = -\text{diag}(p)$ is $z = 0$. We write

$$q_i = \frac{1}{p_i}, \quad d = \min_i p_i, \quad s = \max_i p_i, \quad \vartheta = \frac{N}{\sum_{i=1}^N q_i^2}. \quad (7.2)$$

For the matrix A , we have $\vartheta = N / \text{Tr } Q^2$ with $Q = (zI - A)^{-1}$. We call ϑ the weight of the zero eigenvalue; Lemma 7.1 below explains the name. We distinguish this weight from the low-temperature top fraction θ_N . Let $\mathbf{1} = (1, \dots, 1)^\top$. We define the heat matrix, the heat vector, and its mean by

$$H = -2 \text{diag}(p) + \frac{2}{N} \mathbf{1} \mathbf{1}^\top, \quad h(t) = e^{tH} \mathbf{1}, \quad \bar{h}(t) = \frac{1}{N} \sum_{i=1}^N h_i(t), \quad t \geq 0. \quad (7.3)$$

The heat vector $h(t)$ has positive coordinates (Lemma 7.1 below). We define the variances and the relative errors, and recall the information $\mathcal{I}_N(t) = \mathcal{I}(q, v(t))$ of $v(t)$ relative to q from Definition 3.5:

$$v_i(t) = \frac{h_i(t)}{\bar{h}(t)}, \quad e_i(t) = \frac{v_i(t)}{q_i} - 1 = p_i v_i(t) - 1, \quad \mathcal{I}_N(t) = \frac{1}{2} \sum_{i=1}^N e_i(t)^2. \quad (7.4)$$

For a matrix A , we show in Section 7.2 that $v(t)$ is the covariance evolution of A ; in particular, (7.4) agrees with (1.9) and with Definition 3.5. Finally, for complex ζ with $\zeta \neq -2p_i$ for all i , let

$$G(\zeta) = \frac{1}{N} \sum_{i=1}^N \frac{1}{p_i + \zeta/2}, \quad S(\zeta) = \frac{1}{N} \sum_{i=1}^N \frac{1}{p_i(p_i + \zeta/2)}, \quad (7.5)$$

and for a locally integrable function f on $[0, \infty)$ of at most exponential growth, let $\hat{f}(\zeta) = \int_0^\infty e^{-\zeta t} f(t) dt$ be its Laplace transform, for real ζ large enough.

7.1 The heat matrix

The following lemma contains the algebra used in all later estimates. A real square matrix whose off-diagonal entries are nonnegative is called a Metzler matrix, after [111]. The positivity of its exponential, which we use below, follows from the series argument in the proof of the first part; for background on positive linear systems, see [70, Chapter 2, pp. 7–15].

Lemma 7.1 (The heat matrix). *Let p satisfy (7.1).*

1. The matrix H is symmetric, and its off-diagonal entries are positive. For every $t \geq 0$, the matrix e^{tH} has nonnegative entries, and $h_i(t) \geq e^{-2p_i t} > 0$ for all i .
2. We have $d \leq 1 \leq s$ and $d \leq \vartheta \leq 1$. The matrix H is negative semidefinite, and its kernel is spanned by the vector $q = (q_1, \dots, q_N)^\top$. Every eigenvalue of the restriction of $-H$ to the orthogonal complement of q lies in $[2d, 2s]$.
3. There exists a positive measure ρ on $[2d, 2s]$ with total mass $1 - \vartheta$ such that

$$\bar{h}(t) = \vartheta + \int e^{-\lambda t} \rho(d\lambda), \quad t \geq 0. \quad (7.6)$$

In particular, $\bar{h}$ is completely monotone, $\bar{h}(0) = 1$, $\vartheta \leq \bar{h}(t) \leq 1$, and $\bar{h}(t)$ decreases to ϑ as $t \rightarrow \infty$. For every integer $j \geq 1$ and all $0 \leq t_0 \leq t$ with $t > 0$,

$$\begin{aligned} 0 \leq (-1)^j \bar{h}^{(j)}(t) &\leq (2s)^j (\bar{h}(t) - \vartheta), & (-1)^j \bar{h}^{(j)}(t) &\leq \left(\frac{2j}{et}\right)^j (\bar{h}(t/2) - \vartheta), \\ (-1)^j \bar{h}^{(j)}(t) &\leq e^{-2d(t-t_0)} (-1)^j \bar{h}^{(j)}(t_0). \end{aligned} \quad (7.7)$$

4. For all $t \geq 0$ and $1 \leq i, j \leq N$,

$$(e^{tH})_{ij} = e^{-2p_i t} \delta_{ij} + \frac{2}{N} \int_0^t e^{-2p_i(t-u)} h_j(u) du, \quad (7.8)$$

and in particular

$$h_i(t) = e^{-2p_i t} + 2 \int_0^t e^{-2p_i(t-u)} \bar{h}(u) du. \quad (7.9)$$

With $k(t) = N^{-1} \sum_i e^{-2p_i t}$, we have $\bar{h}(t) = k(t) + 2 \int_0^t k(t-u) \bar{h}(u) du$.

5. For all $t \geq 0$, we have $\sum_i q_i h_i(t) = N$.
6. For every complex $\zeta \notin \{0\} \cup [-2s, -2d]$, the matrix $\zeta I - H$ is invertible, $G(\zeta) \neq 1$, and

$$\frac{1}{N} \mathbf{1}^\top (\zeta I - H)^{-1} \mathbf{1} = \frac{\vartheta}{\zeta} + \int \frac{\rho(d\lambda)}{\zeta + \lambda} = \frac{G(\zeta)}{2(1 - G(\zeta))}, \quad 1 - G(\zeta) = \frac{\zeta}{2} S(\zeta). \quad (7.10)$$

For real $\zeta > 0$, these quantities are equal to $\widehat{\bar{h}}(\zeta)$.

7. For all $t \geq 0$ and all i , we have $\sum_i v_i(t) = N$, $\sum_i q_i e_i(t) = 0$, and

$$h'_i(t) = -2\bar{h}(t) e_i(t), \quad \bar{h}'(t) = -\frac{2\bar{h}(t)}{N} \sum_{j=1}^N e_j(t). \quad (7.11)$$

Proof. We prove the parts in order, except that the lower bound on $h_i(t)$ in the first part uses (7.9), whose proof does not use it.

Part 1. The off-diagonal entries of H are equal to $2/N$. Since the diagonal entries of $H + 2sI$ are $2(s - p_i) + 2/N \geq 0$, the matrix $H + 2sI$ and all its powers have nonnegative entries. The expansion

$$e^{tH} = e^{-2st} \sum_{n=0}^{\infty} \frac{t^n (H + 2sI)^n}{n!}$$

shows that e^{tH} has nonnegative entries for $t \geq 0$. In particular, $h(t)$ has nonnegative coordinates and $\bar{h}(t) \geq 0$. The integral in (7.9) is nonnegative, and $h_i(t) \geq e^{-2p_i t}$.

Part 2. The numbers q_i lie in $[1/s, 1/d]$, and their average is one by (7.1). Hence $d \leq 1 \leq s$. By the Cauchy–Schwarz inequality, we have $N^2 = (\sum_i q_i)^2 \leq N \sum_i q_i^2$, that is, $\vartheta \leq 1$. Also $\sum_i q_i^2 \leq d^{-1} \sum_i q_i = N/d$, that is, $\vartheta \geq d$. For $y \in \mathbb{R}^N$, we have

$$-y^T H y = 2 \sum_i p_i y_i^2 - \frac{2}{N} \left(\sum_i y_i \right)^2. \quad (7.12)$$

Writing $y_i = \sqrt{q_i} \cdot \sqrt{p_i} y_i$ and using the Cauchy–Schwarz inequality and (7.1), we find $(\sum_i y_i)^2 \leq \sum_i q_i \sum_i p_i y_i^2 = N \sum_i p_i y_i^2$, with equality if and only if $(\sqrt{p_i} y_i)_i$ is proportional to $(\sqrt{q_i})_i$, that is, if and only if y is proportional to q . We conclude that $-H$ is positive semidefinite and that $y^T H y = 0$ only for $y \in \text{span}(q)$. Every vector y in the kernel satisfies $y^T H y = 0$, and $(Hq)_i = -2 + 2N^{-1} \sum_j q_j = 0$. Therefore the kernel is spanned by q . Since H is symmetric and $Hq = 0$, the orthogonal complement $q^\perp$ is invariant under H . Let $y \in q^\perp$. Using $p_i q_i = 1$ and $\sum_i q_i y_i = 0$, we find $\sum_i y_i = \sum_i (p_i - d) q_i y_i$, where the numbers $p_i - d$ are nonnegative. Then the Cauchy–Schwarz inequality implies that

$$\left(\sum_i y_i \right)^2 \leq \sum_i (p_i - d) q_i^2 \sum_i (p_i - d) y_i^2 \leq N \sum_i (p_i - d) y_i^2,$$

where in the last step we used $\sum_i (p_i - d) q_i^2 = N - dN/\vartheta \leq N$. Inserting this bound into (7.12) gives $-y^T H y \geq 2d|y|^2$. Also $-y^T H y \leq 2 \sum_i p_i y_i^2 \leq 2s|y|^2$. This shows that the eigenvalues of $-H$ on $q^\perp$ lie in $[2d, 2s]$.

Part 3. By the second part, there exists an orthonormal basis $u_0, u_1, \dots, u_{N-1}$ of $\mathbb{R}^N$ with $u_0 = q/|q|$ and $Hu_k = -\lambda_k u_k$, where $\lambda_0 = 0$ and $\lambda_k \in [2d, 2s]$ for $k \geq 1$. Then

$$\bar{h}(t) = \frac{1}{N} \mathbf{1}^T e^{tH} \mathbf{1} = \sum_{k=0}^{N-1} e^{-\lambda_k t} \frac{(\mathbf{1}^T u_k)^2}{N}.$$

By (7.1), the term $k = 0$ is $(\sum_i q_i)^2 / (N \sum_i q_i^2) = \vartheta$. Let $\rho = \sum_{k \geq 1} N^{-1} (\mathbf{1}^T u_k)^2 \delta_{\lambda_k}$. Its mass is $N^{-1} |\mathbf{1}|^2 - \vartheta = 1 - \vartheta$ by Parseval's identity, and (7.6) holds. Differentiating the finite sum, we have $(-1)^j \bar{h}^{(j)}(t) = \int \lambda^j e^{-\lambda t} \rho(d\lambda) \geq 0$. In particular, $\bar{h}$ is completely monotone and decreases from $\bar{h}(0) = 1$ to its limit ϑ . For $\lambda \in [2d, 2s]$, we have $\lambda^j e^{-\lambda t} \leq (2s)^j e^{-\lambda t}$ and $\lambda^j e^{-\lambda t} \leq e^{-2d(t-t_0)} \lambda^j e^{-\lambda t_0}$. Since $\sup_{x>0} x^j e^{-xt/2} = (2j/(et))^j$, we also have $\lambda^j e^{-\lambda t} \leq (2j/(et))^j e^{-\lambda t/2}$. Integrating these three inequalities against ρ and recalling that $\int e^{-\lambda t} \rho(d\lambda) = \bar{h}(t) - \vartheta$, we obtain (7.7).

Part 4. Since e^{tH} is symmetric, we have $\sum_k (e^{tH})_{kj} = (e^{tH}\mathbf{1})_j = h_j(t)$. The equation $\frac{d}{dt}e^{tH} = He^{tH}$ reads, entrywise,

$$\frac{d}{dt}(e^{tH})_{ij} = -2p_i(e^{tH})_{ij} + \frac{2}{N} \sum_k (e^{tH})_{kj} = -2p_i(e^{tH})_{ij} + \frac{2}{N}h_j(t),$$

and the variation of constants formula with $(e^{0H})_{ij} = \delta_{ij}$ yields (7.8). Summing (7.8) over j gives (7.9), and averaging (7.9) over i gives the equation for $\bar{h}$.

Part 5. Since $Hq = 0$, we have $e^{tH}q = q$. By the symmetry of e^{tH} , we find $\sum_i q_i h_i(t) = q^\top e^{tH}\mathbf{1} = (e^{tH}q)^\top \mathbf{1} = \sum_i q_i = N$.

Part 6. The eigenvalues of $\zeta I - H$ are ζ and $\zeta + \lambda_k$, $k \geq 1$, which do not vanish for $\zeta \notin \{0\} \cup [-2s, -2d]$. Expanding $\mathbf{1}$ in the basis of the third part gives the first identity in (7.10). Let $y = (\zeta I - H)^{-1}\mathbf{1}$ and $\bar{y} = N^{-1} \sum_i y_i$. The i th coordinate of $(\zeta I - H)y = \mathbf{1}$ reads $(\zeta + 2p_i)y_i - 2\bar{y} = 1$. Since $\zeta + 2p_i \neq 0$, we obtain $y_i = (1 + 2\bar{y})/(\zeta + 2p_i)$. Averaging over i , we find $\bar{y} = (1 + 2\bar{y})G(\zeta)/2$, that is,

$$\bar{y}(1 - G(\zeta)) = G(\zeta)/2.$$

If $G(\zeta) = 1$, this identity would read $0 = 1/2$. Therefore $G(\zeta) \neq 1$ and $\bar{y} = G(\zeta)/(2(1 - G(\zeta)))$. By (7.1), $1 - G(\zeta) = N^{-1} \sum_i (q_i - (p_i + \zeta/2)^{-1}) = (\zeta/2)S(\zeta)$. For real $\zeta > 0$, we have $\hat{h}(\zeta) = \vartheta/\zeta + \int \rho(d\lambda)/(\zeta + \lambda)$ by (7.6) and Fubini's theorem.

Part 7. We have $\sum_i v_i = \sum_i h_i/\bar{h} = N$ by (7.3), and $\sum_i q_i e_i = \sum_i v_i - \sum_i q_i = 0$ by (7.1). The equation $h' = Hh$ reads $h'_i = -2p_i h_i + 2\bar{h} = -2\bar{h}(p_i v_i - 1) = -2\bar{h}e_i$. Averaging over i gives the formula for $\bar{h}'$. $\square$

The positivity in the first part holds more generally for the solutions of the linear system $y' = My$ for every real matrix M with nonnegative off-diagonal entries: if $y(0)$ has nonnegative coordinates, then $y(t) = e^{tM}y(0)$ has nonnegative coordinates for all $t \geq 0$. The proof is the same, with $H + 2sI$ replaced by $M + \tau I$ for τ large. We use this in the proof of Lemma 7.6.

Remark 7.2 (Precisions and half-precisions). The critical sections use the half-precisions $r_i = p_i/2$. In terms of these, the trace condition (7.1) and the heat matrix read

$$\sum_{i=1}^N \frac{1}{2r_i} = N, \quad H = -4 \operatorname{diag}(r) + \frac{2}{N} \mathbf{1}\mathbf{1}^\top, \quad h(t) = e^{tH}\mathbf{1}, \quad \sum_{j=1}^N h_j(t) = N\bar{h}(t), \quad (7.13)$$

and the matrix $-4 \operatorname{diag}(r) + (2/N)\mathbf{1}\mathbf{1}^\top$ of the critical sections is the same matrix H . With $k(t) = N^{-1} \sum_i e^{-4r_i t}$, the fourth and sixth parts of Lemma 7.1 read $N\bar{h} = Nk + 2k * (N\bar{h})$ and, for $\zeta > 0$,

$$\widehat{N\bar{h}}(\zeta) = \frac{N\widehat{k}(\zeta)}{1 - 2\widehat{k}(\zeta)}, \quad \widehat{k}(\zeta) = \frac{1}{N} \sum_{i=1}^N \frac{1}{\zeta + 4r_i}, \quad 1 - 2\widehat{k}(\zeta) = \frac{\zeta}{2N} \sum_{i=1}^N \frac{1}{r_i(\zeta + 4r_i)}. \quad (7.14)$$

Indeed $\widehat{2k}(\zeta) = G(\zeta)$ and $S(\zeta) = N^{-1} \sum_i [r_i(\zeta + 4r_i)]^{-1}$. Table 7.1 lists the two conventions side by side.

	precisions p_i	half-precisions $r_i = p_i/2$
trace condition	$\sum_i p_i^{-1} = N$	$\sum_i (2r_i)^{-1} = N$
heat matrix H	$-2 \operatorname{diag}(p) + (2/N) \mathbf{1}\mathbf{1}^\top$	$-4 \operatorname{diag}(r) + (2/N) \mathbf{1}\mathbf{1}^\top$
kernel of H	spanned by $(p_i^{-1})_i$	spanned by $(r_i^{-1})_i$
decay in (7.8)	$e^{-2p_i t}$	$e^{-4r_i t}$
limit of $\sum_j h_j(t)$	$N\vartheta = N^2 / \sum_i p_i^{-2}$	$4N^2 / \sum_i r_i^{-2}$

Table 7.1. The heat matrix in terms of precisions and of half-precisions.

At low temperature the potential is written with the scaled eigenvalue gaps $a_i = c(\lambda_1 - \lambda_i)/2$ of (3.8), and $a_i = r_i - d/2$. More generally, for a real number τ , the matrix built from the coefficients $r_i - \tau/2$ is $-4 \operatorname{diag}(r_i - \tau/2) + (2/N) \mathbf{1}\mathbf{1}^\top = H + 2\tau I$, and its exponential is $e^{2\tau t} e^{tH}$. For $\tau = d$, the zero eigenvalue of H becomes the eigenvalue $2d$, and the exponential grows at rate $2d$ instead of converging.

The next lemma expresses the information through the function $\bar{h}$ alone. This is how the information is computed in Section 14.

Lemma 7.3 (An exact formula for the information). *Let p satisfy (7.1). For every $t \geq 0$,*

$$\mathcal{I}_N(t) = \frac{1}{2} \sum_{i=1}^N e_i(t)^2 = \frac{N \bar{h}''(2t)}{8 \bar{h}(t)^2}. \quad (7.15)$$

Moreover, there exists a finite positive measure ρ supported on $[2d, \infty)$ such that

$$\bar{h}(t) = \vartheta + \int e^{-\lambda t} \rho(d\lambda), \quad \rho([2d, \infty)) = 1 - \vartheta. \quad (7.16)$$

Proof. The measure ρ of Lemma 7.1 has the required properties. Since H is symmetric and commutes with e^{tH} ,

$$\bar{h}''(2t) = \frac{1}{N} \mathbf{1}^\top H^2 e^{2tH} \mathbf{1} = \frac{1}{N} (H e^{tH} \mathbf{1})^\top (H e^{tH} \mathbf{1}) = \frac{1}{N} |Hh(t)|^2.$$

By (7.11), $(Hh(t))_i = h'_i(t) = -2\bar{h}(t)e_i(t)$. Inserting this into the last display gives $\bar{h}''(2t) = 4\bar{h}(t)^2 N^{-1} \sum_i e_i(t)^2 = 8\bar{h}(t)^2 \mathcal{I}_N(t)/N$, which is (7.15). $\square$

At high temperature the relevant times are of order d^{-1} , and the transient part of $\bar{h}$ is best viewed at that scale. The total mass of the measure n_N below is $(1 - \vartheta)/\sqrt{d}$, which need not stay bounded as $N \rightarrow \infty$.

Lemma 7.4 (The transient spectral measure at the scale of the smallest precision). *Let p satisfy (7.1), and let ρ be the measure of (7.16). Let $\mathbf{n}_N$ be $d^{-1/2}$ times the image of ρ under the map $\lambda \mapsto \lambda/(2d) - 1$; it is a finite positive measure on $[0, s/d - 1]$. For $w \geq 0$, let*

$$f_N(w) = \frac{\bar{h}(w/d)}{\sqrt{d}}, \quad j_N(w) = f_N(w) - \frac{\vartheta}{\sqrt{d}} = e^{-2w} \int e^{-2wx} \mathbf{n}_N(dx). \quad (7.17)$$

Then $f_N(w)$ decreases to $\vartheta/\sqrt{d}$ as $w \rightarrow \infty$. For every complex $\zeta \notin \{0\} \cup [-2s/d, -2]$, and in particular for every real $\zeta \in (-2, \infty)$ with $\zeta \neq 0$,

$$\int \frac{\mathbf{n}_N(dx)}{\zeta + 2 + 2x} = \frac{\sqrt{d}}{2(1 - G(d\zeta))} - \frac{\sqrt{d}}{2} - \frac{\vartheta}{\sqrt{d}\zeta}. \quad (7.18)$$

For real $\zeta > 0$, the left side of (7.18) is $\widehat{j}_N(\zeta)$, and $\widehat{f}_N(\zeta) = \sqrt{d}/(2(1 - G(d\zeta))) - \sqrt{d}/2$.

Proof. The measure $\mathbf{n}_N$ is supported in $[0, s/d - 1]$ because ρ is supported in $[2d, 2s]$. By definition of $\mathbf{n}_N$, we have $\int g(\lambda/(2d) - 1) \rho(d\lambda) = \sqrt{d} \int g(x) \mathbf{n}_N(dx)$ for every bounded measurable g . With $\lambda = 2d(1 + x)$, we find from (7.16) that

$$\bar{h}(w/d) - \vartheta = \int e^{-\lambda w/d} \rho(d\lambda) = \sqrt{d} \int e^{-2w(1+x)} \mathbf{n}_N(dx),$$

which is (7.17). By Lemma 7.1, f_N decreases to $\vartheta/\sqrt{d}$. Let $\zeta \notin \{0\} \cup [-2s/d, -2]$. Then $d\zeta \notin \{0\} \cup [-2s, -2d]$, and

$$\int \frac{\rho(d\lambda)}{d\zeta + \lambda} = \frac{1}{d} \int \frac{\rho(d\lambda)}{\zeta + \lambda/d} = \frac{1}{\sqrt{d}} \int \frac{\mathbf{n}_N(dx)}{\zeta + 2 + 2x}.$$

Inserting this into (7.10) at $d\zeta$, multiplying by $\sqrt{d}$, and writing $G/(2(1 - G)) = 1/(2(1 - G)) - 1/2$, we obtain (7.18). For real $\zeta > 0$, the substitution $t = w/d$ yields $\widehat{f}_N(\zeta) = \sqrt{d} \widehat{h}(d\zeta)$, and the formula for $\widehat{f}_N$ follows from the sixth part of Lemma 7.1. Finally, $\widehat{j}_N(\zeta) = \int \mathbf{n}_N(dx)/(\zeta + 2 + 2x)$ by Fubini's theorem. $\square$

7.2 The covariance evolution and the Bingham comparison law

Let A be a real symmetric matrix with eigenvalues $\alpha_1 \geq \dots \geq \alpha_N$. We work in its eigencoordinates, in which $A = \text{diag}(\alpha_1, \dots, \alpha_N)$, and we let z be its Lagrange multiplier and $p_i = z - \alpha_i$ its precisions (Definition 3.3). Then p satisfies (7.1), and $d = p_1 \leq p_2 \leq \dots \leq p_N = s$. All notation of this section refers to this sequence p .

Definition 7.5 (Covariance evolution and Bingham comparison law). Let $m(t) \in \mathbb{R}^N$ be the solution of the linear system

$$m'_i(t) = 2\alpha_i m_i(t) + \frac{2}{N} \sum_{j=1}^N m_j(t), \quad m_i(0) = 1, \quad \text{and let } v_i(t) = \frac{Nm_i(t)}{\sum_j m_j(t)}. \quad (7.19)$$

We define the moving multiplier $z(t)$, the coefficients $b_i(t)$, the matrix B_t , and the Bingham comparison law μ_t by

$$\begin{aligned} z(t) &= 1 + \frac{1}{N} \sum_{i=1}^N \alpha_i v_i(t), & b_i(t) &= z(t) - \frac{1}{v_i(t)}, & B_t &= \text{diag}(b_i(t)), \\ \mu_t(dx) &= \frac{e^{x^\top B_t x/2} \nu_N(dx)}{\int_{S_N} e^{y^\top B_t y/2} \nu_N(dy)}. \end{aligned} \quad (7.20)$$

We call $v(t)$ the covariance evolution of A .

The matrix of (7.19) is $2 \text{diag}(\alpha) + (2/N)\mathbf{1}\mathbf{1}^\top = H + 2zI$, so that $m(t) = e^{2zt}h(t)$. By Lemma 7.1, $m(t)$ has positive coordinates, and $v_i(t) = h_i(t)/\bar{h}(t)$ agrees with (7.4). In particular, $e_i(t) = v_i(t)/q_i - 1$ and $\mathcal{I}_N(t)$ are the relative errors and the information of the covariance evolution. On S_N we have $x^\top B_t x/2 = Nz(t)/2 - \sum_i x_i^2/(2v_i(t))$. Since $\sum_i v_i(t) = N$, the second part of Proposition 5.6 implies that $\mu_t = \mu_{v(t)}$ is the Gaussian law with variances $v_i(t)$ conditioned on S_N .

The covariance evolution has the following interpretation. Consider the linear system $dx_i = (\alpha_i - z(t))x_i dt + \sqrt{2} d\mathcal{B}_i(t)$, driven by independent standard Brownian motions $\mathcal{B}_i$ and started from independent standard Gaussian values, where $z(t)$ is chosen so that $\mathbb{E}|x(t)|^2 = N$ for all t . The variances $\mathbb{E}x_i(t)^2$ then solve $v'_i = 2 + 2(\alpha_i - z(t))v_i$, and the constraint forces $z(t) = 1 + N^{-1} \sum_i \alpha_i v_i(t)$; this is the system (7.21) below. In this system the spherical constraint holds only in expectation. For the spherical Sherrington–Kirkpatrick model, Cugliandolo and Dean solved this system in the limit $N \rightarrow \infty$ by diagonalizing the interaction matrix [51, (II.2)–(II.4)], and they reduced the multiplier to a scalar Volterra equation [51, (II.6)–(II.7)]; see also [156, Section II, (2)–(6)]. In our notation, the function $e^{2zt}\bar{h}(t) = N^{-1} \sum_j m_j(t)$ is equal to $\exp(2 \int_0^t z(u) du)$, by (7.11) and the identity $N^{-1} \sum_j e_j(t) = z - z(t)$, and averaging the variation of constants formula for (7.19) shows that it solves

$$e^{2zt}\bar{h}(t) = \frac{1}{N} \sum_i e^{2\alpha_i t} + 2 \int_0^t \left(\frac{1}{N} \sum_i e^{2\alpha_i(t-u)} \right) e^{2zu}\bar{h}(u) du.$$

This is the finite-dimensional form of the Volterra equation of Cugliandolo and Dean, at temperature one, with unit initial variances, and with the average over the spectrum taken with respect to the empirical spectral measure of A . What we add is the comparison of the Bingham comparison law μ_t with the time- t law of the diffusion on S_N , which is carried out in Section 9.

Lemma 7.6 (Positive covariance evolution and uniform variance bounds). *All $m_i(t)$ are positive, $\sum_i v_i(t) = N$, and, for all $t \geq 0$,*

$$v'_i = 2 + 2(\alpha_i - z(t))v_i, \quad b'_i = z'(t) + \frac{2(\alpha_i - b_i)}{v_i}, \quad \sum_{i=1}^N (\alpha_i - b_i(t))v_i(t) = 0. \quad (7.21)$$

With $q_i = (z - \alpha_i)^{-1}$, $d = z - \alpha_1$, and $s = z - \alpha_N$, we have, for all $t \geq 0$ and all i ,

$$\frac{1}{s} \leq v_i(t) \leq \frac{1}{d}, \quad |v_i(t) - q_i| \leq \frac{s-d}{d^2} e^{-2dt}. \quad (7.22)$$

If $\|A\| \leq M$, then, uniformly in $t \geq 0$,

$$\|B_t\| \leq 2 + 3M, \quad \|A - B_t\| \leq C_M d^{-1} e^{-2dt}, \quad |z(t) - z| \leq (1 + 2M) d^{-1} e^{-2dt}, \quad (7.23)$$

where $C_M = 4M(1 + M)$. Moreover, $\mu_0 = \nu_N$, and $\mu_t \rightarrow \pi_A$ in total variation as $t \rightarrow \infty$, for fixed N and A .

Proof. The proof has five steps: the identities (7.21), a transformed linear system and the decay of its oscillation, the range of v_i , the bounds on $z(t)$ and B_t , and the values at $t = 0$ and $t = \infty$.

Step 1: identities. We showed above that $m_i(t) > 0$. Let $\Lambda(t) = N^{-1} \sum_j m_j(t)$. By (7.19), $\Lambda' = N^{-1} \sum_j (2\alpha_j m_j + 2\Lambda) = 2\Lambda(1 + N^{-1} \sum_j \alpha_j v_j) = 2z(t)\Lambda$. Using $v_i = m_i/\Lambda$, we find

$$v_i' = \frac{m_i'}{\Lambda} - \frac{m_i \Lambda'}{\Lambda^2} = 2\alpha_i v_i + 2 - 2z(t)v_i,$$

which is the first identity in (7.21). Also, $\sum_i v_i = N$ by definition. Differentiating $b_i = z(t) - v_i^{-1}$ and using the first identity, we obtain

$$b_i' = z'(t) + \frac{v_i'}{v_i^2} = z'(t) + \frac{2}{v_i} \left(\frac{1}{v_i} + \alpha_i - z(t) \right) = z'(t) + \frac{2(\alpha_i - b_i)}{v_i}.$$

Finally, $\sum_i (\alpha_i - b_i)v_i = \sum_i \alpha_i v_i - z(t) \sum_i v_i + N = N(z(t) - 1) - Nz(t) + N = 0$.

Step 2: a transformed system. Let $w_i(t) = p_i h_i(t) = e^{-2zt} m_i(t)/q_i$. Recall that $h_i' = -2p_i h_i + 2\bar{h}$ and $\bar{h} = N^{-1} \sum_i h_i$. Multiplying by p_i , we find

$$w_i' = 2p_i(\bar{h} - w_i), \quad w_i(0) = p_i, \quad \bar{h} = \frac{1}{N} \sum_{i=1}^N q_i w_i, \quad v_i = \frac{q_i w_i}{\bar{h}}. \quad (7.24)$$

Note that $\bar{h}(t)$ is an average of the numbers $w_i(t)$ with positive weights q_i/N that sum to one. In particular, $\bar{h}(t)$ lies between $\underline{w}(t) = \min_i w_i(t)$ and $\overline{w}(t) = \max_i w_i(t)$. We claim that

$$\text{osc } w(t) = \overline{w}(t) - \underline{w}(t) \leq (s - d) e^{-2dt}, \quad t \geq 0. \quad (7.25)$$

Fix t . By continuity, for small $\varepsilon > 0$ the maximum of $w_j(t + \varepsilon)$ over j is attained at an index j that maximizes $w_j(t)$. It follows that the right derivative of $\overline{w}$ at t exists and equals $\max\{w_i'(t) : w_i(t) = \overline{w}(t)\}$. For such i , we have $w_i'(t) = 2p_i(\bar{h} - \overline{w}) \leq 2d(\bar{h} - \overline{w})$, since $\bar{h} - \overline{w} \leq 0$ and $p_i \geq d$. In the same way, the right derivative of $\underline{w}$ is at least $2d(\bar{h} - \underline{w})$.

Subtracting, we find that the right derivative of $\text{osc } w$ is at most $-2d \text{ osc } w$. A continuous function whose right derivative exists everywhere and is nonpositive is nonincreasing. We conclude that $e^{2dt} \text{osc } w(t)$ is nonincreasing, and (7.25) follows from $\text{osc } w(0) = s - d$. The same argument shows that $\bar{w}$ is nonincreasing and $\underline{w}$ is nondecreasing; in particular, $w_i(t) \in [d, s]$ for all i and t .

By Lemma 7.1, we have $\bar{h}(t) \geq \vartheta \geq d$. Using also $q_i \leq 1/d$, we obtain

$$|v_i - q_i| = \frac{q_i |w_i - \bar{h}|}{\bar{h}} \leq \frac{\text{osc } w}{d^2} \leq \frac{s - d}{d^2} e^{-2dt},$$

which is the second bound in (7.22).

Step 3: the range of v_i . The right side of (7.9) is a nonincreasing function of p_i , because $\bar{h} \geq 0$. Then $v_1(t) \geq v_2(t) \geq \dots \geq v_N(t)$, and it suffices to show $v_1 \leq 1/d$ and $v_N \geq 1/s$.

For the upper bound, let $k_i = w_i - w_1$ for $2 \leq i \leq N$. We have $\bar{h} - w_1 = N^{-1} \sum_j q_j (w_j - w_1) = N^{-1} \sum_{j \geq 2} q_j k_j$, and (7.24) yields

$$k'_i = 2p_i(\bar{h} - w_i) - 2d(\bar{h} - w_1) = \frac{2(p_i - d)}{N} \sum_{j \geq 2} q_j k_j - 2p_i k_i, \quad k_i(0) = p_i - d \geq 0.$$

This is a linear system whose matrix has the nonnegative off-diagonal entries $2(p_i - d)q_j/N$. As noted after Lemma 7.1, its solution stays nonnegative, since it starts from nonnegative values. Thus $w_1 = \underline{w} \leq \bar{h}$, and $v_1 = q_1 w_1 / \bar{h} \leq q_1 = 1/d$.

For the lower bound, (7.9) with $p_N = s$ implies that

$$w_N(t) = se^{-2st} + 2s \int_0^t e^{-2s(t-u)} \bar{h}(u) du \geq se^{-2st} + \bar{h}(t)(1 - e^{-2st}) \geq \bar{h}(t),$$

where we used that $\bar{h}$ is nonincreasing and $s \geq 1 \geq \bar{h}(t)$ (Lemma 7.1). We conclude that $v_N = q_N w_N / \bar{h} \geq q_N = 1/s$. This proves the first bound in (7.22).

Step 4: the multiplier and the quadratic coefficient matrix. Let $w_{\text{av}} = N^{-1} \sum_i w_i$. Since $N^{-1} \sum_i p_i v_i = w_{\text{av}} / \bar{h}$ and $N^{-1} \sum_i p_i v_i = z - N^{-1} \sum_i \alpha_i v_i = z - z(t) + 1$, we have

$$z(t) - z = \frac{\bar{h} - w_{\text{av}}}{\bar{h}}, \quad v_i^{-1} - p_i = \frac{\bar{h} - w_i}{\bar{h} v_i}, \quad (7.26)$$

where the second identity follows from $v_i = q_i w_i / \bar{h}$. Both $\bar{h}$ and w_{av} lie between $\underline{w}$ and $\bar{w}$. Since also $v_i \geq 1/s$, we conclude that

$$|z(t) - z| \leq \frac{\text{osc } w(t)}{\bar{h}(t)}, \quad |v_i^{-1} - p_i| \leq \frac{s \text{osc } w(t)}{\bar{h}(t)}, \quad \|A - B_t\| \leq \frac{(1+s) \text{osc } w(t)}{\bar{h}(t)}, \quad (7.27)$$

where for the last bound we used $\alpha_i - b_i = (v_i^{-1} - p_i) - (z(t) - z)$. Suppose now that $\|A\| \leq M$. Then $s - d = \alpha_1 - \alpha_N \leq 2M$ and, since $d \leq 1$, also $s \leq 1 + 2M$. By (7.25)

and $\bar{h} \geq d$, we have $|z(t) - z| \leq 2Md^{-1}e^{-2dt}$ and $\|A - B_t\| \leq (2 + 2M)2Md^{-1}e^{-2dt}$, which are the last two bounds in (7.23). Also $|z(t)| \leq 1 + MN^{-1}\sum_i v_i = 1 + M$ and $v_i^{-1} \leq s \leq 1 + 2M$, so that $\|B_t\| = \max_i |b_i(t)| \leq 2 + 3M$.

Step 5: the endpoints. At $t = 0$, $v_i(0) = 1$ and $b_i(0) = z(0) - 1 = N^{-1} \text{Tr } A$ for all i . The density of μ_0 with respect to ν_N is then constant on S_N , and $\mu_0 = \nu_N$. As $t \rightarrow \infty$, we have $b_i(t) \rightarrow \alpha_i$ by (7.22) and (7.27). Then the densities of μ_t with respect to ν_N converge uniformly on S_N to that of π_A , and the total variation distance, which is half the $L^1(\nu_N)$ distance of the densities, tends to zero. $\square$

For the model $A = cW_N$, where $\alpha_i = c\lambda_i$, the first identity in (7.21) shows that v solves the system (1.8) of the introduction. Its right side is a polynomial in v , and in particular locally Lipschitz. Therefore v is the unique solution of that system, and the two definitions agree.

The proof of Lemma 7.6 contains a conserved quantity and a more precise form of the bounds on $z(t) - z$ and $A - B_t$, which we record separately.

Lemma 7.7 (A conserved quantity and the weight of the zero eigenvalue). *For the covariance evolution of Definition 7.5, let $w_i(t) = p_i h_i(t) = e^{-2zt} m_i(t) / q_i$; then $\bar{h}(t) = N^{-1} \sum_i q_i w_i(t)$. For all $t \geq 0$,*

$$\sum_{i=1}^N q_i^2 w_i(t) = N, \quad 1 \geq \bar{h}(t) \geq \vartheta, \quad \bar{h}(t) \longrightarrow \vartheta \quad (t \rightarrow \infty). \quad (7.28)$$

Moreover, with $\Omega_t = (s - d)e^{-2dt}$,

$$\text{osc } w(t) \leq \Omega_t, \quad |z(t) - z| \leq \frac{\Omega_t}{\bar{h}(t)}, \quad \|A - B_t\| \leq \frac{(1 + s)\Omega_t}{\bar{h}(t)}. \quad (7.29)$$

Proof. By the fifth part of Lemma 7.1, we have $\sum_i q_i^2 w_i = \sum_i q_i h_i = N$, and the remaining assertions of (7.28) are contained in the third part of that lemma. The bounds (7.29) are (7.25) and (7.27). $\square$

Note that ϑ is the long-time limit of $\bar{h}$, not a rate of decay. The decay of $\bar{h}$ towards ϑ is governed by the spectral measure ρ on $[2d, 2s]$. Under the hypothesis (7.30) of the next subsection, Lemma 7.10 shows that $\bar{h}(t)$ is comparable to $(1 + t)^{-1/2}$ for $0 \leq t \leq d^{-1}$ and to $\sqrt{d}$ for $t \geq d^{-1}$, with constants depending only on the constant K in (7.30).

Remark 7.8 (The variances are not second moments). The Bingham comparison law μ_t is an explicit comparison law. In general, it is not equal to the time- t law $\nu_N P_t^A$ of the diffusion, since the dynamics does not preserve the family of Bingham laws; the two laws agree at $t = 0$, and for all t when A is a multiple of the identity. The numbers $v_i(t)$ are the variances of the Gaussian law that is conditioned on S_N to obtain μ_t . In general, they

are equal neither to the second moments $\mu_t(x_i^2)$ nor to the second moments of the time- t law, and in general q_i is not equal to $\pi_A(x_i^2)$. The difference is already visible in the first derivative at $t = 0$. On S_N we have

$$L_A x_i^2 = 2 + 2(\alpha_i - 1)x_i^2 - \frac{2}{N} x_i^2 x^\top A x,$$

and under ν_N we have $\int x_i^2 x_j^2 d\nu_N = N(1 + 2\delta_{ij})/(N + 2)$. With $\bar{\alpha} = N^{-1} \text{Tr } A$, we find

$$\left. \frac{d}{dt} \right|_{t=0} \int x_i^2 d(\nu_N P_t^A) = \int L_A x_i^2 d\nu_N = \frac{2N}{N+2} (\alpha_i - \bar{\alpha}), \quad v'_i(0) = 2(\alpha_i - \bar{\alpha}),$$

so the two vectors of initial derivatives differ unless A is a multiple of the identity. For a general nonscalar A , the equations for the second moments of the coordinates under the time- t law involve mixed fourth moments, since $L_A x_i^2$ contains the term $x_i^2 x^\top A x$. Our use of μ_t in Section 9 does not require any identification of moments: it compares the two laws through the forward equation.

7.3 Estimates under two-sided precision-spacing bounds

In this subsection we fix $K \geq 1$ and assume the following two-sided precision-spacing bounds with constant K :

$$\begin{aligned} 0 < d \leq 1, \quad p_1 = d, \quad \frac{1}{K} \left(\frac{i-1}{N} \right)^{2/3} \leq p_i - d \leq K \left(\frac{i-1}{N} \right)^{2/3} \quad (2 \leq i \leq N), \\ \sum_{i=1}^N \frac{1}{p_i} = N, \quad Nd^{3/2} \geq 1. \end{aligned} \quad (7.30)$$

In particular, $d = \min_i p_i$ and $s \leq d + K \leq 1 + K$. Constants denoted C_K or c_K are positive and depend only on K ; their values may change from line to line. For positive X, Y , we write $X \asymp_K Y$ if $c_K Y \leq X \leq C_K Y$. We abbreviate

$$a(t) = \sqrt{d} + (1+t)^{-1/2}, \quad t \geq 0. \quad (7.31)$$

We use the following elementary properties of a : it is decreasing, $a(t/2) \leq \sqrt{2} a(t)$, $\int_0^t a(u) du \leq \sqrt{d} t + 2\sqrt{t}$, $a(t) \leq 3(1+t)^{-1/2}$ for $t \leq d^{-1}$, and $a(t) \leq 2\sqrt{d}$ for $t \geq d^{-1}$.

Remark 7.9 (Precision-spacing bounds and the order of the coordinates). The hypothesis does not depend on the order of the indices $i \geq 2$, in the following sense. Let $g_1 = 0$ and $g_i = p_i - d$, and let $g_{(1)} \leq \dots \leq g_{(N)}$ be the increasing rearrangement. The bounds in (7.30) increase with i . Thus the i numbers $g_1, \dots, g_i$ are at most $K((i-1)/N)^{2/3}$, which implies that $g_{(i)} \leq K((i-1)/N)^{2/3}$. The $N-i+1$ numbers g_j with $j \geq i$ are at least $K^{-1}((i-1)/N)^{2/3}$; at most $i-1$ of the g_j are smaller, and $g_{(i)} \geq K^{-1}((i-1)/N)^{2/3}$.

Hence the increasing rearrangement of p satisfies (7.30) with the same K . For the model at fixed high temperature and in the approach to the critical temperature from above, $p_i - d = c(\lambda_1 - \lambda_i)$, and (7.30) holds at every fixed confidence in the sense of Section 3.4, with K depending on the confidence (and on c at fixed temperature), by Lemma 4.6 and Theorem 4.13. This is checked in Sections 13 and 14.

Lemma 7.10 (Bounds on $\bar{h}$ under precision-spacing bounds). *Assume (7.30). Then, for all $t \geq 0$,*

$$\vartheta \asymp_K \sqrt{d}, \quad \bar{h}(t) \asymp_K \sqrt{d} + (1+t)^{-1/2} = a(t). \quad (7.32)$$

Moreover, $\bar{h}$ is completely monotone and has the representation

$$\bar{h}(t) = \vartheta + \int_{[2d, \infty)} e^{-\lambda t} \rho(d\lambda), \quad \rho \geq 0, \quad \rho([2d, \infty)) = 1 - \vartheta. \quad (7.33)$$

Proof. The representation (7.33) and complete monotonicity are the third part of Lemma 7.1. It remains to prove (7.32). The proof has four steps: a bound on the sum S of (7.5), from which $\vartheta \asymp_K \sqrt{d}$ follows; bounds on the Laplace transform of $\bar{h}$; the upper bound on $\bar{h}$; and the lower bound on $\bar{h}$.

Step 1: a rational sum. We show that, for $0 \leq \zeta \leq 1$,

$$S(\zeta) = \frac{1}{N} \sum_{i=1}^N \frac{1}{p_i(p_i + \zeta/2)} \asymp_K \frac{1}{\sqrt{d} + \sqrt{d + \zeta}}. \quad (7.34)$$

Let $d_\zeta = d + \zeta/2$, and let $\phi(x) = [(d+x)(d_\zeta+x)]^{-1}$ for $x \geq 0$, a decreasing function. For $i \geq 2$ and $y_i = (i-1)/N$, we have $\phi(Ky_i^{2/3}) \leq [p_i(p_i + \zeta/2)]^{-1} \leq \phi(K^{-1}y_i^{2/3})$ by (7.30). For a decreasing function g on $[0, 1]$, we have

$$\int_{1/N}^1 g(y) dy \leq \frac{1}{N} \sum_{j=1}^{N-1} g(j/N) \leq \int_0^1 g(y) dy, \quad \int_0^{1/N} g(y) dy \leq \frac{g(0)}{N}.$$

The summand $i = 1$ in (7.34) equals $\phi(0)/N = (Ndd_\zeta)^{-1}$. Applying these inequalities to $g(y) = \phi(Ky^{2/3})$ and to $g(y) = \phi(K^{-1}y^{2/3})$, we obtain

$$I_K \leq S(\zeta) \leq I_{K^{-1}} + \frac{1}{Ndd_\zeta}, \quad I_L = \int_0^1 \phi(Ly^{2/3}) dy = \frac{3}{2L^{3/2}} \int_0^L \frac{\sqrt{x} dx}{(d+x)(d_\zeta+x)}, \quad (7.35)$$

where we substituted $x = Ly^{2/3}$. The substitution $x = y^2$ and partial fractions give, for $d \neq d_\zeta$, and by continuity also for $d = d_\zeta$,

$$\int_0^\infty \frac{\sqrt{x} dx}{(x+d)(x+d_\zeta)} = \frac{\pi}{\sqrt{d} + \sqrt{d_\zeta}}. \quad (7.36)$$

Since $d_\zeta \leq d + \zeta \leq 2d_\zeta$, we have $\sqrt{d} + \sqrt{d_\zeta} \leq \sqrt{d} + \sqrt{d + \zeta} \leq \sqrt{2}(\sqrt{d} + \sqrt{d_\zeta})$. It therefore suffices to prove (7.34) with d_ζ in place of $d + \zeta$.

For the upper bound, we have $I_{K-1} \leq \frac{3}{2}K^{3/2}\pi(\sqrt{d} + \sqrt{d_\zeta})^{-1}$ by (7.36). For the summand $i = 1$, we use $Nd^{3/2} \geq 1$ and $d \leq d_\zeta$ to find

$$\frac{1}{Ndd_\zeta} = \frac{1}{Nd^{3/2}} \cdot \frac{\sqrt{d}}{d_\zeta} \leq \frac{\sqrt{d}}{d_\zeta} = \frac{1}{\sqrt{d} + \sqrt{d_\zeta}} \left(\frac{d}{d_\zeta} + \sqrt{\frac{d}{d_\zeta}} \right) \leq \frac{2}{\sqrt{d} + \sqrt{d_\zeta}}.$$

For the lower bound, we bound I_K from below. If $2d_\zeta \leq K$, we restrict the integral in (7.35) to $[d_\zeta, 2d_\zeta]$, where $\sqrt{x} \geq \sqrt{d_\zeta}$, $d + x \leq 3d_\zeta$, and $d_\zeta + x \leq 3d_\zeta$. This gives

$$\int_0^K \frac{\sqrt{x} dx}{(d+x)(d_\zeta+x)} \geq d_\zeta \cdot \frac{\sqrt{d_\zeta}}{9d_\zeta^2} = \frac{1}{9\sqrt{d_\zeta}} \geq \frac{1}{9(\sqrt{d} + \sqrt{d_\zeta})}.$$

If $2d_\zeta > K$, then $d_\zeta \geq 1/2$ (since $K \geq 1$) and $d_\zeta \leq 3/2$. Restricting the integral to $[K/4, K/2]$, we find that it is at least a positive constant depending only on K , while $(\sqrt{d} + \sqrt{d_\zeta})^{-1} \leq \sqrt{2}$. In both cases, $I_K \geq c_K(\sqrt{d} + \sqrt{d_\zeta})^{-1}$, which completes the proof of (7.34). At $\zeta = 0$, we have $S(0) = N^{-1} \sum_i q_i^2 = 1/\vartheta$, and (7.34) implies that $\vartheta \asymp_K \sqrt{d}$.

Step 2: the Laplace transform. Let $0 < \zeta \leq 1$. By (7.10), $\hat{h}(\zeta) = G(\zeta)/(\zeta S(\zeta))$. We have $G(\zeta) \leq G(0) = 1$ and $G(\zeta) \geq (1 + K + 1/2)^{-1}$, since $p_i \leq 1 + K$. Using (7.34) and $\sqrt{d} + \sqrt{d + \zeta} \asymp \sqrt{d} + \sqrt{\zeta}$, we obtain

$$\hat{h}(\zeta) \asymp_K \frac{\sqrt{d} + \sqrt{d + \zeta}}{\zeta} \asymp_K \frac{\sqrt{d} + \sqrt{\zeta}}{\zeta}, \quad 0 < \zeta \leq 1. \quad (7.37)$$

Step 3: the upper bound. Since $\bar{h}$ is nonincreasing and nonnegative, for every $t > 0$,

$$\hat{h}(1/t) \geq \int_0^t e^{-u/t} \bar{h}(u) du \geq e^{-1} t \bar{h}(t).$$

For $t \geq 1$, (7.37) with $\zeta = 1/t$ gives $\bar{h}(t) \leq e t^{-1} \hat{h}(1/t) \leq C_K(\sqrt{d} + t^{-1/2}) \leq C_K a(t)$. For $t \leq 1$, we have $\bar{h}(t) \leq 1 \leq \sqrt{2} a(t)$. Thus $\bar{h}(t) \leq C_1 a(t)$ for all $t \geq 0$, for a constant $C_1 = C_1(K)$.

Step 4: the lower bound. By (7.37), there exists $c_1 = c_1(K) > 0$ such that $\hat{h}(1/t) \geq c_1 \sqrt{t}$ for $t \geq 1$. Fix $\varepsilon = \varepsilon(K) \in (0, 1)$ such that $C_1(\varepsilon + 2\sqrt{\varepsilon}) \leq c_1/2$, and let $1 \leq t \leq d^{-1}$. Using Step 3 and $\sqrt{d} t \leq \sqrt{t}$, we have

$$\int_0^{\varepsilon t} e^{-u/t} \bar{h}(u) du \leq C_1 \int_0^{\varepsilon t} a(u) du \leq C_1(\sqrt{d} \varepsilon t + 2\sqrt{\varepsilon t}) \leq C_1(\varepsilon + 2\sqrt{\varepsilon}) \sqrt{t} \leq \frac{c_1}{2} \sqrt{t}.$$

Since $\bar{h}$ is nonincreasing, the remaining part of the integral defining $\hat{h}(1/t)$ is at most $\bar{h}(\varepsilon t) \int_0^\infty e^{-u/t} du = t \bar{h}(\varepsilon t)$. Then $t \bar{h}(\varepsilon t) \geq (c_1/2) \sqrt{t}$, that is, $\bar{h}(v) \geq (c_1 \sqrt{\varepsilon}/2) v^{-1/2}$ for

$v = \varepsilon t \in [\varepsilon, \varepsilon/d]$. On this interval, we have $\sqrt{d} \leq \sqrt{\varepsilon/v}$ and hence $a(v) \leq (1 + \varepsilon^{-1/2})v^{-1/2}$. Then $\bar{h}(v) \geq c_K a(v)$ for $\varepsilon \leq v \leq \varepsilon/d$. For $0 \leq v \leq \varepsilon$, we have $\bar{h}(v) \geq e^{-2sv} \geq e^{-2(1+K)}$ by the first part of Lemma 7.1, while $a(v) \leq 2$. For $v \geq \varepsilon/d$, we have $\bar{h}(v) \geq \vartheta \geq c_K \sqrt{d}$ by Step 1, while $a(v) \leq (1 + \varepsilon^{-1/2})\sqrt{d}$. The three intervals cover $[0, \infty)$, since $d \leq 1$. This proves the lower bound in (7.32). $\square$

Lemma 7.11 (The coordinates of the heat vector). *Assume (7.30). Uniformly for every $t \geq 0$ and $1 \leq i \leq N$,*

$$\bar{h}(t) \asymp_K a(t), \quad h_i(t) \asymp_K a(t) \frac{1+t}{1+p_it}, \quad \max_j h_j(t) \leq \frac{C_K}{\sqrt{d}}.$$

Moreover, for every nonzero $y \in [0, \infty)^N$,

$$\max_j h_j(L-u) \frac{h(u)^\top y}{h(L)^\top y} \leq C_K \frac{a(u)}{d}, \quad 0 \leq u \leq L < \infty. \quad (7.38)$$

Proof. The bound on $\bar{h}$ is Lemma 7.10. We use (7.9), $h_i(t) = e^{-2p_it} + 2 \int_0^t e^{-2p_i(t-u)} \bar{h}(u) du$, and recall that $p_i \leq 1+K$.

For $t \leq 1$, we have $e^{-2(1+K)} \leq h_i(t) \leq 1+2t \leq 3$, since $\bar{h} \leq 1$, while $1/(2(2+K)) \leq a(t)(1+t)/(1+p_it) \leq 4$. This proves the bounds on h_i for $t \leq 1$. Let $t \geq 1$. For the lower bound, we keep only the integral over $[t/2, t]$, where $\bar{h}(u) \geq \bar{h}(t)$:

$$h_i(t) \geq 2\bar{h}(t) \int_0^{t/2} e^{-2p_iv} dv = \bar{h}(t) \frac{1-e^{-p_it}}{p_i} \geq \bar{h}(t) \frac{t}{1+p_it} \geq c_K a(t) \frac{1+t}{1+p_it},$$

where we used $1-e^{-x} \geq x/(1+x)$ and $t \geq (1+t)/2$. For the upper bound, on $[t/2, t]$ we use $\bar{h}(u) \leq C_K a(t/2) \leq C_K a(t)$ and $2 \int_0^{t/2} e^{-2p_iv} dv \leq \min\{t, p_i^{-1}\} \leq 2t/(1+p_it)$. On $[0, t/2]$ we use $e^{-2p_i(t-u)} \leq e^{-p_it} \leq (1+p_it)^{-1}$ and $\int_0^{t/2} \bar{h} \leq C_K(\sqrt{d}t + \sqrt{t}) \leq C_K a(t)(1+t)$. Finally, $e^{-2p_it} \leq (1+p_it)^{-1} \leq a(t)(1+t)/(1+p_it)$, since $a(t)(1+t) \geq \sqrt{1+t} \geq 1$. Adding the three contributions gives the upper bound on $h_i(t)$.

For the maximum, note that $a(t)(1+t)/(1+p_it) \leq a(t)(1+t)/(1+dt)$. Since $d \leq 1$, we have $\sqrt{d}(1+t)/(1+dt) \leq \sqrt{d} \max\{1, d^{-1}\} = d^{-1/2}$. Further, $\sqrt{1+t}/(1+dt)$ is at most $\sqrt{1+d^{-1}} \leq \sqrt{2/d}$ if $dt \leq 1$, and at most $\sqrt{2t}/(dt) \leq \sqrt{2}d^{-1/2}$ if $dt \geq 1$. With the upper bound on h_j , this gives $\max_j h_j(t) \leq C_K d^{-1/2}$.

For (7.38), let $0 \leq u \leq L$. Since $y \neq 0$ has nonnegative coordinates and $h_i(L) > 0$, we have $h(u)^\top y/h(L)^\top y \leq \max_i h_i(u)/h_i(L)$. By the bounds on h_i ,

$$\frac{h_i(u)}{h_i(L)} \leq C_K \frac{a(u)}{a(L)} \cdot \frac{1+p_iL}{1+L} \cdot \frac{1+u}{1+p_iu} \leq C_K \frac{a(u)}{\sqrt{d}} \max\{1, p_i\} \leq C_K \frac{a(u)}{\sqrt{d}}.$$

In the second step, we used $a(L) \geq \sqrt{d}$ and that $x \mapsto (1+p_ix)/(1+x)$ is monotone with values between 1 and p_i ; the ratio of its values at L and at $u \leq L$ is therefore at most $\max\{1, p_i\}$. Multiplying by $\max_j h_j(L-u) \leq C_K d^{-1/2}$ yields (7.38). $\square$

The next lemma bounds the relative errors $e_i(t) = p_i v_i(t) - 1$ of (7.4), and their mean square $N^{-1} \sum_i e_i(t)^2 = 2\mathcal{I}_N(t)/N$, before and after the time d^{-1} .

Lemma 7.12 (Relative errors before and after the time d^{-1}). *Assume (7.30), and let $T = d^{-1}$. There exists $C_K < \infty$ such that, for all $1 \leq i \leq N$,*

$$|e_i(t)| \leq \frac{C_K}{1 + p_i t} \quad (0 \leq t \leq T), \quad (7.39)$$

$$\frac{1}{N} \sum_{j=1}^N e_j(t)^2 \leq C_K (1 + t)^{-3/2} \quad (0 \leq t \leq T), \quad (7.40)$$

$$|e_i(T + \tau)| \leq C_K \frac{d}{p_i} (1 + d\tau) e^{-2d\tau} \quad (\tau \geq 0), \quad (7.41)$$

$$\frac{1}{N} \sum_{j=1}^N e_j(T + \tau)^2 \leq C_K d^{3/2} (1 + d\tau)^2 e^{-4d\tau} \quad (\tau \geq 0). \quad (7.42)$$

The factor $1 + d\tau$ in (7.41) comes from the coordinates with $p_i - d \leq d$, whose own relaxation rate $2p_i$ is comparable to the rate $2d$ at which $\bar{h}$ approaches ϑ ; it is integrated in later estimates and cannot be dropped in this argument.

Proof. We first prove the coordinate bounds before T , then the bound on the mean square before T , and finally the bounds after T . By the third part of Lemma 7.1 with $j = 1$, for every $v > 0$,

$$-\bar{h}'(v) \leq \frac{2}{ev} \bar{h}(v/2). \quad (7.43)$$

Coordinates before T . For $t \leq 1$, we have $h_i(t) \leq 1 + 2t \leq 3$ by (7.9) and $\bar{h} \leq 1$, and $\bar{h}(t) \geq c_K$ by Lemma 7.10. Since $p_i \leq 1 + K$, this gives $|e_i(t)| \leq p_i h_i(t) / \bar{h}(t) + 1 \leq C_K \leq C_K(2 + K)/(1 + p_i t)$. Let now $1 \leq t \leq T$. Then $\bar{h}(t) \asymp_K t^{-1/2}$ by Lemma 7.10, since $a(t) \leq 3(1 + t)^{-1/2}$ for $t \leq d^{-1}$.

If $p_i t \leq 1$, then (7.9) and $\sqrt{d}t \leq \sqrt{t}$ give $h_i(t) \leq 1 + 2 \int_0^t \bar{h} \leq 1 + C_K(\sqrt{d}t + \sqrt{t}) \leq C_K \sqrt{t}$. Then $p_i h_i(t) / \bar{h}(t) \leq C_K p_i t \leq C_K$, and $|e_i(t)| \leq C_K + 1 \leq 2(C_K + 1)/(1 + p_i t)$.

If $p_i t \geq 1$, we split the integral in (7.9) at $t/2$ and integrate the part over $[t/2, t]$ by parts, using $\frac{d}{du} e^{-2p_i(t-u)} = 2p_i e^{-2p_i(t-u)}$. The result is the exact identity

$$\begin{aligned} p_i h_i(t) - \bar{h}(t) &= p_i e^{-2p_i t} + 2p_i \int_0^{t/2} e^{-2p_i(t-u)} \bar{h}(u) du \\ &\quad - e^{-p_i t} \bar{h}(t/2) - \int_{t/2}^t e^{-2p_i(t-u)} \bar{h}'(u) du. \end{aligned} \quad (7.44)$$

We divide (7.44) by $\bar{h}(t) \geq c_K t^{-1/2}$ and bound the four terms on the right. The first contributes at most $C_K p_i \sqrt{t} e^{-2p_i t} \leq C_K p_i t e^{-p_i t}$. Using $\int_0^{t/2} \bar{h} \leq C_K \sqrt{t}$, we see that the second contributes at most $2p_i e^{-p_i t} \int_0^{t/2} \bar{h} / \bar{h}(t) \leq C_K p_i t e^{-p_i t}$. The third

contributes at most $e^{-p_i t} \bar{h}(t/2)/\bar{h}(t) \leq C_K e^{-p_i t}$. For $u \in [t/2, t]$, (7.43) implies that $-\bar{h}'(u) \leq (4/(et))\bar{h}(t/4) \leq C_K \bar{h}(t)/t$, and the last term contributes at most $C_K t^{-1} \int_0^\infty e^{-2p_i v} dv \leq C_K/(p_i t)$. Since $(1+x)e^{-x} \leq C/x$ for $x \geq 1$, we obtain $|e_i(t)| = |p_i h_i(t) - \bar{h}(t)|/\bar{h}(t) \leq C_K/(p_i t) \leq 2C_K/(1+p_i t)$. This proves (7.39).

Mean square before T . By (7.39), it suffices to show that $N^{-1} \sum_j (1+p_j t)^{-2} \leq C_K (1+t)^{-3/2}$ for $t \leq T$. At $t = 0$ the sum equals one. For $0 < t \leq T$, the summand $j = 1$ is at most $1/N \leq d^{3/2}$, and $d^{3/2} \leq \min\{1, t^{-3/2}\} \leq 2^{3/2}(1+t)^{-3/2}$. For $j \geq 2$, we have $p_j \geq K^{-1}((j-1)/N)^{2/3}$, and $y \mapsto (1+K^{-1}ty^{2/3})^{-2}$ is decreasing. For $t \geq 1$, we find

$$\frac{1}{N} \sum_{j \geq 2} \frac{1}{(1+p_j t)^2} \leq \int_0^1 \frac{dy}{(1+K^{-1}ty^{2/3})^2} = \frac{3}{2} \left(\frac{K}{t}\right)^{3/2} \int_0^{t/K} \frac{\sqrt{x} dx}{(1+x)^2} \leq \frac{3\pi}{4} \left(\frac{K}{t}\right)^{3/2},$$

where we substituted $x = K^{-1}ty^{2/3}$ and used $\int_0^\infty \sqrt{x}(1+x)^{-2} dx = \pi/2$. For $t \leq 1$, the sum is at most one. Squaring (7.39) proves (7.40).

After T . Using (7.43) and Lemma 7.10, we find that $-\bar{h}'(T) \leq 2d\bar{h}(T/2)/e \leq C_K d a(T/2) \leq C_K d^{3/2}$. The last bound in (7.7) with $j = 1$ implies that

$$-\bar{h}'(T+u) \leq C_K d^{3/2} e^{-2du}, \quad u \geq 0. \quad (7.45)$$

Let $\kappa_i = p_i h_i - \bar{h} = \bar{h} e_i$. From $h'_i = -2p_i h_i + 2\bar{h}$, we find $\kappa'_i = -2p_i \kappa_i - \bar{h}'$. Solving this equation from time T by variation of constants and dividing by $\bar{h}(T+\tau)$, we obtain

$$e_i(T+\tau) = e^{-2p_i \tau} \frac{\bar{h}(T)}{\bar{h}(T+\tau)} e_i(T) - \frac{1}{\bar{h}(T+\tau)} \int_0^\tau e^{-2p_i(\tau-u)} \bar{h}'(T+u) du. \quad (7.46)$$

By Lemma 7.10, we have $\bar{h}(T) \leq C_K a(T) \leq C_K \sqrt{d}$ and $\bar{h}(T+\tau) \geq \vartheta \geq c_K \sqrt{d}$, so that $\bar{h}(T)/\bar{h}(T+\tau) \leq C_K$. Further, (7.39) at $t = T$ gives $|e_i(T)| \leq C_K/(1+p_i/d) \leq C_K d/p_i$. The first term in (7.46) is at most $C_K(d/p_i)e^{-2d\tau}$. By (7.45), the second term is at most

$$\begin{aligned} C_K \frac{d^{3/2}}{\sqrt{d}} \int_0^\tau e^{-2p_i(\tau-u)} e^{-2du} du &= C_K d e^{-2d\tau} \int_0^\tau e^{-2(p_i-d)(\tau-u)} du \\ &\leq C_K d e^{-2d\tau} \min \left\{ \tau, \frac{1}{2(p_i-d)} \right\}, \end{aligned}$$

where the minimum is τ when $p_i = d$. We claim that $d \min\{\tau, (2(p_i-d))^{-1}\} \leq 2(1+d\tau)d/p_i$. If $p_i - d \leq d$, then $d/p_i \geq 1/2$ and $d\tau \leq 2(1+d\tau)d/p_i$. If $p_i - d > d$, then $2(p_i-d) \geq p_i$ and $d/(2(p_i-d)) \leq d/p_i$. This proves (7.41). Finally, $N^{-1} \sum_j (d/p_j)^2 = d^2/\vartheta \leq C_K d^{3/2}$ by Lemma 7.10, and squaring (7.41) proves (7.42). $\square$

Lemma 7.13 (Derivatives of the heat vector). *Assume (7.30). There exists $C_K < \infty$ such that, for every $t \geq 0$ and $1 \leq i \leq N$,*

$$\max_j p_j h_j(t)^2 \leq C_K(1+dt), \quad \max_j \sqrt{p_j} |h'_j(t)| \leq \frac{C_K}{1+t}, \quad |h'_i(t)| \leq \frac{C_K h_i(t)}{1+t}. \quad (7.47)$$

In particular, for all $t, u \geq 0$,

$$\max_j p_j [h_j(t+u) - h_j(t)]^2 \leq C_K \log^2 \left(1 + \frac{u}{1+t}\right), \quad (7.48)$$

and, for every nonzero $y \in [0, \infty)^N$, $|\frac{d}{dt}h(t)^\top y| \leq C_K h(t)^\top y / (1+t)$.

Proof. Let $T = d^{-1}$. We use Lemma 7.11, the identity $h'_i = -2\bar{h}e_i$ of (7.11), and Lemma 7.12.

For the first bound, let $t \geq 1$. Since $(1+x)^2 \geq 4x$, we have $p_i/(1+p_it)^2 \leq 1/(4t)$. Using Lemma 7.11, we find

$$\begin{aligned} p_i h_i(t)^2 &\leq C_K a(t)^2 \frac{p_i(1+t)^2}{(1+p_it)^2} \leq C_K a(t)^2 \frac{(1+t)^2}{4t} \\ &\leq C_K (d + (1+t)^{-1})(1+t) \leq C_K(1+dt), \end{aligned}$$

where we used $a(t)^2 \leq 2(d + (1+t)^{-1})$, $(1+t)/(4t) \leq 1/2$, and $d \leq 1$. For $t \leq 1$, we use $p_i \leq 1+K$ and $h_i \leq 3$.

For the third bound, note that Lemma 7.11 implies $\bar{h}(t)/h_i(t) \leq C_K(1+p_it)/(1+t)$. If $t \leq T$, then (7.39) yields

$$\frac{|h'_i(t)|}{h_i(t)} = \frac{2\bar{h}(t)|e_i(t)|}{h_i(t)} \leq C_K \frac{1+p_it}{1+t} \cdot \frac{1}{1+p_it} = \frac{C_K}{1+t}.$$

If $t = T + \tau$ with $\tau \geq 0$, we use (7.41) and $d(1+p_it)/p_i \leq 1+dt = 2+d\tau$ to find

$$\frac{|h'_i(t)|}{h_i(t)} \leq C_K \frac{1+p_it}{1+t} \cdot \frac{d}{p_i} (1+d\tau)e^{-2d\tau} \leq \frac{C_K}{1+t} (2+d\tau)(1+d\tau)e^{-2d\tau} \leq \frac{C_K}{1+t}.$$

For the second bound, $\sqrt{p_i}|h'_i| = 2\sqrt{p_i}\bar{h}|e_i|$. For $t \leq 1$, this is at most $2\sqrt{1+K}C_K$. For $1 \leq t \leq T$, we have $\sqrt{p_i}/(1+p_it) \leq 1/(2\sqrt{t})$, since $1+x \geq 2\sqrt{x}$, and $\bar{h}(t) \leq C_K(1+t)^{-1/2}$. Then (7.39) implies that $\sqrt{p_i}|h'_i(t)| \leq C_K/(\sqrt{t}\sqrt{1+t}) \leq C_K/(1+t)$. For $t = T + \tau$, we have $\bar{h}(t) \leq C_K a(t) \leq C_K \sqrt{d}$. Using (7.41) and $p_i \geq d$, we obtain

$$\begin{aligned} \sqrt{p_i}|h'_i(t)| &\leq C_K \frac{d^{3/2}}{\sqrt{p_i}} (1+d\tau)e^{-2d\tau} \leq C_K d(1+d\tau)e^{-2d\tau} \\ &\leq \frac{C_K}{1+t} (2+d\tau)(1+d\tau)e^{-2d\tau} \leq \frac{C_K}{1+t}, \end{aligned}$$

where we used $d(1+t) = d + 1 + d\tau \leq 2 + d\tau$.

Integrating the second bound in (7.47), we find $\sqrt{p_j}|h_j(t+u) - h_j(t)| \leq C_K \int_t^{t+u} (1+v)^{-1} dv = C_K \log(1+u/(1+t))$, and squaring yields (7.48). Finally, for $y \geq 0$, we have $|\frac{d}{dt}h(t)^\top y| \leq \sum_j |h'_j(t)|y_j \leq C_K(1+t)^{-1} \sum_j h_j(t)y_j$ by the third bound. $\square$

7.4 Estimates in the critical window

In the critical window, the smallest half-precision is of order $N^{-2/3}$ and no gap between the first two half-precisions is assumed. The following lemma uses only the trace condition and two-sided bounds of order $(i/N)^{2/3}$ on all half-precisions. These hypotheses are part of the critical spectral conditions (4.36), which hold for the model at every fixed confidence by Lemma 4.17. We write $X \asymp Y$ if $c_* Y \leq X \leq C Y$ for positive constants c_*, C that depend only on the quantities named in the statement.

Lemma 7.14 (Heat estimates in the critical window). *Let $N \geq 2$ and $0 < c_0 \leq C_0$, and let $r_1, \dots, r_N > 0$, not necessarily ordered, satisfy*

$$\sum_{i=1}^N \frac{1}{2r_i} = N, \quad c_0 \left(\frac{i}{N}\right)^{2/3} \leq r_i \leq C_0 \left(\frac{i}{N}\right)^{2/3} \quad (1 \leq i \leq N). \quad (7.49)$$

Let H , $h(t)$, and $\bar{h}(t)$ be as in (7.3) with $p_i = 2r_i$; then H is the matrix (7.13). For every fixed $K < \infty$, uniformly in $0 \leq t \leq KN^{2/3}$ and $1 \leq i \leq N$,

$$\bar{h}(t) = \frac{1}{N} \sum_{j=1}^N h_j(t) \asymp \frac{1}{\sqrt{1+t}}, \quad h_i(t) \asymp \frac{\sqrt{1+t}}{1+r_i t}, \quad (7.50)$$

with constants depending only on c_0, C_0 , and K . There exists $C < \infty$, depending only on c_0 and C_0 , such that for all $u, w \geq 0$,

$$\sum_{i=1}^N \frac{1}{(1+r_i w)^2(1+r_i u)} \leq \frac{CN}{(1+w)^{3/2}} \min \left\{ 1, \frac{1+w}{1+u} \right\}. \quad (7.51)$$

The entries of e^{tH} are given by

$$(e^{tH})_{ij} = e^{-4r_i t} \delta_{ij} + \frac{2}{N} \int_0^t e^{-4r_i(t-u)} h_j(u) du. \quad (7.52)$$

Proof. The identity (7.52) is (7.8) with $p_i = 2r_i$. In this proof, c_* and C denote positive constants depending only on c_0 and C_0 , and C_K also depends on K . Note that $p_i = 2r_i \leq 2C_0$. The proof has five steps: bounds on the Laplace transform of $\bar{h}$, the upper bound on $\bar{h}$, the lower bound on $\bar{h}$, the bounds on h_i , and the bound (7.51).

Step 1: the Laplace transform. By (7.10) and (7.14), for $\zeta > 0$,

$$\hat{\bar{h}}(\zeta) = \frac{G(\zeta)}{\zeta S(\zeta)}, \quad G(\zeta) = \frac{1}{N} \sum_i \frac{1}{2r_i + \zeta/2}, \quad S(\zeta) = \frac{1}{N} \sum_i \frac{1}{r_i(\zeta + 4r_i)}.$$

For $0 < \zeta \leq 1$, we have $(2C_0 + 1/2)^{-1} \leq G(\zeta) \leq G(0) = 1$. We show that

$$S(\zeta) \leq C\zeta^{-1/2} \quad (\zeta > 0), \quad S(\zeta) \geq c_* \zeta^{-1/2} \quad (N^{-2/3} \leq \zeta \leq 1). \quad (7.53)$$

For the upper bound, the lower bound in (7.49) and the monotonicity of $y \mapsto [c_0 y^{2/3}(\zeta + 4c_0 y^{2/3})]^{-1}$ give

$$S(\zeta) \leq \int_0^\infty \frac{dy}{c_0 y^{2/3}(\zeta + 4c_0 y^{2/3})} = \frac{6}{(4c_0)^{3/2} \sqrt{\zeta}} \int_0^\infty \frac{dx}{\sqrt{x}(1+x)} = \frac{3\pi}{4c_0^{3/2} \sqrt{\zeta}},$$

where we substituted $4c_0 y^{2/3} = \zeta x$. For the lower bound, suppose first that $N^{-2/3} \leq \zeta \leq 2^{-2/3}$. The interval $[N\zeta^{3/2}, 2N\zeta^{3/2}]$ has length at least one and lies in $[1, N]$. It contains at least $N\zeta^{3/2}/2$ integers i , and for each of them we have $r_i \leq C_0(2\zeta^{3/2})^{2/3} \leq 2C_0\zeta$ and $[r_i(\zeta + 4r_i)]^{-1} \geq [2C_0(1 + 8C_0)\zeta^2]^{-1}$. Summing over these i yields $S(\zeta) \geq c_* \zeta^{-1/2}$. If $2^{-2/3} \leq \zeta \leq 1$, the bound $r_i \leq C_0$ yields $S(\zeta) \geq [C_0(1 + 4C_0)]^{-1} \geq c_* \zeta^{-1/2}$. Combining (7.53) with the bounds on G , we obtain

$$c_* \zeta^{-1/2} \leq \hat{h}(\zeta) \leq C\zeta^{-1/2}, \quad N^{-2/3} \leq \zeta \leq 1. \quad (7.54)$$

Step 2: the upper bound on $\bar{h}$. As in the proof of Lemma 7.10, $\bar{h}(t) \leq e t^{-1} \hat{h}(1/t)$ for $t > 0$. For $1 \leq t \leq N^{2/3}$, this bound and (7.54) imply that $\bar{h}(t) \leq C t^{-1/2}$. For $t \leq 1$, we have $\bar{h}(t) \leq 1$. For $N^{2/3} \leq t \leq KN^{2/3}$, monotonicity gives $\bar{h}(t) \leq \bar{h}(N^{2/3}) \leq CN^{-1/3} \leq C\sqrt{K} t^{-1/2}$. Altogether, $\bar{h}(t) \leq C_K(1+t)^{-1/2}$ for $0 \leq t \leq KN^{2/3}$.

Step 3: the lower bound on $\bar{h}$. Let $\bar{\rho} = \vartheta\delta_0 + \rho$. By Lemma 7.1, we have $\bar{h}(t) = \int e^{-\lambda t} \bar{\rho}(d\lambda)$ and $\hat{h}(\zeta) = \int \bar{\rho}(d\lambda)/(\zeta + \lambda)$ for $\zeta > 0$. Let c_2, C_2 be the constants in (7.54), and fix $L \geq 1$ such that $2C_2 L^{-1/2} \leq c_2/2$. Let $N^{-2/3} \leq \zeta \leq 1/L$. For $\lambda > L\zeta$, we have $(\zeta + \lambda)^{-1} \leq \lambda^{-1} \leq 2(L\zeta + \lambda)^{-1}$, so

$$\int_{(L\zeta, \infty)} \frac{\bar{\rho}(d\lambda)}{\zeta + \lambda} \leq 2\hat{h}(L\zeta) \leq 2C_2(L\zeta)^{-1/2} \leq \frac{c_2}{2} \zeta^{-1/2},$$

where we used (7.54) at $L\zeta \in [N^{-2/3}, 1]$. It follows that $\int_{[0, L\zeta]} \bar{\rho}(d\lambda)/(\zeta + \lambda) \geq (c_2/2)\zeta^{-1/2}$. The integrand is at most ζ^{-1} , which implies that $\bar{\rho}([0, L\zeta]) \geq (c_2/2)\zeta^{1/2}$. Then

$$\bar{h}(1/\zeta) \geq \int_{[0, L\zeta]} e^{-\lambda/\zeta} \bar{\rho}(d\lambda) \geq e^{-L} \bar{\rho}([0, L\zeta]) \geq e^{-L} \frac{c_2}{2} \zeta^{1/2}.$$

Setting $t = 1/\zeta$, we obtain $\bar{h}(t) \geq c_* t^{-1/2}$ for $L \leq t \leq N^{2/3}$. For $t \leq L$, we have $\bar{h}(t) \geq e^{-4C_0 t} \geq e^{-4C_0 L}$ by the first part of Lemma 7.1. For $t \geq N^{2/3}$, we use $\bar{h}(t) \geq \vartheta$. By (7.2), $\vartheta = N / \sum_i (2r_i)^{-2} = 4N / \sum_i r_i^{-2}$, and (7.49) gives

$$C_0^{-2} N^{4/3} \leq r_1^{-2} \leq \sum_i r_i^{-2} \leq c_0^{-2} N^{4/3} \sum_{i \geq 1} i^{-4/3} \leq 4c_0^{-2} N^{4/3}.$$

This gives $c_0^2 N^{-1/3} \leq \vartheta \leq 4C_0^2 N^{-1/3}$; in particular, $\sum_j h_j(t)$ decreases to $N\vartheta \asymp N^{2/3}$. For $t \geq N^{2/3}$, we have $(1+t)^{-1/2} \leq N^{-1/3}$ and therefore $\bar{h}(t) \geq c_0^2(1+t)^{-1/2}$. We conclude that $\bar{h}(t) \geq c_*(1+t)^{-1/2}$ for all $t \geq 0$, and together with Step 2 this proves the first bound in (7.50).

Step 4: the coordinates. By (7.9), $h_i(t) = e^{-4r_i t} + 2 \int_0^t e^{-4r_i(t-u)} \bar{h}(u) du$. For $t \leq 1$, we have $e^{-4C_0} \leq h_i(t) \leq 3$, while $(1 + C_0)^{-1} \leq \sqrt{1+t}/(1+r_i t) \leq \sqrt{2}$. Let $1 \leq t \leq KN^{2/3}$. For the lower bound, we keep the integral over $[t/2, t]$ and use Step 3 and $1 - e^{-x} \geq x/(1+x)$:

$$h_i(t) \geq 2\bar{h}(t) \int_0^{t/2} e^{-4r_i v} dv = \bar{h}(t) \frac{1 - e^{-2r_i t}}{2r_i} \geq \bar{h}(t) \frac{t}{1 + 2r_i t} \geq c_* \frac{\sqrt{1+t}}{1 + r_i t}.$$

For the upper bound, we have $\bar{h}(u) \leq C_K(1+t)^{-1/2}$ for $u \in [t/2, t]$ by Step 2, and $2 \int_0^{t/2} e^{-4r_i v} dv \leq \min\{t, (2r_i)^{-1}\} \leq 2t/(1+r_i t)$. Thus the integral over $[t/2, t]$ is at most $C_K \sqrt{1+t}/(1+r_i t)$. The integral over $[0, t/2]$ is at most $2e^{-2r_i t} \int_0^{t/2} \bar{h} \leq C_K e^{-2r_i t} \sqrt{t} \leq C_K \sqrt{1+t}/(1+r_i t)$, and $e^{-4r_i t} \leq (1+r_i t)^{-1}$. This proves the second bound in (7.50).

Step 5: the sum (7.51). Let $\bar{w} = 1 + w$ and $\bar{u} = 1 + u$. Since $r_i \leq C_0$, we have $1 + r_i \bar{w} \leq (1 + C_0)(1 + r_i w)$, and similarly for u . The summand in (7.51) is a decreasing function of r_i , and the lower bound in (7.49) gives

$$\sum_{i=1}^N \frac{1}{(1 + r_i w)^2 (1 + r_i u)} \leq (1 + C_0)^3 \sum_{i=1}^N \phi(c_0(i/N)^{2/3}), \quad \phi(x) = \frac{1}{(1 + \bar{w}x)^2 (1 + \bar{u}x)}.$$

Since ϕ is decreasing, the right side is at most $(1 + C_0)^3 N \int_0^1 \phi(c_0 y^{2/3}) dy$, and the substitution $x = c_0 y^{2/3}$ bounds this by $CN \int_0^\infty \sqrt{x} \phi(x) dx$. If $\bar{u} \leq \bar{w}$, we bound the last factor of ϕ by one and substitute $x = x'/\bar{w}$:

$$\int_0^\infty \sqrt{x} \phi(x) dx \leq \int_0^\infty \frac{\sqrt{x} dx}{(1 + \bar{w}x)^2} = \bar{w}^{-3/2} \int_0^\infty \frac{\sqrt{x'} dx'}{(1 + x')^2} = \frac{\pi}{2} \bar{w}^{-3/2}.$$

If $\bar{u} \geq \bar{w}$, we use $(1 + \bar{u}x)^{-1} \leq (\bar{u}x)^{-1}$ and the same substitution:

$$\int_0^\infty \sqrt{x} \phi(x) dx \leq \frac{1}{\bar{u}} \int_0^\infty \frac{dx}{\sqrt{x} (1 + \bar{w}x)^2} = \frac{\bar{w}^{-1/2}}{\bar{u}} \int_0^\infty \frac{dx'}{\sqrt{x'} (1 + x')^2} = \frac{\pi}{2} \bar{w}^{-3/2} \frac{\bar{w}}{\bar{u}}.$$

In both cases, the integral is at most $(\pi/2) \bar{w}^{-3/2} \min\{1, \bar{w}/\bar{u}\}$, and (7.51) follows. $\square$

7.5 Positive measures from real squared resolvents

The following lemma identifies limits of positive measures on $[0, \infty)$ from the convergence of their squared resolvents at real arguments. It is used in Sections 14 and 20, where these squared resolvents are computed from exact resolvent identities such as (7.10). The compact-interval moment method goes back to Hausdorff [81, 82]; see [151, Chapter III, pp. 100–143]. After compactifying $[0, \infty)$ to $[0, 1]$, we recover the moments from the squared resolvents by the finite-difference calculation below. A related use of compactification and Hausdorff moments for generalized Stieltjes transforms appears in

[135, proof of Theorem 2, (13)]. We then approximate continuous functions by Bernstein polynomials [29, pp. 1–2]. We include the short proof because we need versions with varying test functions, with uniformity over families of test functions, and with random measures.

Lemma 7.15 (Positive measures from real squared resolvents). *Let μ_n , $n \geq 1$, and μ be positive Borel measures on $[0, \infty)$, and let*

$$T_n(\zeta) = \int_{[0, \infty)} \frac{\mu_n(dy)}{(\zeta + y)^2}, \quad T(\zeta) = \int_{[0, \infty)} \frac{\mu(dy)}{(\zeta + y)^2}, \quad \zeta \geq 1.$$

Suppose that $T_n(1)$ and $T(1)$ are finite and that $T_n(\zeta) \rightarrow T(\zeta)$ for every rational $\zeta \geq 1$.

1. *For every continuous function $f : [0, 1] \rightarrow \mathbb{R}$,*

$$\int_{[0, \infty)} \frac{f((1+y)^{-1})}{(1+y)^2} \mu_n(dy) \longrightarrow \int_{[0, \infty)} \frac{f((1+y)^{-1})}{(1+y)^2} \mu(dy). \quad (7.55)$$

In particular, $\mu_n \rightarrow \mu$ vaguely.

2. *Let $\varphi_n, \varphi : [0, \infty) \rightarrow \mathbb{R}$ be continuous functions such that $\sup_{0 \leq y \leq R} |\varphi_n(y) - \varphi(y)| \rightarrow 0$ for every $R < \infty$, and*

$$\lim_{R \rightarrow \infty} \sup_n \sup_{y \geq R} (1+y)^2 |\varphi_n(y)| = 0, \quad \lim_{R \rightarrow \infty} \sup_{y \geq R} (1+y)^2 |\varphi(y)| = 0. \quad (7.56)$$

Then $\int \varphi_n d\mu_n \rightarrow \int \varphi d\mu$.

3. *Let Θ be a set, and let $\varphi_{n,\theta}, \varphi_\theta$, $\theta \in \Theta$, be continuous functions on $[0, \infty)$. Suppose that $\sup_\theta \sup_{0 \leq y \leq R} |\varphi_{n,\theta}(y) - \varphi_\theta(y)| \rightarrow 0$ for every R , and that (7.56) holds with $\sup_\theta$ inserted before $\sup_{y \geq R}$ in both conditions. Suppose also that the family $(\varphi_\theta)_{\theta \in \Theta}$ is uniformly bounded and uniformly equicontinuous on $[0, R]$ for every R . Then $\sup_\theta \left| \int \varphi_{n,\theta} d\mu_n - \int \varphi_\theta d\mu \right| \rightarrow 0$.*

4. *Suppose instead that μ_n and μ are random measures, that $T_n(1)$ and $T(1)$ are almost surely finite, and that $T_n(\zeta) - T(\zeta) \rightarrow 0$ in probability for every rational $\zeta \geq 1$. Then the conclusion of the first part holds in probability for every continuous f , and also for every random f that is a measurable random element of $C([0, 1])$. The conclusions of the second and third parts hold in probability for deterministic tests that satisfy their hypotheses. They also hold in probability for random tests under the following conditions.*

- *The tests are measurable random elements of $C([0, \infty))$, with the topology of uniform convergence on compact sets. In the third part, all suprema over θ that appear in the hypotheses and in the conclusion are measurable.*
- *For each n , $\int |\varphi_n| d\mu_n + \int |\varphi| d\mu < \infty$ almost surely, and in the third part the same holds for each θ .*

- For every $R < \infty$, $\sup_{\theta} \sup_{0 \leq y \leq R} |\varphi_{n,\theta}(y) - \varphi_{\theta}(y)| \rightarrow 0$ in probability.
- For every $\varepsilon > 0$, there exist events $\mathcal{E}_n^{\varepsilon}$ with $\liminf_n \mathbb{P}(\mathcal{E}_n^{\varepsilon}) \geq 1 - \varepsilon$ and a deterministic function τ_{ε} with $\tau_{\varepsilon}(R) \rightarrow 0$ as $R \rightarrow \infty$ such that, for each n , on $\mathcal{E}_n^{\varepsilon}$,

$$\sup_{\theta \in \Theta} \sup_{y \geq R} (1+y)^2 (|\varphi_{n,\theta}(y)| + |\varphi_{\theta}(y)|) \leq \tau_{\varepsilon}(R) \quad \text{for every } R \geq 1.$$

In the third part, we also assume that on $\mathcal{E}_n^{\varepsilon}$ the functions φ_{θ} are bounded by a deterministic constant B_{ε} and have, on each interval $[0, R]$, a modulus of continuity bounded by a deterministic function $\omega_{\varepsilon,R}$ with $\omega_{\varepsilon,R}(r) \rightarrow 0$ as $r \downarrow 0$.

In the second part, the suprema over θ are omitted.

Proof. We first compactify, then recover the moments by finite differences, approximate continuous functions by Bernstein polynomials, and finally treat varying tests, families of tests, and random measures.

Step 1: compactification. Let λ_n and λ be the images of the finite measures $(1+y)^{-2}\mu_n(dy)$ and $(1+y)^{-2}\mu(dy)$ under the map $y \mapsto x = (1+y)^{-1}$ from $[0, \infty)$ to $(0, 1]$. They are finite measures on $[0, 1]$ with no atom at 0 and with total masses $T_n(1)$ and $T(1)$. The left side of (7.55) is $\int f d\lambda_n$. For $t \geq 0$, let $F_t(x) = (1+tx)^{-2}$. Since $(1+t/(1+y))^{-2}(1+y)^{-2} = (1+t+y)^{-2}$, we have

$$\int F_t d\lambda_n = T_n(1+t), \quad \int F_t d\lambda = T(1+t). \quad (7.57)$$

In particular, $\lambda_n([0, 1]) = T_n(1) \rightarrow T(1) = \lambda([0, 1])$.

Step 2: moments. Fix an integer $k \geq 1$ and a rational $\delta > 0$, and let

$$D_{k,\delta}(x) = \frac{1}{(-1)^k(k+1)! \delta^k} \sum_{j=0}^k (-1)^{k-j} \binom{k}{j} F_{j\delta}(x).$$

We have $\partial_t^k F_t(x) = (-1)^k(k+1)! x^k(1+tx)^{-k-2}$. The k th forward difference of $t \mapsto F_t(x)$ with step δ is the integral of $\partial_t^k F_t(x)$ at $t = t_1 + \dots + t_k$ over $(t_1, \dots, t_k) \in [0, \delta]^k$, by the fundamental theorem of calculus applied k times. Hence

$$D_{k,\delta}(x) = \delta^{-k} \int_{[0,\delta]^k} \frac{x^k}{(1+(t_1+\dots+t_k)x)^{k+2}} dt_1 \cdots dt_k.$$

For $x \in [0, 1]$ and $0 \leq \tau \leq k\delta$, Bernoulli's inequality $(1+\tau x)^{-k-2} \geq 1 - (k+2)\tau x$ implies that $0 \leq x^k - x^k(1+\tau x)^{-k-2} \leq (k+2)k\delta$. Taking $\tau = t_1 + \dots + t_k$ and averaging over the cube, we obtain

$$\sup_{0 \leq x \leq 1} |D_{k,\delta}(x) - x^k| \leq k(k+2)\delta. \quad (7.58)$$

By (7.57), $\int D_{k,\delta} d\lambda_n$ is a finite linear combination of the numbers $T_n(1 + j\delta)$, at rational arguments, and it converges to $\int D_{k,\delta} d\lambda$. Further, (7.58) gives

$$\left| \int x^k \lambda_n(dx) - \int x^k \lambda(dx) \right| \leq \left| \int D_{k,\delta} d\lambda_n - \int D_{k,\delta} d\lambda \right| + k(k+2)\delta(T_n(1) + T(1)).$$

Letting $n \rightarrow \infty$ and then $\delta \downarrow 0$ through the rationals, we find that every moment of λ_n converges to the corresponding moment of λ . For $k = 0$ this is $T_n(1) \rightarrow T(1)$.

Step 3: continuous functions. For a continuous f on $[0, 1]$ and $m \geq 1$, let $B_m f(x) = \sum_{j=0}^m f(j/m) \binom{m}{j} x^j (1-x)^{m-j}$ be its Bernstein polynomial, and let ω_f be its modulus of continuity. Let J be a binomial random variable with parameters m and x . Then $B_m f(x) - f(x) = \mathbb{E}[f(J/m) - f(x)]$, and J/m has mean x and variance $x(1-x)/m \leq 1/(4m)$. Splitting according to whether $|J/m - x| \leq r$ and using Chebyshev's inequality, we find

$$\|B_m f - f\|_\infty \leq \omega_f(r) + 2\|f\|_\infty \frac{1}{4mr^2} = \omega_f(r) + \frac{\|f\|_\infty}{2mr^2}.$$

With $r = m^{-1/4}$, the right side tends to zero as $m \rightarrow \infty$. Then

$$\left| \int f d\lambda_n - \int f d\lambda \right| \leq \left| \int B_m f d\lambda_n - \int B_m f d\lambda \right| + \|B_m f - f\|_\infty (T_n(1) + T(1)).$$

The first term tends to zero as $n \rightarrow \infty$ by Step 2, since $B_m f$ is a polynomial. Letting $n \rightarrow \infty$ and then $m \rightarrow \infty$ proves (7.55). If φ is continuous with compact support in $[0, \infty)$, then $f(x) = x^{-2}\varphi(x^{-1} - 1)$ for $x \in (0, 1]$ and $f(0) = 0$ defines a continuous function on $[0, 1]$ that vanishes near zero, with $\int f d\lambda_n = \int \varphi d\mu_n$, and vague convergence follows from (7.55).

Step 4: varying tests. Define $f_n(x) = x^{-2}\varphi_n(x^{-1} - 1)$ and $f(x) = x^{-2}\varphi(x^{-1} - 1)$ for $x \in (0, 1]$, and $f_n(0) = f(0) = 0$. With $y = x^{-1} - 1$, we have $|f(x)| = (1+y)^2|\varphi(y)|$. The second condition in (7.56) then shows that f is continuous at zero, and similarly for each f_n . Also $\int \varphi_n d\mu_n = \int f_n d\lambda_n$ and $\int \varphi d\mu = \int f d\lambda$. Let $R > 0$. For $x \geq (1+R)^{-1}$, that is $y \leq R$, we have $|f_n(x) - f(x)| \leq (1+R)^2 \sup_{[0,R]} |\varphi_n - \varphi|$. For $x < (1+R)^{-1}$, we have $|f_n(x) - f(x)| \leq \sup_{y \geq R} (1+y)^2(|\varphi_n(y)| + |\varphi(y)|)$. Choosing first R large and then n large, we find $\|f_n - f\|_\infty \rightarrow 0$. Then

$$\left| \int f_n d\lambda_n - \int f d\lambda \right| \leq \|f_n - f\|_\infty T_n(1) + \left| \int f d\lambda_n - \int f d\lambda \right| \rightarrow 0,$$

by Step 3. This proves the second part.

Step 5: families of tests. Define $f_{n,\theta}$ and f_θ as in Step 4. As in Step 4, we have $\sup_\theta \|f_{n,\theta} - f_\theta\|_\infty \rightarrow 0$. We claim that $\{f_\theta : \theta \in \Theta\}$ is uniformly bounded and uniformly equicontinuous on $[0, 1]$. Let $\tau(R) = \sup_\theta \sup_{y \geq R} (1+y)^2|\varphi_\theta(y)|$, which tends to

zero. On $[0, (1+R)^{-1}]$ we have $|f_\theta| \leq \tau(R)$. On $[(1+R)^{-1}, 1]$, the map $x \mapsto x^{-1} - 1$ is Lipschitz with constant $(1+R)^2$, and $x \mapsto x^{-2}$ is bounded by $(1+R)^2$ and Lipschitz with constant $2(1+R)^3$. Therefore the uniform bound and the uniform modulus of continuity of (φ_θ) on $[0, R]$ yield a uniform bound and a uniform modulus of continuity of (f_θ) there. Given $\varepsilon > 0$, we choose R with $2\tau(R) \leq \varepsilon$; two points of $[0, 1]$ at distance at most r that are not both in $[(1+R)^{-1}, 1]$ can be compared through the point $(1+R)^{-1}$, where $|f_\theta| \leq \tau(R)$. This proves the claim. By the Arzelà–Ascoli theorem, for every $\varepsilon > 0$ there exist finitely many continuous functions $g_1, \dots, g_L$ on $[0, 1]$ such that every f_θ satisfies $\|f_\theta - g_l\|_\infty \leq \varepsilon$ for some l . Then

$$\sup_\theta \left| \int f_\theta d\lambda_n - \int f_\theta d\lambda \right| \leq \max_{l \leq L} \left| \int g_l d\lambda_n - \int g_l d\lambda \right| + \varepsilon(T_n(1) + T(1)),$$

and the maximum tends to zero by Step 3. Since ε is arbitrary and $\sup_\theta \|f_{n,\theta} - f_\theta\|_\infty T_n(1) \rightarrow 0$, the third part follows.

Step 6: random measures. Under the hypotheses of the fourth part, $T_n(1)$ converges in probability to the almost surely finite $T(1)$. The total masses are therefore tight, that is, for every $\varepsilon > 0$ there exists M_ε with $\mathbb{P}(T_n(1) + T(1) > M_\varepsilon) \leq \varepsilon$ for all large n . In Step 2, at fixed k and δ , $\int D_{k,\delta} d\lambda_n - \int D_{k,\delta} d\lambda$ tends to zero in probability. Given $\varepsilon > 0$, we choose δ with $k(k+2)\delta M_\varepsilon \leq \varepsilon$ and then n large; this shows that every moment difference tends to zero in probability. In Step 3, for deterministic f , we choose m with $\|B_m f - f\|_\infty M_\varepsilon \leq \varepsilon$. For a random f that is a measurable random element of $C([0, 1])$, $\int B_m f d\lambda_n - \int B_m f d\lambda$ is a sum of $m+1$ products of the almost surely finite random variables $f(j/m)$ with moment differences of $x^j(1-x)^{m-j}$, which tend to zero in probability. Hence it tends to zero in probability, and $\|B_m f - f\|_\infty \rightarrow 0$ almost surely as $m \rightarrow \infty$. This proves the first part in probability.

For random tests, the integrals in the second and third parts are almost surely defined by the integrability assumption, and the measurability assumptions make the events below measurable. Since the limit tests φ_θ do not depend on n , the event on which their weighted tails are bounded by τ_ε , and in the third part also the bound B_ε and the moduli $\omega_{\varepsilon,R}$ hold, contains $\mathcal{E}_n^\varepsilon$ for every n , and so has probability at least $1 - \varepsilon$. On this event, the transformed limit tests f_θ are continuous at zero, and in the third part they lie in a fixed deterministic set $\mathcal{K}_\varepsilon$ of functions on $[0, 1]$ with a common bound and a common modulus of continuity, as in Step 5. This set is compact in the uniform topology by the Arzelà–Ascoli theorem, so the functions $g_1, \dots, g_L$ of Step 5 can be chosen deterministic. On $\mathcal{E}_n^\varepsilon$, as in Step 4,

$$\sup_\theta \|f_{n,\theta} - f_\theta\|_\infty \leq (1+R)^2 \sup_\theta \sup_{0 \leq y \leq R} |\varphi_{n,\theta}(y) - \varphi_\theta(y)| + \tau_\varepsilon(R)$$

for every $R \geq 1$. We choose R with $\tau_\varepsilon(R)$ small and then let $n \rightarrow \infty$; since the complements of the events above have probability at most ε asymptotically, $\sup_\theta \|f_{n,\theta} - f_\theta\|_\infty \rightarrow 0$

in probability. The first part, in probability, applied to $g_1, \dots, g_L$ and to f , together with the tightness of $T_n(1) + T(1)$, completes the proof as in Steps 4 and 5. $\square$

Functional inequalities on the sphere

In this section, we prove the functional inequalities used in the rest of the paper. They concern the Langevin generator of a quadratic potential on S_N and its Gibbs measure, which is a Bingham distribution. We note that all statements of this section are deterministic. They hold for a fixed matrix, or for a fixed sequence of precisions, that satisfies explicit hypotheses. In the applications, these hypotheses are verified on events of the disorder, following the convention of Section 3.4.

In Section 8.1, we fix the conventions and collect standard tools. In Section 8.2, we write the Bochner formula for the generator in ambient coordinates (Lemma 8.3). With a weight built from the equilibrium covariance, we then prove a weighted gradient estimate for the semigroup and a lower bound on the spectral gap (Lemma 8.5 and Theorem 8.6). These results are used in Sections 9 and 10. Section 8.3 contains a lower bound on the spectral gap that is sharp when many precisions are comparable to the smallest one (Theorem 8.8), and the Poincaré inequality that follows from it (Corollary 8.9). The first is used in Section 9, and the second in Sections 12, 17, and 19. In Section 8.4, we extend functions from the sphere to Gaussian space and obtain logarithmic Sobolev inequalities with an additive error term (Theorem 8.11). We also bound the density of a weighted sum of squared Gaussian variables (Lemma 8.12). Together with a Poincaré inequality, these results give logarithmic Sobolev inequalities (Corollary 8.13). Section 8.5 treats the case where only a few precisions are close to the smallest one (Lemmas 8.18 and 8.19), which is the situation in the critical window. In Section 8.6, we combine these results. We bound the Poincaré and logarithmic Sobolev constants of the Gibbs measure by $CN^{2/3}$ in the critical window, and those of the Bingham comparison law at time $t \leq KN^{2/3}$ by $C_K(1+t)$ (Proposition 8.21). This is used in Sections 12 and 19.

We use the definitions of Section 3, the conditioned Gaussian laws μ_v of Definition 5.4, the hypotheses (4.36) on the critical half-precisions, and the heat estimates of Lemma 7.14. The analytic inputs are classical; we recall them in Lemma 8.2. They are the Bochner formula and the Γ_2 calculus of Bakry and Émery [11, 12], the Gaussian logarithmic Sobolev inequality [79], Rothaus's lemma [125], the Brascamp–Lieb inequality [36] in its version on convex domains [96], and the perturbation argument of Holley and Stroock [84, Lemma 5.1, pp. 1183–1184].

8.1 Conventions and standard tools

Dirichlet forms. Throughout this section, $N \geq 2$. For a function f defined near S_N , we write $\nabla_S f$ for its gradient along S_N and set $\Gamma(f, g) = \langle \nabla_S f, \nabla_S g \rangle$ and $\Gamma(f) = \Gamma(f, f) = |\nabla_S f|^2$. For a probability measure μ on S_N with a smooth positive density with respect to ν_N , we write

$$\mathcal{E}_\mu(f) = \int \Gamma(f) d\mu$$

for the spherical Dirichlet form, defined on the closure $H^1(S_N)$ of the smooth functions in the norm $(\mu(f^2) + \mathcal{E}_\mu(f))^{1/2}$, and $C_P(\mu)$ and $C_{LS}(\mu)$ for the smallest constants such that the Poincaré and logarithmic Sobolev inequalities (3.5) hold. Recall from Section 3 that every such inequality in this paper refers to this spherical form, which is the Dirichlet form of the Langevin generator with noise coefficient $\sqrt{2}$. Ambient Euclidean and Gaussian forms appear only inside proofs, where they are named. The inequalities (3.5) for smooth f extend to $H^1(S_N)$: if $f_n \rightarrow f$ in $H^1(S_N)$ and almost everywhere, then $\text{Var}_\mu(f_n) \rightarrow \text{Var}_\mu(f)$ and $\mathcal{E}_\mu(f_n) \rightarrow \mathcal{E}_\mu(f)$, and Fatou's lemma, applied to $f_n^2 \log f_n^2 + e^{-1} \geq 0$, gives $\text{Ent}_\mu(f^2) \leq \liminf_n \text{Ent}_\mu(f_n^2)$. We use this density argument without further comment.

Quadratic potentials. Let A be a real symmetric $N \times N$ matrix with eigenvalues $\alpha_1 \geq \dots \geq \alpha_N$, and recall the generator L_A , the semigroup P_t^A and the Gibbs measure π_A from (3.3). In the notation of this section, the integration by parts formula (3.4) reads

$$\int f(-L_A g) d\pi_A = \int \Gamma(f, g) d\pi_A \quad (8.1)$$

for smooth f and g . By elliptic spectral theory on the compact connected weighted manifold S_N [130, Section 2, p. 30], the closure of $-L_A$ is self-adjoint and nonnegative on $L^2(\pi_A)$, its spectrum is discrete, its eigenfunctions are smooth, and its kernel consists of the constant functions. In particular, the spectral gap $\lambda_{\text{gap}}(L_A)$ of Section 3 is the smallest $\lambda > 0$ for which there exists a nonconstant smooth f with $L_A f = -\lambda f$, and $C_P(\pi_A) = 1/\lambda_{\text{gap}}(L_A)$.

We recall from Definition 3.3 the Lagrange multiplier $z > \alpha_1$, the precisions $p_i = z - \alpha_i$, the variances $q_i = 1/p_i$, which satisfy $\sum_i q_i = N$, the smallest and largest precisions $d = p_1$ and $s = p_N$, the ratio $\kappa = s/d$, and the half-precisions $r_i = p_i/2$. We also write $Q = (zI - A)^{-1}$. Unless stated otherwise, we work in an orthonormal eigenbasis of A , so that $A = \text{diag}(\alpha_1, \dots, \alpha_N)$, $Q = \text{diag}(q_1, \dots, q_N)$, and $x_1, \dots, x_N$ are the eigencoordinates. Recall from Proposition 5.6 that π_A is the conditioned Gaussian law μ_q of Definition 5.4; on S_N , it is proportional to $\exp(-\frac{1}{2} \sum_i p_i x_i^2) \nu_N = \exp(-\sum_i r_i x_i^2) \nu_N$. By (3.7), adding a multiple of the identity to A changes neither L_A nor π_A . It shifts z by the same amount and leaves the precisions and variances unchanged. In particular, $L_A = L_{-P}$ and $\pi_A = \pi_{-P}$ for $P = \text{diag}(p_1, \dots, p_N)$. Conversely, for every $q \in (0, \infty)^N$ with $\sum_i q_i = N$, the law μ_q equals π_A for $A = -\text{diag}(q_1^{-1}, \dots, q_N^{-1})$, whose Lagrange multiplier is $z = 0$.

and whose variances are $q_1, \dots, q_N$. We use this identification without further comment: a statement about π_A in terms of its precisions is a statement about every μ_q . As in Definition 3.3, the variances q_i are the parameters of the Gaussian vector before conditioning; they need not equal the second moments of π_A .

Reflections. For $1 \leq j \leq N$, let R_j be the reflection that changes the sign of x_j . When A is diagonal, R_j preserves ν_N , π_A , L_A and Γ (see Lemma 3.1). Let μ be a probability measure on S_N with a smooth positive density that is invariant under R_j . We write $H_j^+ = \{x \in S_N : x_j \geq 0\}$ and $\mu_j^+ = \mu(\cdot \mid H_j^+)$, and we equip μ_j^+ with the reflecting form

$$\mathcal{E}_{\mu_j^+}(u) = \int_{H_j^+} \Gamma(u) d\mu_j^+, \quad u \in H^1(H_j^+).$$

No boundary condition is imposed on u . When μ is invariant under every R_i , we similarly equip the orthant $O = \{x \in S_N : x_i \geq 0 \text{ for all } i\}$ with $\mu_O = \mu(\cdot \mid O)$ and the reflecting form. A function invariant under every R_i , that is, a function in the coordinate-even sector of Definition 3.2, is determined by its restriction to O . We call $C_P(\mu_O)$ and $C_{LS}(\mu_O)$ the constants of the coordinate-even sector. When $j = 1$ we write $\mu_+ = \mu_1^+$.

Lemma 8.1 (Reflecting hemispheres and the coordinate-even sector). *Let μ be invariant under R_j , and let $\alpha, \beta \geq 0$. Suppose that*

$$\text{Ent}_\mu(f^2) \leq \alpha \mathcal{E}_\mu(f) + \beta \mu(f^2) \quad \text{for all } f \in H^1(S_N).$$

Then the same inequality holds for μ_j^+ and every $u \in H^1(H_j^+)$. Also, $C_P(\mu_j^+) \leq C_P(\mu)$ and $C_{LS}(\mu_j^+) \leq C_{LS}(\mu)$. If μ is invariant under every R_i , the same conclusions hold for μ_O .

Proof. Let $\phi_j(x)$ be x with x_j replaced by $|x_j|$. For $u \in H^1(H_j^+)$, the even extension $\tilde{u} = u \circ \phi_j$ belongs to $H^1(S_N)$, and $\Gamma(\tilde{u}) = \Gamma(u) \circ \phi_j$ almost everywhere, because R_j is an isometry of S_N whose fixed set is the smooth hypersurface $\{x_j = 0\}$. We have $\int g \circ \phi_j d\mu = \int g d\mu_j^+$ for every measurable $g \geq 0$ on H_j^+ , because μ is R_j -invariant. Applying this to $g = u^2$, $u^2 \log u^2$ and $\Gamma(u)$, we find that the variance, the entropy and the energy of $\tilde{u}$ under μ coincide with those of u under μ_j^+ . This proves the claims for μ_j^+ . For μ_O , use $\phi(x) = (|x_1|, \dots, |x_N|)$. $\square$

Lemma 8.2 (Standard tools). *The following hold.*

- (i) (Bounded perturbation.) *Let μ and $\mu' = \rho\mu$ be probability measures on a finite-dimensional Euclidean space carrying an energy $\int \Gamma(u) d\mu$ with the same pointwise energy density $\Gamma(u) \geq 0$ for both measures. If $\sup \rho / \inf \rho \leq e^\omega$, then $C_P(\mu') \leq e^\omega C_P(\mu)$.*
- (ii) (Rothaus's lemma.) *For every probability measure μ on S_N with a smooth strictly positive density relative to ν_N , every smooth real-valued f on S_N , and every $a \in \mathbb{R}$, we have*

$\text{Ent}_\mu(f^2) \leq \text{Ent}_\mu((f - a)^2) + 2\mu((f - a)^2)$. If $\text{Ent}_\mu(f^2) \leq \alpha \mathcal{E}_\mu(f) + \beta \mu(f^2)$ for all smooth f , with $\alpha, \beta \geq 0$, then $C_{\text{LS}}(\mu) \leq \alpha + (\beta + 2)C_{\text{P}}(\mu)$.

(iii) (Positive test functions.) For every smooth positive w and every smooth u on S_N ,

$$\int u^2 \frac{-L_A w}{w} d\pi_A \leq \int \Gamma(u) d\pi_A.$$

(iv) (Gaussian logarithmic Sobolev inequality.) If γ is the standard Gaussian measure on $\mathbb{R}^m$, then $\text{Ent}_\gamma(F^2) \leq 2 \int |\nabla F|^2 d\gamma$ for every smooth compactly supported F .

(v) (Brascamp–Lieb inequality on a convex domain.) Let $m \geq 1$, let $\Omega \subset \mathbb{R}^m$ be a nonempty bounded convex open set with smooth boundary, let Ω' be an open set containing $\bar{\Omega}$, and let $U \in C^{1,1}(\Omega')$. Suppose that $\nabla^2 U \succeq \Lambda$ almost everywhere on Ω' , where Λ is a continuous field of positive definite matrices on Ω' . If μ is the probability measure on Ω proportional to $e^{-U} dy$, then

$$\text{Var}_\mu(u) \leq \int \langle \Lambda^{-1} \nabla u, \nabla u \rangle d\mu \quad \text{for every } u \in C^1(\mathbb{R}^m).$$

(vi) (Spheres.) The uniform probability measure on the unit sphere $S^{k-1} \subset \mathbb{R}^k$, $k \geq 2$, with its spherical gradient, has Poincaré constant $1/(k-1)$, and so does the uniform measure on the unit hemisphere $\{\omega \in S^{k-1} : \omega_1 \geq 0\}$ with its reflecting form. In particular, $C_{\text{P}}(\nu_N) = N/(N-1)$.

Proof. For (i), we have

$$\begin{aligned} \text{Var}_{\mu'}(f) &\leq \int (f - \mu f)^2 d\mu' \leq (\sup \rho) \text{Var}_\mu(f) \\ &\leq (\sup \rho) C_{\text{P}}(\mu) \int \Gamma(f) d\mu \leq \frac{\sup \rho}{\inf \rho} C_{\text{P}}(\mu) \int \Gamma(f) d\mu'. \end{aligned}$$

This is the Poincaré version of the perturbation argument of Holley and Stroock [84, Lemma 5.1, pp. 1183–1184]. The inequality in (ii) is due to Rothaus [125]; see also [18, Lemma 4]. Applying it with $a = \mu(f)$, and then the hypothesis to $f - \mu(f)$, we obtain

$$\text{Ent}_\mu(f^2) \leq \alpha \mathcal{E}_\mu(f) + (\beta + 2) \text{Var}_\mu(f) \leq (\alpha + (\beta + 2)C_{\text{P}}(\mu)) \mathcal{E}_\mu(f).$$

For a reflecting hemisphere or orthant obtained from a reflection-invariant smooth positive spherical law, the same Rothaus inequality follows by even extension and approximation in H^1 , as in Lemma 8.1. Consequently the tightening implication also holds for each of these reflecting forms, using its own Poincaré constant. For (iii), (8.1) with $f = u^2/w$ and $g = w$ gives

$$\begin{aligned} \int u^2 \frac{-L_A w}{w} d\pi_A &= \int \Gamma\left(\frac{u^2}{w}, w\right) d\pi_A \\ &= \int (2u \Gamma(u, \log w) - u^2 \Gamma(\log w)) d\pi_A \leq \int \Gamma(u) d\pi_A, \end{aligned}$$

where in the last step we used $|\nabla_S u - u \nabla_S \log w|^2 \geq 0$. This elementary inequality is the basis of the Lyapunov-function method; see [12, Section 4.6.1, (4.6.4)–(4.6.5), p. 202]. For the marginal generators below, the diffusion matrices are symmetric and positive definite on the open chart. The same proof then applies once the integration by parts produces no boundary term; we check this in each use. Item (iv) is Gross's inequality [79, Corollary 4.2, (3.4), p. 1073], multiplied by 2.

For (v), we first suppose that U is smooth on $\bar{\Omega}$ and that $\nabla^2 U$ is positive definite there. Then the claim with $\Lambda = \nabla^2 U$ is [96, Theorem 1.2(1)], applied with $1/N = 0$ to the weighted manifold $\bar{\Omega}$ with the Euclidean metric and the measure $e^{-U} dy$. Its boundary is locally convex because Ω is convex, and its generalized Ricci tensor is $\nabla^2 U$. For $\Omega = \mathbb{R}^m$ this is the inequality of Brascamp and Lieb [36]. In general, let $U_\varepsilon = U * \phi_\varepsilon$ for a standard mollifier ϕ_ε , defined on $\bar{\Omega}$ for small ε . The gradient ∇U is locally Lipschitz, and $\nabla^2 U_\varepsilon = (\nabla^2 U) * \phi_\varepsilon$. Fix $\theta \in (0, 1)$. There exists $\varepsilon_0 > 0$ such that $\Lambda(y - w) \succeq (1 - \theta)\Lambda(y)$ for $y \in \bar{\Omega}$ and $|w| \leq \varepsilon_0$, because Λ is continuous and positive definite on a compact neighborhood of $\bar{\Omega}$. Then $\nabla^2 U_\varepsilon \succeq (1 - \theta)\Lambda$ on $\bar{\Omega}$ for $\varepsilon \leq \varepsilon_0$, and the smooth case gives the claim for e^{-U_ε} with the factor $(1 - \theta)^{-1}$. Finally, $U_\varepsilon \rightarrow U$ uniformly on $\bar{\Omega}$, and we may let $\varepsilon \rightarrow 0$ and then $\theta \rightarrow 0$.

For (vi), the eigenvalues of the Laplacian of S^{k-1} are $\ell(\ell + k - 2)$, $\ell \geq 0$, and the first positive one, $k - 1$, has the linear functions as eigenfunctions. For $k \geq 3$, this value also attains the Lichnerowicz bound [12, Section 4.8, (4.8.1), p. 213]. The reflecting form of the hemisphere corresponds to the functions on S^{k-1} that are even in ω_1 , as in Lemma 8.1, and ω_2 is such an eigenfunction. Scaling by $\sqrt{N}$ gives $C_P(v_N) = N/(N - 1)$. $\square$

8.2 Ambient calculus for the quadratic generator

The results of this section use only Euclidean derivatives, matrix algebra and (8.1). We fix a real symmetric matrix A , which we do not assume to be diagonal. For $x \in S_N$, let

$$\Pi_x = I - \frac{xx^\top}{N}, \quad e_A(x) = \frac{x^\top Ax}{N}, \quad b_A(x) = Ax - \left(e_A(x) + \frac{N-1}{N}\right)x.$$

For a smooth function u defined near S_N , we write Du and D^2u for its Euclidean gradient and Hessian. Then

$$L_A u = \text{tr}(\Pi_x D^2 u) + b_A(x) \cdot Du, \quad \Gamma(u, v) = (Du)^\top \Pi_x Dv \quad (8.2)$$

on S_N . The second formula holds because $\Pi_x Du$ is the tangential part of Du . For the first, we note that its right side depends only on the restriction of u to S_N . Indeed, suppose $u = 0$ on S_N . Then $\Pi_x Du = 0$. For a unit vector e orthogonal to x , the curve $\gamma_e(r) = (x + re)/(1 + r^2/N)^{1/2}$ stays on S_N and satisfies $\gamma'_e(0) = e$ and $\gamma''_e(0) = -x/N$. Differentiating $u(\gamma_e(r)) = 0$ twice at $r = 0$ gives $e^\top D^2 u e = x \cdot Du/N$, and summing

over an orthonormal basis of $x^\perp$, we find $\text{tr}(\Pi_x D^2 u) = (N-1)x \cdot Du/N$. Also, Du is parallel to x , which gives $b_A(x) \cdot Du = (x \cdot Du)(x^\top A x - N e_A(x) - (N-1))/N = -(N-1)x \cdot Du/N$. The two terms cancel, and the right side vanishes. It suffices to check (8.2) for an extension that is constant along rays. For such u , the identities $x \cdot Du = 0$ and $x^\top D^2 u x = 0$ imply that $\text{tr}(\Pi_x D^2 u) = \Delta u = \Delta_S u$, and $b_A(x) \cdot Du = (Ax)^\top \Pi_x Du$ is the drift term of L_A .

For a smooth positive function w on S_N , we define

$$\Gamma_2(f) = \frac{1}{2}(L_A \Gamma(f) - 2\Gamma(f, L_A f)), \quad \Gamma_{w,2}(f) = \frac{1}{2}(L_A(w\Gamma(f)) - 2w\Gamma(f, L_A f)).$$

Lemma 8.3 (Bochner formula in ambient coordinates). *Let f be smooth on S_N , and extend it to $\{\sqrt{N}/2 < |x| < 2\sqrt{N}\}$ by $f(x) = f(\sqrt{N}x/|x|)$. On S_N , let*

$$g = Df, \quad E = D^2 f, \quad \nabla_S^2 f = \Pi_x E \Pi_x.$$

Then $x \cdot g = 0$, $\Gamma(f) = |g|^2$, and

$$Ex = -g, \quad E = \nabla_S^2 f - \frac{gx^\top + xg^\top}{N}, \quad (8.3)$$

$$\|E\|_{\text{HS}}^2 = \|\nabla_S^2 f\|_{\text{HS}}^2 + \frac{2|g|^2}{N}, \quad \text{tr}(\Pi_x E^2) = \|\nabla_S^2 f\|_{\text{HS}}^2 + \frac{|g|^2}{N}. \quad (8.4)$$

Moreover,

$$\Gamma_2(f) = \|\nabla_S^2 f\|_{\text{HS}}^2 + g^\top \left[\left(e_A + 1 - \frac{2}{N} \right) I - A \right] g, \quad (8.5)$$

and for every smooth positive w on S_N , with $a = \nabla_S \log w$,

$$\Gamma_{w,2}(f) = w \|\nabla_S^2 f\|_{\text{HS}}^2 + w g^\top \left[\left(e_A + 1 - \frac{2}{N} + \frac{L_A w}{2w} \right) I - A \right] g + 2w \nabla_S^2 f(a, g). \quad (8.6)$$

Here $\nabla_S^2 f(a, g) = a^\top (\nabla_S^2 f) g$. The matrix $\nabla_S^2 f$ represents the Hessian of f on S_N , since the extension is constant along rays, and (8.5) is Bochner's formula for L_A , written in ambient coordinates; compare [40, equations (1.38)–(1.41)], with potential $V = -\frac{1}{2}x^\top A x$. For $A = 0$, the identity (8.5) implies that $\Gamma_2(f) \geq (1 - 2/N)\Gamma(f)$, the curvature bound $\text{Ric} = (N-2)/N$ of S_N . The semigroup interpolation argument based on Γ_2 is the method of Bakry and Émery [11]; see [12, Corollary 3.3.19, p. 163]. The weighted expression used here is computed directly in (8.6) and specialized to the covariance weight in Lemma 8.5.

Proof. Since the extension is constant along rays, Euler's identity implies that $x \cdot Df = 0$. Differentiating it once, we obtain $g + Ex = 0$, and multiplying by $x^\top$ yields $x^\top Ex = 0$. We decompose E with the orthogonal projections Π_x and $xx^\top/N$. Using $Ex = -g$ and $\Pi_x g = g$, we find $\Pi_x E(xx^\top/N) = -gx^\top/N$. Its transpose is $(xx^\top/N)E\Pi_x = -xg^\top/N$, and $(xx^\top/N)E(xx^\top/N) = 0$. This proves (8.3). The three matrices $\nabla_S^2 f$, $gx^\top/N$ and $xg^\top/N$ are orthogonal for the Hilbert–Schmidt inner product, since $(\nabla_S^2 f)x = 0$ and

$x \cdot g = 0$, and the last two have squared norm $|g|^2|x|^2/N^2 = |g|^2/N$. This proves the first identity in (8.4). For the second, we compute $\Pi_x E = \nabla_S^2 f - gx^\top/N$ and

$$\Pi_x E^2 = \left(\nabla_S^2 f - \frac{gx^\top}{N} \right) \left(\nabla_S^2 f - \frac{gx^\top + xg^\top}{N} \right) = (\nabla_S^2 f)^2 - \frac{(\nabla_S^2 f)g x^\top}{N} + \frac{gg^\top}{N},$$

whose trace is $\|\nabla_S^2 f\|_{\text{HS}}^2 + |g|^2/N$.

Let $G = |Df|^2$ on the annulus. The function G is homogeneous of degree -2 , because the gradient of a function that is homogeneous of degree zero is homogeneous of degree -1 . By Euler's identity, we have $x \cdot DG = -2G$ and $x^\top D^2 G x = 6G$. Also $DG = 2Eg$ and $D^2 G = 2E^2 + 2\sum_j g_j D^2(\partial_j f)$; taking the trace, we find $\text{tr}(D^2 G) = 2\|E\|_{\text{HS}}^2 + 2g \cdot D\Delta f$. By (8.2), using $x^\top E g = -|g|^2$, we find

$$L_A G = 2\|E\|_{\text{HS}}^2 + 2g \cdot D\Delta f - \frac{6G}{N} + 2g^\top E A x + 2\left(e_A + \frac{N-1}{N}\right)G.$$

By (8.2), the function $\Delta f + Ax \cdot Df$ is an extension of $L_A f$, since $x \cdot Df = 0$ and $x^\top E x = 0$. Its gradient is $D\Delta f + Ag + EAx$, and taking the inner product with g , we find

$$2\Gamma(f, L_A f) = 2g \cdot D\Delta f + 2g^\top Ag + 2g^\top E A x.$$

Subtracting and using the first identity in (8.4), we have

$$\Gamma_2(f) = \|E\|_{\text{HS}}^2 + \left(e_A + 1 - \frac{4}{N}\right)|g|^2 - g^\top Ag,$$

which is (8.5). Finally, by the product rule $L_A(wG) = wL_A G + GL_A w + 2\Gamma(w, G)$, we have

$$\Gamma_{w,2}(f) = w\Gamma_2(f) + \frac{1}{2}G L_A w + \Gamma(w, G).$$

The tangential part of $DG = 2Eg$ is $2\Pi_x E g = 2(\nabla_S^2 f)g$ by (8.3). Since $a = \nabla_S \log w$, we have $\Gamma(w, G) = 2w \nabla_S^2 f(a, g)$, which proves (8.6). $\square$

Corollary 8.4 (Gradient of the norm of the gradient). *In the notation of Lemma 8.3, for every $\varepsilon > 0$ the function $u_\varepsilon = (\Gamma(f) + \varepsilon^2)^{1/2}$ satisfies*

$$\Gamma(u_\varepsilon) = \frac{|(\nabla_S^2 f)g|^2}{|g|^2 + \varepsilon^2} \leq \|\nabla_S^2 f\|_{\text{HS}}^2.$$

Proof. As shown at the end of the previous proof, the tangential gradient of $\Gamma(f) = G$ is $2(\nabla_S^2 f)g$. Using the chain rule, we find $\nabla_S u_\varepsilon = (\nabla_S^2 f)g/u_\varepsilon$, and the claim follows from $|(\nabla_S^2 f)g| \leq \|\nabla_S^2 f\|_{\text{HS}}|g|$. $\square$

We now use the equilibrium covariance $Q = (zI - A)^{-1}$ as a weight. Let

$$\varphi(x) = \frac{x^\top Q x}{N}, \quad x \in S_N.$$

Note that $1/s \leq \varphi \leq 1/d$ on S_N , because the eigenvalues of Q lie in $[1/s, 1/d]$.

Lemma 8.5 (The weight φ and a weighted curvature bound). *Let $a = \nabla_S \log \varphi$ and $J = (\kappa + \kappa^{-1} - 2)/N$. Then*

$$L_A \varphi = 2(z - 1 - e_A) \varphi, \quad |a|^2 \leq J, \quad (8.7)$$

and, in the notation of Lemma 8.3,

$$\Gamma_{\varphi,2}(f) = \varphi \left\{ \|\nabla_S^2 f\|_{\text{HS}}^2 + 2\nabla_S^2 f(a, g) + g^\top \left[(zI - A) - \frac{2}{N}I \right] g \right\}. \quad (8.8)$$

In particular, with $r_0 = d - (\kappa + \kappa^{-1})/N$,

$$\Gamma_{\varphi,2}(f) \geq r_0 \varphi \Gamma(f). \quad (8.9)$$

The identity (8.8) and the bound (8.9) hold for every symmetric A , also when $r_0 \leq 0$.

Proof. By (8.2) with $D\varphi = 2Qx/N$ and $D^2\varphi = 2Q/N$, and using $\text{tr } Q = N$, we have

$$\begin{aligned} L_A \varphi &= \frac{2}{N} \left(\text{tr } Q - \frac{x^\top Q x}{N} \right) + \frac{2}{N} \left(x^\top A Q x - \left(e_A + \frac{N-1}{N} \right) x^\top Q x \right) \\ &= 2 - 2\varphi + \frac{2}{N} x^\top A Q x - 2e_A \varphi. \end{aligned}$$

Recall that $AQ = zQ - I$. Then $x^\top A Q x = Nz\varphi - N$, and the first identity in (8.7) follows. For the second, we work in an eigenbasis of A and write $\omega_i = x_i^2/N$, so that $\omega_i \geq 0$ and $\sum_i \omega_i = 1$. We have

$$|a|^2 = \frac{4}{N^2 \varphi^2} \left(|Qx|^2 - N\varphi^2 \right) = \frac{4}{N} \left(\frac{\sum_i \omega_i q_i^2}{(\sum_i \omega_i q_i)^2} - 1 \right),$$

where we used that the tangential part of $D\varphi$ is $2Qx/N - 2\varphi x/N$. Let $\ell = 1/s$ and $\bar{u} = 1/d$. Then $q_i \in [\ell, \bar{u}]$, and therefore $q_i^2 \leq (\ell + \bar{u})q_i - \ell\bar{u}$. With $m = \sum_i \omega_i q_i \in [\ell, \bar{u}]$, the ratio above is at most $((\ell + \bar{u})m - \ell\bar{u})/m^2$. As a function of $m > 0$, this is maximized at $m = 2\ell\bar{u}/(\ell + \bar{u})$, where it equals $(\ell + \bar{u})^2/(4\ell\bar{u}) = (\kappa + 2 + \kappa^{-1})/4$. We conclude that $|a|^2 \leq (\kappa + \kappa^{-1} - 2)/N = J$. This is the elementary form of the Kantorovich inequality [93, pp. 142–143]. For the equivalent second-moment form and its earlier history, including the work of Frucht, see [112, Section 1, (1.3)].

Inserting $L_A \varphi / (2\varphi) = z - 1 - e_A$ into (8.6), we find that the function e_A cancels, which proves (8.8). For (8.9), let $T = (ag^\top + ga^\top)/2$. Then $\langle \nabla_S^2 f, T \rangle_{\text{HS}} = \nabla_S^2 f(a, g)$ and $\|T\|_{\text{HS}}^2 = (|a|^2|g|^2 + (a \cdot g)^2)/2 \leq |a|^2|g|^2$. Completing the square, we find

$$\|\nabla_S^2 f\|_{\text{HS}}^2 + 2\nabla_S^2 f(a, g) = \|\nabla_S^2 f + T\|_{\text{HS}}^2 - \|T\|_{\text{HS}}^2 \geq -|a|^2|g|^2 \geq -J|g|^2.$$

The eigenvalues of $zI - A$ are at least d . Inserting these two bounds into (8.8), we obtain $\Gamma_{\varphi,2}(f) \geq \varphi(d - 2/N - J)|g|^2 = r_0 \varphi \Gamma(f)$. $\square$

Theorem 8.6 (Weighted gradient estimate and spectral gap). *Let $r_0 = d - (\kappa + \kappa^{-1})/N$. For every $t \geq 0$ and every smooth f on S_N ,*

$$\varphi \Gamma(P_t^A f) \leq e^{-2r_0 t} P_t^A (\varphi \Gamma(f)), \quad \Gamma(P_t^A f) \leq \kappa e^{-2r_0 t} P_t^A \Gamma(f). \quad (8.10)$$

If $r_0 > 0$, then $\lambda_{\text{gap}}(L_A) \geq r_0$. These conclusions remain true when r_0 is replaced by any smaller number, for instance by $r = d - (\kappa + 1 + \kappa^{-1})/N = r_0 - 1/N$.

Proof. Fix $t > 0$. The function $(u, x) \mapsto P_u^A f(x)$ is smooth on $[0, t] \times S_N$, by parabolic regularity on the compact manifold S_N . For $0 \leq u \leq t$, let

$$\mathcal{K}(u) = e^{-2r_0 u} P_u^A (\varphi \Gamma(P_{t-u}^A f)).$$

Using $\partial_u P_{t-u}^A f = -L_A P_{t-u}^A f$ and the definition of $\Gamma_{\varphi, 2}$, we find

$$\begin{aligned} \mathcal{K}'(u) &= e^{-2r_0 u} P_u^A \left(L_A (\varphi \Gamma(P_{t-u}^A f)) - 2\varphi \Gamma(P_{t-u}^A f, L_A P_{t-u}^A f) - 2r_0 \varphi \Gamma(P_{t-u}^A f) \right) \\ &= 2e^{-2r_0 u} P_u^A \left(\Gamma_{\varphi, 2}(P_{t-u}^A f) - r_0 \varphi \Gamma(P_{t-u}^A f) \right), \end{aligned}$$

which is nonnegative by (8.9) and the positivity of P_u^A . Comparing $\mathcal{K}(0)$ and $\mathcal{K}(t)$ proves the first inequality in (8.10). The second follows from $1/s \leq \varphi \leq 1/d$.

Let f be a nonconstant smooth eigenfunction, $L_A f = -\lambda f$. By (8.1), we have $\int L_A(\varphi \Gamma(f)) d\pi_A = 0$. Using also $\Gamma(f, L_A f) = -\lambda \Gamma(f)$, we find

$$\int \Gamma_{\varphi, 2}(f) d\pi_A = \lambda \int \varphi \Gamma(f) d\pi_A.$$

The last integral is positive. Integrating (8.9), we find $\lambda \geq r_0$. Since this holds for every nonconstant eigenfunction, we conclude that $\lambda_{\text{gap}}(L_A) \geq r_0$. The last assertion is immediate, since $e^{-2r_0 t} \leq e^{-2rt}$ and $r \leq r_0$. $\square$

The rate r_0 is positive only when $d > (\kappa + \kappa^{-1})/N$. Near the critical temperature, κ is of order d^{-1} , and this condition requires Nd^2 to be large. The next section proves a lower bound on $\lambda_{\text{gap}}(L_A)$ without this restriction, and Sections 8.5 and 8.6 treat the critical window, where Nd^2 tends to zero.

8.3 Many precisions comparable to the smallest one

In this section, A is diagonal, with precisions p_i , variances q_i , smallest precision $d = p_1$ and weight $\varphi = x^\top Q x / N$, as in Sections 8.1 and 8.2. Let $L \geq 1$, and let $K \subset \{1, \dots, N\}$ be a set of $k \geq 3$ indices such that

$$p_i \leq L d \quad \text{for every } i \in K. \quad (8.11)$$

We write $\ell = k - 2$ and $S_K(x) = \sum_{i \in K} x_i^2$. The gradient of $\log \varphi$ can be large where the coordinates in K are small. The following inequality, of Hardy type, controls its square on average by $1/S_K$, which is small on average when k is large.

Lemma 8.7 (A Hardy-type inequality). *Assume (8.11). For every smooth u on S_N ,*

$$\pi_A(\Gamma(\log \varphi) u^2) \leq \frac{8L^2 d}{k-2} \pi_A(u^2) + \frac{32L(1+L)}{(k-2)^2} \pi_A(\Gamma(u)). \quad (8.12)$$

Proof. Since $L_A = L_{-P}$ and $\pi_A = \pi_{-P}$, we may assume that $A = -P$, so that $z = 0$ and $e_A = -\bar{p}$, where $\bar{p}(x) = N^{-1} \sum_i p_i x_i^2$. By (8.2), for every i ,

$$L_A x_i^2 = 2 - 2x_i^2 - 2p_i x_i^2 + 2\bar{p} x_i^2.$$

Multiplying by q_i/N and summing over i , and using $\sum_i q_i = N$ and $|x|^2 = N$, we find $L_A \varphi = 2(\bar{p} - 1)\varphi$, in agreement with (8.7). It follows that $L_A \log \varphi = 2\bar{p} - 2 - \Gamma(\log \varphi)$. Summing over $i \in K$ instead, and using $\Gamma(S_K) = 4S_K - 4S_K^2/N$, we find

$$L_A \log S_K = \frac{L_A S_K}{S_K} - \frac{\Gamma(S_K)}{S_K^2} = \frac{2(k-2)}{S_K} - 2 - 2\bar{p}_K + 2\bar{p} + \frac{4}{N}, \quad \bar{p}_K = \frac{\sum_{i \in K} p_i x_i^2}{S_K}.$$

Let $Y = \log(S_K/\varphi)$. Subtracting the two identities, we obtain

$$L_A Y = \frac{2(k-2)}{S_K} - 2\bar{p}_K + \frac{4}{N} + \Gamma(\log \varphi). \quad (8.13)$$

By the computation of $|a|^2$ in the proof of Lemma 8.5,

$$\Gamma(\log \varphi) = \frac{4 \sum_i q_i^2 x_i^2}{(\sum_i q_i x_i^2)^2} - \frac{4}{N} \leq \frac{4q_1}{\sum_{i \in K} q_i x_i^2} \leq \frac{4L}{S_K},$$

where we used $q_i \leq q_1 = 1/d$ for all i and $q_i \geq 1/(Ld)$ for $i \in K$. Together with $\Gamma(\log S_K) \leq 4/S_K$, this gives $\Gamma(Y) \leq 2\Gamma(\log S_K) + 2\Gamma(\log \varphi) \leq 8(1+L)/S_K$.

We claim that

$$\int u^2 L_A Y d\pi_A = -2 \int u \Gamma(u, Y) d\pi_A. \quad (8.14)$$

If $k = N$, then $S_K = N$, the function Y is smooth, and (8.14) is a case of (8.1). If $k < N$, then Y is smooth on $\{S_K > 0\}$, and $|L_A Y|$ and $\Gamma(Y)$ are bounded by a constant multiple of $1/S_K$. Under ν_N , the variable S_K/N has the Beta($k/2, (N-k)/2$) distribution, whose density is proportional to $s^{k/2-1}(1-s)^{(N-k)/2-1}$, and $1/S_K$ is integrable because $k \geq 3$; the same holds under π_A , whose density with respect to ν_N is bounded. For $0 < \varepsilon < N$, we apply the divergence theorem on $\{S_K > \varepsilon\}$. Its boundary $\{S_K = \varepsilon\}$ is a product of spheres of dimensions $k-1$ and $N-k-1$ with radii $\sqrt{\varepsilon}$ and $\sqrt{N-\varepsilon}$, and its $(N-2)$ -dimensional volume is at most $C_{N,k} \varepsilon^{(k-1)/2}$. The normal derivative of Y there is at most $\Gamma(Y)^{1/2} \leq (8(1+L)/\varepsilon)^{1/2}$. The boundary term is $O(\varepsilon^{(k-2)/2})$, which tends to zero, and the two integrals over $\{S_K > \varepsilon\}$ converge to those in (8.14) by dominated convergence. This proves (8.14).

Let $n = \pi_A(u^2)$, $U = \pi_A(u^2/S_K)$ and $E = \pi_A(\Gamma(u))$. In (8.13), we drop the non-negative terms $4/N + \Gamma(\log \varphi)$ and use $\bar{p}_K \leq Ld$. Then (8.14) and the Cauchy–Schwarz inequality give

$$\begin{aligned} 2(k-2)U - 2Ldn &\leq 2(\pi_A(u^2\Gamma(Y)))^{1/2}E^{1/2} \\ &\leq 2(8(1+L)UE)^{1/2} \leq (k-2)U + \frac{8(1+L)}{k-2}E. \end{aligned}$$

Absorbing the term $(k-2)U$, we find $U \leq 2Ldn/(k-2) + 8(1+L)E/(k-2)^2$. Multiplying by $4L$ and using $\Gamma(\log \varphi) \leq 4L/S_K$, we obtain (8.12). $\square$

Theorem 8.8 (Spectral gap when many precisions are comparable to the smallest one). *Assume (8.11), and set*

$$A_H = \frac{8L^2d}{\ell}, \quad B_H = \frac{32L(1+L)}{\ell^2}, \quad \rho_\varepsilon = \frac{1}{2} + \frac{1}{4\varepsilon} \quad (\varepsilon > 0).$$

(i) *For every smooth f on S_N ,*

$$\pi_A(\Gamma(\log \varphi)\Gamma(f)) \leq A_H \pi_A(\Gamma(f)) + B_H \pi_A(\|\nabla_S^2 f\|_{\text{HS}}^2). \quad (8.15)$$

(ii) *Let f be a nonconstant smooth function with $L_A f = -\lambda f$, and write $E_1 = \pi_A(\Gamma(f))$ and $E_2 = \pi_A(\|\nabla_S^2 f\|_{\text{HS}}^2)$. Then for every $\varepsilon > 0$,*

$$\lambda E_1 \geq (1 - \varepsilon - \rho_\varepsilon B_H) E_2 + \left(d - \frac{2}{N} - \rho_\varepsilon A_H\right) E_1. \quad (8.16)$$

(iii) *If $\ell \geq 64L^2$, then*

$$\lambda_{\text{gap}}(L_A) \geq d - \frac{2}{N} - \frac{8L^2d}{\ell}.$$

Proof. For (i), we apply Lemma 8.7 in the form

$$\pi_A(\Gamma(\log \varphi)u^2) \leq A_H \pi_A(u^2) + B_H \pi_A(\Gamma(u)) \quad (8.17)$$

to the smooth function $u_\varepsilon = (\Gamma(f) + \varepsilon^2)^{1/2}$. Using $\Gamma(f) \leq u_\varepsilon^2$ and the bound $\Gamma(u_\varepsilon) \leq \|\nabla_S^2 f\|_{\text{HS}}^2$ of Corollary 8.4, we find

$$\pi_A(\Gamma(\log \varphi)\Gamma(f)) \leq \pi_A(\Gamma(\log \varphi)u_\varepsilon^2) \leq A_H(E_1 + \varepsilon^2) + B_H E_2,$$

and we let $\varepsilon \rightarrow 0$.

For (ii), we again take $A = -P$, so that $e_A = -\bar{p}$ and (8.5) reads

$$\Gamma_2(f) = \|\nabla_S^2 f\|_{\text{HS}}^2 + \sum_i p_i g_i^2 + \left(1 - \frac{2}{N} - \bar{p}\right) |g|^2,$$

with $g = Df$ as in Lemma 8.3. Recall from the proof of Lemma 8.7 that $1 - \bar{p} = -\frac{1}{2}L_A \log \varphi - \frac{1}{2}\Gamma(\log \varphi)$. By (8.1), $\pi_A(\Gamma_2(f)) = \pi_A((L_A f)^2) = \lambda E_1$. Also by (8.1), and since the tangential gradient of $\Gamma(f)$ is $2(\nabla_S^2 f)g$,

$$-\frac{1}{2} \int (L_A \log \varphi) \Gamma(f) d\pi_A = \frac{1}{2} \int \Gamma(\log \varphi, \Gamma(f)) d\pi_A = \int \nabla_S^2 f(a, g) d\pi_A,$$

where $a = \nabla_S \log \varphi$. Integrating the expression for $\Gamma_2(f)$, we obtain

$$\lambda E_1 = E_2 + \pi_A\left(\sum_i p_i g_i^2\right) - \frac{2}{N}E_1 + \pi_A(\nabla_S^2 f(a, g)) - \frac{1}{2}\pi_A(\Gamma(\log \varphi)\Gamma(f)).$$

We have $\sum_i p_i g_i^2 \geq d|g|^2$, and $|\nabla_S^2 f(a, g)| \leq \varepsilon \|\nabla_S^2 f\|_{\text{HS}}^2 + |a|^2|g|^2/(4\varepsilon)$ with $|a|^2 = \Gamma(\log \varphi)$. Inserting these bounds, we have

$$\lambda E_1 \geq (1 - \varepsilon)E_2 + \left(d - \frac{2}{N}\right)E_1 - \rho_\varepsilon \pi_A(\Gamma(\log \varphi)\Gamma(f)),$$

and (8.16) follows from (8.15).

For (iii), we take $\varepsilon = 1/2$, so that $\rho_\varepsilon = 1$. If $\ell \geq 64L^2$, then $B_H \leq 64L^2/\ell^2 \leq 1/(64L^2) < 1/2$, and the coefficient of E_2 in (8.16) is nonnegative. Dropping this term and dividing by $E_1 > 0$, we find $\lambda \geq d - 2/N - A_H$ for every nonconstant eigenfunction, which proves (iii). $\square$

Corollary 8.9 (Poincaré inequality when many precisions are comparable to the smallest one). *Assume (8.11) with $\ell \geq 64L^2$, and suppose that $Nd \geq 8$. Then $C_P(\pi_A) \leq 2/d$. The same bound holds for every coordinate hemisphere with its reflecting form and for the coordinate-even sector. Equivalently, in terms of the half-precisions $r_i = p_i/2$ and $r_{\min} = d/2$: if $r_i \leq Lr_{\min}$ for $i \in K$, $k - 2 \geq 64L^2$, and $Nr_{\min} \geq 4$, then $C_P(\pi_A) \leq 1/r_{\min}$.*

Proof. Under these hypotheses, $8L^2d/\ell \leq d/8$ and $2/N \leq d/4$. Theorem 8.8 implies that $\lambda_{\text{gap}}(L_A) \geq 5d/8 \geq d/2$, that is, $C_P(\pi_A) \leq 2/d$. The other cases follow from Lemma 8.1. $\square$

Several later arguments apply these inequalities to conditional laws on spheres of other radii. We record the scaling.

Remark 8.10 (Spheres of other radii). Let $n \geq 2$, $S > 0$ and $v \in (0, \infty)^n$ with $\sum_i v_i = S$, and let μ_v be the probability measure on $\{y \in \mathbb{R}^n : |y|^2 = S\}$ proportional to $\exp(-\sum_i y_i^2/(2v_i))$ times the surface measure, as in (5.10), with the spherical Dirichlet form of that sphere. The map $x \mapsto (S/n)^{1/2}x$ carries $\mu_{nv/S}$ on S_n to μ_v . If f is a function on $\{|y|^2 = S\}$ and $f'(x) = f((S/n)^{1/2}x)$, then $\Gamma(f') = (S/n)\Gamma(f)$ at corresponding points, while variances and entropies are unchanged. It follows that $C_P(\mu_v) = (S/n)C_P(\mu_{nv/S})$, and similarly for C_{LS} . With $r_i = 1/(2v_i)$ and $r_{\min} = \min_i r_i$,

Corollary 8.9 reads as follows: if $r_i \leq Lr_{\min}$ for k indices i , with $k - 2 \geq 64L^2$, and $Sr_{\min} \geq 4$, then $C_P(\mu_v) \leq 1/r_{\min}$. Likewise, the inequalities of Corollary 8.13 below hold for μ_v in the form $C_{LS}(\mu_v) \leq C_\delta \max_i v_i + C_\delta C_P(\mu_v)$, since this form is invariant under the scaling.

8.4 Logarithmic Sobolev inequalities from a Gaussian extension

In this section, $q \in (0, \infty)^N$ satisfies $\sum_i q_i = N$, and $\mu = \mu_q$; recall from Section 8.1 that $\mu_q = \pi_A$ for $A = -\text{diag}(q_1^{-1}, \dots, q_N^{-1})$, whose smallest precision is $1/\max_i q_i$. We write $|q|^2 = \sum_i q_i^2$. Let $G_1, \dots, G_N$ be independent standard Gaussian variables, and let h_q be the density of $\sum_i q_i G_i^2$. The number $h_q(N)$ is the value of this density at its mean $\mathbb{E} \sum_i q_i G_i^2 = N$.

Recall that μ_q is the conditional law of a Gaussian vector given that its squared norm equals N . In the next proof, we extend a function f on S_N to Gaussian space, in a thin layer outside the sphere, so that the Gaussian norm, energy and entropy of the extension are controlled by those of f . We then apply the Gaussian logarithmic Sobolev inequality to the extension. The extension must compensate for the Jacobian of the coarea formula, which varies along the sphere. We obtain a logarithmic Sobolev inequality with an additive error term, which we then remove with a Poincaré inequality.

Theorem 8.11 (A logarithmic Sobolev inequality with an error term). *After relabeling the coordinates, let $q_1 \geq q_2 \geq \dots \geq q_N$, and write $\text{osc } q = q_1 - q_N$. Fix $\eta \in (0, 1]$ and suppose that*

$$|q|^2 \geq (1 + \eta)(\text{osc } q)^2. \quad (8.18)$$

There exists $C_\eta < \infty$, depending only on η , such that for every $f \in H^1(S_N)$,

$$\text{Ent}_\mu(f^2) \leq C_\eta q_1 \mathcal{E}_\mu(f) + [C_\eta + \log(h_q(N)^2 |q|^2)] \mu(f^2), \quad (8.19)$$

$$\text{Ent}_\mu(f^2) \leq C_\eta q_1 \mathcal{E}_\mu(f) + \left[C_\eta + \log \frac{|q|^2}{q_1 q_2} \right] \mu(f^2). \quad (8.20)$$

Moreover,

$$C_{LS}(\mu) \leq C_\eta q_1 + \left[C_\eta + 2 + \log \frac{|q|^2}{q_1 q_2} \right] C_P(\mu). \quad (8.21)$$

The same three inequalities hold for every coordinate hemisphere and for the coordinate-even sector, with Ent , $\mathcal{E}$ and C_P of that form. In particular, if $q_2 \geq \delta q_1$ and $|q|^2 \leq K' q_1^2$ for some $\delta \in (0, 1]$ and $K' < \infty$, then $C_{LS}(\mu) \leq C_\delta q_1 + C_{\delta, K'} C_P(\mu)$.

In terms of the half-precisions $r_i = 1/(2q_i)$, the first term is $C_\eta q_1 = C_\eta / (2 \min_i r_i)$.

Proof. The proof has seven steps. Step 1 transfers the problem to an ellipsoid in Gaussian coordinates. Step 2 proves a weighted estimate on the ellipsoid. Steps 3–5 construct

the extension along a normal flow and bound its norm and energy. Step 6 applies the Gaussian logarithmic Sobolev inequality and absorbs a logarithmic weight. Step 7 bounds $h_q(N)$ and concludes. We may take f smooth, by the density argument of Section 8.1. By Lemma 8.1, it suffices to prove (8.19) and (8.20) on the full sphere: they then hold for every hemisphere and for the orthant, and (8.21) for each of these forms follows from Lemma 8.2(ii), since the bracket in (8.20) is nonnegative (see Step 7).

Step 1: an ellipsoid in Gaussian coordinates. Let $Q = \text{diag}(q_1, \dots, q_N)$, $\mathcal{Q}(y) = y^\top Q y$, and $\omega(y) = y^\top Q^2 y = |\nabla \mathcal{Q}(y)|^2/4$. For $\ell > 0$, the level set $E_\ell = \{\mathcal{Q} = \ell\}$ is a smooth ellipsoid with unit normal $n = Qy/\omega^{1/2}$. Let γ be the standard Gaussian measure on $\mathbb{R}^N$ and σ_ℓ the surface measure of E_ℓ . By the coarea formula [134, Chapter 2, Section 6, (6.3), p. 66], for bounded measurable g ,

$$\int g d\gamma = \int_0^\infty \int_{E_\ell} g(y) \frac{e^{-|y|^2/2}}{(2\pi)^{N/2} 2\omega(y)^{1/2}} \sigma_\ell(dy) d\ell.$$

In particular, the inner integral with $g = 1$ is $h_q(\ell)$, and the probability measure $\mu^{(\ell)}$ on E_ℓ with density proportional to $e^{-|y|^2/2}\omega^{-1/2}$ with respect to σ_ℓ is the conditional law of a standard Gaussian vector G given $\mathcal{Q}(G) = \ell$. The vector $X = Q^{1/2}G$ has independent $\mathcal{N}(0, q_i)$ coordinates and $|X|^2 = \mathcal{Q}(G)$. By the coarea formula in the coordinates x , where $|\nabla |x|^2|$ is constant on each sphere, the conditional law of X given $|X|^2 = \ell$ is proportional to $e^{-\sum_i x_i^2/(2q_i)}$ times the surface measure; for $\ell = N$ this is μ_q . Both families of conditional laws depend continuously on ℓ , and they describe the same random vector given the same variable. We conclude that $\mu^E = \mu^{(N)}$ is the image of μ_q under $x \mapsto Q^{-1/2}x$. We write $E = E_N$.

Let $\Gamma_E(\hat{f}) = |(I - nn^\top)\nabla \hat{f}|^2$ be the Euclidean tangential energy on E . For f smooth near S_N , let $\hat{f}(y) = f(Q^{1/2}y)$. Using $\partial_{y_i}\hat{f} = q_i^{1/2}\partial_i f$ and $\nabla \mathcal{Q} = 2Qy$, we find, at $x = Q^{1/2}y$,

$$\Gamma_E(\hat{f}) = \min_{b \in \mathbb{R}} |\nabla \hat{f} - 2bQy|^2 = \min_{b \in \mathbb{R}} \sum_i q_i (\partial_i f - 2bx_i)^2 \leq q_1 \min_{b \in \mathbb{R}} \sum_i (\partial_i f - 2bx_i)^2 = q_1 \Gamma(f).$$

Here $\partial_i f$ are the derivatives of any smooth extension of f , and both minima are tangential energies, which do not depend on the extension. The image map preserves Ent and $\mu(f^2)$, and it suffices to show that for every smooth f on E ,

$$\text{Ent}_{\mu^E}(f^2) \leq C_\eta \int \Gamma_E(f) d\mu^E + [C_\eta + \log(h_q(N)^2|q|^2)] \mu^E(f^2). \quad (8.22)$$

Step 2: a weighted estimate on E . For $y \in E$, let $w_i(y) = q_i^2 y_i^2 / \omega(y)$. These weights form a probability vector, and we write $\langle q \rangle_w = \sum_i w_i q_i$ and $\text{Var}_w(q) = \sum_i w_i q_i^2 - \langle q \rangle_w^2$. Then $\langle q \rangle_w \in [q_N, q_1]$ and $\text{Var}_w(q) \leq (\text{osc } q)^2/4$. The generator L_E of the form $\int \Gamma_E d\mu^E$ is

$\Delta_E + \nabla_E \log \rho \cdot \nabla_E$, where $\rho \propto e^{-|y|^2/2} \omega^{-1/2}$ and Δ_E is the Laplace–Beltrami operator of E . For a smooth function f defined near E , we have $\Delta_E f = \text{tr}((I - nn^\top)D^2 f) - (\text{div } n) n \cdot Df$, where $\text{div } n = (\text{tr } Q - \langle q \rangle_w)/\omega^{1/2}$ is the mean curvature of E . The tangential part of $\nabla \log \rho = -y - Q^2 y/\omega$ is $-y + Q(y)Qy/\omega - Q^2 y/\omega + \langle q \rangle_w Qy/\omega$, since $n \cdot y = Q(y)/\omega^{1/2}$ and $n \cdot Q^2 y = \langle q \rangle_w \omega^{1/2}$. Adding the term $-(\text{div } n) n \cdot Df$ and using $\text{tr } Q = N = Q(y)$ on E , we obtain

$$L_E f = \text{tr}((I - nn^\top)D^2 f) + \left(-y - \frac{Q^2 y}{\omega} + \frac{2\langle q \rangle_w Qy}{\omega}\right) \cdot Df.$$

For ω , with $D\omega = 2Q^2 y$ and $D^2 \omega = 2Q^2$, this formula gives

$$\begin{aligned} L_E \omega &= 2|q|^2 - 2\omega - 4\text{Var}_w(q), & \Gamma_E(\omega) &= 4\omega \text{Var}_w(q), \\ L_E \log \omega &= \frac{2|q|^2}{\omega} - 2 - \frac{8\text{Var}_w(q)}{\omega}. \end{aligned} \quad (8.23)$$

The formula for $\Gamma_E(\omega)$ uses only the normal n and holds on every level set E_ℓ . Let $\mathcal{W} = \omega^{-\eta}$. Then

$$\frac{-L_E \mathcal{W}}{\mathcal{W}} = \eta L_E \log \omega - \eta^2 \Gamma_E(\log \omega) = \frac{2\eta|q|^2 - (8\eta + 4\eta^2)\text{Var}_w(q)}{\omega} - 2\eta \geq \frac{\Lambda_\eta}{\omega} - 2\eta,$$

with $\Lambda_\eta = \eta[2|q|^2 - (2 + \eta)(\text{osc } q)^2]$. We claim $\Lambda_\eta \geq \eta^2 q_1^2/4$. If $\text{osc } q \geq q_1/2$, then (8.18) implies that $2|q|^2 - (2 + \eta)(\text{osc } q)^2 \geq \eta(\text{osc } q)^2$, and $\Lambda_\eta \geq \eta^2(\text{osc } q)^2 \geq \eta^2 q_1^2/4$. If $\text{osc } q < q_1/2$, then $|q|^2 \geq q_1^2$ and $(2 + \eta)(\text{osc } q)^2 < 3q_1^2/4$, and $\Lambda_\eta \geq 5\eta q_1^2/4 \geq \eta^2 q_1^2/4$. The manifold E is compact and has no boundary. The argument of Lemma 8.2(iii) applies to L_E and $\mathcal{W}$, and we obtain

$$\int f^2 \frac{q_1^2}{\omega} d\mu^E \leq \frac{4}{\eta^2} \int \Gamma_E(f) d\mu^E + \frac{8}{\eta} \mu^E(f^2). \quad (8.24)$$

We also record that $\int L_E \omega d\mu^E = 0$. Integrating the first identity in (8.23) against μ^E , we find $\mu^E(\omega) = |q|^2 - 2\mu^E(\text{Var}_w(q)) \leq |q|^2$.

Step 3: the normal flow. For $\omega \in E$, let $(Y_\ell(\omega))_{\ell \geq N}$ solve

$$\frac{dY_\ell}{d\ell} = V(Y_\ell), \quad V(y) = \frac{Qy}{2\omega(y)}, \quad Y_N(\omega) = \omega.$$

We have $Q(Y_\ell) = \ell$, because $\nabla Q \cdot V = 1$. On $\{Q \geq N\}$, we have $\omega \geq q_N Q > 0$, and V is smooth with linear growth. The solution exists for all $\ell \geq N$, and the flow, with its inverse obtained by running it backward, maps E diffeomorphically onto E_ℓ . We write $u = \ell - N$, $\omega_\ell(\omega) = \omega(Y_\ell(\omega))$ and $\omega_0 = \omega_N$. Using $d\omega_\ell/d\ell = D\omega \cdot V = \langle q \rangle_w(Y_\ell)$, we find

$$\frac{d\omega_\ell}{d\ell} \in [q_N, q_1], \quad \omega_0 \leq \omega_\ell \leq \omega_0 + q_1 u. \quad (8.25)$$

For tangent vectors ξ, ξ' to E at ω , let $\xi_\ell = DY_\ell(\omega)\xi$ and $\xi'_\ell = DY_\ell(\omega)\xi'$, which are tangent to E_ℓ . Then

$$\frac{d}{d\ell} \langle \xi_\ell, \xi'_\ell \rangle = \langle DV\xi_\ell, \xi'_\ell \rangle + \langle \xi_\ell, DV\xi'_\ell \rangle = \frac{\xi_\ell^\top Q \xi'_\ell}{\omega_\ell},$$

because $DV\xi = Q\xi/(2\omega) - Qy(D\omega \cdot \xi)/(2\omega^2)$ and $Qy \perp \xi'_\ell$. In particular $|\xi_\ell|$ is non-decreasing in ℓ . We say that a function $\tilde{g}$ on E_ℓ is transported from $E_{\ell'}$, $N \leq \ell' \leq \ell$, if $\tilde{g}(Y_\ell(\omega)) = g(Y_{\ell'}(\omega))$ for a function g on $E_{\ell'}$. Writing Γ_ℓ for the tangential energy on E_ℓ , we see from the monotonicity of $|\xi_\ell|$ that

$$\Gamma_\ell(\tilde{g})(Y_\ell(\omega)) \leq \Gamma_{\ell'}(g)(Y_{\ell'}(\omega)). \quad (8.26)$$

Applying (8.23) on each level, we find that $\Gamma_\ell(\log \omega) \leq (\text{osc } q)^2/\omega$ on E_ℓ . Combined with (8.26) and (8.25), this shows that for every $N \leq \ell' \leq \ell$ the function $\log \omega_{\ell'}$, viewed on E_ℓ , satisfies $\Gamma_\ell(\log \omega_{\ell'}) \leq (\text{osc } q)^2/\omega_0$.

We now compute the Gaussian density in the coordinates $(\omega, \ell) \in E \times [N, \infty)$. We write $\gamma(dy) = j_\ell(\omega) \sigma_N(d\omega) d\ell$ on $\{Q \geq N\}$. At $\ell = N$, the factor $j_N(\omega) = (2\pi)^{-N/2} e^{-|\omega|^2/2} / (2\omega_0^{1/2})$ is the coarea density of Step 1, and $j_N(\omega) \sigma_N(d\omega) = h_q(N) \mu^E(d\omega)$. By the Liouville formula, $\partial_\ell \log j_\ell = (\text{div } V - y \cdot V)(Y_\ell)$, where $y \cdot V$ comes from the Gaussian factor. We have $\text{div } V = \text{tr } Q/(2\omega) - \langle q \rangle_w/\omega$ and $y \cdot V = Q/(2\omega)$. On E_ℓ , where $Q = N + u$ and $\text{tr } Q = N$, this gives $\partial_\ell \log j_\ell = -u/(2\omega_\ell) - \langle q \rangle_w/\omega_\ell$. Integrating, and using $\partial_\ell \log \omega_\ell = \langle q \rangle_w/\omega_\ell$, we arrive at

$$\frac{j_\ell}{j_N} = \frac{\omega_0}{\omega_\ell} e^{-I_\ell}, \quad I_\ell = \int_0^u \frac{v}{2\omega_{N+v}} dv. \quad (8.27)$$

For $0 \leq u \leq \omega_0^{1/2}$, we have $0 \leq I_\ell \leq u^2/(4\omega_0) \leq 1/4$.

Step 4: the extension and its norm. Fix once and for all a nonzero $\chi \in C_c^\infty((1/4, 3/4))$, and let $\chi_0 = \int \chi^2$, $\chi_1 = \int (\chi')^2$ and $\chi_2 = \int \chi^2 \log \chi^2$. For smooth f on E , we define $\mathcal{F} = 0$ on $\{Q < N\}$ and

$$\mathcal{F}(Y_\ell(\omega)) = \alpha_\ell(\omega) f(\omega) \chi(u/\omega_0(\omega)^{1/2}), \quad \alpha_\ell = \left(\frac{j_N}{j_\ell}\right)^{1/2} \omega_0^{-1/4} = \left(\frac{\omega_\ell}{\omega_0}\right)^{1/2} e^{I_\ell/2} \omega_0^{-1/4}. \quad (8.28)$$

On the support of $\mathcal{F}$, we have $\omega_0^{1/2}/4 \leq u \leq 3\omega_0^{1/2}/4$. The function ω_0 is bounded above and below on E , and this support is compact and at positive distance from E , and $\mathcal{F}$ is smooth on $\mathbb{R}^N$. The amplitude cancels the Jacobian: $\alpha_\ell^2 j_\ell = j_N \omega_0^{-1/2}$. Substituting $u = \omega_0^{1/2} t$, we find

$$\int \mathcal{F}^2 d\gamma = \int_E j_N(\omega) f(\omega)^2 \int_0^\infty \chi(t)^2 dt \sigma_N(d\omega) = \chi_0 h_q(N) \mu^E(f^2). \quad (8.29)$$

Step 5: the energy of the extension. For a function $\mathcal{G}$ on $\{\mathcal{Q} > N\}$, the gradient splits into its tangential part on E_ℓ and its normal part. As $V = n/(2\omega^{1/2})$, the derivative ∂_ℓ along the flow at fixed ω satisfies $n \cdot \nabla \mathcal{G} = 2\omega_\ell^{1/2} \partial_\ell \mathcal{G}$, and so

$$|\nabla \mathcal{G}|^2 = \Gamma_\ell(\mathcal{G}) + 4\omega_\ell (\partial_\ell \mathcal{G})^2.$$

Let $\tilde{f}$ be f transported along the flow, so that $\partial_\ell \tilde{f} = 0$ and $\Gamma_\ell(\tilde{f}) \leq \Gamma_E(f)$ by (8.26). Let $H_0 = q_1^2/\omega_0$, a function on E , and recall $\text{osc } q \leq q_1$. We bound the derivatives of the two other factors in (8.28) on the support of $\mathcal{F}$.

First, $\log \mathfrak{a}_\ell = \frac{1}{2} \log \omega_\ell - \frac{3}{4} \log \omega_0 + \frac{1}{2} I_\ell$. Tangential differentiation fixes u . By the bound at the end of Step 3 and Minkowski's inequality,

$$\Gamma_\ell(I_\ell)^{1/2} \leq \int_0^u \frac{v}{2\omega_{N+v}} \Gamma_\ell(\log \omega_{N+v})^{1/2} dv \leq \frac{\text{osc } q}{\omega_0^{1/2}} \int_0^u \frac{v}{2\omega_0} dv \leq \frac{\text{osc } q}{4\omega_0^{1/2}},$$

since $u \leq \omega_0^{1/2}$. Using the inequality $(\theta_1 + \theta_2 + \theta_3)^2 \leq 3(\theta_1^2 + \theta_2^2 + \theta_3^2)$ and the bounds of Step 3, we obtain

$$\Gamma_\ell(\log \mathfrak{a}_\ell) \leq 3 \left(\frac{1}{4} + \frac{9}{16} + \frac{1}{4} \cdot \frac{1}{16} \right) \frac{(\text{osc } q)^2}{\omega_0} = \frac{159}{64} \frac{(\text{osc } q)^2}{\omega_0} \leq \frac{159}{64} H_0.$$

In the normal direction, $\partial_\ell \log \mathfrak{a}_\ell = \langle q \rangle_w / (2\omega_\ell) + u / (4\omega_\ell)$. Since $\langle q \rangle_w \leq q_1$, $\omega_\ell \geq \omega_0$ and $u \leq \omega_0^{1/2}$, we have

$$4\omega_\ell (\partial_\ell \log \mathfrak{a}_\ell)^2 \leq \frac{2\langle q \rangle_w^2}{\omega_\ell} + \frac{u^2}{2\omega_\ell} \leq 2H_0 + \frac{1}{2}.$$

Together, $|\nabla \log \mathfrak{a}_\ell|^2 \leq 5H_0 + 1$. Second, let $t = u/\omega_0^{1/2} \leq 1$. Then $\partial_\ell t = \omega_0^{-1/2}$, and the tangential gradient of t is $-(t/2)$ times that of $\log \omega_0$, and $\Gamma_\ell(t) \leq (\text{osc } q)^2 / (4\omega_0) \leq H_0/4$. Using (8.25) and $4H_0^{1/2} \leq 2 + 2H_0$, we find

$$|\nabla \chi(t)|^2 = \chi'(t)^2 \left(\Gamma_\ell(t) + \frac{4\omega_\ell}{\omega_0} \right) \leq \chi'(t)^2 \left(\frac{H_0}{4} + 4 + \frac{4q_1 u}{\omega_0} \right) \leq \chi'(t)^2 \left(6 + \frac{9}{4} H_0 \right).$$

By the product rule and the same elementary inequality,

$$|\nabla \mathcal{F}|^2 \leq 3\mathfrak{a}_\ell^2 \left(\tilde{f}^2 \chi(t)^2 (5H_0 + 1) + \chi(t)^2 \Gamma_E(f) + \tilde{f}^2 \chi'(t)^2 \left(6 + \frac{9}{4} H_0 \right) \right).$$

We integrate against $j_\ell(\omega) \sigma_N(d\omega) d\ell$, use $\mathfrak{a}_\ell^2 j_\ell = j_N \omega_0^{-1/2}$ and $d\ell = \omega_0^{1/2} dt$, and obtain

$$\int |\nabla \mathcal{F}|^2 d\gamma \leq 3h_q(N) \left\{ \chi_0 \int \Gamma_E(f) d\mu^E + (\chi_0 + 6\chi_1) \mu^E(f^2) + \left(5\chi_0 + \frac{9}{4}\chi_1 \right) \mu^E(f^2 H_0) \right\}. \quad (8.30)$$

By (8.24), the right side is at most $\chi_0 h_q(N) (A_\eta \int \Gamma_E(f) d\mu^E + B_\eta \mu^E(f^2))$, where

$$A_\eta = 3 + \frac{12}{\eta^2} \left(5 + \frac{9\chi_1}{4\chi_0} \right), \quad B_\eta = 3 \left(1 + \frac{6\chi_1}{\chi_0} \right) + \frac{24}{\eta} \left(5 + \frac{9\chi_1}{4\chi_0} \right).$$

Step 6: the Gaussian inequality and the absorption of a weight. Let $n_f = \mu^E(f^2)$. Computing $\int \mathcal{F}^2 \log \mathcal{F}^2 d\gamma$ as in (8.29) and using $\int \mathcal{F}^2 d\gamma = \chi_0 h_q(N) n_f$, we obtain

$$\frac{\text{Ent}_\gamma(\mathcal{F}^2)}{\chi_0 h_q(N)} = \text{Ent}_{\mu^E}(f^2) + \frac{1}{\chi_0} \int f^2 \left(\int_0^\infty \chi(t)^2 \log(\alpha_\ell^2 \chi(t)^2) dt \right) d\mu^E - n_f \log(\chi_0 h_q(N)).$$

We have $\log \alpha_\ell^2 = \log(\omega_\ell/\omega_0) + I_\ell - \frac{1}{2} \log \omega_0 \geq -\frac{1}{2} \log \omega_0$, and the inner integral is at least $\chi_2 - \frac{\chi_0}{2} \log \omega_0$. By Lemma 8.2(iv) and Step 5,

$$\text{Ent}_{\mu^E}(f^2) \leq 2A_\eta \int \Gamma_E(f) d\mu^E + C'_\eta n_f + \frac{1}{2} \int f^2 \log \omega d\mu^E + n_f \log h_q(N), \quad (8.31)$$

with $C'_\eta = 2B_\eta + \log \chi_0 - \chi_2/\chi_0$. The weight $\log \omega$ must be paid with the factor f^2 . By the variational formula for the entropy, $\int f^2 g d\mu^E \leq \text{Ent}_{\mu^E}(f^2) + n_f \log \int e^g d\mu^E$ for every bounded measurable real function g ; with $g = \log(\omega/q_1^2)$, which is bounded since E is compact and $\omega > 0$, and the bound $\mu^E(\omega) \leq |q|^2$ from Step 2,

$$\int f^2 \log \frac{\omega}{q_1^2} d\mu^E \leq \text{Ent}_{\mu^E}(f^2) + n_f \log \frac{|q|^2}{q_1^2}.$$

Writing $\log \omega = \log(\omega/q_1^2) + 2 \log q_1$, we find $\frac{1}{2} \int f^2 \log \omega d\mu^E \leq \frac{1}{2} \text{Ent}_{\mu^E}(f^2) + \frac{1}{2} n_f \log |q|^2$. Inserting this into (8.31) and absorbing half of the entropy, we obtain (8.22) with $C_\eta = \max\{4A_\eta, 2C'_\eta, 0\}$, which depends only on η since χ is fixed. Together with Step 1, this proves (8.19).

Step 7: the density at the mean. The variable $q_i G_i^2$ has density $e^{-y/(2q_i)} / (2\pi q_i y)^{1/2}$ on $(0, \infty)$. The density of $q_1 G_1^2 + q_2 G_2^2$ at $y > 0$ is the convolution of these densities, which is at most

$$\frac{1}{2\pi(q_1 q_2)^{1/2}} \int_0^y \frac{dv}{(v(y-v))^{1/2}} = \frac{1}{2(q_1 q_2)^{1/2}}.$$

Convolution with the law of $\sum_{i \geq 3} q_i G_i^2$ preserves this bound. Thus $h_q(N)^2 \leq 1/(4q_1 q_2)$, and (8.19) implies

$$\text{Ent}_\mu(f^2) \leq C_\eta q_1 \mathcal{E}_\mu(f) + \left[C_\eta + \log \frac{|q|^2}{4q_1 q_2} \right] \mu(f^2), \quad (8.32)$$

which gives (8.20). The bracket in (8.20) is nonnegative, because $|q|^2 \geq q_1^2 + q_2^2 \geq 2q_1 q_2$, and Lemma 8.2(ii) implies (8.21). Finally, if $q_2 \geq \delta q_1$ and $|q|^2 \leq K' q_1^2$, then $|q|^2 \geq (1 + \delta^2) q_1^2 \geq (1 + \delta^2) (\text{osc } q)^2$, which is (8.18) with $\eta = \delta^2$, and $\log(|q|^2/(q_1 q_2)) \leq \log(K'/\delta)$. This completes the proof. $\square$

The second inequality is useful only when $|q|^2/q_1^2$ is bounded. When at least three variances are comparable to the largest one, the density $h_q(N)$ in (8.19) is bounded by a constant multiple of $|q|^{-1}$, whatever the size of $|q|^2/q_1^2$. This is the content of the next lemma, which is also used in Sections 12, 17, and 19.

Lemma 8.12 (Densities of weighted sums of squared Gaussians). *Let $b_1, \dots, b_n > 0$, $\beta = \max_i b_i$, $M_1 = \sum_i b_i$ and $|b|^2 = \sum_i b_i^2$. Let $G_1, \dots, G_n$ be independent standard Gaussian variables, let h_b be the density of $Q_b = \sum_i b_i G_i^2$, let f_i be the density of $b_i G_i^2$, and let $h_{b,-i}$ be the density of $Q_b - b_i G_i^2$. Fix $\delta \in (0, 1]$ and suppose that there exist three distinct indices i with $b_i \geq \delta\beta$. Then there exist $c_\delta, C_\delta > 0$, depending only on δ , such that*

$$h_b(M_1) \geq \frac{c_\delta}{|b|}, \quad \sup_y h_b(y) \leq \frac{C_\delta}{|b|}, \quad \sup_y h_{b,-i}(y) \leq \frac{C_\delta}{|b|} \quad \text{for every } i.$$

Moreover, for every i , the conditional expectation at the exact level,

$$\mathbb{E}[b_i G_i^2 \mid Q_b = M_1] = \frac{1}{h_b(M_1)} \int_0^{M_1} y f_i(y) h_{b,-i}(M_1 - y) dy,$$

satisfies $\mathbb{E}[b_i G_i^2 \mid Q_b = M_1] \leq C_\delta b_i$.

Proof. Let $R = |b|^2/\beta^2 \geq 1$. We distinguish bounded and large values of R , with a threshold R_0 fixed below that depends on nothing.

Large R . Let $\sigma^2 = 2|b|^2$, which is the variance of Q_b , let $Z = (Q_b - M_1)/\sigma$, and let $\hat{h}(t) = \mathbb{E}e^{itZ}$. With $u_j = b_j/\beta \leq 1$, we have $\sum_j u_j^2 = R$ and

$$\hat{h}(t) = \prod_j e^{-itb_j/\sigma} (1 - 2itb_j/\sigma)^{-1/2}, \quad |\hat{h}(t)| = \prod_j \left(1 + \frac{2t^2 u_j^2}{R}\right)^{-1/4} \leq \left(1 + \frac{2t^2}{R}\right)^{-R/4},$$

where we used $\log(1 + xs) \geq s \log(1 + x)$ for $x \geq 0$ and $s = u_j^2 \in [0, 1]$, which holds because the left side is concave in s . For $|t| \leq R^{1/2}$, we have $|\hat{h}(t)| \leq e^{-t^2/6}$ by the bound $\log(1 + x) \geq x/3$ on $[0, 2]$. For $|t| \geq R^{1/2}$, the bound $1 + 2y^2 \geq 3y$ for $y \geq 1$ gives $\int_{|t| \geq R^{1/2}} |\hat{h}(t)| dt \leq 2R^{1/2} 3^{-R/4}/(R/4 - 1)$ when $R > 4$. Next, with $w_j = 2itb_j/\sigma$, we have $|w_j| \leq \sqrt{2}|t|R^{-1/2}$ and $-\frac{1}{2}\log(1 - w) = \frac{1}{2}w + \frac{1}{4}w^2 + \frac{1}{2}\epsilon(w)$ with $|\epsilon(w)| \leq \frac{2}{3}|w|^3$ for $|w| \leq \frac{1}{2}$. The linear terms cancel the factors $e^{-itb_j/\sigma}$, the quadratic terms sum to $-t^2|b|^2/\sigma^2 = -t^2/2$, and since $\sum_j b_j^3 \leq \beta|b|^2$, the remainders sum to at most $|t|^3 R^{-1/2}$ in absolute value. It follows that $\hat{h}(t) \rightarrow e^{-t^2/2}$ uniformly on compact sets of t as $R \rightarrow \infty$, uniformly in b . Let $\epsilon_0 = (2\pi)^{-1/2}/4$. We first choose T_0 such that $\pi^{-1} \int_{T_0}^\infty (e^{-t^2/6} + e^{-t^2/2}) dt \leq \epsilon_0/3$, and then $R_F \geq \max\{8, 8T_0^2\}$ such that, for $R \geq R_F$, the integral of $\pi^{-1}|\hat{h}(t) - e^{-t^2/2}|$ over $[0, T_0]$ and the quantity $\pi^{-1}R^{1/2}3^{-R/4}/(R/4 - 1)$ are each at most $\epsilon_0/3$. For $R \geq R_F$, Fourier inversion shows that Z has a continuous density g_Z with $|g_Z(x) - (2\pi)^{-1/2}e^{-x^2/2}| \leq \epsilon_0$ for all x . In particular, $g_Z(0) \geq \frac{3}{4}(2\pi)^{-1/2}$ and $\sup g_Z \leq \frac{5}{4}(2\pi)^{-1/2}$. The identity $h_b(y) = \sigma^{-1}g_Z((y - M_1)/\sigma)$ then yields the first two bounds with absolute constants when $R \geq R_F$.

Let $R_0 = R_F + 1$ and suppose $R > R_0$. Deleting coordinate i leaves weights $b^{(-i)}$ with $|b^{(-i)}|^2 \geq |b|^2 - \beta^2 \geq |b|^2/2$ and maximum at most β , and the ratio of $|b^{(-i)}|^2$ to the

square of this maximum is at least $R - 1 > R_F$. Applying the previous paragraph to these weights, we find that $\sup h_{b,-i} \leq C|b^{(-i)}|^{-1} \leq \sqrt{2}C|b|^{-1}$.

Bounded R . Suppose $R \leq R_0$. Let i_1 be an index with $b_{i_1} = \beta$, let $i_2 \neq i_1$ be one of the distinguished indices, so that $b_{i_2} \geq \delta\beta$, and let U be the sum of the remaining terms $b_j G_j^2$. Then $\text{Var } U \leq 2|b|^2 \leq 2R_0\beta^2$. Let $s_0 = (b_{i_1} + b_{i_2})/2 \geq \beta/2$. By Cantelli's inequality,

$$\mathbb{P}(U - \mathbb{E}U \leq s_0) \geq \frac{s_0^2}{\text{Var } U + s_0^2} \geq \frac{1}{8R_0 + 1} =: p_0,$$

and by Chebyshev's inequality, $\mathbb{P}(U - \mathbb{E}U < -L_0\beta) \leq 2R_0/L_0^2 \leq p_0/2$ if $L_0^2 = 4R_0/p_0$. On the event $\{-L_0\beta \leq U - \mathbb{E}U \leq s_0\}$, which has probability at least $p_0/2$, the variable $M_1 - U = b_{i_1} + b_{i_2} - (U - \mathbb{E}U)$ lies in $[s_0, (2 + L_0)\beta]$. The density of $b_{i_1}G_{i_1}^2 + b_{i_2}G_{i_2}^2$ at y is

$$\frac{1}{2\pi(b_{i_1}b_{i_2})^{1/2}} \int_0^y \frac{e^{-v/(2b_{i_1}) - (y-v)/(2b_{i_2})}}{(v(y-v))^{1/2}} dv \geq \frac{e^{-y/(2b_{i_2})}}{2(b_{i_1}b_{i_2})^{1/2}} \geq \frac{e^{-(2+L_0)/(2\delta)}}{2\beta}$$

for $y \leq (2 + L_0)\beta$, where we used $b_{i_2} \leq b_{i_1}$ and $\int_0^y (v(y-v))^{-1/2} dv = \pi$. Averaging over U , we obtain $h_b(M_1) \geq (p_0/4)e^{-(2+L_0)/(2\delta)}\beta^{-1}$, and $\beta \leq |b| \leq R_0^{1/2}\beta$. For the upper bounds, after deleting any coordinate i , at least two of the three distinguished indices remain, say j and j' . By the computation in Step 7 of the proof of Theorem 8.11, the density of $b_j G_j^2 + b_{j'} G_{j'}^2$ is at most $(2(b_j b_{j'})^{1/2})^{-1} \leq (2\delta\beta)^{-1}$, and convolution with the other terms preserves this bound. Using $|b| \leq R_0^{1/2}\beta$, we find $\sup h_{b,-i} \leq R_0^{1/2}/(2\delta|b|)$.

Conclusion. In both cases, $\sup h_b \leq \sup h_{b,-i}$, since $h_b = h_{b,-i} * f_i$. Finally,

$$\mathbb{E}[b_i G_i^2 \mid Q_b = M_1] \leq \frac{\sup h_{b,-i}}{h_b(M_1)} \int_0^\infty y f_i(y) dy = \frac{\sup h_{b,-i}}{h_b(M_1)} b_i \leq C_\delta b_i.$$

□

Corollary 8.13 (Logarithmic Sobolev inequality when three variances are comparable to the largest one). *Let $q \in (0, \infty)^N$ with $\sum_i q_i = N$, write $q_{\max} = \max_i q_i$, and let $\delta \in (0, 1]$. Suppose that there exist three distinct indices i with $q_i \geq \delta q_{\max}$. There exists C_δ , depending only on δ , such that for the full sphere, every coordinate hemisphere with its reflecting form, and the coordinate-even sector,*

$$\text{Ent}_{\mu_q}(f^2) \leq C_\delta q_{\max} \mathcal{E}_{\mu_q}(f) + C_\delta \mu_q(f^2), \quad C_{\text{LS}}(\mu_q) \leq C_\delta q_{\max} + C_\delta C_P(\mu_q),$$

where $\mathcal{E}$, Ent and C_P refer to the corresponding form.

In terms of half-precisions, $q_{\max} = 1/(2r_{\min})$, and the hypothesis says that three of the r_i are at most $r_{\min}/\delta$.

Proof. At least two of the three indices differ from an index where the maximum is attained, and their squares contribute at least $2\delta^2 q_{\max}^2$ to $|q|^2$. Using $\text{osc } q \leq q_{\max}$, we see that $|q|^2 \geq (1 + 2\delta^2)q_{\max}^2 \geq (1 + \eta)(\text{osc } q)^2$ with $\eta = \min\{1, 2\delta^2\}$. By Lemma 8.12 with $b = q$ and $M_1 = N$, we have $h_q(N) \leq C_\delta |q|^{-1}$, that is, $\log(h_q(N)^2 |q|^2) \leq 2 \log C_\delta$. Inserting this into (8.19) proves the first inequality, for all three forms. The second follows from Lemma 8.2(ii), applied to each form with its own Poincaré constant. $\square$

8.5 A few precisions close to the smallest one

In the critical window, the smallest precision is of order $N^{-2/3}$, the gap between the two smallest precisions is of the same order, and the conditions (4.36) place only a bounded number of precisions within a fixed factor of the smallest one. These conditions do not guarantee the hypotheses of Corollary 8.9. In this section we prove a Poincaré inequality on the hemisphere $\{x_1 \geq 0\}$ with its reflecting form (Lemma 8.18), and then pass to the full sphere (Lemma 8.19).

Setting. Throughout this section, $k \geq 5$ and $N > k$ are integers, and $a_1, \dots, a_N \geq 0$ are quadratic potential coefficients such that, for some $0 < \bar{a} \leq a$,

$$a_1 = 0, \quad 0 < a_2 \leq a_i \leq \bar{a} \quad (2 \leq i \leq k), \quad a_j \geq a \quad (j > k). \quad (8.33)$$

Let $F(x) = \sum_i a_i x_i^2$, let μ be the probability measure proportional to $e^{-F} \nu_N$, and let $\mu_+ = \mu(\cdot \mid x_1 \geq 0)$. Note that $\mu = \pi_A$ with $A = -2 \text{diag}(a_1, \dots, a_N)$: the coefficients a_i are half-precisions shifted so that the smallest one is zero. We write $x = (\sqrt{S} \omega, y)$, where $y = (x_{k+1}, \dots, x_N) \in B_N = \{y \in \mathbb{R}^{N-k} : |y|^2 < N\}$, $S = N - |y|^2 = \sum_{i \leq k} x_i^2$ and ω is in the unit sphere $S^{k-1} \subset \mathbb{R}^k$. Let σ^k be the uniform probability measure on S^{k-1} and σ_+^k that on $S_+^{k-1} = \{\omega \in S^{k-1} : \omega_1 \geq 0\}$. In these coordinates, the metric of S_N is $(I + yy^\top/S) \oplus S g_{S^{k-1}}$, whose volume element with respect to $dy \sigma^k(d\omega)$ is proportional to $S^{k/2-1}$. We obtain

$$\mu(dx) \propto e^{-V(y)} e^{-S\mathcal{A}(\omega)} S^{k/2-1} dy \sigma^k(d\omega), \quad V(y) = \sum_{j>k} a_j y_j^2, \quad \mathcal{A}(\omega) = \sum_{i \leq k} a_i \omega_i^2, \quad (8.34)$$

and, since the inverse of $I + yy^\top/S$ is $M_y = I - yy^\top/N$,

$$\Gamma(f) = \Gamma_y(f) + S^{-1} |\nabla_\omega f|^2, \quad \Gamma_y(f) = (\nabla_y f)^\top M_y \nabla_y f. \quad (8.35)$$

Here $\nabla_\omega f$ is the gradient on S^{k-1} of $\omega \mapsto f(\sqrt{S}\omega, y)$, and $\nabla_y f$ is the gradient of $y \mapsto f(\sqrt{N - |y|^2} \omega, y)$ at fixed ω . The marginal law of y under μ , and also under μ_+ since $\mathcal{A}$ is even in ω_1 , is

$$\mu_{\text{rem}}(dy) \propto e^{-V(y)} S^{k/2-1} Z(S) dy, \quad Z(S) = \int e^{-S\mathcal{A}} d\sigma^k = \int e^{-S\mathcal{A}} d\sigma_+^k.$$

Given y , the conditional law of ω is $\vartheta_S \propto e^{-S\mathcal{A}}\sigma^k$ under μ , and $\vartheta_S^+ = \vartheta_S(\cdot \mid \omega_1 \geq 0)$ under μ_+ . We write $\langle \mathcal{A} \rangle_S$ and $\text{Var}_{\vartheta_S}(\mathcal{A})$ for the mean and variance of $\mathcal{A}$ under ϑ_S , which are the same under ϑ_S^+ . With $\psi = \log Z$, differentiation under the integral gives $\psi'(S) = -\langle \mathcal{A} \rangle_S$ and $\psi''(S) = \text{Var}_{\vartheta_S}(\mathcal{A}) \geq 0$. For functions u of y alone, (8.35) shows that the Dirichlet form of μ restricts to $\int \Gamma_y(u) d\mu_{\text{rem}}$, whose generator on B_N is $L_{\text{rem}}u = \rho^{-1} \text{div}(\rho M_y \nabla u)$, where ρ is the density of μ_{rem} .

Lemma 8.14 (Moments of the angular law). *There exist absolute constants C_1, C_2, C_A, C_V such that the following hold.*

- (i) *For $\alpha \geq 2$ and $\tau \geq 0$, let $\nu_{\alpha, \tau}$ be the probability measure on $(0, 1)$ proportional to $v^{\alpha-1}(1-v)^{-1/2}e^{-\tau v} dv$. Then*

$$\mathbb{E}_{\nu_{\alpha, \tau}}[\tau v] \leq \alpha + C_1 \sqrt{\alpha}, \quad \mathbb{E}_{\nu_{\alpha, \tau}}[(\tau v)^2] \leq C_2 \alpha^2. \quad (8.36)$$

- (ii) *Let $a_1 = 0$ and $a_2, \dots, a_k \geq 0$, with $k \geq 5$, and let $\vartheta_S \propto e^{-S\mathcal{A}}\sigma^k$. Then for every $S > 0$,*

$$S \langle \mathcal{A} \rangle_S \leq \frac{k-1}{2} + C_1 \sqrt{\frac{k-1}{2}} \leq C_A k, \quad S^2 \text{Var}_{\vartheta_S}(\mathcal{A}) \leq S^2 \mathbb{E}_{\vartheta_S}[\mathcal{A}^2] \leq C_V k^2. \quad (8.37)$$

The same bounds hold under ϑ_S^+ .

Proof. Let $p_\alpha(w) = w^{\alpha-1}e^{-w}/\Gamma(\alpha)$ be the Gamma($\alpha, 1$) density. We first show that there exist absolute constants $c, C > 0$ such that, for all $\alpha \geq 2$,

$$\begin{aligned} p_\alpha(w) &\geq \frac{c}{\sqrt{\alpha}} && \text{for } \alpha - \frac{1}{2}\sqrt{\alpha} \leq w \leq \alpha, \\ p_\alpha(w) &\leq \frac{C}{\sqrt{\alpha}} e^{-(w-\alpha)^2/(16\alpha)} && \text{for } \alpha \leq w \leq 4\alpha, \\ p_\alpha(w) &\leq \frac{C}{\sqrt{\alpha}} e^{-w/10} && \text{for } w \geq 2\alpha. \end{aligned} \quad (8.38)$$

Stirling's formula gives $p_\alpha(\alpha) \asymp \alpha^{-1/2}$. For $w = \alpha(1+x)$, we have $\log(p_\alpha(w)/p_\alpha(\alpha)) = (\alpha-1)\log(1+x) - \alpha x$. Using $\log(1+x) \leq x - x^2/(2(1+x))$ for $x \geq 0$ and $\alpha-1 \geq \alpha/2$, we see that this is at most $-\alpha x^2/16$ for $0 \leq x \leq 3$, which proves the second bound. For $w \geq 2\alpha$, it is at most $\alpha(\log(1+x) - x) \leq -\alpha(1+x)/10 = -w/10$, because $\log y - y + 1 \leq -y/10$ for $y \geq 2$; this proves the third bound. For $-1/(2\sqrt{\alpha}) \leq x \leq 0$, the inequality $\log(1+x) \geq x - x^2/(2(1+x))$ shows that the logarithm is at least $-(\alpha-1)x^2 \geq -1/4$, which gives the first bound.

For (i), the claims are immediate if $\tau \leq \alpha$, since $\tau v \leq \tau$. Let $\tau > \alpha$. The variable $w = \tau v$ has density proportional to $p_\alpha(w)(\tau-w)^{-1/2}$ on $(0, \tau)$, with normalizing integral $H(\tau)$. Integrating the first bound in (8.38) over its interval, we find

$$H(\tau) \geq \frac{c}{2} (\tau - \alpha + \sqrt{\alpha})^{-1/2}. \quad (8.39)$$

Then $\mathbb{E}w \leq \alpha + K(\tau)/H(\tau)$ with $K(\tau) = \int_{\alpha}^{\tau} (w - \alpha)(\tau - w)^{-1/2} p_{\alpha}(w) dw$. Suppose first $\alpha < \tau \leq 4\alpha$, and let $T = (\tau - \alpha)/\sqrt{\alpha}$. Substituting $w = \alpha + \sqrt{\alpha}x$ and using the second bound in (8.38), we find

$$K(\tau) \leq C\alpha^{1/4} \int_0^T x(T-x)^{-1/2} e^{-x^2/16} dx \leq \frac{C\alpha^{1/4}}{\sqrt{T+1}}.$$

For the last inequality: if $T \leq 2$, the integral is at most $\int_0^T x(T-x)^{-1/2} dx = \frac{4}{3}T^{3/2}$; if $T > 2$, the part over $[0, T/2]$ is at most $(T/2)^{-1/2} \int_0^{\infty} xe^{-x^2/16} dx = 8(2/T)^{1/2}$, and the part over $[T/2, T]$ is at most $Te^{-T^2/64}(2T)^{1/2}$. Then (8.39) implies that $K/H \leq C\sqrt{\alpha}$, because $\tau - \alpha + \sqrt{\alpha} = \sqrt{\alpha}(T+1)$. Suppose next $\tau \geq 4\alpha$, and split K at $\tau/2 \geq 2\alpha$. On $[\alpha, \tau/2]$, we have $(\tau - w)^{-1/2} \leq (2/\tau)^{1/2}$ and $\int (w - \alpha)_+ p_{\alpha}(w) dw \leq \sqrt{\alpha}$, the standard deviation of p_{α} . On $[\tau/2, \tau]$, the third bound in (8.38) gives at most $\tau \cdot C\alpha^{-1/2}e^{-\tau/20}(2\tau)^{1/2}$. Using $H(\tau) \geq (c/2)\tau^{-1/2}$ and the boundedness of $\tau^2e^{-\tau/20}$, we again find $K/H \leq C\sqrt{\alpha}$. This proves the first bound in (8.36). For the second, suppose first $\alpha < \tau \leq 4\alpha$. The bound $w^2 \leq \alpha^2 + 2\alpha(w - \alpha)_+ + (w - \alpha)_+^2$ gives $\mathbb{E}w^2 \leq \alpha^2 + 2\alpha K(\tau)/H(\tau) + K_2(\tau)/H(\tau)$, where $K_2(\tau) = \int_{\alpha}^{\tau} (w - \alpha)^2(\tau - w)^{-1/2} p_{\alpha}(w) dw$. The same substitution and the same split, with x^2 in place of x , give $K_2(\tau) \leq C\alpha^{3/4}(T+1)^{-1/2}$. Then $K_2/H \leq C\alpha$ and $\mathbb{E}w^2 \leq C\alpha^2$. If $\tau \geq 4\alpha$, the integral of $w^2(\tau - w)^{-1/2} p_{\alpha}(w)$ over $[0, \tau/2]$ is at most $(2/\tau)^{1/2}\alpha(\alpha+1)$, and over $[\tau/2, \tau]$ it is at most $C\alpha^{-1/2}\tau^{5/2}e^{-\tau/20}$. Dividing by $H(\tau) \geq (c/2)\tau^{-1/2}$, and using that $\tau^3e^{-\tau/20}$ is bounded, we find $\mathbb{E}w^2 \leq C\alpha^2$.

For (ii), we write $\omega = (\omega_1, (1 - \omega_1^2)^{1/2}\eta)$ with η in the unit sphere of the coordinates $2, \dots, k$. Under σ^k , the variable $v = 1 - \omega_1^2$ has density proportional to $v^{(k-3)/2}(1-v)^{-1/2}$ and is independent of η , which is uniform. In these coordinates, $\mathcal{A}(\omega) = v \sum_{i \geq 2} a_i \eta_i^2$, and $S\mathcal{A} = \tau v$ with $\tau = S \sum_{i \geq 2} a_i \eta_i^2 \geq 0$. Given η , the law of v under θ_S is then $\nu_{\alpha, \tau}$ with $\alpha = (k-1)/2 \geq 2$. The bounds (8.36) are uniform in τ , and integrating them over η , we obtain (8.37); in the second bound, we bounded the variance by the second moment. The sign of ω_1 plays no role in this argument, and the same bounds hold under θ_S^+ . $\square$

Lemma 8.15 (A drift inequality for the remaining coordinates). *Assume (8.33). For $\zeta \in (0, 1/2]$, let $\tilde{a}_j = a_j - \zeta a \geq a/2$ for $j > k$, and*

$$Q_{\zeta}(y) = \sum_{j>k} \frac{y_j^2}{2\tilde{a}_j}, \quad R_{\zeta} = \frac{Q_{\zeta}}{S}, \quad D_{\zeta} = N - \sum_{j>k} \frac{1}{2\tilde{a}_j}.$$

(i) On B_N ,

$$L_{\text{rem}} R_{\zeta} = 2 - \frac{2D_{\zeta}}{S} - 4\zeta a R_{\zeta} + 4\langle \mathcal{A} \rangle_S R_{\zeta} - \frac{2(k-4)R_{\zeta}}{S}. \quad (8.40)$$

(ii) There exists $k_0(\zeta)$, depending only on ζ , such that if $k \geq k_0(\zeta)$ and

$$D_{\zeta} \geq 0, \quad (8.41)$$

then on B_N ,

$$L_{\text{rem}} R_\zeta \leq 2 - 2\zeta a R_\zeta. \quad (8.42)$$

(iii) We have $\mu_{\text{rem}}(L_{\text{rem}} R_\zeta) = 0$. For $w = (1 + a R_\zeta)^{1/2}$ and every bounded $u \in C^1(B_N)$ with $\int \Gamma_y(u) d\mu_{\text{rem}} < \infty$,

$$\int u^2 \frac{-L_{\text{rem}} w}{w} d\mu_{\text{rem}} \leq \int \Gamma_y(u) d\mu_{\text{rem}}.$$

Proof. For (i), we compute $L_{\text{rem}} u = \text{div}(M_y \nabla u) + \nabla \log \rho \cdot M_y \nabla u$, with

$$\nabla \log \rho = -2(a_j y_j)_{j>k} - 2y \left(\frac{k/2 - 1}{S} - \langle \mathcal{A} \rangle_S \right),$$

where we used $\nabla S = -2y$ and $\psi' = -\langle \mathcal{A} \rangle_S$. For a smooth u , we have $\text{div}(M_y \nabla u) = \Delta u - ((N - k + 1)y \cdot \nabla u + y^\top \nabla^2 u y)/N$. For $u = S$, we have $M_y \nabla S = -2Sy/N$ and $\text{div}(M_y \nabla S) = -2(N - k) + 2(N - k + 2)|y|^2/N$. For $u = Q_\zeta$, we have $M_y \nabla Q_\zeta = \nabla Q_\zeta - 2Q_\zeta y/N$ and $\text{div}(M_y \nabla Q_\zeta) = \sum_{j>k} \tilde{a}_j^{-1} - 2(N - k + 2)Q_\zeta/N$. We also use $\sum_{j>k} a_j y_j^2 / \tilde{a}_j = |y|^2 + 2\zeta a Q_\zeta$, $|y|^2 = N - S$ and $\sum_{j>k} \tilde{a}_j^{-1} = 2N - 2D_\zeta$. Collecting terms, we obtain, with $V_* = V + S\langle \mathcal{A} \rangle_S$,

$$L_{\text{rem}} S = 2k - 2S - 4S\langle \mathcal{A} \rangle_S + \frac{4SV_*}{N}, \quad L_{\text{rem}} Q_\zeta = 2S - 2D_\zeta - (2 + 4\zeta a)Q_\zeta + \frac{4Q_\zeta V_*}{N},$$

$$\Gamma_y(S) = 4S \left(1 - \frac{S}{N} \right), \quad \Gamma_y(Q_\zeta, S) = -\frac{4Q_\zeta S}{N}.$$

The quotient rule $L(Q/S) = LQ/S - QLS/S^2 - 2\Gamma(Q, S)/S^2 + 2Q\Gamma(S)/S^3$ then gives (8.40); the terms containing V_* and the terms of order R_ζ/N cancel.

For (ii), by (8.37), we have $4\langle \mathcal{A} \rangle_S - 2(k - 4)/S \leq (6 + C_4\sqrt{k})/S$ with $C_4 = 2\sqrt{2}C_1$. Let $L_\zeta = \max\{1, (C_4 + 6/\sqrt{5})/(2\zeta)\}$ and $k_0(\zeta) = \max\{8, \lceil 16L_\zeta^2 \rceil\}$. If $S \geq L_\zeta\sqrt{k}/a$, then $(6 + C_4\sqrt{k})/S \leq (6/\sqrt{5} + C_4)a/L_\zeta \leq 2\zeta a$. If $S < L_\zeta\sqrt{k}/a$, we use $\langle \mathcal{A} \rangle_S \leq \bar{a} \leq a$ and $2(k - 4)/(L_\zeta\sqrt{k}) \geq \sqrt{k}/L_\zeta \geq 4$, which hold for $k \geq k_0(\zeta)$; then $4\langle \mathcal{A} \rangle_S - 2(k - 4)/S \leq 4a - 4a \leq 2\zeta a$. In both cases the coefficient of R_ζ in (8.40) is at most $-2\zeta a$. Together with $D_\zeta \geq 0$, this proves (8.42).

For (iii), we fix the finite model; the constants C below may depend on it. Near $\partial B_N = \{S = 0\}$, the density ρ is at most $C S^{k/2-1}$, since $Z \leq 1$ and $V \geq 0$. We have $Q_\zeta \leq C$, and $Q_\zeta \geq c > 0$ near ∂B_N , where $|y|^2$ is close to N . Using $\nabla R_\zeta = \nabla Q_\zeta/S + 2Q_\zeta y/S^2$ and $M_y y = Sy/N$, we find $|M_y \nabla R_\zeta| \leq C/S$ and $\Gamma_y(R_\zeta) \leq C/S^3$, and (8.40) gives $|L_{\text{rem}} R_\zeta| \leq C/S^2$. Integrating $L_{\text{rem}} R_\zeta \rho = \text{div}(\rho M_y \nabla R_\zeta)$ over $\{S > \varepsilon\}$, the boundary term is at most $C\varepsilon^{k/2-1}\varepsilon^{-1} = C\varepsilon^{k/2-2}$, which tends to zero because $k \geq 5$. The function S^{-2} is μ_{rem} -integrable, because $\int_0^1 S^{-2} S^{k/2-1} dS < \infty$ for $k \geq 5$, and dominated convergence gives $\mu_{\text{rem}}(L_{\text{rem}} R_\zeta) = 0$. For $w = (1 + a R_\zeta)^{1/2}$, we have $\nabla \log w = a \nabla R_\zeta / (2(1 + a R_\zeta))$ and

$$\frac{-L_{\text{rem}} w}{w} = \frac{-a L_{\text{rem}} R_\zeta}{2(1 + a R_\zeta)} + \frac{a^2 \Gamma_y(R_\zeta)}{4(1 + a R_\zeta)^2}.$$

Near ∂B_N , we have $1 + aR_\zeta \geq c/S$, which implies that $|L_{\text{rem}}w/w|$ and $\Gamma_y(\log w)$ are at most C/S , a μ_{rem} -integrable function, and $|M_y \nabla w|/w \leq C/S$. As in Lemma 8.2(iii), on $\{S > \varepsilon\}$ we have

$$\int_{\{S > \varepsilon\}} u^2 \frac{-L_{\text{rem}}w}{w} d\mu_{\text{rem}} = \int_{\{S > \varepsilon\}} (2u\Gamma_y(u, \log w) - u^2\Gamma_y(\log w)) d\mu_{\text{rem}} + O(\varepsilon^{k/2-2}),$$

where the error is the boundary term, bounded using $\sup |u|$. The integrand on the right is at most $\Gamma_y(u)$, and it is integrable by the Cauchy–Schwarz inequality. Letting $\varepsilon \rightarrow 0$ proves (iii). $\square$

The next lemma combines a Poincaré inequality on a set Ω with an inequality of the type in Lemma 8.15(iii), which has a gain outside Ω . This is the Lyapunov-function method; see [12, Theorem 4.6.2, pp. 202–203].

Lemma 8.16 (From a local to a global Poincaré inequality). *Let μ be a probability measure and $\mathcal{E}(u) = \int \Gamma(u) d\mu$ an energy with $\Gamma(u) \geq 0$ and $\Gamma(u+c) = \Gamma(u)$ for constants c . Let Ω be measurable with $\mu(\Omega) > 0$, and let $P_{\text{loc}}, \lambda > 0$ and $B \geq 0$. Suppose that, for every u in a class of square-integrable functions of finite energy that is stable under the addition of constants,*

$$\text{Var}_{\mu(\cdot|\Omega)}(u) \leq P_{\text{loc}} \int \Gamma(u) d\mu(\cdot|\Omega), \quad \int u^2 (\lambda \mathbf{1}_{\Omega^c} - B \mathbf{1}_{\Omega}) d\mu \leq \mathcal{E}(u).$$

Then $\text{Var}_{\mu}(u) \leq [\lambda^{-1} + (1 + B/\lambda)P_{\text{loc}}]\mathcal{E}(u)$ for every such u .

Proof. Let $c = \int u d\mu(\cdot|\Omega)$ and $v = u - c$. The first hypothesis and $\Gamma \geq 0$ give $\int_{\Omega} v^2 d\mu \leq P_{\text{loc}} \int_{\Omega} \Gamma(u) d\mu \leq P_{\text{loc}} \mathcal{E}(u)$. The second hypothesis, applied to v , gives $\lambda \int_{\Omega^c} v^2 d\mu \leq \mathcal{E}(u) + B \int_{\Omega} v^2 d\mu$. Adding the two bounds, we find $\int v^2 d\mu \leq [\lambda^{-1} + (1 + B/\lambda)P_{\text{loc}}]\mathcal{E}(u)$, and $\text{Var}_{\mu}(u) \leq \int v^2 d\mu$. $\square$

If the class contains a function u with $\mu(u^2) + \mathcal{E}(u) < \infty$, then applying the second hypothesis to $u + c$, dividing by c^2 , and letting $|c| \rightarrow \infty$, we find $\lambda\mu(\Omega^c) - B\mu(\Omega) \leq 0$, that is, $\mu(\Omega) \geq \lambda/(\lambda + B)$. In particular, the lemma does not apply to sets of arbitrarily small mass with fixed λ and B . We first use it on the unit hemisphere, where it yields the angular part of the estimate.

Lemma 8.17 (Poincaré inequality on a hemisphere with a quadratic potential). *Let $k \geq 5$, let $t_2, \dots, t_k \geq t \geq 8k$, and let ϑ be the probability measure on S_+^{k-1} proportional to $\exp(-\sum_{i \geq 2} t_i \omega_i^2) \sigma_+^k$, with the reflecting form of the unit hemisphere. Then*

$$C_P(\vartheta) \leq \frac{14}{t}. \tag{8.43}$$

No upper bound on the t_i is needed.

Proof. Let $z = \omega_1^2$, $F_t(\omega) = \sum_{i \geq 2} t_i \omega_i^2 \geq t(1 - z)$, and let L_t be the generator of ϑ . On the unit sphere, we have $\Delta \omega_1^2 = 2 - 2kz$, and the tangential inner product of the gradients of F_t and z is $-4F_t z$. Then

$$L_t z = 2 - 2kz + 4F_t z, \quad \Gamma(z) = 4z(1 - z).$$

Let $\varepsilon = 4/t$ and $w = (z + \varepsilon)^{-1}$. Then

$$\frac{-L_t w}{w} = \frac{L_t z}{z + \varepsilon} - \frac{8z(1 - z)}{(z + \varepsilon)^2}.$$

On $\{z \leq 1/4\}$, we have $F_t \geq 3t/4$ and $2k \leq t/4$, so that $L_t z \geq 2 + \frac{5}{2}tz$ and

$$\frac{-L_t w}{w} \geq \frac{2\varepsilon - 6z}{(z + \varepsilon)^2} + \frac{\frac{5}{2}tz}{z + \varepsilon} \geq \frac{t}{2}.$$

For the last inequality, write $s = tz$; after multiplying by $(s + 4)^2/t$, the difference of the two sides is $8 - 6s + \frac{5}{2}s(s + 4) - \frac{1}{2}(s + 4)^2 = 2s^2 \geq 0$. Everywhere, dropping $F_t \geq 0$ and using $z/(z + \varepsilon)^2 \leq 1/(4\varepsilon)$, we have $-L_t w/w \geq -2k - t/2 \geq -t$. The function w depends on ω_1 only through ω_1^2 , and its normal derivative on the boundary $\{\omega_1 = 0\}$ vanishes. The argument of Lemma 8.2(iii) then shows that $\int u^2(-L_t w/w) d\vartheta \leq \int \Gamma(u) d\vartheta$ for every smooth u on the closed hemisphere. This is the second hypothesis of Lemma 8.16, with $\Omega = \{z > 1/4\}$, $\lambda = t/2$ and $B = t$.

On Ω , we use the coordinates $v = (\omega_2, \dots, \omega_k)$ in the ball $\{|v|^2 < 3/4\}$, with $\omega_1 = (1 - |v|^2)^{1/2}$. In these coordinates, the uniform measure has density proportional to $z^{-1/2}$, the metric is $I + vv^\top/z$, and its inverse is $I - vv^\top \succeq zI \succeq \frac{1}{4}I$. The density of ϑ is proportional to e^{-U} with $U(v) = \sum_{i \geq 2} t_i v_i^2 + \frac{1}{2} \log(1 - |v|^2)$, and

$$\nabla^2 U = 2 \operatorname{diag}(t_i) - \frac{I}{z} - \frac{2vv^\top}{z^2} \succeq (2t - 4 - 24)I \succeq tI,$$

since $z \geq 1/4$, $|v|^2 \leq 3/4$ and $t \geq 8k \geq 40$. A smooth spherical test is smooth on a neighborhood of the closure of this chart ball and, after multiplication by a cutoff outside that neighborhood, admits an ambient smooth extension. By Lemma 8.2(v), $\operatorname{Var}_{\vartheta(\cdot|\Omega)}(u) \leq t^{-1} \int |\nabla_v u|^2 d\vartheta(\cdot|\Omega) \leq 4t^{-1} \int \Gamma(u) d\vartheta(\cdot|\Omega)$. Lemma 8.16 with $P_{\text{loc}} = 4/t$ gives $C_P(\vartheta) \leq 2/t + 3 \cdot 4/t = 14/t$. $\square$

Lemma 8.18 (Poincaré inequality on the hemisphere with a few small precisions). *Assume (8.33), and suppose that (8.41) holds for some $\zeta \in (0, 1/2]$. There exist $k_0(\zeta)$ and $C(\zeta) < \infty$, depending only on ζ , such that if $k \geq k_0(\zeta)$, then*

$$C_P(\mu_+) \leq C(\zeta) \exp \left\{ C(\zeta) k^2 + \frac{8k\bar{a}}{a_2} \right\} \left(\frac{k}{a} + \frac{1}{a_2} \right). \quad (8.44)$$

The same bound holds for the coordinate-even sector.

The lemma assumes no upper bound on the coefficients a_j , $j > k$. It does not control the full sphere: functions that are odd in x_1 are treated in Lemma 8.19.

Proof. We decompose the variance of f into the variance of its conditional mean given y and the conditional variance given y . The marginal law μ_{rem} of y satisfies a Poincaré inequality by Lemma 8.16: on a region where S is not small, a convexity argument applies after a bounded modification of the density (Steps 2 and 3), and Lemma 8.15 controls the complement of this region (Step 4). The angular law is controlled by Lemma 8.17 and a bounded perturbation (Step 1). Step 5 combines the two, using an exact formula for the gradient of the conditional mean. Table 8.1 lists the constants, in the order in which they are chosen. By density, it suffices to consider f smooth on the closed hemisphere $\{x_1 \geq 0\}$.

Table 8.1. Constants in the proof of Lemma 8.18, in the order of their choice.

Constant	Choice	Depends on
C_A, C_V	Lemma 8.14	nothing
$L_\zeta, k_0(\zeta)$	Lemma 8.15(ii)	ζ
K_0	$2/\zeta$	ζ
C_T	$\max\{1, 4C_A/\zeta, 32C_V K_0\}$	ζ
λ_0, T	$a/(8K_0), C_T k^2/a$	ζ, k, a
Θ	$\exp(8k\bar{a}/a_2)$	$k, \bar{a}/a_2$

Step 1: the angular law. By Lemma 8.14(ii) and $\langle \mathcal{A} \rangle_S \leq \bar{a}$,

$$\langle \mathcal{A} \rangle_S \leq \min \left\{ \bar{a}, \frac{C_A k}{S} \right\}, \quad \text{Var}_{\vartheta_S}(\mathcal{A}) \leq \frac{C_V k^2}{S^2}. \quad (8.45)$$

Let $\Theta = \exp(8k\bar{a}/a_2)$. If $Sa_2 \geq 8k$, Lemma 8.17 with $t_i = Sa_i$ and $t = Sa_2$ implies that $C_P(\vartheta_S^+) \leq 14/(Sa_2)$. If $Sa_2 < 8k$, then $S\mathcal{A}$ has oscillation at most $S\bar{a} \leq \log \Theta$, and Lemma 8.2(i),(vi) give $C_P(\vartheta_S^+) \leq \Theta/(k-1)$. In both cases, for every $S > 0$ and every smooth g on the closed unit hemisphere,

$$\text{Var}_{\vartheta_S^+}(g) \leq \frac{2\Theta}{k} \int |\nabla_\omega g|^2 d\vartheta_S^+, \quad \text{Var}_{\vartheta_S^+}(g) \leq \frac{14\Theta}{a_2} \int S^{-1} |\nabla_\omega g|^2 d\vartheta_S^+. \quad (8.46)$$

For the first bound we used $14/(Sa_2) \leq 14/(8k)$ and $1/(k-1) \leq 2/k$; for the second, $\Theta/(k-1) \leq (8k/(Sa_2))\Theta/(k-1) \leq 10\Theta/(Sa_2)$ when $Sa_2 < 8k$ and $k \geq 5$.

Step 2: a modification of the marginal density. Recall that $\psi = \log Z$ is convex with $-\bar{a} \leq \psi' \leq 0$. Let $T = C_T k^2/a$, and let $\tilde{\psi}$ be the C^1 convex function equal to ψ on $[T, \infty)$ and to its tangent line at T on $[0, T]$; if $T > N$, only the tangent line is used on $[0, N]$. For

$S < T$, convexity implies that $\psi(S) \geq \tilde{\psi}(S)$, and $\psi(S) - \psi(T) \leq \bar{a}(T - S)$ together with $\psi'(T) \leq 0$ gives $\psi(S) - \tilde{\psi}(S) \leq \bar{a}T$. We conclude that, on the whole interval $[0, N]$,

$$0 \leq \psi - \tilde{\psi} \leq \bar{a}T \leq C_T k^2. \quad (8.47)$$

Let $\tilde{\mu}_{\text{rem}}$ be the probability measure on B_N proportional to $e^{-\tilde{U}} dy$, where $\tilde{U} = V - (\frac{k}{2} - 1) \log S - \tilde{\psi}(S)$. The density of μ_{rem} with respect to $\tilde{\mu}_{\text{rem}}$ is proportional to $e^{\psi - \tilde{\psi}}$.

Step 3: a local inequality for the marginal. Let $K_0 = 2/\zeta$ and $\lambda_0 = a/(8K_0) = \zeta a/16$, and let

$$\Omega = \{R_\zeta < K_0/a\} = \left\{ \sum_{j>k} \left(\frac{1}{2\tilde{a}_j} + \frac{K_0}{a} \right) y_j^2 < \frac{K_0 N}{a} \right\},$$

with R_ζ , Q_ζ and $\tilde{a}_j$ as in Lemma 8.15. This is a nonempty convex ellipsoid with smooth boundary, and its closure lies in B_N , since $Q_\zeta \leq K_0 S/a$ on $\bar{\Omega}$ forces $S > 0$. Similarly, $\Omega' = \{R_\zeta < 2K_0/a\}$ is an open neighborhood of $\bar{\Omega}$ whose closure is a compact subset of B_N . The function $\tilde{U}$ belongs to $C_{\text{loc}}^{1,1}(B_N)$, and in particular to $C^{1,1}(\Omega')$. Almost everywhere,

$$\nabla^2 \tilde{U} = 2 \text{diag}(a_j)_{j>k} + \left(\frac{k-2}{S} + 2\tilde{\psi}' \right) I + \left(\frac{2(k-2)}{S^2} - 4\tilde{\psi}'' \right) yy^\top. \quad (8.48)$$

We claim that $\nabla^2 \tilde{U} \succeq \lambda_0 M_y^{-1}$ almost everywhere on Ω' , where $M_y^{-1} = I + yy^\top/S$. First, $\tilde{\psi}' \geq -C_A k/T$ on $(0, N]$ by (8.45), and $\tilde{\psi}'' = 0$ below T while $0 \leq \tilde{\psi}'' \leq C_V k^2/S^2$ above T . The diagonal part of $\nabla^2 \tilde{U} - \lambda_0 M_y^{-1}$ is

$$\Lambda_y = 2 \text{diag}(a_j) + \left(\frac{k-2}{S} + 2\tilde{\psi}' - \lambda_0 \right) I \succeq 2 \text{diag}(\tilde{a}_j),$$

because $2C_A k/T + \lambda_0 \leq \zeta a/(2k) + \zeta a/16 \leq 2\zeta a$ by the choice of C_T . The remaining part is $\xi yy^\top$ with $\xi = 2(k-2)/S^2 - 4\tilde{\psi}'' - \lambda_0/S$. If $\xi \geq 0$, there is nothing to prove. If $\xi < 0$, then $\Lambda_y + \xi yy^\top \succeq 0$ as soon as $|\xi| y^\top \Lambda_y^{-1} y \leq 1$, by the Cauchy–Schwarz inequality $(y^\top e)^2 \leq (y^\top \Lambda_y^{-1} y)(e^\top \Lambda_y e)$. We have $y^\top \Lambda_y^{-1} y \leq Q_\zeta$ and $|\xi| \leq \lambda_0/S + 4\tilde{\psi}''$. On Ω' with $S \leq T$, this gives $|\xi| y^\top \Lambda_y^{-1} y \leq \lambda_0 R_\zeta < 2\lambda_0 K_0/a = 1/4$. On Ω' with $S > T$, it gives at most $(\lambda_0 + 4C_V k^2/T) R_\zeta < 1/4 + 8C_V K_0/C_T \leq 1/2$. This proves the claim. The matrix field M_y^{-1} is continuous and positive definite on B_N . Applying Lemma 8.2(v) with $\Lambda = \lambda_0 M_y^{-1}$, we obtain, for ambient C^1 tests u ,

$$\text{Var}_{\tilde{\mu}_{\text{rem}}(\cdot|\Omega)}(u) \leq \frac{1}{\lambda_0} \int (\nabla u)^\top M_y \nabla u d\tilde{\mu}_{\text{rem}}(\cdot|\Omega) = \frac{8K_0}{a} \int \Gamma_y(u) d\tilde{\mu}_{\text{rem}}(\cdot|\Omega).$$

By (8.47) and Lemma 8.2(i), applied to the conditioned measures,

$$\text{Var}_{\mu_{\text{rem}}(\cdot|\Omega)}(u) \leq P_{\text{loc}} \int \Gamma_y(u) d\mu_{\text{rem}}(\cdot|\Omega), \quad P_{\text{loc}} = \frac{8K_0}{a} e^{C_T k^2}. \quad (8.49)$$

Step 4: the Poincaré inequality for the marginal. Let $w = (1 + aR_\zeta)^{1/2}$. By (8.42) and $\Gamma_y(R_\zeta) \geq 0$, we have

$$\frac{-L_{\text{rem}} w}{w} \geq \frac{-aL_{\text{rem}} R_\zeta}{2(1 + aR_\zeta)} \geq a \frac{\zeta a R_\zeta - 1}{1 + aR_\zeta}.$$

The right side is at least $-a$ everywhere. It is increasing in aR_ζ , and for $aR_\zeta \geq K_0 = 2/\zeta$ it is at least $a\zeta/(\zeta + 2) \geq \zeta a/4$. With $\lambda = \zeta a/4$, this means that

$$\frac{-L_{\text{rem}} w}{w} \geq \lambda \mathbf{1}_{\Omega^c} - a \mathbf{1}_\Omega. \quad (8.50)$$

By Lemma 8.15(iii), the second hypothesis of Lemma 8.16 holds with $B = a$ for every bounded $u \in C^1(B_N)$ with finite energy, and (8.49) is the first hypothesis. Applying Lemma 8.16, we obtain, for every such u ,

$$\text{Var}_{\mu_{\text{rem}}}(u) \leq P \int \Gamma_y(u) d\mu_{\text{rem}}, \quad P = \frac{1}{\lambda} + \left(1 + \frac{a}{\lambda}\right) P_{\text{loc}} \leq \frac{C'(\zeta)}{a} e^{C_T k^2}, \quad (8.51)$$

where $C'(\zeta)$ depends only on ζ .

Step 5: the conditional mean and the conclusion. For $y \in B_N$, let $m(y) = \int f(\sqrt{S}\omega, y) \vartheta_S^+(d\omega)$. Differentiating the conditional law, whose density is proportional to $e^{-S\mathcal{A}}$, and using $\nabla S = -2y$, we find

$$\nabla_y m = \int \nabla_y f d\vartheta_S^+ + 2y \text{Cov}_{\vartheta_S^+}(f, \mathcal{A}).$$

By Jensen's inequality for the quadratic form M_y , the Cauchy–Schwarz inequality, $y^\top M_y y = |y|^2 S/N \leq S$, (8.45), and the first bound in (8.46),

$$\begin{aligned} \Gamma_y(m) &\leq 2 \int \Gamma_y(f) d\vartheta_S^+ + 8S \text{Var}_{\vartheta_S^+}(\mathcal{A}) \text{Var}_{\vartheta_S^+}(f) \\ &\leq 2 \int \Gamma_y(f) d\vartheta_S^+ + 16C_V k \Theta \int S^{-1} |\nabla_\omega f|^2 d\vartheta_S^+. \end{aligned} \quad (8.52)$$

The function m is bounded and C^1 on B_N , and $\Gamma_y(f) + S^{-1} |\nabla_\omega f|^2 = \Gamma(f)$ is bounded. Further, m has finite energy by (8.52), and (8.51) applies to it. By the conditional variance formula, the second bound in (8.46), (8.51) and (8.52),

$$\begin{aligned} \text{Var}_{\mu_+}(f) &= \int \text{Var}_{\vartheta_S^+}(f) d\mu_{\text{rem}} + \text{Var}_{\mu_{\text{rem}}}(m) \\ &\leq 2P \int \Gamma_y(f) d\mu_+ + \left(\frac{14\Theta}{a_2} + 16C_V k \Theta P\right) \int S^{-1} |\nabla_\omega f|^2 d\mu_+. \end{aligned}$$

By (8.35), the two integrals add up to $\mathcal{E}_{\mu_+}(f)$. Using $\Theta \geq 1$, $2 + 16C_V k \leq (2 + 16C_V)k$ and the bound on P in (8.51), we see that the right side is at most $((2 + 16C_V)C'(\zeta)\Theta e^{C_T k^2} k/a + 14\Theta/a_2)\mathcal{E}_{\mu_+}(f)$. This proves (8.44) with $C(\zeta) = \max\{C_T, 14, (2 + 16C_V)C'(\zeta)\}$. Finally, since μ_+ is invariant under $R_2, \dots, R_N$, the argument of Lemma 8.1 bounds the constant of the coordinate-even sector by $C_P(\mu_+)$. $\square$

Lemma 8.19 (From the hemisphere to the full sphere). *Assume the hypotheses of Lemma 8.18, and suppose in addition that*

$$D_0 = N - \sum_{j>k} \frac{1}{2a_j} \leq \frac{M_0 k}{a}$$

for some $M_0 \geq 0$. Let $M_1 = M_0 + 1/\zeta$ and $T_1 = 2M_1 k/a$. Then

$$C_P(\mu) \leq 5C_P(\mu_+) + \frac{4T_1 e^{\bar{a}T_1}}{k-1}.$$

Proof. We first show that $\mu(S \leq T_1) \geq 1/2$. Let $R_0 = \sum_{j>k} y_j^2 / (2a_j S)$. For $j > k$, we have $\partial_j \log \rho = -2a_j y_j - (k-2)y_j/S + 2\langle \mathcal{A} \rangle_S y_j$, and the divergence of the vector field $\rho(y)y_j e_j / (2a_j)$ is $\rho(1 + y_j \partial_j \log \rho) / (2a_j)$. We integrate it over $\{S > \varepsilon\}$. The boundary term is $O(\varepsilon^{k/2-1})$, and the integrand is $O(1/S)$ near ∂B_N , which is integrable. Letting $\varepsilon \rightarrow 0$ and summing over $j > k$, we obtain

$$\sum_{j>k} \frac{1}{2a_j} = \mathbb{E}_{\mu_{\text{rem}}}[N - S] + (k-2)\mathbb{E}_{\mu_{\text{rem}}}[R_0] - 2\mathbb{E}_{\mu_{\text{rem}}}[S\langle \mathcal{A} \rangle_S R_0],$$

so that $\mathbb{E}S \leq D_0 + (k-2)\mathbb{E}R_0$. We have $R_0 \leq R_\zeta$, because $a_j \geq \bar{a}_j$, and Lemma 8.15(iii) and (8.42) give $0 = \mathbb{E}[L_{\text{rem}} R_\zeta] \leq 2 - 2\zeta a \mathbb{E}R_\zeta$. We obtain $\mathbb{E}R_0 \leq 1/(\zeta a)$ and $\mathbb{E}S \leq M_0 k/a + k/(\zeta a) = M_1 k/a$. By Markov's inequality, the set $B_0 = \{S \leq T_1\}$ satisfies $\mu(B_0) \geq 1/2$.

For $S \leq T_1$, the conditional law ϑ_S on the full unit sphere has potential $S\mathcal{A}$ with oscillation at most $\bar{a}T_1$. By Lemma 8.2(i),(vi), for every smooth g on S^{k-1} ,

$$\text{Var}_{\vartheta_S}(g) \leq \frac{e^{\bar{a}T_1}}{k-1} \int |\nabla_\omega g|^2 d\vartheta_S \leq J_1 \int S^{-1} |\nabla_\omega g|^2 d\vartheta_S, \quad J_1 = \frac{T_1 e^{\bar{a}T_1}}{k-1}.$$

Let f be smooth on S_N , and let f_e and f_o be its even and odd parts under R_1 . Since μ and Γ are invariant under R_1 , the cross terms are odd and integrate to zero, so that $\text{Var}_\mu(f) = \text{Var}_\mu(f_e) + \mu(f_o^2)$ and $\mathcal{E}_\mu(f) = \mathcal{E}_\mu(f_e) + \mathcal{E}_\mu(f_o)$. As in Lemma 8.1, $\text{Var}_\mu(f_e) \leq C_P(\mu_+) \mathcal{E}_\mu(f_e)$. For the odd part, for each y , the function $\omega \mapsto f_o(\sqrt{S}\omega, y)$ is odd in ω_1 , and its mean under ϑ_S vanishes. By the last display and (8.35), we have

$$\int_{B_0} f_o^2 d\mu = \int_{\{S \leq T_1\}} \text{Var}_{\vartheta_S}(f_o) d\mu_{\text{rem}} \leq J_1 \mathcal{E}_\mu(f_o).$$

Let $m = \mu_+(f_o)$. The restriction of f_o to the hemisphere is a function of the reflecting form, with no boundary condition, and f_o^2 , $\Gamma(f_o)$ and $\mathbf{1}_{B_0}$ are even. Using $\mu_+(B_0) = \mu(B_0) \geq 1/2$, we find

$$\frac{m^2}{2} \leq m^2 \mu_+(B_0) \leq 2 \int_{B_0} f_o^2 d\mu_+ + 2 \int_{B_0} (f_o - m)^2 d\mu_+ \leq 2J_1 \mathcal{E}_\mu(f_o) + 2C_P(\mu_+) \mathcal{E}_\mu(f_o).$$

We conclude that

$$\mu(f_o^2) = \mu_+(f_o^2) = \text{Var}_{\mu_+}(f_o) + m^2 \leq (5C_P(\mu_+) + 4J_1)\mathcal{E}_\mu(f_o).$$

Combining the two parts proves the lemma. Note that the indicator of B_0 is only integrated, never differentiated. $\square$

8.6 The critical window

We now prove the Poincaré and logarithmic Sobolev inequalities used in the critical window. In this section, $r_1 \leq r_2 \leq \dots \leq r_N$ are positive half-precisions with $\sum_i (2r_i)^{-1} = N$, and π_r is the probability measure proportional to $e^{-\sum_i r_i x_i^2} v_N$. We abbreviate $b = N^{2/3}$. In the application to the GOE, $r_i = (z_N - c_N \lambda_i)/2$ with $c_N > 0$ and $\lambda_1 \geq \dots \geq \lambda_N$. These half-precisions are ordered, and on the events of Lemma 4.17 they satisfy (4.36).

Lemma 8.20 (Poincaré and logarithmic Sobolev constants for half-precisions of order $(i/N)^{2/3}$). *Let $0 < c_* \leq C_*$ and $\delta_* > 0$. There exist N_0 and C , depending only on c_* , C_* and δ_* , such that the following holds for $N \geq N_0$. Suppose that $0 < r_1 \leq \dots \leq r_N$ satisfy $\sum_i (2r_i)^{-1} = N$ and*

$$c_* \left(\frac{i}{N}\right)^{2/3} \leq r_i \leq C_* \left(\frac{i}{N}\right)^{2/3} \quad (1 \leq i \leq N), \quad r_2 - r_1 \geq \frac{\delta_*}{b}. \quad (8.53)$$

Then $C_P(\pi_r) \leq Cb$ and $C_{LS}(\pi_r) \leq Cb$. The same bounds hold for every coordinate hemisphere with its reflecting form and for the coordinate-even sector.

Proof. Let $a_i = r_i - r_1$, so that π_r is the measure μ of Section 8.5 with these coefficients, since a common shift of the coefficients does not change it. We use Lemmas 8.18 and 8.19 with a fixed number J of coordinates in place of k . The parameters are chosen in the order of Table 8.2.

Table 8.2. Parameters in the proof of Lemma 8.20, in the order of their choice.

Parameter	Choice	Depends on
ζ	$\min\{1/4, c_*^2/(96C_*^2)\}$	c_*, C_*
J	smallest integer with $c_*(J+1)^{2/3} \geq 2C_*$, $J^{2/3} \geq 48C_*^2/c_*^2$ and $J \geq k_0(\zeta)$	c_*, C_*
N_0	$J+1$	c_*, C_*

We take $k = J$, $\bar{a} = a_J$ and $a = a_{J+1}$. By the ordering, we have $a_2 \leq a_i \leq \bar{a} \leq a \leq a_j$ for $2 \leq i \leq J < j$, and $a_2 = r_2 - r_1 \geq \delta_*/b > 0$. This is (8.33). From (8.53) and the choice of J ,

$$\frac{c_*(J+1)^{2/3}}{2b} \leq \frac{c_*(J+1)^{2/3} - C_*}{b} \leq a \leq r_{J+1} \leq \frac{2C_*J^{2/3}}{b}, \quad \bar{a} \leq r_J \leq \frac{C_*J^{2/3}}{b}.$$

We check (8.41). By the choice of ζ and $c_* \leq C_*$, we have $\zeta a \leq c_* J^{2/3}/(48b)$, and $r_1 \leq C_*/b \leq c_*(J+1)^{2/3}/(2b)$. For $j > J$, it follows that $r_1 + \zeta a \leq \frac{3}{4}c_*(J+1)^{2/3}/b \leq \frac{3}{4}r_j$, that is, $\tilde{a}_j = r_j - r_1 - \zeta a \geq r_j/4$. By the trace condition $\sum_i (2r_i)^{-1} = N$, we have

$$D_\zeta = \sum_{i \leq J} \frac{1}{2r_i} - \sum_{j > J} \frac{r_1 + \zeta a}{2r_j \tilde{a}_j}.$$

The first sum is at least $J/(2r_J) \geq bJ^{1/3}/(2C_*)$. Using $\sum_{j > J} j^{-4/3} \leq 3J^{-1/3}$ and $r_1 + \zeta a \leq (C_*/b)(1 + 2\zeta J^{2/3})$, we see that the second sum is at most

$$2(r_1 + \zeta a) \sum_{j > J} \frac{1}{r_j^2} \leq \frac{6C_*b}{c_*^2} (J^{-1/3} + 2\zeta J^{1/3}) \leq \frac{bJ^{1/3}}{4C_*},$$

where the last inequality uses $J^{2/3} \geq 48C_*^2/c_*^2$ for the first term and $\zeta \leq c_*^2/(96C_*^2)$ for the second. Then $D_\zeta > 0$. Moreover, since $a_j \leq r_j$ and $\tilde{a}_j \leq a_j$,

$$0 \leq D_\zeta \leq D_0 \leq \sum_{i \leq J} \frac{1}{2r_i} \leq \frac{b}{2c_*} \sum_{i \leq J} i^{-2/3} \leq \frac{3bJ^{1/3}}{2c_*} \leq \frac{3C_*}{c_*} \frac{J}{a},$$

so the hypothesis of Lemma 8.19 holds with $M_0 = 3C_*/c_*$.

We have $8J\bar{a}/a_2 \leq 8C_*J^{5/3}/\delta_*$, $J/a \leq 2J^{1/3}b/c_*$ and $1/a_2 \leq b/\delta_*$, and Lemma 8.18 implies that $C_P(\pi_{r,+}) \leq Cb$, where $\pi_{r,+} = \pi_r(\cdot \mid x_1 \geq 0)$. In Lemma 8.19, we have $T_1 = 2M_1J/a \leq 4M_1J^{1/3}b/c_*$ and $\bar{a}T_1 \leq 4M_1C_*J/c_*$, and the lemma implies that $C_P(\pi_r) \leq Cb$. For the entropy, the variances $v_i = 1/(2r_i)$ satisfy $\max_i v_i = v_1 \leq b/(2c_*)$ and $v_i/v_1 = r_1/r_i \geq c_*/(3^{2/3}C_*)$ for $i \leq 3$. Corollary 8.13 implies that $C_{LS}(\pi_r) \leq C(v_1 + C_P(\pi_r)) \leq Cb$. The bounds for hemispheres and for the coordinate-even sector follow from Lemma 8.1. $\square$

We now state the result in the form used later. Part (ii) concerns the Bingham comparison law of Definition 7.5 for the matrix with half-precisions r_i . With $H = -4 \operatorname{diag}(r_1, \dots, r_N) + \frac{2}{N} \mathbf{1}\mathbf{1}^\top$ and $h(t) = e^{tH} \mathbf{1}$, as in (7.3) with $p_i = 2r_i$, its Gaussian variance parameters are, by (7.4),

$$v_i(t) = \frac{h_i(t)}{\bar{h}(t)}, \quad \bar{h}(t) = \frac{1}{N} \sum_j h_j(t),$$

and $\mu_t = \mu_{v(t)}$ is proportional to $e^{-\sum_i r_i(t)x_i^2} \nu_N$ with $r_i(t) = 1/(2v_i(t)) = \bar{h}(t)/(2h_i(t))$. Since $\sum_i v_i(t) = N$, these are the half-precisions of μ_t in the sense of Section 8.1, and $\mu_0 = \nu_N$. We emphasize that μ_t is a comparison law; in general, it need not coincide with the law $\nu_N P_t$ of the diffusion.

Proposition 8.21 (Poincaré and logarithmic Sobolev constants in the critical window). *Let $r_1 \leq \dots \leq r_N$ satisfy (4.36) with constants c_0, C_0, δ , that is, $\sum_i (2r_i)^{-1} = N$, $c_0(i/N)^{2/3} \leq r_i \leq C_0(i/N)^{2/3}$ for all i , and $r_2 - r_1 \geq \delta N^{-2/3}$.*

(i) *There exist N_0 and C , depending only on c_0, C_0 and δ , such that for $N \geq N_0$,*

$$C_P(\pi_r) \leq CN^{2/3}, \quad C_{LS}(\pi_r) \leq CN^{2/3}.$$

(ii) *For every $K < \infty$, there exist N_0 and C_K , depending only on K, c_0, C_0 and δ , such that for $N \geq N_0$ and $0 \leq t \leq KN^{2/3}$,*

$$C_P(\mu_t) + C_{LS}(\mu_t) \leq C_K(1+t). \quad (8.54)$$

These bounds hold for the full sphere, for every coordinate hemisphere with its reflecting form, and for the coordinate-even sector.

Proof. Part (i) is Lemma 8.20 with $(c_*, C_*, \delta_*) = (c_0, C_0, \delta)$. For part (ii), we may enlarge the time horizon by replacing K with $\max\{K, 1\}$, and hence assume $K \geq 1$. The hypotheses (7.49) of Lemma 7.14 hold; that lemma uses only the trace condition and the two-sided bounds on the r_i , for every $N \geq 2$. It gives constants $0 < c_K \leq C_K$, depending only on K, c_0 and C_0 , such that for $0 \leq t \leq Kb$ we have $\bar{h}(t) \asymp (1+t)^{-1/2}$ and $h_i(t) \asymp (1+t)^{1/2}/(1+r_it)$, with these constants. After adjusting c_K and C_K , we find

$$c_K \frac{1+r_it}{1+t} \leq r_i(t) \leq C_K \frac{1+r_it}{1+t} \quad (1 \leq i \leq N, 0 \leq t \leq Kb). \quad (8.55)$$

The coordinates of h solve $h'_i = -4r_i h_i + 2\bar{h}$ with $h_i(0) = 1$. By variation of constants, $h_i(t) = \mathfrak{h}_t(r_i)$ and $r_i(t) = \mathfrak{r}_t(r_i)$, where

$$\mathfrak{h}_t(r) = e^{-4rt} + 2 \int_0^t e^{-4r\tau} \bar{h}(t-\tau) d\tau, \quad \mathfrak{r}_t(r) = \frac{\bar{h}(t)}{2\mathfrak{h}_t(r)}.$$

The function $\mathfrak{h}_t$ is decreasing and $\mathfrak{r}_t$ is increasing. Hence $r_1(t) \leq \dots \leq r_N(t)$. The parameters are chosen as follows: c_K and C_K as above; $L = 2C_K/c_K$; k the smallest integer with $k-2 \geq 64L^2$; $\varepsilon_0 = \min\{1, 1/(C_0 k^{2/3})\}$; and finally N_0 , as specified below. All of them depend only on K, c_0, C_0 and δ .

Early times, $0 \leq t \leq \varepsilon_0 b$. For $i \leq k$, we have $r_it \leq C_0(k/N)^{2/3} \varepsilon_0 b \leq 1$, and (8.55) implies that $r_i(t) \leq 2C_K/(1+t) \leq Lr_1(t)$. Also, the trace condition and $r_i(t) \leq Lr_1(t)$ for $i \leq k$ give

$$N = \sum_{i=1}^N \frac{1}{2r_i(t)} \geq \frac{k}{2Lr_1(t)}, \quad \text{that is,} \quad Nr_1(t) \geq \frac{k}{2L} \geq 4,$$

since $k \geq 64L^2 + 2 \geq 8L$. Corollary 8.9, in its form with half-precisions and with $K = \{1, \dots, k\}$, gives $C_P(\mu_t) \leq 1/r_1(t) \leq (1+t)/c_K$. For $i \leq 3$, the variances satisfy $v_i(t) = 1/(2r_i(t)) \geq v_1(t)/L$, and $\max_i v_i(t) = v_1(t) \leq (1+t)/(2c_K)$. We conclude from Corollary 8.13 that $C_{LS}(\mu_t) \leq C(1+t)$.

Late times, $\varepsilon_0 b \leq t \leq Kb$. We take N large enough that $\varepsilon_0 b \geq 1$. Then (8.55) gives $r_i(t) \geq c_K r_i t / (1+t) \geq (c_K/2) r_i \geq (c_K c_0/2) (i/N)^{2/3}$. Also, since $1/t \leq 1/(\varepsilon_0 b) \leq \varepsilon_0^{-1} (i/N)^{2/3}$, we have $r_i(t) \leq C_K(1/t + r_i) \leq C_K(\varepsilon_0^{-1} + C_0)(i/N)^{2/3}$. It remains to bound $r_2(t) - r_1(t)$ from below. Differentiating $\mathfrak{r}_t$, we find

$$\mathfrak{r}'_t(r) = \frac{4\mathfrak{r}_t(r)}{\mathfrak{h}_t(r)} \left(t e^{-4rt} + 2 \int_0^t \tau e^{-4r\tau} \bar{h}(t-\tau) d\tau \right).$$

Let $r \in [r_1, r_2]$. Then $r \leq 2^{2/3} C_0/b$ and $4r\tau \leq 2^{5/3} C_0 K$ for $\tau \leq t/2$. Restricting the integral to $\tau \in [t/4, t/2]$ and using $\bar{h}(t-\tau) \geq c_K(1+t)^{-1/2}$, we see that the bracket is at least $c t^2(1+t)^{-1/2}$, with $c > 0$ depending only on K , c_0 and C_0 . Also $\mathfrak{h}_t(r) \leq \mathfrak{h}_t(r_1) = h_1(t) \leq C_K(1+t)^{1/2}$ and $\mathfrak{r}_t(r) \geq \mathfrak{r}_t(r_1) = r_1(t) \geq c_K/(1+t)$. Combining these bounds, we find $\mathfrak{r}'_t(r) \geq c' t^2/(1+t)^2 \geq c'/4$, since $t \geq 1$. By the mean value theorem, $r_2(t) - r_1(t) \geq (c'/4)(r_2 - r_1) \geq c'\delta/(4b)$. We have shown that the half-precisions $r_i(t)$ satisfy (8.53), and the trace condition, with

$$c_* = \frac{c_K c_0}{2}, \quad C_* = C_K(\varepsilon_0^{-1} + C_0), \quad \delta_* = \frac{c'\delta}{4},$$

which depend only on K , c_0 , C_0 and δ . For $N \geq N_0(c_*, C_*, \delta_*)$, Lemma 8.20 gives $C_P(\mu_t) + C_{LS}(\mu_t) \leq Cb \leq (C/\varepsilon_0)t$.

Taking N_0 to be the largest of the thresholds above proves (8.54). The bounds for hemispheres and for the coordinate-even sector follow from Lemma 8.1. $\square$

The restriction $N \geq N_0$ is used only to ensure that the fixed numbers J and k of eigencoordinates are smaller than N and that $\varepsilon_0 N^{2/3} \geq 1$; in the applications $N \rightarrow \infty$.

Comparison with Bingham laws through the forward equation

In this section, we compare the time- t law $\nu_N P_t^A$ of the Langevin diffusion started from the uniform distribution with the Bingham comparison law μ_t of Section 7. The law μ_t is built from the covariance evolution of the linear dynamics, whose scalar normalization goes back to Cugliandolo and Dean [51, pp. 8–9, (II.2)–(II.7)]. In general, its density does not solve the forward (Fokker–Planck) equation of L_A , but it solves this equation up to an explicit error term, which we call the forward-equation defect (Definition 9.2). In Lemma 9.3, we show that the defect is $-1/N$ times a centered product of two quadratic forms. Duhamel’s formula, tested against the backward evolution $P_{t-u}^A f$ of a bounded function f , then bounds $\|\nu_N P_t^A - \mu_t\|_{\text{TV}}$ by an integral over $u \in [0, t]$ of the size of the defect at time u , measured against the gradient of $P_{t-u}^A f$. This is Theorem 9.5. It has two forms. In the first, the defect is measured in $L^2(\mu_u)$, and a Poincaré inequality for μ_u bounds the variance of $P_{t-u}^A f$ by its Dirichlet energy. In the second, the defect is written as a divergence by solving a Poisson equation for the generator of μ_u , and only the energy of the solution enters.

The section contains three applications. Theorem 9.13 bounds $\|\nu_N P_t^A - \mu_t\|_{\text{TV}}$ by $C_\Lambda d^{-2} e^{-rt}$ when $\|A\| \leq \Lambda$ and Nd^2 is large; it is used at fixed high temperature in Section 13. In Proposition 9.14, we obtain a bound of order ϵ^2 , uniformly in time, when A lies within ϵ of a multiple of the identity; this bound is used for the infinite-temperature crossover in Section 15. Theorem 9.16 bounds the energy of the solution of the Poisson equation $-L_B \psi = R$ for a Bingham law whose variances satisfy explicit conditions at a scale δ . Together with the second form of Theorem 9.5 and Corollary 9.7, it is used in Section 14 near the critical temperature, where Nd^2 may tend to zero.

We use the covariance evolution of Section 7 (Lemmas 7.6 and 7.7); the moment bounds and the bounds on marginal densities for Gaussian laws conditioned on the sphere from Section 5 (Theorems 5.11 and 5.15); and, from Section 8, the weighted gradient estimate of Theorem 8.6 and the lower bound on the spectral gap in Theorem 8.8.

In this section, we compare probability measures on S_N , and the time evolution is always given by the semigroup P_t^A of the diffusion. The numbers $v_i(t)$ are the variance parameters of μ_t as a conditioned Gaussian law. They need not equal the second moments of either μ_t or $\nu_N P_t^A$ (Remark 9.4). Accordingly, the quadratic forms below are centered

at these parameters, and the only centering by an actual expectation is the one in the defect itself.

9.1 Notation

Throughout this section, we fix $N \geq 2$ and a real symmetric $N \times N$ matrix A . We work in an orthonormal eigenbasis of A , with coordinates $x_1, \dots, x_N$, so that $A = \text{diag}(\alpha_i)$ with $\alpha_1 \geq \dots \geq \alpha_N$; all matrices below are diagonal in this basis. For a symmetric matrix D we write $\Phi_D(x) = x^\top D x / 2$. As in Section 3, $L_A = \Delta_{S_N} + \nabla_{S_N} \Phi_A \cdot \nabla_{S_N}$, $P_t^A = e^{tL_A}$ and $\pi_A \propto e^{\Phi_A} \nu_N$, where Δ_{S_N} is the Laplace–Beltrami operator of S_N and ∇_{S_N} is the gradient along S_N . We write $\Gamma(f, g) = \langle \nabla_{S_N} f, \nabla_{S_N} g \rangle$ and $\Gamma(f) = |\nabla_{S_N} f|^2$. Adding a multiple of the identity to D changes Φ_D only by a constant on S_N ; it changes neither L_D nor π_D . Recall that π_D is reversible for L_D , that is, $\pi_D(f(-L_D g)) = \pi_D \Gamma(f, g)$, and that a positive lower bound $\lambda > 0$ on the spectral gap $\lambda_{\text{gap}}(L_D)$, the smallest positive eigenvalue of $-L_D$, is equivalent to the Poincaré inequality $\text{Var}_{\pi_D}(f) \leq \lambda^{-1} \pi_D \Gamma(f)$; see Section 3.1. For a probability measure μ and $p \geq 1$, we denote by $\|\cdot\|_p$ the norm of $L^p(\mu)$ when μ is clear from the context.

We use two formulas of spherical calculus. Let $P_x = I - x x^\top / N$ be the orthogonal projection onto the tangent space of S_N at x , and let $u_D = x^\top D x$. For symmetric matrices D and D' , we have $\nabla_{S_N} \Phi_D(x) = P_x D x$ and

$$\Delta_{S_N} \Phi_D = \text{Tr } D - u_D, \quad \Gamma(\Phi_D, \Phi_{D'}) = x^\top D D' x - \frac{u_D u_{D'}}{N}. \quad (9.1)$$

The second formula follows from $\nabla_{S_N} \Phi_D = P_x D x$ and $P_x^2 = P_x$. The first follows from the formula $\Delta_{S_N} f = \text{Tr}(P_x D^2 f) - (N-1)N^{-1}x \cdot D f$ for the Laplace–Beltrami operator of the sphere of radius $\sqrt{N}$, applied to $f = \Phi_D$, for which $D^2 f = D$ and $x \cdot D f = u_D$.

Recall that the Lagrange multiplier of A is the unique $z > \alpha_1$ such that $\sum_i (z - \alpha_i)^{-1} = N$. We write $p_i = z - \alpha_i$ for the precisions, $q_i = 1/p_i$ for the variances, $Q = \text{diag}(q_i)$, $d = p_1$, $s = p_N$ and $\kappa = s/d$, and we set

$$\vartheta = \frac{N}{\sum_i q_i^2}, \quad \varphi(x) = \frac{x^\top Q x}{N} = \frac{1}{N} \sum_i q_i x_i^2. \quad (9.2)$$

Using $\sum_i q_i = N$ and $q_i \leq 1/d$, we have $N \leq \sum_i q_i^2 \leq N/d$, where the lower bound is the Cauchy–Schwarz inequality. Thus $d \leq \vartheta \leq 1$. Also $1/s \leq \varphi \leq 1/d$ on S_N . The function φ is the weight of Lemma 8.5, and (9.3) below is (8.10). Let

$$r_0 = d - \frac{\kappa + \kappa^{-1}}{N}, \quad r = d - \frac{\kappa + 1 + \kappa^{-1}}{N}.$$

Recall from Theorem 8.6 that, for all $t \geq 0$ and smooth f ,

$$\varphi \Gamma(P_t^A f) \leq e^{-2r_0 t} P_t^A (\varphi \Gamma(f)), \quad \Gamma(P_t^A f) \leq \kappa e^{-2r_0 t} P_t^A \Gamma(f), \quad (9.3)$$

and that $\lambda_{\text{gap}}(L_A) \geq r_0$ when $r_0 > 0$. Both inequalities in (9.3) also hold with r in place of r_0 , because $r \leq r_0$. We define the numbers z , d , s , κ and r_0 in the same way for any symmetric matrix in place of A , in particular for the diagonal matrices B below.

Gaussian laws conditioned on the sphere. For a positive sequence $v = (v_i)$ with $\sum_i v_i = N$, we write μ_v for the law of a vector with independent $\mathcal{N}(0, v_i)$ coordinates conditioned on $|x|^2 = N$ ((5.10)). It is the Bingham law with density proportional to $\exp(-\sum_i x_i^2 / (2v_i))$ with respect to ν_N . In particular, $\pi_A = \mu_q$, and $\mu_v = \pi_B$ for every matrix of the form $B = yI - \text{diag}(v_i^{-1})$ with $y \in \mathbb{R}$. Note that the Lagrange multiplier of such a matrix B is y , because $\sum_i (y - b_i)^{-1} = \sum_i v_i = N$, and that its precisions are the numbers v_i^{-1} . As in Section 5, $S_v = (2\sum_i v_i^2)^{1/2}$ and $\eta_v = \max_i v_i / S_v$. By the Cauchy–Schwarz inequality, $\sum_i v_i^2 \geq N^{-1}(\sum_i v_i)^2 = N$, which gives

$$\eta_v \leq \frac{\max_i v_i}{\sqrt{2N}}. \quad (9.4)$$

The covariance evolution. We now recall the objects of Section 7. Let $H = -2 \text{diag}(p_i) + (2/N)\mathbf{1}\mathbf{1}^\top$ and $h(t) = e^{tH}\mathbf{1}$, and define

$$\begin{aligned} \bar{h}(t) &= \frac{1}{N} \sum_i h_i(t), & v_i(t) &= \frac{h_i(t)}{\bar{h}(t)}, & w_i(t) &= p_i h_i(t), \\ z(t) &= 1 + \frac{1}{N} \sum_i \alpha_i v_i(t), & b_i(t) &= z(t) - \frac{1}{v_i(t)}. \end{aligned} \quad (9.5)$$

Let $B_t = \text{diag}(b_i(t))$ and $C_t = A - B_t = \text{diag}(C_i(t))$, so that $C_i(t) = \alpha_i - b_i(t)$. The Bingham comparison law is

$$\mu_t(dx) = Z_t^{-1} e^{\Phi_{B_t}(x)} \nu_N(dx), \quad Z_t = \nu_N(e^{\Phi_{B_t}}).$$

Note that $\Phi_{B_t}(x) = Nz(t)/2 - \sum_i x_i^2 / (2v_i(t))$ on S_N . Therefore $\mu_t = \mu_{v(t)}$, and $z(t)$ is the Lagrange multiplier of B_t . The functions v_i , z and b_i are smooth on $[0, \infty)$. We use the following facts from Lemmas 7.6 and 7.7: for all $t \geq 0$ and all i ,

$$\sum_i v_i(t) = N, \quad v_i(0) = 1, \quad \frac{1}{s} \leq v_i(t) \leq \frac{1}{d}, \quad (9.6)$$

$$v_i'(t) = 2 + 2(\alpha_i - z(t))v_i(t), \quad \sum_i (\alpha_i - b_i(t))v_i(t) = 0, \quad (9.7)$$

$$b_i'(t) = z'(t) + \frac{2(\alpha_i - b_i(t))}{v_i(t)},$$

$$\sum_i q_i h_i(t) = N, \quad \vartheta \leq \bar{h}(t) \leq 1, \quad \text{osc } w(t) = \max_i w_i(t) - \min_i w_i(t) \leq \Omega_t, \quad (9.8)$$

where $\Omega_t = (s - d)e^{-2dt}$ and $\vartheta = N / \sum_i q_i^2$ is the weight of the zero eigenvalue of H (see (9.2)). The first identity in (9.8) is the conserved quantity $\sum_i q_i^2 w_i(t) = N$ of Lemma 7.7. In particular, $\mu_0 = \nu_N$, since $v(0) = (1, \dots, 1)$.

Let $e_i(t) = p_i v_i(t) - 1 = v_i(t)/q_i - 1$ be the relative errors of the variances, so that $\mathcal{I}_N(t) = \mathcal{I}(q, v(t)) = \sum_i e_i(t)^2/2$ is the information of Definition 3.5. Two consequences of (9.8) are used repeatedly. Dividing the first identity in (9.8) by $\bar{h}(t)$, we have

$$\sum_i q_i v_i(t) = \frac{N}{\bar{h}(t)}. \quad (9.9)$$

Further, $\bar{h}(t) = N^{-1} \sum_j q_j w_j(t)$ is an average of the numbers $w_j(t)$ with the weights q_j/N , which sum to one. It therefore lies between their minimum and their maximum, and $|w_i(t) - \bar{h}(t)| \leq \Omega_t$ for every i . As $e_i(t) = (w_i(t) - \bar{h}(t))/\bar{h}(t)$, we obtain

$$|e_i(t)| \leq \frac{\Omega_t}{\bar{h}(t)} \leq \frac{(s-d)e^{-2dt}}{\vartheta}. \quad (9.10)$$

We record some elementary properties of the Lagrange multiplier.

Lemma 9.1 (The Lagrange multiplier of a bounded or nearly isotropic matrix). *We have $0 < d \leq 1 \leq s$ and $s - d = \alpha_1 - \alpha_N$. In particular, if $\|A\| \leq \Lambda$, then $s \leq 1 + 2\Lambda$. Let $\bar{\alpha} = N^{-1} \text{Tr } A$ and $\epsilon = \|A - \bar{\alpha}I\|$. If $\epsilon \leq 1/4$, then for every i ,*

$$|z - 1 - \bar{\alpha}| \leq 8\epsilon^2, \quad |q_i - 1 - (\alpha_i - \bar{\alpha})| \leq 8\epsilon^2, \quad |d - 1| \leq 2\epsilon, \quad |s - 1| \leq 2\epsilon.$$

Proof. The numbers q_i have average one. Then $\min_i q_i \leq 1 \leq \max_i q_i$, and $d = 1/\max_i q_i \leq 1 \leq 1/\min_i q_i = s$. Also $d > 0$ because $z > \alpha_1$. The identity $s - d = \alpha_1 - \alpha_N$ is immediate, and $\alpha_1 - \alpha_N \leq 2\|A\|$.

For the second part, replacing A by $A - \bar{\alpha}I$ replaces z by $z - \bar{\alpha}$ and leaves the numbers p_i , q_i , d and s unchanged. We may therefore assume that $\bar{\alpha} = 0$, so that $\sum_i \alpha_i = 0$ and $|\alpha_i| \leq \epsilon$. Let $y = z - 1$ and $t_i = y - \alpha_i$, so that $p_i = 1 + t_i$. If $y > \epsilon$, then $p_i > 1$ for every i and $\sum_i q_i < N$; if $y < -\epsilon$, then $p_i < 1$ for every i and $\sum_i q_i > N$. We conclude that $|y| \leq \epsilon$ and $|t_i| \leq 2\epsilon \leq 1/2$. Using $(1+t)^{-1} = 1 - t + t^2/(1+t)$ and $\sum_i t_i = Ny$, the average of the q_i is

$$1 = \frac{1}{N} \sum_i \frac{1}{1+t_i} = 1 - y + \frac{1}{N} \sum_i \frac{t_i^2}{1+t_i}.$$

Each term $t_i^2/(1+t_i)$ lies in $[0, (2\epsilon)^2/(1/2)] = [0, 8\epsilon^2]$, which gives $0 \leq y \leq 8\epsilon^2$. From the same expansion, we have $q_i - 1 - \alpha_i = -y + t_i^2/(1+t_i)$, which is the difference of two numbers in $[0, 8\epsilon^2]$. Finally, $s - d \leq 2\epsilon$ and $d \leq 1 \leq s$ imply $1 - 2\epsilon \leq d \leq 1 \leq s \leq 1 + 2\epsilon$. $\square$

9.2 The forward-equation defect

Definition 9.2 (Forward-equation defect). Let $(\mu_u)_{u \geq 0}$ be a family of probability measures on S_N with densities $g_u = d\mu_u/d\nu_N$ such that $(u, x) \mapsto g_u(x)$ is smooth and positive on

$[0, \infty) \times S_N$. Let L_A^* be the adjoint of L_A in $L^2(\nu_N)$, that is, $L_A^*g = \Delta_{S_N}g - \operatorname{div}(g\nabla_{S_N}\Phi_A)$, so that $\nu_N(fL_A^*g) = \nu_N(gL_Af)$ for smooth f and g . The forward-equation defect of the family (relative to L_A) is

$$R_u = \partial_u \log g_u - \frac{L_A^*g_u}{g_u}.$$

Equivalently, $\partial_u g_u = L_A^*g_u + R_u g_u$, which we also write as

$$\partial_u \mu_u = L_A^* \mu_u + R_u \mu_u. \quad (9.11)$$

The defect is smooth on $[0, \infty) \times S_N$, and it has mean zero:

$$\mu_u(R_u) = \nu_N(\partial_u g_u) - \nu_N(L_A^*g_u) = \partial_u \nu_N(g_u) - \nu_N(g_u L_A 1) = 0.$$

If $\mu_u = \mu_0 P_u^A$ for all u , then g_u solves the forward equation $\partial_u g_u = L_A^*g_u$, and the defect vanishes identically.

Lemma 9.3 (The defect of the Bingham comparison law). *For $t \geq 0$, define the quadratic forms*

$$Q_B(t) = x^\top B_t x - N(z(t) - 1), \quad Q_C(t) = x^\top C_t x. \quad (9.12)$$

The forward-equation defect of the Bingham comparison laws $(\mu_t)_{t \geq 0}$ is

$$R_t = -\frac{Q_B(t)Q_C(t) - \mu_t(Q_B(t)Q_C(t))}{N}. \quad (9.13)$$

Proof. Fix $t \geq 0$. In this proof we write $B = B_t$, $C = C_t$, $g = g_t = e^{\Phi_B}/Z_t$, $\zeta = z(t)$, $\zeta' = z'(t)$, $v_i = v_i(t)$, $b_i = b_i(t)$ and $C_i = C_i(t)$.

We begin with the adjoint. Using $\operatorname{div}(g\nabla_{S_N}\Phi_A) = \Gamma(g, \Phi_A) + g\Delta_{S_N}\Phi_A$, we have $L_A^*g = \Delta_{S_N}g - \Gamma(g, \Phi_A) - g\Delta_{S_N}\Phi_A$. For $g = e^{\Phi_B}/Z_t$, we also have $\Delta_{S_N}g = g(\Delta_{S_N}\Phi_B + \Gamma(\Phi_B))$ and $\Gamma(g, \Phi_A) = g\Gamma(\Phi_B, \Phi_A)$. Recalling that $C = A - B$, we obtain

$$\frac{L_A^*g}{g} = \Delta_{S_N}\Phi_B - \Delta_{S_N}\Phi_A + \Gamma(\Phi_B, \Phi_B) - \Gamma(\Phi_B, \Phi_A) = -\Delta_{S_N}\Phi_C - \Gamma(\Phi_B, \Phi_C),$$

and then (9.1) implies that

$$\frac{L_A^*g}{g} = -\operatorname{Tr} C + u_C - x^\top B C x + \frac{u_B u_C}{N}. \quad (9.14)$$

Next, we compute the time derivative. We have $\partial_t \log g = \partial_t \Phi_{B_t} - \mu_t(\partial_t \Phi_{B_t})$, because $\log g = \Phi_{B_t} - \log Z_t$ and $\partial_t \log Z_t = \mu_t(\partial_t \Phi_{B_t})$. By (9.7) and $1/v_i = \zeta - b_i$, we have $b'_i = \zeta' + 2C_i(\zeta - b_i)$. Using $|x|^2 = N$, we find

$$\partial_t \Phi_{B_t} = \frac{1}{2} \sum_i b'_i x_i^2 = \frac{N\zeta'}{2} + \zeta u_C - x^\top B C x.$$

The constant $N\zeta'/2$ cancels against its mean, and we are left with

$$\partial_t \log g = \zeta u_C - x^\top BCx - \mu_t(\zeta u_C - x^\top BCx). \quad (9.15)$$

We subtract (9.14) from (9.15). The terms $x^\top BCx$ cancel, and we find

$$R_t = (\zeta - 1)u_C - \frac{u_B u_C}{N} + c_t = -\frac{Q_B(t)Q_C(t)}{N} + c_t,$$

where $c_t = \text{Tr } C - \mu_t(\zeta u_C - x^\top BCx)$ does not depend on x . In the second equality, we used $u_B - N(\zeta - 1) = Q_B(t)$ and $u_C = Q_C(t)$. Finally, $\mu_t(R_t) = 0$ implies that $c_t = \mu_t(Q_B(t)Q_C(t))/N$, and (9.13) follows. $\square$

We record two other expressions for the forms in (9.12). On S_N we have $\sum_i (x_i^2 - v_i(t)) = 0$, by (9.6). By the definition of $b_i(t)$, we have $\sum_i b_i(t)v_i(t) = N(z(t) - 1)$, and $\sum_i C_i(t)v_i(t) = 0$ by (9.7). Combining these identities, we have

$$Q_B(t) = \sum_i b_i(t)(x_i^2 - v_i(t)), \quad Q_C(t) = \sum_i C_i(t)(x_i^2 - v_i(t)).$$

We subtract $z(t) \sum_i (x_i^2 - v_i(t)) = 0$ from the first identity and add $(z(t) - z) \sum_i (x_i^2 - v_i(t)) = 0$ to the second. Recalling that $b_i(t) - z(t) = -1/v_i(t)$ and noting that $C_i(t) + z(t) - z = 1/v_i(t) - p_i$, we find

$$Q_B(t) = -\sum_i \left(\frac{x_i^2}{v_i(t)} - 1 \right), \quad Q_C(t) = -\sum_i e_i(t) \left(\frac{x_i^2}{v_i(t)} - 1 \right). \quad (9.16)$$

Both forms are linear combinations of the variables $x_i^2/v_i(t) - 1$, to which Theorem 5.15 applies under $\mu_t = \mu_{v(t)}$ whenever $\eta_{v(t)} \leq \eta_0$, where η_0 is the universal threshold of that theorem. The coefficient vector of $Q_B(t)$ has Euclidean norm $\sqrt{N}$, and that of $Q_C(t)$ has Euclidean norm $(\sum_i e_i(t)^2)^{1/2} = (2\mathcal{I}_N(t))^{1/2}$.

Remark 9.4. The variance parameters $v_i(t)$ need not equal the second moments of either μ_t or $\nu_N P_t^A$. For μ_t , we have $\mu_t(x_i^2) = v_i(t)(1 + O(\eta_{v(t)}^{1/2}))$ by Theorem 5.12 when $\eta_{v(t)}$ is small. For the time- t law, the first derivatives at $t = 0$ differ whenever $\alpha_i \neq \bar{\alpha}$, where $\bar{\alpha} = N^{-1} \text{Tr } A$; thus the vectors of first derivatives differ unless A is a multiple of the identity. Indeed, by (9.1) we have $L_A x_i^2 = 2 - 2x_i^2 + 2\alpha_i x_i^2 - 2u_A x_i^2/N$, and $\nu_N(x_i^2 x_j^2) = N(1 + 2 \cdot \mathbf{1}_{i=j})/(N+2)$. We find

$$\left. \frac{d}{dt} \right|_{t=0} \nu_N P_t^A(x_i^2) = \nu_N(L_A x_i^2) = \frac{2N}{N+2}(\alpha_i - \bar{\alpha}), \quad \text{whereas} \quad v'_i(0) = 2(\alpha_i - \bar{\alpha}).$$

For this reason, the forms (9.12) and (9.16) need not have mean zero under μ_t , and none of the estimates below replaces an expectation under μ_t or $\nu_N P_t^A$ by its value at the parameters $v_i(t)$.

9.3 Comparison through the defect

The following theorem compares μ_t with $\mu_0 P_t^A$, the law at time t of the diffusion started from μ_0 . The argument uses the classical backward-semigroup interpolation underlying Duhamel formulas; compare [12, Section 3.1, pp. 130–131]. The forward-equation defect and resulting estimates are computed below.

Theorem 9.5 (Comparison through the forward-equation defect). *Let $(\mu_u)_{u \geq 0}$ be a family of probability measures as in Definition 9.2, with forward-equation defect R_u . For every $t \geq 0$ and every smooth f ,*

$$\mu_t(f) - \mu_0 P_t^A(f) = \int_0^t \mu_u(R_u P_{t-u}^A f) du. \quad (9.17)$$

Fix $t > 0$. All functions $\mathcal{A}$, $\mathcal{C}$, $\mathcal{G}$ and $\mathcal{J}$ below are Borel measurable.

(a) Suppose that $\mathcal{A}: (0, t) \rightarrow [0, \infty)$ satisfies $\|R_u\|_{L^2(\mu_u)} \leq \mathcal{A}(u)$ for $u \in (0, t)$. Then

$$\|\mu_0 P_t^A - \mu_t\|_{\text{TV}} \leq \min\left\{1, \frac{1}{2} \int_0^t \mathcal{A}(u) du\right\}.$$

(b) Suppose, in addition to the hypothesis of (a), that $\mathcal{C}: (0, t) \rightarrow (0, \infty)$ and $\mathcal{G}: (0, t) \times (0, t) \rightarrow [0, \infty)$ satisfy, for all $u, \tau \in (0, t)$, all smooth g , and all smooth f with $0 \leq f \leq 1$,

$$\text{Var}_{\mu_u}(g) \leq \mathcal{C}(u) \mu_u \Gamma(g), \quad \mu_u \Gamma(P_\tau^A f) \leq \mathcal{G}(u, \tau). \quad (9.18)$$

Then

$$\|\mu_0 P_t^A - \mu_t\|_{\text{TV}} \leq \min\left\{1, \int_0^t \mathcal{A}(u) \min\left(\frac{1}{2}, \sqrt{\mathcal{C}(u) \mathcal{G}(u, t-u)}\right) du\right\}. \quad (9.19)$$

(c) Suppose that $\mathcal{G}: (0, t) \times (0, t) \rightarrow [0, \infty)$ satisfies the second inequality in (9.18), and that for each $u \in (0, t)$ there exists a smooth tangent vector field j_u on S_N such that, for all smooth g ,

$$\mu_u(R_u g) = \mu_u \langle j_u, \nabla_{S_N} g \rangle, \quad \mu_u |j_u|^2 \leq \mathcal{J}(u), \quad (9.20)$$

where $\mathcal{J}: (0, t) \rightarrow [0, \infty)$. Then

$$\|\mu_0 P_t^A - \mu_t\|_{\text{TV}} \leq \min\left\{1, \int_0^t \sqrt{\mathcal{J}(u) \mathcal{G}(u, t-u)} du\right\}. \quad (9.21)$$

If the hypothesis of (a) also holds, the integrand in (9.21) may be replaced by $\min\{\mathcal{A}(u)/2, \sqrt{\mathcal{J}(u) \mathcal{G}(u, t-u)}\}$. If the hypotheses of (a) and (b) also hold, it may be replaced by

$$\min\left\{\frac{\mathcal{A}(u)}{2}, \mathcal{A}(u) \sqrt{\mathcal{C}(u) \mathcal{G}(u, t-u)}, \sqrt{\mathcal{J}(u) \mathcal{G}(u, t-u)}\right\}. \quad (9.22)$$

Part (a) uses no gradient bound, and part (c) uses neither a bound on $\|R_u\|_{L^2(\mu_u)}$ nor a Poincaré inequality for μ_u . The proof uses the smoothness of the family, the diffusion property of L_A and the compactness of S_N .

Proof. We begin with (9.17). Both sides vanish for $t = 0$. Let $t > 0$ and let f be smooth. As L_A is an elliptic operator with smooth coefficients on the compact manifold S_N , the function $(s, x) \mapsto P_s^A f(x)$ is smooth on $[0, \infty) \times S_N$, and $\partial_s P_s^A f = L_A P_s^A f$. Then $\Theta(u) = \mu_u(P_{t-u}^A f) = \nu_N(g_u P_{t-u}^A f)$ is continuously differentiable on $[0, t]$, and by Definition 9.2 we have

$$\Theta'(u) = \nu_N((L_A^* g_u + R_u g_u) P_{t-u}^A f) - \nu_N(g_u L_A P_{t-u}^A f) = \mu_u(R_u P_{t-u}^A f).$$

Integrating over $[0, t]$ and noting that $\Theta(t) = \mu_t(f)$ and $\Theta(0) = \mu_0 P_t^A(f)$, we obtain (9.17).

Next, let f be smooth with $0 \leq f \leq 1$, fix $u \in (0, t)$, and let $g = P_{t-u}^A f$. Then g is smooth and takes values in $[0, 1]$. Using $\mu_u(R_u) = 0$ and the Cauchy–Schwarz inequality, we have

$$|\mu_u(R_u g)| = |\mu_u(R_u(g - \mu_u g))| \leq \|R_u\|_{L^2(\mu_u)} \text{Var}_{\mu_u}(g)^{1/2}.$$

As g takes values in $[0, 1]$, we have $\text{Var}_{\mu_u}(g) \leq \mu_u(g - 1/2)^2 \leq 1/4$. Under the hypotheses of (b), we also have $\text{Var}_{\mu_u}(g) \leq \mathcal{C}(u) \mu_u \Gamma(P_{t-u}^A f) \leq \mathcal{C}(u) \mathcal{G}(u, t - u)$. Under the hypotheses of (c), we have instead, by the Cauchy–Schwarz inequality in $L^2(\mu_u)$,

$$|\mu_u(R_u g)| = |\mu_u \langle j_u, \nabla_{S_N} g \rangle| \leq (\mu_u |j_u|^2 \mu_u \Gamma(g))^{1/2} \leq \sqrt{\mathcal{J}(u) \mathcal{G}(u, t - u)}.$$

Each of these bounds uses only its own hypotheses. We bound $|\mu_u(R_u P_{t-u}^A f)|$ by the minimum of the available bounds and integrate in u . Using (9.17), we find that $|\mu_t(f) - \mu_0 P_t^A(f)|$ is at most the integral in the corresponding statement.

It remains to pass from smooth functions to the total variation distance. Let μ and ρ be Borel probability measures on S_N , let E be a Borel set, and let $\varepsilon > 0$. By the regularity of the finite measure $\mu + \rho$ on the compact metric space S_N , there exist a compact set K and an open set O such that $K \subset E \subset O$ and $(\mu + \rho)(O \setminus K) < \varepsilon$. Let f be a smooth function such that $\mathbf{1}_K \leq f \leq \mathbf{1}_O$. Then $|\mu(E) - \mu(f)| \leq \mu(O \setminus K)$, and similarly for ρ . It follows that $|\mu(E) - \rho(E)| \leq |\mu(f) - \rho(f)| + \varepsilon$, and $\|\mu - \rho\|_{\text{TV}}$ is the supremum of $|\mu(f) - \rho(f)|$ over smooth f with $0 \leq f \leq 1$. Applying this with $\mu = \mu_t$ and $\rho = \mu_0 P_t^A$, and using that the total variation distance is at most one, we obtain each of the stated bounds. $\square$

The flux hypothesis (9.20) is verified through a Poisson equation. This uses the standard spectral inverse on mean-zero functions [12, Section 1.7.2, pp. 31–32]; the energy estimate is included below.

Lemma 9.6 (The Poisson equation for a Bingham law). *Let B be a symmetric matrix, let $\mu = \pi_B$, and let $\lambda_{\text{gap}}(L_B) > 0$ be the smallest positive eigenvalue of $-L_B$. For every smooth R with $\mu(R) = 0$, there exists a unique smooth ψ with $\mu(\psi) = 0$ and $-L_B\psi = R$. It satisfies $\mu(Rg) = \mu\Gamma(\psi, g)$ for every smooth g , and*

$$\mu\Gamma(\psi) \leq \lambda_{\text{gap}}(L_B)^{-1} \|R\|_{L^2(\mu)}^2.$$

Proof. The operator $-L_B$ is symmetric and nonnegative in $L^2(\mu)$, with $\mu(f(-L_B g)) = \mu\Gamma(f, g)$. As an elliptic operator with smooth coefficients on the compact connected manifold S_N , it has discrete spectrum and smooth eigenfunctions, and its kernel consists of the constants. In particular, $-L_B$ is invertible on the orthogonal complement of the constants in $L^2(\mu)$, with inverse of norm at most $\lambda_{\text{gap}}(L_B)^{-1}$, and the solution ψ of $-L_B\psi = R$ with $\mu(\psi) = 0$ exists. It is smooth by elliptic regularity, since R is smooth. Two such solutions differ by a constant with mean zero, which must vanish. For smooth g , we have $\mu(Rg) = \mu(g(-L_B\psi)) = \mu\Gamma(\psi, g)$. Taking $g = \psi$ and using the Poincaré inequality $\|\psi\|_{L^2(\mu)}^2 \leq \lambda_{\text{gap}}(L_B)^{-1} \mu\Gamma(\psi)$, we find

$$\mu\Gamma(\psi) = \mu(R\psi) \leq \|R\|_{L^2(\mu)} (\lambda_{\text{gap}}(L_B)^{-1} \mu\Gamma(\psi))^{1/2},$$

which gives the energy bound. □

Corollary 9.7 (The defect of the Bingham comparison law as a divergence). *For $u \geq 0$, let R_u be the forward-equation defect of the Bingham comparison laws (Lemma 9.3), and let ψ_u be the smooth solution of $-L_{B_u}\psi_u = R_u$ with $\mu_u(\psi_u) = 0$. Then $\mu_u(R_u g) = \mu_u\Gamma(\psi_u, g)$ for every smooth g . In particular, the family (μ_u) satisfies (9.20) with $j_u = \nabla_{S_N}\psi_u$ and any nonnegative Borel measurable function $\mathcal{J}$ satisfying $\mathcal{J}(u) \geq \mu_u\Gamma(\psi_u)$.*

Proof. Note that $\mu_u = \pi_{B_u}$ and $\mu_u(R_u) = 0$. The claim follows from Lemma 9.6 with $B = B_u$ and $R = R_u$. □

9.4 Reverse variance inequalities and exponential decay

To use Theorem 9.5, we need bounds on the energy of $P_\tau^A f$ for bounded f . They follow from gradient estimates of the form (9.3) by the following reverse variance inequalities. The proof uses the semigroup interpolation method of the Bakry–Émery theory [11]; related local Poincaré inequalities are discussed in [12, Section 4.7, (4.7.3), p. 208]. We prove below the precise forms used here, including the constant K and the weight F .

Lemma 9.8 (Reverse variance inequalities). *Let $t > 0$, let f be smooth, and write $V_t(f) = P_t^A(f^2) - (P_t^A f)^2$.*

- (a) Suppose that a smooth positive function F on S_N and constants $D > 0$ and $\rho > 0$ satisfy $F \leq D$ and $F\Gamma(P_s^A f') \leq e^{-2\rho s} P_s^A(F\Gamma(f'))$ for all $s \geq 0$ and smooth f' . Then

$$\Gamma(P_t^A f) \leq \frac{D\rho}{F(e^{2\rho t} - 1)} V_t(f). \quad (9.23)$$

In particular, for every probability measure μ on S_N and every smooth f with $0 \leq f \leq 1$,

$$\mu\Gamma(P_t^A f) \leq \frac{D\rho \mu(F^{-1})}{4(e^{2\rho t} - 1)}. \quad (9.24)$$

- (b) Suppose that constants $K \geq 1$ and $\rho > 0$ satisfy $\Gamma(P_s^A f') \leq K e^{-2\rho s} P_s^A \Gamma(f')$ for all $s \geq 0$ and smooth f' . Then

$$\Gamma(P_t^A f) \leq \frac{K\rho}{e^{2\rho t} - 1} V_t(f). \quad (9.25)$$

In particular, if $0 \leq f \leq 1$, then

$$\sup_{S_N} \Gamma(P_t^A f) \leq \frac{K\rho}{4(e^{2\rho t} - 1)}. \quad (9.26)$$

Proof. For $0 \leq s \leq t$, let $h_s = P_{t-s}^A f$. Using the diffusion property $L_A(h^2) = 2hL_A h + 2\Gamma(h)$, we find

$$\frac{d}{ds} P_s^A(h_s^2) = P_s^A(L_A(h_s^2) - 2h_s L_A h_s) = 2P_s^A \Gamma(h_s).$$

Integrating over $s \in [0, t]$, we obtain

$$V_t(f) = 2 \int_0^t P_s^A \Gamma(P_{t-s}^A f) ds. \quad (9.27)$$

Under the hypotheses of (a), using $F \leq D$, the positivity of P_s^A and the hypothesis with $f' = h_s$, we find

$$P_s^A \Gamma(h_s) \geq D^{-1} P_s^A (F\Gamma(h_s)) \geq D^{-1} e^{2\rho s} F\Gamma(P_s^A h_s) = D^{-1} e^{2\rho s} F\Gamma(P_t^A f).$$

Inserting this into (9.27) and using $2 \int_0^t e^{2\rho s} ds = (e^{2\rho t} - 1)/\rho$, we obtain (9.23). If $0 \leq f \leq 1$, then $V_t(f)(x)$ is the variance of f under the transition probability $P_t^A(x, \cdot)$, which is at most $1/4$. Integrating (9.23) against μ then yields (9.24). The proof of (b) is the same, with $F \equiv 1$, $D = 1$ and the factor K in place of D . $\square$

Lemma 9.9 (An elementary integral). For $\rho > 0$, $a \geq 2\rho$ and $t \geq 0$,

$$\int_0^t \frac{e^{-(a-\rho)u}}{\sqrt{1 - e^{-2\rho(t-u)}}} du \leq \frac{\sqrt{1 - e^{-2\rho t}}}{\rho} \leq \frac{1}{\rho}, \quad (9.28)$$

with equality in the first inequality when $a = 2\rho$.

Proof. As $a - \rho \geq \rho$, the integrand is at most $e^{-\rho u}(1 - e^{-2\rho(t-u)})^{-1/2}$, with equality when $a = 2\rho$. Substituting $\tau = t - u$, we have

$$\int_0^t \frac{e^{-\rho u} du}{\sqrt{1 - e^{-2\rho(t-u)}}} = e^{-\rho t} \int_0^t \frac{e^{2\rho\tau} d\tau}{\sqrt{e^{2\rho\tau} - 1}} = \frac{e^{-\rho t}}{\rho} \sqrt{e^{2\rho t} - 1} = \frac{\sqrt{1 - e^{-2\rho t}}}{\rho},$$

where we used that $\rho^{-1}(e^{2\rho\tau} - 1)^{1/2}$ is a primitive of $e^{2\rho\tau}(e^{2\rho\tau} - 1)^{-1/2}$. The singularity of the integrand at $\tau = 0$ is of order $\tau^{-1/2}$, hence integrable. $\square$

Theorem 9.10 (Comparison under an exponentially decaying defect). *Let $(\mu_u)_{u \geq 0}$ be as in Definition 9.2, with forward-equation defect R_u . Suppose that there exist constants $\mathcal{A} \geq 0$ and $a > 0$ such that*

$$\|R_u\|_{L^2(\mu_u)} \leq \mathcal{A}e^{-au} \quad (u \geq 0). \quad (9.29)$$

Then, for every $t \geq 0$,

$$\|\mu_0 P_t^A - \mu_t\|_{\text{TV}} \leq \frac{\mathcal{A}}{2a}(1 - e^{-at}) \leq \frac{\mathcal{A}}{2} \min(t, a^{-1}). \quad (9.30)$$

Suppose in addition that there exist constants $\mathcal{C} > 0$, $K \geq 1$ and $\rho \in (0, a/2]$ such that $\text{Var}_{\mu_u}(g) \leq \mathcal{C}\mu_u\Gamma(g)$ for all $u \geq 0$ and smooth g , and $\Gamma(P_s^A f) \leq Ke^{-2\rho s}P_s^A\Gamma(f)$ for all $s \geq 0$ and smooth f . Then, for every $t \geq 0$,

$$\|\mu_0 P_t^A - \mu_t\|_{\text{TV}} \leq \frac{\mathcal{A}}{2} \sqrt{\frac{\mathcal{C}K}{\rho}} e^{-\rho t}. \quad (9.31)$$

The first estimate uses only (9.29), and no Poincaré inequality or gradient bound.

Proof. Both estimates are trivial for $t = 0$. Let $t > 0$. By part (a) of Theorem 9.5 with $\mathcal{A}(u) = \mathcal{A}e^{-au}$, we have

$$\|\mu_0 P_t^A - \mu_t\|_{\text{TV}} \leq \frac{\mathcal{A}}{2} \int_0^t e^{-au} du = \frac{\mathcal{A}}{2a}(1 - e^{-at}),$$

and the inequality $1 - e^{-at} \leq \min(at, 1)$ yields (9.30). For (9.31), we use part (b) of Theorem 9.5 with $\mathcal{C}(u) = \mathcal{C}$ and, by (9.26),

$$\mathcal{G}(u, \tau) = \frac{K\rho}{4(e^{2\rho\tau} - 1)}.$$

Writing $(e^{2\rho(t-u)} - 1)^{-1/2} = e^{-\rho(t-u)}(1 - e^{-2\rho(t-u)})^{-1/2}$, we see that the integrand in (9.19) is at most

$$\mathcal{A}(u) \sqrt{\mathcal{C}\mathcal{G}(u, t-u)} = \frac{\mathcal{A}\sqrt{\mathcal{C}K\rho}}{2} e^{-\rho t} \frac{e^{-(a-\rho)u}}{\sqrt{1 - e^{-2\rho(t-u)}}}.$$

By Lemma 9.9, the integral of the last factor over $u \in (0, t)$ is at most $1/\rho$, and (9.31) follows. $\square$

9.5 An inverse moment under conditioned Gaussian laws

The weighted inequality (9.24) involves $\mu(F^{-1})$. For the weight φ and the comparison laws, this is an inverse moment of a nonnegative quadratic form under a conditioned Gaussian law. The following lemma bounds such moments without any lower bound on the variances.

Lemma 9.11 (An inverse moment of a nonnegative quadratic form). *There exist universal constants $\eta_{\text{inv}} > 0$ and $C_{\text{inv}} < \infty$ with the following property. Let v be a positive sequence with $\sum_i v_i = N$ and $\eta_v \leq \eta_{\text{inv}}$. Let $\gamma_1, \dots, \gamma_N \geq 0$ satisfy*

$$H_\gamma = \sum_i \gamma_i v_i > 0, \quad \max_i \gamma_i v_i \leq \frac{H_\gamma}{8}.$$

Then $\sum_i \gamma_i x_i^2 > 0$ for μ_v -almost every x , and

$$\mu_v \left[\left(\sum_i \gamma_i x_i^2 \right)^{-1} \right] \leq \frac{C_{\text{inv}}}{H_\gamma}. \quad (9.32)$$

Proof. We first fix the constants. Let $\eta_1 = \eta(1/3)$ and $K_1 = C_{1/3}$ be the universal constants of the second part of Theorem 5.11, applied with $\varepsilon = 1/3$. Thus, if $\eta_v \leq \eta_1$ and J is a set of indices such that $\sum_{i \notin J} v_i^2 \geq (1/3) \sum_i v_i^2$, then the law of $(x_i / \sqrt{v_i})_{i \in J}$ under μ_v has a density bounded by K_1 with respect to the standard Gaussian measure on $\mathbb{R}^J$. We set $\eta_{\text{inv}} = \min\{\eta_1, 24^{-1/2}\}$ and $C_{\text{inv}} = 4K_1$. Let $\Sigma_2 = \sum_i v_i^2$. Each coordinate carries at most a fraction $v_i^2 / \Sigma_2 = 2(v_i / S_v)^2 \leq 2\eta_v^2 \leq 1/12$ of Σ_2 .

Next, we split the indices into two blocks. Let n be the smallest index such that $\sum_{i \leq n} v_i^2 \geq \Sigma_2/3$, and let $J_0 = \{1, \dots, n\}$. As each coordinate carries at most $\Sigma_2/12$, we have

$$\frac{\Sigma_2}{3} \leq \sum_{i \in J_0} v_i^2 \leq \frac{5\Sigma_2}{12}, \quad \frac{7\Sigma_2}{12} \leq \sum_{i \notin J_0} v_i^2 \leq \frac{2\Sigma_2}{3}.$$

The complement of each of the blocks J_0 and J_0^c carries at least $\Sigma_2/3$, and the density bound of the previous paragraph applies to both blocks.

Let $\omega_i = \gamma_i v_i$. One of the two blocks, which we call J , satisfies $H_J = \sum_{i \in J} \omega_i \geq H_\gamma/2$. Let $\omega_* = \max_{i \in J} \omega_i$. Then $\omega_* > 0$, since $H_J > 0$, and $\omega_* \leq H_\gamma/8 \leq H_J/4$. Let $(G_i)_{i \in J}$ be independent standard Gaussian variables. Using $y^{-1} = \int_0^\infty e^{-\tau y} d\tau$ for $y > 0$, $\mathbb{E} e^{-\tau \omega G^2} = (1 + 2\tau \omega)^{-1/2}$ and Tonelli's theorem, we have

$$\mathbb{E} \left(\sum_{i \in J} \omega_i G_i^2 \right)^{-1} = \int_0^\infty \prod_{i \in J} (1 + 2\tau \omega_i)^{-1/2} d\tau.$$

For fixed $\tau \geq 0$, the function $y \mapsto \log(1 + 2\tau y)$ is concave on $[0, \omega_*]$ and vanishes at zero. This implies that $\log(1 + 2\tau \omega_i) \geq (\omega_i / \omega_*) \log(1 + 2\tau \omega_*)$ for $i \in J$, and the product is at most $(1 + 2\tau \omega_*)^{-H_J/(2\omega_*)}$. Because $H_J/(2\omega_*) \geq 2$, we have

$$\int_0^\infty (1 + 2\tau \omega_*)^{-H_J/(2\omega_*)} d\tau = \frac{1}{H_J - 2\omega_*} \leq \frac{2}{H_J} \leq \frac{4}{H_\gamma}.$$

Finally, $\sum_i \gamma_i x_i^2 \geq \sum_{i \in J} \omega_i (x_i / \sqrt{v_i})^2$, and the law of $(x_i / \sqrt{v_i})_{i \in J}$ under μ_v has density at most K_1 with respect to the law of $(G_i)_{i \in J}$. As the function $y \mapsto (\sum_{i \in J} \omega_i y_i^2)^{-1}$ is nonnegative, we conclude that

$$\mu_v \left[\left(\sum_i \gamma_i x_i^2 \right)^{-1} \right] \leq K_1 \mathbb{E} \left(\sum_{i \in J} \omega_i G_i^2 \right)^{-1} \leq \frac{4K_1}{H_\gamma}.$$

In particular, $\sum_i \gamma_i x_i^2$ is positive μ_v -almost surely. $\square$

9.6 Comparison when Nd^2 is large

In this section, the operator norm of A is bounded by a constant $\Lambda \geq 0$, and Nd^2 is larger than a constant depending only on Λ . We fix the following notation. Let $\eta_0 > 0$ and $C_4 < \infty$ be the universal threshold and the constant for $p = 4$ in Theorem 5.15: if $\eta_v \leq \eta_0$, then $\|\sum_i a_i (x_i^2 / v_i - 1)\|_{L^4(\mu_v)} \leq C_4 (\sum_i a_i^2)^{1/2}$ for all real a_i . Let η_{inv} and C_{inv} be as in Lemma 9.11.

Lemma 9.12 (Poincaré constants and defect bounds for the comparison laws). *Let $\Lambda \geq 0$, suppose that $\|A\| \leq \Lambda$, and suppose that*

$$Nd^2 \geq C_{\text{dim}}(\Lambda), \quad C_{\text{dim}}(\Lambda) = \max\{2(3 + 2\Lambda), (2\eta_0^2)^{-1}\}. \quad (9.33)$$

Then $r \geq d/2$. For every $t \geq 0$, the law μ_t satisfies $\text{Var}_{\mu_t}(g) \leq (2/d)\mu_t\Gamma(g)$ for all smooth g , and its forward-equation defect satisfies

$$\|R_t\|_{L^2(\mu_t)} \leq C_4^2 \frac{\Omega_t}{\bar{h}(t)} \leq 2\Lambda C_4^2 \vartheta^{-1} e^{-2dt}. \quad (9.34)$$

Proof. We first show that $r \geq d/2$. We have $d \leq 1 \leq s \leq 1 + 2\Lambda$ by Lemma 9.1. In particular $d + d^2/s \leq 2$, and

$$\kappa + 1 + \kappa^{-1} = \frac{s + d + d^2/s}{d} \leq \frac{s + 2}{d} \leq \frac{3 + 2\Lambda}{d}.$$

Then (9.33) implies that $d - r = (\kappa + 1 + \kappa^{-1})/N \leq (3 + 2\Lambda)/(Nd) \leq d/2$.

Next, we prove the Poincaré inequality. Fix $t \geq 0$. The Lagrange multiplier of B_t is $z(t)$, and its precisions are the numbers $1/v_i(t)$, which lie in $[d, s]$ by (9.6). In particular, the

smallest precision d_t of B_t satisfies $d_t \geq d$, and its ratio κ_t of largest to smallest precision satisfies $\kappa_t \leq \kappa$. The function $y \mapsto y + y^{-1}$ is nondecreasing on $[1, \infty)$. By Theorem 8.6, applied to B_t , the spectral gap $\lambda_{\text{gap}}(L_{B_t})$ is therefore at least

$$d_t - \frac{\kappa_t + \kappa_t^{-1}}{N} \geq d - \frac{\kappa + \kappa^{-1}}{N} = r_0 \geq r \geq \frac{d}{2}.$$

As $\mu_t = \pi_{B_t}$, this is the Poincaré inequality with constant $2/d$.

Finally, we bound the defect. Using (9.4) and (9.6), we find $\eta_{v(t)} \leq (2Nd^2)^{-1/2} \leq \eta_0$, and Theorem 5.15 applies to $\mu_t = \mu_{v(t)}$. By (9.16) and (9.10), we have

$$\|Q_B(t)\|_4 \leq C_4 \sqrt{N}, \quad \|Q_C(t)\|_4 \leq C_4 \left(\sum_i e_i(t)^2 \right)^{1/2} \leq C_4 \sqrt{N} \frac{\Omega_t}{\bar{h}(t)}.$$

For every square-integrable Y , we have $\|Y - \mu_t(Y)\|_2 \leq \|Y\|_2$. Using (9.13) and Hölder's inequality, we find

$$\|R_t\|_2 \leq \frac{1}{N} \|Q_B(t)Q_C(t)\|_2 \leq \frac{1}{N} \|Q_B(t)\|_4 \|Q_C(t)\|_4 \leq C_4^2 \frac{\Omega_t}{\bar{h}(t)}.$$

This is the first inequality in (9.34). The same argument also gives $\|R_t\|_2 \leq C_4^2 (2\mathcal{I}_N(t)/N)^{1/2}$. The second inequality follows from $\bar{h}(t) \geq \vartheta$ and $s - d = \alpha_1 - \alpha_N \leq 2\Lambda$. $\square$

Theorem 9.13 (Comparison with the Bingham comparison law when Nd^2 is large). *For every $\Lambda \geq 0$, there exist constants $C_w(\Lambda) > 0$ and $C_\Lambda > 0$ such that the following holds. Suppose that $\|A\| \leq \Lambda$ and that*

$$Nd^2 \geq C_w(\Lambda). \quad (9.35)$$

Then $r \geq d/2$, and for every $t \geq 0$,

$$\|\nu_N P_t^A - \mu_t\|_{\text{TV}} \leq C_\Lambda d^{-3/2} \vartheta^{-1/2} e^{-rt} \leq C_\Lambda d^{-2} e^{-rt}. \quad (9.36)$$

In particular, if $d \geq d_0 > 0$, then the left side of (9.36) is at most $C_\Lambda d_0^{-2} e^{-d_0 t/2}$, which tends to zero along every sequence of times tending to infinity.

We choose $C_w(\Lambda)$ sufficiently large that (9.35) implies the following four lower bounds on Nd^2 .

Proof. Choose $C_w(\Lambda) \geq \max\{C_{\text{dim}}(\Lambda), 8, (2\eta_{\text{inv}}^2)^{-1}\}$ and $C_\Lambda \geq 1 + 2\Lambda C_4^2 \sqrt{2 \max\{1, C_{\text{inv}}\}}$. The proof has two steps. We first bound the energy of $P_\tau^A f$ under μ_u , using the weighted gradient estimate (9.3) for the fixed weight φ and an inverse moment of φ under μ_u . We then insert this bound, the Poincaré inequality and the defect bound of Lemma 9.12 into part (b) of Theorem 9.5, applied to the family (μ_u) of Bingham comparison laws, for which $\mu_0 = \nu_N$. Recall from Lemma 9.12 that $r \geq d/2$.

Table 9.1. The conditions in (9.35).

Condition	Use in the proof of Theorem 9.13
$Nd^2 \geq 2(3 + 2\Lambda)$	$r \geq d/2$; Poincaré constant $2/d$ for μ_u (Lemma 9.12)
$Nd^2 \geq (2\eta_0^2)^{-1}$	$\eta_{v(u)} \leq \eta_0$, so Theorem 5.15 applies to μ_u
$Nd^2 \geq 8$	$\max_i q_i v_i(u) \leq (1/8) \sum_i q_i v_i(u)$ in Lemma 9.11
$Nd^2 \geq (2\eta_{\text{inv}}^2)^{-1}$	$\eta_{v(u)} \leq \eta_{\text{inv}}$ in Lemma 9.11

Step 1: the energy of $P_\tau^A f$ under μ_u . Fix $u \geq 0$. We verify the hypotheses of Lemma 9.11 with $v = v(u)$ and $\gamma_i = q_i$. We have $H_\gamma = \sum_i q_i v_i(u) = N/\bar{h}(u)$ by (9.9), and $\max_i q_i v_i(u) \leq d^{-2}$ by (9.6) and $q_i \leq 1/d$. As $\bar{h}(u) \leq 1$ and $Nd^2 \geq 8$, we have $\max_i q_i v_i(u) \leq N/8 \leq H_\gamma/8$. Also $\eta_{v(u)} \leq (2Nd^2)^{-1/2} \leq \eta_{\text{inv}}$ by (9.4). Then Lemma 9.11 implies that $\mu_u[(\sum_i q_i x_i^2)^{-1}] \leq C_{\text{inv}} \bar{h}(u)/N$, that is,

$$\mu_u(\varphi^{-1}) \leq C_{\text{inv}} \bar{h}(u). \quad (9.37)$$

We note that this is a bound for an expectation under μ_u . The number $N/\bar{h}(u)$, which is the value of $\sum_i q_i x_i^2$ at $x_i^2 = v_i(u)$, enters only through the hypotheses of Lemma 9.11. The weight φ satisfies the hypothesis of Lemma 9.8(a) with $\rho = r$ and $D = 1/d$, by (9.3) with r in place of r_0 . By (9.24) and (9.37), for every $\tau > 0$ and smooth f with $0 \leq f \leq 1$,

$$\mu_u \Gamma(P_\tau^A f) \leq \frac{r \mu_u(\varphi^{-1})}{4d(e^{2r\tau} - 1)} \leq \frac{C_{\text{inv}} r \bar{h}(u)}{4d(e^{2r\tau} - 1)} = \mathcal{G}(u, \tau). \quad (9.38)$$

Step 2: the Duhamel integral. Fix $t > 0$. We use part (b) of Theorem 9.5 with $\mathcal{G}$ from (9.38), $\mathcal{C}(u) = 2/d$ and $\mathcal{A}(u) = C_4^2 \Omega_u / \bar{h}(u)$, which are admissible by Lemma 9.12. Using $(e^{2r(t-u)} - 1)^{-1/2} = e^{-r(t-u)}(1 - e^{-2r(t-u)})^{-1/2}$, we find

$$\mathcal{A}(u) \sqrt{\mathcal{C}(u) \mathcal{G}(u, t-u)} = C_4^2 (s-d) \sqrt{\frac{C_{\text{inv}} r}{2}} \frac{e^{-rt}}{d \sqrt{\bar{h}(u)}} \frac{e^{-(2d-r)u}}{\sqrt{1 - e^{-2r(t-u)}}}.$$

We bound $\bar{h}(u) \geq \vartheta$ and $s-d \leq 2\Lambda$. As $r \leq d$, Lemma 9.9 applies with $\rho = r$ and $a = 2d$, and the integral over $u \in (0, t)$ of the last factor is at most $1/r$. Inserting these bounds into (9.19), we obtain

$$\|v_N P_t^A - \mu_t\|_{\text{TV}} \leq 2\Lambda C_4^2 \sqrt{\frac{C_{\text{inv}}}{2r}} \frac{e^{-rt}}{d \sqrt{\vartheta}} \leq C_\Lambda d^{-3/2} \vartheta^{-1/2} e^{-rt},$$

where in the last step we used $r \geq d/2$. This is the first inequality in (9.36), and the second follows from $\vartheta \geq d$. For $t = 0$ the left side of (9.36) is zero. If $d \geq d_0$, then $rt \geq dt/2 \geq d_0 t/2$, which implies the last assertion. $\square$

9.7 Comparison for nearly isotropic matrices

When A is close to a multiple of the identity, the defect is quadratically small, and the comparison holds uniformly in time without any condition relating N and d beyond $N \geq N_0$. This is the input for the infinite-temperature crossover in Section 15.

Proposition 9.14 (Comparison for a nearly isotropic matrix). *There exist universal constants $\epsilon_0 > 0$, N_0 and C_1 such that the following holds. Suppose that $N \geq N_0$ and $\|A - (N^{-1} \text{Tr } A)I\| \leq \epsilon \leq \epsilon_0$. Then $r > 0$, and for every $t \geq 0$,*

$$\|\nu_N P_t^A - \mu_t\|_{\text{TV}} \leq C_1 \epsilon^2 \min(t, 1), \quad \|\nu_N P_t^A - \mu_t\|_{\text{TV}} \leq C_1 \epsilon^2 e^{-rt}. \quad (9.39)$$

The constants do not depend on $N^{-1} \text{Tr } A$, so that no bound on the scalar part of A is needed. One may take $\epsilon_0 = 1/4$, $N_0 = \max\{35, 2\eta_0^{-2}\}$ and $C_1 = 8\sqrt{32}C_4^2$, with η_0 and C_4 as in Section 9.6.

Proof. Let $\epsilon_0 = 1/4$ and $N_0 = \max\{35, 2\eta_0^{-2}\}$. By Lemma 9.1,

$$\frac{1}{2} \leq 1 - 2\epsilon \leq d \leq 1 \leq s \leq 1 + 2\epsilon \leq \frac{3}{2}, \quad s - d \leq 2\epsilon, \quad \kappa \leq 3.$$

By (9.6), every $v_i(t)$ lies in $[1/s, 1/d]$. As $1/d - 1 = (1 - d)/d \leq 4\epsilon$ and $1 - 1/s \leq 2\epsilon$, we have $|1 - v_i(t)| \leq 4\epsilon$. Combining (9.10) with $\vartheta \geq d \geq 1/2$ and $s - d \leq 2\epsilon$, we conclude that $|e_i(t)| \leq 4\epsilon e^{-2dt}$.

We first bound the defect. By (9.4), $\eta_{v(t)} \leq d^{-1}(2N)^{-1/2} \leq (2/N)^{1/2} \leq \eta_0$, and Theorem 5.15 applies to μ_t . On S_N , $\sum_i (x_i^2 - v_i(t)) = \sum_i v_i(t)(x_i^2/v_i(t) - 1) = 0$. Adding this to the first expression in (9.16), we find

$$Q_B(t) = - \sum_i (1 - v_i(t)) \left(\frac{x_i^2}{v_i(t)} - 1 \right).$$

The coefficients of $Q_B(t)$ and $Q_C(t)$ in the variables $x_i^2/v_i(t) - 1$ are bounded in absolute value by 4ϵ and $4\epsilon e^{-2dt}$, respectively. Then Theorem 5.15 implies that $\|Q_B(t)\|_4 \leq 4C_4\epsilon\sqrt{N}$ and $\|Q_C(t)\|_4 \leq 4C_4\epsilon e^{-2dt}\sqrt{N}$. As in the proof of Lemma 9.12, we deduce from (9.13) and Hölder's inequality that

$$\|R_t\|_{L^2(\mu_t)} \leq \frac{1}{N} \|Q_B(t)\|_4 \|Q_C(t)\|_4 \leq 16C_4^2 \epsilon^2 e^{-2dt}.$$

The first estimate in (9.39) now follows from (9.30) with $\mathcal{A} = 16C_4^2 \epsilon^2$ and $a = 2d$, since $\mu_0 = \nu_N$ and $\min(t, 1/(2d)) \leq \min(t, 1)$.

For the second estimate, we verify the remaining hypotheses of Theorem 9.10. For $N \geq 35$ and $\kappa \leq 3$, we have $r \geq 1/2 - (13/3)/N \geq 3/8$. Also $r \leq d$, and therefore $a = 2d \geq 2r$. By (9.3) with r in place of r_0 , the gradient hypothesis holds with $K = \kappa \leq 3$

and $\rho = r$. As in the proof of Lemma 9.12, the precisions $1/v_i(t)$ of B_t lie in $[d, s]$, and $\lambda_{\text{gap}}(L_{B_t}) \geq d - (\kappa + \kappa^{-1})/N \geq 1/2 - (10/3)/N \geq 1/4$. Thus $\mu_t = \pi_{B_t}$ satisfies the Poincaré inequality with $C = 4$. By (9.31), we obtain

$$\|\nu_N P_t^A - \mu_t\|_{\text{TV}} \leq 8C_4^2 \epsilon^2 \sqrt{\frac{4 \cdot 3}{3/8}} e^{-rt} = 8\sqrt{32} C_4^2 \epsilon^2 e^{-rt},$$

which completes the proof. $\square$

9.8 A Poisson equation for Bingham laws

Near the critical temperature, Nd^2 may tend to zero, and the defect bound of Lemma 9.12 is too weak. In Section 14, the comparison is instead made with part (c) of Theorem 9.5, through the Poisson equation of Corollary 9.7. This section bounds the energy of the solution of this Poisson equation for a single Bingham law. The proof constructs an explicit approximate solution, a rational function of the quadratic form $F_v(x) = N^{-1} \sum_i v_i x_i^2$ multiplied by Q_C , and handles the remainder with the spectral gap.

Throughout this section, v is a positive sequence with $\sum_i v_i = N$, and $\mu = \mu_v$. We let $B = -\text{diag}(v_i^{-1})$, so that $\mu = \pi_B$ and the Lagrange multiplier of B is zero. Any other matrix $yI - \text{diag}(v_i^{-1})$ defines the same generator L_B . Let $S_k = \sum_i v_i^k$ for $k = 2, 3, 4$ and $V = \text{diag}(v_i)$, and set

$$F_v(x) = \frac{x^\top V x}{N} = \frac{1}{N} \sum_i v_i x_i^2, \quad F_0 = \frac{S_2}{N}, \quad f(y) = -\frac{1}{2} + \frac{F_0}{2y} \quad (y > 0).$$

Note that F_0 is the value of F_v at $x_i^2 = v_i$. Let $C = \text{diag}(C_i)$ be a real diagonal matrix with $\sum_i C_i v_i = 0$. We define

$$Q_B = N - \sum_i \frac{x_i^2}{v_i}, \quad Q_C = x^\top C x, \quad R = -\frac{Q_B Q_C - \mu(Q_B Q_C)}{N}, \quad \psi_0 = f(F_v) Q_C,$$

and

$$I_C = \sum_i (C_i v_i)^2, \quad J_C = \sum_i C_i^2, \quad K_C = \sum_i C_i^2 v_i.$$

When $v = v(u)$ and $C = C_u$ for some $u \geq 0$, the condition $\sum_i C_i v_i = 0$ holds by (9.7), the forms Q_B and Q_C coincide on S_N with those of (9.12) by (9.16), and $R = R_u$ is the forward-equation defect of the Bingham comparison law at time u .

Definition 9.15 (Variance conditions at scale δ). Let $0 < c_0 < 1 \leq C_0$ and $0 < \delta \leq 1$, and set $M = N\delta^3$. We say that v satisfies the variance conditions at scale δ with constants (c_0, C_0) if

$$\begin{aligned} c_0 \leq v_i \leq C_0 \delta^{-2} \quad (1 \leq i \leq N), \quad c_0 N \delta^{-1} \leq S_2 \leq C_0 N \delta^{-1}, \\ S_3 \leq C_0 N \delta^{-3}, \quad S_4 \leq C_0 N \delta^{-5}, \end{aligned} \tag{9.40}$$

and if there exists a set $\mathcal{S}$ of indices such that

$$|\mathcal{S}| \geq c_0 M, \quad v_i \geq c_0 \delta^{-2} \quad (i \in \mathcal{S}), \quad \sum_{i \notin \mathcal{S}} v_i^2 \geq c_0 S_2. \quad (9.41)$$

The set $\mathcal{S}$ consists of at least $c_0 M$ coordinates whose variances are comparable to the largest one. Equivalently, their precisions v_i^{-1} are comparable to the smallest precision. The last condition in (9.41) says that the remaining coordinates carry a fixed fraction of S_2 . For $M \geq M_0$, where M_0 is the constant of Theorem 9.16 below, these conditions imply that $\lambda_{\text{gap}}(L_B) \geq \delta^2/(2C_0)$, as proved in Step 6 of its proof. They also imply that $\mu\{F_v < F_0/2\} \leq K/M$ for a constant K depending only on c_0 and C_0 , by (9.47) with $q = 2$ and Markov's inequality applied to $(F_v - F_0)^2$.

Theorem 9.16 (Energy of the solution of the Poisson equation). *Let $0 < c_0 < 1 \leq C_0$. There exist constants M_0 and C_* , depending only on c_0 and C_0 , such that the following holds. Let $0 < \delta \leq 1$ and let v satisfy the variance conditions at scale δ with constants (c_0, C_0) , with $M = N\delta^3 \geq M_0$. Let $C = \text{diag}(C_i)$ be real with $\sum_i C_i v_i = 0$. Then the solution ψ of $-L_B \psi = R$ with $\mu(\psi) = 0$ satisfies*

$$\mu\Gamma(\psi) \leq \frac{C_*}{M} \left(K_C + \frac{J_C}{\delta^2} + \delta^2 I_C \right). \quad (9.42)$$

Moreover, the function $\psi_0 = f(F_v)Q_C$ satisfies

$$\|\psi_0\|_{L^2(\mu)} \leq C_* \sqrt{\frac{I_C}{M}}, \quad \mu\Gamma(\psi_0) \leq \frac{C_*}{M} (K_C + \delta^2 I_C). \quad (9.43)$$

Recall that the solution ψ exists by Lemma 9.6, because $\mu(R) = 0$. The bound (9.42) involves only the three sums I_C , J_C and K_C . The theorem assumes no relation between C and v other than $\sum_i C_i v_i = 0$, and no lower bound on $N\delta^4$. The lower bounds on M used in the proof are collected in Table 9.2.

Table 9.2. Lower bounds on $M = N\delta^3$ used in the proof of Theorem 9.16. Here η_0 is the threshold of Theorem 5.15 and $\eta(c_0)$ is the threshold of Theorem 5.11 for complements carrying a fraction at least c_0 of S_2 . The number M_0 is the smallest M satisfying all of them, and depends only on c_0 and C_0 .

	Condition	Use
1	$M \geq 1$ and $C_0^2 \leq 2c_0\eta_0^2 M$	$\eta_v \leq \eta_0$ (Step 1)
2	$C_0^2 \leq 2c_0\eta(c_0)^2 M$	bounded density of $(g_i)_{i \in \mathcal{S}}$ (Step 2)
3	$c_0 M \geq 96$	inverse moments of order $q \leq 24$ (Step 2)
4	$c_0 M \geq 64(C_0/c_0)^2 + 2$ and $M \geq 8C_0$	$\lambda_{\text{gap}}(L_B) \geq \delta^2/(2C_0)$ (Step 6)

Proof. The proof has six steps. Steps 1 and 2 collect moment bounds under μ for quadratic forms and for the inverse of F_v . Step 3 computes $L_B\psi_0$ exactly: ψ_0 solves the Poisson equation up to a remainder $\tilde{E}$ with mean zero. Steps 4 and 5 bound the $L^2(\mu)$ norm of $\tilde{E}$ and the energy of ψ_0 . Step 6 solves the Poisson equation with right side $\tilde{E}$, using the spectral gap of $-L_B$, whose lower bound $\delta^2/(2C_0)$ is proved in Step 6 using Theorem 8.8.

Here, and for the rest of this proof, we write $X \lesssim Y$ if $X \leq KY$ for a constant K depending only on c_0 and C_0 and, where an exponent q appears, on q . All exponents below are at most 24. The number M_0 is chosen so that the conditions in Table 9.2 hold whenever $M \geq M_0$. All norms are in $L^q(\mu)$.

Step 1: moments of quadratic forms. By (9.40),

$$\eta_v = \frac{\max_i v_i}{(2S_2)^{1/2}} \leq \frac{C_0\delta^{-2}}{(2c_0N\delta^{-1})^{1/2}} = \frac{C_0}{(2c_0M)^{1/2}}.$$

The first condition in Table 9.2 gives $\eta_v \leq \eta_0$, the threshold of Theorem 5.15. Writing $\sum_i a_i(x_i^2 - v_i) = \sum_i (a_i v_i)(x_i^2/v_i - 1)$ and using Theorem 5.15, we find that for every real sequence (a_i) and $q \geq 1$,

$$\left\| \sum_i a_i(x_i^2 - v_i) \right\|_q \lesssim \left(\sum_i a_i^2 v_i^2 \right)^{1/2}. \quad (9.44)$$

If all $a_i \geq 0$, then $(\sum_i a_i^2 v_i^2)^{1/2} \leq \sum_i a_i v_i$, and adding the constant $\sum_i a_i v_i$ we find $\|\sum_i a_i x_i^2\|_q \lesssim \sum_i a_i v_i$. We use these bounds for the following forms. Because $\sum_i C_i v_i = 0$, we have $Q_C = \sum_i C_i(x_i^2 - v_i)$. We also set $\tilde{Q}_C = \sum_i C_i(x_i^2/v_i - 1)$. By (9.44), we have

$$\|Q_C\|_q \lesssim \sqrt{I_C}, \quad \|\tilde{Q}_C\|_q \lesssim \sqrt{J_C}, \quad \|x^\top V^2 x\|_q \lesssim S_3, \quad \|x^\top C^2 x\|_q \lesssim K_C. \quad (9.45)$$

Next, $x^\top V C x = \sum_i v_i^2 C_i + \sum_i v_i C_i(x_i^2 - v_i)$. We have $|\sum_i v_i^2 C_i| \leq S_2^{1/2} I_C^{1/2}$ by the Cauchy-Schwarz inequality, and by (9.44), the L^q norm of the second sum is at most a constant times $(\sum_i v_i^4 C_i^2)^{1/2} \leq (\max_i v_i) I_C^{1/2}$. Recalling that $(N\delta^{-1})^{1/2} = \delta^{-2} M^{1/2}$ and $M \geq 1$, we obtain

$$\|x^\top V C x\|_q \lesssim ((N\delta^{-1})^{1/2} + \delta^{-2}) \sqrt{I_C} \lesssim \delta^{-2} \sqrt{M I_C}. \quad (9.46)$$

Finally, $F_v - F_0 = N^{-1} \sum_i v_i(x_i^2 - v_i)$, and (9.44) and (9.40) yield

$$\|F_v - F_0\|_q \lesssim \frac{\sqrt{S_4}}{N} \lesssim \frac{\delta^{-5/2}}{\sqrt{N}} = \frac{\delta^{-1}}{\sqrt{M}} \lesssim \frac{F_0}{\sqrt{M}}, \quad (9.47)$$

where in the last step we used $F_0 = S_2/N \geq c_0\delta^{-1}$.

Step 2: inverse moments of F_v . Let $\mathcal{S}$ be as in (9.41), let $k = |\mathcal{S}|$, and write $g_i = x_i / \sqrt{v_i}$. On S_N , we have

$$F_v = \frac{1}{N} \sum_i v_i^2 g_i^2 \geq \frac{c_0^2}{N\delta^4} \sum_{i \in \mathcal{S}} g_i^2, \quad \text{so} \quad \frac{F_0}{F_v} \leq \frac{C_0}{c_0^2} \frac{M}{\sum_{i \in \mathcal{S}} g_i^2},$$

where we used $F_0 \leq C_0 \delta^{-1}$. The coordinates outside $\mathcal{S}$ carry at least a fraction c_0 of S_2 , which means that the fraction carried by $\mathcal{S}$ is at most $1 - c_0$. By the second part of Theorem 5.11, applied with $\varepsilon = c_0$, there exist $\eta(c_0) > 0$ and $K(c_0) = C_{c_0} < \infty$ such that, if $\eta_v \leq \eta(c_0)$, the law of $(g_i)_{i \in \mathcal{S}}$ under μ has a density bounded by $K(c_0)$ with respect to the standard Gaussian measure on $\mathbb{R}^{\mathcal{S}}$. We have $\eta_v \leq \eta(c_0)$ by the second condition in Table 9.2. Let $q \geq 1$ be an integer such that $k \geq 4q$. Recall that a chi-square variable X with k degrees of freedom has $\mathbb{E}X^{-q} = 2^{-q} \Gamma(k/2 - q) / \Gamma(k/2)$. Also $\Gamma(k/2) / \Gamma(k/2 - q) = \prod_{j=1}^q (k/2 - j) \geq (k/4)^q$. Then

$$\mu \left[\left(\sum_{i \in \mathcal{S}} g_i^2 \right)^{-q} \right] \leq K(c_0) 2^{-q} \left(\frac{4}{k} \right)^q = K(c_0) \left(\frac{2}{k} \right)^q.$$

For real $1 \leq q \leq 24$, we apply this estimate with the integer $m = \lceil q \rceil \leq 24$ and use $\|Y\|_q \leq \|Y\|_m$ for nonnegative Y under a probability measure. Using $k \geq c_0 M$ and the third condition in Table 9.2, which gives $k \geq 96$, we conclude that for every $1 \leq q \leq 24$

$$\left\| \frac{F_0}{F_v} \right\|_q \lesssim 1, \quad \|f(F_v)\|_q \lesssim \frac{1}{\sqrt{M}} \quad (q \leq 12). \quad (9.48)$$

For the second bound, we wrote $f(F_v) = -(F_v - F_0) / (2F_v)$ as $-1/2$ times the product of $(F_v - F_0) / F_0$ and F_0 / F_v , and used Hölder's inequality, (9.47) and the first bound.

Step 3: an explicit approximate solution. We first compute the action of L_B on the coordinates. By (9.1) with $D = \text{diag}((\mathbf{1}_{\{j=i\}})_{j=1}^N)$ and $D' = B$, we have $\Delta_{S_N} x_i^2 = 2 - 2x_i^2$ and $\Gamma(\Phi_B, x_i^2) = 2(-x_i^2/v_i - u_B x_i^2/N)$, where $u_B = x^\top B x = Q_B - N$ on S_N . Inserting these into $L_B x_i^2 = \Delta_{S_N} x_i^2 + \Gamma(\Phi_B, x_i^2)$, we find

$$L_B x_i^2 = 2 - 2x_i^2 - \frac{2x_i^2}{v_i} - \frac{2(Q_B - N)x_i^2}{N} = 2 - \frac{2x_i^2}{v_i} - \frac{2Q_B x_i^2}{N}.$$

We multiply by v_i/N or by C_i and sum over i . Using $\sum_i v_i = \sum_i x_i^2 = N$, we have

$$L_B F_v = -\frac{2Q_B F_v}{N}, \quad L_B Q_C = -2\tilde{Q}_C - \frac{2Q_B Q_C}{N}. \quad (9.49)$$

Note that $F_v \geq \min_i v_i > 0$ on S_N , which makes ψ_0 smooth. By the diffusion property of L_B ,

$$L_B \psi_0 = f(F_v) L_B Q_C + Q_C (f'(F_v) L_B F_v + f''(F_v) \Gamma(F_v)) + 2f'(F_v) \Gamma(F_v, Q_C).$$

We have $f'(y) = -F_0/(2y^2)$ and $f''(y) = F_0/y^3$. Then $f'(F_v)L_B F_v = F_0 Q_B/(NF_v)$, $f''(F_v)\Gamma(F_v) = (F_0/F_v)\Gamma(\log F_v)$ and $2f'(F_v)\Gamma(F_v, Q_C) = -(F_0/F_v)\Gamma(\log F_v, Q_C)$. Inserting these and (9.49), the two terms proportional to $Q_B Q_C/N$ have the total coefficient $F_0/F_v - 2f(F_v) = 1$, and we obtain

$$\begin{aligned} -L_B \psi_0 &= -\frac{Q_B Q_C}{N} + E, \\ E &= 2f(F_v)\tilde{Q}_C - \frac{F_0}{F_v} Q_C \Gamma(\log F_v) + \frac{F_0}{F_v} \Gamma(\log F_v, Q_C). \end{aligned} \quad (9.50)$$

The function f is chosen so that $F_0/y - 2f(y) = 1$, which produces this cancellation, while $f(F_v)$ is small because F_v concentrates around F_0 . As the measure μ is stationary for L_B , we have $\mu(L_B \psi_0) = 0$ and $\mu(Q_B Q_C)/N = \mu(E)$. Writing $\tilde{E} = E - \mu(E)$, we conclude that

$$-L_B \psi_0 = R + \tilde{E}. \quad (9.51)$$

Step 4: the remainder. A direct computation from $\nabla_{S_N} F_v = (2/N)P_x Vx$ and $\nabla_{S_N} Q_C = 2P_x Cx$ shows that

$$\Gamma(\log F_v) = \frac{4x^\top V^2 x}{N^2 F_v^2} - \frac{4}{N} \leq \frac{4x^\top V^2 x}{N^2 F_v^2}, \quad \Gamma(\log F_v, Q_C) = \frac{4x^\top V C x}{N F_v} - \frac{4Q_C}{N}, \quad (9.52)$$

and $\Gamma(Q_C) \leq 4x^\top C^2 x$. By (9.52), Hölder's inequality, (9.45), (9.48) and (9.40), for $q \leq 6$,

$$\|\Gamma(\log F_v)\|_q \leq \frac{4}{N^2 F_0^2} \|x^\top V^2 x\|_{2q} \left\| \frac{F_0}{F_v} \right\|_{4q}^2 \lesssim \frac{S_3}{S_2^2} \lesssim \frac{\delta^{-1}}{N} = \frac{\delta^2}{M}.$$

Similarly, by (9.46) and $F_0 \geq c_0 \delta^{-1}$, for $q \leq 4$,

$$\|\Gamma(\log F_v, Q_C)\|_q \leq \frac{4}{N F_0} \|x^\top V C x\|_{2q} \left\| \frac{F_0}{F_v} \right\|_{2q} + \frac{4}{N} \|Q_C\|_q \lesssim \frac{\sqrt{M I_C}}{N \delta} + \frac{\sqrt{I_C}}{N} \lesssim \delta^2 \sqrt{\frac{I_C}{M}},$$

where in the last step we used $(N\delta)^{-1}\sqrt{M} = \delta^2/\sqrt{M}$ and $N^{-1} = \delta^3/M \leq \delta^2/\sqrt{M}$. We now bound the three terms of E in $L^2(\mu)$ with Hölder's inequality:

$$\begin{aligned} \|2f(F_v)\tilde{Q}_C\|_2 &\leq 2\|f(F_v)\|_4 \|\tilde{Q}_C\|_4 \lesssim \sqrt{\frac{I_C}{M}}, \\ \left\| \frac{F_0}{F_v} Q_C \Gamma(\log F_v) \right\|_2 &\leq \left\| \frac{F_0}{F_v} \right\|_6 \|Q_C\|_6 \|\Gamma(\log F_v)\|_6 \lesssim \frac{\delta^2 \sqrt{I_C}}{M}, \\ \left\| \frac{F_0}{F_v} \Gamma(\log F_v, Q_C) \right\|_2 &\leq \left\| \frac{F_0}{F_v} \right\|_4 \|\Gamma(\log F_v, Q_C)\|_4 \lesssim \delta^2 \sqrt{\frac{I_C}{M}}. \end{aligned}$$

As $M \geq 1$ and centering does not increase the $L^2(\mu)$ norm, we arrive at

$$\|\tilde{E}\|_2^2 \leq \|E\|_2^2 \lesssim \frac{I_C + \delta^4 I_C}{M}. \quad (9.53)$$

Step 5: the approximate solution. By the chain rule,

$$\nabla_{S_N} \psi_0 = f(F_v) \nabla_{S_N} Q_C - \frac{F_0}{2F_v} Q_C \nabla_{S_N} \log F_v,$$

and therefore

$$\Gamma(\psi_0) \leq 2f(F_v)^2 \Gamma(Q_C) + \frac{1}{2} \left(\frac{F_0}{F_v} \right)^2 Q_C^2 \Gamma(\log F_v).$$

By Hölder's inequality, (9.45), (9.48) and the bound on $\Gamma(\log F_v)$ from Step 4,

$$\begin{aligned} \mu[f(F_v)^2 \Gamma(Q_C)] &\leq 4\|f(F_v)\|_4^2 \|x^\top C^2 x\|_2 \lesssim \frac{K_C}{M}, \\ \mu\left[\left(\frac{F_0}{F_v}\right)^2 Q_C^2 \Gamma(\log F_v)\right] &\leq \left\|\frac{F_0}{F_v}\right\|_6^2 \|Q_C\|_6^2 \|\Gamma(\log F_v)\|_3 \lesssim \frac{\delta^2 I_C}{M}. \end{aligned}$$

This proves the second bound in (9.43). The first follows from $\|\psi_0\|_2 \leq \|f(F_v)\|_4 \|Q_C\|_4 \lesssim \sqrt{I_C/M}$.

Step 6: the spectral gap and the solution. The precisions of B are the numbers v_i^{-1} . By (9.40), the smallest one satisfies $d_B = 1/\max_i v_i \geq \delta^2/C_0$, and for $i \in \mathcal{S}$ we have $v_i^{-1} \leq \delta^2/c_0 \leq Ld_B$ with $L = C_0/c_0 \geq 1$. Thus at least k of the precisions are at most Ld_B , and so are the k smallest ones. By the fourth condition in Table 9.2, $k-2 \geq 64L^2$. Then the lower bound of Theorem 8.8, applied to B , implies that

$$\lambda_{\text{gap}}(L_B) \geq d_B - \frac{2}{N} - \frac{8L^2 d_B}{k-2} \geq d_B - \frac{d_B}{4} - \frac{d_B}{8} \geq \frac{d_B}{2} \geq \frac{\delta^2}{2C_0}.$$

In the second inequality, we used $8L^2/(k-2) \leq 1/8$, and $2/N = 2\delta^3/M \leq \delta^2/(4C_0) \leq d_B/4$, which holds because $M \geq 8C_0$ and $\delta \leq 1$. As $\mu(\tilde{E}) = 0$, by Lemma 9.6 there exists a smooth χ with $\mu(\chi) = 0$ and $-L_B \chi = -\tilde{E}$, and by (9.53),

$$\mu \Gamma(\chi) \leq \lambda_{\text{gap}}(L_B)^{-1} \|\tilde{E}\|_2^2 \lesssim \frac{I_C/\delta^2 + \delta^2 I_C}{M}.$$

The function $\psi_0 - \mu(\psi_0) + \chi$ is smooth and has mean zero, and by (9.51), $-L_B(\psi_0 - \mu(\psi_0) + \chi) = R + \tilde{E} - \tilde{E} = R$. By the uniqueness in Lemma 9.6, it equals ψ . Finally, $\Gamma(\psi) \leq 2\Gamma(\psi_0) + 2\Gamma(\chi)$, and (9.42) follows from Step 5 and the last display. This completes the proof. $\square$

In Section 14, Theorem 9.16 is applied at each time u to $v = v(u)$ and $C = C_u$, with a scale $\delta = \delta(u)$ depending on u . Then ψ is the function ψ_u of Corollary 9.7, and (9.42) provides the function $\mathcal{J}$ in part (c) of Theorem 9.5.

The unnormalized process and its moments

This section introduces the stochastic representation of the Langevin dynamics on which our moment bounds rest, and proves those bounds. For real numbers $\rho_1, \dots, \rho_N$, the unnormalized process is the solution Y in $\mathbb{R}^N$ of the stochastic differential equation $dY_i = -2\rho_i Y_i dt + \sqrt{2/N} |Y| dB_i$. We show in Lemma 10.2 that it never vanishes and that its direction $\sqrt{N} Y/|Y|$ is the diffusion on S_N with potential $\sum_i \rho_i x_i^2$, with the same time parameter and with noise coefficient $\sqrt{2}$. If $Y(0)$ is a standard Gaussian vector, this diffusion starts from ν_N . The coordinates of Y have linear drifts, and when $\mathbb{E}|Y(0)|^2 < \infty$, the expectations of the squares Y_i^2 solve the linear equation with matrix $-4 \operatorname{diag}(\rho) + (2/N) \mathbf{1}\mathbf{1}^\top$ studied in Section 7. The only nonlinear feature is the random noise amplitude $|Y|$, and the estimates of this section bound the squared norm $|Y_t|^2$ from above and from below by solutions of the linear equation.

Section 10.1 proves the representation (Lemma 10.2), together with a representation $Y = \Phi X_0$, where Φ is a matrix-valued process independent of the initial point X_0 . Section 10.2 proves a criterion for positive martingales (Lemma 10.4): a uniform bound on conditional relative variances implies bounds on conditional moments of the maximum and of the inverse of the minimum, and on exponential moments of the quadratic variation of the logarithm. In Section 10.3, we bound conditional moments of $h(u)^\top Z_t$, where $Z_i = Y_i^2$ and h is the solution of the linear equation started from $\mathbf{1}$, by the corresponding solution of the linear equation (Lemma 10.5). We also bound moments of single coordinates by moments of the norm (Lemma 10.6). Section 10.4 applies these results in the critical window (Corollaries 10.7 and 10.8 and Lemmas 10.9 and 10.10). Sections 10.5 to 10.7 treat precision sequences whose smallest element d is small but satisfies $Nd^{3/2} \gg 1$, which is the situation at temperatures above and near the critical temperature, on time intervals of length of order $d^{-1} \log(Nd^{3/2})$. In Section 10.5, we prove conditional moment bounds (Lemma 10.11), inverse moment bounds (Lemmas 10.13 and 10.14), and an exponential moment bound for the time integral of $\Gamma(\log \varphi)$ along the path, where φ is the quadratic form defined in (10.40) (Theorem 10.15). Section 10.6 compares the initial state with all future solutions of the linear equation at once when the initial law is a Bingham comparison law (Lemma 10.16). Section 10.7 upgrades the fixed-time inverse moments of $|Y_t|^2$ to a bound for the supremum over the whole time interval (Theorem 10.18). Finally, Section 10.8 proves a Feynman–Kac bound for a weighted gradient of the semigroup (Lemma 10.21).

We use the model and its stochastic differential equation from Section 3, the conditioned Gaussian laws of Section 5 (Proposition 5.6 and Theorem 5.11), the estimates for the linear equation from Section 7 (Lemmas 7.11, 7.13, and 7.14), and the weighted Bochner identity (8.8) from Section 8. The representation is used in Sections 16 to 19. The estimates of Section 10.4 are used in Sections 18 and 19. The estimates of Sections 10.5 to 10.8 are used in Section 14, and through it in Section 20.

All statements in this section are deterministic in the disorder: the numbers ρ_i and the precisions p_i below are fixed, and probabilities and expectations refer only to the initial state and the Brownian motion. We work on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$ satisfying the usual conditions and carrying an $(\mathcal{F}_t)$ -Brownian motion $B = (B_1, \dots, B_N)$. Stopping times are taken with respect to $(\mathcal{F}_t)$. For a random variable ξ and $r > 0$, we write $\|\xi\|_r = (\mathbb{E}|\xi|^r)^{1/r}$; for $r \geq 1$, this is the usual L^r norm. Since B is an $(\mathcal{F}_t)$ -Brownian motion, it is independent of $\mathcal{F}_0$; in particular, every $\mathcal{F}_0$ -measurable initial value below is independent of B . The estimates below are moment bounds for the process itself, obtained from Itô's formula and martingale inequalities. None of them uses a Gaussian approximation for the law of Y_t or for the conditional law of a coordinate given the past.

10.1 The unnormalized process

For $\rho = (\rho_1, \dots, \rho_N) \in \mathbb{R}^N$, we call the numbers ρ_i the quadratic potential coefficients; they are the coefficients of the potential $\sum_i \rho_i x_i^2$. We write $D_\rho = \text{diag}(\rho_1, \dots, \rho_N)$ and

$$L_\rho = \Delta_{S_N} - \nabla_{S_N} \left(\sum_{i=1}^N \rho_i x_i^2 \right) \cdot \nabla_{S_N}. \quad (10.1)$$

In the notation of (3.3), we have $L_\rho = L_A$ with $A = -2D_\rho$, and by (3.7), $L_{\rho+\tau \mathbf{1}} = L_\rho$ for every $\tau \in \mathbb{R}$.

Definition 10.1 (The unnormalized process). Let $\rho \in \mathbb{R}^N$, and let $Y(0)$ be an $\mathcal{F}_0$ -measurable random vector in $\mathbb{R}^N$ with $Y(0) \neq 0$ almost surely. The unnormalized process with quadratic potential coefficients ρ and initial value $Y(0)$ is the solution of

$$dY_i(t) = -2\rho_i Y_i(t) dt + \sqrt{2/N} |Y(t)| dB_i(t), \quad 1 \leq i \leq N, \quad t \geq 0. \quad (10.2)$$

We write

$$Z_i(t) = Y_i(t)^2, \quad \mathcal{A}_t = |Y(t)|^2 = \sum_{i=1}^N Z_i(t), \quad X(t) = \frac{\sqrt{N} Y(t)}{|Y(t)|}, \quad (10.3)$$

and we call $\mathcal{A}_t$ the squared norm.

We use the letter $\mathcal{A}$ for the squared norm to distinguish it from the matrix A of (3.3). When convenient, we also write Y_t , Z_t , and X_t for $Y(t)$, $Z(t)$, and $X(t)$.

Lemma 10.2 (The unnormalized process). *Let $N \geq 2$ and $\rho \in \mathbb{R}^N$.*

1. *For every $\mathcal{F}_0$ -measurable $Y(0)$ with $Y(0) \neq 0$ almost surely, the equation (10.2) has a pathwise unique strong solution. Conditionally on $Y(0) = y$, the process Y has the law of the solution started at y . Almost surely, $Y(t) \neq 0$ for all $t \geq 0$, and*

$$d \log \mathcal{A}_t = \left(2 - 4\bar{\rho}_t - \frac{4}{N}\right) dt + \sqrt{8/N} d\hat{B}_t, \quad \bar{\rho}_t = \frac{1}{\mathcal{A}_t} \sum_{i=1}^N \rho_i Z_i(t), \quad (10.4)$$

where $\hat{B}$ is a standard one-dimensional Brownian motion.

2. *The process X is the diffusion on S_N generated by L_ρ , with noise coefficient $\sqrt{2}$ and initial point $X(0) = \sqrt{N} Y(0) / |Y(0)|$; no change of time is involved. If $Y(0)$ is independent of B and $X(0)$ has law μ , then X is this diffusion with initial law μ . In particular, if $Y(0)$ is independent of B and is either a standard Gaussian vector or a random vector with law v_N , then X is the diffusion generated by L_ρ with initial law v_N .*
3. *Let $\lambda > 0$ be an $\mathcal{F}_0$ -measurable random variable. Then λY is the unnormalized process with coefficients ρ and initial value $\lambda Y(0)$, driven by the same B . For $\tau \in \mathbb{R}$, the process $(e^{-2\tau t} Y(t))_{t \geq 0}$ is the unnormalized process with coefficients $\rho + \tau \mathbf{1}$ and initial value $Y(0)$, driven by the same B . In both cases X is unchanged.*
4. *The squares satisfy*

$$\begin{aligned} dZ_i &= \left(-4\rho_i Z_i + \frac{2}{N} \mathcal{A}\right) dt + dM_i, & dM_i &= 2\sqrt{2/N} Y_i |Y| dB_i, \\ d[M_i, M_j]_t &= \delta_{ij} \frac{8}{N} \mathcal{A}_t Z_i(t) dt. \end{aligned} \quad (10.5)$$

5. *Let $\mathcal{B} = (\mathcal{B}_{ij})_{i,j=1}^N$ be a matrix of independent standard Brownian motions, let X_0 be a random vector independent of $\mathcal{B}$ with $X_0 \neq 0$ almost surely, and let Φ be the solution of the linear matrix equation*

$$d\Phi(t) = -2D_\rho \Phi(t) dt + \sqrt{2/N} d\mathcal{B}(t) \Phi(t), \quad \Phi(0) = I. \quad (10.6)$$

Then almost surely, for all $t \geq 0$,

$$\det \Phi(t) = \exp \left\{ - \left(1 + 2 \sum_{i=1}^N \rho_i\right) t + \sqrt{2/N} \operatorname{Tr} \mathcal{B}(t) \right\} > 0. \quad (10.7)$$

The process Φ is independent of X_0 . The process $Y = \Phi X_0$ is the unnormalized process with coefficients ρ and initial value X_0 , with respect to the filtration generated by X_0 and $\mathcal{B}$ and driven by the Brownian motion $B(t) = \int_0^t d\mathcal{B}(s) Y(s) / |Y(s)|$, which is independent of X_0 . In particular, conditionally on $X_0 = y$, the process Y has the law of the unnormalized process started at y .

Proof. We begin with the first part. The drift $y \mapsto -2D_\rho y$ is linear, and the diffusion coefficient $y \mapsto \sqrt{2/N} |y| I$ is globally Lipschitz:

$$\|\sqrt{2/N} (|y| - |y'|) I\|_{\text{HS}} = \sqrt{2} ||y| - |y'|| \leq \sqrt{2} |y - y'|.$$

By [123, Theorem IX.2.1, p. 375], the equation (10.2) has a pathwise unique strong solution for every deterministic initial value. The same Picard construction, localized on $\{|Y(0)| \leq m\}$ and followed by $m \rightarrow \infty$, applies to every $\mathcal{F}_0$ -measurable initial value. For the conditional law, we use a stability estimate. If Y and Y' are two solutions driven by the same B , with $\mathcal{F}_0$ -measurable initial values, then $\Delta = Y - Y'$ satisfies $d\Delta = -2D_\rho \Delta dt + \sqrt{2/N} (|Y| - |Y'|) dB$. Doob's L^2 inequality [123, Theorem II.1.7], the Lipschitz bound above, and Gronwall's lemma, applied after localization and on events in $\mathcal{F}_0$, give

$$\mathbb{E} \left[\sup_{s \leq t} |\Delta_s|^2 \mid \mathcal{F}_0 \right] \leq C_t |\Delta_0|^2, \quad t \geq 0, \quad (10.8)$$

where C_t depends only on t and $\max_i |\rho_i|$. For deterministic y , let Y^y denote the solution started at y ; it is adapted to the filtration generated by B . Fix $t > 0$ and a bounded continuous function Φ on $C([0, t], \mathbb{R}^N)$, and let $\psi(y) = \mathbb{E}\Phi(Y^y)$; by (10.8), ψ is continuous. For $n \geq 1$, let $Y_n(0)$ be a measurable function of $Y(0)$ that takes countably many values $y_{n,k}$ and satisfies $|Y_n(0) - Y(0)| \leq 1/n$. Since stochastic integrals commute with multiplication by indicators of events in $\mathcal{F}_0$, the process $Y^{(n)} = \sum_k \mathbf{1}_{\{Y_n(0) = y_{n,k}\}} Y^{y_{n,k}}$ solves (10.2) with initial value $Y_n(0)$. As $Y(0)$ is independent of B , we have $\mathbb{E}[\Phi(Y^{(n)})g(Y(0))] = \mathbb{E}[\psi(Y_n(0))g(Y(0))]$ for every bounded measurable g . By (10.8), $Y^{(n)} \rightarrow Y$ uniformly on $[0, t]$ in probability, and $\psi(Y_n(0)) \rightarrow \psi(Y(0))$. We conclude that $\mathbb{E}[\Phi(Y)g(Y(0))] = \mathbb{E}[\psi(Y(0))g(Y(0))]$, which is the second assertion of the first part.

Let $\zeta = \inf\{t \geq 0 : \mathcal{A}_t = 0\}$. For $t < \zeta$, Itô's formula and the identity $\sum_i d[Y_i]_t = 2\mathcal{A}_t dt$ give

$$d\mathcal{A}_t = (2 - 4\bar{\rho}_t) \mathcal{A}_t dt + 2\sqrt{2/N} |Y_t| Y_t \cdot dB_t.$$

Let $\hat{B}_t = \int_0^t \mathbf{1}_{\{s < \zeta\}} |Y_s|^{-1} Y_s \cdot dB_s + \int_0^t \mathbf{1}_{\{s \geq \zeta\}} dB_1(s)$. Its integrands have unit length. Thus $\hat{B}$ is a continuous local martingale with $[\hat{B}]_t = t$, and by Lévy's characterization [123, Theorem IV.3.6], it is a standard Brownian motion. For $t < \zeta$, we obtain (10.4) from Itô's formula for $\log \mathcal{A}_t$; the term $-4/N$ is $-\frac{1}{2} \cdot \frac{8}{N}$. Since $|\bar{\rho}_t| \leq \max_i |\rho_i|$, we have

$$\log \mathcal{A}_t \geq \log \mathcal{A}_0 - \left(2 + 4 \max_i |\rho_i| + \frac{4}{N}\right) t - \sqrt{8/N} \sup_{s \leq t} |\hat{B}_s|, \quad t < \zeta.$$

On the event $\{\zeta < \infty\}$, the left side tends to $-\infty$ as $t \uparrow \zeta$, by continuity of $\mathcal{A}$, while the right side stays bounded below. Hence $\zeta = \infty$ almost surely, and (10.4) holds for all $t \geq 0$.

For the second part, the map $f(y) = \sqrt{N}y/|y|$ is smooth on $\mathbb{R}^N \setminus \{0\}$, where Y stays by the first part. With $\Pi(x) = I - xx^\top/N$ as in (3.2), its derivatives are

$$\partial_k f_i(y) = \frac{\sqrt{N}}{|y|} \Pi(f(y))_{ik}, \quad \sum_{k=1}^N \partial_k^2 f_i(y) = -\frac{(N-1)\sqrt{N}y_i}{|y|^3}.$$

The martingale parts of the coordinates of Y have covariations $(2/N)|Y|^2\delta_{kl}dt$, so the Itô correction in dX_i is $\frac{1}{2} \cdot \frac{2}{N}|Y|^2 \sum_k \partial_k^2 f_i(Y) dt = -\frac{N-1}{N} X_i dt$. Itô's formula gives

$$dX_t = \sqrt{2}\Pi(X_t)dB_t + \left(\Pi(X_t)(-2D_\rho)X_t - \frac{N-1}{N}X_t\right)dt, \quad (10.9)$$

where we used $\sqrt{N}Y/|Y| = X$ in the drift and $(\sqrt{N}/|Y|) \cdot \sqrt{2/N}|Y| = \sqrt{2}$ in the noise. In coordinates, the drift of X_i is $(-2\rho_i + \frac{2}{N}\sum_j \rho_j X_j^2 - \frac{N-1}{N})X_i$. Equation (10.9) is (3.2) with cW_N replaced by $A = -2D_\rho$. As shown in Section 3.1, the solutions of this equation started on S_N remain on S_N and form the diffusion generated by $L_A = L_\rho$. Its coefficients are smooth; the solution is therefore pathwise unique, and its law is determined by the law of its initial point. If $Y(0)$ is independent of B , then so is $X(0)$. This gives the assertion about the initial law μ . If $Y(0)$ is a standard Gaussian vector, then $X(0)$ has law ν_N by the first part of Proposition 5.6; if $Y(0)$ has law ν_N , then $X(0) = Y(0)$.

For the third part, since λ is $\mathcal{F}_0$ -measurable and $\lambda|Y| = |\lambda Y|$, multiplying (10.2) by λ shows that λY solves (10.2) with the same B . The product rule shows that $\tilde{Y}(t) = e^{-2\tau t}Y(t)$ satisfies

$$d\tilde{Y}_i = -2\tau\tilde{Y}_i dt + e^{-2\tau t}(-2\rho_i Y_i dt + \sqrt{2/N}|Y|dB_i) = -2(\rho_i + \tau)\tilde{Y}_i dt + \sqrt{2/N}|\tilde{Y}|dB_i.$$

In both cases Y is multiplied by a positive scalar process, which leaves X unchanged. The fourth part follows from $dZ_i = 2Y_i dY_i + d[Y_i]$ and $d[Y_i, Y_j]_t = \delta_{ij}(2/N)\mathcal{A}_t dt$.

For the last part, the linear equation (10.6) has a unique strong solution, which is a measurable function of $\mathcal{B}$ and is independent of X_0 . The determinant is a polynomial in the entries of Φ that is affine in each row of Φ . The martingale part of $d\Phi_{ij}$ is $\sqrt{2/N}\sum_k \Phi_{kj}d\mathcal{B}_{ik}$, by (10.6). In particular, two entries in different rows have zero covariation. Thus the second-order terms in Itô's formula for $\det \Phi$ vanish, and we have

$$\begin{aligned} d\det \Phi &= \text{Tr}(\text{adj}(\Phi)d\Phi) = \text{Tr}\left(\Phi \text{adj}(\Phi)\left(-2D_\rho dt + \sqrt{2/N}d\mathcal{B}\right)\right) \\ &= \det \Phi \left(-2\sum_i \rho_i dt + \sqrt{2/N}d\text{Tr} \mathcal{B}\right), \end{aligned}$$

where $\text{adj}(\Phi)$ is the adjugate matrix, which satisfies $\Phi \text{adj}(\Phi) = \det(\Phi)I$. The process $\text{Tr} \mathcal{B}$ is a Brownian motion with variance N per unit time. The linear equation $dD = D dS$, $D_0 = 1$, with $S_t = -2t\sum_i \rho_i + \sqrt{2/N}\text{Tr} \mathcal{B}(t)$, has the unique solution $D_t = \exp(S_t - [S]_t/2)$, and $[S]_t = 2t$. This proves (10.7). In particular, $\Phi(t)$ is invertible for all t ,

and $Y = \Phi X_0$ never vanishes. Let $(\mathcal{G}_t)$ be the usual augmentation of the filtration generated by X_0 and $\mathcal{B}$; by independence, $\mathcal{B}$ is a $(\mathcal{G}_t)$ -Brownian motion. The process $B_i(t) = \sum_k \int_0^t Y_k(s) |Y(s)|^{-1} d\mathcal{B}_{ik}(s)$ satisfies $d[B_i, B_j]_t = \delta_{ij} dt$. Lévy's characterization implies that B is a $(\mathcal{G}_t)$ -Brownian motion, and it is independent of $\mathcal{G}_0 \ni X_0$. Multiplying (10.6) by X_0 on the right, we obtain

$$dY_i = -2\rho_i Y_i dt + \sqrt{2/N} \sum_k Y_k d\mathcal{B}_{ik} = -2\rho_i Y_i dt + \sqrt{2/N} |Y| d\mathcal{B}_i,$$

so Y solves (10.2) with initial value X_0 on the filtered space with filtration $(\mathcal{G}_t)$, and the last assertion follows from the first part. $\square$

Remark 10.3 (The two choices of coefficients). Let A be a real symmetric matrix with Lagrange multiplier z and precisions $p_i = z - \alpha_i$ (Definition 3.3), and work in eigencoordinates of A . On S_N we have $-\frac{1}{2} \sum_i p_i x_i^2 = \frac{1}{2} x^\top A x - \frac{zN}{2}$, so $L_\rho = L_A$ for the half-precisions $\rho_i = r_i = p_i/2$. The unnormalized process with these coefficients has drift rates $2r_i = p_i$ and solves

$$dY_i = -2r_i Y_i dt + \sqrt{2/N} |Y| d\mathcal{B}_i = -p_i Y_i dt + \sqrt{2/N} |Y| d\mathcal{B}_i. \quad (10.10)$$

This choice is used in the critical window and above the critical temperature, and it is the equation (1.15) of the introduction. For the model at low temperature, we use instead the scaled eigenvalue gaps $\rho_i = a_i = c(\lambda_1 - \lambda_i)/2$ of (3.8) as coefficients, for which L_ρ is the generator with potential $F = \sum_{i \geq 2} a_i x_i^2$; the drift rates are $2a_i$, and the equation reads

$$dY_i = -2a_i Y_i dt + \sqrt{2\mathcal{A}_t/N} d\mathcal{B}_i, \quad \mathcal{A}_t = |Y(t)|^2. \quad (10.11)$$

Since $a_1 = 0$, the process $Y_1(t) - Y_1(0)$ is a continuous local martingale in this case; if $\mathbb{E}|Y(0)|^2 < \infty$, then Y_1 is a square-integrable martingale on every bounded time interval, because $e^{-2s} \mathcal{A}_s$ is a nonnegative local supermartingale by (10.5), so that $\mathbb{E} \mathcal{A}_s \leq e^{2s} \mathbb{E} \mathcal{A}_0$. For $A = cW_N$, we have $r_i = a_i + d_N/2$, where $d_N = z_N - c\lambda_1$. Using the third part of Lemma 10.2, we see that the two processes with the same initial value and the same Brownian motion are related by $Y^{(r)}(t) = e^{-d_N t} Y^{(a)}(t)$, and that they have the same direction X .

In [51], Cugliandolo and Dean diagonalized the interaction of this model and, in the limit $N \rightarrow \infty$, reduced its dynamics to linear equations for the eigencoordinates and a scalar Volterra equation for the Lagrange multiplier [51, Section II, (II.2)–(II.7)]. The unnormalized process is an exact analogue at finite N : its coordinates have linear drifts, and when $\mathbb{E}|Y(0)|^2 < \infty$, the expectations of their squares solve the closed linear equation of Section 7 (see (10.21) below), while its noise amplitude $|Y|$ is random.

10.2 Positive martingales

The following criterion converts a bound on conditional relative variances of a positive martingale into bounds on its maximum, on the inverse of its minimum, and on the quadratic variation of its logarithm. It is used in Sections 10.4 and 10.5 for martingales built from the squares Z_i , and its hypothesis is verified separately in each case.

Lemma 10.4 (Maximum, minimum, and the quadratic variation of $\log V$). *Let $T > 0$, let $V = (V_t)_{0 \leq t \leq T}$ be a strictly positive continuous square-integrable martingale, and let $\varepsilon \geq 0$. Suppose that for every stopping time $\tau \leq T$,*

$$\mathbb{E}[(V_T - V_\tau)^2 \mid \mathcal{F}_\tau] \leq \varepsilon V_\tau^2. \quad (10.12)$$

1. *For every $p > 0$ with $2^p \cdot 4\varepsilon < 1$ and every stopping time $\tau \leq T$,*

$$\mathbb{E}\left[\sup_{\tau \leq s \leq T} \left(\frac{V_\tau}{V_s}\right)^p \mid \mathcal{F}_\tau\right] \leq \frac{2^p}{1 - 2^p \cdot 4\varepsilon}. \quad (10.13)$$

2. *For every $p > 0$ with $2^p \varepsilon < 1$ and every stopping time $\tau \leq T$,*

$$\mathbb{E}\left[\sup_{\tau \leq s \leq T} \left(\frac{V_s}{V_\tau}\right)^p \mid \mathcal{F}_\tau\right] \leq \frac{2^p}{1 - 2^p \varepsilon}. \quad (10.14)$$

3. *Suppose in addition that $\varepsilon \leq 1/32$, and that the quadratic variation of the logarithm of V has the form*

$$Q_t = \int_0^t V_s^{-2} d[V]_s = \int_0^t q_s ds, \quad 0 \leq t \leq T,$$

where $q \geq 0$ is progressively measurable and $Q_T \leq \bar{Q}$ for a deterministic constant $\bar{Q}$. Then for every stopping time $\tau \leq T$,

$$\mathbb{E}[Q_T - Q_\tau \mid \mathcal{F}_\tau] \leq K, \quad K = 6\sqrt{2\varepsilon}, \quad (10.15)$$

and for every $\lambda \geq 0$ with $\lambda K < 1$,

$$\mathbb{E}[e^{\lambda(Q_T - Q_\tau)} \mid \mathcal{F}_\tau] \leq \frac{1}{1 - \lambda K}. \quad (10.16)$$

The constants in (10.13)–(10.16) depend only on p , ε , and λ ; they do not depend on T or $\bar{Q}$.

The bound $\bar{Q}$ is used only to show that the stochastic logarithm of V is a square-integrable martingale. The successive hitting-time argument in Steps 1–3 follows the classical martingale John–Nirenberg proof pattern; see [123, Chapter II, Exercise (3.16), part 5, pp. 75–76]. We prove the conditional ordered-integral estimate (10.16) in Step 4 below; its Markov-process antecedent is [95, Lemma 3, pp. 337–338].

Proof. The proof has four steps: a bound for the probability that V moves by a fixed fraction of its current value, its iteration at successive hitting times, the resulting moment bounds, and the two bounds on Q .

Step 1: one step. Fix a stopping time $\tau \leq T$ and $\theta > 0$, and let

$$\sigma = \inf\{s \geq \tau : |V_s - V_\tau| \geq \theta V_\tau\} \wedge T, \quad G_\theta = \left\{ \sup_{\tau \leq s \leq T} |V_s - V_\tau| \geq \theta V_\tau \right\}.$$

By optional sampling, $\mathbb{E}[V_T - V_\tau \mid \mathcal{F}_\sigma] = V_\sigma - V_\tau$, and by the conditional Jensen inequality, $\mathbb{E}[(V_T - V_\tau)^2 \mid \mathcal{F}_\sigma] \geq (V_\sigma - V_\tau)^2$. On G_θ , we have $|V_\sigma - V_\tau| = \theta V_\tau$ by continuity of V . Taking conditional expectations given $\mathcal{F}_\tau$ and using (10.12), we obtain

$$\varepsilon V_\tau^2 \geq \mathbb{E}[(V_\sigma - V_\tau)^2 \mid \mathcal{F}_\tau] \geq \theta^2 V_\tau^2 \mathbb{P}(G_\theta \mid \mathcal{F}_\tau),$$

so $\mathbb{P}(G_\theta \mid \mathcal{F}_\tau) \leq \varepsilon/\theta^2$. The events $\{\inf_{\tau \leq s \leq T} V_s \leq V_\tau/2\}$ and $\{\sup_{\tau \leq s \leq T} V_s \geq 2V_\tau\}$ are contained in $G_{1/2}$ and G_1 , respectively. We conclude that

$$\mathbb{P}\left(\inf_{\tau \leq s \leq T} V_s \leq \frac{V_\tau}{2} \mid \mathcal{F}_\tau\right) \leq 4\varepsilon, \quad \mathbb{P}\left(\sup_{\tau \leq s \leq T} V_s \geq 2V_\tau \mid \mathcal{F}_\tau\right) \leq \varepsilon. \quad (10.17)$$

Step 2: iteration. Fix τ , and for $j \geq 0$ let $\sigma_j = \inf\{s \geq \tau : V_s \leq 2^{-j}V_\tau\}$, with $\inf \emptyset = \infty$, and $E_j = \{\sigma_j \leq T\}$. Then $\sigma_0 = \tau$, $E_j \subset E_{j-1}$, and $E_{j-1} \in \mathcal{F}_{\sigma_{j-1} \wedge T}$. On E_{j-1} , continuity gives $V_{\sigma_{j-1}} = 2^{1-j}V_\tau$, and V stays above $2^{-j}V_\tau$ before σ_{j-1} , so

$$E_j = E_{j-1} \cap \left\{ \inf_{\sigma_{j-1} \wedge T \leq s \leq T} V_s \leq \frac{V_{\sigma_{j-1} \wedge T}}{2} \right\}.$$

Applying (10.17) at the stopping time $\sigma_{j-1} \wedge T$ and conditioning on $\mathcal{F}_\tau$, we find $\mathbb{P}(E_j \mid \mathcal{F}_\tau) \leq 4\varepsilon \mathbb{P}(E_{j-1} \mid \mathcal{F}_\tau)$. By induction,

$$\mathbb{P}\left(\sup_{\tau \leq s \leq T} \frac{V_\tau}{V_s} \geq 2^j \mid \mathcal{F}_\tau\right) = \mathbb{P}(E_j \mid \mathcal{F}_\tau) \leq (4\varepsilon)^j, \quad j \geq 0.$$

The same argument with the successive first hitting times of the levels $2^j V_\tau$ and the second bound in (10.17) yields

$$\mathbb{P}\left(\sup_{\tau \leq s \leq T} \frac{V_s}{V_\tau} \geq 2^j \mid \mathcal{F}_\tau\right) \leq \varepsilon^j, \quad j \geq 0.$$

Step 3: moments. Let $R = \sup_{\tau \leq s \leq T} V_\tau/V_s$, which is finite and at least one. Splitting according to the dyadic interval $[2^j, 2^{j+1})$ that contains R , we have

$$\mathbb{E}[R^p \mid \mathcal{F}_\tau] \leq \sum_{j \geq 0} 2^{p(j+1)} \mathbb{P}(R \geq 2^j \mid \mathcal{F}_\tau) \leq 2^p \sum_{j \geq 0} (2^p \cdot 4\varepsilon)^j = \frac{2^p}{1 - 2^p \cdot 4\varepsilon},$$

which is (10.13). With the second bound of Step 2 in place of the first, the same computation gives (10.14).

Step 4: the quadratic variation of the logarithm. Let $\varepsilon \leq 1/32$, fix a stopping time $\tau \leq T$, and let $R' = V_T/V_\tau$. The hypothesis (10.12) states that $\mathbb{E}[(R' - 1)^2 \mid \mathcal{F}_\tau] \leq \varepsilon$. By (10.13) with $p = 2$, for which $2^2 \cdot 4\varepsilon \leq 1/2$, we have $\mathbb{E}[R'^{-2} \mid \mathcal{F}_\tau] \leq 8$, and then $\mathbb{E}[(1 + R'^{-1})^2 \mid \mathcal{F}_\tau] \leq 2(1 + 8) = 18$. For every $r > 0$, we have $|\log r| \leq |r - 1|(1 + r^{-1})$: for $r \geq 1$ this follows from $\log r \leq r - 1$, and for $r < 1$ from $\log(1/r) \leq 1/r - 1 = (1 - r)/r$. Using the conditional Cauchy–Schwarz inequality, we find

$$\mathbb{E}[|\log R'| \mid \mathcal{F}_\tau] \leq \sqrt{\varepsilon} \cdot \sqrt{18} = 3\sqrt{2\varepsilon}.$$

Since $[\mathcal{L}]_T = Q_T \leq \bar{Q}$, the stochastic logarithm $\mathcal{L}_t = \int_0^t V_s^{-1} dV_s$ is a square-integrable martingale on $[0, T]$, and by Itô's formula, $\log V_t = \log V_0 + \mathcal{L}_t - \frac{1}{2}Q_t$. Taking conditional expectations, we have

$$\mathbb{E}[Q_T - Q_\tau \mid \mathcal{F}_\tau] = -2 \mathbb{E}[\log R' \mid \mathcal{F}_\tau] \leq 6\sqrt{2\varepsilon},$$

which is (10.15).

For (10.16), we show by induction that $\mathbb{E}[(Q_T - Q_\tau)^j \mid \mathcal{F}_\tau] \leq j! K^j$ for every stopping time $\tau \leq T$ and every integer $j \geq 1$. The case $j = 1$ is (10.15). Since $s \mapsto (Q_T - Q_s)^j$ is absolutely continuous with derivative $-jq_s(Q_T - Q_s)^{j-1}$, we have

$$(Q_T - Q_\tau)^j = j \int_0^T \mathbf{1}_{\{s > \tau\}} q_s (Q_T - Q_s)^{j-1} ds.$$

The factor $\mathbf{1}_{\{s > \tau\}} q_s$ is $\mathcal{F}_s$ -measurable. By the conditional Tonelli theorem, the tower property, and the induction hypothesis at the deterministic time s , we obtain

$$\begin{aligned} \mathbb{E}[(Q_T - Q_\tau)^j \mid \mathcal{F}_\tau] &= j \int_0^T \mathbb{E}[\mathbf{1}_{\{s > \tau\}} q_s \mathbb{E}[(Q_T - Q_s)^{j-1} \mid \mathcal{F}_s] \mid \mathcal{F}_\tau] ds \\ &\leq j! K^{j-1} \mathbb{E}[Q_T - Q_\tau \mid \mathcal{F}_\tau] \leq j! K^j. \end{aligned}$$

Summing the exponential series by monotone convergence, we find $\mathbb{E}[e^{\lambda(Q_T - Q_\tau)} \mid \mathcal{F}_\tau] \leq \sum_{j \geq 0} (\lambda K)^j = (1 - \lambda K)^{-1}$. $\square$

10.3 Moment bounds with backward semigroup weights

In this section, $\rho \in [0, \infty)^N$. We write

$$H_\rho = -4D_\rho + \frac{2}{N} \mathbf{1}\mathbf{1}^\top, \quad h(t) = e^{tH_\rho} \mathbf{1}, \quad \bar{h}(t) = \frac{1}{N} \mathbf{1}^\top h(t), \quad m(t) = \max_i h_i(t), \quad (10.18)$$

and, for $y \in [0, \infty)^N$, we set $\mathcal{H}_y(t) = h(t)^\top y$. When $\rho_i = p_i/2$ for precisions p_i , the matrix $H_\rho = -2 \operatorname{diag}(p) + (2/N) \mathbf{1}\mathbf{1}^\top$ is the matrix H of Section 7. The matrix H_ρ is symmetric and its off-diagonal entries are nonnegative. For $c_\rho = 4 \max_i \rho_i$, the matrix

$H_\rho + c_\rho I$ has nonnegative entries, so $e^{tH_\rho} = e^{-c_\rho t} e^{t(H_\rho + c_\rho I)}$ has nonnegative entries; this is the positivity used in the first part of Lemma 7.1. By the variation of constants formula for the coordinates of the equation $h' = H_\rho h$, we have

$$h_i(t) = e^{-4\rho_i t} + \frac{2}{N} \int_0^t e^{-4\rho_i(t-s)} N\bar{h}(s) ds \geq e^{-4\rho_i t} > 0. \quad (10.19)$$

In particular, $\mathcal{H}_y(t) > 0$ for every nonzero $y \in [0, \infty)^N$. We have $\mathbf{1}^\top e^{tH_\rho} y = \mathcal{H}_y(t)$ by symmetry of H_ρ , and the same argument applied to $e^{tH_\rho} y$ yields

$$(e^{tH_\rho} y)_i = e^{-4\rho_i t} y_i + \frac{2}{N} \int_0^t e^{-4\rho_i(t-s)} \mathcal{H}_y(s) ds. \quad (10.20)$$

For the unnormalized process with coefficients $\rho \in [0, \infty)^N$ and $\mathbb{E}|Y(0)|^2 < \infty$, taking expectations in (10.5) gives

$$\mathbb{E}Z_t = e^{tH_\rho} \mathbb{E}Z_0, \quad t \geq 0. \quad (10.21)$$

Indeed, for a deterministic initial value, the local martingales M_i in (10.5) are square integrable on bounded intervals by the moment bound (10.26) proved below, so $\mathbb{E}Z_t$ solves the linear equation with matrix H_ρ ; the general case follows by conditioning on $Y(0)$, using the first part of Lemma 10.2.

The squares of the unnormalized process satisfy the following equations with $\epsilon \equiv 0$, by (10.5). The estimates below also allow a nonnegative perturbation ϵ of the drift, which may depend on the past of the process. We say that a continuous adapted process Z with values in $[0, \infty)^N$ and with $\mathcal{A}_t = \sum_i Z_i(t) > 0$ for all t satisfies (10.22) with perturbation bound $E_* \geq 0$ if

$$dZ_i = \left[-4\rho_i Z_i + (2 + \epsilon_i(t)) \frac{\mathcal{A}}{N} \right] dt + dM_i, \quad d[M_i, M_j]_t = \delta_{ij} \frac{8}{N} \mathcal{A}_t Z_i(t) dt, \quad (10.22)$$

where the M_i are continuous local martingales and the ϵ_i are progressively measurable with $\epsilon_i \geq 0$ and $\sum_i \epsilon_i(t) \leq E_*$ for all t .

The following lemma bounds moments of $h(u)^\top Z_t$ by the corresponding power of $\mathcal{H}_{Z(0)}(u+t)$. When $\epsilon \equiv 0$, the process $s \mapsto h(u+t-s)^\top Z_s$, $0 \leq s \leq u+t$, is a local martingale with initial value $\mathcal{H}_{Z(0)}(u+t)$, which we call a martingale with backward semigroup weights, and the lemma bounds its second and higher moments by powers of its initial value. The size of the bound is governed by the quantity

$$\Lambda(\ell, \Theta) = \sup \left\{ \int_0^t m(w-s) \frac{\mathcal{H}_y(s)}{\mathcal{H}_y(w)} ds : y \in [0, \infty)^N \setminus \{0\}, 0 \leq t \leq \ell, t \leq w \leq \Theta \right\}, \quad (10.23)$$

which depends only on ρ , ℓ , and Θ .

Lemma 10.5 (Moment bounds with backward semigroup weights). *Let $\rho \in [0, \infty)^N$, let Z satisfy (10.22) with perturbation bound E_* , let $k \geq 2$ be an integer, and let $0 < \ell \leq \Theta < \infty$. Set*

$$\alpha = \frac{k(E_* + 4k - 4)}{N} \Lambda(\ell, \Theta), \quad (10.24)$$

and suppose that $\alpha < 1$. Then for every bounded stopping time τ , and all $\mathcal{F}_\tau$ -measurable random variables $u \geq 0$ and $t \in [0, \ell]$ with $u + t \leq \Theta$,

$$\mathbb{E}[(h(u)^\top Z_{\tau+t})^k \mid \mathcal{F}_\tau] \leq \frac{\mathcal{H}_{Z_\tau}(u+t)^k}{1-\alpha} \quad \text{almost surely.} \quad (10.25)$$

In particular, if $Z(0) = y$ is deterministic, then $\mathbb{E}[(h(u)^\top Z_t)^k] \leq \mathcal{H}_y(u+t)^k / (1-\alpha)$ for all $t \in [0, \ell]$ and $u \geq 0$ with $u+t \leq \Theta$.

The conditional statement allows the drift perturbation ϵ after time τ to depend on the whole past; no Markov property is used.

Proof. The proof has three steps. We first prove a crude bound on moments of $\mathcal{A}$, which shows that a normalized supremum $\mathcal{B}$ of the left side of (10.25) is finite. We then compute the evolution of $h(w-s)^\top Z_{\tau+s}$ with Itô's formula, and finally we show that $\mathcal{B} \leq 1 + \alpha\mathcal{B}$.

Step 1: a crude moment bound. Let $\mathcal{M} = \sum_i M_i$; as the M_i have zero mutual covariation, $d[\mathcal{M}]_t = (8/N)\mathcal{A}_t^2 dt$. Summing (10.22) and using $\rho \geq 0$ and $\sum_i \epsilon_i \leq E_*$, we see that $\mathcal{A}$ has drift at most $(2 + E_*/N)\mathcal{A}$. For an integer $j \geq 1$, Itô's formula shows that the drift of $\mathcal{A}^j$ is at most $c_j \mathcal{A}^j$, with

$$c_j = 2j + \frac{jE_* + 4j(j-1)}{N}.$$

Fix a bounded stopping time σ . Relative to the filtration $(\mathcal{F}_{\sigma+s})_{s \geq 0}$, Itô's formula shows that

$$R_s = e^{-c_j s} \left(\frac{\mathcal{A}_{\sigma+s}}{\mathcal{A}_\sigma} \right)^j$$

is a nonnegative local supermartingale with $R_0 = 1$; by Fatou's lemma, it is a supermartingale, without any moment assumption on $\mathcal{A}_\sigma$. If $\sigma \leq \sigma' \leq \sigma + \ell$ for a stopping time σ' , then $\sigma' - \sigma$ is a stopping time for the shifted filtration, and optional sampling gives $\mathbb{E}[e^{-c_j(\sigma'-\sigma)} (\mathcal{A}_{\sigma'}/\mathcal{A}_\sigma)^j \mid \mathcal{F}_\sigma] \leq 1$. Since $c_j \geq 0$ and $\mathcal{A}_\sigma$ is $\mathcal{F}_\sigma$ -measurable, we obtain

$$\mathbb{E}[\mathcal{A}_{\sigma'}^j \mid \mathcal{F}_\sigma] \leq e^{c_j \ell} \mathcal{A}_\sigma^j. \quad (10.26)$$

Let $m_\Theta = \max_{s \leq \Theta} m(s)$ and $h_\Theta = \min_{s \leq \Theta} \min_i h_i(s)$, which is positive by (10.19). Call a pair (u, t) of $\mathcal{F}_\tau$ -measurable random variables admissible if $u \geq 0$, $t \in [0, \ell]$, and $u + t \leq \Theta$. For an admissible pair, $\tau + t$ is a bounded stopping time. We set

$$\beta(u, t) = \frac{\mathbb{E}[(h(u)^\top Z_{\tau+t})^k \mid \mathcal{F}_\tau]}{\mathcal{H}_{Z_\tau}(u+t)^k}.$$

We have $h(u)^\top Z_{\tau+t} \leq m_\Theta \mathcal{A}_{\tau+t}$ and $\mathcal{H}_{Z_\tau}(u+t) \geq h_\Theta \mathcal{A}_\tau$, and (10.26) implies that $\beta(u, t) \leq (m_\Theta/h_\Theta)^k e^{c_k \ell}$. Let $\mathcal{B}$ be the essential supremum of $\beta(u, t)$ over all admissible pairs. It is an $\mathcal{F}_\tau$ -measurable random variable, $\beta(u, t) \leq \mathcal{B}$ almost surely for every admissible pair, and

$$\mathcal{B} \leq (m_\Theta/h_\Theta)^k e^{c_k \ell} < \infty. \quad (10.27)$$

Step 2: the backward semigroup weights. Fix an admissible pair (u, t) , let $w = u + t$, and set $X_s = h(w-s)^\top Z_{\tau+s}$ for $0 \leq s \leq t$. Since $h' = H_\rho h$ and H_ρ is symmetric, the terms of (10.22) that involve H_ρ cancel, and we have

$$dX_s = D_s ds + \sum_i h_i(w-s) dM_i(\tau+s), \quad D_s = \frac{\mathcal{A}_{\tau+s}}{N} \sum_i \epsilon_i(\tau+s) h_i(w-s).$$

The drift satisfies $0 \leq D_s \leq (E_*/N) m(w-s) \mathcal{A}_{\tau+s}$, and since $\sum_i h_i^2 Z_i \leq m \sum_i h_i Z_i$, the quadratic variation satisfies

$$d[X]_s = \frac{8}{N} \mathcal{A}_{\tau+s} \sum_i h_i(w-s)^2 Z_i(\tau+s) ds \leq \frac{8}{N} m(w-s) \mathcal{A}_{\tau+s} X_s ds.$$

Applying Itô's formula to X^k , we see that the drift of X^k is at most

$$kX_s^{k-1} D_s + \frac{k(k-1)}{2} X_s^{k-2} \frac{d[X]_s}{ds} \leq \frac{k(E_* + 4k - 4)}{N} m(w-s) \mathcal{A}_{\tau+s} X_s^{k-1}.$$

We stop at the first time that $\mathcal{A}_{\tau+s}$ exceeds n , take conditional expectations given $\mathcal{F}_\tau$, and let $n \rightarrow \infty$, using conditional Fatou on the left and monotone convergence on the right. Since $X_0 = \mathcal{H}_{Z_\tau}(w)$ and $X_t = h(u)^\top Z_{\tau+t}$, this gives

$$\begin{aligned} \mathbb{E}[(h(u)^\top Z_{\tau+t})^k \mid \mathcal{F}_\tau] &\leq \mathcal{H}_{Z_\tau}(w)^k \\ &\quad + \frac{k(E_* + 4k - 4)}{N} \int_0^\ell \mathbf{1}_{\{s < t\}} m(w - (s \wedge t)) \mathbb{E}[\mathcal{A}_{\tau+s} X_{s \wedge t}^{k-1} \mid \mathcal{F}_\tau] ds. \end{aligned} \quad (10.28)$$

Here we used that $\mathbf{1}_{\{s < t\}}$ and $m(w - (s \wedge t))$ are $\mathcal{F}_\tau$ -measurable, which allows us to take them out of the conditional expectation; the argument $w - (s \wedge t)$ is nonnegative, and on $\{s < t\}$ it equals $w - s$.

Step 3: absorption. For $s \leq \ell$, both $(0, s)$ and $(w - s \wedge t, s \wedge t)$ are admissible pairs, and $\mathcal{A}_{\tau+s} = h(0)^\top Z_{\tau+s}$. By the conditional Hölder inequality and the definition of $\mathcal{B}$, on the event $\{s < t\}$,

$$\mathbb{E}[\mathcal{A}_{\tau+s} X_s^{k-1} \mid \mathcal{F}_\tau] \leq \mathbb{E}[\mathcal{A}_{\tau+s}^k \mid \mathcal{F}_\tau]^{1/k} \mathbb{E}[X_s^k \mid \mathcal{F}_\tau]^{(k-1)/k} \leq \mathcal{B} \mathcal{H}_{Z_\tau}(s) \mathcal{H}_{Z_\tau}(w)^{k-1}.$$

Inserting this into (10.28) and dividing by $\mathcal{H}_{Z_\tau}(w)^k$, we obtain

$$\beta(u, t) \leq 1 + \frac{k(E_* + 4k - 4)}{N} \mathcal{B} \int_0^t m(w - s) \frac{\mathcal{H}_{Z_\tau}(s)}{\mathcal{H}_{Z_\tau}(w)} ds \leq 1 + \alpha \mathcal{B},$$

where in the last step we used the definition (10.23) with $y = Z_\tau$, which is a nonzero vector in $[0, \infty)^N$. This holds almost surely for every admissible pair, and we conclude that $\mathcal{B} \leq 1 + \alpha \mathcal{B}$ almost surely. Since $\mathcal{B}$ is finite by (10.27) and $\alpha < 1$, we conclude that $\mathcal{B} \leq (1 - \alpha)^{-1}$, which is (10.25). $\square$

The next lemma bounds the moments of a single square by moments of the squared norm.

Lemma 10.6 (Moments of single coordinates). *Let $\rho \in [0, \infty)^N$, let $k \geq 1$ be an integer, and let Y be the unnormalized process with coefficients ρ and an $\mathcal{F}_0$ -measurable initial value with $\mathbb{E}|Y(0)|^{4k} < \infty$. Then for every i and $t \geq 0$,*

$$\|Z_i(t)\|_k \leq e^{-4\rho_i t} \|Z_i(0)\|_k + \frac{2(2k-1)}{N} \int_0^t e^{-4\rho_i(t-s)} \|\mathcal{A}_s\|_k ds. \quad (10.29)$$

Proof. All moments used below are finite and locally bounded in time, by (10.26) with $\sigma = 0$ and $j = 2k$. Using (10.5) and Itô's formula, we find

$$d(Z_i^k) = \left[-4k\rho_i Z_i^k + \frac{2k}{N} \mathcal{A}Z_i^{k-1} + \frac{k(k-1)}{2} \cdot \frac{8}{N} \mathcal{A}Z_i^{k-1} \right] dt + kZ_i^{k-1} dM_i,$$

and $2k + 4k(k-1) = 2k(2k-1)$. The martingale term has integrable quadratic variation, since $Z_i^{2k-2} \mathcal{A}Z_i \leq \mathcal{A}^{2k}$. Taking expectations, we see that $\phi(t) = \mathbb{E}Z_i(t)^k$ is absolutely continuous, and by Hölder's inequality,

$$\begin{aligned} \phi'(t) &= -4k\rho_i \phi(t) + \frac{2k(2k-1)}{N} \mathbb{E}[\mathcal{A}_t Z_i(t)^{k-1}] \\ &\leq -4k\rho_i \phi(t) + \frac{2k(2k-1)}{N} \|\mathcal{A}_t\|_k \phi(t)^{(k-1)/k}. \end{aligned}$$

For $\eta > 0$, let $\psi_\eta = (\phi + \eta)^{1/k}$. Using $\phi(\phi + \eta)^{1/k-1} \geq (\phi + \eta)^{1/k} - \eta^{1/k}$ and $\phi^{(k-1)/k}(\phi + \eta)^{1/k-1} \leq 1$, we find

$$\psi'_\eta(t) \leq -4\rho_i \psi_\eta(t) + 4\rho_i \eta^{1/k} + \frac{2(2k-1)}{N} \|\mathcal{A}_t\|_k.$$

Integrating this linear differential inequality and letting $\eta \downarrow 0$, we obtain (10.29). $\square$

10.4 The critical window

In this section, we specialize Lemmas 10.5 and 10.6 to the half-precisions of the critical window. Let $0 < r_1 \leq \dots \leq r_N$ be real numbers, and suppose that for fixed constants $0 < c_0 \leq C_0 < \infty$,

$$\sum_{i=1}^N \frac{1}{2r_i} = N, \quad c_0 \left(\frac{i}{N} \right)^{2/3} \leq r_i \leq C_0 \left(\frac{i}{N} \right)^{2/3} \quad (1 \leq i \leq N). \quad (10.30)$$

These are the hypotheses (7.49) of Lemma 7.14, and they are the first two of the critical spectral conditions (4.36); the bound on the first gap in (4.36) is not used in this section. For the model, fix $\sigma_0 > 0$ and $\eta \in (0, 1)$; the second part of Lemma 4.17 supplies constants depending only on σ_0 and η , and events $\mathcal{E}_N$ with $\liminf_{N \rightarrow \infty} \mathbb{P}(\mathcal{E}_N) \geq 1 - \eta$, on which (4.36) holds simultaneously for the half-precisions of $(1 + \sigma N^{-1/3})W_N$, $|\sigma| \leq \sigma_0$, when $N > (2\sigma_0)^3$. Here the probability is over the disorder W_N , and the constants are fixed before $N \rightarrow \infty$. Note that $r_i \leq C_0$ for all i . We take $\rho = r$ in (10.18); then H_r is the matrix H of Section 7 with $p = 2r$. By Lemma 7.14, for every fixed $K < \infty$, uniformly in $0 \leq t \leq KN^{2/3}$,

$$N\bar{h}(t) \asymp \frac{N}{\sqrt{1+t}}, \quad h_i(t) \asymp \frac{\sqrt{1+t}}{1+r_it}, \quad (10.31)$$

with constants depending only on c_0, C_0 , and K .

The first two statements of this section assume only a two-sided bound on the coordinates of h on a general time interval $[0, \Theta]$; they do not use (10.30).

Corollary 10.7 (Second-moment bounds with backward semigroup weights). *Let $N \geq 1$, let $0 < r_i \leq C_0$ for $1 \leq i \leq N$, and let $h(t) = e^{tH_r} \mathbf{1}$ with $H_r = -4 \operatorname{diag}(r) + (2/N) \mathbf{1}\mathbf{1}^\top$. Fix $\Theta > 0$, and assume that for $0 \leq t \leq \Theta$,*

$$c_* \frac{\sqrt{1+t}}{1+r_it} \leq h_i(t) \leq C_* \frac{\sqrt{1+t}}{1+r_it}, \quad 1 \leq i \leq N. \quad (10.32)$$

Let Z satisfy (10.22) with $\rho = r$ and perturbation bound E_ . There exists a constant C depending only on c_*, C_*, C_0 , and E_* such that the following holds. For $0 < \ell \leq \Theta$, let $\alpha = C(1 + \Theta)\sqrt{\ell}/N$. If $\alpha < 1$, then every deterministic nonzero initial state $Z(0) = y \in [0, \infty)^N$ satisfies*

$$\mathbb{E}[(h(u)^\top Z_t)^2] \leq \frac{[h(u+t)^\top y]^2}{1-\alpha}, \quad 0 \leq t \leq \ell, \quad u \geq 0, \quad u+t \leq \Theta. \quad (10.33)$$

The same bound holds conditionally on $\mathcal{F}_\tau$, for every bounded stopping time τ , with y replaced by Z_τ and with $\mathcal{F}_\tau$ -measurable u and t in the same ranges.

Proof. We bound $\Lambda(\ell, \Theta)$ and apply Lemma 10.5 with $k = 2$. Let $0 \leq s \leq w \leq \Theta$ and $y \in [0, \infty)^N \setminus \{0\}$. The ratio $\mathcal{H}_y(s)/\mathcal{H}_y(w)$ is at most $\max_i h_i(s)/h_i(w)$, because $y \geq 0$.

The function $x \mapsto (1 + xw)/(1 + xs)$ is nondecreasing, and for $r_i \leq 1$ it is at most $(1 + w)/(1 + s)$. For $1 \leq r_i \leq C_0$, we have $(1 + r_i w)/(1 + r_i s) \leq r_i(1 + w)/(1 + s)$. Thus

$$\frac{1 + r_i w}{1 + r_i s} \leq \max(1, C_0) \frac{1 + w}{1 + s},$$

and by (10.32), we have

$$\frac{h_i(s)}{h_i(w)} \leq \frac{C_*}{c_*} \sqrt{\frac{1 + s}{1 + w}} \cdot \frac{1 + r_i w}{1 + r_i s} \leq \frac{C_*}{c_*} \max(1, C_0) \sqrt{\frac{1 + w}{1 + s}}.$$

Together with $m(w - s) \leq C_* \sqrt{1 + w}$, this yields

$$m(w - s) \frac{\mathcal{H}_y(s)}{\mathcal{H}_y(w)} \leq C_1 \frac{1 + w}{\sqrt{1 + s}}, \quad C_1 = \frac{C_*^2}{c_*} \max(1, C_0). \quad (10.34)$$

Since $\int_0^t (1 + s)^{-1/2} ds \leq 2\sqrt{t}$, we have $\Lambda(\ell, \Theta) \leq 2C_1(1 + \Theta)\sqrt{\ell}$. With $k = 2$, the quantity (10.24) is at most $(2E_* + 8) \cdot 2C_1(1 + \Theta)\sqrt{\ell}/N$. Thus Lemma 10.5 applies with $C = 2(2E_* + 8)C_1$ and gives (10.33) and its conditional form. $\square$

Corollary 10.8 (Second moments on short intervals in the critical window). *Fix $K > 0$, and suppose that the hypotheses of Corollary 10.7 hold with $\Theta = KN^{2/3}$ and with constants c_* , C_* , C_0 , E_* . For every $\eta > 0$, there exists $\delta > 0$, depending only on η , K , c_* , C_* , C_0 , and E_* , such that every deterministic nonzero initial state $Z(0) = y$ satisfies*

$$\mathbb{E}[(h(u)^\top Z_t)^2] \leq (1 + \eta) [h(u + t)^\top y]^2, \quad 0 \leq t \leq \delta N^{2/3}, \quad u \geq 0, \quad u + t \leq KN^{2/3}. \quad (10.35)$$

The same bound holds conditionally after every bounded stopping time, as in Corollary 10.7.

Proof. Let C be the constant of Corollary 10.7, and choose

$$\delta = \min \left\{ K, \left[\frac{\eta}{(1 + \eta)C(K + 1)} \right]^2 \right\}.$$

Let $\ell = \delta N^{2/3} \leq \Theta$ in Corollary 10.7. Since $N \geq 1$, we have $1 + KN^{2/3} \leq (1 + K)N^{2/3}$, and therefore

$$\alpha = \frac{C(1 + KN^{2/3})\sqrt{\delta N^{2/3}}}{N} \leq C(1 + K)\sqrt{\delta} \leq \frac{\eta}{1 + \eta}.$$

Then $(1 - \alpha)^{-1} \leq 1 + \eta$, and (10.33) implies (10.35). $\square$

For the rest of this section, Y is the unnormalized process with coefficients r , so that by Remark 10.3 its direction is the diffusion generated by L_ρ with $\rho = r$, and its squares satisfy (10.22) with $\epsilon \equiv 0$. We see from (10.31) that the hypotheses of Corollary 10.8 hold with $\Theta = KN^{2/3}$, $E_* = 0$, and constants c_* , C_* depending only on c_0 , C_0 , and K .

Lemma 10.9 (A martingale with small conditional variance). *Assume (10.30), fix $K < \infty$ and a terminal time $T \leq KN^{2/3}$, and let $M_s = h(T-s)^\top Z(s)$ for $0 \leq s \leq T$, where Y has a deterministic nonzero initial value. Then M is a positive continuous square-integrable martingale with $M_T = \mathcal{A}_T$. For every $\epsilon_0 > 0$ there exists $\delta > 0$, depending only on c_0, C_0, K , and ϵ_0 , such that for every interval $[s_1, s_2] \subset [0, T]$ of length at most $\delta N^{2/3}$ and every stopping time τ with $s_1 \leq \tau \leq s_2$,*

$$\mathbb{E}[(M_{s_2} - M_\tau)^2 \mid \mathcal{F}_\tau] \leq \epsilon_0 M_\tau^2. \quad (10.36)$$

This holds uniformly over all deterministic nonzero initial values.

Proof. The process M is a positive continuous local martingale, by Step 2 of the proof of Lemma 10.5 with $\epsilon \equiv 0$. Its quadratic variation satisfies $[M]_T \leq C \int_0^T \mathcal{A}_s^2 ds$, with $C = (8/N) \max_{v \leq T} m(v)^2$, and the right side is integrable by (10.26). Thus M is a square-integrable martingale. We have $M_T = \mathcal{A}_T$, because $h(0) = \mathbf{1}$. Let δ be given by Corollary 10.8 with $\eta = \epsilon_0$ and $E_* = 0$. For a stopping time $\tau \in [s_1, s_2]$, we apply the conditional form of Corollary 10.8 with $u = T - s_2$ and $t = s_2 - \tau$, which is $\mathcal{F}_\tau$ -measurable and at most $\delta N^{2/3}$; note that $u + t = T - \tau \leq KN^{2/3}$. It gives

$$\mathbb{E}[M_{s_2}^2 \mid \mathcal{F}_\tau] = \mathbb{E}[(h(T - s_2)^\top Z_{\tau+t})^2 \mid \mathcal{F}_\tau] \leq (1 + \epsilon_0) (h(T - \tau)^\top Z_\tau)^2 = (1 + \epsilon_0) M_\tau^2.$$

By optional sampling, $\mathbb{E}[M_{s_2} \mid \mathcal{F}_\tau] = M_\tau$, and (10.36) follows. $\square$

Lemma 10.10 (Two-sided moments in the critical window). *Assume (10.30), and fix $p > 0$ and $K < \infty$. There exists $C_{K,p}$, depending only on c_0, C_0, K , and p , such that the following holds for $0 \leq T \leq KN^{2/3}$.*

1. *For every deterministic initial value $Y(0) = y \neq 0$, the two inequalities*

$$\mathbb{E} \mathcal{A}_T^p \leq C_{K,p} \left(\sum_i h_i(T) y_i^2 \right)^p, \quad \mathbb{E} \mathcal{A}_T^{-p} \leq C_{K,p} \left(\sum_i h_i(T) y_i^2 \right)^{-p} \quad (10.37)$$

hold.

2. *Let $Y(0) = g$ be a standard Gaussian vector independent of B . There exists N_0 , depending only on c_0, C_0, K , and p , such that for $N \geq N_0$,*

$$\mathbb{E} \left[\left(\frac{\mathcal{A}_T}{N \bar{h}(T)} \right)^p \right] + \mathbb{E} \left[\left(\frac{N \bar{h}(T)}{\mathcal{A}_T} \right)^p \right] \leq C_{K,p}, \quad (10.38)$$

and, if in addition $p \geq 1$,

$$\|Y_i(T)^2\|_p \leq C_{K,p} h_i(T), \quad 1 \leq i \leq N. \quad (10.39)$$

In the notation of Section 7, $N \bar{h}(T) = \mathbf{1}^\top h(T)$ is the total mass of the solution of the linear equation started from $\mathbf{1}$; by (10.31), it is of order $N/\sqrt{1+T}$.

Proof. For the first part, let $\epsilon_0 > 0$ satisfy $2^p \cdot 4\epsilon_0 \leq 1/2$, and let δ be given by Lemma 10.9. Partition $[0, T]$ into $J \leq \lceil K/\delta \rceil$ intervals $[s_{j-1}, s_j]$ of length at most $\delta N^{2/3}$. On each interval, the martingale $(M_s)_{s_{j-1} \leq s \leq s_j}$, with respect to $(\mathcal{F}_s)$, satisfies the hypothesis (10.12) of Lemma 10.4 with terminal time s_j and $\varepsilon = \epsilon_0$, by (10.36). Then (10.13) and (10.14) at $\tau = s_{j-1}$ imply that

$$\mathbb{E} \left[\left(\frac{M_{s_{j-1}}}{M_{s_j}} \right)^p \middle| \mathcal{F}_{s_{j-1}} \right] \leq 2^{p+1}, \quad \mathbb{E} \left[\left(\frac{M_{s_j}}{M_{s_{j-1}}} \right)^p \middle| \mathcal{F}_{s_{j-1}} \right] \leq 2^{p+1},$$

since $2^p \epsilon_0 \leq 2^p \cdot 4\epsilon_0 \leq 1/2$. Writing M_T/M_0 as a product of the J ratios and taking iterated conditional expectations, we have $\mathbb{E}[(M_T/M_0)^{\pm p}] \leq 2^{(p+1)J}$. Here $M_0 = h(T)^\top Z(0) = \sum_i h_i(T) y_i^2$ is deterministic and $M_T = \mathcal{A}_T$, which proves (10.37) with $C_{K,p} = 2^{(p+1)\lceil K/\delta \rceil}$.

For the second part, let $S = \sum_i h_i(T) g_i^2$, $h_{\max} = \max_i h_i(T)$, and $D = N\bar{h}(T)/(2h_{\max})$. Conditionally on g , the process Y is the solution started at g , by the first part of Lemma 10.2. Then (10.37) implies that $\mathbb{E}[\mathcal{A}_T^{\pm p} | g] \leq C_{K,p} S^{\pm p}$. It remains to bound $\mathbb{E}S^{\pm p}$ by a constant times $(N\bar{h}(T))^{\pm p}$. For the positive moments, we have $\mathbb{E}S^p \leq C_p (N\bar{h}(T))^p$ by Minkowski's inequality when $p \geq 1$ and by Jensen's inequality when $p < 1$, since $\mathbb{E}S = N\bar{h}(T)$. For the inverse moments, note that the function $x \mapsto \log(1+x)$ is concave and vanishes at zero, which implies that $\log(1+2\lambda h_i(T)) \geq (h_i(T)/h_{\max}) \log(1+2\lambda h_{\max})$ for $\lambda \geq 0$. This gives

$$\mathbb{E}e^{-\lambda S} = \prod_i (1+2\lambda h_i(T))^{-1/2} \leq (1+2\lambda h_{\max})^{-D}.$$

Let $\kappa_p = \int_0^\infty x^{p-1} e^{-x} dx$. Using $s^{-p} = \kappa_p^{-1} \int_0^\infty \lambda^{p-1} e^{-\lambda s} d\lambda$ and substituting $x = 2\lambda h_{\max}$, we find

$$\mathbb{E}S^{-p} \leq \frac{(2h_{\max})^{-p}}{\kappa_p} \int_0^\infty x^{p-1} (1+x)^{-D} dx.$$

If $D \geq 2p+2$, then the last integral is at most $C_p D^{-p}$: on $[0, 1]$ we use $(1+x)^{-D} \leq e^{-Dx/2}$, and on $[1, \infty)$ we use $(1+x)^{-D} \leq 2^{-D/2} x^{-D/2}$. We conclude that $\mathbb{E}S^{-p} \leq C_p (2h_{\max} D)^{-p} = C_p (N\bar{h}(T))^{-p}$. Recall from (10.31) that $N\bar{h}(T) \geq c_1 N/\sqrt{1+T}$ and $h_{\max} \leq C\sqrt{1+T}$, with $c_1 > 0$ and C depending only on c_0, C_0 , and K . It follows that $D \geq (c_1/2C)N/(1+T) \geq c_K N^{1/3}$, which is at least $2p+2$ for $N \geq N_0$. This proves (10.38).

For (10.39), let first p be an integer. Then (10.38) implies that $\|\mathcal{A}_s\|_p \leq C_{K,p} N\bar{h}(s)$ for $0 \leq s \leq T$. Using Lemma 10.6 with $k = p$ and $\|g_i^2\|_p \leq C_p$, we obtain

$$\|Z_i(T)\|_p \leq C_p e^{-4r_i T} + \frac{2(2p-1)C_{K,p}}{N} \int_0^T e^{-4r_i(T-s)} N\bar{h}(s) ds \leq C'_{K,p} h_i(T),$$

where in the last step we compared with (10.19). For noninteger p , note that $\|\cdot\|_p \leq \|\cdot\|_{\lceil p \rceil}$. $\square$

10.5 Above the critical temperature: moments and occupation

In Sections 10.5 to 10.7, we fix $K \geq 1$ and a sequence of precisions $0 < p_1 \leq p_2 \leq \dots \leq p_N$ that satisfies the hypothesis (7.30) of Section 7, that is,

$$0 < d \leq 1, \quad p_1 = d, \quad \frac{1}{K} \left(\frac{i-1}{N} \right)^{2/3} \leq p_i - d \leq K \left(\frac{i-1}{N} \right)^{2/3} \quad (2 \leq i \leq N),$$

$$\sum_{i=1}^N \frac{1}{p_i} = N, \quad Nd^{3/2} \geq 1.$$

We write $q_i = 1/p_i$ and $M = Nd^{3/2}$. For a real symmetric matrix A with eigenvalues $\alpha_1 \geq \dots \geq \alpha_N$ and Lagrange multiplier z , the precisions $p_i = z - \alpha_i$ are nondecreasing in i ; for a general sequence, the increasing rearrangement satisfies (7.30) with the same K (Remark 7.9). For the model with $c_N = 1 - \delta_N$, $\delta_N \rightarrow 0$, and $N\delta_N^3 \rightarrow \infty$, and for every $\eta \in (0, 1)$, there exist $K = K_\eta$ and events $\mathcal{E}_N$ with $\liminf_{N \rightarrow \infty} \mathbb{P}(\mathcal{E}_N) \geq 1 - \eta$ on which the precisions satisfy (7.30); the probability is over the disorder, and K_η is fixed before $N \rightarrow \infty$. This is checked in Section 14. Note that $p_N \leq d + K \leq 1 + K$ and $q_i \in [(1 + K)^{-1}, d^{-1}]$.

Throughout these three sections, Y is the unnormalized process with coefficients $\rho_i = p_i/2$, which solves (10.10). Its direction X is the diffusion generated by L_A when $p_i = z - \alpha_i$ (Remark 10.3). The heat matrix (10.18) is $H = -2 \operatorname{diag}(p) + (2/N) \mathbf{1}\mathbf{1}^\top$, as in (7.3). We use the notation $h, \bar{h}, m, \mathcal{H}_y$ of (10.18). As in Sections 7 and 8, we set

$$a(t) = \sqrt{d} + (1+t)^{-1/2}, \quad \vartheta = \frac{N}{\sum_i q_i^2}, \quad \varphi(x) = \frac{1}{N} \sum_i q_i x_i^2, \quad (10.40)$$

so that $\varphi(x) = x^\top (zI - A)^{-1} x / N$ in eigencoordinates of A . Along the unnormalized process, we write

$$\mathcal{W}_t = \sum_i q_i Z_i(t), \quad V_2(t) = \sum_i q_i^2 Z_i(t), \quad \varphi(X_t) = \frac{\mathcal{W}_t}{\mathcal{A}_t}. \quad (10.41)$$

The last identity holds since $X_i(t)^2 = NZ_i(t)/\mathcal{A}_t$. Since $q_i \leq 1/d$ and $1 = p_i q_i \leq p_N q_i$, we have

$$\mathcal{A}_t \leq p_N \mathcal{W}_t, \quad V_2(t) \leq d^{-1} \mathcal{W}_t. \quad (10.42)$$

We use the following properties of the heat vector. Recall from Lemma 7.1 that H is symmetric and negative semidefinite, its kernel is spanned by q , the function $\bar{h}$ decreases to ϑ , and $\sum_i q_i h_i(t) = N$. The vector $h(t)$ converges to the orthogonal projection of $\mathbf{1}$ onto the kernel, which is $(q^\top \mathbf{1} / |q|^2) q = \vartheta q$, and by self-adjointness, $\mathbf{1}^\top e^{tH} e^{uH} \mathbf{1} = \mathbf{1}^\top e^{(t+u)H} \mathbf{1}$. We obtain

$$\lim_{t \rightarrow \infty} h(t) = \vartheta q, \quad \bar{h}(t) \geq \vartheta, \quad \sum_i h_i(t) h_i(u) = N \bar{h}(t+u), \quad \sum_i q_i h_i(t) = N. \quad (10.43)$$

We see from (7.9) that $h_i(t)$ is a nonincreasing function of p_i ; in particular, $h_1(t) \geq h_2(t) \geq \dots \geq h_N(t)$. By Lemma 7.11, there exists $C_h \geq 1$, depending only on K , such that for all $t \geq 0$,

$$C_h^{-1}a(t) \leq \bar{h}(t) \leq C_h a(t), \quad C_h^{-1}a(t) \frac{1+t}{1+p_i t} \leq h_i(t) \leq C_h a(t) \frac{1+t}{1+p_i t}, \quad m(t) \leq \frac{C_h}{\sqrt{d}}, \quad (10.44)$$

and, by (7.38), for every nonzero $y \in [0, \infty)^N$,

$$m(w-s) \frac{\mathcal{H}_y(s)}{\mathcal{H}_y(w)} \leq C_h \frac{a(s)}{d}, \quad 0 \leq s \leq w. \quad (10.45)$$

The statements below hold for all sufficiently large M , uniformly in N, d , the sequence (p_i) , and, in Sections 10.6 and 10.7, the time parameter $u \geq 0$ of the initial law. Table 10.1 lists the parameters and the order in which they are chosen.

Table 10.1. Parameters in Sections 10.5 to 10.7, in the order in which they are fixed.

Parameter	Role	Depends on
K	constant in (7.30)	fixed first
C_T	time range $[0, T_*]$, $T_* = C_T d^{-1} \log M$	fixed
k, r, γ, λ	moment orders and exponential tilt	fixed
$C, C_r, \dots$	constants in the conclusions	K, C_T , orders
$M_1, M_2, \dots$	lower bounds for $M = Nd^{3/2}$	K, C_T , orders
$N, d, (p_i), u$	precisions, initial law	arbitrary, with M large

Lemma 10.11 (Conditional moments on long time intervals). *Assume (7.30). For every integer $k \geq 2$, let $C_k = 8k(k-1)C_h$ and*

$$\alpha_k(T) = \frac{C_k}{Nd} (\sqrt{T} + \sqrt{d} T), \quad T > 0. \quad (10.46)$$

If $\alpha_k(T) < 1$, then for every deterministic nonzero initial value $Y(0)$, with $y = Z(0)$,

$$\mathbb{E}[(h(u)^\top Z_t)^k] \leq \frac{\mathcal{H}_y(u+t)^k}{1 - \alpha_k(T)}, \quad 0 \leq t \leq T, \quad u \geq 0. \quad (10.47)$$

The same bound holds conditionally on $\mathcal{F}_\tau$, for every $\mathcal{F}_0$ -measurable initial value and every bounded stopping time τ , with y replaced by Z_τ and with $\mathcal{F}_\tau$ -measurable $u \geq 0$ and $t \in [0, T]$. If $\alpha_k(T) \leq 1/2$, then for every $\omega \in [0, \infty)^N$,

$$\mathbb{E}[(\omega^\top Z_t)^k] \leq C'_k (\omega^\top e^{tH} y)^k, \quad 0 \leq t \leq T, \quad C'_k = (2^{1/k}(2k-1))^k. \quad (10.48)$$

Proof. Using (10.45), we find that for every $\Theta \geq T$,

$$\Lambda(T, \Theta) \leq \frac{C_h}{d} \int_0^T a(s) ds \leq \frac{C_h}{d} (\sqrt{d} T + 2\sqrt{T}),$$

since $\int_0^T (1+s)^{-1/2} ds \leq 2\sqrt{T}$. The squares satisfy (10.22) with $\epsilon \equiv 0$, and the quantity (10.24) with $E_* = 0$ and $\ell = T$ is at most $4k(k-1)\Lambda(T, \Theta)/N \leq \alpha_k(T)$, for every $\Theta \geq T$. Let $u \geq 0$ and $t \in [0, T]$ be $\mathcal{F}_\tau$ -measurable. For $n \geq 1$, Lemma 10.5 with $\Theta = n + T$ applies to the pair $(u \wedge n, t)$, and on the $\mathcal{F}_\tau$ -measurable event $\{u \leq n\}$ it yields the conditional form of (10.47). Letting $n \rightarrow \infty$ proves the conditional form, and the case $\tau = 0$ is (10.47).

Now let $\alpha_k(T) \leq 1/2$. Then (10.47) with $u = 0$ implies that $\|\mathcal{A}_s\|_k \leq 2^{1/k} \mathcal{H}_y(s)$ for $s \leq T$. Using Lemma 10.6 with $\rho_i = p_i/2$, and then (10.20), we find

$$\|Z_i(t)\|_k \leq e^{-2p_i t} y_i + \frac{2(2k-1)2^{1/k}}{N} \int_0^t e^{-2p_i(t-s)} \mathcal{H}_y(s) ds \leq 2^{1/k} (2k-1) (e^{tH} y)_i.$$

Since $\|\omega^\top Z_t\|_k \leq \sum_i \omega_i \|Z_i(t)\|_k$ by Minkowski's inequality, (10.48) follows. $\square$

For $T \leq C_T d^{-1} \log M$, we have $\sqrt{T}/(Nd) \leq \sqrt{C_T \log M}/M$ and $\sqrt{d} T/(Nd) \leq C_T \log M/M$, so

$$\alpha_k(T) \leq C_k \frac{\sqrt{C_T \log M} + C_T \log M}{M}, \quad T \leq C_T d^{-1} \log M. \quad (10.49)$$

This tends to zero as $M \rightarrow \infty$, for fixed k and C_T .

The vector q is a stationary solution of the linear equation, and weighting the squares by q produces a martingale.

Lemma 10.12 (The martingale $\mathcal{W}_t = \sum_i q_i Z_i(t)$). *Assume (7.30), and let $Y(0)$ be $\mathcal{F}_0$ -measurable with $\mathbb{E}|Y(0)|^4 < \infty$. Then $\mathcal{W}$ is a strictly positive continuous square-integrable martingale on every bounded interval, with*

$$d\mathcal{W}_t = \sum_i q_i dM_i(t), \quad d[\mathcal{W}]_t = \frac{8}{N} \mathcal{A}_t V_2(t) dt. \quad (10.50)$$

Let $k \geq 2$ be an integer and let $T > 0$ satisfy $\alpha_k(T) < 1$. For every bounded stopping time τ and every $\mathcal{F}_\tau$ -measurable $t \in [0, T]$,

$$\mathbb{E}[\mathcal{W}_{\tau+t}^k | \mathcal{F}_\tau] \leq \frac{\mathcal{W}_\tau^k}{1 - \alpha_k(T)}. \quad (10.51)$$

In particular, if $\alpha_2(T) < 1$, then for all stopping times $\tau \leq D \leq T$,

$$\mathbb{E}[(\mathcal{W}_D - \mathcal{W}_\tau)^2 | \mathcal{F}_\tau] \leq \varepsilon_T \mathcal{W}_\tau^2, \quad \varepsilon_T = \frac{\alpha_2(T)}{1 - \alpha_2(T)}. \quad (10.52)$$

Proof. Using (10.5) and $Hq = 0$, we find that the drift of $\mathcal{W} = q^\top Z$ is $q^\top HZ = (Hq)^\top Z = 0$, and (10.50) holds because the M_i have zero mutual covariation. We have $\mathcal{A}_s V_2(s) \leq d^{-2} \mathcal{A}_s^2$ by (10.42), and the right side is integrable on bounded intervals by (10.26). Thus $\mathcal{W}$ is a square-integrable martingale; it is positive because $\mathcal{W} \geq \mathcal{A}/p_N > 0$. For (10.51), apply the conditional form of (10.47) with lag u and let $u \rightarrow \infty$. By (10.43), we have $h(u)^\top Z_{\tau+t} \rightarrow \vartheta \mathcal{W}_{\tau+t}$ and $\mathcal{H}_{Z_\tau}(u+t) \rightarrow \vartheta \mathcal{W}_\tau$. The conditional Fatou lemma then implies (10.51), after division by ϑ^k .

For (10.52), note that $D \leq T \leq \tau + T$. By optional sampling, $\mathcal{W}_D = \mathbb{E}[\mathcal{W}_{\tau+T} \mid \mathcal{F}_D]$, and by the conditional Jensen inequality and (10.51) with $k = 2$ and $t = T$,

$$\mathbb{E}[\mathcal{W}_D^2 \mid \mathcal{F}_\tau] \leq \mathbb{E}[\mathcal{W}_{\tau+T}^2 \mid \mathcal{F}_\tau] \leq \frac{\mathcal{W}_\tau^2}{1 - \alpha_2(T)}.$$

Since $\mathbb{E}[\mathcal{W}_D \mid \mathcal{F}_\tau] = \mathcal{W}_\tau$, the conditional variance is at most $(1 - \alpha_2(T))^{-1} \mathcal{W}_\tau^2 - \mathcal{W}_\tau^2 = \varepsilon_T \mathcal{W}_\tau^2$. $\square$

Lemma 10.13 (Moments of the maximum and inverse minimum of $\mathcal{W}$). *In the setting of Lemma 10.12, let $p > 0$, and suppose that $\alpha_2(T) < 1$ and $2^p \cdot 4\varepsilon_T \leq 1/2$. Then for all stopping times $\tau \leq D \leq T$,*

$$\mathbb{E} \left[\sup_{\tau \leq s \leq D} \left(\frac{\mathcal{W}_\tau}{\mathcal{W}_s} \right)^p \mid \mathcal{F}_\tau \right] \leq 2^{p+1}. \quad (10.53)$$

In particular, if $\mathcal{W}_0 = N$, then $\mathbb{E}[\sup_{s \leq T} (N/\mathcal{W}_s)^p] \leq 2^{p+1}$. Further, for every integer $k \geq 2$ with $\alpha_k(T) < 1$,

$$\mathbb{E} \left[\sup_{\tau \leq s \leq D} \left(\frac{\mathcal{W}_s}{\mathcal{W}_\tau} \right)^k \mid \mathcal{F}_\tau \right] \leq \left(\frac{k}{k-1} \right)^k \frac{1}{1 - \alpha_k(T)}. \quad (10.54)$$

Proof. We apply Lemma 10.4 on $[0, T]$ to the stopped martingale $\mathcal{W}_s^D = \mathcal{W}_{s \wedge D}$, with $\varepsilon = \varepsilon_T$. To verify (10.12), let $\sigma \leq T$ be a stopping time and $G \in \mathcal{F}_\sigma$. Then $G \cap \{\sigma < D\} \in \mathcal{F}_{\sigma \wedge D}$, and $\mathcal{W}_T^D - \mathcal{W}_\sigma^D$ vanishes outside $\{\sigma < D\}$. By (10.52) at the stopping times $\sigma \wedge D \leq D$, we obtain

$$\mathbb{E}[\mathbf{1}_G (\mathcal{W}_D - \mathcal{W}_{\sigma \wedge D})^2] = \mathbb{E}[\mathbf{1}_{G \cap \{\sigma < D\}} (\mathcal{W}_D - \mathcal{W}_{\sigma \wedge D})^2] \leq \varepsilon_T \mathbb{E}[\mathbf{1}_G \mathcal{W}_{\sigma \wedge D}^2],$$

which is (10.12) for $\mathcal{W}^D$. Then (10.13) implies (10.53), since $2^p / (1 - 2^p \cdot 4\varepsilon_T) \leq 2^{p+1}$. For (10.54), Doob's L^k inequality [123, Theorem II.1.7], applied conditionally on $\mathcal{F}_\tau$ to the martingale $(\mathcal{W}_{(\tau+s) \wedge D})_{s \geq 0}$, yields

$$\mathbb{E} \left[\sup_{\tau \leq s \leq D} \mathcal{W}_s^k \mid \mathcal{F}_\tau \right] \leq \left(\frac{k}{k-1} \right)^k \mathbb{E}[\mathcal{W}_D^k \mid \mathcal{F}_\tau].$$

As in the proof of (10.52), optional sampling, the conditional Jensen inequality, and (10.51) give $\mathbb{E}[\mathcal{W}_D^k \mid \mathcal{F}_\tau] \leq \mathbb{E}[\mathcal{W}_{\tau+T}^k \mid \mathcal{F}_\tau] \leq \mathcal{W}_\tau^k / (1 - \alpha_k(T))$. $\square$

Lemma 10.14 (Inverse moments of the squared norm). *Assume (7.30), let $p > 0$, and suppose that $\alpha_2(T) < 1$ and $2^p \cdot 4\varepsilon_T \leq 1/2$, with ε_T as in (10.52). Fix a deterministic $D \in (0, T]$ and $u \geq 0$, and let $V_s = h(u + D - s)^\top Z_s$ for $0 \leq s \leq D$. For every deterministic nonzero initial value, with $y = Z(0)$, and every stopping time $\tau \leq D$,*

$$\mathbb{E} \left[\sup_{\tau \leq s \leq D} \left(\frac{V_s}{V_\tau} \right)^p \mid \mathcal{F}_\tau \right] \leq 2^{p+1}.$$

In particular,

$$\mathbb{E} \mathcal{A}_D^{-p} \leq 2^{p+1} \mathcal{H}_y(D)^{-p}, \quad 0 < D \leq T. \quad (10.55)$$

If $Y(0)$ is random and independent of B , then $\mathbb{E}[\mathcal{A}_D^{-p} \mid Y(0)] \leq 2^{p+1} \mathcal{H}_{Z(0)}(D)^{-p}$.

Proof. The process V is a positive continuous local martingale by Step 2 of the proof of Lemma 10.5, and it is square integrable, as in the proof of Lemma 10.9. Let $\sigma \leq D$ be a stopping time. By the conditional form of (10.47) with $k = 2$, lag u , and duration $D - \sigma$, which is $\mathcal{F}_\sigma$ -measurable and at most T , we have

$$\mathbb{E}[V_D^2 \mid \mathcal{F}_\sigma] = \mathbb{E}[(h(u)^\top Z_{\sigma+(D-\sigma)})^2 \mid \mathcal{F}_\sigma] \leq \frac{\mathcal{H}_{Z_\sigma}(u + D - \sigma)^2}{1 - \alpha_2(T)} = \frac{V_\sigma^2}{1 - \alpha_2(T)}.$$

Since $\mathbb{E}[V_D \mid \mathcal{F}_\sigma] = V_\sigma$, the hypothesis (10.12) holds on $[0, D]$ with $\varepsilon = \varepsilon_T$, and the first assertion follows from (10.13). For $u = 0$, we have $V_D = h(0)^\top Z_D = \mathcal{A}_D$ and $V_0 = \mathcal{H}_y(D)$, and the first assertion at $\tau = 0$ implies (10.55). The last assertion follows from the first part of Lemma 10.2. $\square$

We now turn to the occupation bound. For $x \in S_N$, we compute the tangential gradient of φ and find

$$\begin{aligned} \Gamma(\log \varphi)(x) &= \frac{|\nabla_{S_N} \varphi(x)|^2}{\varphi(x)^2} = \frac{4}{N} \left(\frac{N \sum_i q_i^2 x_i^2}{(\sum_i q_i x_i^2)^2} - 1 \right), \\ \Gamma(\log \varphi)(X_t) &= \frac{4}{N} \left(\frac{\mathcal{A}_t V_2(t)}{\mathcal{W}_t^2} - 1 \right). \end{aligned} \quad (10.56)$$

Indeed, the tangential gradient is $\nabla_{S_N} \varphi(x) = (2/N)((q_i - \varphi(x))x_i)_i$. Using $\sum_i q_i x_i^2 = N\varphi(x)$ and $|x|^2 = N$, we find $|\nabla_{S_N} \varphi(x)|^2 = (4/N^2)(\sum_i q_i^2 x_i^2 - N\varphi(x)^2)$, which is the first formula after division by $\varphi(x)^2$. The second follows from $X_i(t)^2 = NZ_i(t)/\mathcal{A}_t$.

Theorem 10.15 (Exponential moments of $\int_0^T \Gamma(\log \varphi)(X_s) ds$). *Fix $C_T > 0$ and $\lambda > 0$. There exist M_1 , depending only on K , C_T , and λ , and C , depending only on K and C_T , such that every sequence satisfying (7.30) with $M \geq M_1$ and every $T \leq C_T d^{-1} \log M$ satisfy*

$$\sup_{Y(0) \neq 0} \mathbb{E} \exp \left[\lambda \int_0^T \Gamma(\log \varphi)(X_s) ds \right] \leq \frac{1}{1 - 3\sqrt{2} \lambda \sqrt{\varepsilon_T}} \leq 1 + C\lambda \sqrt{\frac{\log M}{M}}, \quad (10.57)$$

where the supremum is over deterministic initial values. The same bound holds for every initial value $Y(0)$ independent of B , and hence for the diffusion generated by L_A with an arbitrary initial law.

Proof. Using (10.49), we choose M_1 such that $M \geq M_1$ implies $\alpha_2(T) \leq 1/64$, $\varepsilon_T \leq 1/32$, and $3\sqrt{2}\lambda\sqrt{\varepsilon_T} \leq 1/2$. Let $Y(0) \neq 0$ be deterministic. The quadratic variation of $\log \mathcal{W}$ is, by (10.50),

$$Q_t = \int_0^t \frac{d[\mathcal{W}]_s}{\mathcal{W}_s^2} = \int_0^t \frac{8}{N} \frac{\mathcal{A}_s V_2(s)}{\mathcal{W}_s^2} ds,$$

and by (10.42) its density is at most $8p_N/(Nd)$; in particular, $Q_T \leq 8p_N T/(Nd)$ deterministically. The process $\mathcal{W}$ satisfies (10.12) on $[0, T]$ with $\varepsilon = \varepsilon_T$, by Lemma 10.12. Then (10.15) and (10.16) apply with $K = 6\sqrt{2\varepsilon_T}$, and we obtain

$$\mathbb{E}[Q_T - Q_\tau \mid \mathcal{F}_\tau] \leq 6\sqrt{2\varepsilon_T}, \quad \mathbb{E}e^{\lambda Q_T/2} \leq \frac{1}{1 - \lambda K/2} = \frac{1}{1 - 3\sqrt{2}\lambda\sqrt{\varepsilon_T}}. \quad (10.58)$$

Using (10.56), we obtain the exact identity

$$\int_0^T \Gamma(\log \varphi)(X_s) ds = \frac{Q_T}{2} - \frac{4T}{N},$$

and dropping the negative term proves the first inequality in (10.57). For the second, we use $(1-x)^{-1} \leq 1+2x$ for $x \leq 1/2$ and $\sqrt{\varepsilon_T} \leq \sqrt{2\alpha_2(T)} \leq C\sqrt{\log M/M}$, by (10.49). The bound does not depend on the deterministic initial value, and the first part of Lemma 10.2 extends it to every initial value independent of B . Every initial law of X arises in this way, by the second part of that lemma; for instance, one may take $Y(0) = X(0)$. $\square$

10.6 Initial laws equal to Bingham comparison laws

We keep the setting of Section 10.5. For $u \geq 0$, let $v_i(u) = h_i(u)/\bar{h}(u)$, so that $\sum_i v_i(u) = N$, and let $\mu_u = \mu_{v(u)}$ be the Gaussian law with variances $v(u)$ conditioned on S_N (Definition 5.4). By the second part of Proposition 5.6, μ_u is the Bingham comparison law at time u of Section 7, and $\mu_0 = \nu_N$. We consider the unnormalized process with the initial value

$$Y(0) = \frac{X(0)}{\sqrt{\varphi(X(0))}}, \quad X(0) \sim \mu_u \text{ independent of } B. \quad (10.59)$$

Then $\mathcal{W}_0 = \sum_i q_i X_i(0)^2 / \varphi(X(0)) = N$, and $\sqrt{N}Y(0)/|Y(0)| = X(0)$ since $|X(0)|^2 = N$. We see from the second part of Lemma 10.2 that X is the diffusion generated by L_A with initial law μ_u . We write

$$\mathcal{H}(t) = \mathcal{H}_{Z(0)}(t) = \sum_i h_i(t) Y_i(0)^2. \quad (10.60)$$

The martingale with backward semigroup weights $s \mapsto h(t-s)^T Z_s$, $0 \leq s \leq t$, is a martingale (see the proof of Lemma 10.14) with terminal value $\mathcal{A}_t$, and we have $\mathbb{E}[\mathcal{A}_t \mid Y(0)] = \mathcal{H}(t)$; that is, $\mathcal{H}(t)$ is the conditional mean of the squared norm given the initial state. The following lemma compares $\mathcal{H}(t)$ with its deterministic counterpart $N\bar{h}(u+t)$ simultaneously for all $t \geq 0$, with one random constant.

Lemma 10.16 (Uniform-in-time comparison of the initial quadratic forms). *For every $r > 0$ there exist M_2 and C_r , depending only on r and K , such that the following holds for every sequence satisfying (7.30) with $M \geq M_2$ and every $u \geq 0$. With the initial value (10.59), there exists a random variable $C_{\text{init}} \geq 1$, measurable with respect to $X(0)$, with $\mathbb{E}C_{\text{init}}^r \leq C_r$ and such that simultaneously for all $t \geq 0$,*

$$C_{\text{init}}^{-1} \leq \frac{\mathcal{H}(t)}{N\bar{h}(u+t)} \leq C_{\text{init}}. \quad (10.61)$$

The point of the lemma is that one random constant serves for all future times. The proof uses one comparison of the initial law with independent Gaussian variables on each of two sets of coordinates, together with Abel summation; it involves no union bound over times.

Proof. The proof has four steps.

Step 1: the weights carry no macroscopic fraction. Let $\zeta_i = X_i(0)/\sqrt{v_i(u)}$ be the standardized coordinates of $X(0)$. For $t \geq 0$, define the weight sequences

$$w_i(t) = h_i(t) v_i(u), \quad w_i^q = q_i v_i(u), \quad 1 \leq i \leq N.$$

Each of them is a positive nonincreasing function of i , since $h_i(t)$, $v_i(u)$, and q_i are. Using (10.43), we compute

$$\sum_i w_i(t) = \frac{N\bar{h}(u+t)}{\bar{h}(u)}, \quad \sum_i w_i^q = \frac{N}{\bar{h}(u)}, \quad \sum_i v_i(u)^2 = \frac{N\bar{h}(2u)}{\bar{h}(u)^2}. \quad (10.62)$$

By (10.44) with $p_1 = d$, we have $h_1(t) \leq C_h a(t)(1+t)/(1+dt) \leq 4C_h a(t)^{-1}$, where we used that $a(t)^2(1+t) \leq 2(d(1+t)+1) \leq 4(1+dt)$ for $d \leq 1$. Using $a \geq \sqrt{d}$ and $\bar{h} \geq C_h^{-1}a$, we find

$$\frac{w_1(t)}{\sum_i w_i(t)} = \frac{h_1(t)h_1(u)}{N\bar{h}(u+t)} \leq \frac{C}{Na(t)a(u)a(u+t)} \leq \frac{C}{M}, \quad \frac{w_1^q}{\sum_i w_i^q} = \frac{h_1(u)}{Nd} \leq \frac{C}{Nda(u)} \leq \frac{C}{M},$$

and similarly $v_1(u)^2/\sum_i v_i(u)^2 = h_1(u)^2/(N\bar{h}(2u)) \leq C/M$. Fix $C_* \geq 1$, depending only on K , such that each of these three fractions is at most C_*/M , for all $t, u \geq 0$. In the notation (5.1) of Section 5, the last bound gives $\eta_{v(u)}^2 \leq C_*/(2M)$.

Step 2: two blocks of coordinates. Let I_1 and I_2 be the sets of odd and of even indices. For a positive nonincreasing weight sequence x , the alternating sum $x_1 - x_2 + x_3 - \dots$ lies in $[0, x_1]$, so $\sum_{I_1} x$ and $\sum_{I_2} x$ both lie in $[\frac{1}{2}(\sum_i x_i - x_1), \frac{1}{2}(\sum_i x_i + x_1)]$. Applied to the three weight sequences of Step 1, this shows that for $M \geq 3C_*$, each block carries between $1/3$ and $2/3$ of the sum of each of them, and that the maximal entry of $w(t)$ or w^q on a block is at most $3C_*/M$ times the sum of that weight sequence over the block. In particular, the fraction $h_{I_b} = \sum_{i \in I_b} v_i(u)^2 / \sum_i v_i(u)^2$ of the variance of the radial constraint carried by the block I_b is at most $2/3$, for $b = 1, 2$. By the second part of Theorem 5.11 with $\epsilon = 1/3$, for M large enough that $\eta_{v(u)} \leq \eta(1/3)$, the law of $(\zeta_i)_{i \in I_b}$ under μ_u has a density with respect to the standard Gaussian measure on $\mathbb{R}^{I_b}$ that is bounded by a universal constant $C_{1/3}$.

Step 3: prefix sums. Fix $b \in \{1, 2\}$, enumerate I_b in increasing order as $j_1 < j_2 < \dots < j_n$, and let $\Sigma_k = \sum_{l \leq k} \zeta_{j_l}^2$. Let $m_0 = \lfloor M/(24C_*) \rfloor$; since $M \leq N$, we have $m_0 \leq n$. We define

$$\underline{s}_b = \min_{m_0 \leq k \leq n} \frac{\Sigma_k}{k}, \quad \bar{s}_b = \max_{m_0 \leq k \leq n} \frac{\Sigma_k}{k}.$$

Under independent standard Gaussian variables, Σ_k has the chi-squared distribution with k degrees of freedom, and Chernoff's bound yields $\mathbb{P}(\Sigma_k \leq kx) \leq (xe^{1-x})^{k/2} \leq (ex)^{k/2}$ for $x \leq 1$, and $\mathbb{P}(\Sigma_k \geq kx) \leq e^{-kx/4} 2^{k/2} \leq e^{-kx/8}$ for $x \geq 8$. Summing over $k \geq m_0$, we obtain

$$\mathbb{P}(\underline{s}_b < x) \leq 2(ex)^{m_0/2} \quad \left(0 < x \leq \frac{1}{4e}\right), \quad \mathbb{P}(\bar{s}_b > x) \leq 2e^{-m_0x/8} \quad (x \geq 8).$$

Integrating these tails, we find that for every $r > 0$ there exists C_r such that $\mathbb{E}\underline{s}_b^{-r} + \mathbb{E}\bar{s}_b^r \leq C_r$ whenever $m_0 \geq 2r + 2$. Using the density bound of Step 2, we find that the same bounds hold under μ_u , with C_r replaced by $C_{1/3}C_r$.

Let x be one of the weight sequences $w(t)$ or w^q , restricted to I_b and written as $x_{j_1} \geq \dots \geq x_{j_n}$, and set $x_{j_{n+1}} = 0$. Abel summation gives

$$\sum_{l=1}^n x_{j_l} \zeta_{j_l}^2 = \sum_{k=1}^n (x_{j_k} - x_{j_{k+1}}) \Sigma_k.$$

The coefficients $x_{j_k} - x_{j_{k+1}}$ are nonnegative. We also have $\sum_k (x_{j_k} - x_{j_{k+1}})k = \sum_l x_{j_l}$ and $\sum_{k < m_0} (x_{j_k} - x_{j_{k+1}})k \leq m_0 \max_l x_{j_l}$. For the lower bound, we use $\Sigma_k \geq 0$ for $k < m_0$ and $\Sigma_k \geq k\underline{s}_b$ for $k \geq m_0$, which gives at least $\underline{s}_b(\sum_l x_{j_l} - m_0 \max_l x_{j_l})$. For the upper bound, we use $\Sigma_k \leq \Sigma_{m_0} \leq m_0 \bar{s}_b$ for $k < m_0$ and $\Sigma_k \leq k\bar{s}_b$ for $k \geq m_0$, which gives at most $\bar{s}_b(\sum_l x_{j_l} + m_0 \max_l x_{j_l})$. Since $m_0 \max_l x_{j_l} \leq \frac{M}{24C_*} \cdot \frac{3C_*}{M} \sum_l x_{j_l} = \frac{1}{8} \sum_l x_{j_l}$ by Step 2, we conclude that

$$\frac{7}{8} \underline{s}_b \sum_{i \in I_b} x_i \leq \sum_{i \in I_b} x_i \zeta_i^2 \leq \frac{9}{8} \bar{s}_b \sum_{i \in I_b} x_i. \quad (10.63)$$

Step 4: conclusion. Let $\mathcal{S} = 1 + \underline{s}_1^{-1} + \bar{s}_1 + \bar{s}_2$. Using the lower bound in (10.63) on I_1 , the upper bounds on both blocks, and the fact that I_1 carries at least a third of each weight sequence, we find that simultaneously for every weight sequence x among $w(t)$, $t \geq 0$, and w^q ,

$$\frac{7}{24} \mathcal{S}^{-1} \sum_i x_i \leq \sum_i x_i \zeta_i^2 \leq \frac{9}{8} \mathcal{S} \sum_i x_i.$$

No independence between the two blocks is needed. Since $X_i(0)^2 = v_i(u) \zeta_i^2$ and $\varphi(X(0)) = N^{-1} \sum_i w_i^q \zeta_i^2$, we have exactly

$$\mathcal{H}(t) = \frac{\sum_i h_i(t) X_i(0)^2}{\varphi(X(0))} = N \frac{\sum_i w_i(t) \zeta_i^2}{\sum_i w_i^q \zeta_i^2},$$

and by (10.62), $\sum_i w_i(t) / \sum_i w_i^q = \bar{h}(u+t)$. This implies (10.61) with $C_{\text{init}} = \frac{27}{7} \mathcal{S}^2$. By Step 3 and the elementary inequality $(\sum_{l=1}^4 x_l)^{2r} \leq 4^{2r} \sum_l x_l^{2r}$, we have $\mathbb{E} C_{\text{init}}^r \leq C_r$ for $M \geq M_2(r)$. All constants depend only on r and K , and not on u , N , or d . $\square$

Corollary 10.17 (Fixed-time moments for Bingham comparison initial laws). *For every $r > 0$ and $C_T > 0$, there exist M_3 and C_r , depending only on r , C_T , and K , such that the following holds for every sequence satisfying (7.30) with $M \geq M_3$, every $u \geq 0$, and every $0 \leq t \leq C_T d^{-1} \log M$. With the initial value (10.59),*

$$\mathbb{E} \left(\frac{\mathcal{A}_t}{N \bar{h}(u+t)} \right)^r + \mathbb{E} \left(\frac{N \bar{h}(u+t)}{\mathcal{A}_t} \right)^r \leq C_r. \quad (10.64)$$

For the diffusion X with initial law μ_u ,

$$\mathbb{E} [(\varphi(X_t) \bar{h}(u+t))^r] + \mathbb{E} [(\varphi(X_t) \bar{h}(u+t))^{-r}] \leq C_r, \quad \mathbb{E} [(\vartheta \varphi(X_t))^r] \leq C_r. \quad (10.65)$$

These are bounds at each fixed time, uniform in the time; a bound for the expectation of the supremum over t is proved in Theorem 10.18.

Proof. Let $T = C_T d^{-1} \log M$. By Hölder's inequality, it suffices to treat integer $r \geq 2$. Using (10.49), we choose $M_3 \geq M_2(2r)$ so that $\alpha_{2r}(T) \leq 1/2$ and $2^{2r} \cdot 4\epsilon_T \leq 1/2$; note that $\alpha_k(T)$ increases with k . We condition on $Y(0)$, which is independent of B . By the conditional form of (10.47) at $\tau = 0$ with $u = 0$, and by the last assertion of Lemma 10.14, we have

$$\mathbb{E} [\mathcal{A}_t^r | Y(0)] \leq 2 \mathcal{H}(t)^r, \quad \mathbb{E} [\mathcal{A}_t^{-r} | Y(0)] \leq 2^{r+1} \mathcal{H}(t)^{-r}.$$

By (10.61), $\mathcal{H}(t)^{\pm r} \leq C_{\text{init}}^r (N \bar{h}(u+t))^{\pm r}$, and taking expectations, we obtain (10.64). Next, we have $\varphi(X_t) \bar{h}(u+t) = (\mathcal{W}_t/N) \cdot N \bar{h}(u+t) / \mathcal{A}_t$ by (10.41). Since $\mathcal{W}_0 = N$, Lemma 10.13 and (10.51) bound $\mathbb{E} (\mathcal{W}_t/N)^{\pm 2r}$, and the first bound in (10.65) follows from the Cauchy-Schwarz inequality and (10.64) at order $2r$. Finally, $\bar{h}(u+t) \geq \vartheta$ by (10.43). Thus $\vartheta \varphi(X_t) \leq \varphi(X_t) \bar{h}(u+t)$ pointwise, and the last bound follows. $\square$

10.7 The maximal inverse of the squared norm

We keep the setting of Section 10.6: the precisions satisfy (7.30), $u \geq 0$, and the initial value is (10.59). Fix $C_T > 0$ and let $T_* = C_T d^{-1} \log M$. In this section, constants depend only on K , C_T , and the indicated moment orders, and the lower bounds for M depend on the same quantities.

Theorem 10.18 (Maximal inverse moments of the squared norm). *For every $r > 0$, there exist M_4 and C_r , depending only on r , C_T , and K , such that for every sequence satisfying (7.30) with $M \geq M_4$ and every $u \geq 0$,*

$$\mathbb{E} \left[\sup_{0 \leq t \leq T_*} \left(\frac{\mathcal{H}(t)}{\mathcal{A}_t} \right)^r \right] \leq C_r, \quad \mathbb{E} \left[\sup_{0 \leq t \leq T_*} \left(\frac{N\bar{h}(u+t)}{\mathcal{A}_t} \right)^r \right] \leq C_r. \quad (10.66)$$

In particular, for the diffusion X with initial law μ_u ,

$$\mathbb{E} \left[\sup_{0 \leq t \leq T_*} (\vartheta \varphi(X_t))^r \right] \leq C_r. \quad (10.67)$$

The proof occupies the rest of this section. Its outline is as follows. We write $\mathcal{A}_t = \mathcal{H}(t) + S_t$, where S_t is a stochastic convolution, and bound the moments of the increments of $J_t = S_t/\mathcal{H}(t)$ on the blocks $[0, 1]$ and $[L, 2L]$, $1 \leq L \leq T_*$ (Lemma 10.19). A chaining argument turns these into bounds for the Hölder norm of J on each block. Combined with the fixed-time inverse moments, this yields small-ball bounds for $\inf \mathcal{A}_t/\mathcal{H}(t)$ on each block (Lemma 10.20), with a factor that is summable over dyadic blocks. A union bound over the blocks then proves Theorem 10.18. Both the Hölder bound and the fixed-time inverse moments are needed.

Inputs. We collect the estimates of Sections 10.5 and 10.6 that are used below, on the doubled interval $[0, 2T_*]$; doubling the interval only replaces C_T by $2C_T$. For every $r > 0$, there exist C_r and a lower bound for M such that, for all $t \geq 0$ in the first two lines and all $s, t \in [0, 2T_*]$ in the last two,

$$\bar{h}(t) \asymp a(t), \quad h_i(t) \asymp a(t) \frac{1+t}{1+p_i t}, \quad (10.68)$$

$$C_{\text{init}}^{-1} \leq \frac{\mathcal{H}(t)}{N\bar{h}(u+t)} \leq C_{\text{init}}, \quad (10.69)$$

$$\mathbb{E} C_{\text{init}}^r \leq C_r, \quad \|\mathcal{A}_s\|_r \leq C_r N\bar{h}(u+s), \quad \left\| \sup_{s \leq 2T_*} \frac{\mathcal{W}_s}{N} \right\|_r \leq C_r, \quad (10.70)$$

$$\mathbb{E} \left[\left(\frac{\mathcal{H}(t)}{\mathcal{A}_t} \right)^r \right] \leq C_r. \quad (10.71)$$

Here (10.68) is (10.44), (10.69) and the first bound of (10.70) are Lemma 10.16, the second bound of (10.70) is (10.64), and the third follows from (10.54) with $\tau = 0$ and $\mathcal{W}_0 = N$.

For (10.71), we condition on $Y(0)$ and apply (10.55) at the deterministic initial state $Y(0)$; no Gaussian approximation of $\mathcal{A}_t$ is involved. We also use the following bounds on the derivatives of h , which hold for all $t, z \geq 0$ by Lemma 7.13:

$$\begin{aligned} \max_i p_i h_i(t)^2 &\leq C(1 + dt), & \max_i \sqrt{p_i} |h'_i(t)| &\leq \frac{C}{1+t}, \\ \max_i p_i [h_i(t+z) - h_i(t)]^2 &\leq C \log^2 \left(1 + \frac{z}{1+t}\right), \end{aligned} \quad (10.72)$$

and $|\mathcal{H}'(t)| \leq C\mathcal{H}(t)/(1+t)$, which follows by summing $|h'_i(t)| \leq Ch_i(t)/(1+t)$ against $Y_i(0)^2 \geq 0$. Finally, we use two elementary facts about a : for all $u, t \geq 0$,

$$\int_0^t a(u+v) dv \leq 2t a(u+t), \quad a(u+v) \leq \sqrt{2} a(u+t) \quad \text{for } t/2 \leq v \leq t. \quad (10.73)$$

The first follows from $\int_0^t (1+u+v)^{-1/2} dv = 2t/(\sqrt{1+u+t} + \sqrt{1+u}) \leq 2t/\sqrt{1+u+t}$, and the second from $1+u+t \leq 2(1+u+v)$.

The stochastic convolution. For fixed t , the martingale with backward semigroup weights $s \mapsto h(t-s)^\top Z_s$, $0 \leq s \leq t$, has initial value $\mathcal{H}(t)$ and terminal value $\mathcal{A}_t$. It follows that

$$\mathcal{A}_t = \mathcal{H}(t) + S_t, \quad S_t = \sum_i \int_0^t h_i(t-s) dM_i(s), \quad J_t = \frac{S_t}{\mathcal{H}(t)}, \quad (10.74)$$

so that $\mathcal{A}_t/\mathcal{H}(t) = 1 + J_t$. The random variable $\mathcal{H}(t)$ is $\mathcal{F}_0$ -measurable.

Lemma 10.19 (Increments of the normalized stochastic convolution). *For every $r \geq 2$ there exist C_r and a lower bound for M such that the following holds. Let $1 \leq L \leq T_*$, and let $L \leq s \leq t \leq 2L$. Then*

$$\|J_L\|_r \leq C_r \sigma_L, \quad \|J_t - J_s\|_r \leq C_r \sigma_L \sqrt{\frac{t-s}{L}}, \quad \sigma_L^2 = \frac{L(1+dL)}{Na(u+L)} \leq \frac{dL + (dL)^2}{M}. \quad (10.75)$$

On the initial block, for $0 \leq s \leq t \leq 1$, we have $J_0 = 0$ and $\|J_t - J_s\|_r \leq C_r \sigma_* \sqrt{t-s}$, with $\sigma_*^2 = [Na(u+1)]^{-1} \leq d/M$. After the linear change of time that maps the block to $[0, 1]$, let $\mathcal{K}$ be the sum of the supremum of $|J|$ and the $1/4$ -Hölder seminorm of J on $[0, 1]$. For every $r > 4$,

$$\|\mathcal{K}\|_r \leq C_r \sigma_L \quad \text{on } [L, 2L], \quad \|\mathcal{K}\|_r \leq C_r \sigma_* \quad \text{on } [0, 1]. \quad (10.76)$$

The restriction $L \leq T_*$ is necessary: the moment inputs above hold only on $[0, 2T_*]$, and the bounds (10.75) fail for large L at fixed N and d . Indeed, we have $\mathcal{A}_t V_2(t) \geq \mathcal{W}_t^2$ by the Cauchy–Schwarz inequality, and (10.50) implies that $\mathbb{E} \mathcal{W}_t^2 \geq N^2 e^{8t/N}$; thus $\|J_t\|_2$ grows exponentially in t/N . Moreover, since $\log \mathcal{W}$ has quadratic variation at least $8t/N$, the martingale $\mathcal{W}_t$ tends to zero almost surely, so that $\mathcal{A}_t/\mathcal{H}(t) \rightarrow 0$.

Proof. We begin with a general bound. Let $\omega_i(v)$ be deterministic functions, and let $\mathcal{N}_t = \sum_i \int_0^t \omega_i(v) dM_i(v)$. Its quadratic variation is $\frac{8}{N} \int_0^t \mathcal{A}_v \sum_i \omega_i(v)^2 Z_i(v) dv$, and $\sum_i \omega_i^2 Z_i = \sum_i (p_i \omega_i^2) q_i Z_i \leq (\max_i p_i \omega_i^2) \mathcal{W}$. By the Burkholder–Davis–Gundy inequality [123, Theorem IV.4.1] and Minkowski’s inequality in $L^{r/2}$, for $r \geq 2$,

$$\|\mathcal{N}_t\|_r^2 \leq \frac{C_r}{N} \int_0^t \max_i p_i \omega_i(v)^2 \|\mathcal{A}_v \mathcal{W}_v\|_{r/2} dv \leq C_r N \int_0^t \max_i p_i \omega_i(v)^2 a(u+v) dv, \quad (10.77)$$

where in the last step we used $\|\mathcal{A}_v \mathcal{W}_v\|_{r/2} \leq \|\mathcal{A}_v\|_r \|\mathcal{W}_v\|_r \leq C_r N^2 a(u+v)$, by (10.68) and (10.70).

Pointwise bound. With $\omega_i(v) = h_i(t-v)$, the first bound in (10.72) and (10.73) give

$$\|S_t\|_r^2 \leq C_r N \int_0^t a(u+v)(1+d(t-v)) dv \leq C_r N a(u+t) t(1+dt), \quad 0 \leq t \leq 2T_*. \quad (10.78)$$

Increments on $[L, 2L]$. Let $L \leq s \leq t \leq 2L$ and $z = t-s$, so that $z \leq L \leq s$. We split $S_t - S_s$ into an old part, the integral over $[0, s]$ with $\omega_i(v) = h_i(t-v) - h_i(s-v)$, and a fresh part, the integral over $[s, t]$ with $\omega_i(v) = h_i(t-v)$. Using (10.77) and (10.72), we find that the squared L^r norm of $S_t - S_s$ is at most a constant times

$$N \int_0^s a(u+v) \log^2 \left(1 + \frac{z}{1+s-v}\right) dv + N \int_s^t a(u+v)(1+d(t-v)) dv.$$

We bound the three integrals separately. For $v \leq s/2$, the logarithm is at most $z/(1+s-v) \leq 2z/(1+s)$, and by (10.73), $\int_0^{s/2} a(u+v) dv \leq s a(u+s/2) \leq \sqrt{2} s a(u+s)$; hence the integral over $[0, s/2]$ is at most $Ca(u+s)z^2/(1+s) \leq Ca(u+s)z$, where we used $z \leq s$. For $s/2 \leq v \leq s$, we use $a(u+v) \leq \sqrt{2}a(u+s)$ and

$$\int_0^\infty \log^2 \left(1 + \frac{z}{1+w}\right) dw \leq 4z. \quad (10.79)$$

To prove (10.79), write the logarithm as $\int_0^z (1+w+x)^{-1} dx$; Minkowski’s inequality in $L^2(dw)$ bounds its norm by $\int_0^z (1+x)^{-1/2} dx \leq 2\sqrt{z}$. For the fresh part, $a(u+v) \leq a(u+s)$ and $\int_s^t (1+d(t-v)) dv \leq z + dz^2$. Altogether,

$$\|S_t - S_s\|_r^2 \leq C_r N a(u+s) (z + dz^2). \quad (10.80)$$

Next, we divide by $\mathcal{H}$. We write

$$J_t - J_s = \frac{S_t - S_s}{\mathcal{H}(t)} + J_s \left(\frac{\mathcal{H}(s)}{\mathcal{H}(t)} - 1 \right).$$

Using Hölder's inequality and (10.69), we find $\|(S_t - S_s)/\mathcal{H}(t)\|_r \leq \|C_{\text{init}}\|_{2r} \|S_t - S_s\|_{2r} / (N\bar{h}(u+t))$. On the block, $a(u+L)/\sqrt{2} \leq a(u+t) \leq a(u+s) \leq a(u+L)$, and $z + dz^2 \leq z(1+dL)$. Together with (10.80) and (10.68), this gives

$$\left\| \frac{S_t - S_s}{\mathcal{H}(t)} \right\|_r \leq C_r \left(\frac{z(1+dL)}{Na(u+L)} \right)^{1/2} = C_r \sigma_L \sqrt{\frac{z}{L}}.$$

In the same way, we find from (10.78) that $\|J_s\|_r \leq C_r (s(1+ds)/(Na(u+s)))^{1/2} \leq C_r \sigma_L$ for $s \in [L, 2L]$, which contains the bound on $\|J_L\|_r$. The bound $|\mathcal{H}'| \leq C\mathcal{H}/(1+t)$ gives $|\log \mathcal{H}(t) - \log \mathcal{H}(s)| \leq Cz/L \leq C$, and hence $|\mathcal{H}(s)/\mathcal{H}(t) - 1| \leq C'z/L \leq C'\sqrt{z/L}$. This proves (10.75). The bound on σ_L^2 follows from $a(u+L) \geq \sqrt{d}$ and $N\sqrt{d} = M/d$.

The initial block. For $0 \leq s \leq t \leq 1$ and $z = t - s$, the second bound in (10.72) implies that $p_i[h_i(t-v) - h_i(s-v)]^2 \leq Cz^2$, and the first gives $p_i h_i(t-v)^2 \leq C$. For $v \leq 1$, we have $a(u+v) \leq a(u) \leq \sqrt{2}a(u+1)$, and (10.77) yields $\|S_t - S_s\|_r^2 \leq C_r Na(u+1)(z^2 + z)$ and $\|S_t\|_r^2 \leq C_r Na(u+1)$. Dividing by $\mathcal{H}(t) \geq C_{\text{init}}^{-1} N\bar{h}(u+t) \geq c_1 C_{\text{init}}^{-1} Na(u+1)$, with $c_1 > 0$ depending only on K , as before, and using $|\mathcal{H}'| \leq C\mathcal{H}$, we have $\|J_t - J_s\|_r \leq C_r \sigma_* \sqrt{z}$. Also $J_0 = 0$, since $S_0 = 0$, and $\sigma_*^2 \leq (N\sqrt{d})^{-1} = d/M$.

Hölder norms. We use the dyadic proof of Kolmogorov's continuity criterion, keeping its moment bound; see [123, Theorem I.2.1 and its proof, pp. 26–28]. Let $\tilde{J}$ be J on one of the blocks after the linear change of time to $[0, 1]$, and let σ be the corresponding σ_L or σ_* . The bounds above give $\|\tilde{J}(\theta') - \tilde{J}(\theta)\|_r \leq C_r \sigma |\theta' - \theta|^{1/2}$ and $\|\tilde{J}(0)\|_r \leq C_r \sigma$. For $n \geq 0$, let $\Delta_n = \max_{0 \leq k < 2^n} |\tilde{J}((k+1)2^{-n}) - \tilde{J}(k2^{-n})|$. Bounding the maximum by the ℓ^r sum of its 2^n terms, we find $\|\Delta_n\|_r \leq C_r \sigma 2^{n(1/r-1/2)}$. For $r > 4$, it follows that

$$\left\| \sum_{n \geq 0} 2^{n/4} \Delta_n \right\|_r \leq C_r \sigma \sum_{n \geq 0} 2^{n(1/r-1/4)} \leq C_r \sigma.$$

Let $\theta < \theta'$ be dyadic points in $[0, 1]$ with $2^{-n-1} < \theta' - \theta \leq 2^{-n}$. Approximating θ and θ' by their dyadic expansions, as in the proof of Kolmogorov's continuity criterion, we have $|\tilde{J}(\theta') - \tilde{J}(\theta)| \leq 2 \sum_{k \geq n} \Delta_k \leq 2^{5/4} (\theta' - \theta)^{1/4} \sum_{k \geq 0} 2^{k/4} \Delta_k$. By continuity of $\tilde{J}$, this bound extends to all $\theta, \theta' \in [0, 1]$, and the $1/4$ -Hölder seminorm of $\tilde{J}$ is at most $2^{5/4} \sum_k 2^{k/4} \Delta_k$. The supremum of $|\tilde{J}|$ is at most $|\tilde{J}(0)|$ plus this seminorm, which proves (10.76). $\square$

Lemma 10.20 (Small values of the normalized squared norm on a block). *Let $1 \leq L \leq T_*$, let $\iota_L = \inf_{L \leq t \leq 2L} \mathcal{A}_t / \mathcal{H}(t)$, and let $\iota_* = \inf_{0 \leq t \leq 1} \mathcal{A}_t / \mathcal{H}(t)$. For every $\gamma > 0$ and $r > 4$, there exist constants $C_\gamma, C_{\gamma,r}$ and a lower bound for M such that for $0 < x \leq 1/2$,*

$$\mathbb{P}(\iota_L \leq x) \leq C_\gamma x^\gamma, \quad \mathbb{P}(\iota_L \leq x) \leq C_{\gamma,r} x^{\gamma/2} \sigma_L^{r/2}, \quad (10.81)$$

and the same bounds hold for ι_* with σ_* in place of σ_L .

Proof. Let $b_t = \mathcal{A}_t / \mathcal{H}(t) = 1 + J_t$ on the block, after the change of time to $[0, 1]$. For $L \leq T_*$, we have $\sigma_L^2 \leq (C_T \log M + C_T^2 \log^2 M) / M$ by (10.75), and for M large every σ_L and σ_* is at most one. By (10.76), every moment of order greater than 4 of the $1/4$ -Hölder seminorm of b , which equals that of J , is bounded by a constant. We choose deterministic grid points in $[0, 1]$ with spacing at most x^8 ; there are at most $2x^{-8}$ of them. If the seminorm is at most x^{-1} , every point of $[0, 1]$ is within $x^{-1}(x^8)^{1/4} = x$ of the value at a grid point. The event $\{\iota_L \leq x\}$ is contained in the union of the event that the seminorm exceeds x^{-1} and the events that $b \leq 2x$ at one of the grid points. By Markov's inequality, the first event has probability at most $C_{\gamma'} x^{\gamma'}$ for every $\gamma' > 4$. We take $\gamma' = \max(\gamma, 5)$. By (10.71) with order $\gamma + 9$, each grid event has probability at most $C(2x)^{\gamma+9}$, and the union over the grid contributes at most $2x^{-8} C(2x)^{\gamma+9} \leq C' x^\gamma$. This proves the first bound. For the second, we have $\{\iota_L \leq 1/2\} \subset \{\sup |J| \geq 1/2\}$, since $b = 1 + J$, and (10.76) and Markov's inequality imply that $\mathbb{P}(\iota_L \leq 1/2) \leq 2^r \mathbb{E} K^r \leq C_r \sigma_L^r$. For $x \leq 1/2$, the probability $\mathbb{P}(\iota_L \leq x)$ is at most the minimum of $C_\gamma x^\gamma$ and $C_r \sigma_L^r$, which is at most their geometric mean. The initial block is treated in the same way, with $b_0 = 1$. $\square$

Proof of Theorem 10.18. We take M large enough that $T_* \geq 1$ and that all the preceding lemmas apply with the finitely many moment orders used below. Cover $[0, T_*]$ by $[0, 1]$ and the dyadic blocks $[2^j, 2^{j+1}]$, $0 \leq j \leq j_*$, where $2^{j_*} \leq T_* < 2^{j_*+1}$. Every block lies in $[0, 2T_*]$, and every $L = 2^j$ satisfies $1 \leq L \leq T_*$; thus Lemmas 10.19 and 10.20 apply. For $r > 4$, the noise scales satisfy

$$\sigma_*^{r/2} + \sum_{j=0}^{j_*} \sigma_{2^j}^{r/2} \leq C_r \left(\frac{1 + (dT_*)^2}{M} \right)^{r/4} \leq C_r \left(\frac{1 + \log^2 M}{M} \right)^{r/4}. \quad (10.82)$$

Indeed, by (10.75), $\sigma_{2^j}^{r/2} \leq C_r [(d2^j)^{r/4} + (d2^j)^{r/2}] M^{-r/4}$, and the two geometric sums over $j \leq j_*$ are bounded by constants times their last terms, which are at most $C_r (dT_*)^{r/4}$ and $C_r (dT_*)^{r/2}$; the initial term is at most $(d/M)^{r/4}$, and $dT_* = C_T \log M$. In particular, the number of blocks does not enter, and no factor depending on $\log(1/d)$ appears.

Let $\iota = \inf_{0 \leq t \leq T_*} \mathcal{A}_t / \mathcal{H}(t)$. Fix $r' > 0$, and apply the second bound of (10.81) on every block with $\gamma = 2r' + 2$ and $r = 8$. By (10.82) and a union bound, for $0 < x \leq 1/2$,

$$\mathbb{P}(\iota \leq x) \leq C x^{r'+1} \left(\frac{1 + \log^2 M}{M} \right)^2.$$

Integrating this tail bound, we find

$$\mathbb{E} \iota^{-r'} \leq 2^{r'} + \int_2^\infty r' y^{r'-1} \mathbb{P}(\iota \leq 1/y) dy \leq 2^{r'} + C \left(\frac{1 + \log^2 M}{M} \right)^2 \leq C_{r'},$$

which is the first bound in (10.66). For the second, we have $\sup_{t \leq T_*} N \bar{h}(u+t) / \mathcal{A}_t \leq C_{\text{init}} \iota^{-1}$ by (10.69), and the claim follows from the Cauchy–Schwarz inequality, the first bound at order $2r'$, and (10.70).

Finally, by (10.41), we have the exact identity

$$\vartheta\varphi(X_t) = \frac{\mathcal{W}_t}{N} \cdot \frac{N\bar{h}(u+t)}{\mathcal{H}(t)} \cdot \frac{\mathcal{H}(t)}{\mathcal{A}_t} \cdot \frac{\vartheta}{\bar{h}(u+t)}.$$

The last factor is at most one, since $\bar{h}(u+t) \geq \vartheta$ by (10.43). Hence $\sup_{t \leq T_*} \vartheta\varphi(X_t) \leq (\sup_{t \leq T_*} \mathcal{W}_t/N) C_{\text{init}} \iota^{-1}$, and Hölder's inequality with three factors, together with (10.70) and the first bound in (10.66), proves (10.67). The laws and times in this argument are those of the diffusion with initial law μ_u , for every $u \geq 0$, including $u = 0$, where $\mu_0 = \nu_N$. $\square$

10.8 A Feynman–Kac bound for a weighted gradient

Let A be a real symmetric matrix with eigenvalues $\alpha_1 \geq \dots \geq \alpha_N$, Lagrange multiplier z , and smallest precision $d = z - \alpha_1$ (Definition 3.3). As in Section 8, let $\varphi(x) = x^\top(zI - A)^{-1}x/N$; in eigencoordinates, this is the quadratic form (10.40). We write $\Gamma(f_1, f_2) = \nabla_{S_N} f_1 \cdot \nabla_{S_N} f_2$, $\Gamma(f) = \Gamma(f, f)$, and

$$\Gamma_{\varphi,2}(f) = \frac{1}{2} \left\{ L_A(\varphi\Gamma(f)) - 2\varphi\Gamma(f, L_A f) \right\}.$$

This is a weighted version of the iterated carré du champ; note that $\varphi\Gamma$ is not, in general, the carré du champ of L_A . Let $g = \nabla_{S_N} f$, let $a = \nabla_{S_N} \log \varphi$, and let $\nabla_S^2 f$ be the tangential Hessian of f , as in Lemma 8.3. By (8.8), for every smooth f on S_N ,

$$\Gamma_{\varphi,2}(f) = \varphi \left\{ \|\nabla_S^2 f\|_{\text{HS}}^2 + 2\nabla_S^2 f(a, g) + g^\top \left[(zI - A) - \frac{2}{N}I \right] g \right\}, \quad (10.83)$$

where $\nabla_S^2 f(a, g) = a^\top (\nabla_S^2 f) g$.

Lemma 10.21 (A Feynman–Kac bound for the weighted gradient). *Let $\bar{r} = d - 2/N$. For every smooth function f on S_N , every $x \in S_N$, and every $t \geq 0$,*

$$\varphi(x) \Gamma(P_t^A f)(x) \leq e^{-2\bar{r}t} \mathbb{E}_x \left[\exp \left(2 \int_0^t \Gamma(\log \varphi)(X_s) ds \right) \varphi(X_t) \Gamma(f)(X_t) \right], \quad (10.84)$$

where X is the diffusion generated by L_A started at x .

The inequality holds whatever the sign of $\bar{r}$. In Section 14, the exponential factor in (10.84) is controlled by Theorem 10.15.

Proof. We begin with a pointwise lower bound for $\Gamma_{\varphi,2}$. For a symmetric matrix H and vectors b, v , the Cauchy–Schwarz inequality gives $2|H(b, v)| \leq 2\|H\|_{\text{HS}}|b||v| \leq \|H\|_{\text{HS}}^2 +$

$|b|^2|v|^2$. With $H = \nabla_S^2 f$, $b = a$, and $v = g$, we have $|b|^2 = \Gamma(\log \varphi)$ and $|v|^2 = \Gamma(f)$. The eigenvalues of $zI - A$ are at least d , and by (10.83), we obtain

$$\Gamma_{\varphi,2}(f) \geq (\bar{r} - \Gamma(\log \varphi)) \varphi \Gamma(f). \quad (10.85)$$

Since L_A is an elliptic operator with smooth coefficients on the compact manifold S_N , the function $(t, x) \mapsto P_t^A f(x)$ is smooth on $[0, \infty) \times S_N$ for smooth f , and so is $\phi(t, x) = \varphi(x) \Gamma(P_t^A f)(x)$. Using $\partial_t P_t^A f = L_A P_t^A f$, the definition of $\Gamma_{\varphi,2}$, and (10.85), we find

$$(\partial_t - L_A)\phi = 2\varphi \Gamma(P_t^A f, L_A P_t^A f) - L_A(\varphi \Gamma(P_t^A f)) = -2\Gamma_{\varphi,2}(P_t^A f) \leq \mathcal{V}\phi,$$

where $\mathcal{V} = 2\Gamma(\log \varphi) - 2\bar{r}$ is smooth and bounded on S_N . Fix $t > 0$ and $x \in S_N$, let X be the diffusion generated by L_A started at x , and let

$$Y_s = \exp\left(\int_0^s \mathcal{V}(X_v) dv\right) \phi(t-s, X_s), \quad 0 \leq s \leq t.$$

By Itô's formula for (3.2) with cW_N replaced by A , whose generator is L_A by the calculation in Section 3.1, the drift of Y is

$$\exp\left(\int_0^s \mathcal{V}(X_v) dv\right) (\mathcal{V}\phi - \partial_t \phi + L_A \phi)(t-s, X_s) \geq 0,$$

and its martingale part is a stochastic integral with a bounded integrand. Thus Y is a submartingale, and $\phi(t, x) = Y_0 \leq \mathbb{E}Y_t$, which is (10.84). This is the Feynman–Kac argument; see [123, Proposition VIII.3.10, pp. 358–359] for the generator identity underlying it, after a constant shift of the bounded potential. $\square$

Score bounds, smoothing, and conditional relative entropy

In this section, we collect the analytic tools that we use to prove upper bounds on the total variation distance at low temperature, in Section 16 (fixed $c > 1$) and Section 17 (the approach to the critical temperature from below). In both regimes, we compare the time- t law with the product of a law of finitely many eigencoordinates and a Gaussian law of the remaining coordinates. The results below control the two factors and their combination.

In Section 11.1, we define weak scores along rotations and record three elementary facts: the effect of a cutoff function on a score (Lemma 11.2), the total variation cost of a path weight (Lemma 11.3), and the smoothness of weighted and killed endpoint measures (Lemma 11.4). Section 11.2 proves a Bismut-type formula for the score over a final time interval: Theorem 11.5 bounds the moments of the derivative along a rotation of the logarithm of the density of a weighted endpoint law, using only the final time interval, and Corollary 11.6 specializes it to the time scale a_2^{-1} . In Section 11.3, we show that weak convergence of finite-dimensional laws implies convergence in total variation when nearby laws have bounded Fisher information (Theorem 11.8). Section 11.4 combines these results and proves total variation convergence of finitely many standardized eigencoordinates (Corollary 11.12). Section 11.5 concerns the conditional relative entropy of the remaining coordinates given the first ones. It contains an exact identity for its time derivative (Proposition 11.18), a logarithmic Sobolev inequality for it (Proposition 11.16), and an entropy dissipation inequality under a cutoff function adapted to the approach to the critical temperature (Theorem 11.19). Finally, Lemma 11.21 in Section 11.6 combines the two factors in total variation.

Besides the conventions of Section 3, in particular Lemmas 3.1 and 3.6, we use only classical tools: Itô's formula, Lévy's characterization of Brownian motion, Hörmander's hypoellipticity theorem, and the Gaussian logarithmic Sobolev inequality. The special case of Girsanov's theorem that we need is proved in Section 11.2. All statements are deterministic in the coefficients of the potential. Probabilities and expectations refer only to the initial point and the driving Brownian motion. Section 16 uses Theorem 11.5, Theorem 11.8 and Lemma 11.7, the results of Sections 11.4 to 11.5, and Lemma 11.21. Section 17 uses Corollary 11.12, Theorem 11.19, and Lemma 11.21; the hypotheses of

these statements are verified there.

11.1 Setting, weak scores, and weighted endpoint measures

The diffusion. Throughout this section, $a_1, \dots, a_N$ are real numbers, coordinates refer to a fixed orthonormal basis of $\mathbb{R}^N$, and

$$F(x) = \sum_{k=1}^N a_k x_k^2, \quad L = \Delta_{S_N} - \nabla_{S_N} F \cdot \nabla_{S_N}. \quad (11.1)$$

For the model at low temperature, the coordinates are the eigencoordinates and the $a_k = c(\lambda_1 - \lambda_k)/2$ are the scaled eigenvalue gaps of (3.8). More generally, by (3.7), the generator L_A of (3.3) is of the form (11.1) with $a_k = (\alpha_1 - \alpha_k)/2$ in an eigenbasis of A . As in (3.2), the diffusion generated by L solves

$$dX_s = \sqrt{2} \Pi(X_s) dB_s + \mathbf{b}(X_s) ds, \quad \mathbf{b}(x) = -\frac{N-1}{N} x - \nabla_{S_N} F(x), \quad \Pi(x) = I - \frac{xx^\top}{N}, \quad (11.2)$$

where $\nabla_{S_N} F(x) = \Pi(x) \nabla F(x)$. We always work on a filtered probability space with a filtration $(\mathcal{F}_s)_{s \geq 0}$, on which B is an $(\mathcal{F}_s)$ -Brownian motion in $\mathbb{R}^N$ and X is $(\mathcal{F}_s)$ -adapted, with $X_0 \sim \nu_N$. All process-specific statements in this chapter concern this original uniformly initialized diffusion; the statements about probability measures and weak scores are independent of a diffusion. Solutions of (11.2) remain on S_N , by the computation following (3.2). The coefficients are smooth and may be modified outside a neighborhood of S_N to become globally Lipschitz. This implies pathwise uniqueness, and that, for $u \geq 0$, the path $(X_s)_{s \geq u}$ is a measurable function of X_u and of the increments of B after time u .

Rotations and reflections. For $i \neq j$, let $\Omega_{ij} = e_j e_i^\top - e_i e_j^\top$ and $\mathcal{R}_\vartheta^{ij} = e^{\vartheta \Omega_{ij}}$, the rotation by the angle ϑ in the (i, j) -plane, and let K_{ij} be the vector field

$$K_{ij}\varphi(x) = \frac{d}{d\vartheta} \varphi(\mathcal{R}_\vartheta^{ij} x) \Big|_{\vartheta=0} = x_i \partial_j \varphi(x) - x_j \partial_i \varphi(x).$$

It is tangent to S_N , and $K_{ij}x = \Omega_{ij}x$. Since each $\mathcal{R}_\vartheta^{ij}$ preserves ν_N , differentiating $\int \varphi \circ \mathcal{R}_\vartheta^{ij} d\nu_N = \int \varphi d\nu_N$ at $\vartheta = 0$, we obtain

$$\int_{S_N} K_{ij} \varphi d\nu_N = 0 \quad \text{for every } \varphi \in C^1(S_N). \quad (11.3)$$

We write $R_1(x) = (-x_1, x_2, \dots, x_N)$ for the reflection of the first coordinate, as in Lemma 3.1. The equation (11.2) is invariant under R_1 , because F is even in x_1 : if X solves it with Brownian motion B , then $R_1 X$ solves it with Brownian motion $R_1 B$.

Weak scores. The following definition makes sense of a logarithmic derivative along a rotation without assuming that a density is differentiable or positive.

Definition 11.1 (Weak score). Let ρ be a probability measure on S_N with $\rho \ll \nu_N$, and let $K = K_{ij}$. A function $s \in L^1(\rho)$ is a *weak K -score* of ρ if

$$\int K\varphi d\rho = - \int \varphi s d\rho \quad \text{for every } \varphi \in C^\infty(S_N). \quad (11.4)$$

We then write $s = K \log(d\rho/d\nu_N)$.

A weak K -score is unique ρ -almost everywhere, because (11.4) determines the signed measure $s\rho$ on the smooth functions, which are dense in $C(S_N)$. Approximating a function of class C^1 uniformly, together with its K -derivative, by smooth functions, we see that (11.4) also holds for $\varphi \in C^1(S_N)$. If $d\rho/d\nu_N = f$ is of class C^1 and positive on an open set U , then $s = Kf/f$ almost everywhere on U . This follows from (11.4) with φ supported in U and (11.3) applied to φf . In particular, the weak score is the usual score where the latter is defined. We note that (11.4) is a global identity. A jump of nonzero size across a smooth hypersurface transverse to K contributes a singular part to Kf , and therefore prevents a weak K -score. For this reason, the cutoff functions below are smooth.

The first lemma computes the score after multiplication by a cutoff function. For $p = 1$, its hypothesis is $|K\eta| \leq B\eta^{1/2}$.

Lemma 11.2 (Multiplying by a cutoff function). Let ρ be a probability measure on S_N with a weak K -score s , where $K = K_{ij}$, and let $p \geq 1$ be an integer with $s \in L^{2p}(\rho)$. Let $\eta \in C^1(S_N)$ take values in $[0, 1]$, with $Z_\eta = \int \eta d\rho > 0$. Suppose that there exists $B \geq 0$ such that

$$|K\eta| \leq B\eta^{1-1/(2p)} \quad \rho\text{-almost everywhere.} \quad (11.5)$$

Then the normalized cutoff-weighted law $\hat{\rho} = \eta\rho/Z_\eta$, with normalizing mass Z_η , has the weak K -score $\hat{s} = s + \mathbf{1}_{\{\eta>0\}}K\eta/\eta$, and

$$\mathbb{E}_{\hat{\rho}}|\hat{s}|^{2p} \leq \frac{2^{2p-1}}{Z_\eta} (\mathbb{E}_\rho|s|^{2p} + B^{2p}). \quad (11.6)$$

Proof. We have $K\eta = 0$ ρ -almost everywhere on $\{\eta = 0\}$, by (11.5). Let $\varphi \in C^\infty(S_N)$. Applying the identity (11.4) to $\eta\varphi \in C^1(S_N)$, we find

$$\int K\varphi d\hat{\rho} = \frac{1}{Z_\eta} \int (K(\eta\varphi) - \varphi K\eta) d\rho = -\frac{1}{Z_\eta} \int \varphi (\eta s + K\eta) d\rho = - \int \varphi \hat{s} d\hat{\rho}.$$

This proves the formula for the score, once we know that $\hat{s} \in L^1(\hat{\rho})$, which follows from the moment bound below. By convexity, $|\hat{s}|^{2p} \leq 2^{2p-1}(|s|^{2p} + \mathbf{1}_{\{\eta>0\}}|K\eta/\eta|^{2p})$. We have $\mathbb{E}_{\hat{\rho}}|s|^{2p} = Z_\eta^{-1} \int \eta|s|^{2p} d\rho \leq Z_\eta^{-1} \mathbb{E}_\rho|s|^{2p}$, and, by (11.5),

$$\mathbb{E}_{\hat{\rho}} \left[\mathbf{1}_{\{\eta>0\}} \left| \frac{K\eta}{\eta} \right|^{2p} \right] = \frac{1}{Z_\eta} \int_{\{\eta>0\}} |K\eta|^{2p} \eta^{1-2p} d\rho \leq \frac{B^{2p}}{Z_\eta}.$$

Combining the two bounds, we obtain (11.6). $\square$

The second lemma bounds the total variation cost of a path weight. We use it in the following form: if $0 \leq U \leq 1$ is a random variable with $\mathbb{E}U > 0$ and Y is a random element, then the laws of Y under $\mathbb{P}$ and under $(U/\mathbb{E}U)\mathbb{P}$ are within $1 - \mathbb{E}U$ in total variation. This is the case $\sigma(A) = \mathbb{E}[U\mathbf{1}_{\{Y \in A\}}]$ and $\rho = \text{Law}(Y)$ below.

Lemma 11.3 (Removing a path weight). *Let ρ be a probability measure and $\sigma \leq \rho$ a measure on the same space with total mass $Z > 0$. Then $\|\rho - \sigma/Z\|_{\text{TV}} \leq 1 - Z$.*

Proof. For both probability measures, $\|\rho - \sigma/Z\|_{\text{TV}}$ is the supremum of $\sigma(A)/Z - \rho(A)$ over measurable sets A , by the first part of Lemma 3.6. Since $\rho(A) \geq \sigma(A)$, we have $\sigma(A)/Z - \rho(A) \leq (Z^{-1} - 1)\sigma(A) \leq 1 - Z$. $\square$

The third lemma records the forward equation and the smoothness of the endpoint measures that appear later: the path is weighted by a variable known at a time T , and possibly killed after T when it leaves an open set.

Lemma 11.4 (Weighted and killed endpoint measures). *Let X solve (11.2) with $X_0 \sim \nu_N$, let $T \geq 0$, let Ψ be an $\mathcal{F}_T$ -measurable random variable with values in $[0, 1]$, and let $D \subset S_N$ be open (possibly $D = S_N$). Let $\tau = \inf\{s \geq T : X_s \notin D\}$, with $\inf \emptyset = \infty$, and for $t \geq T$ define the endpoint subprobability measure*

$$\sigma_t(A) = \mathbb{E}[\Psi \mathbf{1}_{\{\tau > t\}} \mathbf{1}_{\{X_t \in A\}}]. \quad (11.7)$$

1. *For every $\varphi \in C^\infty(S_N)$ with compact support in D and all $T \leq s \leq t$,*

$$\sigma_t(\varphi) - \sigma_s(\varphi) = \int_s^t \sigma_r(L\varphi) dr. \quad (11.8)$$

2. *There exists $u \in C^\infty((T, \infty) \times D)$ such that $\sigma_t(dx) = u(t, x) \nu_N(dx)$ for every $t > T$.*

3. *If D and σ_T are invariant under R_1 , then σ_t is invariant under R_1 for every $t \geq T$.*

Proof. For the first part, note that $X_\tau \notin D$ on $\{\tau < \infty\}$, because D is open and the path is continuous. Let $\chi \in C^\infty(\mathbb{R} \times S_N)$ be such that each $\chi(r, \cdot)$ has its support in a fixed compact subset of D . By Itô's formula,

$$\chi(t \wedge \tau, X_{t \wedge \tau}) - \chi(s \wedge \tau, X_{s \wedge \tau}) = \int_{s \wedge \tau}^{t \wedge \tau} (\partial_r + L)\chi(r, X_r) dr + \sqrt{2} \int_{s \wedge \tau}^{t \wedge \tau} \nabla_{S_N} \chi(r, X_r) \cdot dB_r.$$

On $\{\tau \leq t\}$ we have $\chi(\tau, X_\tau) = 0$, and the same holds with s in place of t . We multiply by Ψ and take expectations. The stochastic integral has a bounded integrand and starts after time T , and its product with the $\mathcal{F}_T$ -measurable variable Ψ has mean zero. Using Fubini's theorem, we obtain

$$\sigma_t(\chi(t, \cdot)) - \sigma_s(\chi(s, \cdot)) = \int_s^t \sigma_r((\partial_r + L)\chi(r, \cdot)) dr, \quad (11.9)$$

and (11.8) is the case of a χ that does not depend on r .

For the second part, we take $\chi \in C_c^\infty((T, \infty) \times D)$, $s = T$, and t larger than every time in the support of χ . Then (11.9) says that the measure $\sigma_r(dx) dr$ on $(T, \infty) \times D$ is annihilated, in the sense of distributions, by the formal adjoint of $\partial_r + L$. In local coordinates on D , this adjoint is $P = -\partial_r + L^*$, where L^* is the formal adjoint of L with respect to the coordinate Lebesgue measure. Its second-order part is that of L , which is $\sum_{k,l} a^{kl} \partial_k \partial_l$ with a smooth positive definite matrix a . Writing $a = ss^T$ with $s = a^{1/2}$ smooth, we have $P = \sum_k V_k^2 + V_0 + c_0$, where $V_k = \sum_l s_{lk} \partial_l$, c_0 is a smooth function, and $V_0 = -\partial_r + V'_0$ with V'_0 a first-order operator in the space variables. The fields $V_1, \dots, V_{N-1}$ span all space directions, and V_0 has a nonzero time component. This implies that $V_0, \dots, V_{N-1}$ are linearly independent at every point. By Hörmander's theorem [85, Theorem 1.1], P is hypoelliptic, and the measure $\sigma_r(dx) dr$ is a smooth function times $dx dr$. In other words, there exists $u \in C^\infty((T, \infty) \times D)$ such that $\sigma_r(dx) dr = u(r, x) \nu_N(dx) dr$. For $\varphi \in C_c^\infty(D)$, both $r \mapsto \sigma_r(\varphi)$ and $r \mapsto \int \varphi u(r, \cdot) d\nu_N$ are continuous on (T, ∞) , the first by (11.8). They agree for almost every r , hence for every $r > T$. We conclude that $\sigma_r = u(r, \cdot) \nu_N$ on D . Since $\sigma_r(D^c) = 0$ for $r \geq T$, the second part follows.

For the third part, let $k_v(x, A)$ be the probability that the solution of (11.2) started at x stays in D during $[0, v]$ and lies in A at time v . By the Markov property at the time T , we have $\sigma_t(A) = \int \sigma_T(dx) k_{t-T}(x, A)$. Since the equation and D are invariant under R_1 , we have $k_v(R_1 x, R_1 A) = k_v(x, A)$. Together with the invariance of σ_T , this gives $\sigma_t(R_1 A) = \sigma_t(A)$. $\square$

Conditioned, weighted, and killed measures. We use the following names for the measures that appear here and in Sections 16 and 17. For an event E with $\mathbb{P}(E) > 0$, the law conditioned on E is $\mathbb{P}(\cdot \mid E) = \mathbb{P}(\cdot \cap E) / \mathbb{P}(E)$, with normalizing mass $\mathbb{P}(E)$. The measure σ_t of (11.7) is the endpoint subprobability measure weighted by Ψ and killed on leaving D . It is bounded by the law of X_t , and its mass $\sigma_t(S_N) \leq 1$ need not be one. For a cutoff function $\eta : S_N \rightarrow [0, 1]$ with $\int \eta d\sigma_t > 0$, the normalized cutoff-weighted law is the probability measure $\eta \sigma_t / \int \eta d\sigma_t$, with normalizing mass $\int \eta d\sigma_t$. If $D = S_N$ and $\Psi = \mathbf{1}_E$ for an event $E \in \mathcal{F}_T$, then $\sigma_t(S_N) = \mathbb{P}(E)$, and $\sigma_t / \sigma_t(S_N)$ is the law of X_t under $\mathbb{P}(\cdot \mid E)$.

11.2 A Bismut-type formula for the score over a final time interval

Formulas that express the derivative of a diffusion semigroup through a stochastic integral, without differentiating the initial data, go back to Bismut [32] and Elworthy and Li [67, Theorems 2.1 and 3.1 and Corollary 3.2]; see also [87, Section 8.3, Theorems 8.3.3–8.3.4]. Theorem 11.5 below is a version of this idea for the logarithmic derivative of an endpoint density along a rotation. Three features are needed in the applications. The derivative is that of the density of an endpoint law weighted by a functional of the

path and normalized. The paths are restricted to a rotation-invariant open set on which the change of drift is small. The bound uses only a final time interval of the original uniformly initialized diffusion. It requires no score bound before that interval and no new initialization at its left endpoint; the law there may be the evolved and prefix-weighted law. The proof differentiates a Girsanov density with respect to a rotation of the path during the final interval.

Setting. Fix $i \neq j$ and write $K = K_{ij}$, $\Omega = \Omega_{ij}$, $\mathcal{R}_\vartheta = \mathcal{R}_\vartheta^{ij}$, and $\Delta = |a_j - a_i|$. Let $D \subset S_N$ be an open set with $\mathcal{R}_\vartheta D = D$ for every $\vartheta \in \mathbb{R}$, and let $H < \infty$ satisfy

$$x_i^2 + x_j^2 \leq H \quad \text{for } x \in D. \quad (11.10)$$

Fix $t > 0$ and $\ell \in (0, t]$, let $I = [t - \ell, t]$, and let $\kappa : I \rightarrow \mathbb{R}$ be absolutely continuous with $\kappa(t - \ell) = 0$, $\kappa(t) = 1$, and $\int_I \kappa'(s)^2 ds < \infty$. For $\varepsilon \in \mathbb{R}$, let Θ_ε act on paths $\omega \in C(I; S_N)$ by $(\Theta_\varepsilon \omega)_s = \mathcal{R}_{\varepsilon \kappa(s)} \omega_s$. Let ξ be an $\mathcal{F}_{t-\ell}$ -measurable random element of a measurable space E , and let $\Phi : E \times C(I; S_N) \rightarrow [0, 1]$ be measurable with

$$\Phi(e, \omega) = 0 \quad \text{unless } \omega_s \in D \text{ for all } s \in I, \quad \Phi(e, \Theta_\varepsilon \omega) = \Phi(e, \omega) \quad \text{for all } \varepsilon \in \mathbb{R}. \quad (11.11)$$

We call $W = \Phi(\xi, X|_I)$ a path weight for the final interval I . For example, ξ may be the path up to time $t - \ell$, together with any other $\mathcal{F}_{t-\ell}$ -measurable variable. We assume $Z = \mathbb{E}W > 0$ and define the weighted endpoint law $\rho^W(A) = Z^{-1} \mathbb{E}[W \mathbf{1}_{\{X_t \in A\}}]$, whose normalizing mass is Z .

Theorem 11.5 (A Bismut-type formula for the score over a final time interval). *For the uniformly initialized diffusion in this setting, the law ρ^W has a weak K-score, and for every integer $p \geq 1$,*

$$\mathbb{E}_{\rho^W} \left| K \log \frac{d\rho^W}{d\nu_N} \right|^{2p} \leq \frac{C_p}{Z} \left[H \int_I (\kappa'(s)^2 + \Delta^2 \kappa(s)^2) ds \right]^p, \quad (11.12)$$

where C_p depends only on p . For the linear ramp $\kappa(s) = (s - t + \ell)/\ell$,

$$\mathbb{E}_{\rho^W} \left| K \log \frac{d\rho^W}{d\nu_N} \right|^{2p} \leq \frac{C_p}{Z} [H(\ell^{-1} + \Delta^2 \ell)]^p. \quad (11.13)$$

The proof has four steps. We first show that rotating the path during I changes its law by an explicit Girsanov density. Using the invariance of W , we deduce the identity (11.19). Differentiating it at $\varepsilon = 0$ represents $\int K \varphi d\rho^W$ through a stochastic integral M . Finally, the moments of M are bounded by its quadratic variation.

Proof. Let $\tau = \inf\{s \geq t - \ell : X_s \notin D\}$. All integrals below may be restricted to $\{\tau > t\}$, because $W = 0$ unless $\tau > t$.

An exponential bound. We use the following elementary bound twice. Let M' be a continuous local martingale with $M'_0 = 0$ and $\langle M' \rangle_t \leq V$ for a deterministic constant $V \geq 0$. If $V = 0$, then $M'_s = 0$ almost surely for all $s \leq t$, so its exponential moments equal one and all its positive moments and tails vanish. Suppose now that $V > 0$. For $\lambda \in \mathbb{R}$, the process $\exp(\lambda M'_s - \lambda^2 \langle M' \rangle_s / 2)$ is a nonnegative local martingale, hence a supermartingale with expectation at most one. Therefore

$$\mathbb{E} e^{\lambda M'_t} \leq e^{\lambda^2 V / 2}, \quad \mathbb{P}(|M'_t| > r) \leq 2e^{-r^2 / (2V)}, \quad \mathbb{E}|M'_t|^{2p} \leq 2p! (2V)^p \quad (11.14)$$

for $\lambda \in \mathbb{R}$, $r > 0$, and integers $p \geq 1$. The second bound follows from the first by Markov's inequality with $\lambda = \pm r / V$, and the third from the second, since $\mathbb{E}|M'_t|^{2p} = \int_0^\infty 2pr^{2p-1} \mathbb{P}(|M'_t| > r) dr$.

Step 1: the change of drift. For $\vartheta \in \mathbb{R}$, the function $y \mapsto F(y) - F(\mathcal{R}_{-\vartheta} y)$ is the quadratic form of a symmetric matrix Q_ϑ that acts only on the (i, j) -plane. In that plane, we write $\text{diag}(a_i, a_j) = \frac{a_i + a_j}{2} I + \frac{a_i - a_j}{2} E$ with $E = \text{diag}(1, -1)$. Then $Q_\vartheta = \frac{a_i - a_j}{2} (E - \mathcal{R}_\vartheta E \mathcal{R}_\vartheta^\top)$, and $\mathcal{R}_\vartheta E \mathcal{R}_\vartheta^\top$ is the reflection matrix with entries $\cos 2\vartheta, \sin 2\vartheta, \sin 2\vartheta, -\cos 2\vartheta$. We have

$$\|Q_\vartheta\| = \Delta |\sin \vartheta|, \quad \|\partial_\vartheta Q_\vartheta\| = \Delta, \quad \|\partial_\vartheta^2 Q_\vartheta\| = 2\Delta, \quad (11.15)$$

since $E - \mathcal{R}_\vartheta E \mathcal{R}_\vartheta^\top$ has eigenvalues $\pm 2|\sin \vartheta|$, and the first two derivatives of the reflection matrix are 2 and 4 times symmetric orthogonal matrices. For $s \in I$ and $y \in S_N$, we define the tangent vector field

$$u_\varepsilon(s, y) = \varepsilon \kappa'(s) \Omega y + \nabla_{S_N} \{F - F \circ \mathcal{R}_{-\varepsilon \kappa(s)}\}(y). \quad (11.16)$$

We have $|\Omega y| = (y_i^2 + y_j^2)^{1/2}$ and $|\nabla(y^\top Q_\vartheta y)| = 2|Q_\vartheta y| \leq 2\|Q_\vartheta\|(y_i^2 + y_j^2)^{1/2}$, and the tangent projection does not increase lengths. Using (11.10) and (11.15), we obtain

$$|u_\varepsilon(s, y)| \leq |\varepsilon| \sqrt{H} (|\kappa'(s)| + 2\Delta |\kappa(s)|) \quad (y \in D). \quad (11.17)$$

We set

$$\Lambda_\varepsilon = \exp \left\{ \frac{1}{\sqrt{2}} \int_{t-\ell}^{\tau \wedge t} u_\varepsilon(s, X_s) \cdot dB_s - \frac{1}{4} \int_{t-\ell}^{\tau \wedge t} |u_\varepsilon(s, X_s)|^2 ds \right\}. \quad (11.18)$$

The integrand is bounded by a deterministic function in $L^2(I)$, by (11.17). The following special case of Girsanov's theorem (see [123, Chapter VIII, Section 1, Theorem (1.4)]) is then elementary. We include the argument. Let $v_s = 2^{-1/2} \mathbf{1}_{\{s < \tau\}} u_\varepsilon(s, X_s)$ for $s \in I$ and $v_s = 0$ for $s < t - \ell$, and let $\Lambda_s = \exp(\int_0^s v_r \cdot dB_r - \frac{1}{2} \int_0^s |v_r|^2 dr)$, so that $\Lambda_\varepsilon = \Lambda_t$. The quadratic variation $\int_0^t |v_r|^2 dr$ is bounded by a deterministic constant V_ε . For $q > 1$ and every stopping time $\sigma \leq t$, we have $\Lambda_\sigma^q = \mathcal{E}_\sigma \exp\{(q^2 - q) \int_0^\sigma |v_r|^2 dr / 2\}$, where $\mathcal{E}$ is the stochastic exponential of $q \int v \cdot dB$, a nonnegative local martingale and hence a supermartingale. Thus $\mathbb{E} \Lambda_\sigma^q \leq e^{(q^2 - q)V_\varepsilon / 2}$, the local martingale Λ is of class (D) on $[0, t]$,

and it is a martingale. In particular, $\mathbb{Q}_\varepsilon = \Lambda_\varepsilon \mathbb{P}$ is a probability measure that agrees with $\mathbb{P}$ on $\mathcal{F}_{t-\ell}$. Using Itô's product formula, we see that each component of $\Lambda_s(B_s - \int_0^s v_r dr)$ is a local $\mathbb{P}$ -martingale. By Bayes' formula, after localization, $\widehat{B}_s = B_s - \int_0^s v_r dr$ is a continuous local $\mathbb{Q}_\varepsilon$ -martingale with quadratic covariation sI_N , and Lévy's characterization [123, Theorem IV.3.6] shows that $\widehat{B}$ is a $\mathbb{Q}_\varepsilon$ -Brownian motion. For $s \in I$, we have $\widehat{B}_s = B_s - 2^{-1/2} \int_{t-\ell}^{s \wedge \tau} u_\varepsilon(r, X_r) dr$. Because u_ε is tangent, we have $\Pi(X_s)u_\varepsilon = u_\varepsilon$, and under $\mathbb{Q}_\varepsilon$ the process X solves $dX_s = \sqrt{2}\Pi(X_s)d\widehat{B}_s + (\mathbf{b} + u_\varepsilon)(s, X_s)ds$ for $s \in [t-\ell, \tau \wedge t]$.

We define $\widehat{X}_s = \mathcal{R}_{-\varepsilon\kappa(s)}X_s$ for $s \in I$. The map $s \mapsto \mathcal{R}_{-\varepsilon\kappa(s)}$ is absolutely continuous and Ω commutes with every $\mathcal{R}_\theta$, and we find, for $s \leq \tau \wedge t$,

$$d\widehat{X}_s = -\varepsilon\kappa'(s)\Omega\widehat{X}_s ds + \mathcal{R}_{-\varepsilon\kappa(s)}[\sqrt{2}\Pi(X_s)d\widehat{B}_s + (\mathbf{b} + u_\varepsilon)(s, X_s)ds].$$

For an orthogonal matrix $\mathcal{R}$, we have $\mathcal{R}^\top \Pi(\mathcal{R}x)\mathcal{R} = \Pi(x)$ and $\mathcal{R}^\top \nabla_{S_N} \varphi(\mathcal{R}x) = \nabla_{S_N}(\varphi \circ \mathcal{R})(x)$ for every smooth φ . With $\mathcal{R} = \mathcal{R}_{\varepsilon\kappa(s)}$, we deduce from these identities that $\mathcal{R}^\top \mathbf{b}(\mathcal{R}\widehat{x}) = -\frac{N-1}{N}\widehat{x} - \nabla_{S_N}(F \circ \mathcal{R})(\widehat{x})$ and $\mathcal{R}^\top u_\varepsilon(s, \mathcal{R}\widehat{x}) = \varepsilon\kappa'(s)\Omega\widehat{x} + \nabla_{S_N}(F \circ \mathcal{R} - F)(\widehat{x})$. The terms with Ω and with $F \circ \mathcal{R}$ cancel, which gives

$$d\widehat{X}_s = \sqrt{2}\Pi(\widehat{X}_s)d\check{B}_s + \mathbf{b}(\widehat{X}_s)ds, \quad \check{B}_s = \int_{t-\ell}^s \mathcal{R}_{-\varepsilon\kappa(r)}d\widehat{B}_r,$$

for $s \in [t-\ell, \tau \wedge t]$. By Lévy's characterization, $\check{B}$ is a $\mathbb{Q}_\varepsilon$ -Brownian motion. Moreover, $\widehat{X}_{t-\ell} = X_{t-\ell}$, and the exit time of $\widehat{X}$ from D is τ , because D is rotation invariant. Thus, under $\mathbb{Q}_\varepsilon$, the path $\widehat{X}$ stopped at its exit time solves (11.2) with the same $\mathcal{F}_{t-\ell}$ -measurable data $(\xi, X_{t-\ell})$, whose joint law is the same under $\mathbb{Q}_\varepsilon$ and $\mathbb{P}$. Pathwise uniqueness implies that the joint law of ξ and the stopped path $\widehat{X}$ under $\mathbb{Q}_\varepsilon$ equals the joint law of ξ and the stopped path X under $\mathbb{P}$.

Let φ be bounded and measurable on S_N . The function $(e, \omega) \mapsto \Phi(e, \omega)\varphi(\mathcal{R}_\varepsilon\omega_t)$ vanishes unless ω stays in D , by (11.11). In particular, it depends only on e and on the path stopped at its exit time. Using the equality of laws, then the invariance in (11.11), and finally $X = \Theta_\varepsilon\widehat{X}$ and $\kappa(t) = 1$, we find

$$\begin{aligned} \mathbb{E}[\Phi(\xi, X|_I)\varphi(\mathcal{R}_\varepsilon X_t)] &= \mathbb{E}_{\mathbb{Q}_\varepsilon}[\Phi(\xi, \widehat{X}|_I)\varphi(\mathcal{R}_\varepsilon \widehat{X}_t)] \\ &= \mathbb{E}_{\mathbb{Q}_\varepsilon}[\Phi(\xi, \Theta_\varepsilon \widehat{X}|_I)\varphi(\mathcal{R}_\varepsilon \widehat{X}_t)] = \mathbb{E}_{\mathbb{Q}_\varepsilon}[W\varphi(X_t)]. \end{aligned}$$

That is,

$$\mathbb{E}[W\varphi(\mathcal{R}_\varepsilon X_t)] = \mathbb{E}[W\Lambda_\varepsilon \varphi(X_t)] \quad (\varepsilon \in \mathbb{R}). \quad (11.19)$$

For $\varphi = 1$, this shows that $\mathbb{E}[W\Lambda_\varepsilon] = Z$ for every ε .

Step 2: differentiation at $\varepsilon = 0$. Let

$$M = \frac{1}{\sqrt{2}} \int_{t-\ell}^{\tau \wedge t} \{\kappa'(s)\Omega X_s + \kappa(s)\nabla_{S_N}(KF)(X_s)\} \cdot dB_s. \quad (11.20)$$

The integrand is $\partial_\varepsilon u_\varepsilon(s, X_s)$ at $\varepsilon = 0$, because $\partial_\varepsilon [F(y) - F(\mathcal{R}_{-\varepsilon\kappa}y)] = \kappa(KF)(\mathcal{R}_{-\varepsilon\kappa}y)$. We claim that $(\Lambda_\varepsilon - 1)/\varepsilon \rightarrow M$ in $L^1(\mathbb{P})$ as $\varepsilon \rightarrow 0$. We write $\Lambda_\varepsilon = \exp(N_\varepsilon - \frac{1}{2}\langle N_\varepsilon \rangle)$, where N_ε is the stochastic integral in (11.18). By (11.17), we have $\langle N_\varepsilon \rangle \leq \varepsilon^2 V_1$ with the deterministic constant $V_1 = \frac{H}{2} \int_I (|\kappa'| + 2\Delta|\kappa|)^2 ds$. The second derivative of u_ε in ε is bounded by $4\Delta\kappa^2\sqrt{H}$ on D , by (11.15). The integrand of $N_\varepsilon - \varepsilon M$ is therefore bounded by $2\varepsilon^2\Delta\kappa(s)^2\sqrt{H}$ before τ , and $\mathbb{E}(N_\varepsilon - \varepsilon M)^2 = O(\varepsilon^4)$, since κ is bounded on I . Let $x_\varepsilon = N_\varepsilon - \frac{1}{2}\langle N_\varepsilon \rangle$. By (11.14), we have $\mathbb{E}e^{\pm 2N_\varepsilon} \leq e^{2\varepsilon^2 V_1}$ and $\mathbb{E}N_\varepsilon^4 \leq 16\varepsilon^4 V_1^2$, and therefore $\mathbb{E}x_\varepsilon^4 = O(\varepsilon^4)$. Using also $|e^x - 1 - x| \leq \frac{1}{2}x^2 e^{|x|}$ and the Cauchy–Schwarz inequality, we obtain $\mathbb{E}|e^{x_\varepsilon} - 1 - x_\varepsilon| = O(\varepsilon^2)$. Since $x_\varepsilon - \varepsilon M = (N_\varepsilon - \varepsilon M) - \frac{1}{2}\langle N_\varepsilon \rangle$, we conclude that $\mathbb{E}|\Lambda_\varepsilon - 1 - \varepsilon M| = O(\varepsilon^2)$, as claimed.

Now let $\varphi \in C^\infty(S_N)$. The left side of (11.19) has derivative $\mathbb{E}[W(K\varphi)(X_t)]$ at $\varepsilon = 0$, by dominated convergence, and the claim shows that the right side has derivative $\mathbb{E}[WM\varphi(X_t)]$. Equating the two derivatives, we find

$$\int K\varphi d\rho^W = Z^{-1}\mathbb{E}[WM\varphi(X_t)]. \quad (11.21)$$

Step 3: identification of the score. Let $\mathbb{Q} = (W/Z)\mathbb{P}$. Since $\mathbb{E}[W|M] \leq \mathbb{E}|M| < \infty$, we may define $m(x) = \mathbb{E}_{\mathbb{Q}}[M | X_t = x]$. Then (11.21) reads $\int K\varphi d\rho^W = \int m\varphi d\rho^W$. The law ρ^W is absolutely continuous with respect to ν_N : it is bounded by Z^{-1} times the law of X_t , which has a density by the second part of Lemma 11.4 applied with $T = 0$, $\Psi = 1$, and $D = S_N$. We conclude that $-m$ is a weak K -score of ρ^W , provided that it lies in $L^1(\rho^W)$, which we show in the next step.

Step 4: moments. For the quadratic potential, we have

$$KF = 2(a_j - a_i)x_i x_j, \quad |\nabla_{S_N}(KF)|^2 \leq |\nabla(KF)|^2 = 4\Delta^2(x_i^2 + x_j^2). \quad (11.22)$$

Together with $|\Omega X_s| = (X_{s,i}^2 + X_{s,j}^2)^{1/2}$ and (11.10), this bounds the integrand in (11.20) by $\sqrt{H}(|\kappa'(s)| + 2\Delta|\kappa(s)|)$ before τ . The quadratic variation of M therefore satisfies the deterministic bound

$$\langle M \rangle \leq \frac{H}{2} \int_I (|\kappa'| + 2\Delta|\kappa|)^2 ds \leq 4H \int_I (\kappa'(s)^2 + \Delta^2\kappa(s)^2) ds = V_*. \quad (11.23)$$

By (11.14), we have $\mathbb{E}|M|^{2p} \leq 2p!(2V_*)^p$. Applying Jensen's inequality for conditional expectations, we see that

$$\mathbb{E}_{\rho^W}|m|^{2p} \leq \mathbb{E}_{\mathbb{Q}}|M|^{2p} = Z^{-1}\mathbb{E}[|W|M|^{2p}] \leq Z^{-1}\mathbb{E}|M|^{2p} \leq 2p!Z^{-1}(2V_*)^p,$$

which yields (11.12) with $C_p = 2 \cdot 8^p p!$. For the linear ramp, $\int_I \kappa'^2 = \ell^{-1}$ and $\int_I \kappa^2 = \ell/3$, and (11.13) follows. $\square$

No derivative of W is taken in the proof. For this reason, the path weight may be any measurable functional satisfying (11.11). We now specialize to the low-temperature time scale. Recall that the first coordinate is singled out by $a_1 = 0$, and that in the applications the parameter θ below is the top fraction θ_N of (3.14).

Corollary 11.6 (Scores on the time scale a_2^{-1}). *Let $J \geq 2$, $C_J \geq 1$, $L > 0$, $z_0 > 0$, and $\theta > 0$. Suppose that $0 = a_1 < a = a_2 \leq a_3 \leq \dots \leq a_J \leq C_J a$, and set $m = \theta N a$. Let $D \subset S_N$ be an open set that is invariant under all rotations of the coordinates $x_1, \dots, x_J$ and contained in*

$$\left\{ x \in S_N : \sum_{k=1}^J x_k^2 < L\theta N \right\}. \quad (11.24)$$

Let $t \geq 1/a$, and let W be a path weight for the final interval $I = [t - 1/a, t]$ and the set D , as in (11.11), which is invariant under the rotations Θ_ε in the $(1, j)$ -plane with the linear ramp, for every $2 \leq j \leq J$ and $\varepsilon \in \mathbb{R}$. Suppose that $Z = \mathbb{E}W \geq z_0$. Then, for every $2 \leq j \leq J$ and every integer $p \geq 1$,

$$\mathbb{E}_{\rho^W} \left| K_{1j} \log \frac{d\rho^W}{dv_N} \right|^{2p} \leq C m^p, \quad (11.25)$$

where C depends only on p , z_0 , L , and C_J .

Proof. We use Theorem 11.5 with (i, j) replaced by $(1, j)$, $\ell = 1/a$, and $H = L\theta N$. The set D is invariant under the rotations of the $(1, j)$ -plane, and (11.10) holds because $x_1^2 + x_j^2 \leq \sum_{k \leq j} x_k^2$. With $\Delta = a_j \leq C_J a$, the bound (11.13) becomes

$$\mathbb{E}_{\rho^W} \left| K_{1j} \log \frac{d\rho^W}{dv_N} \right|^{2p} \leq \frac{C_p}{z_0} [L\theta N(a + C_J^2 a)]^p = \frac{C_p}{z_0} [L(1 + C_J^2)m]^p.$$

This is (11.25). □

A weight that depends on the path only through its values before $t - 1/a$, multiplied by the indicator that the path stays in D from some time $T \leq t - 1/a$ until t , satisfies the hypotheses: the rotations act only on I and preserve D .

11.3 Weak convergence to total variation convergence by Gaussian smoothing

In this section $d \geq 1$ is a fixed integer and densities are taken with respect to Lebesgue measure on $\mathbb{R}^d$. For a probability density p on $\mathbb{R}^d$ that belongs to $W_{\text{loc}}^{1,1}(\mathbb{R}^d)$, its Fisher information is

$$\text{FI}(p) = \int_{\mathbb{R}^d} \frac{|\nabla p|^2}{p} dz,$$

where the integrand is 0 where $p = 0$ and $\nabla p = 0$, and $+\infty$ where $p = 0$ and $\nabla p \neq 0$. For other densities, $\text{FI}(p) = \infty$. We write $\text{FI}(\mu) = \text{FI}(p)$ for a probability measure μ with density p , and $\text{FI}(\mu) = \infty$ if μ has no density. If $\text{FI}(p) < \infty$, then by the Cauchy–Schwarz inequality, we have

$$\int_{\mathbb{R}^d} |\nabla p| dz \leq \left(\int_{\{p>0\}} \frac{|\nabla p|^2}{p} dz \right)^{1/2} \left(\int p dz \right)^{1/2} = \text{FI}(p)^{1/2}, \quad (11.26)$$

so $p \in W^{1,1}(\mathbb{R}^d)$. The first lemma is the classical Fisher information bound for a Gaussian convolution, related to the convolution inequality of Stam [138]. Its proof uses the conditional-score and Cauchy–Schwarz argument of Blachman [33]. We include it because the input law is arbitrary and may, for instance, have no moments.

Lemma 11.7 (Fisher information of a Gaussian convolution). *For every probability measure λ on $\mathbb{R}^d$, the convolution $\mu = \lambda * \mathcal{N}(0, I_d)$ has a smooth positive density p with $\text{FI}(p) \leq d$.*

Proof. Let ϕ_d be the standard Gaussian density on $\mathbb{R}^d$. The density of μ and its gradient are

$$p(y) = \int \phi_d(y - v) \lambda(dv) > 0, \quad \nabla p(y) = - \int (y - v) \phi_d(y - v) \lambda(dv).$$

Every derivative of ϕ_d is bounded, and dominated convergence justifies differentiation under the integral. In particular, p is smooth. Applying the Cauchy–Schwarz inequality with respect to the finite measure $\phi_d(y - v) \lambda(dv)$, we find

$$\frac{|\nabla p(y)|^2}{p(y)} \leq \int |y - v|^2 \phi_d(y - v) \lambda(dv).$$

We integrate in y and use Tonelli’s theorem and $\int |y - v|^2 \phi_d(y - v) dy = d$ for every v . This gives $\text{FI}(p) \leq d$. $\square$

For probability measures μ, μ' on $\mathbb{R}^d$, let

$$d_{\text{BL}}(\mu, \mu') = \sup \left\{ \left| \int \varphi d\mu - \int \varphi d\mu' \right| : \|\varphi\|_\infty \leq 1, \text{Lip}(\varphi) \leq 1 \right\} \quad (11.27)$$

be the bounded Lipschitz distance. The next theorem shows that weak convergence implies convergence in total variation when nearby laws have bounded Fisher information. We smooth with a Gaussian kernel. The Fisher information is required only of approximating laws that are close in total variation. In the applications, these are obtained from the laws of interest by path weights and cutoff functions.

Theorem 11.8 (From weak convergence to total variation). *Let $\mu_N, \mu, \hat{\mu}_N$ be probability measures on $\mathbb{R}^d$. Suppose that $\text{FI}(\mu) \leq I_*$, $\text{FI}(\hat{\mu}_N) \leq C$, and $\|\mu_N - \hat{\mu}_N\|_{\text{TV}} \leq \delta_N$. Then for every $0 < \varepsilon \leq 1$,*

$$\|\mu_N - \mu\|_{\text{TV}} \leq 2\delta_N + C_d \varepsilon (\sqrt{C} + \sqrt{I_*}) + (1 + C_d/\varepsilon) d_{\text{BL}}(\mu_N, \mu), \quad (11.28)$$

with $C_d = \sqrt{d}$. In particular, suppose that $\mu_N \Rightarrow \mu$, that $\text{FI}(\mu) < \infty$, and that for every $\delta > 0$ there exist $C_\delta < \infty$ and probability measures $\hat{\mu}_N^\delta$ with $\text{FI}(\hat{\mu}_N^\delta) \leq C_\delta$ for all large N and $\limsup_N \|\mu_N - \hat{\mu}_N^\delta\|_{\text{TV}} \leq \delta$. Then $\|\mu_N - \mu\|_{\text{TV}} \rightarrow 0$.

Proof. We begin with two estimates for a density q with $\text{FI}(q) < \infty$. Note that $q \in W^{1,1}(\mathbb{R}^d)$, by (11.26). For $q \in C_c^\infty(\mathbb{R}^d)$, we see that $\|q(\cdot + h) - q\|_{L^1} \leq |h| \|\nabla q\|_{L^1}$ by the fundamental theorem of calculus along segments. Smooth compactly supported functions are dense in $W^{1,1}(\mathbb{R}^d)$ [37, Theorem 9.2, p. 265], and the same inequality holds for our q . Together with (11.26), this shows that

$$\|q(\cdot + h) - q\|_{L^1} \leq |h| \text{FI}(q)^{1/2} \quad (h \in \mathbb{R}^d). \quad (11.29)$$

Let ϕ_ε be the density of $\mathcal{N}(0, \varepsilon^2 I_d)$. For a Borel function $f : \mathbb{R}^d \rightarrow [0, 1]$, let $f_\varepsilon = f * \phi_\varepsilon$. Then $0 \leq f_\varepsilon \leq 1$, and $\text{Lip}(f_\varepsilon) \leq \int |\nabla \phi_\varepsilon| = \varepsilon^{-1} \mathbb{E}|g| \leq \sqrt{d}/\varepsilon$, where g is a standard Gaussian vector. The density ϕ_ε is even, so that $\int f_\varepsilon q = \int f(q * \phi_\varepsilon)$. Using (11.29), we deduce that

$$\left| \int (f - f_\varepsilon) q \right| \leq \int \phi_\varepsilon(h) \|q - q(\cdot - h)\|_{L^1} dh \leq \text{FI}(q)^{1/2} \int |h| \phi_\varepsilon(h) dh \leq \sqrt{d} \varepsilon \text{FI}(q)^{1/2}.$$

We now decompose

$$\mu_N(f) - \mu(f) = (\mu_N - \hat{\mu}_N)(f) + \hat{\mu}_N(f - f_\varepsilon) + (\hat{\mu}_N - \mu_N)(f_\varepsilon) + (\mu_N - \mu)(f_\varepsilon) + \mu(f_\varepsilon - f).$$

The first and third terms are at most δ_N in absolute value, by Lemma 3.6, because f and f_ε take values in $[0, 1]$. The second and fifth terms are at most $\sqrt{d} \varepsilon \sqrt{C}$ and $\sqrt{d} \varepsilon \sqrt{I_*}$. The function $f_\varepsilon / (1 + \sqrt{d}/\varepsilon)$ is admissible in (11.27), and the fourth term is at most $(1 + \sqrt{d}/\varepsilon) d_{\text{BL}}(\mu_N, \mu)$. Taking the supremum over f proves (11.28).

For the second statement, we first show that $d_{\text{BL}}(\mu_N, \mu) \rightarrow 0$. This is classical, but we include the argument. Fix $\kappa_0 > 0$, let R satisfy $\mu(|z| \leq R) \geq 1 - \kappa_0$, and let $\psi : \mathbb{R}^d \rightarrow [0, 1]$ be continuous, equal to one on $\{|z| \leq R\}$ and to zero outside $\{|z| \leq R + 1\}$. By weak convergence, $\mu_N(|z| \leq R + 1) \geq \mu_N(\psi) \geq 1 - 2\kappa_0$ for large N . The Arzelà–Ascoli theorem shows that the restrictions to $\{|z| \leq R + 1\}$ of the functions admissible in (11.27) have a finite κ_0 -net $\varphi_1, \dots, \varphi_n$ in the uniform norm, which we may take admissible. For an admissible φ and φ_k within κ_0 of it on $\{|z| \leq R + 1\}$, we have

$$|\mu_N(\varphi) - \mu(\varphi)| \leq |\mu_N(\varphi_k) - \mu(\varphi_k)| + 2\kappa_0 + 2(\mu_N(|z| > R + 1) + \mu(|z| > R + 1)).$$

This gives $\limsup_N d_{\text{BL}}(\mu_N, \mu) \leq 8\kappa_0$, and κ_0 is arbitrary. Now we fix $\delta > 0$ and use (11.28) with $\hat{\mu}_N = \hat{\mu}_N^\delta$. Letting $N \rightarrow \infty$ and then $\varepsilon \rightarrow 0$, we find $\limsup_N \|\mu_N - \mu\|_{\text{TV}} \leq 2\delta$. Since δ is arbitrary, this completes the proof. $\square$

In the second statement, the approximating laws $\hat{\mu}_N^\delta$ need not converge weakly, and their limit points need not be μ . Weak convergence is required only of μ_N . The order of the limits is: δ and the approximating laws are fixed, then $N \rightarrow \infty$, then $\varepsilon \rightarrow 0$, then $\delta \rightarrow 0$.

11.4 Fisher information of finitely many coordinates

We now relate scores along the rotations K_{1j} to derivatives in the coordinates $x_2, \dots, x_N$, viewed as coordinates on the hemisphere $\{x_1 > 0\}$.

The chart. For $x \in S_N$, let $y(x) = (x_2, \dots, x_N) \in \mathbb{R}^{N-1}$. Let $U_N = \{y \in \mathbb{R}^{N-1} : |y|^2 < N\}$, and for $y \in U_N$ let $x_1(y) = (N - |y|^2)^{1/2}$ and $x^+(y) = (x_1(y), y) \in S_N$. For a measure ρ on S_N , let ρ^+ be its image under $x \mapsto y(x)$. If ρ is invariant under R_1 , then ρ^+ determines ρ , and by the second part of Lemma 3.1, for two such probability measures ρ, ρ' ,

$$\|\rho - \rho'\|_{\text{TV}} = \|\rho^+ - \rho'^+\|_{\text{TV}}. \quad (11.30)$$

Indeed, ρ^+ is the image of the law of $(|x_1|, x_2, \dots, x_N)$ under a bijection. We write $|S_N|$ for the surface area of S_N and $\omega_N = \sqrt{N}/|S_N|$.

Lemma 11.9 (Densities and scores in the chart). *Let $\rho = f\nu_N$ be a probability measure on S_N that is invariant under R_1 .*

1. *The measure ρ^+ has the Lebesgue density $f^+(y) = 2\omega_N f(x^+(y))/x_1(y)$ on U_N , and $\rho^+(\mathbb{R}^{N-1} \setminus U_N) = 0$.*
2. *Suppose that ρ has a weak K_{1j} -score s_j for some $2 \leq j \leq N$, and write $s_j^+ = s_j \circ x^+$. Then the distributional derivative $\partial_{y_j} f^+$ belongs to $L^1_{\text{loc}}(U_N)$, and*

$$\partial_{y_j} f^+ = f^+ \left(\frac{s_j^+}{x_1} + \frac{y_j}{x_1^2} \right) \quad \text{almost everywhere on } U_N. \quad (11.31)$$

3. *If, in addition, $\rho(\{x_1^2 < r_0 N\}) = 0$ for some $r_0 \in (0, 1]$, then the extension of f^+ by zero has a weak derivative in y_j on $\mathbb{R}^{N-1}$ which belongs to $L^1(\mathbb{R}^{N-1})$ and is given by (11.31) on U_N and by zero outside U_N .*

The term y_j/x_1^2 in (11.31) comes from the surface element $x_1^{-1}dy$ of the chart, and it has a positive sign. In the situation of the second part, the score is odd under R_1 : since $K_{1j}(\varphi \circ R_1) = -(K_{1j}\varphi) \circ R_1$ and ρ is invariant, the function $-s_j \circ R_1$ also satisfies (11.4), so $s_j \circ R_1 = -s_j$ ρ -almost everywhere by uniqueness. In particular $|s_j|$ is invariant under R_1 , and $\int |s_j^+|^q f^+ dy = \mathbb{E}_\rho |s_j|^q$ for every $q \geq 1$.

Proof. The upper hemisphere is the graph of $y \mapsto x_1(y)$ over U_N . We have $\nabla x_1 = -y/x_1$, so its surface element is $(1 + |y|^2/x_1^2)^{1/2} dy = \sqrt{N} x_1^{-1} dy$, and we have $\int_{\{x_1 > 0\}} \varphi d\nu_N = \omega_N \int_{U_N} \varphi(x^+(y)) x_1(y)^{-1} dy$. The equator has ν_N -measure zero, and by the invariance of ρ we have $\rho^+(A) = 2\rho(\{y(x) \in A, x_1 > 0\})$. The first part follows.

For the second part, let $\psi \in C_c^\infty(U_N)$, and let $\varphi(x) = \psi(y(x))$ if $x_1 > 0$ and $\varphi(x) = 0$ otherwise. Since ψ vanishes near ∂U_N , the function φ vanishes near the equator and is

smooth on S_N , with $K_{1j}\varphi(x) = x_1(\partial_j\psi)(y(x))$ for $x_1 > 0$. Inserting φ in (11.4) and using the first part, we find

$$\int_{U_N} x_1 \partial_j \psi f^+ dy = - \int_{U_N} \psi s_j^+ f^+ dy.$$

In other words, the distributional derivative of $x_1 f^+$ in y_j is $s_j^+ f^+$, which lies in $L^1_{\text{loc}}(U_N)$ because $s_j \in L^1(\rho)$ and x_1^{-1} is locally bounded on U_N . The function x_1^{-1} is smooth on U_N with $\partial_j(x_1^{-1}) = y_j/x_1^3$, and by the product rule we have $\partial_j f^+ = \partial_j(x_1 f^+ \cdot x_1^{-1}) = s_j^+ f^+ / x_1 + f^+ y_j / x_1^2$.

For the third part, $f^+ = 0$ almost everywhere on $\{|y|^2 > (1 - r_0)N\}$. Let $\chi \in C_c^\infty(U_N)$ equal one on $\{|y|^2 \leq (1 - r_0)N\}$. For $\psi \in C_c^\infty(\mathbb{R}^{N-1})$, we have $\int \partial_j \psi f^+ = \int \partial_j(\psi \chi) f^+ - \int \psi \partial_j \chi f^+$, and the last integral vanishes because $\partial_j \chi = 0$ wherever $f^+ \neq 0$. The claim follows by applying the second part to $\psi \chi$. The derivative lies in L^1 because $x_1^2 \geq r_0 N$ on the support of f^+ and $s_j \in L^1(\rho)$. $\square$

We next record the score of a marginal. Let $a_2, \dots, a_N > 0$ and let G be the centered Gaussian product law on $\mathbb{R}^{N-1}$ whose coordinates y_i have variance $(2a_i)^{-1}$, $2 \leq i \leq N$. Its density is

$$\gamma(y) = \prod_{i=2}^N \sqrt{a_i/\pi} e^{-a_i y_i^2}. \quad (11.32)$$

Fix $2 \leq J \leq N-1$ and split $y = (z, v)$ with $z = (y_2, \dots, y_J)$ and $v = (y_{J+1}, \dots, y_N)$, the remaining coordinates. Then $G = G_z \otimes G_v$, the product of the corresponding Gaussian laws on $\mathbb{R}^{J-1}$ and $\mathbb{R}^{N-J}$. For a probability density w with respect to G , let $\bar{w}(z) = \int w(z, v) G_v(dv)$ be the density of its z -marginal with respect to G_z , and, for $\bar{w}(z) > 0$, let $\mathbb{E}_w[\cdot | z] = \bar{w}(z)^{-1} \int \cdot w(z, v) G_v(dv)$ be the conditional expectation given z .

Lemma 11.10 (Marginal score). *Let $\hat{\rho}$ be a probability measure on S_N , invariant under R_1 , with $\hat{\rho}(\{x_1^2 < r_0 N\}) = 0$ for some $r_0 \in (0, 1]$ and with weak K_{1j} -scores ϑ_j for $2 \leq j \leq J$. Let $w = d\hat{\rho}^+ / dG$. Then $\bar{w} \in W_{\text{loc}}^{1,1}(\mathbb{R}^{J-1})$ and, almost everywhere on $\{\bar{w} > 0\}$, for $2 \leq j \leq J$,*

$$\partial_{z_j} \log \bar{w} = \mathbb{E}_w \left[\frac{\vartheta_j \circ x^+}{x_1} + \frac{z_j}{x_1^2} \mid z \right] + 2a_j z_j. \quad (11.33)$$

In particular, with $h = \log \bar{w}$,

$$z_j \partial_j h = \mathbb{E}_w \left[\frac{z_j}{x_1} \vartheta_j \circ x^+ + \frac{z_j^2}{x_1^2} \mid z \right] + 2a_j z_j^2. \quad (11.34)$$

Proof. Let $f^+ = w\gamma$ be the Lebesgue density of $\hat{\rho}^+$, and let $p(z) = \int f^+(z, v) dv$. The third part of Lemma 11.9 implies that $\partial_j f^+$ exists in $L^1(\mathbb{R}^{N-1})$ and is given by (11.31). For $\psi \in C_c^\infty(\mathbb{R}^{J-1})$, we use the definition of the weak derivative with the test function $\psi(z)\chi_n(v)$,

where $\chi_n \in C_c^\infty(\mathbb{R}^{N-J})$ take values in $[0, 1]$ and increase to one, and let $n \rightarrow \infty$ by dominated convergence. This shows that $p \in W^{1,1}(\mathbb{R}^{J-1})$ with $\partial_j p(z) = \int \partial_j f^+(z, v) dv$, that is, $\partial_j p = p \mathbb{E}_w[\vartheta_j \circ x^+ / x_1 + z_j / x_1^2 \mid z]$ on $\{p > 0\}$. Here the conditional law of v given z is the same under $f^+(z, v) dv / p(z)$ and under $w(z, v) G_v(dv) / \bar{w}(z)$. Finally $\bar{w} = p / \gamma_z$, where $\gamma_z(z) = \prod_{j=2}^J \sqrt{a_j / \pi} e^{-a_j z_j^2}$ is smooth and positive, and $\partial_j \log \gamma_z = -2a_j z_j$. We obtain (11.33), and (11.34) follows on multiplying by z_j . $\square$

Standardized coordinates. For the rest of this section we assume

$$0 = a_1 < a = a_2 \leq a_3 \leq \cdots \leq a_J \leq C_J a, \quad \theta > 0, \quad m = \theta N a, \quad (11.35)$$

for a fixed integer $J \geq 2$ and a constant $C_J \geq 1$; the coefficients a_k with $k > J$ are arbitrary real numbers. In the applications, θ is the top fraction θ_N and $m = m_N$, as in (3.14). We write

$$\mathbf{z}(x) = (\sqrt{2a_2} x_2, \dots, \sqrt{2a_J} x_J) \in \mathbb{R}^{J-1} \quad (11.36)$$

for the standardized coordinates. This scaling sends $\otimes_{j=2}^J \mathcal{N}(0, (2a_j)^{-1})$ to $\mathcal{N}(0, I_{J-1})$.

Lemma 11.11 (Fisher information of the standardized coordinates). *Assume (11.35) with $m \geq 1$, and let $s, K_0, C_0 > 0$. Let $\hat{\rho}$ be a probability measure on S_N that is invariant under R_1 and satisfies*

$$\hat{\rho}(\{x_1^2 < s\theta N\} \cup \{|\mathbf{z}(x)| > K_0\}) = 0. \quad (11.37)$$

Suppose that $\hat{\rho}$ has weak K_1 -scores ϑ_j with

$$\mathbb{E}_{\hat{\rho}}|\vartheta_j|^2 \leq C_0 m, \quad 2 \leq j \leq J. \quad (11.38)$$

Then the law of $\mathbf{z}(x)$ under $\hat{\rho}$ has a density p with

$$\text{FI}(p) \leq \frac{(J-1)C_0}{s} + \frac{K_0^2}{2s^2}, \quad \text{supp } p \subset \{|\mathbf{z}| \leq K_0\}. \quad (11.39)$$

Proof. Let f^+ be the Lebesgue density of $\hat{\rho}^+$, and change variables to $(\mathbf{z}, v)$ with $\mathbf{z}_j = \sqrt{2a_j} y_j$ for $2 \leq j \leq J$. The joint density of $(\mathbf{z}, v)$ is $Q(\mathbf{z}, v) = f^+(y) \prod_{j=2}^J (2a_j)^{-1/2}$. We have $\partial_{\mathbf{z}_j} = (2a_j)^{-1/2} \partial_{y_j}$ and $y_j = \mathbf{z}_j / \sqrt{2a_j}$, and by the third part of Lemma 11.9,

$$\partial_{\mathbf{z}_j} \log Q = \frac{\vartheta_j \circ x^+}{\sqrt{2a_j} x_1} + \frac{\mathbf{z}_j}{2a_j x_1^2} \quad (11.40)$$

almost everywhere on $\{Q > 0\}$, and $\partial_{\mathbf{z}_j} Q = 0$ almost everywhere on $\{Q = 0\}$. On the support of $\hat{\rho}$, we have $a_j x_1^2 \geq a s \theta N = sm$, because $a_j \geq a$. Using $(\alpha + \beta)^2 \leq 2\alpha^2 + 2\beta^2$, (11.38), $|\mathbf{z}| \leq K_0$, and $m \geq 1$, we have

$$\sum_{j=2}^J \int Q |\partial_{\mathbf{z}_j} \log Q|^2 d\mathbf{z} dv \leq \sum_{j=2}^J \left(\frac{\mathbb{E}_{\hat{\rho}}|\vartheta_j|^2}{sm} + \frac{\mathbb{E}_{\hat{\rho}}\mathbf{z}_j^2}{2s^2 m^2} \right) \leq \frac{(J-1)C_0}{s} + \frac{K_0^2}{2s^2}. \quad (11.41)$$

Let $p(\mathbf{z}) = \int Q(\mathbf{z}, v) dv$. As in the proof of Lemma 11.10, $\partial_j p(\mathbf{z}) = \int \partial_{\mathbf{z}_j} Q(\mathbf{z}, v) dv$. By the Cauchy–Schwarz inequality on each fiber, we have $|\partial_j p|^2 \leq p \int Q |\partial_{\mathbf{z}_j} \log Q|^2 dv$, and in particular $\partial_j p = 0$ where $p = 0$. Dividing by p , integrating in $\mathbf{z}$, and summing over j , we deduce (11.39) from (11.41). $\square$

We can now upgrade weak convergence of the standardized coordinates to convergence in total variation. A path weight in the sense of (11.11) is written $W = \Phi(\xi, X|_I)$, and for $2 \leq j \leq J$ we write $\Theta_\varepsilon^{(1j)}$ for the path rotations in the $(1, j)$ -plane with the linear ramp on I .

Corollary 11.12 (Total variation convergence of finitely many coordinates). *For each N , let $a_1, \dots, a_N$ and $\theta = \theta_N$ satisfy (11.35), with J and C_J independent of N and $m = \theta N a \rightarrow \infty$. Let X solve (11.2) with $X_0 \sim \nu_N$, let $t_N \geq 1/a$ be deterministic, and let μ_N be the law of $\mathbf{z}(X_{t_N})$. Suppose that $\mu_N \Rightarrow \mu$, where μ is a probability measure on $\mathbb{R}^{J-1}$ with $\text{FI}(\mu) < \infty$. Suppose further that for every $\delta \in (0, 1)$ there exist $s, L, B > 0$, independent of N , and for every N a random variable W_N and a function η_N^0 , with the following properties.*

1. $W_N = \Phi_N(\xi_N, X|_{I_N})$ is a path weight for the final interval $I_N = [t_N - 1/a, t_N]$ and an open set D_N , as in (11.11). The set D_N is invariant under all rotations of $x_1, \dots, x_J$ and contained in $\{\sum_{k \leq J} x_k^2 < L\theta N\}$. For every $2 \leq j \leq J$ and $\varepsilon \in \mathbb{R}$, the function Φ_N is invariant under $\Theta_\varepsilon^{(1j)}$. The measure $A \mapsto \mathbb{E}[W_N \mathbf{1}_{\{X_{t_N} \in A\}}]$ is invariant under R_1 .
2. $\eta_N^0 \in C^1(S_N)$ takes values in $[0, 1]$, is invariant under R_1 , vanishes on $\{x_1^2 < s\theta N\}$, and satisfies $|K_{1j}\eta_N^0| \leq B(\eta_N^0)^{1/2}$ for $2 \leq j \leq J$.
3. $\liminf_{N \rightarrow \infty} \mathbb{E}[W_N \eta_N^0(X_{t_N})] \geq 1 - \delta$.

Then $\|\mu_N - \mu\|_{\text{TV}} \rightarrow 0$.

Proof. Fix $\delta \in (0, 1/6)$ and the corresponding s, L, B, W_N , and η_N^0 . We construct approximating laws for Theorem 11.8.

A cutoff in the standardized coordinates. Let $\psi \in C^\infty(\mathbb{R}; [0, 1])$ equal one on $(-\infty, 1]$ and zero on $[2, \infty)$, and for $K_0 > 0$ let $\psi_{K_0}(\mathbf{z}) = \psi(|\mathbf{z}|/K_0)^2$. Then $|\nabla \psi_{K_0}| \leq (2\|\psi'\|_\infty/K_0)\psi_{K_0}^{1/2}$. We fix K_0 such that $\mu(|\mathbf{z}| \geq K_0) \leq \delta$. The function $1 - \psi_{K_0}$ is bounded, continuous, and vanishes on $\{|\mathbf{z}| \leq K_0\}$, and weak convergence implies that $\limsup_N \mathbb{E}[1 - \psi_{K_0}(\mathbf{z}(X_{t_N}))] \leq \delta$. Let $\eta_N = \eta_N^0 \psi_{K_0} \circ \mathbf{z}$ and $U_N = W_N \eta_N(X_{t_N})$. From the third hypothesis, we conclude that

$$\liminf_N \mathbb{E} U_N \geq \liminf_N \mathbb{E}[W_N \eta_N^0(X_{t_N})] - \limsup_N \mathbb{E}[1 - \psi_{K_0}(\mathbf{z}(X_{t_N}))] \geq 1 - 2\delta. \quad (11.42)$$

Scores of the weighted endpoint law. Let $\rho_N = \rho^{W_N}$ be the endpoint law weighted by W_N , with normalizing mass $\mathbb{E} W_N$. We have $\mathbb{E} W_N \geq \mathbb{E} U_N \geq 1/2$ for large N , and Corollary 11.6 with $p = 1$ and $z_0 = 1/2$ gives $\mathbb{E}_{\rho_N} |K_{1j} \log(dp_N/d\nu_N)|^2 \leq Cm$ for $2 \leq j \leq J$,

with C depending only on L and C_J . The law ρ_N is concentrated on D_N , where $x_1^2 \leq L\theta N$. We have $K_{1j}\mathbf{z} = \sqrt{2a_j}x_1e_j$ and $a_j \leq C_Ja$, so that $|K_{1j}\mathbf{z}| \leq (2C_JLm)^{1/2}$ on D_N . Using the product rule, we find that, ρ_N -almost everywhere,

$$|K_{1j}\eta_N| \leq B(\eta_N^0)^{1/2}\psi_{K_0} + \eta_N^0 \frac{2\|\psi'\|_\infty}{K_0} \psi_{K_0}^{1/2}(2C_JLm)^{1/2} \leq B'\sqrt{m}\eta_N^{1/2} \quad (11.43)$$

for $m \geq 1$, where B' depends on B , K_0 , C_J , and L . Here we used $\psi_{K_0} \leq \psi_{K_0}^{1/2}$ and $\eta_N^0 \leq (\eta_N^0)^{1/2}$. The normalizer of $\eta_N\rho_N$ is $\int \eta_N d\rho_N = \mathbb{E}U_N/\mathbb{E}W_N \geq \mathbb{E}U_N \geq 1/2$ for large N . Lemma 11.2 with $p = 1$ implies that the normalized cutoff-weighted law $\hat{\rho}_N = \eta_N\rho_N / \int \eta_N d\rho_N$ has weak K_{1j} -scores with second moments at most $C'm$, where C' depends on δ through B, K_0, L . It is invariant under R_1 , by the first two hypotheses and because $\mathbf{z}$ does not involve x_1 , and it satisfies (11.37) with s and $2K_0$ in place of K_0 . Using Lemma 11.11, we find that the law $\hat{\mu}_N$ of $\mathbf{z}$ under $\hat{\rho}_N$ satisfies $\text{FI}(\hat{\mu}_N) \leq C_\delta$ for large N , with C_δ independent of N .

Conclusion. The law $\hat{\mu}_N$ is the law of $\mathbf{z}(X_{t_N})$ under $(U_N/\mathbb{E}U_N)\mathbb{P}$. Combining Lemma 11.3 and (11.42), we find $\limsup_N \|\mu_N - \hat{\mu}_N\|_{\text{TV}} \leq \limsup_N (1 - \mathbb{E}U_N) \leq 2\delta$. The second statement of Theorem 11.8, applied with 2δ in place of δ , completes the proof. $\square$

Remark 11.13 (Gaussian convolutions as limits). By Lemma 11.7, every limit of the form $\mu = \lambda * \mathcal{N}(0, I_{J-1})$ satisfies $\text{FI}(\mu) \leq J - 1$, for an arbitrary probability measure λ . In the low-temperature sections, μ is the law of the sum of a standard Gaussian vector and an independent multiple of a standard Cauchy variable in the first coordinate. Its Fisher information is finite although the Cauchy variable has no moments.

11.5 Conditional relative entropy of the remaining coordinates

In this section, $a_1 = 0 < a_i$ for $2 \leq i \leq N$, so that $F(x) = \sum_{i \geq 2} a_i x_i^2$ and the Gaussian product law G of (11.32) is defined. We fix $2 \leq J \leq N - 1$ and use the splitting $y = (z, v)$ of Section 11.4. We study the relative entropy of the remaining coordinates v given z , with respect to their Gaussian law G_v . In Section 11.5, we write the generator in the chart and prove a logarithmic Sobolev inequality for this conditional relative entropy. Section 11.5 contains moment bounds for logarithms of densities. Section 11.5 computes the time derivative of the conditional relative entropy along the diffusion, and Section 11.5 bounds it under a cutoff function adapted to the approach to the critical temperature.

The generator in the chart and a logarithmic Sobolev inequality

For $y \in U_N$, let

$$\mathbf{g}(y) = I - \frac{yy^\top}{N}, \quad \Gamma(\varphi, \psi) = \nabla \varphi^\top \mathbf{g} \nabla \psi, \quad \Gamma(\varphi) = \Gamma(\varphi, \varphi).$$

The matrix $\mathbf{g}(y)$ is the inverse of the metric $I + yy^\top/x_1^2$ of S_N in the chart, with eigenvalues 1 and x_1^2/N . If φ is a function on S_N with $\varphi(x) = \psi(y(x))$ for $x_1 > 0$, then $|\nabla_{S_N} \varphi|^2(x^+(y)) = \Gamma(\psi)(y)$, so Γ is the carré du champ of L in the chart. For a smooth vector field V on U_N , let $\operatorname{div}_G V = \gamma^{-1} \operatorname{div}(\gamma V)$, and let $L_G \psi = \operatorname{div}_G(\mathbf{g} \nabla \psi)$. By the divergence theorem, for $\varphi, \psi \in C^2(U_N)$, one of which has compact support in U_N ,

$$\int \varphi L_G \psi dG = - \int \Gamma(\varphi, \psi) dG. \quad (11.44)$$

Lemma 11.14 (The generator in the chart). *For $\psi \in C^2(\mathbb{R}^{N-1})$, let $\varphi = \psi \circ y$ on S_N . Then $L\varphi = (L\psi) \circ y$ on $\{x_1 > 0\}$, where*

$$L\psi = \operatorname{div}_G(\mathbf{g} \nabla \psi) + \frac{y}{N} \cdot \nabla \psi. \quad (11.45)$$

Proof. Using (11.2), we see that the generator acts on $\varphi = \psi(x_2, \dots, x_N)$ as $\sum_{k,l \geq 2} \Pi(x)_{kl} \partial_k \partial_l \psi + \sum_{k \geq 2} \mathbf{b}_k(x) \partial_k \psi$. For $k, l \geq 2$ we have $\Pi(x)_{kl} = \mathbf{g}(y)_{kl}$. Using $\nabla F = 2(a_k x_k)_k$ and $x \cdot \nabla F = 2F$, we find $\nabla_{S_N} F = \nabla F - (2F/N)x$ and $\mathbf{b}_k = [-\frac{N-1}{N} - 2a_k + \frac{2F}{N}]y_k$ for $k \geq 2$. On the other hand, $L_G \psi = \mathbf{g} : \nabla^2 \psi + (\operatorname{div} \mathbf{g} + \mathbf{g} \nabla \log \gamma) \cdot \nabla \psi$. In dimension $N-1$, we have $(\operatorname{div} \mathbf{g})_k = -N^{-1} \sum_l \partial_l (y_k y_l) = -y_k$, and $\mathbf{g} \nabla \log \gamma = -2 \operatorname{diag}(a)y + (2F/N)y$ because $y \cdot \operatorname{diag}(a)y = F$. Then the drift of L_G is $[-1 - 2a_k + 2F/N]y_k$, which differs from $\mathbf{b}_k$ by y_k/N . This proves (11.45). In the uses below, a test in $C_c^\infty(D^+)$ is extended by zero to $\mathbb{R}^{N-1}$, so it belongs to this class; invariance under R_1 transfers the resulting integral identity to the folded measure. $\square$

The coefficient one in front of $(y/N) \cdot \nabla \psi$ reflects the dimension $N-1$ of the chart. It produces the term 1 in the factor $1 + S$ in (11.57) below.

Conditional relative entropy. Let w be a probability density with respect to G . With $\bar{w}$ and $\mathbb{E}_w[\cdot | z]$ as in Section 11.4, let $h = \log \bar{w}$ on $\{\bar{w} > 0\}$ and $g = \log w - h$ on $\{w > 0\}$. We write $\mathbb{E}_w$ for expectations under the probability measure wG , and

$$\mathcal{H}(w) = \int w g dG, \quad \mathcal{D}(w) = \int w \Gamma(g) dG, \quad (11.46)$$

where $\mathcal{D}(w)$ is defined when $w \in C^\infty(\mathbb{R}^{N-1})$ has compact support in U_N . Then $\mathcal{H}(w) = \operatorname{KL}(wG \| (\bar{w}G_z) \otimes G_v) \in [0, \infty]$ is the relative entropy of the law wG with respect to the product of its own z -marginal and the Gaussian law of the remaining coordinates.

Equivalently, it is the average over z of the relative entropy of the conditional law of v given z with respect to G_v . We call it the *conditional relative entropy of the remaining coordinates* with respect to G_v . If $\text{KL}(wG\|G) < \infty$, then

$$\text{KL}(wG\|G) = \text{KL}(\bar{w}G_z\|G_z) + \mathcal{H}(w). \quad (11.47)$$

This is the chain rule for relative entropy (see [46, Theorem 2.5.3 and Section 8.5]), and we include its short proof. By Jensen's inequality on each fiber, applied to the convex function $x \log x$, we have $\int \bar{w} \log \bar{w} dG_z \leq \int w \log w dG < \infty$; since $x \log x \geq -1/e$, both $w \log w$ and $\bar{w} \log \bar{w}$ are integrable. On $\{w > 0\}$, we have $\log w = \log \bar{w} + g$, and Fubini's theorem gives $\int w \log \bar{w} dG = \int \bar{w} \log \bar{w} dG_z$. Hence $\int w \log w dG = \int \bar{w} \log \bar{w} dG_z + \int w g dG$, which is (11.47). The quantity $\mathcal{D}(w)$ uses the full metric $\mathfrak{g}$ applied to the gradient of g in all variables. The next lemma separates its part in the v variables.

Lemma 11.15 (The metric of the remaining coordinates). *Let $y = (z, v) \in U_N$ and $x_1^2 = N - |z|^2 - |v|^2$.*

1. *Let $A_v = I - vv^\top / (N - |z|^2)$. For $\pi = (\pi_z, \pi_v) \in \mathbb{R}^{J-1} \times \mathbb{R}^{N-J}$, let $\pi'_z = \pi_z - z(v \cdot \pi_v) / (N - |z|^2)$. Then*

$$\pi^\top \mathfrak{g} \pi = \pi_z'^\top \left(I - \frac{zz^\top}{N} \right) \pi_z' + \pi_v^\top A_v \pi_v. \quad (11.48)$$

In particular $\Gamma(\varphi) \geq \nabla_v \varphi^\top A_v \nabla_v \varphi$. Moreover A_v is positive definite and

$$v^\top A_v^{-1} v = \frac{(N - |z|^2) |v|^2}{x_1^2}.$$

2. *Let $b > 0$ with $b \leq a_i$ for all $i > J$, and suppose that $b \sum_{i>J} v_i^2 / a_i \leq d x_1^2$ for some $d > 0$. Then*

$$A_v \geq \frac{b}{1+d} \text{diag}(a_i^{-1})_{i>J}. \quad (11.49)$$

Proof. We write $\mathfrak{g}$ in blocks: $\mathfrak{g}_{zz} = I - zz^\top / N$, $\mathfrak{g}_{zv} = -zv^\top / N$, and $\mathfrak{g}_{vv} = I - vv^\top / N$. By the rank-one inverse formula of Bartlett [19], usually called the Sherman–Morrison formula [131], $\mathfrak{g}_{zz}^{-1} = I + zz^\top / (N - |z|^2)$, and a direct computation shows that $\mathfrak{g}_{zz}^{-1} \mathfrak{g}_{zv} \pi_v = -z(v \cdot \pi_v) / (N - |z|^2)$ and

$$\mathfrak{g}_{vv} - \mathfrak{g}_{zv}^\top \mathfrak{g}_{zz}^{-1} \mathfrak{g}_{zv} = I - \frac{vv^\top}{N} - \frac{|z|^2 vv^\top}{N(N - |z|^2)} = A_v.$$

The identity in (11.48) is the completion of the square $\pi^\top \mathfrak{g} \pi = (\pi_z + \mathfrak{g}_{zz}^{-1} \mathfrak{g}_{zv} \pi_v)^\top \mathfrak{g}_{zz} (\pi_z + \mathfrak{g}_{zz}^{-1} \mathfrak{g}_{zv} \pi_v) + \pi_v^\top A_v \pi_v$, and the first term is nonnegative. Using the Sherman–Morrison formula again, we find $A_v^{-1} = I + vv^\top / x_1^2$ and $v^\top A_v^{-1} v = |v|^2 + |v|^4 / x_1^2 = |v|^2 (N - |z|^2) / x_1^2$.

For the second part, let $D_0 = \text{diag}(1 - b/((1+d)a_i))_{i>J}$, a positive definite diagonal matrix. For a positive definite D_0 and $s_0 > 0$, the matrix $D_0 - vv^\top/s_0$ is positive semidefinite if and only if $v^\top D_0^{-1}v \leq s_0$, as one sees by conjugating with $D_0^{-1/2}$. Since $b \leq a_i$, we have $1/(1 - b/((1+d)a_i)) = 1 + b/((1+d)a_i - b) \leq 1 + b/(da_i)$, and therefore

$$v^\top D_0^{-1}v \leq |v|^2 + \frac{b}{d} \sum_{i>J} \frac{v_i^2}{a_i} \leq |v|^2 + x_1^2 = N - |z|^2.$$

We conclude that $D_0 \geq vv^\top/(N - |z|^2)$, which is equivalent to (11.49). $\square$

The Gaussian logarithmic Sobolev inequality is due to Gross [79]. We use it in the following form: if G is the standard Gaussian measure on $\mathbb{R}^n$ and $\phi \geq 0$ is smooth with $0 < \int \phi dG < \infty$, then

$$\int \phi \log \frac{\phi}{\int \phi dG} dG \leq \frac{1}{2} \int \frac{|\nabla \phi|^2}{\phi} dG. \quad (11.50)$$

Here $0 \log 0 = 0$, and $|\nabla \phi|^2/\phi$ is set to zero on $\{\phi = 0\}$, where $\nabla \phi = 0$ because ϕ is smooth and nonnegative; the energy integral may equal $+\infty$. The identically zero function has entropy and energy zero. See [76, Theorem 2.7, (9), p. 5]. Applying (11.50) to $\phi(\Sigma^{1/2} \cdot)$ with $\Sigma = \text{diag}((2a_i)^{-1})_{i>J}$, we obtain, for smooth $\phi \geq 0$ on $\mathbb{R}^{N-J}$ with $\int \phi dG_v = 1$,

$$\int \phi \log \phi dG_v \leq \sum_{i>J} \frac{1}{4a_i} \int \frac{(\partial_{v_i} \phi)^2}{\phi} dG_v. \quad (11.51)$$

Proposition 11.16 (A logarithmic Sobolev inequality for the conditional relative entropy). *Let $w \in C^\infty(\mathbb{R}^{N-1})$ be a nonnegative probability density with respect to G , with compact support in U_N , and let b, d be as in the second part of Lemma 11.15. Suppose that $b \sum_{i>J} v_i^2/a_i \leq d x_1^2$ at every point of $\{w > 0\}$. Then*

$$\mathcal{H}(w) \leq \frac{1+d}{4b} \mathcal{D}(w). \quad (11.52)$$

Proof. Fix z with $\bar{w}(z) > 0$ and let $\phi(v) = w(z, v)/\bar{w}(z)$, a smooth nonnegative function with $\int \phi dG_v = 1$. Then $\int \phi \log \phi dG_v = \mathbb{E}_w[g | z]$ and, on $\{\phi > 0\}$, $(\partial_{v_i} \phi)^2/\phi = \phi (\partial_{v_i} g)^2$. Where $\phi = 0$, also $\nabla \phi = 0$. By (11.51), (11.49), and (11.48), we have

$$\begin{aligned} \mathbb{E}_w[g | z] &\leq \frac{1}{4} \mathbb{E}_w[\nabla_v g^\top \text{diag}(a_i^{-1}) \nabla_v g | z] \\ &\leq \frac{1+d}{4b} \mathbb{E}_w[\nabla_v g^\top A_v \nabla_v g | z] \leq \frac{1+d}{4b} \mathbb{E}_w[\Gamma(g) | z]. \end{aligned}$$

Multiplying by $\bar{w}(z)$ and integrating with respect to G_z yields (11.52). $\square$

Keeping the matrix A_v leads to the constant $(1+d)/(4b)$, close to the Gaussian value $1/(4b)$ when d is small. Replacing A_v by its smallest eigenvalue $x_1^2/(N - |z|^2)$ would lose a factor of order one, which the applications cannot afford.

Weighted logarithmic moments

The entropy identity below contains logarithms of densities multiplied by powers of a cutoff function. The next lemma bounds them using only an L^p bound on the density, with no lower bound on it. We formulate this measure-theoretic statement on an arbitrary product of probability spaces.

Lemma 11.17 (Weighted logarithmic moments). *Let $G = G_Z \otimes G_V$ be a product probability measure on a product space $E_Z \times E_V$. Let $f \geq 0$ be measurable with $\int f dG \leq 1$ and*

$$\log \int f^p dG \leq C_p A$$

for some $p > 1$, $C_p > 0$, and $A \geq 2$. Let η be measurable with values in $[0, 1]$, let $Z = \int \eta f dG \geq z_0 > 0$, and let $w = \eta f / Z$ and $\bar{w}(z) = \int w(z, v) G_V(dv)$. On $E = \{0 < w < \infty, 0 < \bar{w} < \infty\}$ set $g = \log w - \log \bar{w}$, and set $g = 0$ elsewhere. Then for every $r \in [1, \infty)$ and $\alpha_0 \in (0, 1]$,

$$\sup_{\alpha \in [\alpha_0, 1]} \int_E f \eta^\alpha (|g|^r + |\log w|^r + |\log \bar{w}|^r) dG \leq C_r A^r, \quad (11.53)$$

where C_r depends only on r, α_0, p, C_p , and z_0 . Moreover $0 \leq \int w g dG \leq CA$, with C depending only on p, C_p , and z_0 .

For every $\alpha > 0$, the set E has full $f \eta^\alpha G$ -measure on $\{w > 0\}$. Indeed, the L^p assumption makes f finite almost everywhere, $\int \bar{w} dG_Z = 1$ makes $\bar{w}$ finite almost everywhere, and Tonelli's theorem gives $w = 0$ almost everywhere over the fibers where $\bar{w} = 0$. On the latter fibers this also gives $f \eta^\alpha = 0$ almost everywhere. Thus the estimates do not depend on how the logarithms are defined on the exceptional set.

Proof. Throughout, C denotes a constant depending only on r, α_0, p, C_p, z_0 . Note that $z_0 \leq Z \leq 1$. All density identities and fiber inequalities below are understood almost everywhere. In integrals over the product space, expressions involving $\log w$ or $\log \bar{w}$ are taken to be zero off E .

Step 1: moments slightly above one. Hölder's inequality shows that $\log \int f^{1+e} dG$ is a convex function of $e \in [0, p-1]$. Since it is at most 0 at $e = 0$ and at most $C_p A$ at $e = p-1$, we have $\int f^{1+e} dG \leq \exp(C_p A e / (p-1))$. We choose $e = (p-1)/A \leq (p-1)/2$, so that $\int f^{1+e} dG \leq e^{C_p}$. Since $w \leq f/z_0$, $z_0 \leq 1$, and $1+e \leq (p+1)/2$, we find $\int w^{1+e} dG \leq z_0^{-(1+e)} \int f^{1+e} dG \leq z_0^{-(p+1)/2} e^{C_p}$, and by Jensen's inequality on each fiber the functions $\bar{w}$ and $\bar{f}(z) = \int f(z, v) G_V(dv)$ satisfy

$$\int \bar{w}^{1+e} dG_Z \leq z_0^{-(p+1)/2} e^{C_p}, \quad \int \bar{f}^{1+e} dG_Z \leq e^{C_p}. \quad (11.54)$$

Step 2: the logarithms of f , η , and Z . On $\{f \leq 1\}$, we have $f|\log f|^r \leq (r/e_E)^r$, where e_E is Euler's number. On $\{f > 1\}$, by Markov's inequality, we have $\int_{\{\log f > x\}} f dG \leq e^{-(p-1)x} \int f^p dG \leq e^{C_p A - (p-1)x}$, and this integral is also at most 1. Let $x_0 = 2C_p A / (p-1)$. For $x \geq x_0$, we have $C_p A - (p-1)x \leq -(p-1)x/2$, and therefore

$$\int f(\log f)_+^r dG = \int_0^\infty r x^{r-1} \int_{\{\log f > x\}} f dG dx \leq x_0^r + \int_0^\infty r x^{r-1} e^{-(p-1)x/2} dx \leq C A^r.$$

Further, $\eta^\alpha |\log \eta|^r \leq (r/(\alpha_0 e_E))^\alpha$ for $\alpha \geq \alpha_0$, and $|\log Z| \leq \log(1/z_0)$. We have $\log w = \log f + \log \eta - \log Z$ on $\{w > 0\}$, and these bounds prove the part of (11.53) that concerns $\log w$.

Step 3: the logarithm of the marginal. Let $k_\alpha(z) = \int f \eta^\alpha dG_v$. By Hölder's inequality on the fiber with exponents $1/\alpha$ and $1/(1-\alpha)$,

$$k_\alpha = \int (f\eta)^\alpha \bar{f}^{1-\alpha} dG_v \leq (Z\bar{w})^\alpha \bar{f}^{1-\alpha},$$

and in particular $k_\alpha = 0$ where $\bar{w} = 0$. On $\{0 < \bar{w} \leq 1\}$, we have $k_\alpha |\log \bar{w}|^r \leq \bar{w}^\alpha |\log \bar{w}|^r \bar{f}^{1-\alpha} \leq C \bar{f}^{1-\alpha}$, and $\int \bar{f}^{1-\alpha} dG_z \leq 1$ by Jensen's inequality. On $\{\bar{w} > 1\}$, the inequality $(\log x)^r \leq (r/e_E)^r x$ for $x \geq 1$, applied to $x = \bar{w}^e$, gives $(\log \bar{w})^r \leq (rA/((p-1)e_E))^r \bar{w}^e$. Using Hölder's inequality with exponents $(1+e)/(\alpha+e)$ and $(1+e)/(1-\alpha)$ and (11.54), we find

$$\begin{aligned} \int_{\{\bar{w} > 1\}} k_\alpha (\log \bar{w})^r dG_z &\leq C A^r \int \bar{w}^{\alpha+e} \bar{f}^{1-\alpha} dG_z \\ &\leq C A^r \left(\int \bar{w}^{1+e} dG_z \right)^{\frac{\alpha+e}{1+e}} \left(\int \bar{f}^{1+e} dG_z \right)^{\frac{1-\alpha}{1+e}} \leq C A^r; \end{aligned}$$

for $\alpha = 1$, the second factor is absent. We have $\int f \eta^\alpha |\log \bar{w}|^r dG = \int k_\alpha |\log \bar{w}|^r dG_z$, and this proves the part of (11.53) that concerns $\log \bar{w}$, and the bound for $g = \log w - \log \bar{w}$ follows.

Step 4: the entropy. The integral $\int w g dG$ is the relative entropy $\text{KL}(wG \| (\bar{w}G_z) \otimes G_v) \geq 0$. The functions $w \log w$ and $w \log \bar{w}$ are integrable by Steps 2 and 3 with $\alpha = 1$. On $\{w > 0\}$, we have $\log w = \log \bar{w} + g$, and Fubini's theorem gives $\int w \log \bar{w} dG = \int \bar{w} \log \bar{w} dG_z \geq 0$, by Jensen's inequality for the convex function $x \log x$ and $\int \bar{w} dG_z = 1$. Hence $\int w g dG \leq \int w \log w dG = \text{KL}(wG \| G)$. By Jensen's inequality under the probability measure wG and $w \leq f/z_0$,

$$\int w \log w dG = \frac{\int w \log w^{p-1} dG}{p-1} \leq \frac{\log \int w^p dG}{p-1} \leq \frac{C_p A + p \log(1/z_0)}{p-1} \leq C A.$$

This completes the proof. □

The time derivative of the conditional relative entropy

We adapt the classical entropy dissipation calculation for diffusion semigroups [12, Proposition 5.2.2, (5.2.3), p. 245] to the conditional entropy and cutoff-weighted endpoint law; all additional terms are derived below. Let $T \geq 0$, let Ψ be an $\mathcal{F}_T$ -measurable random variable with values in $[0, 1]$, let $D \subset S_N$ be open (possibly $D = S_N$), and let σ_t be the endpoint subprobability measures (11.7) of Lemma 11.4. We assume that D and σ_T are invariant under R_1 , so that every σ_t is. Let

$$D^+ = \{y \in U_N : x^+(y) \in D\}.$$

For $t > T$, the measure σ_t^+ vanishes outside D^+ up to a null set, and, by Lemma 11.4 and the first part of Lemma 11.9, it has a density f_t with respect to G such that $(t, y) \mapsto f_t(y)$ is smooth on $(T, \infty) \times D^+$. Applying (11.8) to $\varphi = \psi \circ y$ for $\psi \in C_c^\infty(D^+)$, and using Lemma 11.14 and the invariance under R_1 , we obtain the weak forward equation

$$\frac{d}{dt} \int \psi f_t dG = \int f_t L\psi dG \quad (t > T, \psi \in C_c^\infty(D^+)), \quad (11.55)$$

where L is the operator (11.45). Let $\zeta \in C^\infty(\mathbb{R}^{N-1})$ take values in $[0, 1]$ and have compact support in D^+ , let $k \geq 3$ be an integer, and let $\eta = \zeta^k$. Let $\mathcal{T} \subset (T, \infty)$ be an open interval on which $Z_t = \int \eta f_t dG > 0$. For $t \in \mathcal{T}$, let

$$w_t = \frac{\eta f_t}{Z_t}, \quad \bar{w}_t(z) = \int w_t(z, v) G_v(dv), \quad h_t = \log \bar{w}_t, \quad g_t = \log w_t - h_t, \quad (11.56)$$

and let $\mathcal{H}(t) = \mathcal{H}(w_t)$, $\mathcal{D}(t) = \mathcal{D}(w_t)$, and $S_t = z \cdot \nabla h_t$. The functions w_t are smooth with compact support in D^+ , so (11.46) applies. We write $\mathbb{E}_w$ for $\mathbb{E}_{w_t}$ and omit the index t when no confusion arises. Recall that $\Gamma(\varphi, \log \eta) = \Gamma(\varphi, \eta)/\eta$ on $\{\eta > 0\}$, and that expectations $\mathbb{E}_w$ only see the set $\{w > 0\} \subset \{\eta > 0\}$.

Proposition 11.18 (Time derivative of the conditional relative entropy). *In this setting, $\mathcal{H}(t)$ and $\mathcal{D}(t)$ are finite for $t \in \mathcal{T}$, the function $\mathcal{H}$ is absolutely continuous on every compact subinterval of $\mathcal{T}$, and $Z'_t = \int f_t L\eta dG$. For almost every $t \in \mathcal{T}$,*

$$\mathcal{H}' = -\mathcal{D} + \frac{1}{N} \mathbb{E}_w[(1 + S) v \cdot \nabla_v g] + 2\mathbb{E}_w \Gamma(g, \log \eta) + \frac{1}{Z} \int f(g - \mathcal{H}) L\eta dG, \quad (11.57)$$

and

$$\begin{aligned} \mathcal{H}' = & -\mathcal{D} + \frac{1}{N} \mathbb{E}_w[(1 + S) v \cdot \nabla_v g] + \mathbb{E}_w[(1 - g + \mathcal{H}) \Gamma(g, \log \eta)] \\ & - \mathbb{E}_w[(g - \mathcal{H}) \Gamma(h, \log \eta)] + \mathbb{E}_w[(g - \mathcal{H}) \Gamma(\log \eta)] + \frac{1}{Z} \int f(g - \mathcal{H}) \frac{y}{N} \cdot \nabla \eta dG. \end{aligned} \quad (11.58)$$

In the last integrals of (11.57) and (11.58), the integrand is set to zero where $\eta f = 0$.

After the second integration by parts, (11.58) contains only first derivatives of η . The explicit drift contribution involving ∇F has been absorbed into the carré du champ terms; the remaining first-order term is $(y/N) \cdot \nabla \eta$.

Proof. We first give the formal computation and then justify it. The justification uses that, at fixed N , f is smooth on a neighborhood of the support of η , and no lower bound on f .

Formal computation. We have $\int w_t dG = 1$ and $\int \bar{w}_t dG_z = 1$. Differentiating $\mathcal{H} = \int w \log w dG - \int \bar{w} \log \bar{w} dG_z$, we find $\mathcal{H}' = \int \partial_t w g dG$. By (11.55) with $\psi = \eta$, we have $Z' = \int f L \eta dG$, and $\partial_t w = \eta \partial_t f / Z - w Z' / Z$. Applying (11.55) with $\psi = \eta g$, we find

$$\mathcal{H}' = \frac{1}{Z} \int f L(\eta g) dG - \frac{Z'}{Z} \mathcal{H}.$$

We expand $L(\eta g) = \eta Lg + g L\eta + 2\Gamma(\eta, g)$, write $L = L_G + (y/N) \cdot \nabla$, and integrate the term $\int f \eta L_G g dG = Z \int w L_G g dG$ by parts with (11.44). Using $\nabla w = w \nabla(g + h)$ on $\{w > 0\}$, we obtain

$$\mathcal{H}' = -\mathcal{D} - \mathbb{E}_w \Gamma(h, g) + \frac{1}{N} \mathbb{E}_w [y \cdot \nabla g] + 2\mathbb{E}_w \Gamma(g, \log \eta) + \frac{1}{Z} \int f(g - \mathcal{H}) L \eta dG. \quad (11.59)$$

Now we use the conditional structure. On each fiber with $\bar{w}(z) > 0$, we have $\int w \nabla_z g dG_v = \int \nabla_z w dG_v - \bar{w} \nabla_z h = \nabla_z \bar{w} - \nabla_z \bar{w} = 0$, that is,

$$\mathbb{E}_w [\nabla_z g \mid z] = 0. \quad (11.60)$$

The function h depends only on z . In $\Gamma(h, g) = \nabla_z h \cdot \nabla_z g - (z \cdot \nabla_z h)(z \cdot \nabla_z g)/N - (z \cdot \nabla_z h)(v \cdot \nabla_v g)/N$, the first two terms are functions of z multiplied by components of $\nabla_z g$. Then (11.60) implies that $\mathbb{E}_w \Gamma(h, g) = -N^{-1} \mathbb{E}_w [S v \cdot \nabla_v g]$. In the same way, $\mathbb{E}_w [y \cdot \nabla g] = \mathbb{E}_w [v \cdot \nabla_v g]$. Inserting these two identities into (11.59) gives (11.57).

For (11.58), we write $L\eta = L_G \eta + (y/N) \cdot \nabla \eta$ in the last term of (11.57), and we integrate the part with L_G by parts once more:

$$\frac{1}{Z} \int f(g - \mathcal{H}) L_G \eta dG = -\frac{1}{Z} \int [(g - \mathcal{H}) \Gamma(f, \eta) + f \Gamma(g, \eta)] dG.$$

On $\{w > 0\}$, we have $\log f = g + h - \log \eta + \log Z$ and $\Gamma(f, \eta) = f[\Gamma(g, \eta) + \Gamma(h, \eta) - \Gamma(\eta)/\eta]$. Using $f\Gamma(\cdot, \eta)/Z = w\Gamma(\cdot, \log \eta)$ and $f\Gamma(\eta)/(Z\eta) = w\Gamma(\log \eta)$, we see that the right side equals

$$\begin{aligned} & -\mathbb{E}_w [(g - \mathcal{H}) \Gamma(g, \log \eta)] - \mathbb{E}_w [(g - \mathcal{H}) \Gamma(h, \log \eta)] \\ & + \mathbb{E}_w [(g - \mathcal{H}) \Gamma(\log \eta)] - \mathbb{E}_w \Gamma(g, \log \eta). \end{aligned}$$

Together with the term $2\mathbb{E}_w \Gamma(g, \log \eta)$ of (11.57), the terms containing $\Gamma(g, \log \eta)$ combine to $\mathbb{E}_w [(1 - g + \mathcal{H}) \Gamma(g, \log \eta)]$, and (11.58) follows.

Justification. Fix a compact interval $[\alpha, \beta] \subset \mathcal{T}$. Below, “finite” and “bounded” refer to bounds that may depend on N , α , and β , and that are uniform in $t \in [\alpha, \beta]$. Let $\mathcal{K}$ be a compact neighborhood of $\text{supp } \zeta$ in D^+ . On $[\alpha, \beta] \times \mathcal{K}$, the function f and all its derivatives are bounded, Z_t is bounded below by a positive constant, and w_t and $\bar{w}_t$ are smooth with bounded derivatives. The function $f \geq 0$ is of class C^2 , and Taylor’s formula along $-\nabla f$ shows that $|\nabla f|^2 \leq C_{\mathcal{K}} f$ on $\text{supp } \zeta$. Indeed, if ρ_0 is the distance from $\text{supp } \zeta$ to the complement of $\mathcal{K}$ and M_2 bounds $\nabla^2 f$ on $\mathcal{K}$, then $0 \leq f - r|\nabla f| + M_2 r^2/2$ for $0 \leq r \leq \rho_0$, and the choices $r = |\nabla f|/M_2$ or $r = \rho_0$ yield $|\nabla f|^2 \leq 2M_2 f$ or $|\nabla f| \leq 2f/\rho_0$. The following quantities are then finite:

- (a) $\int w \Gamma(\log w) dG \leq 2 \int w \Gamma(\log \eta) dG + 2 \int w \Gamma(\log f) dG$, where $w \Gamma(\log \eta) = k^2 Z^{-1} f \zeta^{k-2} \Gamma(\zeta)$ and $w \Gamma(\log f) \leq Z^{-1} \eta |\nabla f|^2 / f$ are bounded;
- (b) $\int \bar{w} |\nabla h|^2 dG_z \leq \int w |\nabla_z \log w|^2 dG$, by the Cauchy–Schwarz inequality on each fiber; hence $\mathcal{D} \leq 2 \int w \Gamma(\log w) dG + 2 \int \bar{w} |\nabla h|^2 dG_z$ is finite;
- (c) $\int f \zeta^{k-2} (|\log w|^2 + |\log \bar{w}|^2) dG$ on $\{w > 0\}$, by Lemma 11.17 applied to $f \mathbf{1}_{\mathcal{K}}$, η , and $\alpha = (k-2)/k \geq 1/3$; and $\mathcal{H} \leq \int w (|\log w| + |\log \bar{w}|) dG$.

Let $0 < e \leq 1$ and $g_e = \log(w+e) - \log(\bar{w}+e)$, a smooth function. Let $\mathcal{H}_e = \int \Phi_e(w) dG - \int \Phi_e(\bar{w}) dG_z$ with $\Phi_e(x) = (x+e) \log(x+e) - x$, so that $\Phi'_e(x) = \log(x+e)$. Then $\mathcal{H}_e$ is continuously differentiable on $[\alpha, \beta]$ with $\mathcal{H}'_e = \int \partial_t w g_e dG$, and $\eta g_e \in C_c^\infty(D^+)$. The formal computation up to (11.59), with g_e in place of g inside the integrals and before any use of $\nabla w = w \nabla(g+h)$, involves only smooth functions and is valid. We obtain $\mathcal{H}_e(\beta) - \mathcal{H}_e(\alpha) = \int_\alpha^\beta R_e(t) dt$, where

$$\begin{aligned} R_e = & - \int \Gamma(w, g_e) dG + \int w \frac{y}{N} \cdot \nabla g_e dG \\ & + \frac{1}{Z} \int f g_e L\eta dG + \frac{2}{Z} \int f \Gamma(\eta, g_e) dG - \frac{Z'}{Z} \int w g_e dG. \end{aligned}$$

Now let $e \rightarrow 0$. We have $\nabla g_e = \nabla w/(w+e) - \nabla \bar{w}/(\bar{w}+e)$, and $\nabla w = 0$ where $w = 0$, $\nabla \bar{w} = 0$ where $\bar{w} = 0$. The terms of R_e converge pointwise to the corresponding terms with g , and they are dominated as follows.

- By item (a), $\Gamma(w)/(w+e) \leq \Gamma(w)/w$ is integrable.
- By items (a) and (b), the functions $|\Gamma(w, \bar{w})|/(\bar{w}+e) \leq (\Gamma(w)/w)^{1/2} w^{1/2} |\nabla \bar{w}|/\bar{w}$ and $w |\nabla g_e| \leq |\nabla w| + w |\nabla \bar{w}|/\bar{w}$ are integrable, and so is $|f \Gamma(\eta, g_e)| \leq Z w \Gamma(\log \eta)^{1/2} (\Gamma(\log w)^{1/2} + |\nabla h|)$.
- Since $|\log(x+e)| \leq |\log x| + \log 2$ for $x > 0$ and $|L\eta| \leq C \zeta^{k-2}$, the remaining terms are dominated using item (c).

Here we used that $f = 0$ wherever $\eta > 0$ and $w = 0$, and that $L\eta = 0$ where $\zeta = 0$ because $k \geq 3$. By dominated convergence, $\mathcal{H}_e(t) \rightarrow \mathcal{H}(t)$ and $R_e(t) \rightarrow R(t)$, the right side of (11.59), with $|R_e(t)|$ bounded uniformly in e and t . We conclude that $\mathcal{H}(\beta) - \mathcal{H}(\alpha) =$

$\int_{\alpha}^{\beta} R(t) dt$. In particular, $\mathcal{H}$ is absolutely continuous and (11.59) holds almost everywhere. The identity (11.60) holds because $w \nabla_z g$ is integrable by items (a) and (b). For the second integration by parts, we apply (11.44) to the smooth function $f(g_e - \mathcal{H})$ and to η , and pass to the limit in the same way; here $|g_e \Gamma(f, \eta)| \leq C |g_e| f^{1/2} \zeta^{k-1}$ by $|\nabla f|^2 \leq C_K f$, which is dominated by item (c) and the Cauchy–Schwarz inequality. This proves the proposition. $\square$

Entropy dissipation uniformly in the top fraction

We now bound the right side of (11.58) for a specific cutoff function. It is adapted to the approach to the critical temperature from below, where the top fraction θ tends to zero and all estimates must be uniform in θ . The point is that every error term is expressed through the dimensionless quantity $m = \theta N a_2$ and through a bound p_* on the mass of a region outside which the cutoff function equals one.

Setting. We assume that $0 = a_1 < a = a_2 \leq a_3 \leq \dots \leq a_N$, and let

$$\theta = 1 - \frac{1}{N} \sum_{i=2}^N \frac{1}{2a_i}, \quad m = \theta N a, \quad b = a_{J+1}, \quad (11.61)$$

where $J \geq 2$ is fixed with $a_J \leq C_J a$. For the scaled eigenvalue gaps of the model, θ is the top fraction θ_N of (3.14). We assume $\theta > 0$, $m \geq 2$, and $N\theta^3 \leq C_0 m^3$. In the chart, we write

$$|x_{\leq J}|^2 = \sum_{k=1}^J x_k^2 = N - |v|^2, \quad u = \frac{x_1^2}{\theta N}, \quad \bar{u} = \frac{|x_{\leq J}|^2}{\theta N}, \quad \zeta = \frac{b \sum_{i>J} v_i^2 / a_i}{|x_{\leq J}|^2}. \quad (11.62)$$

Set $\zeta = 0$ where $|x_{\leq J}| = 0$. The quantities $\bar{u}$ and ζ are invariant under all rotations of $(x_1, \dots, x_J)$. Fix $0 < s < L_0 < L_1 < L$ and functions $\chi_1, \chi_2, \chi_3 \in C^\infty(\mathbb{R}; [0, 1])$ such that $\chi_1 = 0$ on $(-\infty, s/2]$ and $\chi_1 = 1$ on $[s, \infty)$, $\chi_2 = 1$ on $(-\infty, L_0]$ and $\chi_2 = 0$ on $[L_1, \infty)$, and $\chi_3 = 1$ on $(-\infty, 1/2]$ and $\chi_3 = 0$ on $[1, \infty)$. Fix an integer $k \geq 16$ and $0 < \delta \leq 1$, and let

$$\eta = \zeta^k, \quad \zeta = \chi_1(u) \chi_2(\bar{u}) \chi_3(\zeta/\delta). \quad (11.63)$$

Since $\zeta = 0$ on $\{u < s/2\}$, an open set containing $\{|x_{\leq J}| = 0\}$ and a neighborhood of ∂U_N , the function ζ is smooth. Let

$$\mathcal{C} = \{u < s\} \cup \{\bar{u} > L_0\} \cup \{\zeta > \delta/2\}. \quad (11.64)$$

Then $\eta = 1$ outside $\mathcal{C}$, and all derivatives of ζ and η vanish outside $\mathcal{C}$. Thus $\mathcal{C}$ is a region outside which the cutoff function equals one, and it contains the set $\{\eta = 0\}$. We regard η and $\mathcal{C}$ also as a function and a set on S_N , through $x \mapsto y(x)$.

Endpoint measures with weighting and killing. Let X solve (11.2) with $X_0 \sim \nu_N$, and let $\rho_t = \nu_N P_t$ be its time- t law. Let $\pi \propto e^{-F} \nu_N$. Fix a deterministic time $T \geq 0$, an $\mathcal{F}_T$ -measurable random variable Ψ with values in $[0, 1]$ such that the measure $A \mapsto \mathbb{E}[\Psi \mathbf{1}_{\{X_T \in A\}}]$ is invariant under R_1 , and the open set

$$D = \{x \in S_N : |x_{\leq J}|^2 < L\theta N\}.$$

Let σ_t , $t \geq T$, be the endpoint subprobability measures (11.7): the path is weighted by Ψ and killed when it leaves D after time T . Then $\sigma_t \leq \rho_t$ as measures. With ζ as in (11.63), which has compact support in D^+ , let f_t , Z_t , w_t , $\bar{w}_t$, h_t , g_t , $\mathcal{H}(t)$, $\mathcal{D}(t)$, and S_t be as in Section 11.5.

Hypotheses. Let $t_0 \geq T + 1/a$ and $t_1 \geq t_0$. We assume that there exist constants $p > 1$, C_p , C_π , $z_0 > 0$, C_4 , and a number $p_* \geq 0$ such that for every $t \in [t_0, t_1]$:

$$\log \int \left(\frac{d\rho_t}{d\pi} \right)^p d\pi \leq C_p N \theta^3; \quad (11.65)$$

$$\frac{d\pi^+}{dG} \leq C_\pi \quad \text{on } \{y \in U_N : x_1(y)^2 \geq s\theta N/4\}; \quad (11.66)$$

$$Z_t \geq z_0, \quad \sigma_t(\mathcal{C}) \leq p_*; \quad (11.67)$$

$$\left(\int x_j^4 d\sigma_t \right)^{1/2} \leq C_4 \{ \theta N e^{-4a_j(t-T)} + a_j^{-1} \} \quad (2 \leq j \leq J). \quad (11.68)$$

Here (11.66) does not depend on t . The roles of the parameters are summarized in Table 11.1.

Table 11.1. Parameters in Theorem 11.19. The constants in the conclusion depend only on the fixed parameters in the first three rows.

Parameter	Role
J, C_J, C_0	number of retained coordinates; ratio bound $a_J \leq C_J a$; bound $N\theta^3 \leq C_0 m^3$
$s, L_0, L_1, L, \chi_1, \chi_2, \chi_3, k, \varepsilon$	cutoff function (11.63), killing set D , and the fraction ε of dissipation spent on errors
p, C_p, C_π, z_0, C_4	constants in the hypotheses (11.65)–(11.68)
$N, \theta, a_2, \dots, a_N$	free, subject to the setting; $m = \theta N a_2 \geq 2$
$\delta \in (0, 1], p_* \geq 0$	width of the cone condition $\zeta \leq \delta$; bound on $\sigma_t(\mathcal{C})$
$T \leq t_0 - 1/a, t_1 \geq t_0$	time of the weight Ψ and of the start of killing; time interval

Theorem 11.19 (Entropy dissipation for the remaining coordinates). *In this setting, the function $\mathcal{H}$ is absolutely continuous on $[t_0, t_1]$. For every $\varepsilon \in (0, 1/4)$, for almost every $t \in [t_0, t_1]$,*

$$\mathcal{H}' \leq -(1 - 2\varepsilon)\mathcal{D} + \frac{Ca}{m}\{1 + m^2e^{-8a(t-T)}\} + Cam^5\delta^{-2}p_*^{1/2}, \quad (11.69)$$

and for every $t \in [t_0, t_1]$,

$$\mathcal{H} \leq \frac{1 + 2L_1\delta/s}{4b}\mathcal{D}, \quad 0 \leq \mathcal{H} \leq Cm^3. \quad (11.70)$$

The constants C depend only on the fixed parameters of Table 11.1; in particular, they do not depend on N , θ , δ , p_* , T , t_0 , or t_1 .

Remark 11.20. Theorem 11.19 concerns the endpoint subprobability measures σ_t . The mass removed by the weight Ψ and by killing does not enter (11.69); it is accounted for separately, in total variation, with Lemma 11.3 and Lemma 11.21. The hypothesis (11.68) enters only through the bound (11.76) on $\|S\|_{L^2(w)}$. Its proof below shows that an additional term $\theta N p_*^{1/2}$ on the right side of (11.68) changes (11.69) only by a constant factor, provided $m p_*^{1/2} \leq C_*$ for a fixed constant C_* ; the constant C in (11.69) may then also depend on C_* .

The proof bounds each term of (11.58). Step 1 bounds weighted logarithms of f_t by m^3 , using (11.65) and (11.66). Step 2 bounds the scores of the normalized cutoff-weighted law $\eta\sigma_t/Z_t$ along the rotations K_{1j} , using Corollary 11.6. Step 3 converts them into bounds for $S = z \cdot \nabla h$, using Lemma 11.10 and (11.68). Step 4 computes the derivatives of the cutoff function. Steps 5 and 6 bound the term with $1 + S$ and the terms supported in $\mathcal{C}$, and Step 7 concludes with Proposition 11.16.

Proof. Throughout the proof, C denotes a constant depending only on the fixed parameters of Table 11.1. We may assume $p_* \leq 1$, since $\sigma_t(\mathcal{C}) \leq \sigma_t(S_N) \leq 1$. Note that $\theta < 1$, that $1/(\theta N) = a/m$, and that on $\text{supp } \zeta$ we have

$$x_1^2 \geq s\theta N/2, \quad |x_{\leq J}|^2 \leq L_1\theta N, \quad |z|^2 \leq L_1\theta N, \quad \xi \leq \delta. \quad (11.71)$$

Step 0: the entropy identity. The set D is open and invariant under R_1 , and ζ has compact support in D^+ by (11.71), because $L_1 < L$. By (11.67) and continuity, $Z_t > 0$ on an open interval $\mathcal{T} \supset [t_0, t_1]$, and $t_0 > T$. Proposition 11.18 shows that $\mathcal{H}$ is absolutely continuous on $[t_0, t_1]$ and that (11.58) holds for almost every $t \in [t_0, t_1]$. We fix $t \in [t_0, t_1]$. Steps 1–4 below hold for every such t . In Steps 5–7, we also assume that (11.58) holds at t .

Step 1: weighted logarithms. Let $\bar{f} = f \mathbf{1}_{\{x_1^2 \geq s\theta N/4\}}$, which agrees with f on $\text{supp } \eta$. Let $r_t = d\sigma_t/d\pi \leq d\rho_t/d\pi$. Since σ_t and π are invariant under R_1 , the density of σ_t^+ with respect to π^+ is $r_t \circ x^+$, and $\int (r_t \circ x^+)^p d\pi^+ = \int r_t^p d\pi$. By (11.66) and (11.65),

$$\int \bar{f}^p dG = \int_{\{x_1^2 \geq s\theta N/4\}} (r_t \circ x^+)^p \left(\frac{d\pi^+}{dG} \right)^{p-1} d\pi^+ \leq C_\pi^{p-1} \int \left(\frac{d\rho_t}{d\pi} \right)^p d\pi \leq e^{Cm^3}, \quad (11.72)$$

where we used $N\theta^3 \leq C_0 m^3$ and $m \geq 2$. Also $\int \bar{f} dG \leq \sigma_t(S_N) \leq 1$ and $\int \eta \bar{f} dG = Z_t \geq z_0$. We may therefore use Lemma 11.17 with $A = Cm^3$, and we obtain, for every $r \in [1, \infty)$ and $\alpha_0 \in (0, 1]$,

$$\sup_{\alpha \in [\alpha_0, 1]} \int f \eta^\alpha |g|^r dG \leq Cm^{3r}, \quad 0 \leq \mathcal{H} \leq Cm^3, \quad (11.73)$$

where C also depends on r and α_0 . This proves the second inequality in (11.70). We write $A_* = Cm^3$ for the constant in the second bound.

Step 2: scores after weighting by the cutoff. Let $2 \leq j \leq J$. The normalized measure $\sigma_t/\sigma_t(S_N)$ is the endpoint law weighted by $W = \Psi \mathbf{1}_{\{\tau > t\}}$, where τ is the exit time from D after T . This is a path weight for the final interval $[t - 1/a, t]$ and the set D in the sense of (11.11): Ψ and the event that the path stays in D during $[T, t - 1/a]$ are $\mathcal{F}_{t-1/a}^-$ -measurable, and the event that the path stays in D during $[t - 1/a, t]$ is invariant under the rotations $\Theta_\varepsilon^{(1j)}$, because D is invariant under all rotations of $(x_1, \dots, x_J)$. Its mass is $\sigma_t(S_N) \geq Z_t \geq z_0$, and Corollary 11.6 implies that

$$\mathbb{E}_{\sigma_t/\sigma_t(S_N)} \left| K_{1j} \log \frac{d\sigma_t}{d\nu_N} \right|^4 \leq Cm^2; \quad (11.74)$$

here the normalization by $\sigma_t(S_N)$ does not change the score. The quantities $\bar{u}$ and $\bar{\zeta}$ are invariant under the rotations generated by K_{1j} , and $K_{1j}u = -2x_1x_j/(\theta N)$. On $\text{supp } \zeta$, we have $|K_{1j}u| \leq (x_1^2 + x_j^2)/(\theta N) \leq \bar{u} \leq L_1$, and therefore

$$|K_{1j}\eta| = k\zeta^{k-1} |\chi'_1(u)| |K_{1j}u| \chi_2 \chi_3 \leq C\eta^{1-1/k} \leq C\eta^{3/4}, \quad (11.75)$$

since $k \geq 16$. Let $\hat{\sigma}_t = \eta\sigma_t/Z_t$ be the normalized cutoff-weighted law, with normalizing mass $Z_t = \int \eta d\sigma_t$; it is a probability measure with $\hat{\sigma}_t^+ = w_t G$. Lemma 11.2 with $p = 2$ implies that its weak K_{1j} -score ϑ_j satisfies $\mathbb{E}_{\hat{\sigma}_t} |\vartheta_j|^4 \leq Cm^2$.

Step 3: the marginal score. The measure $\hat{\sigma}_t$ is invariant under R_1 and satisfies $\hat{\sigma}_t(\{x_1^2 < s\theta N/2\}) = 0$. By Lemma 11.10, we have (11.34) with $w = w_t$, and $S = \sum_{j=2}^J z_j \partial_j h$ is the sum of the three terms

$$T_{1,j} = \mathbb{E}_w \left[\frac{z_j}{x_1} \vartheta_j \circ x^+ \mid z \right], \quad T_{2,j} = \mathbb{E}_w \left[\frac{z_j^2}{x_1^2} \mid z \right], \quad T_{3,j} = 2a_j z_j^2.$$

For the first term, by Jensen's inequality, (11.71), and the Cauchy–Schwarz inequality,

$$\mathbb{E}_w T_{1,j}^2 \leq \frac{2}{s\theta N} \mathbb{E}_w [z_j^2 \vartheta_j^2] \leq \frac{2}{s\theta N} (\mathbb{E}_w z_j^4)^{1/2} (\mathbb{E}_w |\vartheta_j \circ x^+|^4)^{1/2}.$$

We have $\mathbb{E}_w z_j^4 = Z_t^{-1} \int \eta x_j^4 d\sigma_t \leq z_0^{-1} \int x_j^4 d\sigma_t$ by (11.68), so $(\mathbb{E}_w z_j^4)^{1/2} \leq C\{\theta N e^{-4a_j(t-T)} + a_j^{-1}\}$. With $\mathbb{E}_w |\vartheta_j \circ x^+|^4 = \mathbb{E}_{\hat{\sigma}_t} \vartheta_j^4 \leq Cm^2$, by the remark after Lemma 11.9, and $a_j \geq a$, we find $\mathbb{E}_w T_{1,j}^2 \leq C\{me^{-4a(t-T)} + m/(\theta Na_j)\} \leq C\{1 + me^{-4a(t-T)}\}$. The second term satisfies $0 \leq T_{2,j} \leq |z|^2/x_1^2 \leq 2L_1/s$ by (11.71). For the third term, $\mathbb{E}_w T_{3,j}^2 = 4a_j^2 \mathbb{E}_w z_j^4 \leq C\{\theta Na_j e^{-4a(t-T)} + 1\}^2$, and $\theta Na_j \leq C_J m$. Summing over $j \leq J$, we have

$$\|S\|_{L^2(w)} \leq C\{1 + me^{-4a(t-T)}\}, \quad \|S\|_{L^4(w)} \leq Cm. \quad (11.76)$$

For the second bound, we have $|T_{1,j}|^4 \leq \mathbb{E}_w [(z_j/x_1)^4 \vartheta_j^4 \mid z]$ with $(z_j/x_1)^4 \leq (2L_1/s)^2$, which gives $\mathbb{E}_w T_{1,j}^4 \leq Cm^2$. Further, $T_{2,j}$ is bounded, and $0 \leq T_{3,j} \leq 2a_j L_1 \theta N \leq Cm$ by (11.71). This bound uses no moment of x_j .

Step 4: derivatives of the cutoff function. We write $\mathfrak{d} = y/N$. The following computations involving ξ are on $\{|x_{\leq J}| > 0\}$, which contains $\text{supp } \xi$ by (11.71). A direct computation from (11.62) shows that

$$\begin{aligned} \Gamma(u) &= \frac{4u(1 - \theta u)}{\theta N}, & \Gamma(\bar{u}) &= \frac{4\bar{u}(1 - \theta \bar{u})}{\theta N}, \\ \Gamma(\xi) &= \frac{4(v^\top D_b^2 v + \xi^2 |x_{\leq J}|^2)}{|x_{\leq J}|^4} \leq \frac{4\xi(1 + \xi)}{|x_{\leq J}|^2}, \end{aligned} \quad (11.77)$$

where $D_b = \text{diag}(b/a_i)_{i>J}$. For the last one, we use $\nabla_v \xi = 2(D_b v + \xi v)/|x_{\leq J}|^2$ and $v \cdot \nabla_v \xi = 2N\xi/|x_{\leq J}|^2$. Subtracting the rank-one part of $\mathfrak{g}$ leaves the displayed expression, and the inequality follows from $D_b^2 \leq D_b$, which holds because $b \leq a_i$. Also

$$|\mathfrak{d} \cdot \nabla u| \leq \frac{2}{\theta N}, \quad |\mathfrak{d} \cdot \nabla \bar{u}| \leq \frac{2}{\theta N}, \quad \mathfrak{d} \cdot \nabla \xi = \frac{2\xi}{|x_{\leq J}|^2}, \quad (11.78)$$

and, for every function h of z alone, with $S = z \cdot \nabla h$,

$$\Gamma(h, u) = -\frac{2uS}{N}, \quad \Gamma(h, \bar{u}) = \frac{2|v|^2 S}{\theta N^2}, \quad \Gamma(h, \xi) = -\frac{2\xi S}{|x_{\leq J}|^2}. \quad (11.79)$$

On $\text{supp } \xi$, the derivatives of $\chi_1(u)$, $\chi_2(\bar{u})$, and $\chi_3(\xi/\delta)$ are nonzero only where $u \leq s$, where $\bar{u} \leq L_1$, and where $\delta/2 \leq \xi \leq \delta$, respectively. In the last case, $\Gamma(\xi)/\delta^2 \leq 8/(\delta|x_{\leq J}|^2) \leq 16/(s\theta N\delta)$ and $|\mathfrak{d} \cdot \nabla \xi|/\delta \leq 4/(s\theta N)$ by (11.71). All these derivatives

vanish outside $\mathcal{C}$. Since $\eta = \zeta^k$, $\Gamma(\eta)/\eta = k^2\zeta^{k-2}\Gamma(\zeta)$, and $1/(\theta N) = a/m$, we conclude that

$$\frac{\Gamma(\eta)}{\eta} \leq \frac{Ca}{m\delta^2} \eta^{1-2/k} \mathbf{1}_{\mathcal{C}}, \quad |\mathfrak{d} \cdot \nabla \eta| \leq \frac{Ca}{m} \eta^{1-1/k} \mathbf{1}_{\mathcal{C}}, \quad |\Gamma(h, \eta)| \leq \frac{Ca}{m} |S| \eta^{1-1/k} \mathbf{1}_{\mathcal{C}}. \quad (11.80)$$

In the first bound, δ^{-1} would suffice, but the weaker bound with δ^{-2} is enough for our purposes.

Step 5: the term with $1 + S$. By the first part of Lemma 11.15 and (11.71), we have $|v \cdot \nabla_v g| \leq (v^\top A_v^{-1} v)^{1/2} \Gamma(g)^{1/2}$ and $v^\top A_v^{-1} v \leq N^2/x_1^2 \leq 2N/(s\theta)$. Young's inequality and (11.76) then yield

$$\left| \frac{1}{N} \mathbb{E}_w[(1+S) v \cdot \nabla_v g] \right| \leq \varepsilon \mathcal{D} + \frac{1}{2\varepsilon s \theta N} \mathbb{E}_w(1+S)^2 \leq \varepsilon \mathcal{D} + \frac{Ca}{m} \{1 + m^2 e^{-8a(t-T)}\}. \quad (11.81)$$

Step 6: the terms supported in $\mathcal{C}$. Each of the remaining four terms of (11.58) has the form $\mathbb{E}_w[\dots]$ or $Z^{-1} \int f \dots dG$ with an integrand that vanishes outside $\mathcal{C}$, by (11.80). We first record two consequences of (11.67) and (11.73). If $\beta_1 \geq 0$, then $\int f \eta^{\beta_1} \mathbf{1}_{\mathcal{C}} dG \leq \sigma_t(\mathcal{C}) \leq p_*$. Splitting $\eta^{1-2/k} = \eta^{1/4} \cdot \eta^{3/4-2/k}$ and using the Cauchy–Schwarz inequality, (11.73) with $r = 4$ and $\alpha_0 = 1/2$, and $3/2 - 4/k \geq 0$, we find

$$\int f \eta^{1-2/k} (1+g^2) \mathbf{1}_{\mathcal{C}} dG \leq p_* + \left(\int f \eta^{1/2} g^4 dG \right)^{1/2} \left(\int f \eta^{3/2-4/k} \mathbf{1}_{\mathcal{C}} dG \right)^{1/2} \leq CA_*^2 p_*^{1/2}, \quad (11.82)$$

and in the same way $\int f \eta^{1-2/k} |g| \mathbf{1}_{\mathcal{C}} dG \leq CA_* p_*^{1/2}$ and $\int f \eta^{1-1/k} |g| \mathbf{1}_{\mathcal{C}} dG \leq CA_* p_*^{1/2}$. Recall also that $\mathcal{H} \leq A_*$.

For the first term, we have $|\mathbb{E}_w[(1-g+\mathcal{H})\Gamma(g, \log \eta)]| \leq \varepsilon \mathcal{D} + (4\varepsilon)^{-1} \mathbb{E}_w[(1-g+\mathcal{H})^2 \Gamma(\log \eta)]$, by the Cauchy–Schwarz and Young inequalities. Since $w\Gamma(\log \eta) = Z^{-1} f \Gamma(\eta)/\eta$, we deduce from (11.80) and (11.82) that

$$|\mathbb{E}_w[(1-g+\mathcal{H})\Gamma(g, \log \eta)]| \leq \varepsilon \mathcal{D} + \frac{Ca}{m\delta^2} A_*^2 p_*^{1/2}. \quad (11.83)$$

For the second term, $w\Gamma(h, \log \eta) = Z^{-1} f \Gamma(h, \eta)$, and (11.80) reduces it to $(Ca/m) \int f \eta^{1-1/k} |S| |g - \mathcal{H}| \mathbf{1}_{\mathcal{C}} dG$. By Hölder's inequality with exponents 4, 4, and 2, with the weights $\eta^{1/4}$, $\eta^{1/8}$, and $\eta^{5/8-1/k}$, whose product is $\eta^{1-1/k}$,

$$\begin{aligned} \int f \eta^{1-1/k} |S| |g - \mathcal{H}| \mathbf{1}_{\mathcal{C}} dG &\leq \left(\int f \eta |S|^4 dG \right)^{1/4} \left(\int f \eta^{1/2} |g - \mathcal{H}|^4 dG \right)^{1/4} \left(\int f \eta^{5/4-2/k} \mathbf{1}_{\mathcal{C}} dG \right)^{1/2} \\ &\leq CmA_* p_*^{1/2}, \end{aligned}$$

where we used $\int f \eta |S|^4 dG = Z \|S\|_{L^4(w)}^4 \leq Cm^4$ by (11.76). After multiplication by Ca/m , the second term is at most $CA_* p_*^{1/2}$. Similarly, (11.80) and the bounds after (11.82) give the bound $(Ca/(m\delta^2)) A_* p_*^{1/2}$ for the third term and the bound $(Ca/m) A_* p_*^{1/2}$ for the last term. Since $A_* = Cm^3$, $m \geq 2$, and $\delta \leq 1$, all four terms are at most $\varepsilon \mathcal{D} + Cam^5 \delta^{-2} p_*^{1/2}$ in total.

Step 7: conclusion. Inserting (11.81) and the bounds of Step 6 into (11.58), we obtain (11.69). For the first inequality of (11.70), note that on $\text{supp } \zeta$ we have $\zeta \leq \delta$ and, by (11.71), $|x_{\leq J}|^2 \leq L_1 \theta N \leq (2L_1/s)x_1^2$. Hence

$$b \sum_{i>J} \frac{v_i^2}{a_i} = \zeta |x_{\leq J}|^2 \leq \frac{2L_1 \delta}{s} x_1^2 \quad \text{on } \{w > 0\},$$

and Proposition 11.16 with $d = 2L_1 \delta/s$ implies that $\mathcal{H} \leq (1 + 2L_1 \delta/s) \mathcal{D}/(4b)$. This completes the proof. $\square$

11.6 Combining the coordinates

The last lemma combines a conditional relative entropy bound for a normalized weighted law with a total variation bound for the first coordinates of the unweighted law. In the applications, it is applied after the chart of Section 11.4, with (11.30); the stationary law π is then compared with G separately.

Lemma 11.21 (Combining a conditional relative entropy bound with the first coordinates). *Let ρ be a probability measure on a product space $E_Z \times E_V$, and let $\sigma \leq \rho$ be a measure with mass $Z > 0$ and $\hat{\rho} = \sigma/Z$. Let $G = G_Z \otimes G_V$ be a product probability measure, and let μ be a probability measure on E_Z . Write ρ_Z and $\hat{\rho}_Z$ for the marginals on E_Z . If $\hat{\rho} \ll G$, let $H_V = \text{KL}(\hat{\rho} \| \hat{\rho}_Z \otimes G_V)$. Then*

$$\|\rho - \mu \otimes G_V\|_{\text{TV}} \leq 2(1 - Z) + \sqrt{H_V/2} + \|\rho_Z - \mu\|_{\text{TV}}. \quad (11.84)$$

For every probability measure π on $E_Z \times E_V$,

$$\left| \|\rho - \pi\|_{\text{TV}} - \|\mu - G_Z\|_{\text{TV}} \right| \leq 2(1 - Z) + \sqrt{H_V/2} + \|\rho_Z - \mu\|_{\text{TV}} + \|\pi - G\|_{\text{TV}}. \quad (11.85)$$

Proof. By Lemma 11.3, $\|\rho - \hat{\rho}\|_{\text{TV}} \leq 1 - Z$, and the same bound holds for the marginals ρ_Z and $\hat{\rho}_Z$ by the second part of Lemma 3.6. Pinsker's inequality (Lemma 3.6) gives $\|\hat{\rho} - \hat{\rho}_Z \otimes G_V\|_{\text{TV}} \leq \sqrt{H_V/2}$. For probability measures α, α' on E_Z , we have $\|\alpha \otimes G_V - \alpha' \otimes G_V\|_{\text{TV}} = \|\alpha - \alpha'\|_{\text{TV}}$: the inequality $\geq$ follows by projection, and $\leq$ from $(\alpha - \alpha') \otimes G_V(B) = \int (\alpha - \alpha')(B^v) G_V(dv)$, where B^v is the section of B at v . Using the triangle inequality along $\rho, \hat{\rho}, \hat{\rho}_Z \otimes G_V, \rho_Z \otimes G_V, \mu \otimes G_V$, we obtain

$$\|\rho - \mu \otimes G_V\|_{\text{TV}} \leq (1 - Z) + \sqrt{H_V/2} + (1 - Z) + \|\rho_Z - \mu\|_{\text{TV}},$$

which is (11.84). For (11.85), the reverse triangle inequality gives $|\|\rho - \pi\|_{\text{TV}} - \|\mu \otimes G_V - G\|_{\text{TV}}| \leq \|\rho - \mu \otimes G_V\|_{\text{TV}} + \|\pi - G\|_{\text{TV}}$, and $\|\mu \otimes G_V - G_Z \otimes G_V\|_{\text{TV}} = \|\mu - G_Z\|_{\text{TV}}$. $\square$

If $H_V = \infty$, there is nothing to prove. The constant 2 in front of $1 - Z$ is not optimal, but suffices for the applications.

Density bounds relative to time-dependent Bingham laws

In this section, we bound the density of the time- t law with respect to the Gibbs measure in every space L^p , at times when the time- t law is still far from equilibrium in total variation. The general result is Theorem 12.8. We prove it for deterministic half-precisions $r_1, \dots, r_N$ that satisfy the hypotheses of Definition 12.2 on a time interval $[0, S]$, and it states that, for every fixed $p > 1$,

$$\log \int_{S_N} \left(\frac{d(\nu_N P_t^r)}{d\pi_r} \right)^p d\pi_r \leq C_p N (1+t)^{-3/2}, \quad 0 \leq t \leq S,$$

where π_r and P_t^r are the Gibbs measure and the semigroup with these half-precisions. In the critical window, the hypotheses hold on $[0, KN^{2/3}]$ for every fixed K and all large N , and the right side is bounded at times of order $N^{2/3}$. We record the consequences in Section 12.6. The density of the time- t law has bounded $L^p(\pi_r)$ norm at every fixed positive multiple of $N^{2/3}$ (Theorem 12.13). Together with the stationary Poincaré inequality, this gives exponential decay of the total variation distance on the scale $N^{2/3}$ (Proposition 12.14).

We compare $\nu_N P_t^r$ with the Bingham comparison law μ_s of Section 7, evaluated at the earlier time $s = \gamma t$ for a small constant $\gamma > 0$. The logarithm of the p -th moment of the density of $\nu_N P_t^r$ with respect to $\mu_{\gamma t}$ is a multiple of a Rényi divergence. We differentiate it in t along the forward equation (Lemma 12.7). The derivative contains a dissipation term and two error terms. Using a logarithmic Sobolev inequality for $\mu_{\gamma t}$, we bound the Fisher information in the dissipation term below by a multiple of the divergence itself. The error terms come from the difference between the potential of the dynamics and that of $\mu_{\gamma t}$, and from the motion of $\mu_{\gamma t}$; both are quadratic functions of the coordinates. Their coefficients are small, because in each of them a term common to all coordinates cancels exactly (Lemma 12.4). We control their exponential moments under $\mu_{\gamma t}$ with the exact formula of Lemma 5.13 (Lemma 12.5) and close the resulting differential inequality by a scalar argument (Lemma 12.6). Finally, we change the reference measure from $\mu_{\gamma t}$ to π_r , using the same exponential moments. All statements concern the time- t law of the diffusion started from ν_N . The comparison laws enter only as reference measures: μ_s need not coincide with the time- s law, and no such identification is used.

We use the conventions of Section 3, the conditioned Gaussian laws of Section 5 (in particular Lemmas 5.5 and 5.13), the comparison laws of Section 7 and the estimate Lemma 7.14, and Lemma 8.12 and Proposition 8.21 of Section 8. In Section 17, we use Theorem 12.8 below the critical temperature to prove Theorem 17.20. Theorem 12.13 and Proposition 12.14 are used for the critical profile in Section 19.

Rényi divergences were introduced in [122, Section 3, (3.3), p. 554]; we follow the presentation of van Erven and Harremoës [147]. In [39, Theorem 1.2], Cao, Lu, and Lu proved exponential decay of Rényi divergences along Fokker–Planck equations with uniformly convex potentials, $D^2V \succeq \kappa I$ for some $\kappa > 0$, for smooth initial data that decay sufficiently fast. Vempala and Wibisono computed the time derivative of a Rényi divergence along the Langevin dynamics, and proved its exponential decay when the stationary measure satisfies a logarithmic Sobolev inequality [148, Lemmas 5 and 6, Theorem 2]. In our argument, the reference measure need not be the stationary measure. It moves, its logarithmic Sobolev constant grows linearly in time, and we must estimate the error terms that its motion produces. We compute the exponential moments of these error terms exactly through the radial densities of Section 5, and this is where the scale $N(1+t)^{-3/2}$ comes from.

12.1 Setting and notation

Half-precisions and the dynamics. Throughout this section, $N \geq 3$ and $r = (r_1, \dots, r_N)$ is a vector of positive numbers such that

$$\sum_{i=1}^N \frac{1}{2r_i} = N. \quad (12.1)$$

Let $A_r = -2 \operatorname{diag}(r_1, \dots, r_N)$. Note that, by (12.1), the Lagrange multiplier of A_r (Definition 3.3) is $z = 0$, so that its precisions are $2r_1, \dots, 2r_N$ and its half-precisions are $r_1, \dots, r_N$. We write $F_r(x) = \sum_i r_i x_i^2$ and

$$L_r = L_{A_r} = \Delta_{S_N} - \nabla_{S_N} F_r \cdot \nabla_{S_N}, \quad P_t^r = e^{tL_r}, \quad \pi_r(dx) = \frac{1}{Z_r} e^{-F_r(x)} \nu_N(dx), \quad (12.2)$$

where $Z_r = \int_{S_N} e^{-F_r} d\nu_N$. Recall from (3.9) and (3.12) that the model with rescaled inverse temperature c is the case where $r_i = (z_N(c) - c\lambda_i)/2$ are its half-precisions; then $L_r = L$, $P_t^r = P_t$, $\pi_r = \pi_{N,c}$, and $\nu_N P_t^r$ is the time- t law. Recall from Proposition 5.6 that π_r is the conditioned Gaussian law μ_q with variances $q_i = (2r_i)^{-1}$. We write $\Gamma(f) = |\nabla_{S_N} f|^2$, and logarithmic Sobolev inequalities are understood as in (3.5). Recall from (3.4) that $-\int f L_r g d\pi_r = \int \langle \nabla_{S_N} f, \nabla_{S_N} g \rangle d\pi_r$ for smooth f and g . We also use this identity for $f \in C^1(S_N)$ and $g \in C^2(S_N)$: approximate them by smooth functions uniformly together with their derivatives through orders one and two, respectively, and pass to the limit

on the compact sphere. The same approximation through order one extends every logarithmic Sobolev inequality below from its smooth core to C^1 tests, because both the entropy and the energy converge under this approximation.

The comparison law. Let $\mathbf{1} = (1, \dots, 1)^\top$, and let

$$H = -4 \operatorname{diag}(r_1, \dots, r_N) + \frac{2}{N} \mathbf{1} \mathbf{1}^\top, \quad h(t) = e^{tH} \mathbf{1}, \quad (12.3)$$

and, for $t \geq 0$,

$$\bar{h}(t) = \frac{1}{N} \sum_{i=1}^N h_i(t), \quad v_i(t) = \frac{h_i(t)}{\bar{h}(t)}. \quad (12.4)$$

The matrix H is the matrix of the covariance evolution of Section 7 for A_r , whose precisions are $2r_i$; the vector $h(t)$ is its heat vector, and $v_1(t), \dots, v_N(t)$ are its normalized variances. They are positive, $\sum_i v_i(t) = N$, and $v_i(0) = 1$. Recall from Definition 7.5 and Proposition 5.6 that the Bingham comparison law of A_r at time $t \geq 0$ is the conditioned Gaussian law $\mu_t = \mu_{v(t)}$ of Definition 5.4, that is,

$$\mu_t(dx) = \frac{1}{\mathcal{Z}(t)} \exp\left(-\sum_{i=1}^N \frac{x_i^2}{2v_i(t)}\right) v_N(dx), \quad (12.5)$$

with $\mathcal{Z}(t)$ the normalizing constant. Note that $\mu_0 = v_N$, since $v_i(0) = 1$. The numbers $v_i(t)$ need not equal the second moments of μ_t or of the time- t law. We define

$$\ell(t) = \frac{\bar{h}'(t)}{\bar{h}(t)}, \quad \zeta_i(t) = r_i - \frac{1}{2v_i(t)} + \frac{\ell(t)}{4}. \quad (12.6)$$

Using $\sum_i x_i^2 = N$ on S_N , we find that $F_r(x) - \sum_i x_i^2 / (2v_i(t)) = \sum_i \zeta_i(t) x_i^2 - N\ell(t)/4$ there, which yields

$$\frac{d\mu_t}{d\pi_r}(x) = \frac{\exp\left(\sum_i \zeta_i(t) x_i^2\right)}{\int_{S_N} \exp\left(\sum_i \zeta_i(t) y_i^2\right) \pi_r(dy)}. \quad (12.7)$$

The coefficients $\zeta_i(t)$ measure the difference between μ_t and π_r . The common term $\ell(t)/4$ does not change (12.7); it is included because it makes all $\zeta_i(t)$ small at once (Lemma 12.4).

Rényi divergences. For probability measures $\mu \ll \eta$ and $p > 1$, we write

$$\mathcal{R}_p(\mu \parallel \eta) = \log \int \left(\frac{d\mu}{d\eta}\right)^p d\eta \in [0, \infty]. \quad (12.8)$$

This is $p - 1$ times the Rényi divergence of order p of μ from η [147, Definition 2, (9)]. It is nonnegative by Jensen's inequality. By Jensen's inequality under μ , we also have

$$\operatorname{KL}(\mu \parallel \eta) = \frac{1}{p-1} \int \log \left(\frac{d\mu}{d\eta}\right)^{p-1} d\mu \leq \frac{1}{p-1} \log \int \left(\frac{d\mu}{d\eta}\right)^{p-1} d\mu = \frac{\mathcal{R}_p(\mu \parallel \eta)}{p-1}, \quad (12.9)$$

which is the monotonicity of Rényi divergences in their order [147, Theorems 3 and 5].

Gaussian squares. For positive $w = (w_1, \dots, w_N)$, let p_w be the continuous density of $T_w = \sum_i w_i (g_i^2 - 1)$ from Section 5.1 and Lemma 5.2, where $g_1, \dots, g_N$ are independent standard Gaussian variables. Thus $p_w(N - \sum_j w_j)$ is the density of $\sum_i w_i g_i^2$ at N , and $p_w(0)$ is its density at its mean. We use two facts. The first is Lemma 5.13 with $t = 1$ and with $2v_i \omega_i$ in place of the coefficients w_i there, written in the coordinates $x_i = \sqrt{v_i} Z_i$: if $\sum_i v_i = N$ and $\omega_1, \dots, \omega_N$ are real numbers with $2v_i \omega_i < 1$ for all i , then

$$\int_{S_N} e^{\sum_i \omega_i x_i^2} \mu_v(dx) = \prod_{i=1}^N (1 - 2v_i \omega_i)^{-1/2} \frac{p_w(N - \sum_j w_j)}{p_w(0)}, \quad w_i = \frac{v_i}{1 - 2v_i \omega_i}. \quad (12.10)$$

The second fact is the following restatement of Lemma 8.12.

Lemma 12.1 (Weighted sums of squares). *Let $\delta \in (0, 1]$, let $w = (w_1, \dots, w_N)$ be positive, and suppose that there exist three distinct indices i with $w_i \geq \delta \max_j w_j$. There exists $C_\delta \geq 1$, depending only on δ , such that*

$$p_w(0) \geq C_\delta^{-1} \left(\sum_i w_i^2 \right)^{-1/2}, \quad \sup_{y \in \mathbb{R}} p_w(y) \leq C_\delta \left(\sum_i w_i^2 \right)^{-1/2}. \quad (12.11)$$

If moreover $\sum_i w_i = N$, then $\int_{S_N} x_i^2 \mu_w(dx) \leq C_\delta w_i$ for every i .

Proof. Since T_w is the translate of $Q_w = \sum_i w_i g_i^2$ by $\sum_i w_i$, the lower bound in (12.11) is the lower bound of Lemma 8.12 for the density of Q_w at its mean, with $b = w$. For the upper bound, fix an index k and write $Q_w = Q_{w,-k} + w_k g_k^2$, with independent summands. If h_{-k} is the density of $Q_{w,-k}$ and f_k is the density of $w_k g_k^2$, then the density of Q_w is $h_{-k} * f_k$, which is continuous because h_{-k} is bounded and f_k is integrable. Hence $\sup p_w = \|h_{-k} * f_k\|_\infty \leq \|h_{-k}\|_\infty \|f_k\|_1 = \|h_{-k}\|_\infty$, and the bound of Lemma 8.12 for the density with one coordinate removed gives the upper bound in (12.11). For the last claim, Lemma 8.12 bounds by $C_\delta w_i$ the conditional expectation of $w_i g_i^2$ given $\sum_j w_j g_j^2 = N$, defined there through the joint density at the level N . By the third part of Lemma 5.5, applied with $F(y) = y_i^2$, the integral $\int x_i^2 \mu_w(dx)$ is the same ratio of densities at the level N : in both cases the numerator is the value at that level of a continuous density of the measure $B \mapsto \mathbb{E}[w_i g_i^2; \sum_j w_j g_j^2 \in B]$, and two continuous densities of the same measure agree. This proves the last claim. $\square$

Hypotheses. The general results of this section are stated under the following hypotheses.

Definition 12.2 (Hypotheses on the time interval $[0, S]$). Let $S > 0$ and $C_* \geq 1$. We say that r satisfies (H1)–(H4) on $[0, S]$ with constant C_* if the following hold.

(H1) $\max_i r_i \leq C_*$ and $N(1 + S)^{-3/2} \geq C_*^{-1}$.

(H2) For all $t \in [0, S]$ and all i ,

$$\frac{1}{C_*\sqrt{1+t}} \leq \bar{h}(t) \leq \frac{C_*}{\sqrt{1+t}}, \quad \frac{\sqrt{1+t}}{C_*(1+r_it)} \leq h_i(t) \leq \frac{C_*\sqrt{1+t}}{1+r_it}.$$

(H3) For all $t \in [0, S]$, $\sum_i (1+r_it)^{-2} \leq C_*N(1+t)^{-3/2}$.

(H4) There exist three distinct indices i with $r_iS \leq C_*$.

We say that r satisfies (H1)–(H5) on $[0, S]$ with constant C_* if moreover the following holds.

(H5) For every $t \in [0, S]$, the comparison law μ_t satisfies a logarithmic Sobolev inequality with constant $C_*(1+t)$:

$$\text{Ent}_{\mu_t}(f^2) \leq C_*(1+t) \int_{S_N} \Gamma(f) d\mu_t \quad \text{for all smooth } f \text{ on } S_N.$$

The heat bounds (H2) say that $h(t)$ behaves as in the critical window, where they are Lemma 7.14. The counting bound (H3) holds when the half-precisions grow like $(i/N)^{2/3}$, by Lemma 12.3 below. The hypothesis (H4) ensures that three variances $v_i(t)$ are comparable to the largest one, so that Lemma 12.1 applies to $v(t)$. We verify all five hypotheses in the critical window in Lemma 12.10.

12.2 The coefficients of the comparison law

We begin with an elementary counting bound.

Lemma 12.3 (Sums over the half-precisions). *Let $c_0 > 0$ and $k_0 \geq 0$, and suppose that $\#\{i : r_i \leq u\} \leq k_0 + N(u/c_0)^{3/2}$ for every $u > 0$. Then for every $j > 3/2$ there exists $C = C(j, c_0)$ such that*

$$\sum_{i=1}^N \frac{1}{(1+r_is)^j} \leq k_0 + CN(1+s)^{-3/2} \quad \text{for all } s \geq 0.$$

The hypothesis holds with $k_0 = 0$ if $r_i \geq c_0(i/N)^{2/3}$ for all i , and with $k_0 = 1$ if $r_i \geq c_0((i-1)/N)^{2/3}$ for all $i \geq 2$.

Proof. If $r_i \geq c_0(i/N)^{2/3}$ and $r_i \leq u$, then $i \leq N(u/c_0)^{3/2}$, which gives the hypothesis with $k_0 = 0$; the case $k_0 = 1$ is the same, with $i-1$ in place of i . For $s \leq 1$, the sum is at most $N \leq 2^{3/2}N(1+s)^{-3/2}$. Let $s \geq 1$. For each i , we have $(1+r_is)^{-j} = \int_{r_i}^{\infty} js(1+us)^{-j-1} du$. Summing over i and exchanging the sum and the integral, we obtain

$$\sum_{i=1}^N \frac{1}{(1+r_is)^j} = \int_0^{\infty} \frac{js \#\{i : r_i \leq u\}}{(1+us)^{j+1}} du \leq k_0 + \frac{N}{c_0^{3/2}} \int_0^{\infty} \frac{js u^{3/2}}{(1+us)^{j+1}} du.$$

The substitution $y = us$ shows that the last integral is $s^{-3/2} j \int_0^\infty y^{3/2} (1+y)^{-j-1} dy$, and the integral in y is finite because $j > 3/2$. Since $s^{-3/2} \leq 2^{3/2} (1+s)^{-3/2}$ for $s \geq 1$, this completes the proof. $\square$

The next lemma is the main deterministic input. The comparability (12.12) is a restatement of (H2). The estimates (12.13) and (12.14) use cancellations. For $t \geq 1$ and $r_i t \geq 1$, both r_i and $(2v_i(t))^{-1}$ are comparable to $(1+r_i t)/(1+t)$, whereas the difference $\zeta_i(t)$, which includes the common term $\ell(t)/4$, is bounded by this scale times $C(1+r_i t)^{-2}$.

Lemma 12.4 (Coefficients of the comparison law). *Let $S > 0$ and $C_* \geq 1$, and suppose that $\max_i r_i \leq C_*$ and that r satisfies (H2) on $[0, S]$ with constant C_* . There exists $C = C(C_*)$ such that, for all $t \in [0, S]$ and all i ,*

$$C^{-1} \frac{1+t}{1+r_i t} \leq v_i(t) \leq C \frac{1+t}{1+r_i t}, \quad (12.12)$$

$$|\zeta_i(t)| \leq \frac{C}{(1+t)(1+r_i t)}, \quad (12.13)$$

$$|\zeta'_i(t)| \leq \frac{C}{(1+t)^2(1+r_i t)}. \quad (12.14)$$

If r also satisfies (H3) on $[0, S]$ with constant C_* , then, for all $t \in [0, S]$,

$$\sum_i v_i(t) \zeta_i(t)^2 \leq CN(1+t)^{-5/2}, \quad \sum_i v_i(t) |\zeta'_i(t)| \leq CN(1+t)^{-5/2}. \quad (12.15)$$

Proof. The proof has five steps. We first record bounds on the derivatives of $\bar{h}$. Then we derive an exact equation for v_i , which settles the bounded times and the indices with $r_i t \leq 1$. For $t \geq 1$ and $r_i t \geq 1$, the bound (12.13) comes from an exact cancellation in the Duhamel formula for h_i , and (12.14) from a second-order expansion of h_i combined with the exact equation for v_i . Throughout the proof, C denotes a constant depending only on C_* , whose value may change from line to line. The comparability (12.12) follows from (H2), with $C = C_*^2$, because $v_i = h_i / \bar{h}$.

Step 1: derivatives of $\bar{h}$. For $y \in \mathbb{R}^N$, we have $(\sum_i y_i)^2 \leq \sum_i (2r_i)^{-1} \sum_i 2r_i y_i^2 = 2N \sum_i r_i y_i^2$, by the Cauchy–Schwarz inequality and (12.1). We obtain

$$-4C_* |y|^2 \leq -4 \sum_i r_i y_i^2 \leq y^\top H y = -4 \sum_i r_i y_i^2 + \frac{2}{N} \left(\sum_i y_i \right)^2 \leq 0.$$

Using the spectral theorem, we write $\bar{h}(t) = N^{-1} \langle \mathbf{1}, e^{tH} \mathbf{1} \rangle = \int e^{-\lambda t} m(d\lambda)$, where m is a positive measure supported in $[0, 4C_*]$; namely, m puts mass $N^{-1} \langle \mathbf{1}, u \rangle^2$ at λ for each unit eigenvector u of $-H$ with eigenvalue λ . Differentiating under the integral, we find $(-1)^j \bar{h}^{(j)}(t) = \int \lambda^j e^{-\lambda t} m(d\lambda) \geq 0$; in particular, $\bar{h}$ is nonincreasing. For $t \geq 1$, the

inequality $x^j e^{-x/2} \leq (2j)^j$ gives $\lambda^j e^{-\lambda t} \leq (2j)^j t^{-j} e^{-\lambda t/2}$, so that $|\bar{h}^{(j)}(t)| \leq (2j)^j t^{-j} \bar{h}(t/2)$. For $t \leq 1$, we have $|\bar{h}^{(j)}(t)| \leq (4C_*)^j \bar{h}(t)$. Also, (H2) implies that $\bar{h}(u) \leq C\bar{h}(t)$ whenever $t/4 \leq u \leq t \leq S$. Combining these bounds, we find, for $j = 1, 2, 3$,

$$|\bar{h}^{(j)}(u)| \leq C(1+t)^{-j} \bar{h}(t), \quad t/2 \leq u \leq t \leq S. \quad (12.16)$$

In particular, $|\ell(t)| \leq C(1+t)^{-1}$, and $\ell' = \bar{h}''/\bar{h} - \ell^2$ satisfies $|\ell'(t)| \leq C(1+t)^{-2}$.

Step 2: an exact equation. Differentiating (12.3), we see that the coordinates of h solve

$$h'_i(t) = -4r_i h_i(t) + 2\bar{h}(t), \quad h_i(0) = 1, \quad (12.17)$$

and variation of constants gives

$$h_i(t) = e^{-4r_i t} + 2 \int_0^t e^{-4r_i(t-u)} \bar{h}(u) \, du. \quad (12.18)$$

We divide (12.17) by h_i , subtract ℓ , and use $2\bar{h}/h_i = 2/v_i$. This gives $(\log v_i)' = h'_i/h_i - \ell = -4r_i + 2/v_i - \ell = -4\zeta_i$, that is,

$$v'_i(t) = -4\zeta_i(t)v_i(t), \quad 4\zeta_i(t) = \ell(t) - \frac{h'_i(t)}{h_i(t)}. \quad (12.19)$$

Differentiating (12.6) and using (12.19), we find

$$\zeta'_i = \frac{v'_i}{2v_i^2} + \frac{\ell'}{4} = -\frac{2\zeta_i}{v_i} + \frac{\ell'}{4}. \quad (12.20)$$

Step 3: bounded times and small $r_i t$. Let $t \leq 1$. Then $v_i(t) \geq C^{-1}$ by (12.12), so that $|\zeta_i(t)| \leq r_i + (2v_i)^{-1} + |\ell|/4 \leq C$ and, by (12.20), $|\zeta'_i(t)| \leq C$. These are the claimed bounds, because $(1+t)(1+r_i t) \leq 2(1+C_*)$. Now let $t \geq 1$ and $r_i t \leq 1$. Then $v_i(t) \geq C^{-1}t$, so that $(2v_i)^{-1} \leq C/t$, and also $r_i \leq 1/t$ and $|\ell| \leq C/t$. It follows that $|\zeta_i| \leq C/t$, and (12.20) implies that $|\zeta'_i| \leq 2|\zeta_i|/v_i + |\ell'|/4 \leq C/t^2$. As $1+r_i t \leq 2$ and $1+t \leq 2t$, these are again the claimed bounds.

In the remaining two steps, we fix i and $t \in [1, S]$ with $\xi = r_i t \geq 1$. Recall from (H2) that $h_i(t) \geq C^{-1}\sqrt{1+t}/(1+\xi) \geq C^{-1}\sqrt{t}/\xi$, and that $\bar{h}(u) \leq Ct^{-1/2}$ for $u \in [t/2, t]$.

Step 4: the first cancellation. We split (12.18) at $u = t/2$ and write $h_i = I_1 + I_2$, where

$$I_1(t) = e^{-4r_i t} + 2 \int_0^{t/2} e^{-4r_i(t-u)} \bar{h}(u) \, du, \quad I_2(t) = 2 \int_0^{t/2} e^{-4r_i \tau} \bar{h}(t-\tau) \, d\tau.$$

Each integral has a moving endpoint, and each endpoint contributes $e^{-2\xi} \bar{h}(t/2)$ to the derivative. Using $\bar{h}'(t-\tau) = \ell(t-\tau)\bar{h}(t-\tau)$ and subtracting $\ell(t)h_i(t) = \ell(t)I_1(t) + \ell(t)I_2(t)$, we obtain the exact identity

$$\begin{aligned} h'_i(t) - \ell(t)h_i(t) &= -(4r_i + \ell(t))I_1(t) + 2e^{-2\xi} \bar{h}(t/2) \\ &\quad + 2 \int_0^{t/2} e^{-4r_i \tau} \bar{h}(t-\tau) (\ell(t-\tau) - \ell(t)) \, d\tau. \end{aligned} \quad (12.21)$$

The term $\ell(t)h_i(t)$, common to all coordinates, has been removed exactly; only the variation of ℓ over $[t/2, t]$ remains. We bound the three terms on the right side. For $u \leq t/2$, we have $e^{-4r_i(t-u)} \leq e^{-2\zeta}$, and $\int_0^{t/2} \bar{h}(u) du \leq 2C_*\sqrt{1+t}$ by (H2). This gives $I_1(t) \leq e^{-4\zeta} + 4C_*e^{-2\zeta}\sqrt{1+t} \leq Ce^{-2\zeta}\sqrt{t}$. Using $4r_i = 4\zeta/t$ and $|\ell(t)| \leq C/t \leq C\zeta/t$, we see that the first term on the right side of (12.21) is at most $C\zeta e^{-2\zeta}t^{-1/2}$ in absolute value. The second term is at most $Ce^{-2\zeta}t^{-1/2}$, since $\bar{h}(t/2) \leq Ct^{-1/2}$. For the third term, the bound on ℓ' in Step 1 gives $|\ell(t-\tau) - \ell(t)| \leq C\tau t^{-2}$ for $0 \leq \tau \leq t/2$, and $\bar{h}(t-\tau) \leq Ct^{-1/2}$. Using $\int_0^\infty \tau e^{-4r_i\tau} d\tau = (4r_i)^{-2}$, we find that the third term is at most $Ct^{-5/2}r_i^{-2} = Ct^{-1/2}\zeta^{-2}$. By (12.19), we have $4|\zeta_i(t)| = |h'_i(t) - \ell(t)h_i(t)|/h_i(t)$. Dividing the three bounds by $h_i(t) \geq C^{-1}\sqrt{t}/\zeta$, we conclude that

$$|\zeta_i(t)| \leq \frac{C}{t} \left(\zeta^2 e^{-2\zeta} + \zeta e^{-2\zeta} + \zeta^{-1} \right) \leq \frac{C}{t\zeta},$$

where we used that $\zeta^3 e^{-2\zeta}$ is bounded. Since $(1+t)(1+\zeta) \leq 4t\zeta$, (12.13) follows.

Step 5: the second cancellation. To bound ζ'_i , we cannot differentiate the error terms of Step 4. Instead, we expand h_i to second order and use an exact equation. For $u \in [t/2, t]$, let

$$\varphi(u) = \frac{1}{2r_i} \left(\bar{h}(u) - \frac{\bar{h}'(u)}{4r_i} + \frac{\bar{h}''(u)}{16r_i^2} \right), \quad \eta(u) = h_i(u) - \varphi(u).$$

Using (12.17), we find that the terms with $\bar{h}$, $\bar{h}'$ and $\bar{h}''$ cancel in $\eta' + 4r_i\eta$, so that $\eta'(u) + 4r_i\eta(u) = -\bar{h}'''(u)/(32r_i^3)$. Variation of constants on $[t/2, t]$ gives

$$\eta(t) = e^{-2\zeta}\eta(t/2) - \frac{1}{32r_i^3} \int_{t/2}^t e^{-4r_i(t-u)} \bar{h}'''(u) du.$$

We have $h_i(t/2) \leq C\sqrt{t}/\zeta$ and $\bar{h}(t)/r_i \geq C^{-1}t^{-1/2}r_i^{-1} = C^{-1}\sqrt{t}/\zeta$, by (H2). By (12.16) and $4r_it \geq 4$, we have $|\varphi(t/2)| \leq C\bar{h}(t)/r_i$. Combining the last two bounds, we have $|\eta(t/2)| \leq C\bar{h}(t)/r_i$. Also $|\bar{h}'''(u)| \leq Ct^{-3}\bar{h}(t)$ for $u \in [t/2, t]$, by (12.16). Inserting these bounds into the formula for $\eta(t)$ and using $e^{-2\zeta} \leq C\zeta^{-3}$, we find

$$|\eta(t)| \leq \frac{C\bar{h}(t)}{r_i} e^{-2\zeta} + \frac{C\bar{h}(t)}{r_i^4 t^3} \leq \frac{C\bar{h}(t)}{r_i^4 t^3}.$$

Let $\Theta = 2r_iv_i = 2r_i h_i/\bar{h}$ and $d = \bar{h}''/\bar{h}$. The definitions of φ and η give

$$\Theta = 1 - \frac{\ell}{4r_i} + \frac{d}{16r_i^2} + \epsilon, \quad \epsilon = \frac{2r_i\eta}{\bar{h}}, \quad |\epsilon| \leq \frac{C}{\zeta^3}, \quad (12.22)$$

all evaluated at time t . Since $|\ell| \leq C/t$ and $|d| \leq C/t^2$, we have $|\Theta - 1| \leq C/\zeta$; also $C^{-1} \leq \Theta \leq C$, by (12.12). On the other hand, by (12.19), Θ satisfies the exact equation

$\Theta' = -4\zeta_i\Theta = 4r_i - (4r_i + \ell)\Theta$. Inserting (12.22) into its right side, without differentiating ϵ , and using $\ell' = d - \ell^2$, we find

$$\Theta' = -\frac{\ell'}{4r_i} - \frac{\ell d}{16r_i^2} - (4r_i + \ell)\epsilon.$$

The last two terms are at most $C/(r_i^2 t^3)$ in absolute value, since $|\ell d| \leq Ct^{-3}$ and $(4r_i + |\ell|)|\epsilon| \leq Cr_i\zeta^{-3}$. Differentiating $(2v_i)^{-1} = r_i/\Theta$, we have $((2v_i)^{-1})' = -r_i\Theta'/\Theta^2$, and therefore

$$\zeta'_i = \frac{\ell'}{4} - \left(\frac{1}{2v_i}\right)' = \frac{\ell'}{4} \left(1 - \frac{1}{\Theta^2}\right) + \epsilon_2, \quad |\epsilon_2| \leq \frac{C}{r_i t^3},$$

where $\epsilon_2 = -r_i\Theta^{-2}(\ell d/(16r_i^2) + (4r_i + \ell)\epsilon)$. Using $|1 - \Theta^{-2}| \leq C|\Theta - 1| \leq C/\zeta$ and $|\ell'| \leq Ct^{-2}$, we conclude that $|\zeta'_i(t)| \leq C/(t^2\zeta)$. Since $(1+t)^2(1+\zeta) \leq 8t^2\zeta$, (12.14) follows.

Finally, assume (H3). By (12.12), (12.13), and (12.14), we have

$$\sum_i v_i \zeta_i^2 \leq \frac{C}{1+t} \sum_i \frac{1}{(1+r_i t)^3}, \quad \sum_i v_i |\zeta'_i| \leq \frac{C}{1+t} \sum_i \frac{1}{(1+r_i t)^2},$$

and (H3) bounds both sums by $CN(1+t)^{-3/2}$. This proves (12.15). $\square$

The identities in Steps 2, 4 and 5 are exact consequences of (12.17); they can be checked by expanding both sides in terms of h_i , $\bar{h}$ and the derivatives of $\bar{h}$.

12.3 Exponential moments of quadratic functions

The next lemma controls exponential moments under the comparison law of quadratic functions whose coefficients have the size of the bounds in Lemma 12.4. The constant Λ need not be small, and the coefficients may have either sign.

Lemma 12.5 (Exponential moments under the comparison law). *Let $S > 0$ and $C_* \geq 1$, and suppose that r satisfies (H1)–(H4) on $[0, S]$ with constant C_* . Let $\Lambda \geq 0$ and $s \in [0, S]$, and let $\omega_1, \dots, \omega_N$ be real numbers with*

$$|\omega_i| \leq \frac{\Lambda}{(1+s)(1+r_i s)} \quad \text{for all } i. \quad (12.23)$$

There exists $C = C(\Lambda, C_*)$ such that

$$\log \int_{S_N} \exp \left(\sum_i \omega_i x_i^2 \right) \mu_s(dx) \leq CN(1+s)^{-3/2}. \quad (12.24)$$

Proof. Throughout the proof, C denotes a constant depending only on Λ and C_* . We write $v_i = v_i(s)$ and $T = 1 + s$. Recall from (12.12) that $C^{-1}T/(1 + r_i s) \leq v_i \leq CT/(1 + r_i s)$. Thus $v_i \leq CT$ for all i , and $v_i \geq C^{-1}T$ for the three indices of (H4), since $r_i s \leq r_i S \leq C_*$ for them. Combining the coefficient bound (12.23) with (H3), we obtain

$$\sum_i v_i |\omega_i| \leq C \sum_i \frac{1}{(1 + r_i s)^2} \leq CNT^{-3/2}, \quad \frac{1}{T^2} \sum_i v_i^2 \leq C \sum_i \frac{1}{(1 + r_i s)^2} \leq CNT^{-3/2}. \quad (12.25)$$

We first shift all coefficients by the same amount. On S_N , we have $\sum_i \omega_i x_i^2 = \Lambda N/T + \sum_i (\omega_i - \Lambda/T) x_i^2$. Let $u_i = 2v_i(\Lambda/T - \omega_i)$. The coefficient bound (12.23) implies that $0 \leq u_i \leq 4\Lambda v_i/T \leq C$. Applying (12.10) with the coefficients $\omega_i - \Lambda/T$, for which $1 - 2v_i(\omega_i - \Lambda/T) = 1 + u_i > 0$, we have

$$\log \int e^{\sum_i \omega_i x_i^2} d\mu_s = \frac{\Lambda N}{T} - \frac{1}{2} \sum_i \log(1 + u_i) + \log \frac{p_w(N - \sum_j w_j)}{p_v(0)}, \quad w_i = \frac{v_i}{1 + u_i}. \quad (12.26)$$

The shift ensures that all the variances w_i are positive, whatever the signs of the ω_i .

We bound the last term in (12.26). Since $0 \leq u_i \leq C$, we have $(1 + C)^{-1}v_i \leq w_i \leq v_i$. Then $w_i \geq C^{-1} \max_j w_j$ for the three indices of (H4), and $\sum_i w_i^2 \geq (1 + C)^{-2} \sum_i v_i^2$; the same three indices satisfy $v_i \geq C^{-1} \max_j v_j$. Then Lemma 12.1, applied to v for the lower bound at its mean N and to w for the upper bound on its density, implies that

$$\frac{p_w(N - \sum_j w_j)}{p_v(0)} \leq \frac{C(\sum_i w_i^2)^{-1/2}}{C^{-1}(\sum_i v_i^2)^{-1/2}} \leq C. \quad (12.27)$$

No lower bound for the density of $\sum_i w_i g_i^2$ away from its mean is needed.

We now bound the first two terms in (12.26). Using $\log(1 + u) \geq u - u^2/2$ for $u \geq 0$, and $\sum_i u_i = 2\Lambda N/T - 2\sum_i v_i \omega_i$ (which follows from $\sum_i v_i = N$), we find

$$\frac{\Lambda N}{T} - \frac{1}{2} \sum_i \log(1 + u_i) \leq \frac{\Lambda N}{T} - \frac{1}{2} \sum_i u_i + \frac{1}{4} \sum_i u_i^2 = \sum_i v_i \omega_i + \frac{1}{4} \sum_i u_i^2.$$

The terms $\Lambda N/T$ cancel exactly. Using (12.25) and $u_i^2 \leq 16\Lambda^2 v_i^2/T^2$, we see that the right side is at most $CNT^{-3/2}$. Finally, the logarithm of (12.27) is at most $C \leq CC_* N(1 + s)^{-3/2}$, because $N(1 + s)^{-3/2} \geq N(1 + S)^{-3/2} \geq C_*^{-1}$ by (H1). This proves (12.24). $\square$

12.4 A differential inequality for Rényi divergences

The following scalar lemma closes the differential inequality for the Rényi divergence. It is stated for general probability measures μ_t on S_N ; in our application, μ_t will be the comparison law at time γt .

Lemma 12.6 (A differential inequality for Rényi divergences). *Fix $p > 1$, $C_1 > 0$, $C_2, C_3 \geq 0$, an integer $N \geq 3$, and γ with $0 < \gamma \leq \min\{1, (4pC_1)^{-1}\}$. Let $t_1 > 0$. For $t \in [0, t_1]$, let μ_t be a probability measure on S_N , let f_t be a strictly positive function of class C^2 on S_N with $\int f_t d\mu_t = 1$, and let $U_t \geq 0$ be measurable. Let*

$$\mathcal{R}(t) = \log \int f_t^p d\mu_t, \quad \psi_t = f_t^p e^{-\mathcal{R}(t)}, \quad J(t) = \int \psi_t \Gamma(\log \psi_t) d\mu_t.$$

Assume that $J(t)$ and $\text{Ent}_{\mu_t}(\psi_t)$ are finite, that $\mathcal{R}$ is continuously differentiable on $[0, t_1]$ with $\mathcal{R}(0) = 0$, and that for every $t \in [0, t_1]$,

$$\text{Ent}_{\mu_t}(\psi_t) \leq \frac{C_1(1 + \gamma t)}{4} J(t), \quad (12.28)$$

$$\log \int \exp(pC_1(1 + \gamma t)U_t) d\mu_t \leq C_2 N(1 + \gamma t)^{-3/2}, \quad (12.29)$$

$$\mathcal{R}'(t) \leq -\frac{p-1}{2p} J(t) + (p-1) \int U_t \psi_t d\mu_t + (p-1)C_3 N(1 + \gamma t)^{-5/2}. \quad (12.30)$$

Let $a = (pC_1\gamma)^{-1}$ and $Q = (p-1)[C_2/(pC_1) + C_3]$. Then $a \geq 4$, and for every $t \in [0, t_1]$,

$$\mathcal{R}(t) \leq \frac{Q}{\gamma(a - 3/2)} N(1 + \gamma t)^{-3/2} \leq 2(p-1)(C_2 + pC_1C_3)N(1 + \gamma t)^{-3/2}.$$

Proof. Fix t , and write $\mu = \mu_t$, $\psi = \psi_t$, and $\lambda = pC_1(1 + \gamma t)$. We first bound the forcing term. By Jensen's inequality under the probability measure $\psi\mu$, applied to $\min(U_t, M)$ and followed by the limit $M \rightarrow \infty$, we have

$$\log \int e^{\lambda U_t} d\mu \geq \log \int e^{\lambda U_t - \log \psi} \psi d\mu \geq \lambda \int U_t \psi d\mu - \text{Ent}_{\mu}(\psi).$$

This is the classical duality between relative entropy and exponential moments. Combined with (12.28) and (12.29), this gives

$$\int U_t \psi d\mu \leq \frac{\text{Ent}_{\mu}(\psi)}{\lambda} + \frac{C_2 N(1 + \gamma t)^{-3/2}}{\lambda} \leq \frac{J(t)}{4p} + \frac{C_2}{pC_1} N(1 + \gamma t)^{-5/2}.$$

Inserting this into (12.30), we obtain

$$\mathcal{R}'(t) \leq -\frac{p-1}{4p} J(t) + QN(1 + \gamma t)^{-5/2}. \quad (12.31)$$

Next, we bound $J(t)$ below by $\mathcal{R}(t)$. Since $f_t = e^{\mathcal{R}(t)/p} \psi^{1/p}$ and $\int f_t d\mu = 1$, we have

$$e^{-\mathcal{R}(t)/p} = \int \psi^{1/p} d\mu = \int \exp\left(-\frac{p-1}{p} \log \psi\right) \psi d\mu \geq \exp\left(-\frac{p-1}{p} \text{Ent}_{\mu}(\psi)\right),$$

where the inequality is Jensen's inequality under $\psi\mu$. Taking logarithms, we find $\text{Ent}_\mu(\psi) \geq \mathcal{R}(t)/(p-1)$. Together with (12.28), we find that $J(t) \geq 4\mathcal{R}(t)/((p-1)C_1(1+\gamma t))$, and (12.31) becomes

$$\mathcal{R}'(t) \leq -\frac{\mathcal{R}(t)}{pC_1(1+\gamma t)} + QN(1+\gamma t)^{-5/2} = -\frac{a\gamma \mathcal{R}(t)}{1+\gamma t} + QN(1+\gamma t)^{-5/2}. \quad (12.32)$$

Finally, we integrate. Since $\gamma \leq (4pC_1)^{-1}$, we have $a \geq 4$. Then (12.32) implies that the derivative of $(1+\gamma t)^a \mathcal{R}(t)$ is at most $QN(1+\gamma t)^{a-5/2}$. Integrating from 0 and using $\mathcal{R}(0) = 0$, we have

$$(1+\gamma t)^a \mathcal{R}(t) \leq QN \int_0^t (1+\gamma u)^{a-5/2} du = \frac{QN}{\gamma(a-3/2)} \left((1+\gamma t)^{a-3/2} - 1 \right),$$

which gives the first bound. For the second, note that $\gamma a = (pC_1)^{-1}$ and $a-3/2 \geq 5a/8$, so that $Q/(\gamma(a-3/2)) \leq 8pC_1Q/5 \leq 2(p-1)(C_2 + pC_1C_3)$. $\square$

The next lemma computes the derivative of the Rényi divergence of the time- t law from a time-dependent Bingham law, and estimates it in the form (12.30). The two real numbers a and b are free; they express that adding a common multiple of $\sum_i x_i^2$ to the potential of the reference measure, or to its time derivative, changes nothing on S_N .

Lemma 12.7 (Rényi divergence from a time-dependent Bingham law). *Let $r_1, \dots, r_N > 0$, let $\gamma > 0$ and $t_1 > 0$, and let $k : [0, \gamma t_1] \rightarrow (0, \infty)^N$ be continuously differentiable with $k_i(0) = 1/2$ for all i . For $s \in [0, \gamma t_1]$, let μ_s^k be the probability measure on S_N proportional to $\exp(-\sum_i k_i(s)x_i^2) \nu_N(dx)$. Fix $p > 1$, and for $t \in [0, t_1]$ let*

$$\begin{aligned} f_t &= \frac{d(\nu_N P_t^r)}{d\mu_{\gamma t}^k}, & \mathcal{R}(t) &= \log \int f_t^p d\mu_{\gamma t}^k, \\ \psi_t &= f_t^p e^{-\mathcal{R}(t)}, & J(t) &= \int \psi_t \Gamma(\log \psi_t) d\mu_{\gamma t}^k. \end{aligned}$$

Then, for each t , f_t is a positive function of class C^2 on S_N ; moreover, $\mathcal{R}$ is continuously differentiable on $[0, t_1]$, and $\mathcal{R}(0) = 0$. Moreover, for every $t \in [0, t_1]$ and all real numbers a and b ,

$$\mathcal{R}'(t) \leq -\frac{p-1}{2p} J(t) + (p-1) \int U_t \psi_t d\mu_{\gamma t}^k + (p-1)\gamma \int \sum_i |k'_i(\gamma t) - b| x_i^2 d\mu_{\gamma t}^k, \quad (12.33)$$

where

$$U_t(x) = \sum_i \left(2p(r_i - k_i(\gamma t) + a)^2 + \gamma |k'_i(\gamma t) - b| \right) x_i^2.$$

Proof. We begin with the regularity. Let $g_0 = dv_N/d\pi_r = Z_r e^{F_r}$, a smooth positive function on S_N , and let $g_t = P_t^r g_0$. By the symmetry (3.4) of L_r in $L^2(\pi_r)$, we have $(v_N P_t^r)(\phi) = \pi_r(g_0 P_t^r \phi) = \pi_r(\phi g_t)$ for bounded measurable ϕ , so that $g_t = d(v_N P_t^r)/d\pi_r$. Since L_r is an elliptic operator with smooth coefficients on the compact manifold S_N and g_0 is smooth, the function g_t is of class C^2 in space and C^1 in time on $[0, \infty) \times S_N$, with the time derivative at zero understood from the right, and solves $\partial_t g_t = L_r g_t$; this is standard parabolic regularity, and it also follows by expanding g_0 in eigenfunctions of the self-adjoint operator L_r on $L^2(\pi_r)$, whose coefficients decay faster than any power of the eigenvalues, together with elliptic regularity. Also $g_t \geq \min g_0 > 0$, since P_t^r is a Markov semigroup. Let $W_s(x) = \sum_i k_i(s) x_i^2$. The density of μ_s^k with respect to π_r is $\theta_s = e^{F_r - W_s}/\pi_r(e^{F_r - W_s})$, which is continuously differentiable in s , smooth in x , and positive. Then $f_t = g_t/\theta_{\gamma t}$ is continuously differentiable in t , twice continuously differentiable in x , and positive, and

$$\mathcal{R}(t) = \log \int g_t^p \theta_{\gamma t}^{1-p} d\pi_r.$$

The integrand and its first time derivative are continuous on $[0, t_1] \times S_N$, so compactness permits differentiation under the integral, and $\mathcal{R}$ is continuously differentiable. At $t = 0$, we have $v_N P_0^r = v_N = \mu_0^k$, since $W_0 = N/2$ on S_N ; hence $f_0 = 1$ and $\mathcal{R}(0) = 0$.

We now compute the derivative. Fix t , and write $\mu = \mu_{\gamma t}^k$, $\theta = \theta_{\gamma t}$, $W = W_{\gamma t}$, $f = f_t$, $\psi = \psi_t$, and $\nabla = \nabla_{S_N}$. The derivative of $\int g_t^p \theta_{\gamma t}^{1-p} d\pi_r$ is the sum of

$$p \int f^{p-1} L_r g_t d\pi_r \quad \text{and} \quad (1-p) \int f^p \partial_t \log \theta_{\gamma t} d\mu.$$

For the first term, we use (3.4), $g_t = f\theta$, and $\nabla \theta = \theta \nabla(F_r - W)$:

$$p \int f^{p-1} L_r g_t d\pi_r = -p(p-1) \int f^{p-2} \langle \nabla f, \nabla f + f \nabla(F_r - W) \rangle d\mu.$$

For the second term, $\log \theta_s = F_r - W_s - \log \pi_r(e^{F_r - W_s})$ gives $\partial_s \log \theta_s = -\partial_s W_s + \int \partial_s W_s d\mu_s^k$, and $\partial_t \log \theta_{\gamma t} = \gamma(\partial_s \log \theta_s)|_{s=\gamma t}$. Let $W'(x) = \sum_i k'_i(\gamma t) x_i^2$. Dividing by $e^{\mathcal{R}(t)} = \int f^p d\mu$, and using $f^{p-2} |\nabla f|^2 = f^p |\nabla \log f|^2$ and $\nabla \log \psi = p \nabla \log f$, we obtain the exact identity

$$\begin{aligned} \mathcal{R}'(t) = & -\frac{p-1}{p} J(t) - (p-1) \int \psi \langle \nabla \log \psi, \nabla(F_r - W) \rangle d\mu \\ & + (p-1) \gamma \left(\int \psi W' d\mu - \int W' d\mu \right). \end{aligned} \quad (12.34)$$

Using Young's inequality in the form $|\langle \nabla \log \psi, \nabla(F_r - W) \rangle| \leq (2p)^{-1} |\nabla \log \psi|^2 + (p/2) |\nabla(F_r - W)|^2$, we find that the second term on the right side of (12.34) is at most

$$\frac{p-1}{2p} J(t) + \frac{p(p-1)}{2} \int \psi \Gamma(F_r - W) d\mu.$$

On S_N , we have $F_r - W = \sum_i (r_i - k_i(\gamma t) + a)x_i^2 - aN$, and the constant aN has zero gradient. The tangential gradient of a function on $\mathbb{R}^N$ is the projection of its Euclidean gradient, so that $\Gamma(F_r - W) \leq 4 \sum_i (r_i - k_i(\gamma t) + a)^2 x_i^2$. Similarly, $W' = \sum_i (k'_i(\gamma t) - b)x_i^2 + bN$ on S_N , and the constant bN cancels in the difference of the two integrals in (12.34). We conclude that

$$\int \psi W' d\mu - \int W' d\mu \leq \int \psi \sum_i |k'_i(\gamma t) - b| x_i^2 d\mu + \int \sum_i |k'_i(\gamma t) - b| x_i^2 d\mu.$$

Inserting the last three bounds into (12.34), we obtain (12.33). $\square$

12.5 Density bounds on the time interval $[0, S]$

We now combine the previous lemmas.

Theorem 12.8 (Density bounds on the time interval $[0, S]$). *Let $S > 0$ and $C_* \geq 1$, and suppose that r satisfies (H1)–(H5) on $[0, S]$ with constant C_* . Fix $p > 1$, and let $\gamma = \min\{1, (4pC_*)^{-1}\}$. There exists $C = C(p, C_*)$ such that, for every $t \in [0, S]$,*

$$\log \int_{S_N} \left(\frac{d(v_N P_t^r)}{d\mu_{\gamma t}} \right)^p d\mu_{\gamma t} \leq CN(1+t)^{-3/2}, \quad (12.35)$$

$$\log \int_{S_N} \left(\frac{d(v_N P_t^r)}{d\pi_r} \right)^p d\pi_r \leq CN(1+t)^{-3/2}. \quad (12.36)$$

Further, the function $t \mapsto \int (d(v_N P_t^r)/d\pi_r)^p d\pi_r$ is nonincreasing on $[0, \infty)$, and for every $t \geq 0$,

$$\text{KL}(v_N P_t^r \| \pi_r) \leq \frac{1}{p-1} \log \int_{S_N} \left(\frac{d(v_N P_t^r)}{d\pi_r} \right)^p d\pi_r. \quad (12.37)$$

Proof. We first prove (12.35). We use Lemma 12.7 for the comparison laws and verify the three hypotheses of Lemma 12.6: the differential inequality (12.30) with a reference term of order $N(1+\gamma t)^{-5/2}$, the exponential moment bound (12.29), and the logarithmic Sobolev inequality (12.28). Then we change the reference measure from $\mu_{\gamma t}$ to π_r , and finally we treat later times. The parameters are chosen in the following order: p and C_* are given, then $\gamma = \min\{1, (4pC_*)^{-1}\}$, then the constants of Lemmas 12.1, 12.4, and 12.5, which depend only on C_* and on a constant $\Lambda = \Lambda(p, C_*)$ specified below. In this proof, C denotes a constant depending only on p and C_* , and none of the constants depends on N , S or t . The proof is deterministic.

Step 1: the derivative. For $s \in [0, S]$, let $k_i(s) = (2v_i(s))^{-1}$ be the half-precisions of the comparison law μ_s . The map $s \mapsto k(s)$ is smooth, with values in $(0, \infty)^N$, and $k_i(0) = 1/2$. The measure μ_s^k of Lemma 12.7 is then μ_s . Note that $\gamma t \leq S$ for $t \leq S$, since $\gamma \leq 1$. We apply Lemma 12.7 with $t_1 = S$, $a = \ell(\gamma t)/4$, and $b = \ell'(\gamma t)/4$. Then

$r_i - k_i(\gamma t) + a = \zeta_i(\gamma t)$ and $k'_i(\gamma t) - b = -\zeta'_i(\gamma t)$, by (12.6). With f_t , $\mathcal{R}(t)$, ψ_t , and $J(t)$ as in Lemma 12.7, the inequality (12.33) reads

$$\mathcal{R}'(t) \leq -\frac{p-1}{2p}J(t) + (p-1) \int U_t \psi_t d\mu_{\gamma t} + (p-1)\gamma \sum_i |\zeta'_i(\gamma t)| \int x_i^2 d\mu_{\gamma t},$$

where $U_t(x) = \sum_i (2p\zeta_i(\gamma t)^2 + \gamma|\zeta'_i(\gamma t)|)x_i^2$.

Step 2: the reference term. Let $s = \gamma t$. Using (12.12) and (H4), we find three distinct indices i with $v_i(s) \geq \delta \max_j v_j(s)$, where $\delta \in (0, 1]$ depends only on C_* . Since $\sum_i v_i(s) = N$, we have $\int x_i^2 d\mu_s \leq C v_i(s)$ for every i by Lemma 12.1. By (12.15) and $\gamma \leq 1$, we have

$$\gamma \sum_i |\zeta'_i(s)| \int x_i^2 d\mu_s \leq C \sum_i v_i(s) |\zeta'_i(s)| \leq C_3 N (1 + \gamma t)^{-5/2}, \quad (12.38)$$

for a constant $C_3 = C_3(C_*)$. Thus (12.30) holds with this C_3 .

Step 3: exponential moments of the forcing. Next, (12.13), (12.14), and $\gamma \leq 1$ imply that the coefficients of the quadratic function $pC_*(1+s)U_t$ are at most

$$pC_*(1+s) \left(\frac{2pC_\zeta^2}{(1+s)^2(1+r_is)^2} + \frac{C_\zeta}{(1+s)^2(1+r_is)} \right) \leq \frac{\Lambda}{(1+s)(1+r_is)},$$

where $C_\zeta = C_\zeta(C_*)$ is the constant of Lemma 12.4 and $\Lambda = pC_*C_\zeta(2pC_\zeta + 1)$. By Lemma 12.5, there exists $C_2 = C_2(p, C_*)$ such that

$$\log \int \exp(pC_*(1+\gamma t)U_t) d\mu_{\gamma t} \leq C_2 N (1 + \gamma t)^{-3/2}.$$

This is (12.29) with $C_1 = C_*$.

Step 4: the logarithmic Sobolev inequality. The function ψ_t is positive and of class C^2 , and $\Gamma(\sqrt{\psi_t}) = \psi_t \Gamma(\log \psi_t)/4$. Applying the C^1 extension of (H5) at time γt with $f = \sqrt{\psi_t}$, we have $\text{Ent}_{\mu_{\gamma t}}(\psi_t) \leq C_*(1+\gamma t)J(t)/4$, which is (12.28) with $C_1 = C_*$. All the integrals involved are finite, because their integrands are continuous on the compact manifold S_N .

Step 5: conclusion for the comparison law. We conclude from Steps 1–4 that Lemma 12.6 applies with μ_t there equal to $\mu_{\gamma t}$, $C_1 = C_*$, and the constants C_2, C_3 above. Recall that $\mathcal{R}(0) = 0$ by Lemma 12.7. We conclude that $\mathcal{R}(t) \leq CN(1+\gamma t)^{-3/2}$. Since $\gamma \leq 1$, we have $1 + \gamma t \geq \gamma(1+t)$, and (12.35) follows with a constant $C\gamma^{-3/2}$.

Step 6: change of reference measure. Let $\gamma_2 = \min\{1, (8pC_*)^{-1}\}$, which is the value of γ for the exponent $2p$, and let $s = \gamma_2 t$. We have $d\mu_s/d\pi_r = e^Q/\pi_r(e^Q)$ with $Q(x) = \sum_i \zeta_i(s)x_i^2$, by (12.7). Also $d\pi_r/d\mu_s = e^{-Q}/\mu_s(e^{-Q})$, which gives $\pi_r(e^Q) = 1/\mu_s(e^{-Q})$, and hence

$$\theta := \frac{d\mu_s}{d\pi_r} = \left(\int e^{-Q} d\mu_s \right) e^Q. \quad (12.39)$$

Let $f = d(v_N P_t^r)/d\mu_s$, so that $d(v_N P_t^r)/d\pi_r = f\theta$. By the Cauchy–Schwarz inequality under μ_s ,

$$\int (f\theta)^p d\pi_r = \int f^p \theta^{p-1} d\mu_s \leq \left(\int f^{2p} d\mu_s \right)^{1/2} \left(\int \theta^{2(p-1)} d\mu_s \right)^{1/2}.$$

Using (12.35) with the exponent $2p$, we see that the logarithm of the first integral on the right side is at most $CN(1+t)^{-3/2}$. For the second integral, we have

$$\int \theta^{2(p-1)} d\mu_s = \left(\int e^{-Q} d\mu_s \right)^{2(p-1)} \int e^{2(p-1)Q} d\mu_s.$$

By (12.13), the coefficients of $-Q$ and $2(p-1)Q$ satisfy (12.23) with $\Lambda = \max\{1, 2(p-1)\}C_\zeta$. Then Lemma 12.5 implies that the logarithm of the second integral is at most $CN(1+\gamma_2 t)^{-3/2} \leq C\gamma_2^{-3/2}N(1+t)^{-3/2}$. This proves (12.36).

Step 7: later times and relative entropy. Let $g_t = d(v_N P_t^r)/d\pi_r$. As in the proof of Lemma 12.7, the reversibility of π_r gives $g_{t+u} = P_u^r g_t$ for $t, u \geq 0$. Since P_u^r is a Markov operator, we have $(P_u^r g_t)^p \leq P_u^r(g_t^p)$ by Jensen's inequality, and the invariance of π_r gives

$$\int g_{t+u}^p d\pi_r \leq \int P_u^r(g_t^p) d\pi_r = \int g_t^p d\pi_r.$$

Thus $t \mapsto \int g_t^p d\pi_r$ is nonincreasing. Finally, (12.37) is (12.9). $\square$

Remark 12.9 (The comparison law at the earlier time γt). The comparison law is run at the slower speed γ for the following reason. In (12.32), the dissipation rate $(pC_*(1+\gamma t))^{-1}$ comes from the logarithmic Sobolev constant of $\mu_{\gamma t}$, while the forcing decays like $(1+\gamma t)^{-5/2}$. In the variable $\tau = \gamma t$, the divergence therefore relaxes at rate $a/(1+\tau)$ with $a = (pC_*\gamma)^{-1}$, and integrating the forcing produces the decay $(1+\tau)^{-3/2}$ only when $a > 3/2$. Taking γ small makes a large, at the cost of comparing the time- t law with the comparison law at the earlier time γt .

In Section 17, Theorem 12.8 is applied below the critical temperature, on a time interval of length of order θ_N^{-2} . There the counting bound (H3) follows from Lemma 12.3 with $k_0 = 1$, the extra term being absorbed by the second part of (H1).

12.6 The critical window

We now specialize Theorem 12.8 to the critical window. Recall from Definition 4.16 that a nondecreasing sequence $r_1 \leq \dots \leq r_N$ satisfies the critical spectral conditions with constants (c_0, C_0, δ) if (4.36) holds: $\sum_i (2r_i)^{-1} = N$, $c_0(i/N)^{2/3} \leq r_i \leq C_0(i/N)^{2/3}$ for all i , and $r_2 - r_1 \geq \delta N^{-2/3}$. Some of the results below use only the first two of these conditions, without the ordering and the lower bound on $r_2 - r_1$; we then say that r satisfies the

first two conditions in (4.36) with constants (c_0, C_0) . The ordering and the lower bound on $r_2 - r_1$ enter only through Proposition 8.21. For the model with $c_N = 1 + \sigma N^{-1/3}$, by Lemma 4.17, the half-precisions satisfy the critical spectral conditions at every fixed confidence (in the sense of Section 3.4), simultaneously for all σ in a fixed compact interval, with constants depending only on the confidence and on the interval. All results below are deterministic. In their statements, a constant or a threshold N_0 for the dimension depends only on (c_0, C_0, δ) and on the parameters written as its subscripts or arguments; it does not depend on N , on the time, or on σ . The results therefore apply on the events of Lemma 4.17, uniformly in σ in a compact interval. Table 12.1 lists the parameters in the order in which they are fixed.

Table 12.1. Parameters in the critical window, in the order in which they are fixed.

Parameter	Role	Depends on
c_0, C_0, δ	constants in (4.36)	for the GOE: the confidence and the range of σ (Lemma 4.17)
K	time range $t \leq KN^{2/3}$	fixed
p	order of the moment	fixed
$C_{*,K}$	constant in (H1)–(H5)	c_0, C_0, δ, K
$\gamma_{p,K}$	speed of the comparison law	$p, C_{*,K}$
$C_{p,K}$	constant in the conclusions	p, K, c_0, C_0, δ
$N_0(K)$	threshold for the dimension	c_0, C_0, δ, K ; the bounds hold uniformly in $N \geq N_0(K)$

Lemma 12.10 (The hypotheses in the critical window). *Let $K > 0$ and $S = KN^{2/3}$.*

1. *If r satisfies the first two conditions in (4.36) with constants (c_0, C_0) , then r satisfies (H1)–(H4) on $[0, S]$ with a constant C_* depending only on c_0, C_0 , and K .*
2. *If r satisfies the critical spectral conditions with constants (c_0, C_0, δ) , then there exists $N_0(K)$ such that for $N \geq N_0(K)$, r satisfies (H1)–(H5) on $[0, S]$ with a constant $C_{*,K}$.*

Proof. For (H1), we have $r_i \leq C_0$, and $N(1+S)^{-3/2} \geq N((1+K)N^{2/3})^{-3/2} = (1+K)^{-3/2}$. The heat bounds (H2) are Lemma 7.14, whose constants depend only on c_0, C_0 , and K . The counting bound (H3) is Lemma 12.3 with $k_0 = 0$ and $j = 2$. For (H4), the indices $i = 1, 2, 3$ satisfy $r_i S \leq C_0(3/N)^{2/3}KN^{2/3} = 3^{2/3}C_0K$. This proves the first part. For the second part, (H5) is the bound $C_{LS}(\mu_t) \leq C_K(1+t)$ for $0 \leq t \leq KN^{2/3}$ in the second part

of Proposition 8.21, which holds for $N \geq N_0(K)$; both $N_0(K)$ and C_K depend only on c_0 , C_0 , δ , and K . $\square$

The next corollary restates Lemmas 12.4 and 12.5 in the critical window.

Corollary 12.11 (The comparison law in the critical window). *Suppose that r satisfies the first two conditions in (4.36) with constants (c_0, C_0) , and fix $K > 0$ and $\Lambda \geq 0$. Then (12.12), (12.13), (12.14), and (12.15) hold for $0 \leq t \leq KN^{2/3}$, with a constant depending only on c_0 , C_0 , and K . Further, for $0 \leq s \leq KN^{2/3}$ and all real $\omega_1, \dots, \omega_N$ with $|\omega_i| \leq \Lambda(1+s)^{-1}(1+r_is)^{-1}$,*

$$\log \int_{S_N} \exp \left(\sum_i \omega_i x_i^2 \right) \mu_s(dx) \leq C_{\Lambda, K} N(1+s)^{-3/2}, \quad (12.40)$$

where $C_{\Lambda, K}$ depends only on Λ , K , c_0 , and C_0 .

Proof. By the first part of Lemma 12.10, r satisfies (H1)–(H4) on $[0, KN^{2/3}]$, with a constant depending only on c_0 , C_0 , and K . The corollary therefore follows from Lemmas 12.4 and 12.5. $\square$

Proposition 12.12 (Rényi bounds relative to the comparison law). *Suppose that r satisfies the critical spectral conditions with constants (c_0, C_0, δ) , and fix $p > 1$ and $0 < K < \infty$. There exist $N_0(K)$, $\gamma_{p, K} \in (0, 1]$, and $C_{p, K}$ such that, with $\gamma = \gamma_{p, K}$, for $N \geq N_0(K)$ and $0 \leq t \leq KN^{2/3}$,*

$$\log \int_{S_N} \left(\frac{d(v_N P_t^r)}{d\mu_{\gamma t}} \right)^p d\mu_{\gamma t} \leq C_{p, K} N(1+t)^{-3/2}. \quad (12.41)$$

Proof. By the second part of Lemma 12.10, for $N \geq N_0(K)$, r satisfies (H1)–(H5) on $[0, KN^{2/3}]$ with the constant $C_{*, K}$. The bound (12.41) is then (12.35), with $\gamma_{p, K} = \min\{1, (4pC_{*, K})^{-1}\}$. $\square$

Theorem 12.13 (Density moments in the critical window). *Suppose that r satisfies the critical spectral conditions with constants (c_0, C_0, δ) , and fix $p > 1$ and $0 < K < \infty$. There exist $N_0(K)$ and $C_{p, K}$ such that for $N \geq N_0(K)$,*

$$\log \int_{S_N} \left(\frac{d(v_N P_t^r)}{d\pi_r} \right)^p d\pi_r \leq C_{p, K} N(1+t)^{-3/2}, \quad 0 \leq t \leq KN^{2/3}. \quad (12.42)$$

In particular, for every fixed $s_0 > 0$, there exists C_{p, s_0} such that for $N \geq N_0(s_0)$,

$$\sup_{t \geq s_0 N^{2/3}} \left\| \frac{d(v_N P_t^r)}{d\pi_r} \right\|_{L^p(\pi_r)} \leq C_{p, s_0}.$$

Proof. The bound (12.42) follows from the second part of Lemma 12.10 and (12.36). For the second claim, we apply (12.42) with $K = s_0$ at $t = s_0 N^{2/3}$. Since $N(1 + s_0 N^{2/3})^{-3/2} \leq$

$s_0^{-3/2}$, the logarithm of the p -th moment at this time is at most $C_{p,K}s_0^{-3/2}$ with $K = s_0$, and we may take $C_{p,s_0} = \exp(C_{p,K}s_0^{-3/2}/p)$. The p -th moment is nonincreasing in t by Theorem 12.8, and this gives the bound for all $t \geq s_0N^{2/3}$. $\square$

The applications in Section 19 use Theorem 12.13 with $p = 2$ and $p = 4$; by the triangle inequality, it also bounds $\|d(\nu_N P_t^r)/d\pi_r - 1\|_{L^p(\pi_r)}$ by $C_{p,s_0} + 1$ for $t \geq s_0N^{2/3}$.

Proposition 12.14 (Exponential relaxation in the critical window). *Fix $0 < c_0 \leq C_0$ and $\delta > 0$. For every $a > 0$, there exist $C_a, \kappa > 0$ and N_0 such that, for every $N \geq N_0$ and every r satisfying the critical spectral conditions with constants (c_0, C_0, δ) ,*

$$\|\nu_N P_{sN^{2/3}}^r - \pi_r\|_{\text{TV}} \leq C_a e^{-\kappa(s-a)}, \quad s \geq a. \quad (12.43)$$

The constants and the threshold depend only on c_0, C_0, δ , and a .

Proof. Write $b = N^{2/3}$ and $g_t = d(\nu_N P_t^r)/d\pi_r$. By Theorem 12.13 with $p = 4$, there exist $M_a > 0$ and a dimension threshold such that $\|g_{ab/2}\|_{L^4(\pi_r)} \leq M_a$. Since π_r is a probability measure, this also bounds $\|g_{ab/2}\|_{L^2(\pi_r)}$. By Proposition 8.21, after increasing the threshold, π_r has Poincaré constant at most $K_P b$, where $K_P > 0$ depends only on c_0, C_0, δ .

The density satisfies $g_{t+u} = P_u^r g_t$ and $\int g_t d\pi_r = 1$. At positive times it has the regularity used in the proof of Lemma 12.7. Thus (3.4) and the Poincaré inequality give

$$\frac{d}{dt} \|g_t - 1\|_{L^2(\pi_r)}^2 = -2 \int \Gamma(g_t) d\pi_r \leq -\frac{2}{K_P b} \|g_t - 1\|_{L^2(\pi_r)}^2.$$

Integrating from $ab/2$ to sb , where $s \geq a$, yields

$$\|g_{sb} - 1\|_{L^2(\pi_r)} \leq e^{-(s-a/2)/K_P} \|g_{ab/2} - 1\|_{L^2(\pi_r)} \leq (M_a + 1) e^{-(s-a/2)/K_P}.$$

Finally, Lemma 3.6 and the Cauchy–Schwarz inequality imply

$$\|\nu_N P_{sb}^r - \pi_r\|_{\text{TV}} = \frac{1}{2} \|g_{sb} - 1\|_{L^1(\pi_r)} \leq \frac{M_a + 1}{2} e^{-(s-a)/K_P}.$$

This proves (12.43) with $C_a = (M_a + 1)/2$ and $\kappa = K_P^{-1}$. $\square$

References

- [1] David Aldous. Random walks on finite groups and rapidly mixing Markov chains. In *Séminaire de Probabilités XVII*, volume 986 of *Lecture Notes in Mathematics*, pages 243–297. Springer, Berlin, 1983. doi: 10.1007/BFb0068322.
- [2] David Aldous and Persi Diaconis. Shuffling cards and stopping times. *The American Mathematical Monthly*, 93(5):333–348, 1986. doi: 10.2307/2323590.
- [3] Nima Anari, Kuikui Liu, and Shayan Oveis Gharan. Spectral independence in high-dimensional expanders and applications to the hardcore model. In *2020 IEEE 61st Annual Symposium on Foundations of Computer Science*, pages 1319–1330. IEEE, 2020. doi: 10.1109/FOCS46700.2020.00125.
- [4] Greg W. Anderson, Alice Guionnet, and Ofer Zeitouni. *An Introduction to Random Matrices*, volume 118 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, 2010. doi: 10.1017/CBO9780511801334. URL https://assets.cambridge.org/97805211/94525/excerpt/9780521194525_excerpt.pdf.
- [5] Juhan Aru and Aleksandra Korzhenkova. The spherical model and the large- n limit of the spin $o(n)$ model via the Gaussian free field, 2024. URL <https://arxiv.org/abs/2405.04501v2>. Version 2, revised 3 September 2025.
- [6] Jinho Baik and Ji Oon Lee. Fluctuations of the free energy of the spherical Sherrington-Kirkpatrick model. *Journal of Statistical Physics*, 165(2):185–224, 2016. doi: 10.1007/s10955-016-1610-0. Locators refer to arXiv:1505.07349v5.
- [7] Jinho Baik and Ji Oon Lee. Fluctuations of the free energy of the spherical Sherrington-Kirkpatrick model with ferromagnetic interaction. *Annales Henri Poincaré*, 18(6):1867–1917, 2017. doi: 10.1007/s00023-017-0562-5.
- [8] Jinho Baik and Ji Oon Lee. Free energy of bipartite spherical Sherrington-Kirkpatrick model. *Annales de l’Institut Henri Poincaré, Probabilités et Statistiques*, 56(4):2897–2934, 2020. doi: 10.1214/20-AIHP1062.
- [9] Jinho Baik, Ji Oon Lee, and Hao Wu. Ferromagnetic to paramagnetic transition in spherical spin glass. *Journal of Statistical Physics*, 173(5):1484–1522, 2018. doi: 10.1007/s10955-018-2150-6.
- [10] Jinho Baik, Elizabeth Collins-Woodfin, Pierre Le Doussal, and Hao Wu. Spherical spin glass model with external field. *Journal of Statistical Physics*, 183(2):31, 2021. doi: 10.1007/s10955-021-02757-7.

- [11] D. Bakry and Michel Émery. Diffusions hypercontractives. In *Séminaire de Probabilités XIX, 1983/84*, volume 1123 of *Lecture Notes in Mathematics*, pages 177–206. Springer, 1985. doi: 10.1007/BFb0075847.
- [12] Dominique Bakry, Ivan Gentil, and Michel Ledoux. *Analysis and Geometry of Markov Diffusion Operators*, volume 348 of *Grundlehren der mathematischen Wissenschaften*. Springer, Cham, 2014. doi: 10.1007/978-3-319-00227-9.
- [13] Damien Barbier, Pedro H. de Freitas Pimenta, Leticia F. Cugliandolo, and Daniel A. Stariolo. Finite size effects and loss of self-averageness in the relaxational dynamics of the spherical Sherrington-Kirkpatrick model. *Journal of Statistical Mechanics: Theory and Experiment*, 2021(7):073301, 2021. doi: 10.1088/1742-5468/ac0900. Statement locators refer to v2.
- [14] Gerardo Barrera and Milton Jara. Abrupt convergence for stochastic small perturbations of one dimensional dynamical systems. *Journal of Statistical Physics*, 163(1): 113–138, 2016. doi: 10.1007/s10955-016-1468-1.
- [15] Gerardo Barrera and Milton Jara. Thermalisation for small random perturbations of dynamical systems. *The Annals of Applied Probability*, 30(3):1164–1208, 2020. doi: 10.1214/19-AAP1526.
- [16] Gerardo Barrera, Michael A. Högele, and Juan Carlos Pardo. Cutoff thermalization for Ornstein–Uhlenbeck systems with small Lévy noise in the Wasserstein distance. *Journal of Statistical Physics*, 184:27, 2021. doi: 10.1007/s10955-021-02815-0. Article 27, 54 pages.
- [17] Gerardo Barrera, Michael A. Högele, and Juan Carlos Pardo. The cutoff phenomenon in total variation for nonlinear Langevin systems with small layered stable noise. *Electronic Journal of Probability*, 26(119):1–76, 2021. doi: 10.1214/21-EJP685.
- [18] F. Barthe and C. Roberto. Sobolev inequalities for probability measures on the real line. *Studia Mathematica*, 159(3):481–497, 2003. doi: 10.4064/sm159-3-9.
- [19] M. S. Bartlett. An inverse matrix adjustment arising in discriminant analysis. *The Annals of Mathematical Statistics*, 22(1):107–111, 1951. doi: 10.1214/aoms/1177729698.
- [20] Roland Bauerschmidt and Thierry Bodineau. A very simple proof of the LSI for high temperature spin systems. *Journal of Functional Analysis*, 276(8):2582–2588, 2019. doi: 10.1016/j.jfa.2019.01.007.
- [21] Dave Bayer and Persi Diaconis. Trailing the dovetail shuffle to its lair. *The Annals of Applied Probability*, 2(2):294–313, 1992. doi: 10.1214/aoap/1177005705.
- [22] David Belius and Nicola Kistler. The TAP–Plefka variational principle for the spherical SK model. *Communications in Mathematical Physics*, 367:991–1017, 2019. doi: 10.1007/s00220-019-03304-y.

- [23] Gérard Ben Arous, Amir Dembo, and Alice Guionnet. Cugliandolo–Kurchan equations for dynamics of spin-glasses. *Probability Theory and Related Fields*, 136(4): 619–660, 2006. doi: 10.1007/s00440-005-0491-y. Statement locators refer to v1.
- [24] Gérard Ben Arous and Aukosh Jagannath. Spectral gap estimates in mean field spin glasses. *Communications in Mathematical Physics*, 361(1):1–52, 2018. doi: 10.1007/s00220-018-3152-6.
- [25] Gérard Ben Arous and Aukosh Jagannath. Shattering versus metastability in spin glasses. *Communications on Pure and Applied Mathematics*, 77(1):139–176, 2024. doi: 10.1002/cpa.22133. Locators refer to arXiv:2104.08299v1.
- [26] Gérard Ben Arous, Amir Dembo, and Alice Guionnet. Aging of spherical spin glasses. *Probability Theory and Related Fields*, 120(1):1–67, 2001. doi: 10.1007/PL00008774.
- [27] Gérard Ben Arous, Reza Gheissari, and Aukosh Jagannath. Bounding flows for spherical spin glass dynamics. *Communications in Mathematical Physics*, 373:1011–1048, 2020. doi: 10.1007/s00220-019-03649-4.
- [28] T. H. Berlin and M. Kac. The spherical model of a ferromagnet. *Physical Review*, 86(6):821–835, 1952. doi: 10.1103/PhysRev.86.821.
- [29] S. Bernstein. Démonstration du théorème de weierstrass fondée sur le calcul des probabilités. *Communications de la Société mathématique de Kharkow, 2e série*, 13(1):1–2, 1912. URL <https://www.mathnet.ru/eng/khmo107>.
- [30] Patrick Billingsley. *Convergence of Probability Measures*. Wiley Series in Probability and Statistics. John Wiley and Sons, New York, 2 edition, 1999. doi: 10.1002/9780470316962.
- [31] Christopher Bingham. An antipodally symmetric distribution on the sphere. *The Annals of Statistics*, 2(6):1201–1225, 1974. doi: 10.1214/aos/1176342874.
- [32] Jean-Michel Bismut. *Large Deviations and the Malliavin Calculus*, volume 45 of *Progress in Mathematics*. Birkhäuser, Boston, 1984.
- [33] N. M. Blachman. The convolution inequality for entropy powers. *IEEE Transactions on Information Theory*, 11(2):267–271, 1965. doi: 10.1109/TIT.1965.1053768.
- [34] Vladimir I. Bogachev. *Measure Theory*, volume 1–2. Springer, Berlin, Heidelberg, 2007. doi: 10.1007/978-3-540-34514-5.
- [35] Jeanne Boursier, Djalil Chafaï, and Cyril Labbé. Universal cutoff for Dyson Ornstein Uhlenbeck process. *Probability Theory and Related Fields*, 185(1–2):449–512, 2023. doi: 10.1007/s00440-022-01158-5. Locators refer to arXiv:2107.14452v3.
- [36] 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. *Journal of Functional Analysis*, 22(4):366–389, 1976. doi: 10.1016/0022-1236(76)90004-5.

- [37] Haim Brezis. *Functional Analysis, Sobolev Spaces and Partial Differential Equations*. Universitext. Springer, New York, 2011. doi: 10.1007/978-0-387-70914-7.
- [38] Francesco Caltagirone, Giorgio Parisi, and Tommaso Rizzo. Critical off-equilibrium dynamics in glassy systems. *Physical Review E*, 87(3):032134, 2013. doi: 10.1103/PhysRevE.87.032134.
- [39] Yu Cao, Jianfeng Lu, and Yulong Lu. Exponential decay of Rényi divergence under Fokker–Planck equations. *Journal of Statistical Physics*, 176(5):1172–1184, 2019. doi: 10.1007/s10955-019-02339-8. Statement locators refer to v2.
- [40] Djalil Chafaï and Max Fathi. On cutoff via rigidity for high dimensional curved diffusions. *Comptes Rendus. Mathématique*, 363:1103–1121, 2025. doi: 10.5802/crmath.776. Locators refer to the postpublication arXiv version 2412.15969v4.
- [41] Guan-Yu Chen and Laurent Saloff-Coste. The cutoff phenomenon for ergodic Markov processes. *Electronic Journal of Probability*, 13(3):26–78, 2008. doi: 10.1214/EJP.v13-474.
- [42] Louis H. Y. Chen, Larry Goldstein, and Qi-Man Shao. *Normal Approximation by Stein’s Method*. Probability and Its Applications. Springer, Berlin, Heidelberg, 2011. doi: 10.1007/978-3-642-15007-4.
- [43] Wei-Kuo Chen. The aizenman-sims-starr scheme and Parisi formula for mixed p-spin spherical models. *Electronic Journal of Probability*, 18(94):1–14, 2013. doi: 10.1214/EJP.v18-2580.
- [44] S. Ciuchi and F. de Pasquale. Nonlinear relaxation and ergodicity breakdown in random anisotropy spin glasses. *Nuclear Physics B*, 300:31–60, 1988. doi: 10.1016/0550-3213(88)90585-8.
- [45] Elizabeth Collins-Woodfin. Overlaps of a spherical spin glass model with microscopic external field. *Electronic Journal of Probability*, 26(146):1–22, 2021. doi: 10.1214/21-EJP722.
- [46] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. John Wiley and Sons, 2 edition, 2006. doi: 10.1002/047174882X.
- [47] A. Crisanti and H.-J. Sommers. The spherical p-spin interaction spin glass model: the statics. *Zeitschrift fuer Physik B Condensed Matter*, 87(3):341–354, 1992. doi: 10.1007/BF01309287.
- [48] A. Crisanti, A. Sarracino, and M. Zannetti. Condensation versus ordering: From the spherical models to Bose-Einstein condensation in the canonical and grand canonical ensemble. *Physical Review Research*, 1(2):023022, 2019. doi: 10.1103/PhysRevResearch.1.023022.
- [49] Paul Cuff, Jian Ding, Oren Louidor, Eyal Lubetzky, Yuval Peres, and Allan Sly. Glauber dynamics for the mean-field Potts model. *Journal of Statistical Physics*, 149(3):432–477, 2012. doi: 10.1007/s10955-012-0599-2.

- [50] L. F. Cugliandolo and J. Kurchan. Analytical solution of the off-equilibrium dynamics of a long-range spin-glass model. *Physical Review Letters*, 71(1):173–176, 1993. doi: 10.1103/PhysRevLett.71.173.
- [51] Leticia F. Cugliandolo and David S. Dean. Full dynamical solution for a spherical spin-glass model. *Journal of Physics A: Mathematical and General*, 28(15):4213–4234, 1995. doi: 10.1088/0305-4470/28/15/003. Locators and preprint page numbers refer to arXiv:cond-mat/9502075v1.
- [52] Leticia F. Cugliandolo, Jorge Kurchan, and Luca Peliti. Energy flow, partial equilibration, and effective temperatures in systems with slow dynamics. *Physical Review E*, 55(4):3898–3914, 1997. doi: 10.1103/PhysRevE.55.3898.
- [53] Leticia F. Cugliandolo, David S. Dean, and Hajime Yoshino. Nonlinear susceptibilities of spherical models. *Journal of Physics A: Mathematical and Theoretical*, 40(16):4285–4303, 2007. doi: 10.1088/1751-8113/40/16/003.
- [54] Leticia F. Cugliandolo, Gustavo S. Lozano, Nicolás Nessi, Marco Picco, and Alessandro Tartaglia. Quenched dynamics of classical isolated systems: the spherical spin model with two-body random interactions or the Neumann integrable model. *Journal of Statistical Mechanics: Theory and Experiment*, 2018(6):063206, 2018. doi: 10.1088/1742-5468/aac2fe.
- [55] Amir Dembo, Alice Guionnet, and Christian Mazza. Limiting dynamics for spherical models of spin glasses at high temperature. *Journal of Statistical Physics*, 128(4):847–881, 2007. doi: 10.1007/s10955-006-9239-z. Locators refer to arXiv:math/0609546v1.
- [56] Persi Diaconis. The cutoff phenomenon in finite Markov chains. *Proceedings of the National Academy of Sciences of the USA*, 93(4):1659–1664, 1996. doi: 10.1073/pnas.93.4.1659.
- [57] Persi Diaconis and David Freedman. A dozen de finetti-style results in search of a theory. *Annales de l’Institut Henri Poincaré, Probabilités et Statistiques*, 23(S2):397–423, 1987. URL https://www.numdam.org/item/AIHPB_1987__23_S2_397_0/.
- [58] Persi Diaconis and Mehrdad Shahshahani. Generating a random permutation with random transpositions. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 57(2):159–179, 1981. doi: 10.1007/BF00535487.
- [59] Persi Diaconis, R. L. Graham, and J. A. Morrison. Asymptotic analysis of a random walk on a hypercube with many dimensions. *Random Structures & Algorithms*, 1(1):51–72, 1990. doi: 10.1002/rsa.3240010105.
- [60] Jian Ding, Eyal Lubetzky, and Yuval Peres. The mixing time evolution of Glauber dynamics for the mean-field Ising model. *Communications in Mathematical Physics*, 289(2):725–764, 2009. doi: 10.1007/s00220-009-0781-9.

- [61] Jian Ding, Eyal Lubetzky, and Yuval Peres. Censored Glauber dynamics for the mean field Ising model. *Journal of Statistical Physics*, 137(3):407–458, 2009. doi: 10.1007/s10955-009-9859-1.
- [62] Hang Du and Brice Huang. Overlap distribution of the critical Sherrington-Kirkpatrick model, 2026. URL <https://arxiv.org/abs/2608.08752v1>. Version 1; preprint.
- [63] Ioana Dumitriu and Alan Edelman. Matrix models for beta ensembles. *Journal of Mathematical Physics*, 43(11):5830–5847, 2002. doi: 10.1063/1.1507823.
- [64] Alan Edelman and Brian D. Sutton. From random matrices to stochastic operators. *Journal of Statistical Physics*, 127(6):1121–1165, 2007. doi: 10.1007/s10955-006-9226-4.
- [65] Ahmed El Alaoui, Andrea Montanari, and Mark Sellke. Sampling from the Sherrington–Kirkpatrick Gibbs measure via algorithmic stochastic localization. In *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science*, pages 323–334. IEEE, 2022. doi: 10.1109/FOCS54457.2022.00038.
- [66] Ronen Eldan, Frederic Koehler, and Ofer Zeitouni. A spectral condition for spectral gap: fast mixing in high-temperature Ising models. *Probability Theory and Related Fields*, 182(3–4):1035–1051, 2022. doi: 10.1007/s00440-021-01085-x.
- [67] K. D. Elworthy and Xue-Mei Li. Formulae for the derivatives of heat semigroups. *Journal of Functional Analysis*, 125(1):252–286, 1994. doi: 10.1006/jfan.1994.1124. Author-posted version of the 1994 article; locators refer to arXiv:1911.10971v1.
- [68] Klaus-Jochen Engel and Rainer Nagel. *One-Parameter Semigroups for Linear Evolution Equations*, volume 194 of *Graduate Texts in Mathematics*. Springer, New York, 2000. doi: 10.1007/b97696.
- [69] László Erdős, Horng-Tzer Yau, and Jun Yin. Rigidity of eigenvalues of generalized Wigner matrices. *Advances in Mathematics*, 229(3):1435–1515, 2012. doi: 10.1016/j.aim.2011.12.010. Locators refer to arXiv version 7.
- [70] Lorenzo Farina and Sergio Rinaldi. *Positive Linear Systems: Theory and Applications*. John Wiley and Sons, 2000. doi: 10.1002/9781118033029.
- [71] William Feller. *An Introduction to Probability Theory and Its Applications*, volume 2. John Wiley and Sons, New York, 2 edition, 1971.
- [72] Peter J. Forrester and Eric M. Rains. Interrelationships between orthogonal, unitary and symplectic matrix ensembles. In Pavel Bleher and Alexander Its, editors, *Random Matrix Models and Their Applications*, volume 40 of *Mathematical Sciences Research Institute Publications*, pages 171–208. Cambridge University Press, 2001.
- [73] Yan V. Fyodorov, Anthony Perret, and Grégory Schehr. Large time zero temperature dynamics of the spherical $p = 2$ -spin glass model of finite size. *Journal of Statistical Mechanics: Theory and Experiment*, 2015(11):P11017, 2015. doi: 10.1088/1742-5468/2015/11/P11017.

- [74] Rong Ge, Holden Lee, Jianfeng Lu, and Andrej Risteski. Efficient sampling from the Bingham distribution. In Vitaly Feldman, Katrina Ligett, and Sivan Sabato, editors, *Proceedings of the 32nd International Conference on Algorithmic Learning Theory*, volume 132 of *Proceedings of Machine Learning Research*, pages 673–685. PMLR, 2021. URL <https://proceedings.mlr.press/v132/ge21a.html>.
- [75] Giuseppe Genovese and Daniele Tantari. Legendre duality of spherical and Gaussian spin glasses. *Mathematical Physics, Analysis and Geometry*, 18:10, 2015. doi: 10.1007/s11040-015-9181-x.
- [76] Ivan Gentil. Logarithmic Sobolev inequality for diffusion semigroups, 2010. Lecture notes; locators refer to v1.
- [77] Reza Gheissari and Aukosh Jagannath. On the spectral gap of spherical spin glass dynamics. *Annales de l’Institut Henri Poincaré, Probabilités et Statistiques*, 55(2): 756–776, 2019. doi: 10.1214/18-AIHP897.
- [78] Reza Gheissari, Curtis Grant, and Tianmin Yu. Mixing times of Langevin dynamics for spiked matrix models, 2026. URL <https://arxiv.org/abs/2604.20008v2>. Version 2, 20 May 2026.
- [79] Leonard Gross. Logarithmic Sobolev inequalities. *American Journal of Mathematics*, 97(4):1061–1083, 1975. doi: 10.2307/2373688.
- [80] Jonas Gustavsson. Gaussian fluctuations of eigenvalues in the GUE. *Annales de l’Institut Henri Poincaré, Probabilités et Statistiques*, 41(2):151–178, 2005. doi: 10.1016/j.anihpb.2004.04.002.
- [81] Felix Hausdorff. Summationsmethoden und momentfolgen. i. *Mathematische Zeitschrift*, 9(1–2):74–109, 1921. doi: 10.1007/BF01378337.
- [82] Felix Hausdorff. Summationsmethoden und momentfolgen. ii. *Mathematische Zeitschrift*, 9(3–4):280–299, 1921. doi: 10.1007/BF01279032.
- [83] Jonathan Hermon. A technical report on hitting times, mixing and cutoff, 2015. URL <https://arxiv.org/abs/1501.01869v6>. Version 6, 18 January 2018.
- [84] Richard Holley and Daniel Stroock. Logarithmic Sobolev inequalities and stochastic Ising models. *Journal of Statistical Physics*, 46(5-6):1159–1194, 1987. doi: 10.1007/BF01011161.
- [85] Lars Hörmander. Hypoelliptic second order differential equations. *Acta Mathematica*, 119:147–171, 1967. doi: 10.1007/BF02392081.
- [86] Daniel Hsu, Sham M. Kakade, and Tong Zhang. A tail inequality for quadratic forms of subgaussian random vectors. *Electronic Communications in Probability*, 17(52):1–6, 2012. doi: 10.1214/ECP.v17-2079. URL <https://arxiv.org/abs/1110.2842v1>. Statement locators refer to arXiv version 1.

- [87] Elton P. Hsu. *Stochastic Analysis on Manifolds*, volume 38 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2002. doi: 10.1090/gsm/038.
- [88] Brice Huang, Andrea Montanari, and Huy Tuan Pham. Sampling from spherical spin glasses in total variation via algorithmic stochastic localization, 2024. URL <https://arxiv.org/abs/2404.15651v1>. Version 1, 24 April 2024.
- [89] Brice Huang, Sidhanth Mohanty, Amit Rajaraman, and David X. Wu. Weak Poincaré inequalities, simulated annealing, and sampling from spherical spin glasses. In *Proceedings of the 57th Annual ACM Symposium on Theory of Computing*, pages 915–923. ACM, 2025. doi: 10.1145/3717823.3718236. Locators refer to arXiv:2411.09075v2.
- [90] Vishesh Jain and Clayton Mizgerd. Total variation cutoff for Kac’s walk on the sphere, 2026. URL <https://arxiv.org/abs/2607.13401v1>. Version 1, 15 July 2026.
- [91] Iain M. Johnstone, Yegor Klochkov, Alexei Onatski, and Damian Pavlyshyn. Spin glass to paramagnetic transition and triple point in spherical SK model. *Journal of Statistical Physics*, 191:99, 2024. doi: 10.1007/s10955-024-03296-7. Locators refer to arXiv:2104.07629v4.
- [92] Mark Kac and Colin J. Thompson. Correlation functions in the spherical and mean spherical models. *Journal of Mathematical Physics*, 18(8):1650–1653, 1977. doi: 10.1063/1.523452.
- [93] L. V. Kantorovich. Functional analysis and applied mathematics. *Uspekhi Matematicheskikh Nauk*, 3(6(28)):89–185, 1948. URL https://www.mathnet.ru/php/archive.phtml?jrnid=rm&option_lang=eng&paperid=8775&wshow=paper.
- [94] Tosio Kato. *Perturbation Theory for Linear Operators*. Classics in Mathematics. Springer, Berlin, Heidelberg, 2 edition, 1995. doi: 10.1007/978-3-642-66282-9. Reprint of the 1980 edition.
- [95] R. Z. Khas’minskii. On positive solutions of the equation $\mathfrak{A}u + Vu = 0$. *Teoriya Veroyatnostei i ee Primeneniya*, 4(3):332–341, 1959. doi: 10.1137/1104030. English translation: *Theory of Probability and its Applications* 4 (1959), no. 3, pp. 309–318.
- [96] Alexander V. Kolesnikov and Emanuel Milman. Brascamp–Lieb-type inequalities on weighted Riemannian manifolds with boundary. *The Journal of Geometric Analysis*, 27(2):1680–1702, 2017. doi: 10.1007/s12220-016-9736-5. Locators refer to arXiv:1310.2526v7.
- [97] J. M. Kosterlitz, D. J. Thouless, and Raymund C. Jones. Spherical model of a spin-glass. *Physical Review Letters*, 36(20):1217–1220, 1976. doi: 10.1103/PhysRevLett.36.1217.
- [98] A. Kume and Andrew T. A. Wood. Saddlepoint approximations for the Bingham and Fisher–Bingham normalising constants. *Biometrika*, 92(2):465–476, 2005. doi: 10.1093/biomet/92.2.465.

- [99] Béatrice Lachaud. Cut-off and hitting times of a sample of Ornstein–Uhlenbeck processes and its average. *Journal of Applied Probability*, 42(4):1069–1080, 2005. doi: 10.1239/jap/1134587817.
- [100] Hubert Lacoin. The cutoff profile for the simple exclusion process on the circle. *The Annals of Probability*, 44(5):3399–3430, 2016. doi: 10.1214/15-AOP1053.
- [101] Benjamin Landon. Free energy fluctuations of the two-spin spherical SK model at critical temperature. *Journal of Mathematical Physics*, 63(3):033301, 2022. doi: 10.1063/5.0054298. Locators refer to arXiv version 1.
- [102] Benjamin Landon and Philippe Sosoe. Fluctuations of the overlap at low temperature in the 2-spin spherical SK model. *Annales de l’Institut Henri Poincaré, Probabilités et Statistiques*, 58(3):1426–1459, 2022. doi: 10.1214/21-AIHP1205. Locators refer to arXiv version 1.
- [103] Benjamin Landon and Philippe Sosoe. Fluctuations of the 2-spin SSK model with magnetic field. *The Annals of Applied Probability*, 35(5):3543–3619, 2025. doi: 10.1214/25-AAP2203.
- [104] Lucien Le Cam. Locally asymptotically normal families of distributions. *University of California Publications in Statistics*, 3(2):37–98, 1960.
- [105] David A. Levin and Yuval Peres. *Markov Chains and Mixing Times*. American Mathematical Society, Providence, RI, 2 edition, 2017. ISBN 9781470429621. doi: 10.1090/mbk/107. URL <https://pages.uoregon.edu/dlevin/MARKOV/mcmt2e.pdf>. With contributions by Elizabeth L. Wilmer.
- [106] David A. Levin, Malwina J. Luczak, and Yuval Peres. Glauber dynamics for the mean-field Ising model: cut-off, critical power law, and metastability. *Probability Theory and Related Fields*, 146(1–2):223–265, 2010. doi: 10.1007/s00440-008-0189-z.
- [107] H. W. Lewis and G. H. Wannier. Spherical model of a ferromagnet. *Physical Review*, 88(3):682–683, 1952. doi: 10.1103/PhysRev.88.682.2.
- [108] Eyal Lubetzky and Allan Sly. Information percolation and cutoff for the stochastic Ising model. *Journal of the American Mathematical Society*, 29(3):729–774, 2016. doi: 10.1090/jams/841.
- [109] Jani Lukkarinen. Multi-state condensation in Berlin–Kac spherical models. *Communications in Mathematical Physics*, 373(1):389–433, 2020. doi: 10.1007/s00220-019-03659-2.
- [110] George Marsaglia. Ratios of normal variables and ratios of sums of uniform variables. *Journal of the American Statistical Association*, 60(309):193–204, 1965. doi: 10.1080/01621459.1965.10480783.
- [111] Lloyd A. Metzler. Stability of multiple markets: The hicks conditions. *Econometrica*, 13(4):277–292, 1945. doi: 10.2307/1906922.

- [112] Mohammad Sal Moslehian. Operator Kantorovich type inequality. *RIMS Kôkyûroku*, 1839:119–131, 2013. URL <https://www.kurims.kyoto-u.ac.jp/~kyodo/kokyuroku/contents/pdf/1839-16.pdf>.
- [113] Pierre-Loïc Méliot. The cut-off phenomenon for Brownian motions on compact symmetric spaces. *Potential Analysis*, 40(4):427–509, 2014. doi: 10.1007/s11118-013-9356-7.
- [114] Evita Nestoridi and Sam Olesker-Taylor. Limit profiles for reversible Markov chains. *Probability Theory and Related Fields*, 182(1–2):157–188, 2022. doi: 10.1007/s00440-021-01061-5.
- [115] Vu Lan Nguyen and Philippe Sosoe. Central limit theorem near the critical temperature for the overlap in the 2-spin spherical SK model. *Journal of Mathematical Physics*, 60(10):103302, 2019. doi: 10.1063/1.5065525. Locators refer to arXiv:1809.03675v1.
- [116] Ivan Nourdin and Giovanni Peccati. *Normal Approximations with Malliavin Calculus: From Stein’s Method to Universality*, volume 192 of *Cambridge Tracts in Mathematics*. Cambridge University Press, 2012. doi: 10.1017/CBO9781139084659.
- [117] Frank W. J. Olver, Daniel W. Lozier, Ronald F. Boisvert, and Charles W. Clark, editors. *NIST Handbook of Mathematical Functions*. Cambridge University Press, 2010. ISBN 978-0-521-19225-5.
- [118] Dmitry Panchenko and Michel Talagrand. On the overlap in the multiple spherical SK models. *The Annals of Probability*, 35(6):2321–2355, 2007. doi: 10.1214/009117907000000015.
- [119] Francesco Pedrotti and Justin Salez. The local product condition implies cutoff, 2026. URL <https://arxiv.org/abs/2607.05345v2>. Version 2, 13 July 2026.
- [120] Natesh S. Pillai and Aaron Smith. Kac’s walk on n -sphere mixes in $n \log n$ steps. *The Annals of Applied Probability*, 27(1):631–650, 2017. doi: 10.1214/16-AAP1214.
- [121] José A. Ramírez, Brian Rider, and Bálint Virág. Beta ensembles, stochastic Airy spectrum, and a diffusion. *Journal of the American Mathematical Society*, 24(4):919–944, 2011. doi: 10.1090/S0894-0347-2011-00703-0. Locators refer to arXiv version 4.
- [122] Alfréd Rényi. On measures of entropy and information. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 547–561, Berkeley, 1961. University of California Press.
- [123] Daniel Revuz and Marc Yor. *Continuous Martingales and Brownian Motion*, volume 293 of *Grundlehren der mathematischen Wissenschaften*. Springer, Berlin, 3 edition, 1999. doi: 10.1007/978-3-662-06400-9.
- [124] G. J. Rodgers and M. A. Moore. Distribution of barrier heights in infinite-range spin glass models. *Journal of Physics A: Mathematical and General*, 22(8):1085–1100, 1989. doi: 10.1088/0305-4470/22/8/022.

- [125] O. S. Rothaus. Analytic inequalities, isoperimetric inequalities and logarithmic Sobolev inequalities. *Journal of Functional Analysis*, 64(2):296–313, 1985. doi: 10.1016/0022-1236(85)90079-5.
- [126] Justin Salez. Cutoff for non-negatively curved Markov chains. *Journal of the European Mathematical Society*, 26(11):4375–4392, 2024. doi: 10.4171/JEMS/1348.
- [127] Justin Salez. Cutoff for non-negatively curved diffusions. *Inventiones mathematicae*, 246(1):1–11, 2026. doi: 10.1007/s00222-026-01431-7.
- [128] Laurent Saloff-Coste. Precise estimates on the rate at which certain diffusions tend to equilibrium. *Mathematische Zeitschrift*, 217:641–677, 1994. doi: 10.1007/BF02571965.
- [129] Henry Scheffé. A useful convergence theorem for probability distributions. *The Annals of Mathematical Statistics*, 18(3):434–438, 1947. doi: 10.1214/aoms/1177730390.
- [130] Alberto G. Setti. Eigenvalue estimates for the weighted Laplacian on a Riemannian manifold. *Rendiconti del Seminario Matematico della Università di Padova*, 100:27–55, 1998. URL https://www.numdam.org/item/RSMUP_1998__100__27_0/.
- [131] Jack Sherman and Winifred J. Morrison. Adjustment of an inverse matrix corresponding to a change in one element of a given matrix. *The Annals of Mathematical Statistics*, 21(1):124–127, 1950. doi: 10.1214/aoms/1177729893.
- [132] David Sherrington and Scott Kirkpatrick. Solvable model of a spin-glass. *Physical Review Letters*, 35(26):1792–1796, 1975. doi: 10.1103/PhysRevLett.35.1792.
- [133] Prabodh Shukla and Surjit Singh. Spin glass dynamics in the spherical model. *Journal of Physics C: Solid State Physics*, 14(4):L81–L84, 1981. doi: 10.1088/0022-3719/14/4/004.
- [134] Leon Simon. Introduction to geometric measure theory, 2014. URL <https://web.stanford.edu/class/math285/ts-gmt.pdf>. Author lecture notes.
- [135] Alan D. Sokal. Real-variables characterization of generalized Stieltjes functions. *Expositiones Mathematicae*, 28(2):179–185, 2010. doi: 10.1016/j.exmath.2009.06.004. URL <https://arxiv.org/abs/0902.0065v2>. Equation locators refer to arXiv version 2.
- [136] Alexander Soshnikov. Universality at the edge of the spectrum in Wigner random matrices. *Communications in Mathematical Physics*, 207(3):697–733, 1999. doi: 10.1007/s002200050743.
- [137] Alexander B. Soshnikov. Gaussian fluctuation for the number of particles in Airy, Bessel, sine, and other determinantal random point fields. *Journal of Statistical Physics*, 100(3-4):491–522, 2000. doi: 10.1023/A:1018672622921.

- [138] A. J. Stam. Some inequalities satisfied by the quantities of information of Fisher and Shannon. *Information and Control*, 2(2):101–112, 1959. doi: 10.1016/S0019-9958(59)90348-1.
- [139] A. J. Stam. Limit theorems for uniform distributions on spheres in high-dimensional Euclidean spaces. *Journal of Applied Probability*, 19(1):221–228, 1982. doi: 10.2307/3213932.
- [140] H. E. Stanley. Spherical model as the limit of infinite spin dimensionality. *Physical Review*, 176(2):718–722, 1968. doi: 10.1103/PhysRev.176.718.
- [141] Charles Stein. A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, volume 2, pages 583–602, Berkeley, 1972. University of California Press.
- [142] Eliran Subag. Free energy landscapes in spherical spin glasses. *Duke Mathematical Journal*, 173(7):1291–1357, 2024. doi: 10.1215/00127094-2023-0033.
- [143] Michel Talagrand. Free energy of the spherical mean field model. *Probability Theory and Related Fields*, 134:339–382, 2006. doi: 10.1007/s00440-005-0433-8.
- [144] Lucas Teyssier. Limit profile for random transpositions. *The Annals of Probability*, 48(5):2323–2343, 2020. doi: 10.1214/20-AOP1424.
- [145] Craig A. Tracy and Harold Widom. On orthogonal and symplectic matrix ensembles. *Communications in Mathematical Physics*, 177(3):727–754, 1996. doi: 10.1007/BF02099545.
- [146] A. W. van der Vaart. *Asymptotic Statistics*, volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, 1998. doi: 10.1017/CBO9780511802256.
- [147] Tim van Erven and Peter Harremoës. Rényi divergence and kullback–leibler divergence. *IEEE Transactions on Information Theory*, 60(7):3797–3820, 2014. doi: 10.1109/TIT.2014.2320500. Statement locators refer to v2.
- [148] Santosh S. Vempala and Andre Wibisono. Rapid convergence of the unadjusted Langevin algorithm: isoperimetry suffices, 2019. Version 4, 2 March 2022; statement locators refer to v4.
- [149] Woldemar Voigt. Über das gesetz der intensitätsverteilung innerhalb der linien eines gasspektrums. *Sitzungsberichte der mathematisch-physikalischen Klasse der K. B. Akademie der Wissenschaften zu München*, pages 603–620, 1912. URL <https://publikationen.badw.de/de/003395768>.
- [150] Ding Wang and Lei-Han Tang. Exact finite-size scaling of the maximum likelihood spectra in the quenched and annealed Sherrington-Kirkpatrick spin glass. *SciPost Physics*, 17(4):106, 2024. doi: 10.21468/SciPostPhys.17.4.106. Locators refer to arXiv:2405.14215v2.

- [151] David Vernon Widder. *The Laplace Transform*, volume 6 of *Princeton Mathematical Series*. Princeton University Press, Princeton, 1941.
- [152] Eugene P. Wigner. Characteristic vectors of bordered matrices with infinite dimensions. *Annals of Mathematics*, 62(3):548–564, 1955. doi: 10.2307/1970079.
- [153] Eugene P. Wigner. On the distribution of the roots of certain symmetric matrices. *Annals of Mathematics*, 67(2):325–327, 1958. doi: 10.2307/1970008.
- [154] David Williams. *Probability with Martingales*. Cambridge Mathematical Textbooks. Cambridge University Press, Cambridge, 1991. doi: 10.1017/CBO9780511813658.
- [155] Tingzhou Yu. Analyzing dynamics and average case complexity in the spherical Sherrington–Kirkpatrick model: a focus on extreme eigenvectors, 2024. URL <https://arxiv.org/abs/2401.03668v1>. Version 1, 8 January 2024.
- [156] W. Zippold, R. Kühn, and H. Horner. Non-equilibrium dynamics of simple spherical spin models. *The European Physical Journal B*, 13(3):531–537, 2000. doi: 10.1007/s100510050065. Locators refer to arXiv:cond-mat/9904329v1.