

hexagon:2610.00155v1

The Parisi Formula for the Edwards–Anderson Model in the High-Dimensional Limit

P. M. Aronow

Patrick Lopatto

AUTHORS' NOTE

With the exception of this note, the entire document is machine written. We have taken care to supervise this writing to ensure that the proofs meet a minimum standard of readability and that previous literature is appropriately cited. However, the resulting document 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.

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

ABSTRACT

We study the Edwards–Anderson spin glass on the torus $(\mathbb{Z}/L\mathbb{Z})^d$, with independent Gaussian couplings of variance $1/(2d)$ between nearest neighbors. We prove that, as $d \rightarrow \infty$, its free energy converges to that of the Sherrington–Kirkpatrick model, given by the Parisi formula, at every temperature and in every uniform external field, uniformly in the side length L . The same holds on the hypercube $\{0,1\}^d$, and the ground-state energies converge to the Sherrington–Kirkpatrick ground-state energy. The upper bound holds on every regular graph of large degree. Its proof follows Mourrat’s Hamilton–Jacobi approach for an enriched free energy, and it controls the negative part of the spectrum of the adjacency matrix with Panchenko’s synchronization of overlaps, in a finitary form due to Mourrat. For the lower bound, we construct a vector of magnetizations by message passing, and we show that the pressure is at least its Thouless–Anderson–Palmer free energy using a sublinear relative-entropy bound for the associated planted model.

Contents

1	Introduction	1
1.1	Background	1
1.2	Main results	3
1.3	Ideas of the proofs	6
1.4	Related work	9
1.5	Organization	11
2	Preliminaries	13
2.1	The model	13
2.2	Classical tools	15
2.3	The Sherrington–Kirkpatrick model and the Parisi formula	16
3	Finite cascades and the scalar transform	23
3.1	Notation	23
3.2	Cascade recursions	24
3.3	The scalar transform	26
3.4	Complete monotonicity	27
3.5	Square-root convexity.	32
3.6	The Hopf–Lax formula and the Parisi functional.	35
4	The upper bound on regular graphs	37
4.1	Hamiltonians on the cone and comparison	38
4.2	The subsolution	42
4.3	The enriched free energy	44
4.4	Synchronization	49
4.5	Gaussian perturbations and the envelope	50
4.6	Sparsification and the drift	52
4.7	Proof of Proposition 4.1.	54
5	Message-passing centers on tori and hypercubes	56
5.1	Message passing with deterministic Onsager coefficients	58
5.2	Two imported results	59
5.3	Tame polynomials and coefficient replacement	61
5.4	Proof of the state evolution	64
5.5	Energy of the final message.	68
5.6	The two-phase scheme and controls.	69

5.7	The Parisi control	73
5.8	Homogenization	76
5.9	Proof of Proposition 5.2.	79
6	Gaussian conditioning and the Onsager term	81
6.1	The planted model and the main result	83
6.2	Gaussian conditioning	84
6.3	Expansion around the center	86
6.4	A band of typical configurations	87
6.5	Relative entropy of the restricted and projected model.	90
6.6	Proof of the main result.	93
7	Relative entropy of the planted product model	95
7.1	Gaussian channels	97
7.2	The scalar transform	99
7.3	The scalar criterion for the Parisi center	101
7.4	The relative-entropy bound	103
7.5	Interpolating channels	104
7.6	The score covariance identity	106
7.7	Concentration	107
7.8	The time derivative at contact points	109
7.9	The scalar Hopf–Lax formula	111
7.10	Proof of the relative-entropy bound	113
8	Proofs of the main theorems	118
8.1	The lower bound at a fixed side length	118
8.2	The side length	119
8.3	Proofs of the main results.	121
	References	124

Introduction

1.1 Background

The Edwards–Anderson (EA) model [58] is the basic model of a short-range spin glass. Ising spins $\sigma_x \in \{-1, 1\}$ sit at the sites of a lattice, and each pair of neighboring spins interacts through an independent centered random coupling, so that some interactions favor alignment and others favor opposition. Its mean-field counterpart is the Sherrington–Kirkpatrick (SK) model [127], in which every pair of spins interacts. For the SK model, Parisi’s theory of replica symmetry breaking [118, 119] has to a large extent been made rigorous. The free energy is given by the Parisi formula [76, 134], its minimizer is unique [7], and Panchenko proved that overlap arrays satisfying the Ghirlanda–Guerra identities are ultrametric [113, 114]. For the EA model in a fixed dimension, by contrast, the existence of a spin-glass phase at positive temperature remains open. At zero temperature, Chatterjee has proved several signatures of glassy behavior, including disorder chaos and low-energy macroscopic excitations [30]. There are competing physical descriptions of the low-temperature states: the replica-symmetry-breaking picture [102], the droplet picture [68], and intermediate pictures such as the trivial–nontrivial picture [95, 111]; see [129] for a discussion. The pictures disagree in particular about a uniform external field: replica symmetry breaking predicts a phase transition at a de Almeida–Thouless line [53, 72], whereas the droplet picture predicts that there is no transition in any positive field.

A natural way to relate the two models is to let the number of neighbors of each spin grow. On the torus $(\mathbb{Z}/L\mathbb{Z})^d$ with nearest-neighbor couplings of variance $1/(2d)$, the local field $\sum_{y \sim x} J_{xy} \sigma_y$ acting on a spin, for a fixed configuration, is a sum of $2d$ small independent terms, as in the SK model. The variance of the Hamiltonian per site is $1/2$, which agrees with that of the SK model up to its finite-size correction. Mean-field theory is expected to describe the model increasingly well as d grows. In zero field, the critical behavior is predicted to be that of mean-field theory above the upper critical dimension six [82], and the low-temperature phase of the EA model on $\mathbb{Z}^d$ has been studied through an expansion in powers of $1/d$ around mean-field theory [72]. The EA model on diluted hypercubes of large dimension has also been studied numerically as a mean-field spin glass [67]. In particular, the free energy per site of the EA model is expected to converge to that of the SK model as $d \rightarrow \infty$, at every temperature and in every field.

Rigorous results on this limit have so far required high temperature, interactions of diverging range, or graphs that are random or locally tree-like. For $\beta \leq 1$ in zero field, the SK free energy equals the annealed value $\log 2 + \beta^2/4$ [1]. The same limit holds on any regular graph of diverging degree. For $\beta < 1$, this graph extension follows from the second-moment method and the Gaussian concentration of the free energy. On a D -regular graph with n vertices, the ratio of the second moment of the partition function to the square of its mean is at most $\det(I - \beta^2 S_+)^{-1/2}$, where S_+ is the positive part of the adjacency matrix divided by D . Since $\text{tr } S_+ \leq n/\sqrt{D}$, this ratio is $e^{o(n)}$ as $D \rightarrow \infty$. The case $\beta = 1$ follows by monotonicity in β . In every field, Hachem recently proved that for $\beta^2 < \log 2$ the free energy of the SK model with a sparse doubly stochastic variance profile, which includes the EA model on regular graphs of diverging degree, converges to the replica-symmetric SK value [81, Corollary 3]. Fröhlich and Zegarlinski [71] proved high-temperature mean-field limits of short-range spin glasses as the interaction range tends to infinity. In the Kac limit, the dimension is fixed, the strength of the couplings decays to zero over a distance γ^{-1} , and $\gamma \rightarrow 0$ after the thermodynamic limit. For positive-semidefinite interaction kernels, Franz and Toninelli [69, 70] proved that the free energy converges to the SK free energy. One of the two inequalities, proved earlier by Guerra and Toninelli [77], requires the interaction kernel to be positive semidefinite. For sparse random graphs, Dembo, Montanari, and Sen [54] showed that the maximum cut of random regular and Erdős–Rényi graphs of large average degree is governed by the SK ground-state energy. Their comparison uses the randomness of the graph. It was extended to other optimization problems on sparse random hypergraphs by Sen [126], and to inhomogeneous random graphs, whose large-degree limit is an inhomogeneous Potts spin glass, by Jagannath, Ko, and Sen [89]. El Cheairi and Gamarnik showed that algorithms given by low-degree polynomials with a certain tree structure perform equally well on the SK model and on sparse Erdős–Rényi graphs of large average degree [63]. Approximating message passing for the SK model by such polynomials, they obtained nearly maximal cuts on these graphs, assuming that the support of the Parisi measure is an interval $[0, q_*]$ for all large β . On deterministic regular graphs of large degree and large girth, El Alaoui, Montanari, and Sellke [62] constructed cuts of the same size by local algorithms, under a similar assumption at zero temperature.

None of these methods applies to a deterministic lattice at low temperature. The lattice is not locally tree-like: it has many short cycles. The methods that prove the Parisi formula, namely Guerra’s interpolation [76], the Aizenman–Sims–Starr scheme [2], and Panchenko’s analysis of the overlap distribution [114], use the fact that the covariance of the SK Hamiltonian is a function of the overlap of two configurations. On the lattice, the covariance of $H(\sigma)$ and $H(\tau)$ is instead a function of the fraction of edges xy with $\sigma_x \sigma_y = \tau_x \tau_y$. Further, the nearest-neighbor lattice with an even side length is bipartite, and flipping all spins on one sublattice reverses the sign of the Hamiltonian. Its normalized

adjacency matrix then has the eigenvalue -1 , and the comparison arguments that require this matrix to be close to positive semidefinite do not apply.

The goal of this paper is to prove that the free energy of the EA model on the torus $(\mathbb{Z}/L\mathbb{Z})^d$ converges to the SK free energy as $d \rightarrow \infty$, at every temperature and in every uniform field, uniformly in the side length L . We prove the same statement for the hypercube $\{0,1\}^d$.

1.2 Main results

The model. Throughout, $G = (V, E)$ is a finite simple D -regular graph with $n = |V|$ vertices and $D \geq 1$. It need not be connected. We write $\mathcal{G}_D$ for the set of all such graphs. Given G , let $(g_{xy})_{xy \in E}$ be independent standard Gaussian variables and $J_{xy} = g_{xy}/\sqrt{D}$. For $\beta \geq 0$ and $h \in \mathbb{R}$, the Hamiltonian, the pressure, and the ground-state energy per site are

$$\begin{aligned} H_G(\sigma) &= \sum_{xy \in E} J_{xy} \sigma_x \sigma_y, & p_G(\beta, h) &= \frac{1}{n} \mathbb{E} \log \sum_{\sigma \in \{-1,1\}^V} \exp \left(\beta H_G(\sigma) + h \sum_{x \in V} \sigma_x \right), \\ e(G; h) &= \frac{1}{n} \mathbb{E} \max_{\sigma \in \{-1,1\}^V} \left(H_G(\sigma) + h \sum_{x \in V} \sigma_x \right). \end{aligned} \quad (1.1)$$

As in the literature on the SK model, the field enters the pressure through the Gibbs exponent and is not multiplied by β , so that $p_G(\beta, \beta h)/\beta \rightarrow e(G; h)$ as $\beta \rightarrow \infty$. The normalization gives $\mathbb{E} H_G(\sigma)^2 = |E|/D = n/2$ for every σ , and $\sum_{y \sim x} \mathbb{E} J_{xy}^2 = 1$ for every vertex x . Our main examples are the following.

- The torus $\mathbb{T}_L^d = (\mathbb{Z}/L\mathbb{Z})^d$, in which two vertices are adjacent if they differ by ± 1 in exactly one coordinate, is simple and $2d$ -regular for $L \geq 3$, with $n = L^d$. On it, H_G is the EA Hamiltonian with couplings $g_{xy}/\sqrt{2d}$. We write $p_{d,L} = p_{\mathbb{T}_L^d}$.
- The hypercube $Q_d = \{0,1\}^d$, in which two vertices are adjacent if they differ in exactly one coordinate, is d -regular with $n = 2^d$.

No relation between n and D is assumed. For example, $D = 2d$ is proportional to $\log n$ on $\mathbb{T}_3^d$, and $D = \log_2 n$ on Q_d .

The SK model. The SK Hamiltonian on N spins is $H_N^{\text{SK}}(\sigma) = N^{-1/2} \sum_{1 \leq i < j \leq N} g_{ij} \sigma_i \sigma_j$, with independent standard Gaussian variables g_{ij} . Its variance $(N-1)/2$ agrees with that of H_G up to the factor $(N-1)/N$. The limits

$$\mathcal{P}(\beta, h) = \lim_{N \rightarrow \infty} \frac{1}{N} \mathbb{E} \log \sum_{\sigma} e^{\beta H_N^{\text{SK}}(\sigma) + h \sum_i \sigma_i}, \quad \mathcal{P}_*(h) = \lim_{N \rightarrow \infty} \frac{1}{N} \mathbb{E} \max_{\sigma} \left(H_N^{\text{SK}}(\sigma) + h \sum_i \sigma_i \right) \quad (1.2)$$

exist. The first exists by [75, 134] and is given by the Parisi formula [76, 134], which we recall in Section 2.3. The second exists by a comparison with the first (see (2.5)),

and $P_*(h) = \lim_{\beta \rightarrow \infty} \mathcal{P}(\beta, \beta h) / \beta$. Numerical evaluations of the Parisi formula give $P_*(0) \approx 0.7632$ [51]; see also [54].

Results. Our main theorem states that the pressure of the EA model on the torus converges to the SK pressure, uniformly in the side length.

Theorem 1.1. *For every $\beta \geq 0$ and $h \in \mathbb{R}$,*

$$\lim_{d \rightarrow \infty} \sup_{L \geq 3} |p_{d,L}(\beta, h) - \mathcal{P}(\beta, h)| = 0.$$

The same holds on the hypercube.

Theorem 1.2. *For every $\beta \geq 0$ and $h \in \mathbb{R}$,*

$$\lim_{d \rightarrow \infty} p_{Q_d}(\beta, h) = \mathcal{P}(\beta, h).$$

Letting the temperature tend to zero, we obtain the corresponding statement for ground states.

Corollary 1.3. *For every $h \in \mathbb{R}$,*

$$\lim_{d \rightarrow \infty} \sup_{L \geq 3} |e(\mathbb{T}_L^d; h) - P_*(h)| = 0 \quad \text{and} \quad \lim_{d \rightarrow \infty} e(Q_d; h) = P_*(h).$$

We make several remarks on these results.

Remark 1.4. The pressures p_G and $\mathcal{P}$ are 1-Lipschitz in h and $(B/2)$ -Lipschitz in $\beta \in [0, B]$ (Lemma 2.1). Then the convergence in Theorems 1.1 and 1.2 is uniform for (β, h) in compact subsets of $[0, \infty) \times \mathbb{R}$, and the convergence in Corollary 1.3 is uniform for h in compact subsets of $\mathbb{R}$. We give the short argument at the end of Section 8.

Remark 1.5. Theorem 1.1 imposes no relation between L and d , and it covers both even side lengths, for which the torus is bipartite, and odd side lengths. In particular, the limits may be taken in either order. For every fixed d , the limit $p_d^\infty(\beta, h) = \lim_{L \rightarrow \infty} p_{d,L}(\beta, h)$ exists, and it is also the limit of the pressures of the boxes $\{0, \dots, N-1\}^d$ with free boundary conditions (Lemma 8.3). Then Theorem 1.1 implies that $\lim_{d \rightarrow \infty} p_d^\infty(\beta, h) = \mathcal{P}(\beta, h)$; this is the statement that the free energy of the EA model on $\mathbb{Z}^d$, normalized as above, converges to the SK free energy.

Remark 1.6. Theorem 1.1 also identifies the limit of the link overlap. Let σ^1 and σ^2 be two independent samples from the Gibbs measure on $\mathbb{T}_L^d$, and let $Q = |E|^{-1} \sum_{xy \in E} \sigma_x^1 \sigma_x^2 \sigma_y^1 \sigma_y^2$ be their link overlap. By Gaussian integration by parts, $\partial_\beta p_{d,L} = (\beta/2)(1 - \mathbb{E}\langle Q \rangle)$, and the derivative of the SK pressure has the same form, with the square of the site overlap $R = N^{-1} \sum_i \sigma_i^1 \sigma_i^2$ in place of Q (Lemma 2.1). These pressures are convex in β , and

the derivatives of convex functions that converge pointwise converge at every point where the limit is differentiable. Thus, at every $\beta > 0$ at which $\mathcal{P}(\cdot, h)$ is differentiable, $\mathbb{E}\langle Q \rangle$ on $\mathbb{T}_L^d$ converges, uniformly in L , to the limit of $\mathbb{E}\langle R^2 \rangle$ in the SK model. Since $\mathcal{P}(\cdot, h)$ is convex, this holds for all but countably many β . For the EA model in a fixed dimension, Contucci and Giardinà proved, in temperature average, Ghirlanda–Guerra identities for the link overlap [43], and Contucci, Mingione, and Starr proved distributional identities after suitable generic Gaussian perturbations whose covariances are powers of the link overlap [48]. By Panchenko’s theorem [113], these distributional Ghirlanda–Guerra identities yield ultrametricity of the limiting link-overlap arrays of the generically perturbed models.

Remark 1.7. The proof of the lower bound uses a recent theorem on the support of the Parisi measure in a positive field [99], which we recall in Theorem 2.2: for $h > 0$ the support is an interval $[q_-, q_+]$ with $q_- > 0$. The case $h = 0$ follows from the case $h > 0$ by continuity in h . In zero field, Parisi predicted that the support is an interval $[0, q_+]$ for $\beta > 1$ (full replica symmetry breaking); this was proved by Zhou for β close to 1 [141] and for every $\beta > 1$ in [100]. At zero temperature, the Stieltjes measure associated with the canonical Parisi order parameter has infinitely many points in its support [10], and its support is $[0, 1)$ [32]. We do not use these results. The upper bound uses no information on the support of the Parisi measure.

The upper bound in Theorems 1.1 and 1.2 holds for every regular graph of large degree: Proposition 4.1 states that

$$\limsup_{D \rightarrow \infty} \sup_{G \in \mathcal{G}_D} p_G(\beta, h) \leq \mathcal{P}(\beta, h) \quad (1.3)$$

for every $\beta \geq 0$ and $h \in \mathbb{R}$. The supremum is over all finite simple D -regular graphs, with no relation between n and D and no assumption of transitivity or of a spectral condition. The bound is attained, for example, by the complete graphs K_{D+1} . The lower bound uses the structure of tori and hypercubes, through a homogenization statement for the output of a message-passing algorithm (Section 5). We do not know whether the lower bound holds for every sequence of regular graphs of diverging degree.

Remark 1.8. The complete bipartite graph $K_{D,D}$ is D -regular, and the model on it is the bipartite SK model with two species of D spins each. For it, the upper bound (1.3) is attained in every field: for every $\beta \geq 0$ and $h \in \mathbb{R}$,

$$\lim_{D \rightarrow \infty} p_{K_{D,D}}(\beta, h) = \mathcal{P}(\beta, h).$$

The lower bound follows from a Gaussian interpolation with the SK model on $2D$ spins, since the normalized adjacency matrix of $K_{D,D}$ is at most the projection onto the constant

vectors; we give the proof in Section 8.3. In the limit $D \rightarrow \infty$, this lower bound is a special case of the lower bound of Bates and Sohn for balanced multispecies models [19, Theorem 1.3], which holds in every field [19, Remark 1.8]. For the bipartite SK model in zero field, a Gaussian interpolation with independent SK models on the two species, which gives a lower bound of this kind, was used by Barra, Genovese, and Guerra [16, Section 3.1], and for models with exchangeable species this comparison was proved by Issa [86, Theorem 38]. In zero field, the limit of the free energy of balanced multispecies models, which include this one, was identified in [37, 83]; see Section 1.4.

To our knowledge, for interactions of fixed range, no deterministic family of lattices was previously known on which the free energy of the EA model at low temperature, or its ground-state energy, converges to the SK value. Our results concern the limit in which the dimension tends to infinity, taken after or jointly with the thermodynamic limit; they do not decide between the competing pictures of the EA model in a fixed dimension.

1.3 Ideas of the proofs

The upper and lower bounds use different methods. We describe them in turn.

The upper bound. The interpolation identity recalled in Section 1.4 gives the upper bound with an error of at most $\beta^2\epsilon/4$ when $\lambda_{\min}(S) \geq -\epsilon$, where $S = A_G/D$. On bipartite graphs, S has the eigenvalue -1 , so this error does not vanish. We follow instead the Hamilton–Jacobi approach to the Parisi formula developed by Mourrat and coauthors [56, 107, 109], and in particular the strategy of Mourrat’s upper bounds for models with nonconvex interactions [106, 108]. There, an enriched free energy is shown to be an approximate viscosity supersolution of a Hamilton–Jacobi equation, with Ghirlanda–Guerra identities and synchronization enforced at the points where a test function touches it. Fix a finite Ruelle cascade. We give each vertex x a Gaussian field indexed by the leaves of the cascade, whose covariance is determined by a nondecreasing path q_x , and we let the couplings be independent across the first-level branches of the cascade. Let $F(t, q)$ be minus the free energy of the resulting model at inverse temperature $\beta\sqrt{t}$, plus $\beta^2 t/4$. At $q = 0$ and $t = 1$, it is at most $\beta^2/4 - p_G(\beta, h)$, and at $t = 0$ it is an explicit function of the paths. Its time derivative equals a quadratic Hamiltonian, evaluated at the gradient of F in q , plus a term of the form $\beta^2(4n)^{-1} \sum_j w_j \text{tr}(SC_j)$, where C_j is the covariance matrix of $\sigma^1 \odot \sigma^2$ given that the two replicas meet at depth j of the cascade. For the SK model the analogous term is nonnegative, and Guerra’s bound follows. Here S has negative eigenvalues, and the term has no sign.

We show instead that it is small. For every $\eta > 0$, we have $|\text{tr}(SC_j)| \leq \eta \text{tr} C_j + (4\eta)^{-1} \text{tr}(S^2 C_j)$, with $\text{tr} C_j \leq n$. Thus it suffices to control $\text{tr}(S^2 C_j)$, a sum of conditional variances of the row overlaps $(S(\sigma^1 \odot \sigma^2))_x$, $x \in V$. A sparsification theorem of Cohen, Nelson, and Woodruff [42] reduces this sum to a set of $O(n/D)$ rows. For these rows,

small Gaussian perturbations of the Hamiltonian enforce approximate Ghirlanda–Guerra identities for the pair formed by the overlap in the cascade and the row overlap. A finitary form of Panchenko’s synchronization theorem [115], due to Mourrat [106], then bounds the conditional variance of each of these row overlaps, given the overlap in the cascade, by a multiple of the mesh of the cascade. The perturbations change the free energy by $O(s^2/D)$ for a perturbation of strength s , and the identities require $s \rightarrow \infty$; this is where we use that $D \rightarrow \infty$. The independence of the couplings across first-level branches makes the fluctuations of the free energy of order $1/n$. This allows us to enforce the identities at every point where a test function touches a deterministic regularization of F , and we find that this regularization is, up to small errors, a viscosity supersolution of a Hamilton–Jacobi equation on the cone of V -indexed nondecreasing paths, with the quadratic Hamiltonian $\beta^2(4n)^{-1} \sum_j w_j p_j \cdot S p_j$. A comparison principle bounds it below by a subsolution built from the Hopf–Lax formula for the SK model. That this function is a subsolution of the equation with the indefinite matrix S follows from the inequality between the arithmetic and the harmonic mean, which uses only that S has nonnegative entries and unit row sums. Its initial value lies below that of F because the free energy of a finite cascade is convex in the square roots of its variances, which we derive in Section 3 from complete monotonicity. Finally, the value of the Hopf–Lax formula at time one is given by the Parisi formula, as observed by Mourrat for $h = 0$ [107]. A comparison of this kind with a one-species Hamilton–Jacobi equation was used by Chen, Issa, and Mourrat [37, Appendix B] for balanced multispecies models in zero field; see also [86, Section 7]. There the initial condition is convex in the paths themselves [37, 83]. This convexity fails in a strong field [106, Section 6.1], whereas the convexity in the square roots holds in every field.

The lower bound. For $m \in [-1, 1]^V$ and $\kappa_x = 1 - m_x^2$, let

$$\Theta_G(m) = \beta H_G(m) + h \sum_{x \in V} m_x + \sum_{x \in V} \text{ent}(m_x) + \frac{\beta^2}{4} \langle \kappa, S \kappa \rangle,$$

where $S = A_G/D$. This is the Thouless–Anderson–Palmer (TAP) free energy [135] of the graph, that is, the expansion to second order in the couplings of the free energy at fixed magnetizations [73, 121], with each squared coupling replaced by its mean $1/D$. Jensen’s inequality for the product measure on $\{-1, 1\}^V$ with means m gives $\log Z_G \geq \Theta_G(m) - (\beta^2/4) \langle \kappa, S \kappa \rangle$ for every m , where Z_G is the partition function. The last term of $\Theta_G(m)$, the Onsager correction, is not visible to this argument, and it is of the same order as the other terms. The proof of the lower bound has three steps.

First, we construct m by approximate message passing (AMP), following the algorithms of Montanari [103] and Sellke [125], whose nonlinearities are given by the Parisi partial differential equation (PDE). Assuming that the support of the Parisi measure is an interval $[0, q_*]$ for all large β , Montanari also showed that at low temperature his

algorithm constructs approximate solutions of the TAP equations. The output m is computed from finitely many products of the coupling matrix W with vectors, each a function of the earlier products. On tori and hypercubes, Hachem's state evolution for sparse matrices [80] shows that $n^{-1}\mathbb{E} \Theta_G(m)$ is close to the right side of an identity for $\mathcal{P}(\beta, h)$ in terms of the Parisi martingale (Lemma 2.6), and that the empirical law of the coordinates of m is close to the law μ_+ of $\tanh X_{q_+}$, where X is the Parisi diffusion and q_+ is the largest point of the support of the Parisi measure. Only this step uses the theorem on the support of the Parisi measure, and only for $h > 0$; the case $h = 0$ follows by continuity in h .

Second, we condition on the products computed by the algorithm. Conditioning on the iterates of message passing goes back to Bolthausen [25], and it was used to compute the free energy of the SK model at high temperature in [26, 29]. Given these products, the vector of couplings is the sum of its conditional mean and a Gaussian vector projected onto a subspace of bounded codimension per vertex. We expand the partition function around m , writing $\sigma = m + v$. The unobserved part of the disorder enters through the inner product of a Gaussian vector with the edge vector $a(v)_{xy} = D^{-1/2}v_x v_y$. Averaging the exponential over the Gaussian vector produces a variance term from which the Onsager correction is extracted, whereas the logarithm of this average differs from the average of the logarithm by a relative entropy. This relative entropy is controlled by $\mathcal{D}(m)$, the relative entropy of the law of $\beta a(v^*) + Z$, where $v^* = \sigma^* - m$ for a random configuration σ^* with independent coordinates of means m , with respect to the law of the noise Z . Using Talagrand's transport inequality for the Gaussian measure [132], we show that the loss in the pressure is at most $4\beta(n^{-1}\mathbb{E} \mathcal{D}(m))^{1/2}$ (Theorem 6.2). Thus the lower bound reduces to the sublinear relative-entropy estimate $\mathbb{E} \mathcal{D}(m) = o(n)$ for the planted edge model. In [26, 29], the corresponding error is bounded by a conditional second-moment computation at high temperature; the planted model and the transport inequality take its place here. For spherical models, Huang and Sellke [85] proved the lower bound in the Parisi formula by constructing an ultrametric tree of pure states, each with approximately the same free energy as the whole model. They build the tree with an optimization algorithm and a truncated second-moment argument.

Third, we prove this relative-entropy estimate (Theorem 7.5). We interpolate between the edge observations and independent scalar Gaussian observations of the coordinates of v^* , and we compare the relative entropy along the interpolation with a Hopf–Lax formula, as in the Hamilton–Jacobi approach to the inference of low-rank matrices [56, 104, 105]. The Hopf–Lax formula involves the relative entropy $\psi_\mu(r)$ of a scalar Gaussian channel with signal-to-noise ratio r , averaged over the law μ of the coordinates of m , and it shows that $n^{-1}\mathcal{D}(m)$ is asymptotically at most $\sup_{0 \leq s \leq 1} [\psi_\mu(\beta^2 s) - \beta^2 s^2/4]$. The comparison requires the empirical averages of ψ_{m_x} to be close to ψ_μ , uniformly over a set of signal-to-noise profiles; on tori and hypercubes, this follows from the finite range of the algorithm

and the fact that all but an exponentially small fraction of the eigenvalues of S are close to zero. It uses that $S \preceq I$, but not that S is positive semidefinite. Finally, for $\mu = \mu_+$, the supremum vanishes. For $0 \leq s \leq 1 - q_+$, the Doob-transform representation of the Parisi diffusion above q_+ identifies $\psi_{\mu_+}(\beta^2 s) - \beta^2 s^2/4$ with a difference of values of the first variation of the Parisi functional at its minimizer, computed by Jagannath and Tobasco [88] and by Chen [38]. The optimality of the Parisi measure shows that this difference is nonpositive. For $s \geq 1 - q_+$, the bound $\Gamma'_{\mu_+}(s) \leq \beta^2(1 - q_+ - s) \leq 0$ extends the inequality to the remaining range (Proposition 7.4). For the SK model, the function $s \mapsto 2\psi_\mu(\beta^2 s) - \beta^2 s^2/2$ is the function Γ_μ introduced by Chen, Panchenko, and Subag [41]. With $q = \int m^2 \mu(dm) < 1$, they showed that it is nonpositive on $[0, 1 - q]$ if and only if the variational formula for their generalized TAP correction at the law of $|m|$ is minimized by the point mass at zero, and that the correction is then the classical one [41, Proposition 13].

The lower bound obtained in this way holds for tori of a fixed side length and for hypercubes. In Section 8, a Gaussian interpolation that adds or removes edges shows that the torus pressure changes by at most $\beta^2/(4L)$ when the side length is changed from L to infinity, uniformly in d , which gives the uniformity in L . This interpolation was used by Contucci and Graffi [44, 45] to study the thermodynamic limit and the surface pressure of the EA model.

1.4 Related work

Mean-field comparisons. Gaussian interpolation and comparison arguments are classical [75, 76]. For the SK model, they give the identity

$$p_n^{\text{SK}}(\beta, h) - p_G(\beta, h) = \frac{\beta^2}{4} \int_0^1 \mathbb{E} \langle Q - R^2 \rangle_\theta d\theta$$

along the interpolation $\sqrt{1 - \theta} H_G + \sqrt{\theta} H_n^{\text{SK}}$ between a graph Hamiltonian and an independent SK Hamiltonian on the same vertex set. We have $Q - R^2 = n^{-1} \langle u - R\mathbf{1}, S(u - R\mathbf{1}) \rangle$. If $\lambda_{\min}(S) \geq -\epsilon$, then $Q - R^2 \geq -\epsilon(1 - R^2) \geq -\epsilon$, and the identity gives $p_G(\beta, h) \leq p_n^{\text{SK}}(\beta, h) + \beta^2 \epsilon/4$. For every $G \in \mathcal{G}_D$, the matrix S is nonzero with zero trace, so it cannot be positive semidefinite. Positive-semidefinite interaction kernels give the upper bound without this error for Kac models, as in Guerra and Toninelli [77]. If the eigenvalues of S on the orthogonal complement of $\mathbf{1}$ tend to zero, then $Q - R^2$ tends to zero uniformly, and the identity gives both bounds. Neither condition holds on tori and hypercubes. Their matrices S have eigenvalues close to 1 on this complement, and on bipartite graphs S has the eigenvalue -1 and $Q - R^2$ takes negative values. Proposition 4.1 has no spectral hypothesis.

Kac limits and mean-field approximations. High-temperature results for diverging-range mean-field limits were obtained earlier by Fröhlich and Zegarlinski [71]. In the Kac limit,

the lower bound of Franz and Toninelli [69, 70] interpolates between the Kac model and independent SK models on boxes whose side is large compared with the lattice spacing and small compared with the range of the interaction. It uses that each spin interacts with many spins through a kernel that varies slowly on the scale of the boxes. For nearest-neighbor interactions there are no such boxes, and our lower bound uses message passing instead. For Ising and Potts models with an interaction matrix A whose entries may have either sign, Basak and Mukherjee proved, under a boundedness condition on A , that the mean-field approximation of the free energy is asymptotically exact when $\text{tr}(A^2) = o(n)$ [18, Theorem 1.1]. In particular, on regular graphs of diverging degree D , the free energy of the ferromagnetic Ising model with couplings $1/D$ converges to that of the Curie–Weiss model [18, Theorem 2.1]. For the EA model, $\text{tr}((\beta W)^2)$ is of order n , and the mean-field approximation misses the Onsager correction.

The Parisi formula. The upper bound in the Parisi formula was proved by Guerra [76] and the lower bound by Talagrand [134]. Panchenko proved ultrametricity from the Ghirlanda–Guerra identities [74, 113, 114]. For multispecies models with a positive-definite interaction matrix, Barra, Contucci, Mingione, and Tantari [17] proved the upper bound in the Parisi formula, and Panchenko proved the lower bound by showing that the overlaps of different species are synchronized [115]. Synchronization was extended to Potts and vector spins in [116, 117]. Connections between the free energy of spin glasses and Hamilton–Jacobi equations were pointed out by Guerra and coauthors; see [15]. Mourrat and coauthors developed an approach to the Parisi formula and its extensions through Hamilton–Jacobi equations [56, 105, 107–109], and Chen and Xia studied these equations on convex cones [36]. Our upper bound combines these tools. The rows of S play the role of species, but their number is of order n/D even after sparsification, so we do not use a multispecies formula; each row overlap is synchronized separately with the overlap in a Ruelle cascade. The lower bound uses the variational representation of the Parisi functional of Auffinger and Chen [7] and the characterization of the Parisi measure through the first variation of the Parisi functional, obtained by Jagannath and Tobasco [88] and independently by Chen [38].

Multispecies models with a nonconvex covariance. For such models, which include the bipartite SK model [16], the Hamilton–Jacobi approach gives an upper bound on the free energy [106, 108]. Chen and Mourrat showed that, if the limit of the free energy of a vector spin glass with nonconvex interactions exists, it is a critical value of a functional of Parisi type [33], and Chen extended this to multispecies models [31]. For the bipartite spherical SK model, the limit of the free energy was computed by Auffinger and Chen at high temperature [5] and by Baik and Lee in zero field, at every noncritical temperature [11]; the critical value follows by continuity. Subag computed it for pure multispecies spherical models through a TAP representation, assuming that certain free energies converge [130]. For Ising spins, in zero field the limit of the free energy was recently identified by Chen,

Issa, and Mourrat [37]. For balanced models, such as the bipartite SK model with species of equal sizes, this limit is given by a one-species formula. The corresponding lower bound is due to Bates and Sohn [19], and for models with exchangeable species to Issa [86]. The matching upper bound was proved in [37, 83], and the lower bound of Bates and Sohn holds in every field [19, Remark 1.8]. The model on the complete bipartite graph $K_{D,D}$ is a bipartite SK model with two species of equal size. Proposition 4.1 gives the upper bound for it in every field, and together with the interpolation identity above it gives the limit of its free energy in every field (Remark 1.8).

Message passing and the TAP free energy. AMP was introduced by Donoho, Maleki, and Montanari [57]. For the SK model, an iteration with Gaussian state evolution was introduced by Bolthausen [25], and general AMP algorithms were analyzed in [21, 22, 91]. Hachem extended polynomial AMP to sparse variance profiles [80]. Montanari [103] showed that AMP with nonlinearities given by the Parisi PDE approximately maximizes the SK Hamiltonian if, for all large β , the support of the Parisi measure is an interval $[0, q_*]$, and Sellke [125] extended the algorithm to nonzero external fields under a corresponding no-overlap-gap assumption. El Alaoui, Montanari, and Sellke [60] extended these algorithms to mixed p -spin models and studied the energy that they reach through a stochastic control problem; we use this formulation in Section 5. El Alaoui, Montanari, and Sellke also combined AMP with stochastic localization to sample from the SK Gibbs measure at high temperature [61]. Fan, Li, and Sen [66] proved the TAP equations for spin glasses with orthogonally invariant couplings at high temperature, through a conditional second-moment analysis of the free energy restricted to a thin band around the output of AMP. The functional $\beta H(m) + h \sum_x m_x + \sum_x \text{ent}(m_x) + (\beta^2/4) \sum_{x,y} S_{xy}(1 - m_x^2)(1 - m_y^2)$, whose value at the output of our algorithm we bound below, is the TAP free energy of the graph. For mixed p -spin models with Ising spins, Chen and Panchenko proved that the free energy is asymptotically the maximum of the TAP free energy over magnetizations whose self-overlap lies to the right of the support of the Parisi measure [39, Theorem 1], and this representation was extended by Chen, Panchenko, and Subag [40, 41]; see also [9, 65]. For spherical models, a TAP representation of the free energy was proved by Subag [131].

1.5 Organization

Section 2 collects notation and the facts about the SK model and the Parisi formula that we use. Sections 3 and 4 prove the upper bound (1.3) for all regular graphs. Sections 5–7 prove the lower bound for tori of fixed side length and for hypercubes: Section 5 constructs the centers by message passing, Section 6 reduces the lower bound to a sublinear relative-entropy bound for a planted model, and Section 7 proves this statement. Section 8 removes the restriction to a fixed side length and proves Theorems 1.1 and 1.2

and Corollary 1.3.

Preliminaries

This section fixes notation and collects the inputs used in the rest of the paper. Section 2.1 defines the model and records its elementary properties. Section 2.2 lists classical facts about Gaussian vectors. Section 2.3 recalls the Sherrington–Kirkpatrick model, the Parisi formula, and the properties of the Parisi PDE that we use. It ends with Lemma 2.6, which writes $\mathcal{P}(\beta, h)$ in terms of the Parisi martingale; this identity is the target of the lower bound.

2.1 The model

For $D \geq 1$ let $\mathcal{G}_D$ be the set of finite simple D -regular graphs $G = (V, E)$, with any number $n = |V|$ of vertices, connected or not. Vertices are denoted by x, y, z and edges by e or xy . For $G \in \mathcal{G}_D$ we write $S = A_G/D$ for the normalized adjacency matrix. It is symmetric, has nonnegative entries and zero diagonal, and satisfies

$$S\mathbf{1} = \mathbf{1}, \quad \|S\| \leq 1, \quad \text{tr } S^2 = \frac{n}{D}, \quad \max_{x,y \in V} S_{xy} = \frac{1}{D}, \quad (2.1)$$

where $\|\cdot\|$ is the operator norm. The bound on the norm follows from Jensen's inequality, since $\sum_x (\sum_y S_{xy} w_y)^2 \leq \sum_{x,y} S_{xy} w_y^2 = \sum_y w_y^2$, and the trace identity follows by summing the squares of the entries.

Let $(g_e)_{e \in E}$ be independent standard Gaussian variables. The couplings are $J_{xy} = g_{xy}/\sqrt{D}$, and the coupling matrix W is the symmetric $V \times V$ matrix with $W_{xy} = J_{xy}$ if $xy \in E$ and $W_{xy} = 0$ otherwise; in particular, $W_{xx} = 0$. We evaluate the Hamiltonian also at real vectors. The Hamiltonian, the pressure at inverse temperature $\beta \geq 0$ and field $h \in \mathbb{R}$, and the ground-state energy are

$$\begin{aligned} H_G(w) &= \sum_{xy \in E} J_{xy} w_x w_y = \frac{1}{2} \langle w, Ww \rangle \quad (w \in \mathbb{R}^V), \\ p_G(\beta, h) &= \frac{1}{n} \mathbb{E} \log \sum_{\sigma \in \{-1,1\}^V} \exp \left(\beta H_G(\sigma) + h \sum_{x \in V} \sigma_x \right), \\ e(G; h) &= \frac{1}{n} \mathbb{E} \max_{\sigma \in \{-1,1\}^V} \left(H_G(\sigma) + h \sum_{x \in V} \sigma_x \right). \end{aligned} \quad (2.2)$$

The field is not multiplied by β . For all σ , we have $\mathbb{E} H_G(\sigma)^2 = |E|/D = n/2$.

The torus $\mathbb{T}_L^d = (\mathbb{Z}/L\mathbb{Z})^d$ with $L \geq 3$, in which two vertices are adjacent if they differ by ± 1 in exactly one coordinate, belongs to $\mathcal{G}_{2d}$ and has $n = L^d$ vertices. We write $p_{d,L}$ for its pressure. The hypercube $Q_d = \{0, 1\}^d$, in which two vertices are adjacent if they differ in exactly one coordinate, belongs to $\mathcal{G}_d$ and has $n = 2^d$ vertices.

For fixed disorder, $\langle \cdot \rangle$ denotes the average under the Gibbs measure, which gives σ weight proportional to $\exp(\beta H_G(\sigma) + h \sum_x \sigma_x)$. We use it for one or several independent replicas $\sigma^1, \sigma^2, \dots$ under the same disorder. We write $u = \sigma^1 \odot \sigma^2$ for the coordinatewise product of two replicas, and

$$R = \frac{1}{n} \sum_{x \in V} u_x, \quad Q = \frac{1}{|E|} \sum_{xy \in E} u_x u_y = \frac{1}{n} \langle u, Su \rangle$$

for the site overlap and the link overlap. The second expression for Q uses $|E| = nD/2$.

The following notation is used in the lower bound. For $m \in [-1, 1]$ we write

$$\text{ent}(m) = -\frac{1+m}{2} \log \frac{1+m}{2} - \frac{1-m}{2} \log \frac{1-m}{2} \in [0, \log 2],$$

with $0 \log 0 = 0$, for the entropy of a $\{-1, 1\}$ -valued random variable with mean m . For $m \in [-1, 1]^V$ we write π_m for the product law on $\{-1, 1\}^V$ with means m , and $\kappa_x = 1 - m_x^2$ for its variances. If $m_x = \tanh y$, then

$$\text{ent}(m_x) = \log 2 \cosh y - y \tanh y. \quad (2.3)$$

Lemma 2.1. *Let $G \in \mathcal{G}_D$.*

(a) *The function $(\beta, h) \mapsto p_G(\beta, h)$ is convex on $\mathbb{R}^2$ and even in h , and $p_G(0, h) = \log 2 \cosh h$.*

(b) *For $\beta \geq 0$ and $h \in \mathbb{R}$,*

$$\partial_\beta p_G(\beta, h) = \frac{\beta}{2} (1 - \mathbb{E}\langle Q \rangle) \in \left[0, \frac{\beta}{2}\right], \quad \partial_h p_G(\beta, h) = \frac{1}{n} \sum_{x \in V} \mathbb{E}\langle \sigma_x \rangle \in [-1, 1].$$

(c) *For $\beta \geq 0$ and $h \in \mathbb{R}$, we have $\log 2 \cosh h \leq p_G(\beta, h) \leq \log 2 \cosh h + \beta^2/4$.*

(d) *For $B \geq 0$, $\beta, \beta' \in [0, B]$, and $h, h' \in \mathbb{R}$, we have $|p_G(\beta, h) - p_G(\beta', h')| \leq B|\beta - \beta'|/2 + |h - h'|$.*

Proof. For fixed disorder, the function $(\beta, h) \mapsto \log \sum_\sigma \exp(\beta H_G(\sigma) + h \sum_x \sigma_x)$ is convex, since it is a log-sum-exp of affine functions, and its partial derivatives are $\langle H_G \rangle$ and $\sum_x \langle \sigma_x \rangle$. Since these are bounded by $\max_\sigma |H_G(\sigma)|$ and n , which are integrable, we may differentiate under the expectation. The substitution $\sigma \mapsto -\sigma$ shows that p_G is even in h , and at $\beta = 0$ the sum factorizes and equals $(2 \cosh h)^n$. This proves (a) and the formula for $\partial_h p_G$.

For the β -derivative, Gaussian integration by parts ((F1) below) in g_{xy} gives

$$\mathbb{E}[g_{xy} \langle \sigma_x \sigma_y \rangle] = \mathbb{E} \partial_{g_{xy}} \langle \sigma_x \sigma_y \rangle = \frac{\beta}{\sqrt{D}} \mathbb{E}(1 - \langle \sigma_x \sigma_y \rangle^2).$$

Since $\langle \sigma_x \sigma_y \rangle^2 = \langle u_x u_y \rangle$ for two replicas, summing over the $|E| = nD/2$ edges, we obtain

$$\partial_\beta p_G = \frac{1}{n} \sum_{xy \in E} \frac{1}{\sqrt{D}} \mathbb{E}[g_{xy} \langle \sigma_x \sigma_y \rangle] = \frac{\beta}{nD} \sum_{xy \in E} (1 - \mathbb{E} \langle u_x u_y \rangle) = \frac{\beta}{2} (1 - \mathbb{E} \langle Q \rangle).$$

Since $\langle Q \rangle = |E|^{-1} \sum_{xy \in E} \langle \sigma_x \sigma_y \rangle^2 \in [0, 1]$, this proves (b). Part (c) follows by integrating the bounds of (b) in β from 0 and using (a), and (d) follows from (b). $\square$

2.2 Classical tools

For a random variable Y we write $\|Y\|_p = (\mathbb{E}|Y|^p)^{1/p}$. A function on $\mathbb{R}^k$ has *polynomial growth* if it is bounded by $C(1 + |w|^p)$ for some C, p . We use the following classical facts.

(F1) *Gaussian integration by parts.* Let $(U_1, \dots, U_k)$ be a centered Gaussian vector, with degenerate covariance allowed, and let $\Psi: \mathbb{R}^k \rightarrow \mathbb{R}$ be C^1 , with Ψ and $\nabla \Psi$ of polynomial growth. Then for all $j \leq k$,

$$\mathbb{E}[U_j \Psi(U_1, \dots, U_k)] = \sum_{\ell=1}^k \text{Cov}(U_j, U_\ell) \mathbb{E} \partial_\ell \Psi(U_1, \dots, U_k).$$

For one standard Gaussian variable this is Stein's identity [128], and the general case follows by writing $U_\ell = a_\ell U_j + V_\ell$ with $a_\ell = \text{Cov}(U_\ell, U_j) / \text{Var } U_j$ and (V_ℓ) independent of U_j , and applying Stein's identity conditionally on (V_ℓ) ; if $\text{Var } U_j = 0$, both sides vanish.

(F2) *Method of moments.* A Gaussian law on $\mathbb{R}^k$, degenerate or not, is determined by its moments. If random vectors Z_ℓ satisfy $\mathbb{E} \psi(Z_\ell) \rightarrow \mathbb{E} \psi(Z)$ for every polynomial ψ , with Z Gaussian, then $Z_\ell \rightarrow Z$ in law. In dimension one this is [24, Theorems 30.1 and 30.2], and the general case follows by the Cramér–Wold device [24, Theorem 29.4].

(F3) *Polynomial density.* For every Gaussian measure γ on $\mathbb{R}^k$, including a degenerate one, and every $1 \leq p < \infty$, polynomials are dense in $L^p(\gamma)$. It suffices to work on the affine support of γ , where γ is nondegenerate. If density failed there, duality would give a nonzero $f \in L^{p'}(\gamma)$, with $1/p + 1/p' = 1$, such that $\int f P d\gamma = 0$ for every polynomial P . Since $\int |f(w)| e^{c|w|} \gamma(dw) < \infty$ for all $c > 0$ by Hölder's inequality and the Gaussian exponential moments, we may integrate the exponential series term by term, and we obtain $\int f(w) e^{i\langle t, w \rangle} \gamma(dw) = 0$ for every t . Since $\int f d\gamma = 0$, the positive and negative parts of $f\gamma$ have equal mass and equal Fourier transforms. If this mass is positive, normalize the two parts and apply Fourier uniqueness for probability measures [24, Section 29]; if it is zero, both parts vanish. In either case $f\gamma = 0$, a contradiction.

- (F4) *Gaussian concentration.* Let $\Psi: \mathbb{R}^k \rightarrow \mathbb{R}$ be L -Lipschitz for the Euclidean norm and let g be a standard Gaussian vector in $\mathbb{R}^k$. Then $\text{Var } \Psi(g) \leq L^2$, and $\mathbb{P}(|\Psi(g) - \mathbb{E}\Psi(g)| \geq t) \leq 2e^{-t^2/(2L^2)}$ for every $t \geq 0$; see [28, Chapter 3 and Section 5.4]. In particular, $\|\Psi(g) - \mathbb{E}\Psi(g)\|_p \leq C_p L$ for every $p \geq 1$, where C_p depends only on p .
- (F5) *Transport inequality.* Let γ_k be the standard Gaussian measure on $\mathbb{R}^k$. For every probability measure ρ on $\mathbb{R}^k$, the quadratic Wasserstein distance satisfies $W_2(\rho, \gamma_k)^2 \leq 2\text{KL}(\rho \parallel \gamma_k)$ [132].
- (F6) *Maximum of the Hamiltonian.* Let $G \in \mathcal{G}_D$. Since H_G has no diagonal terms, it is affine in each coordinate separately, and $\sup_{w \in [-1,1]^V} |H_G(w)| = \max_\sigma |H_G(\sigma)|$. For every $p \geq 1$ there exists C_p , depending only on p , such that $\mathbb{E} \max_\sigma |H_G(\sigma)|^p \leq C_p n^p$. Indeed, since $\max_\sigma |H_G(\sigma)|$ is the maximum of the 2^{n+1} centered Gaussian variables $\pm H_G(\sigma)$, each of variance $n/2$, its mean is at most $(2 \cdot (n/2) \log 2^{n+1})^{1/2} \leq 2n$. Further, it is a $(n/2)^{1/2}$ -Lipschitz function of (g_e) , and (F4) bounds its fluctuations.

2.3 The Sherrington–Kirkpatrick model and the Parisi formula

The constants. Let $H_N^{\text{SK}}(\sigma) = N^{-1/2} \sum_{1 \leq i < j \leq N} g_{ij} \sigma_i \sigma_j$ be the Sherrington–Kirkpatrick (SK) Hamiltonian on $\{-1, 1\}^N$, with (g_{ij}) independent standard Gaussian variables, and let

$$p_N^{\text{SK}}(\beta, h) = \frac{1}{N} \mathbb{E} \log \sum_{\sigma} e^{\beta H_N^{\text{SK}}(\sigma) + h \sum_i \sigma_i}, \quad m_N(h) = \frac{1}{N} \mathbb{E} \max_{\sigma} \left(H_N^{\text{SK}}(\sigma) + h \sum_i \sigma_i \right).$$

The proof of Lemma 2.1 applies to p_N^{SK} without change, with the site overlap R of two SK replicas and R^2 in place of Q ; it gives $\partial_\beta p_N^{\text{SK}} = (\beta/2)(1 - \mathbb{E}\langle R^2 \rangle)$, and the conclusions (a)–(d) hold for p_N^{SK} . The limit $\mathcal{P}(\beta, h) = \lim_{N \rightarrow \infty} p_N^{\text{SK}}(\beta, h)$ exists by the theorem of Guerra and Toninelli [75], and it is given by the Parisi formula (2.7) below. By Lemma 2.1, $\mathcal{P}$ is convex, even in h , and satisfies the bounds of (c) and (d).

Bounding the partition function at inverse temperature β and field βh below by its largest term and above by 2^N times its largest term, we find that for all $\beta > 0$,

$$\beta m_N(h) \leq \frac{1}{N} \mathbb{E} \log \sum_{\sigma} e^{\beta(H_N^{\text{SK}}(\sigma) + h \sum_i \sigma_i)} \leq \beta m_N(h) + \log 2. \quad (2.4)$$

By the same argument, we have $\beta e(G; h) \leq p_G(\beta, \beta h) \leq \beta e(G; h) + \log 2$ for every $G \in \mathcal{G}_D$. Letting $N \rightarrow \infty$ in (2.4), we obtain $\limsup_N m_N(h) \leq \mathcal{P}(\beta, \beta h)/\beta \leq \liminf_N m_N(h) + (\log 2)/\beta$ for every $\beta > 0$, and letting $\beta \rightarrow \infty$ shows that $m_N(h)$ converges. We write $P_*(h)$ for its limit and $P_* = P_*(0)$. Then $P_*(h) = \lim_{\beta \rightarrow \infty} \mathcal{P}(\beta, \beta h)/\beta$, and for every $\beta > 0$,

$$\beta P_*(h) \leq \mathcal{P}(\beta, \beta h) \leq \beta P_*(h) + \log 2. \quad (2.5)$$

The Parisi formula. For a probability measure μ on $[0, 1]$ we write $\mu(t) = \mu([0, t])$, and let Φ_μ be the solution of the Parisi PDE

$$\partial_t \Phi + \frac{\beta^2}{2} \left(\partial_x^2 \Phi + \mu(t) (\partial_x \Phi)^2 \right) = 0 \quad \text{on } [0, 1] \times \mathbb{R}, \quad \Phi(1, x) = \log 2 \cosh x. \quad (2.6)$$

For atomic μ , Auffinger and Chen [7, (4)–(5)] define Φ_μ by successive Cole–Hopf transforms; it is smooth and satisfies (2.6) classically between consecutive atoms. They extend $\mu \mapsto \Phi_\mu$ to all μ by continuity, using a Lipschitz estimate of Guerra [76], and this is the function we use. The PDE does not involve the field. The Parisi functional is

$$\mathcal{P}_{\beta,h}(\mu) = \Phi_\mu(0, h) - \frac{\beta^2}{2} \int_0^1 t \mu(t) dt.$$

Parisi predicted the formula (2.7) below [118, 119]; Guerra proved the upper bound [76], and Talagrand proved the formula [134]. By Talagrand’s theorem and the uniqueness theorem of Auffinger and Chen [7, Corollary 1], for every $\beta > 0$ and $h \in \mathbb{R}$,

$$\mathcal{P}(\beta, h) = \min_{\mu} \mathcal{P}_{\beta,h}(\mu) = \mathcal{P}_{\beta,h}(\nu_{\beta,h}) \quad (2.7)$$

for a unique minimizer $\nu_{\beta,h}$, the Parisi measure. Since $\Phi_\mu(t, \cdot)$ is even, $\mathcal{P}_{\beta,h} = \mathcal{P}_{\beta,-h}$ and $\nu_{\beta,-h} = \nu_{\beta,h}$. We write $q_- = \min \text{supp } \nu_{\beta,h}$ and $q_+ = \max \text{supp } \nu_{\beta,h}$.

Most of the cited papers use the terminal condition $\log \cosh x$ and add $\log 2$ to the functional, with covariance $N\xi(R)$ and $\xi(s) = \beta^2 s^2/2$. In [99, Sections 1.1 and 2] the Gibbs weight is $\exp(\beta H_N^{\text{SK}}(\sigma) + h \sum_i \sigma_i)$, and the field parameter is the same as ours. The time variable there is $\beta^2 t$: in the notation of that paper, $\Phi(t, x) = \log 2 + U(\beta^2 t, x)$, and the Parisi diffusion below is a Brownian time change of the diffusion used there. That paper therefore describes our $\mathcal{P}(\beta, h)$ and $\nu_{\beta,h}$, and we use the following theorem from it [99, Theorem 1.1]. We need it only for $h > 0$; the lower bound in zero field follows by continuity in h (Proposition 8.1).

Theorem 2.2. *For every $\beta > 0$ and $h > 0$, there exist $0 < q_- \leq q_+ < 1$ such that $\text{supp } \nu_{\beta,h} = [q_-, q_+]$.*

The Parisi diffusion. For the rest of this section we fix $\beta > 0$ and $h \geq 0$, and write $\nu = \nu_{\beta,h}$ and $\Phi = \Phi_\nu$. Let

$$v(t, x) = \beta^2 \nu(t) \partial_x \Phi(t, x), \quad a(t, x) = \beta \partial_x^2 \Phi(t, x),$$

let B be a standard Brownian motion, let X solve

$$dX_t = v(t, X_t) dt + \beta dB_t, \quad X_0 = h, \quad (2.8)$$

and let $M_t = \partial_x \Phi(t, X_t)$. The solution X exists and is pathwise unique because v is bounded, measurable, and Lipschitz in x by (P1) below. This diffusion appears in [7,

Theorem 3], and it is called the Auffinger–Chen stochastic differential equation (SDE) in [88]. We call X the Parisi diffusion and M the Parisi martingale. Since $v(t) = 0$ for $t < q_-$,

$$X_t = h + \beta B_t \quad (0 \leq t \leq q_-). \quad (2.9)$$

We use the following facts. Property (P1) and Lemma 2.5 hold for every probability measure μ in place of ν .

(P1) *Regularity* [7, Proposition 2(i),(ii),(iv)]. For $j \leq 4$, $\partial_x^j \Phi$ exists and is continuous on $[0, 1] \times \mathbb{R}$; further, $|\partial_x \Phi| \leq 1$, $0 < \partial_x^2 \Phi \leq 1$, and $|\partial_x^3 \Phi| \leq 4$. If $\mu_\ell \rightarrow \mu$ weakly, then $\partial_x^j \Phi_{\mu_\ell} \rightarrow \partial_x^j \Phi_\mu$ uniformly for $j \leq 2$. In particular, $|v| \leq \beta^2$, $0 \leq \partial_x v \leq \beta^2$, $0 < a \leq \beta$, and $|\partial_x a| \leq 4\beta$. In [7] these facts are deduced from [6, Propositions 1 and 2] and an estimate of Talagrand. Jagannath and Tobasco [87, Theorem 4 and Lemma 16] gave a PDE proof of the well-posedness and regularity of the Parisi PDE.

(P2) *Martingales* [7, Lemma 2]. For $0 \leq s \leq t \leq 1$,

$$\begin{aligned} M_t - M_s &= \int_s^t a(r, X_r) dB_r, \\ \partial_x^2 \Phi(t, X_t) - \partial_x^2 \Phi(s, X_s) &= - \int_s^t \beta^2 v(r) (\partial_x^2 \Phi(r, X_r))^2 dr + \int_s^t \beta \partial_x^3 \Phi(r, X_r) dB_r. \end{aligned}$$

(P3) *Variational representation at the optimizer* [7, Theorem 3], applied on $[0, q_+]$ from $x = h$, where the maximizer is $r \mapsto M_r$:

$$\Phi(0, h) = \mathbb{E} \Phi(q_+, X_{q_+}) - \frac{\beta^2}{2} \int_0^{q_+} v(r) \mathbb{E} M_r^2 dr.$$

In [7, Theorem 3], $\Phi_\mu(s, x)$ is written as the value of a stochastic optimal-control problem; Jagannath and Tobasco [87, Lemma 18] gave a second proof of this representation by dynamic programming.

(P4) *Cole–Hopf above q_+* . Since $v(t) = 1$ for $t \geq q_+$, we have $\Phi(t, x) = \log 2 \cosh x + \beta^2(1-t)/2$ for $t \in [q_+, 1]$. For an atomic measure with all atoms in $[0, q_+]$ this is the Cole–Hopf formula on $[q_+, 1]$ (see [7, (4)]), since $\mathbb{E} 2 \cosh(x + \beta \sqrt{1-t} Z) = 2 \cosh(x) e^{\beta^2(1-t)/2}$ for $Z \sim N(0, 1)$. It passes to ν by approximating ν weakly by atomic measures supported in $[0, q_+]$ and using (P1).

Optimality of the Parisi measure. Let $\Sigma(t) = \mathbb{E} a(t, X_t)^2$. It is continuous on $[0, 1]$, because a is bounded and continuous by (P1) and $t \mapsto X_t$ is continuous. By (P2) and the Itô isometry,

$$\mathbb{E} M_t^2 = M_0^2 + \int_0^t \Sigma(r) dr \quad (t \in [0, 1]). \quad (2.10)$$

We define

$$\Psi(x) = \frac{\beta^2}{2} \int_x^1 (\mathbb{E} M_t^2 - t) dt \quad (x \in [0, 1]). \quad (2.11)$$

By (2.10), Ψ is twice continuously differentiable on $[0, 1]$, with one-sided derivatives at the endpoints, and

$$\Psi'(x) = -\frac{\beta^2}{2}(\mathbb{E}M_x^2 - x), \quad \Psi''(x) = -\frac{\beta^2}{2}(\Sigma(x) - 1). \quad (2.12)$$

Jagannath and Tobasco showed that Ψ is the first variation of the Parisi functional at ν : by [88, Lemma 3.2 and Example 3.4], for every probability measure ρ on $[0, 1]$ the function $\varepsilon \mapsto \mathcal{P}_{\beta, h}((1 - \varepsilon)\nu + \varepsilon\rho)$ is right differentiable at $\varepsilon = 0$, with derivative $\int \Psi d\rho - \int \Psi d\nu$. Since ν minimizes $\mathcal{P}_{\beta, h}$, this derivative is nonnegative for every ρ , which gives the following theorem [88, Proposition 1.1]; see also [88, Corollary 3.6]. Chen obtained the first variation and this characterization independently [38, Theorem 2]. For the spherical models, a characterization of this type was obtained earlier by Talagrand; see [88, Remark 1.4].

Theorem 2.3. *Let $\beta > 0$ and $h \geq 0$. Then $\Psi(q) = \min_{x \in [0, 1]} \Psi(x)$ for every $q \in \text{supp } \nu_{\beta, h}$.*

The statement in [88, Proposition 1.1] is that $\nu(\{x : \Psi(x) = \min \Psi\}) = 1$; since Ψ is continuous, this set is closed, and as a closed set of full measure it contains $\text{supp } \nu$. The normalizations agree. In [88], the model is $\xi = \beta^2 \xi_0$, and the SK model corresponds to $\xi_0(t) = t^2/2$. The Parisi PDE there is (2.6) with terminal condition $\log \cosh x$, and its solution u_μ is the unique weak solution in the sense of [88, Section 8.1], where the results of [87] on the Parisi PDE are recalled. For atomic μ , the function $\Phi_\mu - \log 2$ is a weak solution, since it is continuous, has a bounded spatial derivative, and solves the PDE classically between consecutive atoms; it equals u_μ by the uniqueness in [88, Proposition 8.1]. Both $\mu \mapsto u_\mu$ and $\mu \mapsto \Phi_\mu$ are continuous under weak convergence, by [88, Proposition 8.2] and (P1); for the first, weak convergence gives convergence of distribution functions at all but countably many points, which implies convergence in the metric $d(\mu, \mu') = \int_0^1 |\mu([0, s]) - \mu'([0, s])| ds$ used there. Approximating μ weakly by atomic measures, we obtain $u_\mu = \Phi_\mu - \log 2$ for all μ . Then the Parisi functional of [88] is $\mathcal{P}_{\beta, h} - \log 2$, the SDE [88, (1.1.2)] is (2.8), the field is the same, and the function G_ν of [88, (1.1.1)] is Ψ .

The conditions $\mathbb{E}M_t^2 = t$ and $\Sigma(t) \leq 1$ in the next corollary are the self-consistency conditions of [88, Proposition 1.1] and [38, Proposition 1]. If ν has one atom, they are the conditions of de Almeida and Thouless [53]; see [88, Remarks 1.2 and 1.4] and [136]. In zero field, they appear earlier in [6, Theorem 5], where the expectations are written under a Girsanov change of measure. We include the short derivation from Theorem 2.3; the bound $q_+ < 1$ is [88, Lemma 3.8], with the same proof.

Corollary 2.4. *Let $\beta > 0$ and $h \geq 0$. For every $t \in \text{supp } \nu \cap (0, 1)$, we have $\mathbb{E}M_t^2 = t$ and $\Sigma(t) \leq 1$. Further, $q_+ < 1$ and $\mathbb{E}M_{q_+}^2 = q_+$.*

Proof. If $t \in \text{supp } \nu \cap (0, 1)$, then t is an interior minimum point of Ψ by Theorem 2.3, so $\Psi'(t) = 0$ and $\Psi''(t) \geq 0$, and (2.12) gives the first claim. If $q_+ = 1$, then the left derivative of Ψ at 1 is nonpositive, so $\mathbb{E}M_1^2 \geq 1$ by (2.12); however, $M_1 = \partial_x \Phi(1, X_1) = \tanh X_1$ and $|\tanh X_1| < 1$ almost surely, a contradiction. Thus $q_+ < 1$. If $q_+ > 0$, the first claim gives $\mathbb{E}M_{q_+}^2 = q_+$. If $q_+ = 0$, then the right derivative of Ψ at 0 is nonnegative, which gives $\mathbb{E}M_0^2 \leq 0$ and $\mathbb{E}M_{q_+}^2 = 0 = q_+$. $\square$

We also need the regularity of $\partial_x \Phi$ and $\partial_x^2 \Phi$ in time, which follows from (P1).

Lemma 2.5. *For every probability measure μ on $[0, 1]$, all $s, t \in [0, 1]$, and $x \in \mathbb{R}$,*

$$|\partial_x \Phi_\mu(t, x) - \partial_x \Phi_\mu(s, x)| \leq 3\beta^2 |t - s|, \quad |\partial_x^2 \Phi_\mu(t, x) - \partial_x^2 \Phi_\mu(s, x)| \leq 4\sqrt{3}\beta |t - s|^{1/2}.$$

Proof. We write $\Phi = \Phi_\mu$. First let μ be atomic. Between consecutive atoms, Φ is smooth and satisfies (2.6) classically, with $\mu(t)$ constant. Differentiating in x , we obtain

$$\partial_t \partial_x \Phi = -\frac{\beta^2}{2} (\partial_x^3 \Phi + 2\mu(t) \partial_x \Phi \partial_x^2 \Phi),$$

which is at most $\beta^2(4 + 2)/2 = 3\beta^2$ in absolute value by (P1). Since $\partial_x \Phi$ is continuous on $[0, 1] \times \mathbb{R}$, the first bound holds on $[0, 1]$. For the second, fix s, t and let $f = \partial_x \Phi(t, \cdot) - \partial_x \Phi(s, \cdot)$. Then $\|f\|_\infty \leq 3\beta^2 |t - s|$, and $\|f''\|_\infty \leq 8$ by (P1). If $\|f''\|_\infty = 0$, then f is bounded and affine, hence constant, and the claim is immediate. Otherwise, Taylor's formula at $x \pm \epsilon$ gives $2\epsilon |f'(x)| \leq 2\|f\|_\infty + \epsilon^2 \|f''\|_\infty$, and the choice $\epsilon = (2\|f\|_\infty / \|f''\|_\infty)^{1/2}$ gives Landau's inequality $\|f'\|_\infty^2 \leq 2\|f\|_\infty \|f''\|_\infty \leq 48\beta^2 |t - s|$. This is the second bound, since $f' = \partial_x^2 \Phi(t, \cdot) - \partial_x^2 \Phi(s, \cdot)$. For general μ , we approximate μ weakly by atomic measures; both bounds pass to the limit by the uniform convergence in (P1). $\square$

The law of the Parisi center. We write μ_+ for the law of M_{q_+} , and call it the law of the Parisi center. We also set

$$\mathcal{E}(\beta, h) = \int_0^{q_+} \mathbb{E} a(t, X_t) dt.$$

The next lemma collects the properties of the Parisi martingale used in the lower bound. Part (c) writes $\mathcal{P}(\beta, h)$ as the sum of an energy, a field term, an entropy, and a correction, and it identifies the value that the lower bound must reach. For $h > 0$ and a measurable $f: \mathbb{R} \rightarrow [-1, 1]$, we define $\phi_f(t) = \mathbb{E} f(Y) f(Y')$ for $t \in [0, q_-]$, where (Y, Y') is a centered Gaussian pair with variances q_- and covariance t . The right side of (c) has the form of the TAP free energy of Thouless, Anderson, and Palmer [135] at a magnetization whose coordinates have law μ_+ and whose energy per site is $\mathcal{E}(\beta, h)$. That the free energy of mixed p -spin models equals the maximum of the TAP free energy over magnetizations with self-overlap close to q_+ was proved by Chen and Panchenko [39, Theorem 1], and a computation close to our proof of (c) appears in the proof of [40, Theorem 7], without

external field. After the change of variables $s = \beta^2 t$, the inequality $\phi_f(t) > t$ in (a) is the analogue of the property of the root map in the first phase of Sellke's algorithm [125, Lemma 3.2]. There the root map is defined through the Parisi PDE at zero temperature. If $q_- = q_+$, then $f(y) = \tanh(h + \beta y)$, and this inequality follows from [25, Lemma 2.2 and Corollary 2.3]. Our proof uses the same convexity argument as [25, 125]. The proof of (b) follows [103, Lemma 3.7]. Since $a(q_+, x) = \beta(1 - \tanh^2 x)$, the inequality $\mathbb{E}a(q_+, X_{q_+})^2 \leq 1$ used there is Plefka's condition [121] for the law of M_{q_+} ; see [41, Proposition 14].

Lemma 2.6. *Let $\beta > 0$ and $h > 0$.*

(a) *We have $\mathbb{E}M_t^2 = t$ for $t \in [q_-, q_+]$. If $q_- < q_+$, then $\mathbb{E}a(t, X_t)^2 = 1$ for $t \in [q_-, q_+]$. The function $f(y) = \partial_x \Phi(q_-, h + \beta y)$ is C^1 with values in $(-1, 1)$ and $f' \in (0, \beta]$, it satisfies $f(B_{q_-}) = M_{q_-}$ and $\phi_f(q_-) = q_-$, and $\phi_f(t) > t$ for every $t \in [0, q_-]$.*

(b) *We have $M_{q_+} = \tanh X_{q_+}$ and $\beta^2(1 - q_+)^2 \leq 1$.*

(c) *We have*

$$\mathcal{P}(\beta, h) = \beta \mathcal{E}(\beta, h) + h \mathbb{E}M_{q_+} + \mathbb{E} \text{ent}(M_{q_+}) + \frac{\beta^2}{4}(1 - q_+)^2. \quad (2.13)$$

Proof. (a) By Theorem 2.2, $[q_-, q_+] = \text{supp } \nu \subset (0, 1)$, and Corollary 2.4 implies that $\mathbb{E}M_t^2 = t$ for $t \in [q_-, q_+]$. If $q_- < q_+$, differentiating (2.10) gives $\Sigma = 1$ on (q_-, q_+) , and by continuity also on $[q_-, q_+]$.

Next, by (P1), $f'(y) = \beta \partial_x^2 \Phi(q_-, h + \beta y) \in (0, \beta]$, and f takes values in $[-1, 1]$; since f is strictly increasing, its values lie in $(-1, 1)$. By (2.9), $M_t = \partial_x \Phi(t, h + \beta B_t)$ for $t \leq q_-$, and in particular $f(B_{q_-}) = M_{q_-}$ and $\phi_f(q_-) = \mathbb{E}M_{q_-}^2 = q_-$. Fix $t \in [0, q_-]$. Since M is a martingale and M_t is a function of B_t , we have $\mathbb{E}[f(B_{q_-}) \mid B_t] = M_t$. Let B' be a Brownian motion independent of B and let $A' = B_t + B'_{q_-} - B'_t$. Then (B_{q_-}, A') is a centered Gaussian pair with variances q_- and covariance t , and given B_t its two coordinates are independent with the same law. Hence

$$\phi_f(t) = \mathbb{E}[\mathbb{E}[f(B_{q_-}) \mid B_t]^2] = \mathbb{E}M_t^2 = q_- - \int_t^{q_-} \Sigma(r) dr.$$

It remains to show that $\Sigma(r) < 1$ for $r < q_-$. Since $\nu(r) = 0$ for $r < q_-$, we see from the second identity in (P2) that $\partial_x^2 \Phi(t, X_t)$ is a martingale on $[0, q_-]$, and the Itô isometry gives

$$\Sigma(t) = \Sigma(0) + \beta^4 \int_0^t \mathbb{E}(\partial_x^3 \Phi(r, X_r))^2 dr \quad (t \leq q_-).$$

The integrand is positive for $r \in (0, q_-]$. Indeed, $X_r = h + \beta B_r$ has a positive density, and $\partial_x^3 \Phi(r, \cdot)$ is continuous and not identically zero; otherwise $\partial_x^2 \Phi(r, \cdot)$ would be a positive constant by (P1), and $\partial_x \Phi(r, \cdot)$ would be unbounded. By the last display, $\Sigma(r) < \Sigma(q_-) \leq 1$ for $r < q_-$, where the last inequality is Corollary 2.4 at q_- . Then $\phi_f(t) > q_- - (q_- - t) = t$.

(b) By (P4), we have $\partial_x \Phi(q_+, x) = \tanh x$ and $a(q_+, x) = \beta(1 - \tanh^2 x)$. Then $M_{q_+} = \tanh X_{q_+}$, and (a) gives $\mathbb{E} \tanh^2 X_{q_+} = q_+$. Hence $\beta(1 - q_+) = \mathbb{E} a(q_+, X_{q_+})$. Since $q_+ \in (0, 1)$, Jensen's inequality and Corollary 2.4 at q_+ give $\beta^2(1 - q_+)^2 \leq \mathbb{E} a(q_+, X_{q_+})^2 \leq 1$.

(c) We compute each term of $\mathcal{P}_{\beta, h}(\nu)$, which equals $\mathcal{P}(\beta, h)$ by (2.7).

- Since $\nu(r) = 0$ for $r < q_-$ and $\mathbb{E} M_r^2 = r$ on $[q_-, q_+]$, properties (P3) and (P4) give

$$\Phi(0, h) = \mathbb{E} \log 2 \cosh X_{q_+} + \frac{\beta^2}{2}(1 - q_+) - \frac{\beta^2}{2} \int_0^{q_+} t \nu(t) dt.$$

- Since $\nu(t) = 1$ on $[q_+, 1]$, we have

$$\frac{\beta^2}{2} \int_0^1 t \nu(t) dt = \frac{\beta^2}{2} \int_0^{q_+} t \nu(t) dt + \frac{\beta^2}{4}(1 - q_+^2).$$

- By (2.3) and (b), $\mathbb{E} \log 2 \cosh X_{q_+} = \mathbb{E} \text{ent}(M_{q_+}) + \mathbb{E}[X_{q_+} M_{q_+}]$.
- By (2.8), $X_{q_+} = h + \int_0^{q_+} \beta^2 \nu(t) M_t dt + \beta B_{q_+}$. Since M is a bounded martingale, $\mathbb{E}[M_t M_{q_+}] = \mathbb{E} M_t^2$, which equals t wherever $\nu(t) \neq 0$, by (a). By (P2) and the Itô isometry, $\mathbb{E}[B_{q_+} M_{q_+}] = \int_0^{q_+} \mathbb{E} a(t, X_t) dt = \mathcal{E}(\beta, h)$. Combining these identities, we obtain

$$\mathbb{E}[X_{q_+} M_{q_+}] = h \mathbb{E} M_{q_+} + \beta^2 \int_0^{q_+} t \nu(t) dt + \beta \mathcal{E}(\beta, h).$$

When we substitute these expressions into the Parisi functional, the integrals $\int_0^{q_+} t \nu(t) dt$ cancel, and we obtain

$$\mathcal{P}(\beta, h) = \mathbb{E} \text{ent}(M_{q_+}) + h \mathbb{E} M_{q_+} + \beta \mathcal{E}(\beta, h) + \frac{\beta^2}{2}(1 - q_+) - \frac{\beta^2}{4}(1 - q_+^2).$$

This is (2.13), because $2(1 - q_+) - (1 - q_+^2) = (1 - q_+)^2$. □

By Lemma 2.6(a,b), μ_+ is the law of $\tanh X_{q_+}$, $\mu_+((-1, 1)) = 1$, and $\int m^2 \mu_+(dm) = q_+$.

Finite cascades and the scalar transform

The goal of this section is to develop the calculus of finite cascades that is used in the proof of Proposition 4.1. Section 3.1 fixes the notation. Section 3.2 records derivative identities for cascade recursions with auxiliary randomness; they are applied to the scalar transform in Section 3.3 and to an enriched free energy in Section 4. Section 3.4 proves that the curvature and the complements of the transported squared magnetizations of an Ising cascade are completely monotone functions of the square of the field. Section 3.5 deduces that the scalar transform is convex in the square roots of its variances, for every value of the field. Section 3.6 studies a Hopf–Lax formula built from the scalar transform and compares its value at the origin with the Parisi formula.

3.1 Notation

Throughout this section and the next, we fix an integer $K \geq 1$ and real numbers

$$0 < \zeta_0 < \zeta_1 < \cdots < \zeta_{K-1} < \zeta_K = 1, \quad (3.1)$$

and we set $\zeta_{-1} = 0$,

$$w_j = \zeta_j - \zeta_{j-1} \quad (0 \leq j \leq K), \quad \bar{w} = \max_{0 \leq j \leq K} w_j. \quad (3.2)$$

Then $w_j > 0$ and $\sum_j w_j = 1$. For $x, y \in \mathbb{R}^{K+1}$ we write $\langle x, y \rangle_w = \sum_{j=0}^K w_j x_j y_j$, $\|x\|_{1,w} = \sum_j w_j |x_j|$, and $\|x\|_{2,w} = \langle x, x \rangle_w^{1/2}$. Since $\sum_j w_j = 1$, we have $\|x\|_{1,w} \leq \|x\|_{2,w}$. The ordered cone is

$$\mathcal{C}_K = \{q \in \mathbb{R}^{K+1} : 0 \leq q_0 \leq q_1 \leq \cdots \leq q_K\},$$

and its dual cone is $\mathcal{C}_K^* = \{z \in \mathbb{R}^{K+1} : \langle z, q \rangle_w \geq 0 \text{ for all } q \in \mathcal{C}_K\}$. Every $q \in \mathcal{C}_K$ can be written as $q = \sum_{\ell=0}^K (q_\ell - q_{\ell-1}) \mathbf{1}_{[\ell, K]}$ with $q_{-1} = 0$, where $\mathbf{1}_{[\ell, K]}$ is the vector whose coordinates with index at least ℓ are equal to one and whose other coordinates vanish. Since the coefficients are nonnegative, we have

$$\mathcal{C}_K^* = \left\{ z \in \mathbb{R}^{K+1} : \sum_{j=\ell}^K w_j z_j \geq 0 \text{ for } 0 \leq \ell \leq K \right\}. \quad (3.3)$$

We write $z \succeq_* z'$ if $z - z' \in \mathcal{C}_K^*$. A function $u : \mathcal{C}_K \rightarrow \mathbb{R}$ is called *differentiable on $\mathcal{C}_K$* if there exists a continuous map $Du : \mathcal{C}_K \rightarrow \mathbb{R}^{K+1}$ such that $u(q') = u(q) + \langle Du(q), q' - q \rangle_w +$

$o(|q' - q|)$ as $q' \rightarrow q$ within $\mathcal{C}_K$; we call Du the weighted gradient of u . Its coordinates are $(Du)_j = w_j^{-1} \partial u / \partial q_j$. If u is differentiable on $\mathcal{C}_K$, then for $q, q' \in \mathcal{C}_K$ the segment from q to q' lies in $\mathcal{C}_K$ and

$$u(q') - u(q) = \int_0^1 \langle Du(q + \theta(q' - q)), q' - q \rangle_w d\theta. \quad (3.4)$$

We write $Z, Z_0, Z_1, \dots$ for standard Gaussian variables and, for $b \geq 0$ and a continuous function g of at most exponential growth, $\gamma_b * g(x) = \mathbb{E}g(x + \sqrt{b} Z)$. If g is smooth and all its derivatives have at most exponential growth, then $b \mapsto \gamma_b * g(x)$ is smooth on $[0, \infty)$, with one-sided derivatives at $b = 0$, and

$$\partial_b(\gamma_b * g) = \frac{1}{2} \gamma_b * g'' = \frac{1}{2} \partial_x^2(\gamma_b * g). \quad (3.5)$$

3.2 Cascade recursions

We consider recursions in which, at each level, a Gaussian increment of the field is integrated together with an auxiliary random element. Let $(\Omega_\ell, \mathcal{A}_\ell, P_\ell)$, $1 \leq \ell \leq K$, be probability spaces, let $\omega = (\omega_1, \dots, \omega_K)$ have independent coordinates with $\omega_\ell \sim P_\ell$, and let $Z_1, \dots, Z_K$ be independent standard Gaussian variables, independent of ω . We write $\omega_{\leq k} = (\omega_1, \dots, \omega_k)$. Let $L: \mathbb{R} \times \Omega_1 \times \dots \times \Omega_K \rightarrow \mathbb{R}$ be measurable and such that $L(\cdot, \omega)$ is smooth for each ω , with

$$|\partial_x L| \leq 1, \quad \partial_x^2 L = 1 - (\partial_x L)^2, \quad \mathbb{E} \exp(c|L(0, \omega)|) < \infty \text{ for every } c > 0. \quad (3.6)$$

The two main examples are $L(x) = \log 2 \cosh x$, without auxiliary randomness, and $L(x, \omega) = \log \sum_\sigma \exp(x\sigma_{x_0} + A(\sigma, \omega))$, where σ ranges over a finite set of spin configurations and x_0 is a fixed site; then $\partial_x L = \langle \sigma_{x_0} \rangle$ and $\partial_x^2 L = 1 - \langle \sigma_{x_0} \rangle^2$ for the corresponding Gibbs average. The second relation in (3.6) implies, by induction, that every derivative of $L(\cdot, \omega)$ of order $r \geq 1$ is a polynomial in $\partial_x L$, so it is bounded by a constant depending only on r .

Let $\Delta_1, \dots, \Delta_K \geq 0$. We set $L_K = L$ and, for $1 \leq \ell \leq K$,

$$L_{\ell-1}(x, \omega_{\leq \ell-1}) = \frac{1}{\zeta_{\ell-1}} \log \mathbb{E} \left[\exp \left(\zeta_{\ell-1} L_\ell(x + \sqrt{\Delta_\ell} Z_\ell, \omega_{\leq \ell-1}, \omega_\ell) \right) \right], \quad (3.7)$$

where the expectation is over (Z_ℓ, ω_ℓ) only. The function L_0 is then a deterministic function of x . All recursions and identities involving auxiliary randomness are understood for almost every auxiliary prefix; on exceptional null sets, we choose arbitrary versions of the recursive functions and transition kernels. For $0 \leq k < K$, given $(X_k, \omega_{\leq k})$, we let (X_{k+1}, ω_{k+1}) have the law of $(X_k + \sqrt{\Delta_{k+1}} Z_{k+1}, \omega_{k+1})$ reweighted by the density

$$\exp \left(\zeta_k (L_{k+1}(X_{k+1}, \omega_{\leq k+1}) - L_k(X_k, \omega_{\leq k})) \right), \quad (3.8)$$

which has unit mean by (3.7). This defines a Markov chain $(X_k, \omega_{\leq k})_{k \leq K}$ once the law of the initial state X_0 is fixed; we write $\widehat{\mathbb{E}}$ for expectations with respect to it. We set

$$M_j = \partial_x L_j(X_j, \omega_{\leq j}), \quad \Lambda_{k,j}(x, \omega_{\leq k}) = \widehat{\mathbb{E}}[M_j^2 \mid X_k = x, \omega_{\leq k}] \quad (0 \leq k \leq j \leq K).$$

In particular $\Lambda_{k,k} = (\partial_x L_k)^2$.

Lemma 3.1. *The following hold for every $0 \leq k \leq K$.*

(a) *The function $L_k(\cdot, \omega_{\leq k})$ is smooth, $|\partial_x L_k| \leq 1$, each derivative of order $r \geq 1$ is bounded by a constant depending only on r and K , and $\partial_x L_k(X_k, \omega_{\leq k}) = \widehat{\mathbb{E}}[M_{k+1} \mid X_k, \omega_{\leq k}]$ if $k < K$. Further, $0 \leq \Lambda_{k,k} \leq \Lambda_{k,k+1} \leq \dots \leq \Lambda_{k,K} \leq 1$.*

(b) *We have*

$$\partial_x^2 L_k = 1 - \zeta_k \Lambda_{k,k} - \sum_{j=k+1}^K w_j \Lambda_{k,j}. \quad (3.9)$$

(c) *For $k < \ell \leq K$, the function L_k is differentiable in $\Delta_\ell \in [0, \infty)$, with a one-sided derivative at $\Delta_\ell = 0$, and*

$$\partial_{\Delta_\ell} L_k = \frac{1}{2} \left(1 - \sum_{j=\ell}^K w_j \Lambda_{k,j} \right). \quad (3.10)$$

Proof. For $\ell \leq K$, suppose that L_ℓ satisfies the bounds in (a) and that $\mathbb{E} \exp(c|L_\ell(0, \omega_{\leq \ell})|) < \infty$ for all $c > 0$. Since $|L_\ell(x, \cdot) - L_\ell(0, \cdot)| \leq |x|$, differentiation under the expectation in (3.7) is justified by dominated convergence, and

$$\partial_x L_{\ell-1} = \widehat{\mathbb{E}}_{\ell-1}[\partial_x L_\ell], \quad \partial_x^2 L_{\ell-1} = \widehat{\mathbb{E}}_{\ell-1}[\partial_x^2 L_\ell] + \zeta_{\ell-1} \left(\widehat{\mathbb{E}}_{\ell-1}[(\partial_x L_\ell)^2] - (\partial_x L_{\ell-1})^2 \right), \quad (3.11)$$

where $\widehat{\mathbb{E}}_{\ell-1}$ denotes the expectation of the step from level $\ell-1$ to level ℓ in (3.8). Higher derivatives of $L_{\ell-1}$ are polynomials, with coefficients bounded in terms of the order, in tilted moments of derivatives of L_ℓ , and they are bounded by constants depending only on the order and on K . Let $Y = |L_\ell(0, \cdot)| + \sqrt{\Delta_\ell} |Z_\ell|$. By (3.7) and Jensen's inequality, $|L_{\ell-1}(0, \cdot)| \leq \zeta_{\ell-1}^{-1} \log \mathbb{E}_\ell e^{\zeta_{\ell-1} Y}$, where $\mathbb{E}_\ell$ integrates (Z_ℓ, ω_ℓ) . For $c \geq \zeta_{\ell-1}$, using Jensen's inequality again, we obtain $\mathbb{E} \exp(c|L_{\ell-1}(0, \cdot)|) \leq \mathbb{E} e^{cY} < \infty$. By downward induction, this proves the first two assertions in (a) and the martingale identity. Since $M_k = \widehat{\mathbb{E}}[M_j \mid X_k, \omega_{\leq k}]$ for $k \leq j$, the monotonicity of $j \mapsto \Lambda_{k,j}$ follows from the conditional form of Jensen's inequality, and $|M_j| \leq 1$ implies that $\Lambda_{k,K} \leq 1$.

We prove (b) by downward induction. For $k = K$ it is the second relation in (3.6), since $\zeta_K = 1$. Suppose that (3.9) holds at level $k+1$. By the Markov property, $\widehat{\mathbb{E}}_k[\Lambda_{k+1,j}] = \Lambda_{k,j}$ for $j \geq k+1$, and inserting this into (3.11) we obtain

$$\partial_x^2 L_k = 1 - \zeta_{k+1} \Lambda_{k,k+1} - \sum_{j \geq k+2} w_j \Lambda_{k,j} + \zeta_k \Lambda_{k,k+1} - \zeta_k \Lambda_{k,k}.$$

Since $\zeta_{k+1} - \zeta_k = w_{k+1}$, this is (3.9) at level k .

For (c), let first $\ell = k + 1$. By (3.5), applied for fixed $(\omega_{\leq k+1})$ and then averaged over ω_{k+1} , we have

$$\partial_{\Delta_{k+1}} L_k = \frac{1}{2\zeta_k} \frac{\partial_x^2 \exp(\zeta_k L_k)}{\exp(\zeta_k L_k)} = \frac{1}{2} (\partial_x^2 L_k + \zeta_k (\partial_x L_k)^2),$$

and (3.9) turns the right side into $(1 - \sum_{j \geq k+1} w_j \Lambda_{k,j})/2$. If $\ell > k + 1$, then $\partial_{\Delta_\ell} L_k = \widehat{\mathbb{E}}_k[\partial_{\Delta_\ell} L_{k+1}]$, because Δ_ℓ enters (3.7) at level k only through L_{k+1} . By induction on $\ell - k$ and the identity $\widehat{\mathbb{E}}_k[\Lambda_{k+1,j}] = \Lambda_{k,j}$, we obtain (3.10). $\square$

3.3 The scalar transform

We now specialize Lemma 3.1 to $L(x) = \log 2 \cosh x$, without auxiliary randomness. For $v \in \mathcal{C}_K$ we take $\Delta_\ell = v_\ell - v_{\ell-1}$ for $1 \leq \ell \leq K$, write $\phi_k = L_k$, and let the chain start from $X_0 = h + \sqrt{v_0} Z_0$. The *compensated scalar transform* is

$$\Psi_h(v) = \frac{v_K}{2} - \mathbb{E} \phi_0(h + \sqrt{v_0} Z_0), \quad \psi_h(q) = \Psi_h(2q) \quad (v, q \in \mathcal{C}_K). \quad (3.12)$$

A level of ψ_h with endpoint q_j has Gaussian variance $2q_j$. For $h = 0$ and up to the additive constant $\log 2$, the function ψ_h is the initial condition of the Hamilton–Jacobi equation for the Sherrington–Kirkpatrick model in [107]; see also [56, Lemma 6.4]. We also write

$$T_j(h) = T_j(h; v) = \mathbb{E} \Lambda_{0,j}(h + \sqrt{v_0} Z_0) = \widehat{\mathbb{E}}[\phi_j'(X_j)^2] \quad (0 \leq j \leq K). \quad (3.13)$$

Lemma 3.2. *For every $h \in \mathbb{R}$, the function Ψ_h is differentiable on $\mathcal{C}_K$, and*

$$\partial_{v_j} \Psi_h(v) = \frac{w_j}{2} T_j(h; v) \quad (0 \leq j \leq K), \quad 0 \leq T_0 \leq T_1 \leq \dots \leq T_K \leq 1. \quad (3.14)$$

The function ψ_h is differentiable on $\mathcal{C}_K$ with $D\psi_h(q) = (T_j(h; 2q))_{j \leq K} \in \mathcal{C}_K \cap [0, 1]^{K+1}$. In particular, $|\psi_h(q) - \psi_h(q')| \leq \|q - q'\|_{1,w}$ for $q, q' \in \mathcal{C}_K$, and $\psi_h(q) \leq \psi_h(q')$ if moreover $q \leq q'$ coordinatewise.

Proof. Iterating (3.5) through the finite recursion gives all mixed derivatives of ϕ_0 in x and in the durations $\Delta_\ell \in [0, \infty)$. On compact sets of durations, the differentiated integrands have a common Gaussian-integrable exponential bound, by Lemma 3.1(a). Dominated convergence therefore gives joint continuity of these derivatives, with one-sided derivatives when a duration is zero. Applying the same argument to the outer convolution $\gamma_{v_0} * \phi_0$ shows that Ψ_h is smooth in $(v_0, \Delta_1, \dots, \Delta_K)$ on $[0, \infty)^{K+1}$ in this sense, and differentiable on $\mathcal{C}_K$. For $1 \leq j < K$, the endpoint v_j enters only Δ_j and Δ_{j+1} , and by Lemma 3.1(c) we have

$$\partial_{v_j} \Psi_h = -\mathbb{E}[(\partial_{\Delta_j} - \partial_{\Delta_{j+1}}) \phi_0(X_0)] = -\frac{1}{2} \mathbb{E} \left[\sum_{i \geq j+1} w_i \Lambda_{0,i}(X_0) - \sum_{i \geq j} w_i \Lambda_{0,i}(X_0) \right] = \frac{w_j}{2} T_j.$$

For $j = K$, the endpoint enters Δ_K and the term $v_K/2$, and

$$\partial_{v_K} \Psi_h = \frac{1}{2} - \frac{1}{2} \mathbb{E}[1 - w_K \Lambda_{0,K}(X_0)] = \frac{w_K}{2} T_K.$$

The endpoint v_0 enters the variance of X_0 and Δ_1 . By (3.5), (3.9), and (3.10),

$$\begin{aligned} \partial_{v_0} \Psi_h &= -\frac{1}{2} \mathbb{E}[\phi_0''(X_0)] + \mathbb{E}[\partial_{\Delta_1} \phi_0(X_0)] \\ &= -\frac{1}{2} \mathbb{E}\left[1 - \zeta_0 \Lambda_{0,0} - \sum_{j \geq 1} w_j \Lambda_{0,j}\right] + \frac{1}{2} \mathbb{E}\left[1 - \sum_{j \geq 1} w_j \Lambda_{0,j}\right], \end{aligned}$$

which equals $w_0 T_0/2$ because $\zeta_0 = w_0$. The ordering of the T_j follows from Lemma 3.1(a). The assertions about ψ_h follow from the chain rule and (3.4). $\square$

Up to a change of variables, the formula (3.14) is [112, Lemma 1], and for $h = 0$ it also appears in [37, Proposition 2.5]. The nonnegativity and the ordering of the derivatives are a special case of [108, Proposition 3.6].

3.4 Complete monotonicity

A function $f: [0, \infty) \rightarrow \mathbb{R}$ is *completely monotone* if it is smooth on $[0, \infty)$, with one-sided derivatives at 0, and $(-1)^k f^{(k)} \geq 0$ on $[0, \infty)$ for every integer $k \geq 0$. A function $g: [0, \infty) \rightarrow [0, \infty)$ is a *Bernstein function* if it is smooth and g' is completely monotone. For an even function F on $\mathbb{R}$ we write $\tilde{F}(s) = F(\sqrt{s})$, $s \geq 0$, and we say that F is completely monotone in x^2 if $\tilde{F}$ is completely monotone. If F is smooth and even, then

$$\partial_s \tilde{F}(s) = (\mathcal{D}F)(\sqrt{s}), \quad (\mathcal{D}F)(x) = \frac{1}{2} \int_0^1 F''(x\theta) d\theta, \quad (3.15)$$

because $F'(0) = 0$. Since the operator $\mathcal{D}$ maps smooth even functions to smooth even functions, $\tilde{F}$ is smooth on $[0, \infty)$, its derivatives are $\partial_s^k \tilde{F}(s) = (\mathcal{D}^k F)(\sqrt{s})$, and $|\partial_s^k \tilde{F}(s)| \leq 2^{-k} \sup_{|y| \leq \sqrt{s}} |F^{(2k)}(y)|$.

We use the following facts.

- Lemma 3.3.** (i) *Products and nonnegative linear combinations of completely monotone functions are completely monotone. If f is completely monotone, then so are $s \mapsto f(as)$ for $a \geq 0$ and $s \mapsto \int_0^1 f(s\theta^2) d\theta$.*
- (ii) *If g is a Bernstein function, then e^{-g} is completely monotone. If moreover $g > 0$, then $1/g$ is completely monotone.*
- (iii) *A function f is completely monotone if and only if $f(s) = \int_{[0, \infty)} e^{-\lambda s} \mu(d\lambda)$ for a finite positive measure μ with $\int \lambda^k \mu(d\lambda) < \infty$ for every k .*

(iv) If G is completely monotone in x^2 and $b \geq 0$, then $\gamma_b * G$ is completely monotone in x^2 , and $b \mapsto \gamma_b * G(0)$ is completely monotone on $[0, \infty)$.

Proof. Part (i) follows from the Leibniz rule and differentiation under the integral sign. For (ii), let $H = e^{-g}$. Then $H' = -g'H$ and, by the Leibniz rule,

$$(-1)^{k+1}H^{(k+1)} = \sum_{i=0}^k \binom{k}{i} [(-1)^i g^{(i)}] [(-1)^{k-i} H^{(k-i)}],$$

which is nonnegative by induction on k . If $g > 0$ and $H = 1/g$, then $H' = -g'H^2$, and the same induction, applied to H and to H^2 simultaneously, shows that $(-1)^k H^{(k)} \geq 0$. Part (iii) is Bernstein's theorem [124, Theorem 1.4]; the measure is finite because $f(0) < \infty$, and its moments are finite because f is smooth at 0. For (iv), we use (iii) to write $\tilde{G}(s) = \int e^{-\lambda s} \mu(d\lambda)$. For $\lambda \geq 0$ and $b \geq 0$ we have

$$\mathbb{E} \exp(-\lambda(x + \sqrt{b}Z)^2) = (1 + 2b\lambda)^{-1/2} \exp\left(-\frac{\lambda x^2}{1 + 2b\lambda}\right).$$

Integrating against μ represents $\gamma_b * G$ as a mixture of the functions $x \mapsto e^{-\lambda' x^2}$ with $\lambda' = \lambda/(1 + 2b\lambda) \geq 0$, with finite positive weights; differentiation under the integral is justified because $\lambda' \leq \lambda$ and μ has finite moments. At $x = 0$ we obtain $\int (1 + 2b\lambda)^{-1/2} \mu(d\lambda)$, and $b \mapsto (1 + 2b\lambda)^{-1/2}$ is completely monotone. $\square$

The derivatives that we must control satisfy linear parabolic equations in the variable $s = x^2$, which degenerate at $s = 0$. We use the following maximum principle.

Lemma 3.4. *Let $d > 0$ and let $z: [0, d] \times [0, \infty) \rightarrow \mathbb{R}$ be continuous, such that $\partial_b z$, $\partial_s z$, and $\partial_s^2 z$ exist and are continuous on $(0, d] \times [0, \infty)$, with one-sided derivatives at $s = 0$. Suppose that $|z(b, s)| \leq C_z e^{\sqrt{s}}$, that $z(0, \cdot) \geq 0$, and that*

$$\partial_b z \geq 2s \partial_s^2 z + \beta(b, s) \partial_s z + V(b, s) z \quad \text{on } (0, d] \times [0, \infty),$$

where $0 \leq \beta(b, s) \leq C_\beta(1 + \sqrt{s})$ and $|V(b, s)| \leq C_V$. Then $z \geq 0$.

Proof. Let $\Phi(s) = \cosh(2\sqrt{s}) = \sum_{k \geq 0} 4^k s^k / (2k)!$, which is smooth on $[0, \infty)$. We have $\Phi'(s) = \sinh(2\sqrt{s}) / \sqrt{s} \geq 0$ and $2s\Phi''(s) = 2\cosh(2\sqrt{s}) - \sinh(2\sqrt{s}) / \sqrt{s}$. Using $\sinh y \leq y \cosh y$ and $\sinh y \leq \cosh y$ for $y \geq 0$, we find

$$2s\Phi'' + \beta\Phi' \leq 2\Phi + C_\beta \left(\frac{\sinh(2\sqrt{s})}{\sqrt{s}} + \sinh(2\sqrt{s}) \right) \leq (2 + 3C_\beta)\Phi.$$

Set $C_\Phi = 2 + 3C_\beta$, fix $\Lambda > C_V$, and for $\epsilon > 0$ let $z_\epsilon(b, s) = e^{-\Lambda b} z(b, s) + \epsilon e^{C_\Phi b} \Phi(s)$. Since $\Phi(s) \geq e^{2\sqrt{s}}/2$, the function z_ϵ tends to $+\infty$ as $s \rightarrow \infty$, uniformly in $b \in [0, d]$, so it

attains its minimum on $[0, d] \times [0, \infty)$ at some point (b_1, s_1) . Suppose that this minimum is negative. Then $b_1 > 0$, because $z_\epsilon(0, \cdot) > 0$. At (b_1, s_1) we have $\partial_b z_\epsilon \leq 0$; moreover $2s\partial_s^2 z_\epsilon + \beta\partial_s z_\epsilon \geq 0$, since either $s_1 > 0$, in which case $\partial_s z_\epsilon = 0 \leq \partial_s^2 z_\epsilon$, or $s_1 = 0$, in which case $\partial_s z_\epsilon \geq 0$ and the second-order term vanishes. On the other hand, by the assumed inequality, we have

$$\partial_b z_\epsilon \geq (2s\partial_s^2 z_\epsilon + \beta\partial_s z_\epsilon) + \epsilon e^{C_\Phi b} (C_\Phi \Phi - 2s\Phi'' - \beta\Phi') + (V - \Lambda)e^{-\Lambda b} z.$$

At (b_1, s_1) , the first two terms are nonnegative and $e^{-\Lambda b} z = z_\epsilon - \epsilon e^{C_\Phi b} \Phi < 0$, so the last term is positive. Then $\partial_b z_\epsilon > 0$ there, a contradiction. We conclude that $z_\epsilon \geq 0$ for all $\epsilon > 0$, and $z \geq 0$. $\square$

We now show that one step of the recursion (3.7), without auxiliary randomness, preserves the relevant complete monotonicity properties.

Lemma 3.5. *Let $0 < m \leq m' \leq 1$. Let $\phi_{\text{in}}: \mathbb{R} \rightarrow \mathbb{R}$ be smooth, even, and convex, with $|\phi'_{\text{in}}| \leq 1$ and with bounded derivatives of every order $r \geq 1$. Let $\Lambda_{\text{in}}^\iota$, $\iota \in \mathcal{I}$, be finitely many smooth even functions with values in $[0, 1]$ and bounded derivatives of every order. Suppose that for every $c \in [0, m']$ and $\iota \in \mathcal{I}$ the functions*

$$\phi_{\text{in}}'', \quad e^{c\phi_{\text{in}}} \phi_{\text{in}}'', \quad e^{c\phi_{\text{in}}} (1 - (\phi'_{\text{in}})^2), \quad e^{c\phi_{\text{in}}} (1 - \Lambda_{\text{in}}^\iota) \quad (3.16)$$

are completely monotone in x^2 . For $b \geq 0$ let

$$\phi(b, \cdot) = \frac{1}{m} \log(\gamma_b * e^{m\phi_{\text{in}}}), \quad \Lambda^\iota(b, \cdot) = \frac{\gamma_b * (e^{m\phi_{\text{in}}} \Lambda_{\text{in}}^\iota)}{\gamma_b * e^{m\phi_{\text{in}}}}.$$

Then, for every $b \geq 0$, the functions $\phi(b, \cdot)$ and $\Lambda^\iota(b, \cdot)$ satisfy the same assumptions with m' replaced by m .

Proof. Fix $d > 0$; we prove the claim for $b \in [0, d]$. The function $\phi(b, \cdot)$ is even and smooth, and (3.11), with $\zeta_{\ell-1}$ replaced by m , shows that $|\partial_x \phi| \leq 1$, $\partial_x^2 \phi \geq 0$, and that the derivatives of order $r \geq 1$ are bounded uniformly in $b \in [0, d]$. Similarly, $\Lambda^\iota(b, \cdot)$ takes values in $[0, 1]$ and has bounded derivatives, since it is a tilted average of $\Lambda_{\text{in}}^\iota$. Let $u = \partial_x \phi$ and $\psi = \partial_x^2 \phi$. By (3.5), for $b > 0$,

$$\partial_b \phi = \frac{1}{2} (\partial_x^2 \phi + m(\partial_x \phi)^2), \quad \partial_b u = \frac{1}{2} \partial_x^2 u + mu \partial_x u, \quad \partial_b \psi = \frac{1}{2} \partial_x^2 \psi + mu \partial_x \psi + m\psi^2, \quad (3.17)$$

and $2\partial_b \Lambda^\iota = \partial_x^2 \Lambda^\iota + 2mu \partial_x \Lambda^\iota$. For $c \in [0, m]$ we set $H_c = e^{c\phi} \psi$, $G_c = e^{c\phi} (1 - u^2)$, and $\Theta_c^\iota = e^{c\phi} (1 - \Lambda^\iota)$. Using (3.17), we compute

$$\begin{aligned} \partial_b H_c &= \frac{1}{2} \partial_x^2 H_c + (m - c)u \partial_x H_c + \left(m\psi - \frac{1}{2}c(m - c)u^2\right) H_c, \\ \partial_b G_c &= \frac{1}{2} \partial_x^2 G_c + (m - c)u \partial_x G_c - \frac{1}{2}c(m - c)u^2 G_c + H_c \psi, \\ \partial_b \Theta_c^\iota &= \frac{1}{2} \partial_x^2 \Theta_c^\iota + (m - c)u \partial_x \Theta_c^\iota - \frac{1}{2}c(m - c)u^2 \Theta_c^\iota. \end{aligned} \quad (3.18)$$

We pass to the variable $s = x^2$. Let $f(b, s) = \psi(b, \sqrt{s})$ and $b_*(b, s) = \sqrt{s} u(b, \sqrt{s})$. For an even function $F(b, \cdot)$ we have $\partial_x^2 F = 2\partial_s \tilde{F} + 4s \partial_s^2 \tilde{F}$ and $\gamma u \partial_x F = 2\gamma b_* \partial_s \tilde{F}$, where $\tilde{F}(b, s) = F(b, \sqrt{s})$. Since $u(b, 0) = 0$, we have $u(b, x) = x \int_0^1 \psi(b, x\theta) d\theta$, which implies

$$\partial_s b_* = \frac{1}{2} \left[f(s) + \int_0^1 f(s\theta^2) d\theta \right], \quad \partial_s (u(\sqrt{s})^2) = f(s) \int_0^1 f(s\theta^2) d\theta. \quad (3.19)$$

Since $\phi(b, \cdot)$ is even and convex, $u(b, x)$ has the sign of x , and $0 \leq b_* \leq \sqrt{s}$ because $|u| \leq 1$. Since ψ is nonnegative and bounded, we see from (3.19) that $\partial_s b_*$ is nonnegative and bounded.

All functions below are smooth in $(b, s) \in (0, d] \times [0, \infty)$ and continuous on $[0, d] \times [0, \infty)$ together with their s -derivatives, by (3.15) and the smoothness of the heat semigroup. Since $\phi(b, x) \leq \phi(b, 0) + |x|$ and $\phi(b, 0)$ is bounded on $[0, d]$, every x -derivative of H_c , G_c , and Θ_c^t is bounded by $Ce^{c|x|} \leq Ce^{|x|}$, and by (3.15) every s -derivative of their transforms is bounded by $Ce^{\sqrt{s}}$. The functions f , b_* , and u^2 have bounded derivatives.

Step 1: $f(b, \cdot)$ is completely monotone. Set $\omega_k = (-1)^k \partial_s^k f$ and, for $j \geq 1$, $B_j = (-1)^{j-1} \partial_s^j b_*$. By (3.19),

$$B_j = \frac{1}{2} \left[\omega_{j-1}(s) + \int_0^1 \theta^{2j-2} \omega_{j-1}(s\theta^2) d\theta \right].$$

In the variable s , the last equation in (3.17) reads $\partial_b f = 2s \partial_s^2 f + (1 + 2mb_*) \partial_s f + mf^2$. Differentiating $k \geq 1$ times in s , we obtain

$$\begin{aligned} \partial_b \omega_k &= 2s \partial_s^2 \omega_k + (2k + 1 + 2mb_*) \partial_s \omega_k + (2mk \partial_s b_* + 2mf) \omega_k \\ &\quad + m \sum_{i=1}^{k-1} \binom{k}{i} \omega_i \omega_{k-i} + 2m \sum_{i=2}^k \binom{k}{i} B_i \omega_{k-i+1}. \end{aligned} \quad (3.20)$$

We have $\omega_0 = f \geq 0$ because $\phi(b, \cdot)$ is convex. Suppose that $\omega_0, \dots, \omega_{k-1} \geq 0$ on $[0, d] \times [0, \infty)$. Then $B_2, \dots, B_k \geq 0$, and the two sums in (3.20) are nonnegative, because they involve only $\omega_1, \dots, \omega_{k-1}$ and $B_2, \dots, B_k$. Further, $\omega_k(0, \cdot) \geq 0$ by (3.16). Lemma 3.4, applied with $\beta = 2k + 1 + 2mb_*$ and $V = 2mk \partial_s b_* + 2mf$, implies that $\omega_k \geq 0$. By induction, $f(b, \cdot)$ is completely monotone. By (3.19) and Lemma 3.3(i), $(-1)^{j-1} \partial_s^j b_* \geq 0$ and $(-1)^{j-1} \partial_s^j (u^2) \geq 0$ for all $j \geq 1$.

Step 2: a general induction. Let F be one of H_c , G_c , Θ_c^t , with $c \in [0, m]$ and $\gamma = m - c \geq 0$. By (3.18), $\tilde{F}$ satisfies

$$\partial_b \tilde{F} = 2s \partial_s^2 \tilde{F} + (1 + 2\gamma b_*) \partial_s \tilde{F} + P\tilde{F} + J,$$

where $P = mf - c\gamma u^2/2$ and $J = 0$ for H_c , $P = -c\gamma u^2/2$ and $J = \widetilde{H}_c f$ for G_c , and $P = -c\gamma u^2/2$ and $J = 0$ for Θ_c^l . By Step 1, $(-1)^j P^{(j)} \geq 0$ for every $j \geq 1$, where derivatives are in s . Let $z_k = (-1)^k \partial_s^k \widetilde{F}$. Differentiating k times, we find

$$\begin{aligned} \partial_b z_k &= 2s \partial_s^2 z_k + (2k + 1 + 2\gamma b_*) \partial_s z_k + (2\gamma k \partial_s b_* + P) z_k \\ &\quad + 2\gamma \sum_{i=2}^k \binom{k}{i} B_i z_{k-i+1} + \sum_{i=1}^k \binom{k}{i} (-1)^i P^{(i)} z_{k-i} + (-1)^k J^{(k)}. \end{aligned}$$

If J is completely monotone and $z_0, \dots, z_{k-1} \geq 0$, then all source terms are nonnegative, $z_k(0, \cdot) \geq 0$ by (3.16), and Lemma 3.4 gives $z_k \geq 0$. The base case $z_0 = \widetilde{F} \geq 0$ holds because $\psi \geq 0$, $|u| \leq 1$, and $\Lambda' \leq 1$.

We apply Step 2 first to H_c , with $J = 0$, which shows that $H_c(b, \cdot)$ is completely monotone in x^2 . Then $J = \widetilde{H}_c f$ is completely monotone, and Step 2 applies to G_c . Finally it applies to Θ_c^l . Together with Step 1, this proves the lemma. $\square$

Lemma 3.6. *Let $v \in \mathcal{C}_K$, and let ϕ_k and $\Lambda_{k,j}$ be as in Section 3.3. For every $0 \leq k \leq j \leq K$ and $c \in [0, \zeta_k]$, the functions*

$$\phi_k'', \quad e^{c\phi_k} \phi_k'', \quad e^{c\phi_k} (1 - \Lambda_{k,j})$$

are completely monotone in x^2 . Further, ϕ_k' is odd and nondecreasing, $\Lambda_{k,j}$ is even and nondecreasing in $|x|$, and for every j there exists a finite positive measure ν_j on $[0, \infty)$, with finite moments and total mass $1 - T_j(0)$, such that

$$1 - T_j(h) = \int_{[0, \infty)} e^{-\lambda h^2} \nu_j(d\lambda) \quad (h \in \mathbb{R}). \quad (3.21)$$

Proof. We prove the complete-monotonicity claims by downward induction, carrying also that ϕ_k and $\Lambda_{k,j}$ are smooth and even, that their derivatives of every positive order are bounded, and that $0 \leq \Lambda_{k,j} \leq 1$. For $k = K$ these additional properties hold, and we have $\phi_K = \log 2 \cosh x$, $\phi_K'' = 1 - (\phi_K')^2 = \text{sech}^2 x$, $\Lambda_{K,K} = \tanh^2 x$, and $\zeta_K = 1$. The three functions are then sech^2 and 2^csech^{2-c} with $2 - c \geq 1$. By the product formula $\cosh x = \prod_{i \geq 1} (1 + 4x^2 / ((2i - 1)^2 \pi^2))$ [55, (4.36.2)],

$$\partial_s \log \cosh \sqrt{s} = \sum_{i \geq 1} \frac{4}{(2i - 1)^2 \pi^2 + 4s},$$

where the series of derivatives of every order converges uniformly on $[0, \infty)$. Since each summand is completely monotone, $\log \cosh \sqrt{s}$ is a Bernstein function, and $\text{sech}^p \sqrt{s} = e^{-p \log \cosh \sqrt{s}}$ is completely monotone for all $p > 0$ by Lemma 3.3(ii).

Suppose that the strengthened claim holds at level $k + 1 \leq K$. We apply Lemma 3.5 with $m = \zeta_k$, $m' = \zeta_{k+1}$, $\phi_{\text{in}} = \phi_{k+1}$, the family $\Lambda_{\text{in}}^j = \Lambda_{k+1,j}$ for $k + 1 \leq j \leq K$, and $b = \Lambda_{k+1}$. The hypotheses hold by the induction hypothesis and Lemma 3.1(a):

the function ϕ_{k+1} is even, because $\log 2 \cosh$ is even and the Gaussian increments are symmetric; it is convex because ϕ_{k+1}'' , being completely monotone in x^2 , is nonnegative; and $(\phi_{k+1}')^2 = \Lambda_{k+1,k+1}$. By the definition of the tilted chain, $\phi(\Delta_{k+1}, \cdot) = \phi_k$ and $\Lambda^j(\Delta_{k+1}, \cdot) = \Lambda_{k,j}$ for $j \geq k+1$, and $\Lambda_{k,k} = (\phi_k')^2$. Lemma 3.5 preserves the additional regularity and range bounds, which also hold for $(\phi_k')^2$, and implies the claim at level k .

Since ϕ_k is even, ϕ_k' is odd, and it is nondecreasing because $\phi_k'' \geq 0$. Since $1 - \Lambda_{k,j}$ is completely monotone in x^2 , it is nonincreasing in $|x|$. Finally, $1 - \Lambda_{0,j}$ is completely monotone in x^2 , so $1 - T_j = \gamma_{v_0} * (1 - \Lambda_{0,j})$ is completely monotone in h^2 by Lemma 3.3(iv), and (3.21) follows from Lemma 3.3(iii). $\square$

The monotonicity of ϕ_k' and of $\Lambda_{k,j}$ in $|x|$ also follows from [112, Lemma 2]. The complete monotonicity is the additional information used in Lemma 3.7 and Proposition 3.8.

3.5 Square-root convexity

The next estimate bounds the curvature of T_j at $h = 0$ in terms of the total variance v_j . We write T_j'' for the second derivative in h .

Lemma 3.7. *For every $v \in \mathcal{C}_K$ and $0 \leq j \leq K$, we have $v_j T_j''(0) \leq 2T_j(0)$.*

Proof. Fix j . We follow the response $\Lambda_{k,j}$ as k decreases from j to 0, and then through the outer Gaussian average. Let $0 \leq k < j$, and consider the step from level $k+1$ to level k as a function of its duration $b \in [0, \Delta_{k+1}]$, as in Lemma 3.5 with $m = \zeta_k$, $\phi_{\text{in}} = \phi_{k+1}$, and $\Lambda_{\text{in}} = \Lambda_{k+1,j}$. We write $D = e^{\zeta_k \phi_{k+1}}$, $B(b) = \gamma_b * D(0)$, and

$$g(b) = \Lambda(b, 0), \quad 1 - g(b) = \frac{N(b)}{B(b)}, \quad N(b) = \gamma_b * (D(1 - \Lambda_{k+1,j}))(0).$$

By Lemma 3.6 at level $k+1$ with $c = \zeta_k$, the function $D(1 - \Lambda_{k+1,i})$ is completely monotone in x^2 for every $i \geq k+1$. By Lemma 3.3(iv), N and the functions $N_i(b) = \gamma_b * (D(1 - \Lambda_{k+1,i}))(0)$ are completely monotone in b . By (3.9) at level $k+1$ and $\zeta_k + (\zeta_{k+1} - \zeta_k) + \sum_{i>k+1} w_i = 1$,

$$\begin{aligned} D'' &= \zeta_k D(\phi_{k+1}'' + \zeta_k (\phi_{k+1}')^2) \\ &= \zeta_k^2 D + \zeta_k (\zeta_{k+1} - \zeta_k) D(1 - \Lambda_{k+1,k+1}) + \zeta_k \sum_{i>k+1} w_i D(1 - \Lambda_{k+1,i}). \end{aligned}$$

By (3.5),

$$B' = \frac{1}{2} \gamma_b * D''(0) = \frac{\zeta_k^2}{2} B + \frac{\zeta_k}{2} \sum_{i \geq k+1} c_i N_i, \quad c_i \geq 0.$$

The function $\tilde{B} = e^{-\zeta_k^2 b/2} B$ is then positive with completely monotone derivative. By Lemma 3.3(ii), $1/\tilde{B}$ is completely monotone, and $1 - g = N e^{-\zeta_k^2 b/2} / \tilde{B}$ is completely

monotone by Lemma 3.3(i). It follows that g is nondecreasing and concave on $[0, \Delta_{k+1}]$. For the outer Gaussian average the same holds with $g(b) = \gamma_b * \Lambda_{0,j}(0)$, $b \in [0, v_0]$, directly from Lemma 3.3(iv).

Since $\partial_x \phi(b, 0) = 0$, we have $2g'(b) = \partial_x^2 \Lambda(b, 0)$ by the transport equation for Λ . Suppose that at the start of a step the response satisfies $b_0 g'(0) \leq g(0)$ for some $b_0 \geq 0$. By concavity, $bg'(b) \leq g(b) - g(0)$ and $g'(b) \leq g'(0)$. Using these bounds, we obtain

$$(b_0 + b)g'(b) \leq g(b) - g(0) + b_0 g'(0) \leq g(b).$$

At level j the response is $\Lambda_{j,j} = (\phi_j')^2$, which vanishes at 0, so the inequality holds with $b_0 = 0$. At the junction of two consecutive steps, the values of g and g' are $\Lambda_{k,j}(0)$ and $\Lambda_{k,j}''(0)/2$ for both steps. Applying the previous display through the steps of durations $\Delta_j, \dots, \Delta_1$ and then through the outer average of variance v_0 , whose total is v_j , we obtain $v_j T_j''(0) \leq 2T_j(0)$. $\square$

Proposition 3.8. *For every $h \in \mathbb{R}$, the map $s \mapsto \Psi_h(s_0^2, \dots, s_K^2)$ is convex on $\{s \in \mathbb{R}^{K+1} : 0 \leq s_0 \leq s_1 \leq \dots \leq s_K\}$. The same holds for $s \mapsto \psi_h(s_0^2, \dots, s_K^2)$.*

Proof. The second statement follows from the first, since $\psi_h(s^2) = \Psi_h((\sqrt{2}s)^2)$. Let $\mathcal{O} = \{s : 0 < s_0 < s_1 < \dots < s_K\}$, an open convex set whose closure is the set in the statement. Since $s \mapsto \Psi_h(s^2)$ is continuous on the closure, it suffices to show that its Hessian is positive semidefinite on $\mathcal{O}$. Fix $s \in \mathcal{O}$ and $v = s^2$, so that $0 < v_0 < \dots < v_K$, and set $J_{ij} = \partial_{v_j} T_i(h; v)$.

We first show that $J_{ij} \leq 0$ for $i < j$. Under the tilted chain, the law of $(X_0, \dots, X_i)$ has density $\prod_{\ell < i} \exp(\zeta_\ell(\phi_{\ell+1}(X_{\ell+1}) - \phi_\ell(X_\ell)))$ with respect to the Gaussian path with variances $v_0, \Delta_1, \dots, \Delta_i$, which do not depend on v_j . For $\ell < j$, we have $\partial_{v_j} \phi_\ell = c_j - w_j \Lambda_{\ell,j}/2$ by Lemma 3.1(c), where $c_j = 1/2$ if $j = K$ and $c_j = 0$ otherwise. Differentiating $T_i = \widehat{\mathbb{E}}[\phi_i'(X_i)^2]$ under the expectation, we obtain

$$J_{ij} = -w_j \widehat{\mathbb{E}}[\phi_i'(X_i) \partial_x \Lambda_{i,j}(X_i)] - \frac{w_j}{2} \sum_{\ell < i} \zeta_\ell \widehat{\mathbb{E}}[\phi_i'(X_i)^2 (\Lambda_{\ell+1,j}(X_{\ell+1}) - \Lambda_{\ell,j}(X_\ell))],$$

where the constants c_j cancel between the two terms of each tilt. By the Markov property, $\widehat{\mathbb{E}}[\phi_i'(X_i)^2 | X_{\ell+1}] = \Lambda_{\ell+1,i}(X_{\ell+1})$ and $\Lambda_{\ell,j}(X_\ell) = \widehat{\mathbb{E}}[\Lambda_{\ell+1,j}(X_{\ell+1}) | X_\ell]$. The expectation in the sum is the mean of the conditional covariance of $\Lambda_{\ell+1,i}(X_{\ell+1})$ and $\Lambda_{\ell+1,j}(X_{\ell+1})$ given X_ℓ . By Lemma 3.6, both are nondecreasing functions of $|X_{\ell+1}|$, so this covariance is nonnegative: for two such functions F_1, F_2 and an independent copy Y' of a variable Y , $2 \text{Cov}(F_1(Y), F_2(Y)) = \mathbb{E}[(F_1(Y) - F_1(Y'))(F_2(Y) - F_2(Y'))] \geq 0$. Also, by Lemma 3.6, $\phi_i'(x)$ and $\partial_x \Lambda_{i,j}(x)$ have the sign of x . This shows that $J_{ij} \leq 0$. Since the Hessian of Ψ_h in v is symmetric, (3.14) implies that $w_i J_{ij} = w_j J_{ji}$. Then $J_{ij} \leq 0$ for all $i \neq j$.

Adding the same number to every endpoint leaves $\Delta_1, \dots, \Delta_K$ unchanged and increases v_0 . Using (3.5), we find

$$\sum_{j=0}^K J_{ij} = \frac{d}{d\epsilon} \Big|_{\epsilon=0} \mathbb{E} \Lambda_{0,i}(h + \sqrt{v_0 + \epsilon} Z) = \frac{1}{2} T_i''(h). \quad (3.22)$$

By Lemmas 3.6 and 3.7, with $y = \lambda h^2$ and v_i as in (3.21),

$$\begin{aligned} T_i(h) + v_i T_i''(h) &= 1 - \int e^{-\lambda h^2} v_i(d\lambda) - v_i \int \lambda(4y - 2)e^{-y} v_i(d\lambda) \\ &\geq T_i(0) - 4e^{-3/2} v_i \int \lambda v_i(d\lambda) \geq (1 - 4e^{-3/2}) T_i(0) \geq 0. \end{aligned}$$

Here we used that v_i has mass $1 - T_i(0)$, that $\sup_{y \geq 0} (4y - 2)e^{-y} = 4e^{-3/2} < 1$, and that $T_i''(0) = 2 \int \lambda v_i(d\lambda)$, so that Lemma 3.7 reads $v_i \int \lambda v_i(d\lambda) \leq T_i(0)$.

Let M be the Hessian of $s \mapsto \Psi_h(s^2)$ at s . By (3.14), $\partial_{s_i} \Psi_h(s^2) = w_i s_i T_i$, which gives

$$M_{ij} = \delta_{ij} w_i T_i + 2w_i s_i s_j J_{ij}.$$

The off-diagonal entries are nonpositive. Let $q_i = 1/s_i > 0$. By (3.22),

$$(Mq)_i = \frac{w_i T_i}{s_i} + 2w_i s_i \sum_j J_{ij} = \frac{w_i}{s_i} (T_i + v_i T_i''(h)) \geq 0.$$

For $y \in \mathbb{R}^{K+1}$ and $\xi_i = q_i y_i$, using the symmetry of M , we have

$$\xi^T M \xi = \sum_{i < j} (-M_{ij}) q_i q_j (y_i - y_j)^2 + \sum_i q_i (Mq)_i y_i^2 \geq 0,$$

which shows that M is positive semidefinite. $\square$

For $h = 0$, Chen, Issa, and Mourrat [37, Proposition 2.2] and Ho [83, Theorem 1.8] proved that ψ_0 is convex on $\mathcal{C}_K$. Since ψ_0 is nondecreasing, this is stronger than Proposition 3.8 at $h = 0$. Convexity of ψ_0 along rays through the origin follows from an earlier result of Auffinger and Chen [8, Theorem 1]; see [83, Section 3]. Convexity on $\mathcal{C}_K$ fails for large $|h|$, already for $K = 0$, where $\psi_h(s) = s - \mathbb{E} \log 2 \cosh(h + \sqrt{2s} Z)$. The computation in [106, Section 6.1], with the uniform measure on $\{-1, 1\}$ replaced by the measure with mean $\tanh h$, gives $\psi_h''(0) = 2(1 - \tanh^2 h)(1 - 3 \tanh^2 h)$, which is negative if $\tanh^2 h > 1/3$. The same threshold is stated in [37, Section 1]. For spherical spins, the analogous initial condition is convex in the paths in every field [34, Lemma 2.9]. The proof in [37, Section 2] has the same structure as the proof of Proposition 3.8: the Hessian of ψ_0 has nonpositive off-diagonal entries, which is deduced from [112, Theorem 2], and nonnegative row sums. For $h \neq 0$, the row sums of the Hessian of Ψ_h are proportional to $T_i''(h)$ by (3.22), and they may be negative. In the proof of Proposition 3.8, the change of variables $v = s^2$ and Lemma 3.7 compensate for this.

3.6 The Hopf–Lax formula and the Parisi functional

Fix $\beta > 0$ and $h \in \mathbb{R}$. For $t > 0$ and $y \in \mathcal{C}_K$ we define

$$U(t, y) = \sup_{y' \in \mathcal{C}_K} \left\{ \psi_h(y') - \frac{\|y' - y\|_{2,w}^2}{\beta^2 t} \right\}, \quad U(0, y) = \psi_h(y). \quad (3.23)$$

This is the Hopf–Lax formula for the equation $\partial_t U - (\beta^2/4) \|DU\|_{2,w}^2 = 0$ with initial condition ψ_h ; see [56, Theorem 3.8] for the corresponding statement on $\mathbb{R}^{K+1}$ and [36, Proposition 6.2] for cones. We use only the elementary properties in the next lemma.

Lemma 3.9. *The following hold.*

(a) *For $t > 0$ and $y \in \mathcal{C}_K$ the supremum in (3.23) is attained, and every maximizer y' satisfies*

$$P = \frac{2(y' - y)}{\beta^2 t} = D\psi_h(y') \in \mathcal{C}_K \cap [0, 1]^{K+1}. \quad (3.24)$$

(b) *For $t \geq 0$ and $y, \tilde{y} \in \mathcal{C}_K$, we have $|U(t, y) - U(t, \tilde{y})| \leq \|y - \tilde{y}\|_{1,w}$.*

(c) *For $0 \leq t \leq t'$ and $y \in \mathcal{C}_K$, we have $0 \leq U(t', y) - U(t, y) \leq \beta^2(t' - t)/2$.*

In particular, U is continuous on $[0, \infty) \times \mathcal{C}_K$.

Proof. (a) By Lemma 3.2, $\psi_h(y') \leq \psi_h(y) + \|y' - y\|_{2,w}$, so the function maximized in (3.23) is continuous on the closed set $\mathcal{C}_K$ and tends to $-\infty$ as $|y'| \rightarrow \infty$. Let y' be a maximizer and $P = D\psi_h(y')$. For $y'' \in \mathcal{C}_K$, the segment from y' to y'' lies in $\mathcal{C}_K$, and the right derivative at y' of the maximized function in the direction $y'' - y'$ is nonpositive:

$$\left\langle P - \frac{2(y' - y)}{\beta^2 t}, y'' - y' \right\rangle_w \leq 0.$$

Since $P \in \mathcal{C}_K$ by Lemma 3.2, we may take $y'' = y + \beta^2 t P/2 \in \mathcal{C}_K$. With this choice, the left side equals $(\beta^2 t/2) \|P - 2(y' - y)/(\beta^2 t)\|_{2,w}^2$, which must vanish. This proves (3.24).

(b) For $t = 0$ this is Lemma 3.2. Let $t > 0$, let y' be a maximizer for y , and write $y' = y + \beta^2 t P/2$ as in (a). Since $P \in \mathcal{C}_K$, the point $\tilde{y} + \beta^2 t P/2$ lies in $\mathcal{C}_K$. Taking this point in the supremum, we obtain

$$\begin{aligned} U(t, \tilde{y}) &\geq \psi_h\left(\tilde{y} + \frac{\beta^2 t}{2} P\right) - \frac{\beta^2 t}{4} \|P\|_{2,w}^2 \\ &\geq \psi_h(y') - \|y - \tilde{y}\|_{1,w} - \frac{\beta^2 t}{4} \|P\|_{2,w}^2 = U(t, y) - \|y - \tilde{y}\|_{1,w}. \end{aligned}$$

Exchanging y and $\tilde{y}$ proves (b).

(c) If $t > 0$ and y' is a maximizer at time t , then $U(t', y) \geq \psi_h(y') - \|y' - y\|_{2,w}^2 / (\beta^2 t') \geq U(t, y)$; the case $t = 0$ is similar with $y' = y$. Let $y'' = y + \beta^2 t' P'' / 2$ be a maximizer at time t' , with $P'' \in \mathcal{C}_K \cap [0, 1]^{K+1}$. Since the point $y + \beta^2 t P'' / 2$ lies in $\mathcal{C}_K$, we have

$$\begin{aligned} U(t', y) - U(t, y) &\leq \psi_h\left(y + \frac{\beta^2 t'}{2} P''\right) - \psi_h\left(y + \frac{\beta^2 t}{2} P''\right) - \frac{\beta^2 (t' - t)}{4} \|P''\|_{2,w}^2 \\ &\leq \frac{\beta^2 (t' - t)}{2} \|P''\|_{1,w}, \end{aligned}$$

which is at most $\beta^2 (t' - t) / 2$. $\square$

The next lemma compares $U(1, 0)$ with the Parisi formula. Its identity (3.25) is a finite-dimensional form of the representation of the Parisi formula by a Hopf–Lax formula, due to Mourrat [107] for $h = 0$; see [56, Theorem 6.7]. We use the Parisi equation (2.6), the Parisi functional $\mathcal{P}_{\beta,h}$, and the formula (2.7) for $\mathcal{P}(\beta, h)$, and we write $\mu_r = \sum_{j=0}^K w_j \delta_{r_j}$ for $r \in \mathcal{C}_K \cap [0, 1]^{K+1}$.

Lemma 3.10. *Let $\beta > 0$ and $h \in \mathbb{R}$. For every $r \in \mathcal{C}_K \cap [0, 1]^{K+1}$,*

$$\psi_h\left(\frac{\beta^2}{2} r\right) - \frac{\beta^2}{4} \|r\|_{2,w}^2 = \frac{\beta^2}{4} - \mathcal{P}_{\beta,h}(\mu_r), \quad (3.25)$$

and $U(1, 0) \geq \beta^2 / 4 - \mathcal{P}_{\beta,h}(\mu_r)$. Further, the map $\mu \mapsto \mathcal{P}_{\beta,h}(\mu)$ is continuous for weak convergence of probability measures on $[0, 1]$.

Proof. We first prove (3.25). Let $\Phi = \Phi_{\mu_r}$ solve the Parisi equation (2.6). Since μ_r is atomic, Φ is obtained by successive Cole–Hopf transforms [7]: on $[r_K, 1]$ we have $\mu_r([0, t]) = 1$ and $\Phi(t, x) = \log 2 \cosh x + \beta^2 (1 - t) / 2$; on an interval $[r_j, r_{j+1})$ of positive length we have $\mu_r([0, t]) = \zeta_j$ and $\Phi(r_j, x) = \zeta_j^{-1} \log \mathbb{E} \exp(\zeta_j \Phi(r_{j+1}, x + \beta \sqrt{r_{j+1} - r_j} Z))$; and $\Phi(0, x) = \mathbb{E} \Phi(r_0, x + \beta \sqrt{r_0} Z)$. Let ϕ_j be the recursion of Section 3.3 for the endpoints $v = \beta^2 r$, so that $\Delta_{j+1} = \beta^2 (r_{j+1} - r_j)$. Since constants commute with the Cole–Hopf transforms, we have $\Phi(r_j, \cdot) = \phi_j + \beta^2 (1 - r_K) / 2$ for every j by downward induction, including the case of repeated endpoints. Then $\Phi(0, h) = \mathbb{E} \phi_0(h + \beta \sqrt{r_0} Z) + \beta^2 (1 - r_K) / 2$. If $X \sim \mu_r$, then

$$\int_0^1 t \mu_r([0, t]) dt = \mathbb{E} \int_X^1 t dt = \frac{1 - \|r\|_{2,w}^2}{2}.$$

We conclude that

$$\mathcal{P}_{\beta,h}(\mu_r) = \mathbb{E} \phi_0(h + \beta \sqrt{r_0} Z) + \frac{\beta^2}{4} - \frac{\beta^2}{2} r_K + \frac{\beta^2}{4} \|r\|_{2,w}^2.$$

Since $\psi_h(\beta^2 r / 2) = \Psi_h(\beta^2 r) = \beta^2 r_K / 2 - \mathbb{E} \phi_0(h + \beta \sqrt{r_0} Z)$, this is (3.25). Taking $y' = \beta^2 r / 2$ in (3.23) with $t = 1$ and $y = 0$, for which $\|y'\|_{2,w}^2 / \beta^2 = \beta^2 \|r\|_{2,w}^2 / 4$, we obtain the lower bound on $U(1, 0)$.

Finally, $\Phi_\mu(0, h)$ is continuous in μ for weak convergence by (P1), and so is $\int_0^1 t \mu([0, t]) dt = (1 - \int x^2 \mu(dx)) / 2$. This proves the continuity of $\mathcal{P}_{\beta,h}$. $\square$

The upper bound on regular graphs

The goal of this section is to prove the following proposition, which states that the pressure of the Edwards–Anderson model on any D -regular graph is asymptotically at most the Parisi pressure.

Proposition 4.1. *For every $\beta \geq 0$ and $h \in \mathbb{R}$,*

$$\limsup_{D \rightarrow \infty} \sup_{G \in \mathcal{G}_D} p_G(\beta, h) \leq \mathcal{P}(\beta, h).$$

The supremum is over all finite simple D -regular graphs, with any number of vertices; no assumption on the spectrum, the girth, or the symmetries of G is made. We now outline the proof. The strategy is that of Mourrat [106, 108], who proved upper bounds for mean-field models with nonconvex interactions by showing that an enriched free energy is an approximate viscosity supersolution of a Hamilton–Jacobi equation, with Ghirlanda–Guerra identities and synchronization enforced at contact points. Fix a finite cascade as in Section 3. On the cone of V -indexed families of nondecreasing paths we consider an enriched free energy in which every vertex receives a Gaussian field with the covariance structure of a Ruelle cascade, and in which each first-level branch of the cascade carries an independent copy of the couplings. At time t this free energy is at most $\beta^2 t/4 - p_G(\beta\sqrt{t}, h)$ at the origin, and at time zero it is given by the scalar transform of Section 3.3. Its time derivative differs from a quadratic Hamiltonian evaluated at its gradient by $\beta^2(4n)^{-1} \sum_j w_j \operatorname{tr}(SC_j)$, where C_j is a conditional covariance matrix of the products $\sigma_x^1 \sigma_x^2$. Since S has negative eigenvalues, this term has no sign. For every $\eta > 0$, we use $|\operatorname{tr}(SC_j)| \leq \eta \operatorname{tr} C_j + (4\eta)^{-1} \operatorname{tr}(S^2 C_j)$, with $\operatorname{tr} C_j \leq n$. We reduce $\operatorname{tr}(S^2 C_j)$ to the conditional variances of the overlaps of $O(n/D)$ rows of S by a sparsification theorem of Cohen, Nelson, and Woodruff, and control these variances by Mourrat’s finitary form of Panchenko’s synchronization theorem, using Ghirlanda–Guerra identities produced by small Gaussian perturbations. Because the couplings are independent across first-level branches, the free energy has fluctuations of order $1/n$ at every fixed parameter, which allows these identities to be enforced at every point where a test function touches a deterministic regularization of the free energy. A comparison principle on the cone then bounds this regularization from below by an explicit subsolution built from the Hopf–Lax formula of Section 3.6, and Lemma 3.10 compares the value of the subsolution with the Parisi formula.

Section 4.1 introduces the Hamiltonians on the cone and proves the comparison principle. Section 4.2 constructs the subsolution. Section 4.3 defines the enriched free energy and records its properties. Section 4.4 states Mourrat's bound on conditional variances, Section 4.5 introduces the Gaussian perturbations, and Section 4.6 bounds the drift term. Section 4.7 completes the proof of Proposition 4.1.

Throughout this section we fix $D \geq 1$, a graph $G = (V, E) \in \mathcal{G}_D$ with $n = |V|$ vertices, and the notation of (2.2). In particular $S = A_G/D$ is symmetric, nonnegative, with zero diagonal, $S\mathbf{1} = \mathbf{1}$, $\|S\| \leq 1$, and $\text{tr } S^2 = n/D$. We also fix a cascade as in (3.1) and (3.2). All constants are independent of D , G , and n unless stated otherwise. We write

$$Z_G(\beta, h) = \sum_{\sigma \in \{-1, 1\}^V} \exp \left(\beta H_G(\sigma) + h \sum_x \sigma_x \right), \quad p_G(\beta, h) = \frac{1}{n} \mathbb{E} \log Z_G(\beta, h).$$

4.1 Hamiltonians on the cone and comparison

Let $V_K = V \times \{0, \dots, K\}$. For $p \in \mathbb{R}^{V_K}$ we write $p_x = (p_{xj})_{j \leq K}$ and $p_j = (p_{xj})_{x \in V} \in \mathbb{R}^V$. We use the inner product and norms

$$\langle p, q \rangle_V = \frac{1}{n} \sum_{x \in V} \langle p_x, q_x \rangle_w, \quad \|p\|_1 = \frac{1}{n} \sum_{x \in V} \|p_x\|_{1,w}, \quad \|p\|_2 = \langle p, p \rangle_V^{1/2},$$

so that $\|p\|_1 \leq \|p\|_2$, because the weights w_j/n sum to one. The cone is $\mathcal{C}_K^V = \{q \in \mathbb{R}^{V_K} : q_x \in \mathcal{C}_K \text{ for every } x\}$, and its dual cone for $\langle \cdot, \cdot \rangle_V$ is $(\mathcal{C}_K^*)^V$. We write $p \succeq_* p'$ if $p_x \succeq_* p'_x$ for every x . For a function u of $q \in \mathbb{R}^{V_K}$ we write $D_q u$ for its gradient with respect to $\langle \cdot, \cdot \rangle_V$, that is, $(D_q u)_{xj} = n w_j^{-1} \partial u / \partial q_{xj}$; for functions on $\mathcal{C}_K^V$, differentiability is understood as in Section 3.1.

Hamiltonians

For $p \in \mathbb{R}^{V_K}$ we define

$$H_S(p) = \frac{\beta^2}{4n} \sum_{j=0}^K w_j p_j \cdot S p_j \tag{4.1}$$

and

$$H_{\text{ext}}(p) = \inf \{ H_S(A) : A \in \mathcal{C}_K^V, A \succeq_* p \}. \tag{4.2}$$

This is the monotone modification constructed in the proof of [36, Lemma 2.4], applied to H_S . For $z \in \mathbb{R}^{K+1}$ and $0 \leq \ell \leq K+1$ we write $\tau_\ell(z) = \sum_{j \geq \ell} w_j z_j$, $\bar{\tau}_\ell(z) = \max_{\ell \leq \ell' \leq K+1} \tau_{\ell'}(z)$, and $W_\ell = \tau_\ell(\mathbf{1})$. Then $\tau_{K+1}(z) = 0$, so that $\bar{\tau}_\ell(z) \geq 0$, and $z \succeq_* z'$ if and only if $\tau_\ell(z) \geq \tau_\ell(z')$ for every ℓ , by (3.3).

Lemma 4.2. *The following hold.*

- (a) If $A, B \in \mathcal{C}_K^V$ and $B \succeq_* A$, then $H_S(B) \geq H_S(A)$.
- (b) Let $p \in \mathbb{R}^{V_K}$. For each x , let $\mathcal{C}_x$ be the least concave majorant on $[0, 1]$ of the points $(W_\ell, \bar{\tau}_\ell(p_x))$, $0 \leq \ell \leq K+1$, and let $A_{xj} = (\mathcal{C}_x(W_j) - \mathcal{C}_x(W_{j+1}))/w_j$. Then $A \in \mathcal{C}_K^V$, $A \succeq_* p$, every $B \in \mathcal{C}_K^V$ with $B \succeq_* p$ satisfies $B \succeq_* A$, and $H_{\text{ext}}(p) = H_S(A)$. Further, $0 \leq w_K A_{xK} \leq \|p_x\|_{1,w}$ for every x . If $p \in (0, \infty)^{V_K}$, then $A \in (0, \infty)^{V_K}$, and every $j_0 \leq K$ belongs to a block $\{i, \dots, i' - 1\} \ni j_0$ on which A_x is constant, equal to the weighted mean of p_x over the block, and such that $\sum_{j=\ell}^{i'-1} w_j p_{xj} \leq A_{xj_0} \sum_{j=\ell}^{i'-1} w_j$ for every $i \leq \ell < i'$.
- (c) The function H_{ext} is finite, nondecreasing for $\succeq_*$, and equal to H_S on $\mathcal{C}_K^V$. For $R > 0$ and $p, p' \in \mathbb{R}^{V_K}$ with $\|p\|_2, \|p'\|_2 \leq R$, we have $|H_{\text{ext}}(p) - H_{\text{ext}}(p')| \leq \beta^2 R w_K^{-1} \|p - p'\|_2$.

Proof. (a) For $P \in \mathcal{C}_K^V$, we have $(DH_S(P))_{xj} = (\beta^2/2) \sum_y S_{xy} P_{yj}$, which is nonnegative and nondecreasing in j , because S has nonnegative entries and P_{yj} is nonnegative and nondecreasing in j . In other words, $DH_S(P) \in \mathcal{C}_K^V$. Since the segment from A to B lies in $\mathcal{C}_K^V$ and $B - A \in (\mathcal{C}_K^*)^V$, we have $H_S(B) - H_S(A) = \int_0^1 \langle DH_S(A + \theta(B - A)), B - A \rangle_V d\theta \geq 0$.

(b) Fix x and let $a_\ell = \bar{\tau}_\ell(p_x)$. As ℓ increases from 0 to $K+1$, the abscissae W_ℓ decrease strictly from 1 to 0, and a_ℓ is nonincreasing, with $a_{K+1} = 0$. The least concave majorant $\mathcal{C}_x$ is the upper boundary of the convex hull of the points. It is piecewise linear, its vertices are among the points, $\mathcal{C}_x(0) = a_{K+1} = 0$, and $\mathcal{C}_x(1) = a_0$. Since the maximum of $\mathcal{C}_x$ is attained at a vertex and $a_0 = \max_\ell a_\ell$, it is attained at 1, and the concave function $\mathcal{C}_x$ is nondecreasing. Its slopes are then nonnegative and nonincreasing in W , which gives $A_x \in \mathcal{C}_K$. Since $\mathcal{C}_x(0) = 0$, we have $\tau_\ell(A_x) = \mathcal{C}_x(W_\ell) \geq a_\ell \geq \tau_\ell(p_x)$, that is, $A_x \succeq_* p_x$. Moreover $w_K A_{xK} = \mathcal{C}_x(W_K) \in [0, a_0]$, and $a_0 \leq \max_\ell |\tau_\ell(p_x)| \leq \|p_x\|_{1,w}$. Let $B_x \in \mathcal{C}_K$ satisfy $B_x \succeq_* p_x$. The piecewise linear interpolation Φ of the points $(W_\ell, \tau_\ell(B_x))$ is concave and nondecreasing, because its slopes B_{xj} are nonnegative and nondecreasing in j . For $\ell \leq \ell'$ we have $\Phi(W_\ell) \geq \Phi(W_{\ell'}) = \tau_{\ell'}(B_x) \geq \tau_{\ell'}(p_x)$, so that $\Phi(W_\ell) \geq a_\ell$. By the minimality of the least concave majorant, $\Phi \geq \mathcal{C}_x$, and $B_x \succeq_* A_x$. By (a), $H_S(B) \geq H_S(A)$ for every B admissible in (4.2), and A is admissible, which shows that $H_{\text{ext}}(p) = H_S(A)$.

Let now $p \in (0, \infty)^{V_K}$. Then $\tau_\ell(p_x)$ is strictly decreasing in ℓ , and $a_\ell = \tau_\ell(p_x)$. Let $i \leq j_0 < i'$ be such that $W_{i'}$ and W_i are consecutive abscissae of vertices of $\mathcal{C}_x$. On $[W_{i'}, W_i]$ the function $\mathcal{C}_x$ is linear, with slope $A_{xj} = A_{xj_0}$ for $i \leq j < i'$. Since $\mathcal{C}_x(W_i) = \tau_i(p_x)$ and $\mathcal{C}_x(W_{i'}) = \tau_{i'}(p_x)$, this slope equals the weighted mean of p_x over $\{i, \dots, i' - 1\}$, which is positive. For $i \leq \ell < i'$ we have $\tau_\ell(p_x) \leq \mathcal{C}_x(W_\ell) = \tau_{i'}(p_x) + A_{xj_0}(W_\ell - W_{i'})$, which is the last assertion.

(c) By (b), H_{ext} is finite. If $p' \succeq_* p$, then every A admissible for p' is admissible for p , and $H_{\text{ext}}(p) \leq H_{\text{ext}}(p')$. If $p \in \mathcal{C}_K^V$, then p is admissible and every admissible A satisfies $A \succeq_* p$. By (a), we have $H_{\text{ext}}(p) = H_S(p)$.

For the Lipschitz bound, let $\|p\|_2, \|p'\|_2 \leq R$, let A be as in (b) for p , and set $\lambda_x = \bar{\tau}_0(p'_x - p_x) \leq \|p'_x - p_x\|_{1,w}$ and $c_x = \lambda_x w_K^{-1} \mathbf{1}_{\{K\}} \in \mathcal{C}_K$. Then $\tau_\ell(c_x) = \lambda_x \geq \tau_\ell(p'_x - p_x)$

for $\ell \leq K$, and $A + c \in \mathcal{C}_K^V$ is admissible for p' . Since $c_j = 0$ for $j < K$ and $\|S\| \leq 1$, we have

$$H_{\text{ext}}(p') - H_{\text{ext}}(p) \leq H_S(A + c) - H_S(A) \leq \frac{\beta^2 w_K}{4n} (2|c_K| \|A_K\| + |c_K|^2),$$

where $|\cdot|$ is the Euclidean norm on $\mathbb{R}^V$. By (b) and the inequality $\|z\|_{1,w} \leq \|z\|_{2,w}$, we have $w_K^2 |A_K|^2 \leq \sum_x \|p_x\|_{1,w}^2 \leq n \|p\|_2^2$, and similarly $w_K^2 |c_K|^2 \leq n \|p' - p\|_2^2$. Since $\|p\|_2 \leq R$ and $\|p' - p\|_2 \leq 2R$, the right side of the last display is at most $\beta^2 R w_K^{-1} \|p' - p\|_2$. Exchanging p and p' , we obtain the bound. $\square$

If p_x has positive entries, then A_x is the isotonic regression of p_x with weights (w_j) , that is, the nondecreasing vector closest to p_x for $\|\cdot\|_{2,w}$, and Lemma 4.2(b) is equivalent to its classical description by the greatest convex minorant of cumulative sums [122].

Lemma 4.3. *Let $P \in \mathcal{C}_K$ and $r \in (0, \infty)^{V_K}$. Set $M_j = n^{-1} \sum_x r_{xj}$ and $b_{xj} = P_j M_j / r_{xj}$. Then*

$$H_{\text{ext}}(b) \geq \frac{\beta^2}{4} \sum_{j=0}^K w_j P_j^2.$$

Proof. Suppose first that $P_j > 0$ for every j , so that $b \in (0, \infty)^{V_K}$, and let A be as in Lemma 4.2(b) for $p = b$. Fix x and a block $\{i, \dots, i' - 1\}$ as there, on which A_x equals a constant a . For $i \leq \ell < i'$, the weighted mean of b_{xj} over $\{\ell, \dots, i' - 1\}$ is at most a . By Jensen's inequality for the convex function $1/y$, we have $\sum_{j=\ell}^{i'-1} w_j (a^{-1} - b_{xj}^{-1}) \leq 0$. Since a terminal interval $\{\ell, \dots, K\}$ is the union of such a terminal part of one block and of the later blocks, we obtain $\tau_\ell(A_x^{-1} - b_x^{-1}) \leq 0$ for every ℓ , where inverses are taken coordinatewise. Let $v_j = n^{-1} \sum_x A_{xj}^{-1}$. Since $n^{-1} \sum_x b_{xj}^{-1} = P_j^{-1}$, averaging over x gives $\tau_\ell(v - P^{-1}) \leq 0$ for all ℓ , that is, $P^{-1} - v \in \mathcal{C}_K^*$.

The weights S_{xy}/n form a probability measure on $V \times V$ with both marginals uniform. For $a, a' > 0$, the inequality between the arithmetic and geometric means of a^{-1} and a'^{-1} gives $aa' \geq ((a^{-1} + a'^{-1})/2)^{-2}$. Using this and Jensen's inequality for the convex function $a \mapsto a^{-2}$ on $(0, \infty)$, we obtain

$$\frac{1}{n} \sum_{x,y} S_{xy} A_{xj} A_{yj} \geq \frac{1}{n} \sum_{x,y} S_{xy} \left(\frac{A_{xj}^{-1} + A_{yj}^{-1}}{2} \right)^{-2} \geq v_j^{-2}.$$

The tangent line of the convex function $a \mapsto a^{-2}$ at P_j^{-1} gives $v_j^{-2} \geq P_j^2 - 2P_j^3(v_j - P_j^{-1})$. Since the vector $(P_j^3)_j$ belongs to $\mathcal{C}_K$ and $P^{-1} - v \in \mathcal{C}_K^*$, we have

$$\sum_j w_j v_j^{-2} \geq \sum_j w_j P_j^2 + 2 \langle (P_j^3)_j, P^{-1} - v \rangle_w \geq \sum_j w_j P_j^2.$$

Combining the last two displays with the identity $H_{\text{ext}}(b) = H_S(A)$ from Lemma 4.2(b), we obtain the claim in this case. In general, for $\theta > 0$ let $P^{(\theta)} = P + \theta \mathbf{1}$, which belongs to

$\mathcal{C}_K \cap (0, \infty)^{K+1}$, and let $b^{(\theta)}$ be the corresponding vector. Then $b^{(\theta)} \rightarrow b$ as $\theta \rightarrow 0$, and the continuity of H_{ext} from Lemma 4.2(c) proves the claim. $\square$

Viscosity solutions and comparison

Viscosity solutions were introduced by Crandall and Lions [49]; see also [50]. The following definition, in which test points on the boundary of the cone are allowed, follows [36, Definition 1.1].

Definition 4.4. Let $H: \mathbb{R}^{V_K} \rightarrow \mathbb{R}$ be continuous, let $c \in \mathbb{R}$, and let $u: [0, 2] \times \mathcal{C}_K^V \rightarrow \mathbb{R}$ be continuous. We say that u is a *subsolution* of $\partial_t u - H(D_q u) \leq c$ if, for every $(t_0, q_0) \in (0, 2) \times \mathcal{C}_K^V$ and every function φ that is continuously differentiable in a neighborhood of (t_0, q_0) in $\mathbb{R} \times \mathbb{R}^{V_K}$ and such that $u - \varphi$ has a local maximum at (t_0, q_0) relative to $(0, 2) \times \mathcal{C}_K^V$, we have $\partial_t \varphi(t_0, q_0) - H(D_q \varphi(t_0, q_0)) \leq c$. We say that u is a *supersolution* of $\partial_t u - H(D_q u) \geq c$ if the same holds with local minima and the reverse inequality.

The tests are relative to the closed cone: points on its boundary are included, and no boundary condition is imposed. We use the following elementary fact repeatedly. Let u be differentiable in t at $t_0 \in (0, 2)$ and differentiable on $\mathcal{C}_K^V$ in q , and let $u - \varphi$ have a local minimum at (t_0, q_0) relative to $(0, 2) \times \mathcal{C}_K^V$. Since the point $q_0 + \epsilon r$ lies in $\mathcal{C}_K^V$ for $r \in \mathcal{C}_K^V$ and $\epsilon \geq 0$, we have

$$\partial_t u(t_0, q_0) = \partial_t \varphi(t_0, q_0), \quad D_q u(t_0, q_0) - D_q \varphi(t_0, q_0) \in (\mathcal{C}_K^*)^V. \quad (4.3)$$

Lemma 4.5. Let $H: \mathbb{R}^{V_K} \rightarrow \mathbb{R}$ be such that for every $R > 0$ there exists L_R with $|H(p) - H(p')| \leq L_R \|p - p'\|_2$ whenever $\|p\|_2, \|p'\|_2 \leq R$. Let $\tau \geq 0$, $c_0 \in \mathbb{R}$, and let $V, E: [0, 2] \times \mathcal{C}_K^V \rightarrow \mathbb{R}$ be continuous functions with the following properties, for constants L_t, L_q :

- (i) V is a subsolution of $\partial_t u - H(D_q u) \leq 0$, and E is a supersolution of $\partial_t u - H(D_q u) \geq -\tau$;
- (ii) $|E(t, q) - E(t', q')| \leq L_t |t - t'| + L_q \|q - q'\|_1$ and $|V(t, q) - V(t', q)| \leq L_t |t - t'|$;
- (iii) $V(0, q) \leq E(0, q) + c_0$ for every $q \in \mathcal{C}_K^V$.

Then $V(t, q) \leq E(t, q) + c_0 + \tau t$ for every $t \in [0, 2]$ and $q \in \mathcal{C}_K^V$.

The proof is the doubling-of-variables argument of [49, 50]; see [56, Theorem 3.5] and [36, Proposition 3.1] for comparison principles on $\mathbb{R}^d$ and on cones. We include the proof, because the Hamiltonian is only locally Lipschitz and V is not assumed to be Lipschitz in q .

Proof. Suppose that $V(\bar{t}, \bar{q}) - E(\bar{t}, \bar{q}) - c_0 - \tau \bar{t} = 4\theta > 0$ for some $\bar{t} < 2$; by (iii), $\bar{t} > 0$. Let $L = \max\{1, (L_t^2 + L_q^2)^{1/2}\}$. Fix $T' \in (\bar{t}, 2)$, then $\rho > 0$ with $\rho \bar{t} + \rho/(T' - \bar{t}) \leq \theta$, and then $\gamma \in (0, 1]$ with $\gamma(1 + \|\bar{q}\|_2^2)^{1/2} \leq \theta$ and $\gamma L_{2L+1} < \rho$. For $b > 0$ and $(t, q, s, r) \in [0, T'] \times \mathcal{C}_K^V \times [0, 2] \times \mathcal{C}_K^V$ let

$$\Phi_b(t, q, s, r) = V(t, q) - E(s, r) - c_0 - (\tau + \rho)t - \frac{\rho}{T' - t} - \gamma \langle q \rangle - \frac{|t - s|^2 + \|q - r\|_2^2}{2b},$$

where $\langle q \rangle = (1 + \|q\|_2^2)^{1/2}$. By (ii) and (iii), we have $V(t, q) - E(s, r) \leq c_0 + 2L_t t + L_t |t - s| + L_q \|q - r\|_2$, and Φ_b is continuous and tends to $-\infty$ as $\|q\|_2 \rightarrow \infty$, as $\|q - r\|_2 \rightarrow \infty$, or as $t \rightarrow T'$. Since the dimension is finite, Φ_b attains its maximum at some point (t_b, q_b, s_b, r_b) , and $\Phi_b(t_b, q_b, s_b, r_b) \geq \Phi_b(\bar{t}, \bar{q}, \bar{t}, \bar{q}) \geq 2\theta$.

Comparing with $\Phi_b(t_b, q_b, t_b, q_b)$ and using (ii), we obtain, with $X = (|t_b - s_b|^2 + \|q_b - r_b\|_2^2)^{1/2}$,

$$\frac{X^2}{2b} \leq E(t_b, q_b) - E(s_b, r_b) \leq LX, \quad \text{so} \quad X \leq 2bL.$$

If $t_b = 0$, then by (iii) and (ii), $\Phi_b \leq E(0, q_b) - E(s_b, r_b) \leq LX \leq 2bL^2$. If $s_b = 0$, then $\Phi_b \leq V(t_b, q_b) - V(0, q_b) + E(0, q_b) - E(0, r_b) \leq 2LX \leq 4bL^2$. Both contradict $\Phi_b \geq 2\theta$ if $b < \theta/(2L^2)$. Also $s_b \leq t_b + 2bL < 2$ for small b . We conclude that $t_b \in (0, T')$ and $s_b \in (0, 2)$ for small b .

The function $(t, q) \mapsto V(t, q) - \varphi_1(t, q)$, with

$$\varphi_1(t, q) = (\tau + \rho)t + \frac{\rho}{T' - t} + \gamma \langle q \rangle + \frac{|t - s_b|^2 + \|q - r_b\|_2^2}{2b},$$

has a local maximum at (t_b, q_b) relative to $(0, 2) \times \mathcal{C}_K^V$, and the function $(s, r) \mapsto E(s, r) - \varphi_2(s, r)$, with $\varphi_2(s, r) = -(|t_b - s|^2 + \|q_b - r\|_2^2)/(2b)$, has a local minimum at (s_b, r_b) . Let $p = (q_b - r_b)/b$. Since $D_q \langle q \rangle = q/\langle q \rangle$, by (i) we have

$$\tau + \rho + \frac{\rho}{(T' - t_b)^2} + \frac{t_b - s_b}{b} - H\left(p + \frac{\gamma q_b}{\langle q_b \rangle}\right) \leq 0 \leq \frac{t_b - s_b}{b} - H(p) + \tau.$$

We have $\|p\|_2 \leq X/b \leq 2L$ and $\|\gamma q_b/\langle q_b \rangle\|_2 \leq \gamma \leq 1$. Rearranging, we obtain $\rho \leq H(p + \gamma q_b/\langle q_b \rangle) - H(p) \leq L_{2L+1}\gamma < \rho$, a contradiction. $\square$

4.2 The subsolution

In this subsection $\beta > 0$ and $h \in \mathbb{R}$ are fixed, ψ_h is the scalar transform (3.12), and U is the function (3.23). We evaluate U at an average of the paths q_x . A similar construction appears in the Hamilton–Jacobi comparison argument for balanced multispecies models of Chen, Issa, and Mourrat [37, Appendix B], where the weighted mean of the paths is used and the initial conditions are compared through the convexity of ψ_0 . They note that this argument is close in spirit to the Hamilton–Jacobi comparison for permutation-invariant vector spin glasses in [86, Section 7]. Here we average the square roots of the paths, which Proposition 3.8 allows for every h , and the subsolution property follows from Lemma 4.3. For $\delta \in (0, 1)$ and $q \in \mathcal{C}_K^V$ we set

$$r_{xj}(q) = (q_{xj} + \delta)^{1/2}, \quad M_j(q) = \frac{1}{n} \sum_x r_{xj}(q), \quad g_j(q) = M_j(q)^2 - \delta. \quad (4.4)$$

Since r_{xj} is nondecreasing in j and $M_j \geq \sqrt{\delta}$, we have $g(q) \in \mathcal{C}_K$. For $q \in \mathcal{C}_K^V$ we write $\tilde{q}$ for the element of $\mathcal{C}_K^V$ with $\tilde{q}_{x0} = 0$ and $\tilde{q}_{xj} = q_{xj}$ for $1 \leq j \leq K$. We define

$$V_\delta(t, q) = U(t, g(\tilde{q})) - \frac{\beta^2 \zeta_0 t}{4}. \quad (4.5)$$

Lemma 4.6. *The function V_δ is continuous on $[0, 2] \times \mathcal{C}_K^V$, and the following hold.*

- (a) $|V_\delta(t, q) - V_\delta(t', q)| \leq \beta^2 |t - t'|$.
- (b) V_δ is a subsolution of $\partial_t u - H_{\text{ext}}(D_q u) \leq 0$.
- (c) $V_\delta(0, q) \leq n^{-1} \sum_x \psi_h(\tilde{q}_x) + \delta$ for every $q \in \mathcal{C}_K^V$.
- (d) $V_\delta(1, 0) = U(1, 0) - \beta^2 \zeta_0 / 4$.

Proof. Continuity and (a) follow from Lemma 3.9(b),(c), and (d) holds because $g(0) = 0$.

(b) Let φ be continuously differentiable near $(t_0, q_0) \in (0, 2) \times \mathcal{C}_K^V$ and let $V_\delta - \varphi$ have a local maximum at (t_0, q_0) ; we may assume that its value there is zero. Let y' be a maximizer in (3.23) for $U(t_0, g(\tilde{q}_0))$, and let $P = 2(y' - g(\tilde{q}_0)) / (\beta^2 t_0) \in \mathcal{C}_K \cap [0, 1]^{K+1}$ be as in Lemma 3.9(a). The function

$$B(t, q) = \psi_h(y') - \frac{\|y' - g(\tilde{q})\|_{2,w}^2}{\beta^2 t} - \frac{\beta^2 \zeta_0 t}{4}$$

is smooth near (t_0, q_0) , satisfies $B \leq V_\delta$ on $(0, 2) \times \mathcal{C}_K^V$ and $B(t_0, q_0) = V_\delta(t_0, q_0)$. Then $B - \varphi \leq V_\delta - \varphi \leq 0$ near (t_0, q_0) , with equality at that point. Applying (4.3) to $\varphi - B$, we obtain $\partial_t \varphi = \partial_t B$ and $D_q \varphi \succeq_* D_q B$ at (t_0, q_0) . Since $\tilde{q}$ does not depend on the coordinates q_{x0} , a direct computation at (t_0, q_0) gives $(D_q B)_{x0} = 0$ and

$$\begin{aligned} \partial_t B &= \frac{\|y' - g(\tilde{q}_0)\|_{2,w}^2}{\beta^2 t_0^2} - \frac{\beta^2 \zeta_0}{4} = \frac{\beta^2}{4} (\|P\|_{2,w}^2 - \zeta_0), \\ (D_q B)_{xj} &= \frac{P_j M_j(\tilde{q}_0)}{r_{xj}(\tilde{q}_0)} \quad (1 \leq j \leq K). \end{aligned}$$

Let $P' = (0, P_1, \dots, P_K) \in \mathcal{C}_K$. Then $D_q B$ is the vector b of Lemma 4.3 for P' and $r = r(\tilde{q}_0)$, and that lemma gives $H_{\text{ext}}(D_q B) \geq (\beta^2/4) \sum_{j \geq 1} w_j P_j^2$. Since $w_0 = \zeta_0$ and $P_0 \leq 1$, we have $\partial_t B \leq (\beta^2/4) \sum_{j \geq 1} w_j P_j^2$. By Lemma 4.2(c), we conclude that $\partial_t \varphi = \partial_t B \leq H_{\text{ext}}(D_q B) \leq H_{\text{ext}}(D_q \varphi)$.

(c) We have $V_\delta(0, q) = \psi_h(g(\tilde{q}))$. The vectors $g(\tilde{q})$ and $M(\tilde{q})^2 = (M_j(\tilde{q})^2)_j$ belong to $\mathcal{C}_K$, with $g(\tilde{q}) \leq M(\tilde{q})^2$ coordinatewise, and Lemma 3.2 implies that $\psi_h(g(\tilde{q})) \leq \psi_h(M(\tilde{q})^2)$. Since the vectors $r_x(\tilde{q})$ belong to the convex set $\{s : 0 \leq s_0 \leq \dots \leq s_K\}$ and $M(\tilde{q})$ is their average, we have $\psi_h(M(\tilde{q})^2) \leq n^{-1} \sum_x \psi_h(\tilde{q}_x + \delta \mathbf{1})$ by Proposition 3.8 and Jensen's inequality. Finally $\psi_h(\tilde{q}_x + \delta \mathbf{1}) \leq \psi_h(\tilde{q}_x) + \delta$ by Lemma 3.2, since $\|\delta \mathbf{1}\|_{1,w} = \delta$. $\square$

4.3 The enriched free energy

Ruelle cascades

Let $\mathbb{A} = \bigcup_{k=0}^K \mathbb{N}^k$ be the rooted tree of depth K with infinite degree, with root $\emptyset$ and leaves $\mathbb{N}^K$. For $\alpha \in \mathbb{N}^K$ and $k \leq K$ let $\alpha|_k$ be the ancestor of α at depth k , and for leaves α, α' let $\alpha \wedge \alpha' = \max\{k : \alpha|_k = \alpha'|_k\}$. For each nonleaf vertex $\gamma \in \mathbb{N}^k$, $k < K$, let $(u_{\gamma i})_{i \geq 1}$ be the decreasing enumeration of a Poisson point process on $(0, \infty)$ with intensity $\zeta_k x^{-1-\zeta_k} dx$, independently over γ . Set $w_\alpha = \prod_{k=1}^K u_{\alpha|_k}$ for $\alpha \in \mathbb{N}^K$; the sum $\sum_\alpha w_\alpha$ is finite almost surely [56, Lemma 5.23], and the *Ruelle cascade* [123] with masses $(\zeta_k)_{k < K}$ is $v_\alpha = w_\alpha / \sum_{\alpha'} w_{\alpha'}$. We use the following facts, in which the masses are indexed as in (3.1); in [56] the mass ζ_k is denoted ζ_{k+1} .

Theorem 4.7. *Let $(\omega_\gamma)_{\gamma \in \mathbb{A} \setminus \{\emptyset\}}$ be independent random elements, independent of $(u_{\gamma i})$, such that ω_γ has law P_k if $\gamma \in \mathbb{N}^k$. Let $X: \Omega_1 \times \cdots \times \Omega_K \rightarrow \mathbb{R}$ be measurable with $\mathbb{E} \exp(c|X|) < \infty$ for every $c > 0$, set $X_K = X$ and $X_{k-1} = \zeta_{k-1}^{-1} \log \mathbb{E}_{\omega_k} \exp(\zeta_{k-1} X_k)$ for $1 \leq k \leq K$, where $\mathbb{E}_{\omega_k}$ integrates a variable $\omega_k \sim P_k$. Then*

$$\mathbb{E} \log \sum_{\alpha \in \mathbb{N}^K} v_\alpha \exp X(\omega_{\alpha|_1}, \dots, \omega_{\alpha|_K}) = X_0.$$

Moreover, if $\langle \cdot \rangle$ denotes the random probability measure on $\mathbb{N}^K$ with weights proportional to $v_\alpha \exp X(\omega_{\alpha|_1}, \dots, \omega_{\alpha|_K})$, and α^1, α^2 are independent samples from it, then

$$\mathbb{E} \langle \mathbf{1}_{\{\alpha^1 \wedge \alpha^2 = j\}} \rangle = w_j \quad (0 \leq j \leq K). \quad (4.6)$$

The first statement is [56, Theorem 5.25], applied with a trivial root variable; see also [114, Theorem 2.9]. For $X = 0$, (4.6) is [56, (5.93) in Theorem 5.28]. For general X , it is proved in [56, Section 6.4, before Lemma 6.2] for the enriched free energy considered there, by observing that the proof of [56, Theorem 5.28], which rests on the invariance property [56, Lemma 5.27] of Bolthausen and Sznitman [27], remains valid when each w_α is multiplied by a function of variables attached to the vertices of the path of α that are independent of (w_α) and whose joint law is invariant under bijections of $\mathbb{N}^K$ preserving $\wedge$. This applies verbatim to the weights $w_\alpha \exp X(\omega_{\alpha|_1}, \dots, \omega_{\alpha|_K})$.

We also use Gaussian integration by parts for Gibbs measures [56, Theorem 4.6]. Let Σ be a countable set, μ a finite measure on Σ , and $(y(\zeta))_{\zeta \in \Sigma}, (z(\zeta))_{\zeta \in \Sigma}$ centered jointly Gaussian processes with bounded covariances. Let $\langle \cdot \rangle$ be the Gibbs measure with density proportional to $e^{\mathcal{Y}}$ with respect to μ , and $C(\zeta, \zeta') = \mathbb{E} z(\zeta) y(\zeta')$. Then for all $r \geq 1$ and all bounded $f: \Sigma^r \rightarrow \mathbb{R}$,

$$\mathbb{E} \langle f(\zeta^1, \dots, \zeta^r) z(\zeta^1) \rangle = \mathbb{E} \left\langle f(\zeta^1, \dots, \zeta^r) \left(\sum_{\ell=1}^r C(\zeta^1, \zeta^\ell) - r C(\zeta^1, \zeta^{r+1}) \right) \right\rangle. \quad (4.7)$$

Finally, we use the following invariance of Poisson point processes [56, Theorem 5.19 and Proposition 5.18]. Let (u_i) be the decreasing enumeration of a Poisson point process with intensity $\zeta x^{-1-\zeta} dx$, $\zeta \in (0, 1)$. Then $\mathbb{E}(\sum_i u_i)^a < \infty$ for $0 < a < \zeta$, and if (Y_i) are independent and identically distributed positive random variables, independent of (u_i) , with $\mathbb{E}Y_1^\zeta < \infty$, then

$$\sum_i u_i Y_i \quad \text{has the law of} \quad (\mathbb{E}Y_1^\zeta)^{1/\zeta} \sum_i u_i. \quad (4.8)$$

The model

Fix reference values $z_j = j/K$, $0 \leq j \leq K$. For leaves α, α' we write $\vartheta(\alpha, \alpha') = z_{\alpha \wedge \alpha'}$; this is the inner product of $\sum_{k=1}^K (z_k - z_{k-1})^{1/2} e_{\alpha|_k}$ and the analogous vector for α' , for an orthonormal family (e_γ) , which shows that the kernel ϑ is positive semidefinite. We attach the following independent variables to the vertices of $\mathbb{A}$, independently of the cascade:

- to each $a \in \mathbb{N}$ (depth one), an independent copy $(g_e^{(a)})_{e \in E}$ of the couplings, defining $H^{(a)}(\sigma) = \sum_{xy \in E} D^{-1/2} g_{xy}^{(a)} \sigma_x \sigma_y$;
- to each vertex $\gamma \neq \emptyset$ and each $x \in V$, a standard Gaussian variable $\xi_{x,\gamma}$;
- to each vertex $\gamma \neq \emptyset$, the Gaussian variables used to construct the processes $G_{x,\iota}$ of Lemma 4.8 below.

There is no variable attached to the root. Let $\mathcal{S} \subseteq V$ and a finite set I index the processes $G_{x,\iota}$, $x \in \mathcal{S}$, $\iota \in I$, and let $s \geq 0$. For $t \geq 0$, $q \in \mathcal{C}_K^V$, and $y \in \mathcal{Y} = [1/2, 3/2]^{\mathcal{S} \times I}$ we define, for $\alpha \in \mathbb{N}^K$ and $\sigma \in \{-1, 1\}^V$,

$$\mathcal{H}(\alpha, \sigma) = \beta \sqrt{t} H^{(\alpha_1)}(\sigma) + h \sum_x \sigma_x + \sum_x (Y_x(\alpha) \sigma_x - q_{xK}) + s \sum_{x \in \mathcal{S}, \iota \in I} y_{x\iota} G_{x,\iota}(\alpha, \sigma), \quad (4.9)$$

where

$$Y_x(\alpha) = (2q_{x1})^{1/2} \xi_{x,\alpha|_1} + \sum_{k=2}^K (2(q_{xk} - q_{x,k-1}))^{1/2} \xi_{x,\alpha|_k}.$$

We have $\mathbb{E}Y_x(\alpha)Y_x(\alpha') = 2q_{x,\alpha \wedge \alpha'}$ if $\alpha \wedge \alpha' \geq 1$, and 0 otherwise; the coordinate q_{x0} does not enter. Let $Z = \sum_\alpha v_\alpha \sum_\sigma \exp \mathcal{H}(\alpha, \sigma)$ and

$$\tilde{F}(t, q, y) = -\frac{1}{n} \log Z + \frac{\beta^2 t}{4}, \quad F(t, q, y) = \mathbb{E} \tilde{F}(t, q, y), \quad F_0(t, q) = F(t, q, y)|_{s=0}. \quad (4.10)$$

We write $\langle \cdot \rangle$ for the Gibbs measure on $\mathbb{N}^K \times \{-1, 1\}^V$ with weights proportional to $v_\alpha \exp \mathcal{H}(\alpha, \sigma)$, and $(\alpha^\ell, \sigma^\ell)_{\ell \geq 1}$ for independent samples from it. For replicas ℓ, ℓ' we write

$$\begin{aligned} \vartheta_{\ell\ell'} &= z_{\alpha^\ell \wedge \alpha^{\ell'}}, & u_{\ell\ell'}^* &= \mathbf{1}\{\alpha_1^\ell = \alpha_1^{\ell'}\} \sigma^\ell \odot \sigma^{\ell'}, \\ \varrho_{x, \ell\ell'} &= (Su_{\ell\ell'}^*)_x = \mathbf{1}\{\alpha_1^\ell = \alpha_1^{\ell'}\} \sum_y S_{xy} \sigma_y^\ell \sigma_y^{\ell'}, \end{aligned}$$

and $u^* = u_{12}^*$. The vector u^* is the product $u = \sigma^1 \odot \sigma^2$ when the two replicas lie in the same first-level branch, and it vanishes otherwise. For an event or function f of finitely many replicas we write $\mathbb{E}\langle f \mid \vartheta_{12} = z_j \rangle = w_j^{-1} \mathbb{E}\langle f \mathbf{1}\{\vartheta_{12} = z_j\} \rangle$; by Theorem 4.7, applied with $X = \log \sum_\sigma \exp \mathcal{H}(\alpha, \sigma)$, this is a conditional expectation under the probability measure $\mathbb{E}\langle \cdot \rangle$.

Lemma 4.8. *Let $x \in V$, $p \in \mathbb{N}$ with $p \geq 1$, and $c \in [0, 1]$, and let $k(\vartheta, \varrho) = (c\vartheta + (1-c)\varrho)^p$. There exists a centered Gaussian process $(G(\alpha, \sigma))$, which is a linear function of independent Gaussian variables attached to the vertices $\gamma \neq \emptyset$ of $\mathbb{A}$, such that*

$$\mathbb{E} G(\alpha, \sigma) G(\alpha', \sigma') = k\left(z_{\alpha \wedge \alpha'}, \mathbf{1}\{\alpha_1 = \alpha'_1\} \sum_y S_{xy} \sigma_y \sigma'_y\right).$$

The covariance is at most one in absolute value, equals one on the diagonal, and vanishes if $\alpha_1 \neq \alpha'_1$.

Proof. Expanding the power, we have $k = \sum_{l=0}^p \binom{p}{l} c^{p-l} (1-c)^l \vartheta^{p-l} \varrho^l$. For $1 \leq l \leq p$, the term $\vartheta^{p-l} \varrho^l$ evaluated at the pair is $a_l(\alpha \wedge \alpha') \langle F(\sigma)^{\otimes l}, F(\sigma')^{\otimes l} \rangle$, where $F(\sigma) = (S_{xy}^{1/2} \sigma_y)_{y \in V}$ and $a_l(k) = z_k^{p-l} \mathbf{1}\{k \geq 1\}$; for $l = 0$ it is $a_0(\alpha \wedge \alpha')$ with $a_0(k) = z_k^p$. Each a_l is nondecreasing on $\{0, \dots, K\}$ and vanishes at 0. If a is such a function and F' is a map from $\{-1, 1\}^V$ into a finite-dimensional Euclidean space, then $\sum_{k=1}^K (a(k) - a(k-1))^{1/2} \langle \xi_{\alpha|_k}, F'(\sigma) \rangle$, with independent standard Gaussian vectors ξ_γ , has covariance $a(\alpha \wedge \alpha') \langle F'(\sigma), F'(\sigma') \rangle$. Summing independent such processes gives G . Since $|\vartheta|, |\varrho| \leq 1$, the covariance is at most one in absolute value; on the diagonal $\vartheta = 1$ and $\varrho = \sum_y S_{xy} = 1$; and if $\alpha_1 \neq \alpha'_1$ then $\vartheta = \varrho = 0$. $\square$

For $(x, \iota) \in \mathcal{S} \times I$ with $\iota = (c_\iota, p_\iota)$, we let $G_{x, \iota}$ be the process of Lemma 4.8 for x and $k_\iota(\vartheta, \varrho) = (c_\iota \vartheta + (1-c_\iota) \varrho)^{p_\iota}$, built from variables independent over (x, ι) . We write $k_{x, \ell\ell'} = k_\iota(\vartheta_{\ell\ell'}, \varrho_{x, \ell\ell'})$. Up to normalization, these kernels are those of the perturbations in [106, Section 4], with the row overlap ϱ_x in place of the overlap of one type of spins.

Lemma 4.9. *The following hold for $t \geq 0$, $q \in C_K^V$, and $y \in Y$.*

- (a) *F is continuous, differentiable in $t > 0$, differentiable on C_K^V in q , and infinitely differentiable in y , with continuous derivatives. For every bounded function f of finitely many replicas, $\mathbb{E}\langle f \rangle$ is a continuous function of (t, q, y) .*

- (b) For $t > 0$, $\partial_t F = \beta^2 (4n)^{-1} \mathbb{E} \langle u^* \cdot Su^* \rangle$; in particular $|\partial_t F| \leq \beta^2/4$.
- (c) $D_q F = p$, where $p_{x0} = 0$ and $p_{xj} = \mathbb{E} \langle u_x^* | \vartheta_{12} = z_j \rangle$ for $1 \leq j \leq K$. Moreover $p \in \mathcal{C}_K^V \cap [0, 1]^{V_K}$, and consequently $|F(t, q, y) - F(t, q', y)| \leq \|q - q'\|_1$.
- (d) F is concave in y , and $\partial_{y_{x_l}} F = -n^{-1} s \mathbb{E} \langle G_{x,l} \rangle = -n^{-1} s^2 y_{x_l} (1 - \mathbb{E} \langle k_{x_l, 12} \rangle)$. Moreover $|F(t, q, y) - F_0(t, q)| \leq r_s$, where $r_s = 9|S||I|s^2/(4n)$.
- (e) We have

$$F_0(t, 0) = \frac{\beta^2 t}{4} - \frac{1}{n\zeta_0} \log \mathbb{E} Z_G(\beta\sqrt{t}, h)^{\zeta_0} \leq \frac{\beta^2 t}{4} - p_G(\beta\sqrt{t}, h).$$

- (f) We have $F_0(0, q) = n^{-1} \sum_x \psi_h(\tilde{q}_x)$, with $\tilde{q}$ as in Section 4.2.

Proof. (a) We write every Gaussian variable in (4.9) as an amplitude times a fixed standard Gaussian variable; the amplitudes are $\beta\sqrt{t}$, the numbers $(2q_{x1})^{1/2}$ and $(2(q_{xk} - q_{x,k-1}))^{1/2}$, and sy_{x_l} times fixed coefficients. By Theorem 4.7 with $X = \log \sum_{\sigma} \exp \mathcal{H}(\alpha, \sigma)$, we have $F = -X_0/n + \beta^2 t/4$, where X_0 is obtained from X by K successive integrations as in (3.7). By (3.5) and differentiation under the integral sign, X_0 is a smooth function of $\beta\sqrt{t}$, of y , and of the variances $2q_{x1}$, $2(q_{xk} - q_{x,k-1})$ in $[0, \infty)$. This gives the differentiability of F and the continuity of its derivatives. For the Gibbs averages, fix a compact set K of parameters. The variables $\bar{Z}_\alpha = \sup_K \sum_{\sigma} \exp \mathcal{H}(\alpha, \sigma)$ have the same law for all α , have finite mean, and are independent of (v_α) , so $\mathbb{E} \sum_\alpha v_\alpha \bar{Z}_\alpha = \mathbb{E} \bar{Z}_\alpha < \infty$. Then, almost surely, the series defining $\langle \cdot \rangle$ converge uniformly on K , the Gibbs weights are continuous on K , and so is $\langle f \rangle$; by dominated convergence, $\mathbb{E} \langle f \rangle$ is continuous as well. For a fixed amplitude a , multiplying the exponential envelope by $|\partial_a \mathcal{H}|$ or $|\partial_a \mathcal{H}|^2$ preserves integrability, by Gaussian exponential moments. The same domination therefore gives almost-sure twice differentiability in a . The first derivative of $\log Z$ is $\langle \partial_a \mathcal{H} \rangle$. Since $\mathcal{H} + \sum_x q_{xK}$ is affine in each amplitude, the second derivative of $\log Z + \sum_x q_{xK}$ is $\langle (\partial_a \mathcal{H} - \langle \partial_a \mathcal{H} \rangle)^2 \rangle$.

The random function $\log Z + \sum_x q_{xK}$ is convex in each amplitude, as a logarithm of a sum of exponentials of affine functions. Its difference quotients are monotone, and $\mathbb{E} |\log Z| < \infty$ by Lemma 4.10 below. Subtracting the deterministic compensator, we conclude that the derivative of F in each amplitude is the expectation of the derivative of $\tilde{F}$. We now compute these expectations by (4.7), applied conditionally on (v_α) with $\Sigma = \mathbb{N}^K \times \{-1, 1\}^V$.

- (b) The process $H^{(\alpha_1)}(\sigma)$ has covariance $\mathbf{1}\{\alpha_1 = \alpha'_1\}(\sigma \odot \sigma') \cdot S(\sigma \odot \sigma')/2$, which equals $n/2$ on the diagonal. By (4.7) with $r = 1$,

$$\partial_t \mathbb{E} \log Z = \frac{\beta}{2\sqrt{t}} \mathbb{E} \langle H^{(\alpha_1)}(\sigma) \rangle = \frac{\beta^2}{2} \left(\frac{n}{2} - \frac{1}{2} \mathbb{E} \langle u^* \cdot Su^* \rangle \right),$$

which gives the formula. Since $\|S\| \leq 1$ and $|u_x^*| \leq 1$, we have $|u^* \cdot Su^*| \leq n$.

(c) Suppose first that all the increments $q_{x1}, q_{xk} - q_{x,k-1}$ are positive. For $1 \leq j \leq K$, by (4.7) the amplitude derivatives give, exactly as in [56, proof of Lemma 6.2],

$$\begin{aligned} \partial_{q_{xj}} \mathbb{E} \log Z &= \mathbb{E} \langle 1 - \mathbf{1}\{\alpha^1 \wedge \alpha^2 \geq j\} u_x \rangle \\ &\quad - \mathbb{E} \langle 1 - \mathbf{1}\{\alpha^1 \wedge \alpha^2 \geq j+1\} u_x \rangle \mathbf{1}\{j < K\} - \mathbf{1}\{j = K\} \\ &= -\mathbb{E} \langle \mathbf{1}\{\alpha^1 \wedge \alpha^2 = j\} u_x \rangle, \end{aligned}$$

where $u = \sigma^1 \odot \sigma^2$. Since both sides are continuous on $\mathcal{C}_K^V$ by (a), this identity holds everywhere. For $j \geq 1$ we have $\mathbf{1}\{\alpha^1 \wedge \alpha^2 = j\} u_x = \mathbf{1}\{\vartheta_{12} = z_j\} u_x^*$, and F does not depend on q_{x0} . With (4.6), this proves $D_q F = \mathbf{p}$.

It remains to show that $\mathbf{p}_x \in \mathcal{C}_K \cap [0, 1]^{K+1}$. For a model in which all leaves carry the same Hamiltonian, the corresponding statement is [108, Proposition 3.6], and we follow its proof through the cascade recursion. Fix x and write the leaf function as $X = L(Y_x, \omega)$, where $Y_x = Y_x(\alpha)$,

$$L(s, \omega) = \log \sum_{\sigma} \exp(s\sigma_x + \mathcal{H}'(\sigma)),$$

$\mathcal{H}'$ is $\mathcal{H}$ without the term $Y_x(\alpha)\sigma_x$, and ω collects all the vertex variables on the path of α other than $\xi_{x,\alpha|_k}$. The function L satisfies (3.6). The integrations defining X_0 in Theorem 4.7 are then exactly the recursion (3.7) with $\Delta_1 = 2q_{x1}$, $\Delta_k = 2(q_{xk} - q_{x,k-1})$ for $k \geq 2$, and starting point 0; that is, $X_0 = L_0(0)$. By Lemma 3.1(c), for $1 \leq j < K$,

$$\partial_{q_{xj}} X_0 = 2(\partial_{\Delta_j} - \partial_{\Delta_{j+1}}) L_0(0) = -w_j \Lambda_{0,j}(0),$$

and $\partial_{q_{xK}} X_0 = 2\partial_{\Delta_K} L_0(0) - 1 = -w_K \Lambda_{0,K}(0)$, where -1 comes from the term $-q_{xK}$ in $\mathcal{H}'$. Comparing with the first part, we find that $\mathbf{p}_{xj} = \Lambda_{0,j}(0)$ for $1 \leq j \leq K$, and Lemma 3.1(a) shows that $0 \leq \mathbf{p}_{x1} \leq \dots \leq \mathbf{p}_{xK} \leq 1$. Since $\mathbf{p}_{x0} = 0$, we conclude that $\mathbf{p}_x \in \mathcal{C}_K \cap [0, 1]^{K+1}$. The Lipschitz bound follows from (3.4).

(d) Concavity holds because $\log Z$ is convex in y . By (4.7) and Lemma 4.8, the covariance of $G_{x,\iota}$ with the Gaussian part of $\mathcal{H}$ is $s y_{x\iota} k_{\iota}$, which gives the formula for $\partial_{y_{x\iota}} F$. Replacing s by θs in the same formula, we obtain

$$|\partial_{\theta} F| \leq \frac{2\theta}{n} \sum_{x,\iota} s^2 y_{x\iota}^2 \leq \frac{9\theta}{2n} |\mathcal{S}| |I| s^2,$$

and integrating over $\theta \in [0, 1]$ gives the last bound.

(e) If $q = 0$ and $s = 0$, the leaf function $X = \log Z_G^{(\alpha_1)}$, where $Z_G^{(a)}$ is the partition function of $Z_G(\beta\sqrt{t}, h)$ with couplings $g^{(a)}$, depends only on the variable at the depth-one ancestor. Then $X_k = X$ for $k \geq 1$ and $X_0 = \zeta_0^{-1} \log \mathbb{E} Z_G(\beta\sqrt{t}, h) \zeta_0$, and the identity follows from Theorem 4.7. By Jensen's inequality, $\log \mathbb{E} Z_G^{\zeta_0} \geq \zeta_0 \mathbb{E} \log Z_G$, which gives the inequality.

(f) If $t = 0$ and $s = 0$, then $X = \sum_x (\log 2 \cosh(h + Y_x(\alpha)) - q_{xK})$ is a sum of functions of independent variables, one for each x . The integrations in Theorem 4.7 then factorize, as in [56, Corollary 5.26], so that $X_0 = \sum_x X_{0,x}$, where $X_{0,x}$ is the recursion of Section 3.3 with $v = 2\tilde{q}_x = (0, 2q_{x1}, \dots, 2q_{xK})$ evaluated at h , minus q_{xK} . By (3.12), $X_{0,x} = -\psi_h(\tilde{q}_x)$. $\square$

The next lemma is the reason for attaching independent copies of the couplings to the first-level branches: the fluctuations of $\log Z$ do not grow with n .

Lemma 4.10. *There exists C_{fl} , depending only on ζ_0 , such that $\text{Var}(\log Z) \leq C_{\text{fl}}^2$ for all parameters (t, q, y) . In particular,*

$$\mathbb{E}|\tilde{F}(t, q, y) - F(t, q, y)| \leq \frac{C_{\text{fl}}}{n}.$$

Proof. Let $Z_\alpha = \sum_\sigma \exp \mathcal{H}(\alpha, \sigma)$. For a vertex $\gamma \in \mathbb{N}^k$, $1 \leq k \leq K$, let

$$U_\gamma = \sum_{\alpha: \alpha|_k = \gamma} \left(\prod_{i=k+1}^K u_{\alpha|_i} \right) Z_\alpha,$$

and let T_γ be the same sum with Z_α replaced by one. Then $\log Z = \log \sum_{a \geq 1} u_{\mathcal{O}a} U_a - \log \sum_{a \geq 1} u_{\mathcal{O}a} T_a$. Since no variable is attached to the root, the pairs (U_a, T_a) , $a \geq 1$, are independent and identically distributed, and independent of $(u_{\mathcal{O}a})_a$.

We claim that $\mathbb{E}U_\gamma^{\zeta_{k-1}} < \infty$ for $\gamma \in \mathbb{N}^k$. For $k = K$ this holds because Z_α has moments of every order. If it holds at depth $k+1$ and $\gamma \in \mathbb{N}^k$, then $U_\gamma = \sum_i u_{\gamma i} U_{\gamma i}$, and conditionally on the variables attached to the ancestors of γ and to γ , the $U_{\gamma i}$ are independent and identically distributed with conditional moment m_γ of order ζ_k . By (4.8), $\mathbb{E}U_\gamma^{\zeta_{k-1}} = \mathbb{E}[m_\gamma^{\zeta_{k-1}/\zeta_k}] \mathbb{E}(\sum_i u_{\gamma i})^{\zeta_{k-1}}$, which is finite by Jensen's inequality and the moment bound in (4.8). The same holds for T_γ .

Let $S = \sum_a u_{\mathcal{O}a}$. By (4.8), $\log \sum_a u_{\mathcal{O}a} U_a$ and $\log \sum_a u_{\mathcal{O}a} T_a$ both have the law of $\log S$ plus a constant. This implies that $\text{Var}(\log Z) \leq 4 \text{Var}(\log S) \leq 4\mathbb{E}(\log S)^2$. By Campbell's formula, $\mathbb{E}e^{-\lambda S} = \exp(-\Gamma(1 - \zeta_0)\lambda^{\zeta_0})$ for $\lambda \geq 0$, and integrating over $\lambda \in (0, \infty)$ we obtain

$$\mathbb{E}S^{-1} = \int_0^\infty \exp(-\Gamma(1 - \zeta_0)\lambda^{\zeta_0}) d\lambda < \infty.$$

Further, $\mathbb{E}S^{\zeta_0/2} < \infty$ by the moment bound in (4.8). Since $(\log s)^2 \leq 16\zeta_0^{-2}(s^{\zeta_0/2} + s^{-1})$ for $s > 0$, the second moment of $\log S$ is bounded by a constant depending only on ζ_0 , and so is $\text{Var}(\log Z)$. The bound on $\mathbb{E}|\tilde{F} - F|$ follows from the Cauchy-Schwarz inequality, since $\tilde{F} - F = -n^{-1}(\log Z - \mathbb{E} \log Z)$. $\square$

4.4 Synchronization

We use the following bound of Mourrat on conditional variances, which is [106, Proposition 5.5], with the hypothesis of [106, Theorem 5.3]. It is a finitary form of the synchro-

nization theorem of Panchenko [115, Theorem 4], which was extended to Potts and vector spins in [116, 117]. Its proof uses Panchenko's ultrametricity theorem [113] for measures that satisfy the Ghirlanda–Guerra identities [74]. Let $(c_i)_{i \geq 1}$ be a fixed enumeration of $\mathbb{Q} \cap [0, 1]$.

Proposition 4.11. *For every $\varepsilon > 0$ there exists $\delta > 0$ such that the following holds. Let $\mathfrak{G}$ be a random probability measure supported on the Cartesian product of the unit balls of two Hilbert spaces. Let $\langle \cdot \rangle$ be the expectation associated with $\mathfrak{G}^{\otimes \mathbb{N}}$, with canonical random variables $(\sigma_1^\ell, \sigma_2^\ell)_{\ell \geq 1}$, and let $\mathbb{E}$ be the expectation with respect to the randomness of $\mathfrak{G}$. For $a \in \{1, 2\}$ and $\ell, \ell', r \geq 1$, let $O_a^{\ell\ell'} = \sigma_a^\ell \cdot \sigma_a^{\ell'}$ and $O^{\leq r} = (O_a^{\ell\ell'})_{a \in \{1, 2\}, \ell, \ell' \leq r}$. Suppose that for all $r, i_1, i_2, p \in \{1, \dots, \lfloor \delta^{-1} \rfloor\}$ and every continuous $f: \mathbb{R}^{2 \times r \times r} \rightarrow \mathbb{R}$ with $|f| \leq 1$, writing $k_{\ell\ell'} = (c_{i_1} O_1^{\ell\ell'} + c_{i_2} O_2^{\ell\ell'})^p$, we have*

$$\left| \mathbb{E} \langle f(O^{\leq r}) k_{1,r+1} \rangle - \frac{1}{r} \mathbb{E} \langle f(O^{\leq r}) \rangle \mathbb{E} \langle k_{12} \rangle - \frac{1}{r} \sum_{\ell=2}^r \mathbb{E} \langle f(O^{\leq r}) k_{1\ell} \rangle \right| \leq \delta. \quad (4.11)$$

Suppose further that the law of O_2^{12} is $k^{-1} \sum_{\ell=1}^k \delta_{q_\ell}$ for some integer $k \geq 1$ and $-1 = q_0 < q_1 < \dots < q_k \leq 1$. Then

$$\mathbb{E} \left\langle (O_1^{12} - \mathbb{E} \langle O_1^{12} | O_2^{12} \rangle)^2 \right\rangle \leq \frac{12}{k} + \varepsilon k^2 \max_{0 \leq \ell < k} (q_{\ell+1} - q_\ell)^{-1}.$$

The law of O_2^{12} and the conditional expectation are taken under the probability measure $\mathbb{E} \langle \cdot \rangle$. In [106] the number of replicas is denoted by n , the enumeration by (λ_h) , the overlaps $O_a^{\ell\ell'}$ by $R_a^{\ell, \ell'}$, and the random measure by G . For $\delta > 0$ we define the finite set

$$I_\delta = \left\{ \left(\frac{c_{i_2}}{c_{i_1} + c_{i_2}}, p \right) : 1 \leq i_1, i_2, p \leq \lfloor \delta^{-1} \rfloor, c_{i_1} + c_{i_2} > 0 \right\} \subseteq [0, 1] \times \mathbb{N}. \quad (4.12)$$

For $x \in V$ we write

$$\mathcal{V}_x = \sum_{j=0}^K w_j \text{Var} (q_{x,12} | \vartheta_{12} = z_j),$$

where the variance is computed under $\mathbb{E} \langle \cdot | \vartheta_{12} = z_j \rangle$.

4.5 Gaussian perturbations and the envelope

In this subsection $\mathcal{S}$, I , and s are as in the definition of the model, and $\varepsilon \in (0, 1/8]$. We define the deterministic envelope

$$F_\star(t, q) = \min_{y \in Y} \left\{ F(t, q, y) + \frac{s^2}{n\varepsilon} \sum_{x \in \mathcal{S}, \iota \in I} (y_{x\iota} - 1)^2 \right\}. \quad (4.13)$$

By Lemma 4.9(a), the minimum is attained. By Lemma 4.9, since the penalty does not depend on (t, q) , vanishes at $y = \mathbf{1}$, and is nonnegative,

$$|F_*(t, q) - F_*(t', q')| \leq \frac{\beta^2}{4}|t - t'| + \|q - q'\|_1, \quad |F_*(t, q) - F_0(t, q)| \leq r_s. \quad (4.14)$$

Small Gaussian perturbations that enforce the Ghirlanda–Guerra identities [74] through the concentration of the free energy are a standard tool; see [133] and [115, Section 3]. The proof of Lemma 4.12 follows Steps 1–3 of the proof of [106, Theorem 4.1]. There, the lower bound on the second derivative of the free energy in the perturbation parameters comes from a test function that touches the free energy; here it comes from the minimality of y_* in (4.13). Because of Lemma 4.10, the strength s can be chosen independently of n .

Lemma 4.12. *Suppose that $s \geq 2$. Let $(t, q) \in (0, 2) \times C_K^V$ and let y_* be a minimizer in (4.13). Then $|y_{*x\iota} - 1| \leq 3\epsilon/2$ for every (x, ι) , and, under the Gibbs measure at (t, q, y_*) , for every $(x, \iota) \in \mathcal{S} \times I$, every $r \geq 1$, and every measurable function f of r replicas with $|f| \leq 1$,*

$$\left| \mathbb{E} \langle f k_{x\iota, 1, r+1} \rangle - \frac{1}{r} \mathbb{E} \langle f \rangle \mathbb{E} \langle k_{x\iota, 12} \rangle - \frac{1}{r} \sum_{\ell=2}^r \mathbb{E} \langle f k_{x\iota, 1\ell} \rangle \right| \leq \frac{C_{GG}}{s\sqrt{\epsilon}}, \quad (4.15)$$

where C_{GG} depends only on ζ_0 .

Proof. Let $\lambda = s^2/(n\epsilon)$. By Lemma 4.9(d) and Lemma 4.8, $|\partial_{y_{x\iota}} F| \leq 2s^2 y_{x\iota}/n \leq 3s^2/n$. If $y_{*x\iota} = 3/2$, minimality would require $\partial_{y_{x\iota}} F + \lambda \leq 0$, which implies $\lambda \leq 3s^2/n$ and contradicts $\epsilon \leq 1/8$; the case $y_{*x\iota} = 1/2$ is excluded in the same way. This shows that y_* is interior, and stationarity gives $|y_{*x\iota} - 1| = |\partial_{y_{x\iota}} F|/(2\lambda) \leq 3\epsilon/2 \leq 3/16$. In particular $y_* \pm ze_{x\iota} \in Y$ for $z \in [0, 1/4]$.

Fix (x, ι) and write $g(y) = F(t, q, y)$ and $\tilde{g}(y) = \tilde{F}(t, q, y)$, where y equals y_* except that $y_{x\iota} = y$. By global minimality of y_* and stationarity, we have, for $|z| \leq 1/4$,

$$g(y_* + z) \geq g(y_*) + g'(y_*)z - \lambda z^2, \quad (4.16)$$

since the penalty changes by $2\lambda(y_* - 1)z + \lambda z^2$ and $g'(y_*) = -2\lambda(y_* - 1)$. Because g is smooth, it follows from (4.16) that $g''(y_*) \geq -2\lambda$. The random function $\tilde{g}$ is concave and, by the domination argument in the proof of Lemma 4.9(a), almost surely twice differentiable, with $\tilde{g}' = -n^{-1}s\langle G_{x,\iota} \rangle$ and $\tilde{g}'' = -n^{-1}s^2\langle (G_{x,\iota} - \langle G_{x,\iota} \rangle)^2 \rangle$, and $\mathbb{E}\tilde{g}' = g'$ by the argument in the proof of Lemma 4.9(a). By Fatou's lemma applied to the nonnegative quotients $(\tilde{g}'(y_*) - \tilde{g}'(y_* + \epsilon'))/\epsilon'$ as $\epsilon' \downarrow 0$,

$$\frac{s^2}{n} \mathbb{E} \langle (G_{x,\iota} - \langle G_{x,\iota} \rangle)^2 \rangle \leq -g''(y_*) \leq 2\lambda, \quad \text{so} \quad \mathbb{E} \langle (G_{x,\iota} - \langle G_{x,\iota} \rangle)^2 \rangle \leq \frac{2}{\epsilon}. \quad (4.17)$$

Let $z \in (0, 1/4]$ and $\Delta = \tilde{g} - g$. By concavity of $\tilde{g}$ and by (4.16) at $\pm z$,

$$\tilde{g}'(y_*) - g'(y_*) \leq \frac{\Delta(y_*) - \Delta(y_* - z)}{z} + \lambda z, \quad \tilde{g}'(y_*) - g'(y_*) \geq \frac{\Delta(y_* + z) - \Delta(y_*)}{z} - \lambda z.$$

Since the three points y_* , $y_* \pm z$ are deterministic, Lemma 4.10 applies at each of them, and we obtain $\mathbb{E}|\tilde{g}'(y_*) - g'(y_*)| \leq 3C_{\text{fl}}/(nz) + \lambda z$, that is,

$$\mathbb{E}|\langle G_{x,l} \rangle - \mathbb{E}\langle G_{x,l} \rangle| \leq \frac{3C_{\text{fl}}}{sz} + \frac{sz}{\epsilon} = \frac{3C_{\text{fl}} + 1}{\sqrt{\epsilon}}$$

for the choice $z = s^{-1}\sqrt{\epsilon}$, which is at most $1/4$ because $s \geq 2$ and $\epsilon \leq 1/8$. Combining this with (4.17), we obtain

$$\mathbb{E}|\langle G_{x,l} \rangle - \mathbb{E}\langle G_{x,l} \rangle| \leq \frac{3C_{\text{fl}} + 1 + \sqrt{2}}{\sqrt{\epsilon}}. \quad (4.18)$$

Let f be a function of r replicas with $|f| \leq 1$. By (4.7) and Lemma 4.8, with $k_{x_l,11} = 1$,

$$\mathbb{E}\langle G_{x,l}(\alpha^1, \sigma^1) f \rangle = sy_{*x_l} \mathbb{E}\left\langle f\left(\sum_{\ell=1}^r k_{x_l,1\ell} - rk_{x_l,1,r+1}\right) \right\rangle, \quad \mathbb{E}\langle G_{x,l} \rangle = sy_{*x_l} (1 - \mathbb{E}\langle k_{x_l,12} \rangle).$$

Subtracting $\mathbb{E}\langle G_{x,l} \rangle \mathbb{E}\langle f \rangle$ from the first identity and dividing by $-rsy_{*x_l}$, we see that the left side of (4.15) equals $|\mathbb{E}\langle (G_{x,l}(\alpha^1, \sigma^1) - \mathbb{E}\langle G_{x,l} \rangle) f \rangle| / (rsy_{*x_l})$. Since $y_{*x_l} \geq 13/16$, (4.15) follows from (4.18), with $C_{\text{GG}} = 16(3C_{\text{fl}} + 1 + \sqrt{2})/13$. $\square$

4.6 Sparsification and the drift

We use the following theorem of Cohen, Nelson, and Woodruff on row sparsification [42, Theorem 5].

Theorem 4.13. *There exists an absolute constant $C_0 \geq 1$ such that the following holds. Let B be a real $n \times d$ matrix with $\|B\| \leq 1$ and $\|B\|_{\text{F}}^2 \leq r$, where $r > 0$, and let $\lambda \in (0, 1)$. Then there exists a diagonal $n \times n$ matrix Θ with nonnegative entries and at most $C_0(1 + r/\lambda^2)$ nonzero entries such that $\|B^\top \Theta^2 B - B^\top B\| \leq \lambda$.*

In [42, Theorem 5] the diagonal matrix has $O(r/\lambda^2)$ nonzero entries; replacing it by its absolute value does not change $B^\top \Theta^2 B$. The proof of [42, Theorem 5] adapts the deterministic construction of Batson, Spielman, and Srivastava [20], in which the number of nonzero entries is bounded in terms of the rank of B . For $B = S$ the rank may be of order n , whereas $\|S\|_{\text{F}}^2 = n/D$. We apply Theorem 4.13 to $B = S$, with $r = \text{tr } S^2 = n/D \geq 1$, and obtain Θ with

$$\|S\Theta^2 S - S^2\| \leq \lambda, \quad |\{x : \Theta_{xx} \neq 0\}| \leq \frac{2C_0 n}{D\lambda^2}. \quad (4.19)$$

Lemma 4.14. *Let Θ satisfy (4.19) and let $\{x : \Theta_{xx} \neq 0\} \subseteq \mathcal{S}$. Let $t > 0$, $q \in \mathcal{C}_K^V$, $y \in \mathcal{Y}$, and $\varepsilon_0 > 0$, and suppose that $\mathcal{V}_x \leq \varepsilon_0$ at (t, q, y) for every $x \in \mathcal{S}$. Then for every $\eta > 0$,*

$$|\partial_t F(t, q, y) - H_S(D_q F(t, q, y))| \leq \tau, \quad \tau = \frac{\beta^2}{4} \left[\eta + \frac{\lambda + (1 + \lambda)\varepsilon_0}{4\eta} \right].$$

Proof. For $1 \leq j \leq K$ let $p_j = \mathbb{E}\langle u^* | \vartheta_{12} = z_j \rangle \in \mathbb{R}^V$ and let C_j be the covariance matrix of u^* under $\mathbb{E}\langle \cdot | \vartheta_{12} = z_j \rangle$, which is positive semidefinite with $\text{tr } C_j \leq n$. Since $u^* = 0$ on $\{\vartheta_{12} = z_0\}$, Lemma 4.9(b),(c) and (4.6) give

$$\partial_t F = \frac{\beta^2}{4n} \sum_{j=1}^K w_j \mathbb{E}\langle u^* \cdot S u^* | \vartheta_{12} = z_j \rangle = H_S(p) + \frac{\beta^2}{4n} \sum_{j=1}^K w_j \text{tr}(S C_j),$$

with $p = D_q F$. For the Sherrington–Kirkpatrick model, the analogous identity is [56, Lemma 6.2]; there the error term is the conditional variance of the overlap given the cascade overlap, which is nonnegative. For each eigenvalue μ of S we have $|\mu| \leq \eta + \mu^2/(4\eta)$. Since C_j is positive semidefinite, we deduce that $|\text{tr}(S C_j)| \leq \eta \text{tr } C_j + \text{tr}(S^2 C_j)/(4\eta)$. By (4.19), $\text{tr}(S^2 C_j) \leq \lambda \text{tr } C_j + \text{tr}(S \Theta^2 S C_j)$, and

$$\text{tr}(S \Theta^2 S C_j) = \sum_x \Theta_{xx}^2 (S C_j S)_{xx} = \sum_x \Theta_{xx}^2 \text{Var}(\varrho_{x,12} | \vartheta_{12} = z_j).$$

Let $\omega_x = \Theta_{xx}^2/n$. Testing (4.19) on $\mathbf{1}$ and using $S\mathbf{1} = \mathbf{1}$, we find $\sum_x \omega_x \leq 1 + \lambda$. We conclude that

$$\sum_j w_j \frac{\text{tr}(S^2 C_j)}{n} \leq \lambda + \sum_{x \in \mathcal{S}} \omega_x \mathcal{V}_x \leq \lambda + (1 + \lambda)\varepsilon_0,$$

and the claim follows. $\square$

Lemma 4.15. *Suppose that $\zeta_j = (j + 1)/(K + 1)$ for $0 \leq j \leq K$. Let δ_M be given by Proposition 4.11 for $\varepsilon = K^{-1}(K + 1)^{-3}$, and let $I = I_{\delta_M}$ be as in (4.12). Let Θ satisfy (4.19), let $\mathcal{S} = \{x : \Theta_{xx} \neq 0\}$, and suppose that*

$$s \geq 2 \quad \text{and} \quad \frac{2^{\lfloor \delta_M^{-1} \rfloor} C_{GG}}{s\sqrt{\varepsilon}} \leq \delta_M. \quad (4.20)$$

Then $F_\star$ is a supersolution of $\partial_t u - H_{\text{ext}}(D_q u) \geq -\tau$, with τ as in Lemma 4.14 for $\varepsilon_0 = 13/(K + 1)$.

Proof. Let φ be continuously differentiable near $(t_0, q_0) \in (0, 2) \times \mathcal{C}_K^V$, and let $F_\star - \varphi$ have a local minimum at (t_0, q_0) . Let $y_\star$ be a minimizer in (4.13) at (t_0, q_0) ; it is a deterministic point. The function $(t, q) \mapsto F(t, q, y_\star) + s^2(n\varepsilon)^{-1} \sum_{x,l} (y_{\star xl} - 1)^2$ is at least $F_\star$ and equals it at (t_0, q_0) . Then $F(\cdot, \cdot, y_\star) - \varphi$ also has a local minimum at (t_0, q_0) . By Lemma 4.9(a) and (4.3), $\partial_t \varphi = \partial_t F$ and $D_q F \succeq_* D_q \varphi$ at $(t_0, q_0, y_\star)$.

Fix $x \in \mathcal{S}$, and let $\mathfrak{G}$ be the image of the Gibbs measure at (t_0, q_0, y_*) under the map $(\alpha, \sigma) \mapsto (F_x(\alpha, \sigma), \mathbf{e}(\alpha))$, where

$$F_x(\alpha, \sigma) = (S_{xy}^{1/2} \sigma_y e_{\alpha_1})_{y \in V}, \quad \mathbf{e}(\alpha) = \sum_{k=1}^K (z_k - z_{k-1})^{1/2} e_{\alpha|_k},$$

for orthonormal families $(e_a)_{a \in \mathbb{N}}$ and $(e_\gamma)_{\gamma \in \mathbb{A} \setminus \{\emptyset\}}$. Both components are unit vectors, since $\sum_y S_{xy} = 1$ and $z_K - z_0 = 1$. For the replicas $(\alpha^\ell, \sigma^\ell)$, the overlaps of Proposition 4.11 are $O_1^{\ell\ell'} = \varrho_{x, \ell\ell'}$ and $O_2^{\ell\ell'} = \vartheta_{\ell\ell'}$. We check the hypotheses of Proposition 4.11 with $\delta = \delta_M$. Let $r, i_1, i_2, p \leq \lfloor \delta_M^{-1} \rfloor$ and let f be as there. If $c_{i_1} + c_{i_2} = 0$, then $k_{\ell\ell'} = 0$ for all ℓ, ℓ' , and the left side of (4.11) vanishes. Otherwise

$$(c_{i_1} \varrho + c_{i_2} \vartheta)^p = (c_{i_1} + c_{i_2})^p k_\iota(\vartheta, \varrho), \quad \iota = \left(\frac{c_{i_2}}{c_{i_1} + c_{i_2}}, p \right) \in I,$$

and $f(O^{\leq r})$ is a measurable function of r replicas with $|f(O^{\leq r})| \leq 1$. Then the left side of (4.11) is $(c_{i_1} + c_{i_2})^p \leq 2^{\lfloor \delta_M^{-1} \rfloor}$ times the left side of (4.15), and it is at most δ_M by Lemma 4.12 and (4.20). By (4.6), the law of ϑ_{12} is $(K+1)^{-1} \sum_{j=0}^K \delta_{z_j}$. This is the law in Proposition 4.11 with $k = K+1$ and $q_\ell = z_{\ell-1}$ for $1 \leq \ell \leq k$, and the spacings $q_{\ell+1} - q_\ell$ are 1 for $\ell = 0$ and $1/K$ otherwise. Since each event $\{\vartheta_{12} = z_j\}$ has probability $w_j > 0$, the left side of the conclusion of Proposition 4.11 equals $\mathcal{V}_x$, and the proposition implies that

$$\mathcal{V}_x \leq \frac{12}{K+1} + \varepsilon(K+1)^2 K = \frac{13}{K+1}.$$

This holds for every $x \in \mathcal{S}$, and Lemma 4.14 implies that $\partial_t F \geq H_S(D_q F) - \tau$ at (t_0, q_0, y_*) .

Since $D_q F \in \mathcal{C}_K^V$ by Lemma 4.9(c), we have $H_S(D_q F) = H_{\text{ext}}(D_q F) \geq H_{\text{ext}}(D_q \varphi)$ by Lemma 4.2(c). Therefore $\partial_t \varphi - H_{\text{ext}}(D_q \varphi) \geq -\tau$ at (t_0, q_0) . $\square$

4.7 Proof of Proposition 4.1

If $\beta = 0$, then $p_G(0, h) = \log 2 \cosh h$ by Lemma 2.1, and $\mathcal{P}(0, h) = \log 2 \cosh h$ by the same lemma for the Sherrington–Kirkpatrick model. Let $\beta > 0$, $h \in \mathbb{R}$, and $v \in (0, 1)$.

We first make all choices that do not depend on D or G . Let $\nu = \nu_{\beta, h}$ be the Parisi measure, and let $Q(\theta) = \min\{s \in [0, 1] : \nu([0, s]) \geq \theta\}$ for $\theta \in (0, 1)$. Extend Q to $[0, 1]$ by $Q(0) = \min \text{supp } \nu$ and $Q(1) = \max \text{supp } \nu$. The function Q is nondecreasing, and the image of the Lebesgue measure on $(0, 1)$ under Q is ν . For $K \geq 1$, let $r_j^{(K)} = Q((j+1/2)/(K+1))$ for $0 \leq j \leq K$, so that $r^{(K)} \in \mathcal{C}_K \cap [0, 1]^{K+1}$, and let $\mu^{(K)} = (K+1)^{-1} \sum_{j=0}^K \delta_{r_j^{(K)}}$. For $f \in C([0, 1])$, the integral $\int f d\mu^{(K)}$ is a Riemann sum of $f \circ Q$ for the partition of $[0, 1]$ into $K+1$ intervals of equal length. The function $f \circ Q$ is bounded and continuous outside the set of discontinuities of Q , which is countable.

By Lebesgue's criterion it is Riemann integrable, and these Riemann sums converge to $\int_0^1 f(Q(\theta)) d\theta = \int f dv$ as $K \rightarrow \infty$. Then $\mu^{(K)} \rightarrow \nu$ weakly, and Lemma 3.10 and (2.7) imply that $\mathcal{P}_{\beta,h}(\mu^{(K)}) \rightarrow \mathcal{P}(\beta, h)$.

Let $\eta = 2v/(3\beta^2)$, so that $\beta^2\eta/4 = v/6$. Let $K \geq 1$ be so large that, with $\zeta_j = (j+1)/(K+1)$ (so that $w_j = \bar{w} = \zeta_0 = 1/(K+1) \leq 1/2$ and $\mu_{r(K)} = \mu^{(K)}$), we have

$$\mathcal{P}_{\beta,h}(\mu^{(K)}) \leq \mathcal{P}(\beta, h) + \frac{v}{6}, \quad \frac{\beta^2\zeta_0}{4} \leq \frac{v}{6} \quad \text{and} \quad \frac{\beta^2}{16\eta} \cdot 21\bar{w} \leq \frac{v}{6}.$$

Let $\lambda = \bar{w}$ and $\varepsilon_0 = 13\bar{w}$, so that $\lambda + (1+\lambda)\varepsilon_0 \leq 21\bar{w}$ and the constant τ of Lemma 4.14 satisfies $\tau \leq v/3$. Let $\delta = v/6$ and $\varepsilon = 1/8$. Let δ_M and $I = I_{\delta_M}$ be as in Lemma 4.15, and let C_{GG} be as in Lemma 4.12. These objects depend only on β, h , and v .

Now let $D \geq 1$, let $s = D^{1/4}$, and let D_0 be such that (4.20) holds for $D \geq D_0$. Let $D \geq D_0$ and $G \in \mathcal{G}_D$. Let Θ be given by (4.19), let $\mathcal{S} = \{x : \Theta_{xx} \neq 0\}$, and let the model of Section 4.3 be constructed with these $\mathcal{S}, I$, and s . By (4.19),

$$r_s = \frac{9}{4} \frac{|\mathcal{S}||I|s^2}{n} \leq \frac{9C_0|I|}{2\lambda^2\sqrt{D}}. \quad (4.21)$$

We apply Lemma 4.5 with $H = H_{\text{ext}}, V = V_\delta$, and $E = F_\star$, restricted to $[0, 2] \times \mathcal{C}_K^V$. By Lemma 4.2(c), H_{ext} is Lipschitz on bounded sets. Hypothesis (i) holds by Lemma 4.6(b) and Lemma 4.15; (ii) holds with $L_t = \beta^2$ and $L_q = 1$ by Lemma 4.6(a) and (4.14). For (iii), by Lemma 4.6(c), Lemma 4.9(f), and (4.14),

$$V_\delta(0, q) \leq \frac{1}{n} \sum_x \psi_h(\tilde{q}_x) + \delta = F_0(0, q) + \delta \leq F_\star(0, q) + \delta + r_s.$$

The lemma then gives $V_\delta(1, 0) \leq F_\star(1, 0) + \delta + r_s + \tau$. By Lemma 4.6(d), (4.14), and Lemma 4.9(e),

$$U(1, 0) - \frac{\beta^2\zeta_0}{4} \leq F_0(1, 0) + \delta + 2r_s + \tau \leq \frac{\beta^2}{4} - p_G(\beta, h) + \delta + 2r_s + \tau.$$

By Lemma 3.10 with $r = r^{(K)}$ and the choice of K , we have $U(1, 0) \geq \beta^2/4 - \mathcal{P}_{\beta,h}(\mu^{(K)}) \geq \beta^2/4 - \mathcal{P}(\beta, h) - v/6$. Combining these bounds with the choices above and (4.21), we obtain

$$p_G(\beta, h) \leq \mathcal{P}(\beta, h) + \frac{v}{6} + \frac{\beta^2\zeta_0}{4} + \delta + \tau + 2r_s \leq \mathcal{P}(\beta, h) + \frac{5v}{6} + \frac{9C_0|I|}{\lambda^2\sqrt{D}}.$$

The right side does not depend on $G \in \mathcal{G}_D$ or on n . Taking the supremum over $G \in \mathcal{G}_D$ and letting $D \rightarrow \infty$, we obtain $\limsup_{D \rightarrow \infty} \sup_{G \in \mathcal{G}_D} p_G(\beta, h) \leq \mathcal{P}(\beta, h) + 5v/6$. Since v is arbitrary, this completes the proof of Proposition 4.1.

Message-passing centers on tori and hypercubes

The goal of this section is to prove Proposition 5.2, which constructs the centers used in the lower bound on tori and hypercubes. A center is a vector $m \in (-1, 1)^V$ computed from finitely many products of the coupling matrix W with vectors, each of which is a function of the earlier products. The following definition records this structure, which is the form required by Theorem 6.2.

Definition 5.1. Let $G \in \mathcal{G}_D$ and let $T \geq 1$ be an integer. A *query history of depth T* on G consists of a random variable U independent of $(g_e)_{e \in E}$ and random vectors $x^1, \dots, x^T \in \mathbb{R}^V$ such that, for each $t \leq T$, the vector x^t is measurable with respect to the σ -algebra generated by U and $Wx^1, \dots, Wx^{t-1}$. We write $\mathcal{H}$ for the σ -algebra generated by U and $Wx^1, \dots, Wx^T$. A *center* of the history is an $\mathcal{H}$ -measurable vector $m \in (-1, 1)^V$ such that $m = x^t$ for some $t \leq T$.

Throughout this section, $(G_d)_{d \geq 1}$ denotes one of two sequences of graphs: either the tori $G_d = \mathbb{T}_L^d$ for a fixed $L \geq 3$, with $D = 2d$ and $n = L^d$, or the hypercubes $G_d = Q_d$, with $D = d$ and $n = 2^d$. Both are vertex-transitive, since they are Cayley graphs of $(\mathbb{Z}/L\mathbb{Z})^d$ and $(\mathbb{Z}/2\mathbb{Z})^d$. We use the notation of (2.2) on G_d , and we write $\|w\|_n = (n^{-1} \sum_x w_x^2)^{1/2}$ for the normalized Euclidean norm of $w \in \mathbb{R}^V$. The Parisi measure $\nu = \nu_{\beta, h}$, the endpoints $q_- \leq q_+$ of its support, the functions a and v , the Parisi diffusion X_t , the martingale $M_t = \partial_x \Phi(t, X_t)$ and the law $\mu_+ = \text{Law}(M_{q_+})$ are those introduced with (2.6) and (2.7).

Proposition 5.2. Let $\beta > 0$ and $h > 0$. Let $(G_d)_{d \geq 1}$ be either the tori $\mathbb{T}_L^d$ for a fixed $L \geq 3$, with $D = 2d$, or the hypercubes Q_d , with $D = d$. For every $\epsilon > 0$ there exist an integer $T \geq 1$, a probability measure μ_ϵ on $[-1, 1]$ and, for every d , a query history of depth T on G_d with center $m = m^{(d)}$, such that the following three statements hold.

(a) With $\kappa_x = 1 - m_x^2$, we have

$$\liminf_{d \rightarrow \infty} \frac{1}{n} \mathbb{E} \left[\beta H_{G_d}(m) + h \sum_x m_x + \sum_x \text{ent}(m_x) + \frac{\beta^2}{4} \langle \kappa, S\kappa \rangle \right] \geq \mathcal{P}(\beta, h) - \epsilon.$$

(b) We have $W_1(\mu_\epsilon, \mu_+) \leq \epsilon$, where W_1 is the Wasserstein distance of order one on $[-1, 1]$.

(c) Let $Y: [-1, 1] \times [0, \infty) \rightarrow \mathbb{R}$ be continuous, and let $L_Y \geq 0$ be such that $|Y(m', r) - Y(m', r')| \leq L_Y |r - r'|$ for all $m' \in [-1, 1]$ and $r, r' \geq 0$. Let $a, c, B \geq 0$. Then

$$\lim_{d \rightarrow \infty} \mathbb{E} \sup_{r \in a\mathbf{1} + cS[0, B]^V} \left| \frac{1}{n} \sum_x Y(m_x, r_x) - \frac{1}{n} \sum_x \int Y(m', r_x) \mu_\epsilon(\mathrm{d}m') \right| = 0.$$

Remark 5.3. We record three comments on the statement.

- The integer T and the measure μ_ϵ depend on β, h, ϵ , and, for tori, on L . The history is given by the same formulas for every d : polynomials, real coefficients, and a continuous output function, none of which depend on d .
- All terms in (a) are integrable, since $|H_{G_d}(m)| \leq \sum_e |J_e|$.
- The supremum in (c) is a random variable: since $a\mathbf{1} + cS[0, B]^V$ is compact and the expression inside the absolute value is continuous in r , the supremum may be taken over a countable dense subset.

The centers are produced by approximate message passing with full memory and deterministic Onsager coefficients. The Gaussian state evolution of message passing was first proved by Bolthausen [25] and by Bayati and Montanari [21]. Section 5.1 introduces this recursion and states its state evolution, Proposition 5.6. We deduce Proposition 5.6 from a theorem of Hachem [80] on a recursion whose Onsager coefficients are computed from the disorder. Section 5.2 states this theorem and a bound on $\|W\|$, Section 5.3 proves moment bounds and a coefficient-replacement estimate, and Section 5.4 proves Proposition 5.6. Section 5.5 computes the energy of the output. Sections 5.6 and 5.7 adapt to this recursion the two-phase algorithm of Sellke [125]. Its first phase is an iteration of the type used by Bolthausen [25], and its second phase is the incremental message passing of Montanari [103] and El Alaoui, Montanari, and Sellke [60], whose value is described by the stochastic control problem of [60, Section 4]. For the SK model without field, Montanari showed that the incremental algorithm constructs approximate solutions of the TAP equations for all large β , under the assumption that the support of the Parisi measure is an interval $[0, q_*]$ for all large β [103, Theorem 5]. Section 5.8 proves the homogenization estimate, and Section 5.9 completes the proof of Proposition 5.2.

The Onsager coefficients of our recursion are real numbers computed from the state evolution. Using Hachem's coefficients, which explicitly involve squared individual couplings, would require information beyond the earlier products and would not in general satisfy Definition 5.1. Deterministic Onsager coefficients for sparse variance profiles also appear in Theorems 2 and 4 of [80], which treat exact and approximate iterates, respectively, under a nondegeneracy assumption. At high temperature, [81, Theorem 5] treats an exact recursion with the additional sparsity condition $K_n \geq \log n$. These recursions have nonlinearities that act only on the last iterate; the incremental phase below needs the full history. We instead import the sampled-coefficient polynomial

result [80, Proposition 8] and prove the replacement by deterministic coefficients in Section 5.4.

Throughout this section, Z denotes a standard Gaussian variable, $\|\xi\|_p = (\mathbb{E}|\xi|^p)^{1/p}$ for a random variable ξ , and a function on $\mathbb{R}^k$ has *polynomial growth* if it is bounded by $C(1 + |z|^p)$ for some $C, p \geq 0$. Functions act on vectors coordinatewise: for $f: \mathbb{R}^t \rightarrow \mathbb{R}$ and $u^1, \dots, u^t \in \mathbb{R}^V$, the vector $f(u^1, \dots, u^t) \in \mathbb{R}^V$ has coordinates $f(u_x^1, \dots, u_x^t)$. We write $w \odot w'$ for the coordinatewise product.

5.1 Message passing with deterministic Onsager coefficients

Fix an integer $T \geq 1$, a constant $f_0 \in \mathbb{R}$ and polynomials $f_t: \mathbb{R}^t \rightarrow \mathbb{R}$ for $1 \leq t \leq T-1$. For a sequence $u = (u^1, u^2, \dots)$ we write $\bar{u}^t = (u^1, \dots, u^t)$, with $\bar{u}^0$ empty and $f_0(\bar{u}^0) = f_0$. The *state evolution* is the centered Gaussian vector $U = (U_1, \dots, U_T)$ whose covariance is given recursively by

$$\mathbb{E} U_{r+1} U_{s+1} = \mathbb{E} f_r(\bar{U}^r) f_s(\bar{U}^s) \quad (0 \leq r, s \leq T-1). \quad (5.1)$$

The recursion (5.1) defines the law of U , since the covariance of $\bar{U}^{t+1}$ is determined by the law of $\bar{U}^t$; degenerate covariances are allowed. The *Onsager coefficients* are the real numbers

$$b_{t,s} = \mathbb{E} \partial_s f_t(\bar{U}^t) \quad (1 \leq s \leq t \leq T-1). \quad (5.2)$$

On a graph $G \in \mathcal{G}_D$ we define $u^1, \dots, u^T \in \mathbb{R}^V$ by $u^1 = f_0 W \mathbf{1}$ and

$$u^{t+1} = W f_t(\bar{u}^t) - \sum_{s=1}^t b_{t,s} f_{s-1}(\bar{u}^{s-1}) \quad (1 \leq t \leq T-1), \quad (5.3)$$

where $f_0(\bar{u}^0) = f_0 \mathbf{1}$. The coefficients $b_{t,s}$ depend only on $f_0, \dots, f_{T-1}$; they do not depend on G or on the couplings. If $b_{t,s}$ is replaced by the empirical average $n^{-1} \sum_x \partial_s f_t(\bar{u}_x^t)$, then (5.3) is the message-passing iteration of [103, (2.1)], whose state evolution is (5.1).

Lemma 5.4. *Let $G \in \mathcal{G}_D$, let $1 \leq K \leq T$, and let $\theta: \mathbb{R}^K \rightarrow (-1, 1)$ be measurable. Set $x^t = f_{t-1}(\bar{u}^{t-1})$ for $1 \leq t \leq K$, and $x^{K+1} = \theta(\bar{u}^K)$. Then $x^1, \dots, x^{K+1}$, with U constant, form a query history of depth $K+1$ on G , and $m = \theta(\bar{u}^K)$ is a center of it.*

Proof. We show by induction on $t \leq K$ that $\bar{u}^t$ is measurable with respect to $\sigma(Wx^1, \dots, Wx^t)$. Since $x^1 = f_0 \mathbf{1}$, we have $u^1 = Wx^1$. If the claim holds for some $t < K$, then $x^{t+1} = f_t(\bar{u}^t)$ is measurable with respect to $\sigma(Wx^1, \dots, Wx^t)$, and (5.3) reads $u^{t+1} = Wx^{t+1} - \sum_{s \leq t} b_{t,s} x^s$, which is measurable with respect to $\sigma(Wx^1, \dots, Wx^{t+1})$. This completes the induction. In particular, x^1 is deterministic and, for $2 \leq t \leq K+1$, the vector x^t is a function of $\bar{u}^{t-1}$, which is measurable with respect to $\sigma(Wx^1, \dots, Wx^{t-1})$. Finally, $m = x^{K+1}$ takes values in $(-1, 1)^V$ and is $\mathcal{H}$ -measurable. $\square$

The next lemma is the covariance isometry behind the state evolution.

Lemma 5.5. *For $1 \leq k \leq T$, let $\mathfrak{F}_k$ be the linear span of $f_0, f_1(\bar{U}^1), \dots, f_{k-1}(\bar{U}^{k-1})$ in L^2 , and let $\mathfrak{U}_k$ be the span of $U_1, \dots, U_k$. The linear map $\iota_k: \mathfrak{F}_k \rightarrow \mathfrak{U}_k$ with $\iota_k(f_j(\bar{U}^j)) = U_{j+1}$ is well defined and is an isometry onto $\mathfrak{U}_k$, and ι_{k+1} extends ι_k for $k < T$. In particular, for $k < T$, the L^2 distance from U_{k+1} to $\mathfrak{U}_k$ equals the L^2 distance from $f_k(\bar{U}^k)$ to $\mathfrak{F}_k$, and ι_{k+1} maps every $\xi \in \mathfrak{F}_{k+1}$ orthogonal to $\mathfrak{F}_k$ to an element of $\mathfrak{U}_{k+1}$ orthogonal to $\mathfrak{U}_k$.*

Proof. By (5.1), for all real coefficients c_j, c'_j , we have

$$\mathbb{E} \left[\left(\sum_j c_j f_j(\bar{U}^j) \right) \left(\sum_j c'_j f_j(\bar{U}^j) \right) \right] = \mathbb{E} \left[\left(\sum_j c_j U_{j+1} \right) \left(\sum_j c'_j U_{j+1} \right) \right].$$

If $\sum_j c_j f_j(\bar{U}^j) = 0$ in L^2 , then the case $c' = c$ implies that $\sum_j c_j U_{j+1} = 0$ in L^2 . This means that ι_k is well defined, and by the display it preserves inner products. It is onto by definition, and ι_{k+1} extends ι_k . Since ι_{k+1} is an isometry of $\mathfrak{F}_{k+1}$ onto $\mathfrak{U}_{k+1}$ that maps $\mathfrak{F}_k$ onto $\mathfrak{U}_k$, it maps the orthogonal complement of $\mathfrak{F}_k$ in $\mathfrak{F}_{k+1}$ onto the orthogonal complement of $\mathfrak{U}_k$ in $\mathfrak{U}_{k+1}$. Applying this to the component of $f_k(\bar{U}^k)$ orthogonal to $\mathfrak{F}_k$, whose image is the component of U_{k+1} orthogonal to $\mathfrak{U}_k$, we obtain the statement about distances. $\square$

The following proposition is the state evolution of (5.3) on tori and hypercubes. It is proved in Section 5.4.

Proposition 5.6. *Let (G_d) be the tori $\mathbb{T}_L^d$ for a fixed $L \geq 3$ or the hypercubes Q_d , and run (5.3) on G_d with T, f_0 , and $(f_t)_{1 \leq t \leq T-1}$ fixed. Let $1 \leq t \leq T$.*

- (a) *For every polynomial $\psi: \mathbb{R}^t \rightarrow \mathbb{R}$, we have $\sup_d \max_{x \in V} \mathbb{E} |\psi(\bar{u}_x^t)| < \infty$.*
- (b) *For every continuous function $\psi: \mathbb{R}^t \rightarrow \mathbb{R}$ of polynomial growth and every choice of $x = x_d \in V$, as $d \rightarrow \infty$,*

$$\begin{aligned} \mathbb{E} \psi(\bar{u}_x^t) &\rightarrow \mathbb{E} \psi(\bar{U}^t), \\ \mathbb{E} \left| \frac{1}{n} \sum_{y \in V} \psi(\bar{u}_y^t) - \mathbb{E} \psi(\bar{U}^t) \right| &\rightarrow 0, \quad \mathbb{E} \left| \frac{1}{D} \sum_{y \sim x} \psi(\bar{u}_y^t) - \mathbb{E} \psi(\bar{U}^t) \right| \rightarrow 0. \end{aligned}$$

By Lemma 5.17 below, the quantities in (b) have laws that do not depend on x . In particular, the convergence in (b) is uniform in x .

5.2 Two imported results

We use the following special case of a theorem of Hachem [80, Proposition 8], which extends the polynomial state evolution of Bayati, Lelarge, and Montanari [22] to sparse

variance profiles. In Hachem's recursion, the Onsager term is computed from the squared entries of the matrix and from the current iterates. State evolution for matrices with a variance profile was also proved by Wang, Zhong, and Fan [140] and by Bao, Han, and Xu [13], for entries of variance $O(1/n)$. On tori and hypercubes, the nonzero entries of W have variance $1/D$, and $D/n \rightarrow 0$.

Theorem 5.7. *Let $(n_k)_{k \geq 1}$ and $(D_k)_{k \geq 1}$ be sequences of positive integers with $D_k \leq n_k$ and $D_k \rightarrow \infty$. For each k , let $S^{(k)}$ be a symmetric $n_k \times n_k$ matrix with nonnegative entries and zero diagonal, having at most D_k nonzero entries in each row, each of them at most D_k^{-1} . Let $(g_{xy})_{1 \leq x < y \leq n_k}$ be independent standard Gaussian variables, and let $W^{(k)}$ be the symmetric matrix with zero diagonal and entries $W_{xy} = (S_{xy}^{(k)})^{1/2} g_{xy}$ for $x < y$. Fix integers $q, T \geq 1$ and polynomial maps $\mathbf{f}_t = (\mathbf{f}_{t,1}, \dots, \mathbf{f}_{t,q}) : \mathbb{R}^q \rightarrow \mathbb{R}^q$ for $0 \leq t \leq T-1$, not depending on k , and set $\mathbf{f}_{-1} = 0$. Define $\mathbf{z}_x^t \in \mathbb{R}^q$, for $x \in [n_k]$ and $0 \leq t \leq T$, by $\mathbf{z}_x^{-1} = \mathbf{z}_x^0 = 0$ and*

$$\mathbf{z}_x^{t+1}(r) = \sum_y W_{xy} \mathbf{f}_{t,r}(\mathbf{z}_y^t) - \sum_{s=1}^q \mathbf{f}_{t-1,s}(\mathbf{z}_x^{t-1}) \sum_y W_{xy}^2 \partial_s \mathbf{f}_{t,r}(\mathbf{z}_y^t) \quad (r \in [q]). \quad (5.4)$$

Define positive-semidefinite $q \times q$ matrices by

$$\Xi_x^1 = \sum_y S_{xy}^{(k)} \mathbf{f}_0(0) \mathbf{f}_0(0)^\top, \quad \Xi_x^{t+1} = \sum_y S_{xy}^{(k)} \mathbb{E}[\mathbf{f}_t(\zeta_y^t) \mathbf{f}_t(\zeta_y^t)^\top],$$

where $\zeta_y^t \sim N(0, \Xi_y^t)$. Then for every $1 \leq t \leq T$ and every polynomial $\psi : \mathbb{R}^q \rightarrow \mathbb{R}$ the following hold.

- (i) We have $\sup_k \max_x \|\Xi_x^t\| < \infty$ and $\sup_k \max_x \mathbb{E}|\psi(\mathbf{z}_x^t)| < \infty$.
- (ii) For every $C > 0$ and every sequence of nonempty sets $I_k \subset [n_k]$ with $|I_k| \leq CD_k$, as $k \rightarrow \infty$,

$$\frac{1}{D_k} \sum_{x \in I_k} (\psi(\mathbf{z}_x^t) - \mathbb{E}\psi(\zeta_x^t)) \rightarrow 0 \quad \text{and} \quad \frac{1}{n_k} \sum_{x \in [n_k]} (\psi(\mathbf{z}_x^t) - \mathbb{E}\psi(\zeta_x^t)) \rightarrow 0$$

in probability.

In the notation of [80], Theorem 5.7 is Proposition 8, equations (19)–(21), with sparsity parameter $K_n = D_k$, constants $C_{\text{card}} = C_S = 1$, standard Gaussian entries (which satisfy Assumption 1 there), polynomial maps and test polynomials that do not depend on the vertex or on n , and initial vector zero. Assumption 2 of [80] requires only $K_n \rightarrow \infty$, $K_n \leq n$, the bound on the number of nonzero entries in each row, and the bound on the entries; in particular, no relation between K_n and $\log n$ is needed. Equation (20) there bounds the moments of monomials, which implies the moment bound in (i) for polynomials. The normalization in (ii) is D_k^{-1} for every admissible set, as in equation (21a) there; we use this local conclusion only for neighborhoods, which have exactly D_k elements. The theorem requires neither a nondegeneracy assumption on the covariances

nor a bound on $\|W\|$; these enter only Hachem's results for nonpolynomial activation functions.

The second imported result is a bound of Bandeira and van Handel [12, Theorem 1.1] on the norm of a Gaussian matrix with a variance profile.

Theorem 5.8. *Let X be the $n \times n$ symmetric matrix with $X_{xy} = b_{xy}g_{xy}$, where $\{g_{xy} : x \geq y\}$ are independent standard Gaussian variables and $\{b_{xy} : x \geq y\}$ are given real numbers, with $b_{xy} = b_{yx}$. Set $\sigma = \max_x (\sum_y b_{xy}^2)^{1/2}$ and $\sigma_* = \max_{x,y} |b_{xy}|$. Then for every $0 < \epsilon' \leq 1/2$,*

$$\mathbb{E}\|X\| \leq (1 + \epsilon') \left(2\sigma + \frac{6}{\sqrt{\log(1 + \epsilon')}} \sigma_* \sqrt{\log n} \right).$$

Lemma 5.9. *Let $G \in \mathcal{G}_D$ have $n \geq 2$ vertices. Then $\mathbb{E}\|W\| \leq 3 + 15(D^{-1} \log n)^{1/2}$. For the tori $\mathbb{T}_L^d$ with a fixed $L \geq 3$ and for the hypercubes Q_d , we have $\sup_d \mathbb{E}\|W\|^p < \infty$ for every $p \geq 1$.*

Proof. We apply Theorem 5.8 with $b_{xy} = S_{xy}^{1/2}$ and $b_{xx} = 0$, so that $\sigma = 1$ and $\sigma_* = D^{-1/2}$. With $\epsilon' = 1/2$ we have $9(\log(3/2))^{-1/2} \leq 15$, which gives the first bound. Next, we view $\|W\|$ as a function of $g = (g_e)_{e \in E}$. For two values g, g' ,

$$|\|W(g)\| - \|W(g')\|| \leq \|W(g) - W(g')\|_{\text{HS}} = \left(\frac{2}{D} \sum_{e \in E} (g_e - g'_e)^2 \right)^{1/2},$$

that is, $\|W\|$ is $(2/D)^{1/2}$ -Lipschitz. By Gaussian concentration (F4), we have $|\|W\| - \mathbb{E}\|W\||_p \leq C_p(2/D)^{1/2}$, where C_p depends only on p . For tori, $D^{-1} \log n = (\log L)/2$, and for hypercubes, $D^{-1} \log n = \log 2$. By the first bound, we have $\sup_d \mathbb{E}\|W\| < \infty$, and the moment bound follows from Minkowski's inequality. $\square$

5.3 Tame polynomials and coefficient replacement

In this subsection $G \in \mathcal{G}_D$ is arbitrary, and the entries $W_e = J_e$, $e \in E$, are independent with law $N(0, D^{-1})$. Moment bounds that do not depend on D come from a structural property: every monomial of an iterate at x is supported on a connected set of edges attached to x . There are at most CD^s such sets with s edges, and each edge carrying a nonzero even power contributes a factor at most C/D . This bookkeeping follows the moment method of [22]. The letter C denotes constants that depend only on the indicated quantities, never on D , on G , or on the chosen vertices; its value may change from line to line.

A *monomial* is a product $M = \prod_{e \in E} W_e^{k_e}$ with integers $k_e \geq 0$, finitely many of them nonzero. Its degree is $\deg M = \sum_e k_e$, and its support is $\text{supp } M = \{e : k_e \geq 1\}$. A set of edges is *connected* if the graph formed by these edges and their endpoints is connected.

Definition 5.10. Let $x \in V$, let $k \geq 0$ be an integer, and let $C_0 \geq 0$. A finite sum $Y = \sum_M c_M M$ of monomials is *tame rooted at x , of type (k, C_0)* , if every monomial M with $c_M \neq 0$ satisfies $\deg M \leq k$ and $|c_M| \leq C_0$, and $\text{supp } M$ is either empty or a connected set of edges at least one of which has x as an endpoint.

Lemma 5.11. *Let $x \in V$.*

- (a) *If Y and Y' are tame rooted at x , of types (k, C_0) and (k', C'_0) , then $Y + Y'$ is tame rooted at x , of type $(\max(k, k'), C_0 + C'_0)$, and YY' is tame rooted at x , of type $(k + k', 2^{k+k'} C_0 C'_0)$. A polynomial of degree at most m , with coefficients bounded by C_1 , in j tame polynomials rooted at x of type (k, C_0) is tame rooted at x , of a type depending only on k, C_0, m, C_1, j .*
- (b) *If Y_y is tame rooted at y , of type (k, C_0) , for every neighbor y of x , then $\sum_{y \sim x} W_{xy} Y_y$ and $\sum_{y \sim x} W_{xy}^2 Y_y$ are tame rooted at x , of types $(k + 1, (k + 1)C_0)$ and $(k + 2, (k + 2)C_0)$.*

Proof. (a) The statement on sums is immediate. The coefficient of M in YY' is $\sum_{M_1 M_2 = M} c_{M_1} c'_{M_2}$, and M has at most $\prod_e (k_e + 1) \leq 2^{\deg M}$ factorizations. Further, $\text{supp}(M_1 M_2) = \text{supp } M_1 \cup \text{supp } M_2$ is empty or connected with an edge at x , because both sets are, and each of them that is nonempty contains the vertex x . Constants are tame rooted at x , of type $(0, |c|)$, which gives the statement on polynomials.

(b) Let M_y be a monomial of Y_y . The support of $W_{xy} M_y$ is $\text{supp } M_y \cup \{xy\}$. It is connected and contains the edge xy at x : either $\text{supp } M_y$ is empty, or it is connected and contains an edge at y , which shares the vertex y with xy . A given monomial M arises as $W_{xy} M_y$ only for neighbors y with $xy \in \text{supp } M$, which are at most $\deg M \leq k + 1$ in number, and for each of them $M_y = M / W_{xy}$ is determined. The coefficient of M is therefore at most $(k + 1)C_0$. The same argument applies to $W_{xy}^2 M_y$. $\square$

Lemma 5.12. *Let $x \in V$ and let $v \geq 2$. The number of connected sets of edges that contain an edge at x and span exactly v vertices is at most $4^v 2^{v^2} D^{v-1}$.*

Proof. Such a set $\mathcal{S}$ has a spanning tree. We root it at x and list its vertices in breadth-first order. There are at most 4^v rooted plane trees with v vertices. Given the shape, there are at most D^{v-1} labelings, since the root is x and each other vertex is a neighbor of its parent, which is listed before it. Finally, $\mathcal{S}$ is a set of pairs of these v vertices, which leaves at most 2^{v^2} choices. $\square$

Lemma 5.13. *If $Y_1, \dots, Y_j$ are tame rooted at the same vertex x , of type (k, C_0) , then $|\mathbb{E}[Y_1 \cdots Y_j]| \leq C$, where C depends only on j, k , and C_0 . In particular, if Y is tame rooted at x , of type (k, C_0) , then $\mathbb{E}|Y|^p \leq C$ for every $p \geq 1$, with C depending only on p, k , and C_0 .*

Proof. We expand $\mathbb{E} \prod_i Y_i = \sum \prod_i c_{M_i} \mathbb{E} \prod_i M_i$ over j -tuples of monomials. For a tuple, let $\mathcal{S} = \bigcup_i \text{supp } M_i$ and $\mu_e = \sum_i k_e(M_i)$ for $e \in \mathcal{S}$. By independence, $\mathbb{E} \prod_i M_i = \prod_{e \in \mathcal{S}} \mathbb{E} W_e^{\mu_e}$.

This vanishes unless every μ_e is even, and then

$$\left| \mathbb{E} \prod_i M_i \right| = \prod_{e \in \mathcal{S}} (\mu_e - 1)!! D^{-\mu_e/2} \leq (jk)! D^{-|\mathcal{S}|},$$

since $\mu_e \geq 2$ for $e \in \mathcal{S}$. If $\mathcal{S}$ is empty, all M_i are constant, and the tuple contributes at most C_0^j . Otherwise $\mathcal{S}$ is connected and contains an edge at x , because each nonempty $\text{supp } M_i$ does and contains x ; it spans some number $2 \leq v \leq jk + 1$ of vertices, and $|\mathcal{S}| \geq v - 1$. Given $\mathcal{S}$, there are at most $(k + 1)^{j|\mathcal{S}|} \leq (k + 1)^{j^2 k}$ tuples of monomials of degree at most k supported in $\mathcal{S}$. By Lemma 5.12,

$$|\mathbb{E}[Y_1 \cdots Y_j]| \leq C_0^j \left(1 + \sum_{v=2}^{jk+1} 4^v 2^{v^2} D^{v-1} (k + 1)^{j^2 k} (jk)! D^{-(v-1)} \right),$$

which depends only on j, k , and C_0 . The second statement follows by taking $Y_1 = \cdots = Y_{2p'} = Y$ for an integer $p' \geq p/2$ and using Hölder's inequality. $\square$

The next lemma replaces squared couplings by their variances in an Onsager coefficient. For dense matrices, the corresponding fourth-moment bound is proved in [22, Appendix A.2]. Hachem stated a version for sparse variance profiles in [80, Lemma 21] and omitted its proof. We give a complete proof for tame polynomials, along the lines of [22].

Lemma 5.14. *Let $x \in V$, and for every neighbor y of x let Y_y be tame rooted at y , of type (k, C_0) . Then*

$$\mathbb{E} \left(\sum_{y \sim x} (W_{xy}^2 - D^{-1}) Y_y \right)^4 \leq \frac{C}{D^2},$$

where C depends only on k and C_0 .

Proof. Expanding the fourth power and each Y_y , we write the left side as a sum over 4-tuples $(y_a, M_a)_{a \leq 4}$, where $y_a \sim x$ and M_a is a monomial of Y_{y_a} , of

$$\prod_{a=1}^4 c_{M_a}^{(y_a)} \cdot \mathbb{E} \prod_{a=1}^4 (W_{xy_a}^2 - D^{-1}) M_a.$$

Fix a tuple. Let $e_a = xy_a$, let $\mathcal{S} = \bigcup_a (\text{supp } M_a \cup \{e_a\})$, and for $e \in \mathcal{S}$ let $\nu_e = |\{a : e_a = e\}|$ and $\mu_e = \sum_a k_e(M_a)$. By independence,

$$\mathbb{E} \prod_{a=1}^4 (W_{xy_a}^2 - D^{-1}) M_a = \prod_{e \in \mathcal{S}} w_e, \quad w_e = \mathbb{E}[W_e^{\mu_e} (W_e^2 - D^{-1})^{\nu_e}].$$

Writing $W_e = D^{-1/2} g_e$, we have $w_e = D^{-(\mu_e + 2\nu_e)/2} \mathbb{E}[g_e^{\mu_e} (g_e^2 - 1)^{\nu_e}]$. We claim that either some w_e vanishes, or

$$\left| \prod_{e \in \mathcal{S}} w_e \right| \leq C D^{-|\mathcal{S}|-2}. \quad (5.5)$$

Suppose that no w_e vanishes. Then every μ_e is even. If $\nu_e = 0$, then $\mu_e \geq 2$, since $e \in \mathcal{S}$, and $|w_e| \leq CD^{-1}$. If $\nu_e = 1$, then $\mu_e \geq 2$, since $\mathbb{E}[g_e^2 - 1] = 0$, and $|w_e| \leq CD^{-2}$ because $\mu_e + 2\nu_e \geq 4$. If $\nu_e \geq 2$, then $|w_e| \leq CD^{-\nu_e}$. In all cases with $\nu_e \geq 1$, we obtain $|w_e| \leq CD^{-1-\theta_e}$ with $\theta_e = \max(1, \nu_e - 1)$. Since $\sum_e \nu_e = 4$, either a single edge has $\nu_e = 4$ and $\theta_e = 3$, or at least two edges have $\nu_e \geq 1$; in both cases $\sum_{e:\nu_e \geq 1} \theta_e \geq 2$. This proves (5.5).

Each set $\text{supp } M_a \cup \{e_a\}$ is connected and contains the edge e_a at x , because M_a is tame rooted at y_a and $e_a = xy_a$. It follows that $\mathcal{S}$ is connected, contains an edge at x , has at most $4(k+1)$ edges, and spans some number $2 \leq v \leq 4k+5$ of vertices, with $|\mathcal{S}| \geq v-1$. Given $\mathcal{S}$, the number of tuples is at most $|\mathcal{S}|^4(k+1)^{4|\mathcal{S}|} \leq C$. By Lemma 5.12 and (5.5), the left side of the lemma is at most

$$C_0^4 \sum_{v=2}^{4k+5} 4^v 2^{v^2} D^{v-1} \cdot C \cdot C D^{-(v-1)-2} \leq \frac{C}{D^2}. \quad \square$$

5.4 Proof of the state evolution

We compare (5.3) with the recursion in which the Onsager coefficients are computed from the disorder: $\tilde{u}^1 = f_0 W \mathbf{1}$ and, for $1 \leq t \leq T-1$,

$$\tilde{u}^{t+1} = W f_t(\tilde{u}^t) - \sum_{s=1}^t \tilde{b}_{t,s} \odot f_{s-1}(\tilde{u}^{s-1}), \quad (\tilde{b}_{t,s})_x = \sum_y W_{xy}^2 \partial_s f_t(\tilde{u}_y^t). \quad (5.6)$$

This recursion is used only in the proof; it is not a query history.

Lemma 5.15. *Let $G \in \mathcal{G}_D$, $x \in V$, and $1 \leq t \leq T$. Every coordinate of $\tilde{u}_x^t$ and of $\tilde{u}_x^t$ is tame rooted at x , of a type depending only on t and $f_0, \dots, f_{T-1}$. For every polynomial $\psi: \mathbb{R}^{2t} \rightarrow \mathbb{R}$ and every $p \geq 1$, we have*

$$\sup_{D \geq 1} \sup_{G \in \mathcal{G}_D} \max_{x \in V} \mathbb{E} |\psi(\tilde{u}_x^t, \tilde{u}_x^t)|^p < \infty.$$

In particular, Proposition 5.6(a) holds.

Proof. We induct on t . The vectors $u_x^1 = \tilde{u}_x^1 = f_0 \sum_{y \sim x} W_{xy}$ are tame rooted at x , by Lemma 5.11(b) with $Y_y = f_0$. Suppose that the claim holds up to time t . By Lemma 5.11(a), $f_t(\tilde{u}_y^t)$, $f_t(\tilde{u}_y^t)$, and $\partial_s f_t(\tilde{u}_y^t)$ are tame rooted at y , and $f_{s-1}(\tilde{u}_x^{s-1})$ and $f_{s-1}(\tilde{u}_x^{s-1})$ are tame rooted at x . By Lemma 5.11(b), $\sum_{y \sim x} W_{xy} f_t(\tilde{u}_y^t)$, $\sum_{y \sim x} W_{xy} f_t(\tilde{u}_y^t)$, and $(\tilde{b}_{t,s})_x = \sum_{y \sim x} W_{xy}^2 \partial_s f_t(\tilde{u}_y^t)$ are tame rooted at x . Then Lemma 5.11(a), applied to (5.3) and (5.6), implies that u_x^{t+1} and $\tilde{u}_x^{t+1}$ are tame rooted at x . The types depend only on t , on the polynomials, and on the numbers $b_{t,s}$, which are determined by the polynomials. The moment bound follows from Lemma 5.11(a) and Lemma 5.13. $\square$

The next lemma identifies (5.6) as an instance of (5.4). The vector iterate stores the history of the scalar iterates; once a coordinate has appeared, the recursion recomputes the same value at every later step. The same reduction of a recursion with memory to a vector recursion appears in [103, Appendix A], where it is combined with [91, Theorem 1].

Lemma 5.16. *Let (G_d) be the tori $\mathbb{T}_L^d$ for a fixed $L \geq 3$ or the hypercubes Q_d . Let $1 \leq t \leq T$, let $\psi: \mathbb{R}^t \rightarrow \mathbb{R}$ be a polynomial, and let $x = x_d \in V$. Then, as $d \rightarrow \infty$,*

$$\frac{1}{D} \sum_{y \sim x} \psi(\tilde{u}_y^t) \rightarrow \mathbb{E} \psi(\bar{U}^t) \quad \text{and} \quad \frac{1}{n} \sum_{y \in V} \psi(\tilde{u}_y^t) \rightarrow \mathbb{E} \psi(\bar{U}^t)$$

in probability.

Proof. The graphs G_d satisfy the hypotheses of Theorem 5.7, with $k = d$, $n_k = n$, $D_k = D$, and $S^{(k)} = S$: each row of S has exactly D nonzero entries, all equal to D^{-1} , and $D \leq n$. Further, $W^{(k)}$ has the law of W , if we let g_{xy} for nonadjacent x, y be independent dummy variables. Let $q = T$, and for $0 \leq t' \leq T - 1$ define $\mathbf{f}_{t',r}: \mathbb{R}^T \rightarrow \mathbb{R}^T$ by

$$\mathbf{f}_{t',r}(z) = f_{r-1}(z_1, \dots, z_{r-1}) \quad (r \leq t' + 1), \quad \mathbf{f}_{t',r}(z) = 0 \quad (r > t' + 1).$$

We claim that $\mathbf{z}_x^{t'} = (\tilde{u}_x^1, \dots, \tilde{u}_x^{t'}, 0, \dots, 0)$ for $0 \leq t' \leq T$ and $x \in V$. This holds for $t' = 0$, and for $t' = 1$ since $\mathbf{f}_0(0) = (f_0, 0, \dots, 0)$. Suppose it holds at times $t' - 1$ and t' , for some $1 \leq t' \leq T - 1$. If $r > t' + 1$, then $\mathbf{f}_{t',r} = 0$, so $\mathbf{z}_x^{t'+1}(r) = 0$. Let $r \leq t' + 1$. Then $\mathbf{f}_{t',r}(\mathbf{z}_y^{t'}) = f_{r-1}(\tilde{u}_y^{r-1})$, because $r - 1 \leq t'$. The derivative $\partial_s \mathbf{f}_{t',r}$ equals $\partial_s f_{r-1}$ for $s \leq r - 1$ and vanishes for $s \geq r$, and for $s \leq r - 1 \leq t'$ we have $\mathbf{f}_{t'-1,s}(\mathbf{z}_x^{t'-1}) = f_{s-1}(\tilde{u}_x^{s-1})$. Inserting these identities into (5.4), we find

$$\mathbf{z}_x^{t'+1}(r) = \sum_y W_{xy} f_{r-1}(\tilde{u}_y^{r-1}) - \sum_{s=1}^{r-1} f_{s-1}(\tilde{u}_x^{s-1}) \sum_y W_{xy}^2 \partial_s f_{r-1}(\tilde{u}_y^{r-1}) = \tilde{u}_x^r,$$

by (5.6) at time $r - 1$ (for $r = 1$, by $\tilde{u}^1 = f_0 W \mathbf{1}$). This proves the claim.

Next, we claim that $\Xi_x^{t'}$ is the covariance matrix of $(U_1, \dots, U_{t'}, 0, \dots, 0)$ for all x . For $t' = 1$ this holds because $\sum_y S_{xy} = 1$ and $\mathbb{E} U_1^2 = f_0^2$. If it holds at time t' , then $\zeta_y^{t'}$ has the law of $(\bar{U}^{t'}, 0, \dots, 0)$ for every y , and, using $\sum_y S_{xy} = 1$ again, $\Xi_x^{t'+1}$ has entries $\mathbb{E} f_{r-1}(\bar{U}^{r-1}) f_{r'-1}(\bar{U}^{r'-1}) = \mathbb{E} U_r U_{r'}$ for $r, r' \leq t' + 1$, by (5.1), and zero entries otherwise.

Finally, we apply the local conclusion of Theorem 5.7(ii) at time t to the polynomial $z \mapsto \psi(z_1, \dots, z_t)$, with I_k the set of neighbors of x_d , which has D elements. For the vertex average, we use the separate global conclusion of that theorem. By the two claims, $\psi(\mathbf{z}_y^t) = \psi(\tilde{u}_y^t)$ and $\mathbb{E} \psi(\zeta_y^t) = \mathbb{E} \psi(\bar{U}^t)$ for all y . $\square$

We next use the symmetry of tori and hypercubes.

Lemma 5.17. *Let $G = G_d$, let $1 \leq t \leq T$, let $\psi: \mathbb{R}^{2t} \rightarrow \mathbb{R}$ be measurable, and set $\xi_y = \psi(\bar{u}_y^t, \tilde{u}_y^t)$ for $y \in V$. Then the laws of ξ_x , of $D^{-1} \sum_{y \sim x} \xi_y$, and of $(W\xi)_x$ do not depend on $x \in V$.*

Proof. Let φ be an automorphism of G , and let $W_{xy}^\varphi = W_{\varphi(x)\varphi(y)}$. Since φ permutes the edges and the couplings are independent and identically distributed, W^φ has the law of W . The recursions (5.3) and (5.6) use the same polynomials and the same real coefficients at every vertex. By induction on t , computing them from W^φ instead of W produces the vectors $(u_{\varphi(x)}^t)_x$ and $(\tilde{u}_{\varphi(x)}^t)_x$; for instance, $\sum_y W_{xy}^\varphi f_t(\bar{u}_y^t) = \sum_{y'} W_{\varphi(x)y'} f_t(\bar{u}_{y'}^t)$. Writing $\xi(W)$ for the vector ξ computed from W , we obtain $\xi(W^\varphi)_x = \xi(W)_{\varphi(x)}$,

$$\frac{1}{D} \sum_{y \sim x} \xi(W^\varphi)_y = \frac{1}{D} \sum_{y' \sim \varphi(x)} \xi(W)_{y'}, \quad (W^\varphi \xi(W^\varphi))_x = (W\xi(W))_{\varphi(x)}.$$

Since G is vertex-transitive, for all x, x' there exists an automorphism φ with $\varphi(x) = x'$, and the lemma follows. $\square$

Lemma 5.18. *Let (G_d) be the tori $\mathbb{T}_L^d$ for a fixed $L \geq 3$ or the hypercubes Q_d . For $1 \leq s \leq t \leq T-1$ and $x \in V$, $\lim_{d \rightarrow \infty} \mathbb{E} |(\tilde{b}_{t,s})_x - b_{t,s}|^4 = 0$.*

Proof. Since $S_{xy} = D^{-1}$ for $y \sim x$ and $W_{xy} = 0$ otherwise,

$$(\tilde{b}_{t,s})_x - b_{t,s} = \sum_{y \sim x} (W_{xy}^2 - D^{-1}) \partial_s f_t(\tilde{u}_y^t) + \left(\frac{1}{D} \sum_{y \sim x} \partial_s f_t(\tilde{u}_y^t) - b_{t,s} \right).$$

By Lemmas 5.11 and 5.15, the polynomial $\partial_s f_t(\tilde{u}_y^t)$ is tame rooted at y , of a type that does not depend on d or y . Then Lemma 5.14 implies that the fourth moment of the first term is at most CD^{-2} . The second term tends to zero in probability by Lemma 5.16 with $\psi = \partial_s f_t$, since $b_{t,s} = \mathbb{E} \partial_s f_t(\bar{U}^t)$. Its eighth moment is bounded uniformly in d , by Jensen's inequality and Lemma 5.15, so its fourth power is uniformly integrable, and its fourth moment tends to zero. The lemma follows from Minkowski's inequality. $\square$

Proof of Proposition 5.6. Part (a) is contained in Lemma 5.15. For the proof of (b), we use two elementary facts. First, if random variables ξ_d satisfy $\mathbb{E} \xi_d^2 \rightarrow 0$ and $\sup_d \mathbb{E} |\xi_d|^{2p} < \infty$ for every $p \geq 1$, then $\mathbb{E} |\xi_d|^p \leq (\mathbb{E} \xi_d^2)^{1/2} (\mathbb{E} |\xi_d|^{2p-2})^{1/2} \rightarrow 0$ for every $p \geq 1$. Second, for every polynomial $P: \mathbb{R}^t \rightarrow \mathbb{R}$ there exist C, k with $|P(z) - P(z')| \leq C|z - z'| (1 + |z|^k + |z'|^k)$.

Step 1: comparison. Let $w^t = u^t - \tilde{u}^t$. We show by induction on $t \leq T$ that $\mathbb{E}(w_x^t)^2 \rightarrow 0$ for all x . By Lemma 5.17, the law of w_x^t does not depend on x . We have $w^1 = 0$. Suppose that $\mathbb{E}(w_x^s)^2 \rightarrow 0$ for all $s \leq t$, where $t \leq T-1$. By Lemma 5.15 and the two facts above, for every polynomial $P: \mathbb{R}^t \rightarrow \mathbb{R}$ and every $p \geq 1$,

$$\mathbb{E} |P(\bar{u}_x^t) - P(\tilde{u}_x^t)|^p \leq C(\mathbb{E} |\bar{u}_x^t - \tilde{u}_x^t|^{2p})^{1/2} (\mathbb{E}(1 + |\bar{u}_x^t|^k + |\tilde{u}_x^t|^k)^{2p})^{1/2} \rightarrow 0. \quad (5.7)$$

By (5.3) and (5.6), $w^{t+1} = W\xi - \sum_{s=1}^t \rho^s$, where $\xi = f_t(\bar{u}^t) - f_t(\tilde{u}^t)$ and

$$\rho^s = b_{t,s}(f_{s-1}(\bar{u}^{s-1}) - f_{s-1}(\tilde{u}^{s-1})) + (b_{t,s}\mathbf{1} - \tilde{b}_{t,s}) \odot f_{s-1}(\tilde{u}^{s-1}).$$

By Lemma 5.17 and the Cauchy–Schwarz inequality,

$$\mathbb{E}(W\xi)_x^2 = \frac{1}{n}\mathbb{E}\|W\xi\|^2 \leq (\mathbb{E}\|W\|^4)^{1/2} \left(\mathbb{E}\left(\frac{1}{n}\sum_y \xi_y^2\right)^2 \right)^{1/2} \leq (\mathbb{E}\|W\|^4)^{1/2} (\mathbb{E}\xi_x^4)^{1/2},$$

where in the last inequality we used Jensen's inequality and that the law of ξ_y does not depend on y . The right side tends to zero, by Lemma 5.9 and (5.7). The first part of ρ_x^s tends to zero in L^2 by (5.7). For the second part, we have

$$\mathbb{E}[(b_{t,s} - (\tilde{b}_{t,s})_x)^2 f_{s-1}(\tilde{u}_x^{s-1})^2] \leq (\mathbb{E}|b_{t,s} - (\tilde{b}_{t,s})_x|^4)^{1/2} (\mathbb{E}f_{s-1}(\tilde{u}_x^{s-1})^4)^{1/2} \rightarrow 0,$$

by Lemmas 5.15 and 5.18. We conclude that $\mathbb{E}(w_x^{t+1})^2 \rightarrow 0$, which completes the induction. In particular, (5.7) holds for all $t \leq T$.

Step 2: polynomial test functions. Let $\psi: \mathbb{R}^t \rightarrow \mathbb{R}$ be a polynomial. By Lemma 5.17,

$$\mathbb{E}\left|\frac{1}{n}\sum_y \psi(\bar{u}_y^t) - \mathbb{E}\psi(\bar{U}^t)\right| \leq \mathbb{E}|\psi(\bar{u}_x^t) - \psi(\tilde{u}_x^t)| + \mathbb{E}\left|\frac{1}{n}\sum_y \psi(\tilde{u}_y^t) - \mathbb{E}\psi(\bar{U}^t)\right|.$$

The first term tends to zero by (5.7). The random variable in the second term tends to zero in probability by Lemma 5.16, and it tends to zero in L^1 because its second moment is bounded uniformly in d by Jensen's inequality and Lemma 5.15. The same argument applies to $D^{-1}\sum_{y \sim x} \psi(\bar{u}_y^t)$. Finally, by Lemma 5.17, $\mathbb{E}\psi(\bar{u}_x^t) = \mathbb{E}[n^{-1}\sum_y \psi(\bar{u}_y^t)] \rightarrow \mathbb{E}\psi(\bar{U}^t)$. This proves (b) for polynomials.

Step 3: continuous test functions. By Step 2, all moments of $\bar{u}_x^t$ converge to those of the Gaussian vector $\bar{U}^t$, and (F2) implies that $\bar{u}_x^t \rightarrow \bar{U}^t$ in law. Let φ be continuous with polynomial growth. By (a), we have $\sup_d \mathbb{E}\varphi(\bar{u}_x^t)^2 < \infty$. Then the variables $\varphi(\bar{u}_x^t)$ are uniformly integrable, and $\mathbb{E}\varphi(\bar{u}_x^t) \rightarrow \mathbb{E}\varphi(\bar{U}^t)$. This is the first assertion of (b). Now let ψ be continuous with polynomial growth, and let $\eta > 0$. By (F3), there exists a polynomial P with $\mathbb{E}|\psi(\bar{U}^t) - P(\bar{U}^t)| \leq \eta$. By Lemma 5.17,

$$\mathbb{E}\left|\frac{1}{n}\sum_y \psi(\bar{u}_y^t) - \mathbb{E}\psi(\bar{U}^t)\right| \leq \mathbb{E}|(\psi - P)(\bar{u}_x^t)| + \mathbb{E}\left|\frac{1}{n}\sum_y P(\bar{u}_y^t) - \mathbb{E}P(\bar{U}^t)\right| + \mathbb{E}|(P - \psi)(\bar{U}^t)|.$$

As $d \rightarrow \infty$, the first term tends to $\mathbb{E}|(\psi - P)(\bar{U}^t)| \leq \eta$, by the first assertion of (b) applied to $\varphi = |\psi - P|$, and the second term tends to zero by Step 2. The left side therefore has $\limsup$ at most 2η , and η was arbitrary. The neighbor average is treated in the same way, using $\mathbb{E}|D^{-1}\sum_{y \sim x} (\psi - P)(\bar{u}_y^t)| \leq \mathbb{E}|(\psi - P)(\bar{u}_x^t)|$. $\square$

5.5 Energy of the final message

The energy of the final polynomial message is determined by the next iterate and the Onsager term. A similar computation appears in the proof of [103, Lemma 2.3].

Lemma 5.19. *In the setting of Proposition 5.6, let $1 \leq K \leq T - 1$, and let $F = f_K(\bar{u}^K) \in \mathbb{R}^V$ and $\mathbf{F} = f_K(\bar{U}^K)$. Then*

$$\lim_{d \rightarrow \infty} \frac{1}{n} \mathbb{E} H_{G_d}(F) = \lim_{d \rightarrow \infty} \frac{1}{2n} \mathbb{E} \langle F, WF \rangle = \mathbb{E}[\mathbf{F} U_{K+1}].$$

Proof. Let $O = \sum_{s=1}^K b_{K,s} f_{s-1}(\bar{u}^{s-1})$. By (5.3), we have $WF = u^{K+1} + O$ and

$$\frac{1}{n} \langle F, WF \rangle = \frac{1}{n} \sum_x f_K(\bar{u}_x^K) \left(u_x^{K+1} + \sum_{s=1}^K b_{K,s} f_{s-1}(\bar{u}_x^{s-1}) \right),$$

which is the empirical average of a polynomial in $\bar{u}_x^{K+1}$. By Proposition 5.6(b), its expectation converges to

$$\mathbb{E}[\mathbf{F} U_{K+1}] + \sum_{s=1}^K b_{K,s} \mathbb{E}[\mathbf{F} f_{s-1}(\bar{U}^{s-1})].$$

By Gaussian integration by parts (F1), (5.1), and (5.2), we have

$$\mathbb{E}[U_{K+1} f_K(\bar{U}^K)] = \sum_{s=1}^K \text{Cov}(U_{K+1}, U_s) \mathbb{E} \partial_s f_K(\bar{U}^K) = \sum_{s=1}^K \mathbb{E}[f_K(\bar{U}^K) f_{s-1}(\bar{U}^{s-1})] b_{K,s}.$$

We conclude that the limit equals $2\mathbb{E}[\mathbf{F} U_{K+1}]$. Since $2H_{G_d}(F) = \langle F, WF \rangle$, this proves the lemma. $\square$

The Onsager term contributes exactly half of the energy. The center is a bounded function of the history close to F , and its energy is controlled by the norm of W .

Lemma 5.20. *In the setting of Lemma 5.19, let $\theta: \mathbb{R}^K \rightarrow [-1, 1]$ be continuous, and set $m = \theta(\bar{u}^K)$ and $\Theta = \theta(\bar{U}^K)$. Let $M_W = \sup_d (\mathbb{E} \|W\|^2)^{1/2}$, which is finite by Lemma 5.9. Then*

$$\limsup_{d \rightarrow \infty} \left| \frac{1}{n} \mathbb{E} H_{G_d}(m) - \mathbb{E}[\mathbf{F} U_{K+1}] \right| \leq \frac{M_W}{2} \|\Theta - \mathbf{F}\|_2 (\|\Theta\|_2 + \|\mathbf{F}\|_2).$$

Proof. Since $2(H_{G_d}(m) - H_{G_d}(F)) = \langle m - F, W(m + F) \rangle$, we have $2n^{-1} |H_{G_d}(m) - H_{G_d}(F)| \leq \|W\| \Xi$, where $\Xi = \|m - F\|_n (\|m\|_n + \|F\|_n)$. By the Cauchy-Schwarz inequality, $\mathbb{E}[\|W\| \Xi] \leq M_W (\mathbb{E} \Xi^2)^{1/2}$. By Lemma 5.19, it suffices to show that $\mathbb{E} \Xi^2 \rightarrow \|\Theta - \mathbf{F}\|_2^2 (\|\Theta\|_2 + \|\mathbf{F}\|_2)^2$.

Let $A_1 = \|m - F\|_n^2$, $A_2 = \|m\|_n^2$, and $A_3 = \|F\|_n^2$. Each A_i is an empirical average $n^{-1} \sum_y \psi_i(\bar{u}_y^K)$, where $\psi_1 = (\theta - f_K)^2$, $\psi_2 = \theta^2$, and $\psi_3 = f_K^2$ are continuous and have polynomial growth. By Proposition 5.6(b), A_1 , A_2 , and A_3 converge in L^1 and in probability to $\|\Theta - \mathbf{F}\|_2^2$, $\|\Theta\|_2^2$, and $\|\mathbf{F}\|_2^2$. Since $\Xi^2 = A_1(A_2^{1/2} + A_3^{1/2})^2$, it converges in probability to $\|\Theta - \mathbf{F}\|_2^2(\|\Theta\|_2 + \|\mathbf{F}\|_2)^2$. Further, for every $p \geq 1$, we have $A_i^p \leq n^{-1} \sum_y \psi_i(\bar{u}_y^K)^p$ by Jensen's inequality, and ψ_i^p is bounded by the polynomial $2^p(1 + f_K^2)^{\lceil p \rceil}$, because $|\theta| \leq 1$. Then Proposition 5.6(a) implies that $\sup_d \mathbb{E} A_i^p < \infty$, and by Hölder's inequality, $\sup_d \mathbb{E} \Xi^4 < \infty$. Then Ξ^2 is uniformly integrable, and $\mathbb{E} \Xi^2$ converges to the same limit. $\square$

5.6 The two-phase scheme and controls

We now choose the polynomials f_t so that the state evolution follows a stochastic control. We use the root-then-incremental architecture of Sellke [125], with the incremental nonlinearities of Montanari [103]: a fixed-point iteration first produces a field of variance $q_0 > 0$ and a prescribed function of it. Since (5.1) involves only the polynomials f_t , the computations of this subsection concern Gaussian vectors alone.

For $q_0 \in (0, 1)$ and $\tau \in (0, 1]$, let $N = N_{q_0, \tau}$ be the law of $(Y, \Delta_1, \Delta_2, \dots)$, where $Y \sim N(0, q_0)$ and the Δ_j are independent $N(0, \tau)$ variables, independent of Y . For a function p with $\mathbb{E} p(\sqrt{q_0} Z)^2 < \infty$ and $c \in [0, q_0]$, let $\phi_p(c) = \mathbb{E}[p(A)p(A')]$, where (A, A') is a centered Gaussian pair with $\text{Var } A = \text{Var } A' = q_0$ and $\text{Cov}(A, A') = c$. Writing $A = \sqrt{c} Z_0 + \sqrt{q_0 - c} Z_1$ and $A' = \sqrt{c} Z_0 + \sqrt{q_0 - c} Z_2$ with independent standard Gaussian Z_0, Z_1, Z_2 , we have $\phi_p(c) = \mathbb{E}[(\mathbb{E}[p(A) | Z_0])^2] \geq 0$.

Definition 5.21. Let $q_0 \in (0, 1)$, $\tau \in (0, 1]$, and let $K_2 \geq 0$ be an integer with $q_0 + K_2\tau \leq 1$. Let $K_1 \geq 1$ be an integer and p a nonconstant polynomial with $\mathbb{E} p(\sqrt{q_0} Z)^2 = q_0$. Let $P_j(y, z_1, \dots, z_j)$, $0 \leq j < K_2$, be polynomials with

$$\mathbb{E}_N P_j(Y, \Delta_1, \dots, \Delta_j)^2 = 1 \quad (0 \leq j < K_2). \quad (5.8)$$

The *two-phase scheme* is the recursion (5.3) with $T = K_1 + K_2 + 1$ and the following $f_0, \dots, f_{T-1}$. Set $K = K_1 + K_2$.

- *Root phase.* Let $f_0 = \sqrt{q_0}$ if $\mathbb{E} p(\sqrt{q_0} Z) \geq 0$ and $f_0 = -\sqrt{q_0}$ otherwise, and $f_k(\bar{u}^k) = p(u^k)$ for $1 \leq k \leq K_1$. Let $y = u^{K_1}$ and $\hat{F}_0 = p(y)$.
- *Innovations.* Let $\sum_{j \leq K_1} \gamma_j U_j$ be the orthogonal projection of U_{K_1+1} onto $\mathfrak{U}_{K_1}$, let $\tau_1 = \mathbb{E}(U_{K_1+1} - \sum_{j \leq K_1} \gamma_j U_j)^2$, and set $\Delta^1 = (\tau/\tau_1)^{1/2}(u^{K_1+1} - \sum_{j \leq K_1} \gamma_j u^j)$. For $r \geq 2$, let $\Delta^r = u^{K_1+r} - u^{K_1+r-1}$.
- *Incremental phase.* For $1 \leq r \leq K_2$,

$$f_{K_1+r}(\bar{u}^{K_1+r}) = \hat{F}_0 + \sum_{i=1}^r P_{i-1}(y, \Delta^1, \dots, \Delta^{i-1}) \Delta^i. \quad (5.9)$$

On the state-evolution side, we write $Y = U_{K_1}$, $\hat{F}_0 = p(Y)$, $\Delta_1 = (\tau/\tau_1)^{1/2}(U_{K_1+1} - \sum_{j \leq K_1} \gamma_j U_j)$, and $\Delta_r = U_{K_1+r} - U_{K_1+r-1}$ for $r \geq 2$.

The numbers γ_j and τ_1 depend only on $f_0, \dots, f_{K_1}$, and $\tau_1 > 0$ by the next lemma. Each Δ^i is a linear combination of $u^1, \dots, u^{K_1+i}$ with real coefficients, so each f_t is a polynomial in $\bar{u}^t$, and Proposition 5.6 and Lemma 5.4 apply. By part (b) of the next lemma, the notation $(Y, \Delta_1, \Delta_2, \dots)$ is consistent with the law $\mathbf{N}$.

Lemma 5.22. *In the setting of Definition 5.21, the following hold.*

- (a) *The variables $U_1, \dots, U_{K_1+1}$ are $N(0, q_0)$ variables with a nonsingular covariance matrix; in particular $\tau_1 > 0$. Further, $\varrho_k = \mathbb{E}U_k U_{k+1}$ satisfies $\varrho_1 = \sqrt{q_0} |\mathbb{E}p(\sqrt{q_0}Z)| < q_0$, $\varrho_{k+1} = \phi_p(\varrho_k)$ for $1 \leq k < K_1$, and $0 \leq \varrho_k \leq q_0$ for $k \leq K_1$. These numbers depend on p and q_0 but not on K_1 .*
- (b) *The variables $\Delta_1, \dots, \Delta_{K_2+1}$ are independent $N(0, \tau)$ variables, independent of $U_1, \dots, U_{K_1}$. In particular, $(Y, \Delta_1, \dots, \Delta_{K_2})$ has law $\mathbf{N}$.*
- (c) *The final message $\mathbf{F} = f_K(\bar{U}^K) = \hat{F}_0 + \sum_{i=1}^{K_2} P_{i-1} \Delta_i$ satisfies $\mathbb{E}\mathbf{F}^2 \leq 1$ and*

$$\mathbb{E}[\mathbf{F}U_{K+1}] = \varrho_{K_1} \mathbb{E}p'(Y) + (\tau\tau_1)^{1/2} \mathbb{E}P_0 + \tau \sum_{j=1}^{K_2-1} \mathbb{E}P_j,$$

where we abbreviate $P_j = P_j(Y, \Delta_1, \dots, \Delta_j)$, and the terms containing P_j are absent if $j \geq K_2$.

Proof. (a) We have $\text{Var } U_1 = f_0^2 = q_0$, and inductively $\text{Var } U_{k+1} = \mathbb{E}p(U_k)^2 = q_0$ for $k \leq K_1$. Next, $\varrho_1 = \mathbb{E}[f_0 p(U_1)] = \sqrt{q_0} |\mathbb{E}p(\sqrt{q_0}Z)|$ by the choice of the sign of f_0 . This is less than q_0 by the Cauchy–Schwarz inequality, which is strict because $p(\sqrt{q_0}Z)$ is not almost surely constant and $\mathbb{E}p(\sqrt{q_0}Z)^2 = q_0$. Further, $\varrho_{k+1} = \mathbb{E}[p(U_k)p(U_{k+1})] = \phi_p(\varrho_k)$, and $0 \leq \varrho_k \leq q_0$ by the Cauchy–Schwarz inequality and $\phi_p \geq 0$. For nonsingularity we induct on $k \leq K_1$; U_1 is nondegenerate. If $(U_1, \dots, U_k)$ is nondegenerate, then $U_k = \Lambda + \varsigma \xi$, where Λ is a linear combination of $U_1, \dots, U_{k-1}$, $\varsigma > 0$, and $\xi \sim N(0, 1)$ is independent of $U_1, \dots, U_{k-1}$. Conditionally on $U_1, \dots, U_{k-1}$, the message $p(U_k)$ is a nonconstant polynomial of a nondegenerate Gaussian variable, and it is therefore not almost surely equal to a function of $U_1, \dots, U_{k-1}$. Since every element of $\mathfrak{F}_k$ is such a function, we have $p(U_k) \notin \mathfrak{F}_k$. By Lemma 5.5, the variable U_{k+1} has a nonzero component orthogonal to $\mathfrak{U}_k$, which implies that $(U_1, \dots, U_{k+1})$ is nondegenerate. At $k = K_1$, the squared norm of this component is τ_1 .

(b) Let $\mathcal{G}$ be the σ -algebra generated by $U_1, \dots, U_{K_1}$. Then $\Delta_1 \sim N(0, \tau)$ is independent of $\mathcal{G}$, since the residual $U_{K_1+1} - \sum_j \gamma_j U_j$ is Gaussian, uncorrelated with $U_1, \dots, U_{K_1}$, and of variance τ_1 . Now let $1 \leq r \leq K_2$, and assume that $\Delta_1, \dots, \Delta_r$ are independent $N(0, \tau)$ variables, independent of $\mathcal{G}$. Let $\mathcal{G}_r$ be the σ -algebra generated by $\mathcal{G}$ and $\Delta_1, \dots, \Delta_{r-1}$. By (5.9), $f_{K_1+r} - f_{K_1+r-1} = P_{r-1} \Delta_r$, where $f_{K_1} = \hat{F}_0$. Every message f_j with $j < K_1 + r$ is $\mathcal{G}_r$ -measurable, and so is P_{r-1} , whereas Δ_r is centered and independent of $\mathcal{G}_r$. Then

$P_{r-1}\Delta_r$ is orthogonal to $\mathfrak{F}_{K_1+r}$, and by Lemma 5.5 its image $\Delta_{r+1} = U_{K_1+r+1} - U_{K_1+r}$ is orthogonal to $\mathfrak{U}_{K_1+r}$. The variables $U_1, \dots, U_{K_1+r}$ generate $\mathcal{G}$ and $\Delta_1, \dots, \Delta_r$, and by joint Gaussianity Δ_{r+1} is independent of them. Its variance is $\mathbb{E}[P_{r-1}^2 \Delta_r^2] = \tau \mathbb{E}P_{r-1}^2 = \tau$, by the induction hypothesis and (5.8). This completes the induction.

(c) By (b), a product $P_{i-1}\Delta_i$ is orthogonal to every $\mathcal{G}_i$ -measurable square-integrable variable, which includes $\hat{F}_0$, every variable of $\mathcal{G}$, $\Delta_1, \dots, \Delta_{i-1}$, and $P_{j-1}\Delta_j$ for $j < i$, and it is orthogonal to every Δ_r with $r > i$. Further, $\mathbb{E}[P_{i-1}\Delta_i^2] = \tau \mathbb{E}P_{i-1}$. This gives $\mathbb{E}F^2 = \mathbb{E}\hat{F}_0^2 + K_2\tau \leq q_0 + (1 - q_0) = 1$. For the correlation, we write $U_{K+1} = U_{K_1+1} + \sum_{r=2}^{K_2+1} \Delta_r$ and $U_{K_1+1} = \sum_j \gamma_j U_j + (\tau_1/\tau)^{1/2} \Delta_1$. The term $\hat{F}_0$ correlates only with U_{K_1+1} ; the term $P_0\Delta_1$ correlates only with $(\tau_1/\tau)^{1/2} \Delta_1$, giving $(\tau\tau_1)^{1/2} \mathbb{E}P_0$; and for $i \geq 2$ the term $P_{i-1}\Delta_i$ correlates only with Δ_i , giving $\tau \mathbb{E}P_{i-1}$. Finally, (F1) applied to the pair (U_{K_1}, U_{K_1+1}) gives $\mathbb{E}[\hat{F}_0 U_{K_1+1}] = \mathbb{E}[p(U_{K_1}) U_{K_1+1}] = \varrho_{K_1} \mathbb{E}p'(U_{K_1})$. $\square$

A control specifies the ideal root message, the ideal coefficients, of which the P_j are polynomial approximations, and a bounded output.

Definition 5.23. Let $q_0 \in (0, 1)$, $\tau \in (0, 1]$, and $K_2 \geq 0$ be as in Definition 5.21. A control of step τ and length K_2 with root variance q_0 consists of the following data.

- (a) A root map: a nonconstant C^1 function $f: \mathbb{R} \rightarrow [-1, 1]$ with bounded derivative, such that $\mathbb{E}f(\sqrt{q_0}Z)^2 = q_0$ and $\phi_f(c) > c$ for every $c \in [0, q_0]$.
- (b) Random variables $G_j = g_j(Y, \Delta_1, \dots, \Delta_j)$, $0 \leq j < K_2$, with g_j measurable and $\mathbb{E}_N G_j^2 = 1$.
- (c) A continuous output function $\Theta: \mathbb{R}^{1+K_2} \rightarrow (-1, 1)$.

Let $F_0 = f(Y)$. The stochastic integral, value, and defect of the control are

$$F = F_0 + \sum_{j=0}^{K_2-1} G_j \Delta_{j+1}, \quad V = \mathbb{E}[Y F_0] + \tau \sum_{j=0}^{K_2-1} \mathbb{E}G_j, \quad \vartheta = \|F - \Theta\|_2, \quad (5.10)$$

where Θ stands for $\Theta(Y, \Delta_1, \dots, \Delta_{K_2})$ and all expectations are under N . We write μ_Θ for the law of Θ under N .

If $K_2 = 0$, the control consists of the root map and the output only, with $F = F_0$ and $V = \mathbb{E}[Y F_0]$.

The algorithm. Fix a control and accuracy parameters: an integer $\ell \geq 2$ and a number $\eta \in (0, q_0)$.

- (i) *Root polynomial.* By dominated convergence, the function ϕ_f is continuous on $[0, q_0]$, and $c_\eta = \min_{c \in [0, q_0 - \eta]} (\phi_f(c) - c) > 0$. Since f is continuous and nonconstant and $N(0, q_0)$ has full support, $\|f - \mathbb{E}f\|_2 > 0$. By (F3), there exists a polynomial $Q_\star$ with

$$\|Q_\star - f\|_2 \leq \min \left\{ \frac{1}{\ell}, \frac{c_\eta}{8\sqrt{q_0}}, \frac{\|f - \mathbb{E}f\|_2}{3} \right\}$$

in $L^2(N(0, q_0))$. This gives $\|Q_\star\|_2 \geq \|f\|_2 - \|Q_\star - f\|_2 \geq 2\sqrt{q_0}/3 > 0$, so we set $p = \sqrt{q_0} Q_\star / \|Q_\star\|_2$. Let K_1 be the least $k \geq 1$ with $q_k > q_0 - \eta$, where q_k is given by Lemma 5.22(a).

- (ii) *Coefficient polynomials.* By (F3), there exist polynomials Q_j in $(y, z_1, \dots, z_j)$ with $\|Q_j - G_j\|_2 \leq 1/\ell$ under N . Since $\|Q_j\|_2 \geq 1 - 1/\ell \geq 1/2$, we set $P_j = Q_j / \|Q_j\|_2$.
- (iii) *Center.* We run the two-phase scheme of Definition 5.21 with p , K_1 , and (P_j) on G_d , and we set $m = \Theta(y, \Delta^1, \dots, \Delta^{K_2}) \in (-1, 1)^V$. By Lemma 5.4, applied with θ the composition of Θ with the linear map $\bar{u}^K \mapsto (y, \Delta^1, \dots, \Delta^{K_2})$, the vectors $x^t = f_{t-1}(\bar{u}^{t-1})$, $1 \leq t \leq K$, and $x^{K+1} = m$ form a query history of depth $K+1$ with center m .

The update rule depends only on the control and the accuracy parameters, not on d .

Lemma 5.24. *In the algorithm, p is a nonconstant polynomial with $\mathbb{E}p(\sqrt{q_0}Z)^2 = q_0$ and $\|p - f\|_2 \leq 2/\ell$, the index K_1 exists, $K_1 \leq 1 + 2q_0/c_\eta$, and $q_0 - \eta < q_{K_1} \leq q_0$. For every $j < K_2$, the polynomial P_j satisfies (5.8) and $\|P_j - G_j\|_2 \leq 2/\ell$.*

Proof. Since $|\|Q_\star\|_2 - \sqrt{q_0}| \leq \|Q_\star - f\|_2$, we have $\|p - Q_\star\|_2 = |\sqrt{q_0} - \|Q_\star\|_2| \leq \|Q_\star - f\|_2$ and $\|p - f\|_2 \leq 2\|Q_\star - f\|_2 \leq 2/\ell$. Further, p is nonconstant, since $\|(p - \mathbb{E}p) - (f - \mathbb{E}f)\|_2 \leq \|p - f\|_2 \leq (2/3)\|f - \mathbb{E}f\|_2$, and $\mathbb{E}p(\sqrt{q_0}Z)^2 = q_0$ by construction. By the Cauchy–Schwarz inequality, for $c \in [0, q_0]$ we have

$$|\phi_p(c) - \phi_f(c)| \leq \|p - f\|_2(\|p\|_2 + \|f\|_2) = 2\sqrt{q_0}\|p - f\|_2 \leq 4\sqrt{q_0}\|Q_\star - f\|_2 \leq \frac{c_\eta}{2},$$

which implies that $\phi_p(c) \geq c + c_\eta/2$ on $[0, q_0 - \eta]$. By Lemma 5.22(a), $q_1 \in [0, q_0)$ and $q_{k+1} = \phi_p(q_k)$, and these numbers do not depend on K_1 . Then q_k increases by at least $c_\eta/2$ as long as $q_k \leq q_0 - \eta$, and the least k with $q_k > q_0 - \eta$ is at most $1 + 2q_0/c_\eta$, and $q_{K_1} \leq q_0$ by Lemma 5.22(a). For the coefficients, $\|Q_j\|_2 \geq 1 - 1/\ell > 0$, and $\|P_j - G_j\|_2 \leq |1 - \|Q_j\|_2| + \|Q_j - G_j\|_2 \leq 2/\ell$. $\square$

Proposition 5.25. *Let (G_d) be the tori $\mathbb{T}_L^d$ for a fixed $L \geq 3$ or the hypercubes Q_d , and let M_W be as in Lemma 5.20. Fix a control with root variance q_0 , step τ , length K_2 , value V , and defect ϑ , and fix accuracy parameters ℓ and η . Let m be the center produced by the algorithm on G_d , and set $\kappa_x = 1 - m_x^2$. Then the following hold.*

- (a) *As $d \rightarrow \infty$, we have $n^{-1}\mathbb{E}\sum_x m_x \rightarrow \mathbb{E}\Theta$, $n^{-1}\mathbb{E}\sum_x \text{ent}(m_x) \rightarrow \mathbb{E}\text{ent}(\Theta)$, and $n^{-1}\mathbb{E}\langle \kappa, S\kappa \rangle \rightarrow (1 - \mathbb{E}\Theta^2)^2$, where Θ has law μ_Θ .*
- (b) *We have*

$$\limsup_{d \rightarrow \infty} \left| \frac{1}{n} \mathbb{E} H_{G_d}(m) - V \right| \leq 2\sqrt{\tau} \mathbf{1}_{\{K_2 \geq 1\}} + \eta + \frac{4}{\ell} + M_W \left(\vartheta + \frac{4}{\ell} \right).$$

- (c) *For every $x \in V$, the law of m_x does not depend on x , and it converges weakly to μ_Θ as $d \rightarrow \infty$.*

Proof. Let θ be as in Step (iii) of the algorithm, so that $m = \theta(\bar{u}^K)$. By Lemma 5.22(b), $\theta(\bar{U}^K) = \Theta(Y, \Delta_1, \dots, \Delta_{K_2})$ has law μ_Θ , and we denote it by Θ .

(a) and (c). Since the functions θ and $\text{ent} \circ \theta$ are bounded and continuous, the first two limits follow from Proposition 5.6(b). Part (c) follows in the same way, because $\varphi \circ \theta$ is bounded and continuous for every bounded continuous φ ; the law of m_x does not depend on x by Lemma 5.17. Let $\bar{\kappa} = \mathbb{E}(1 - \Theta^2)$. We have $(S\kappa)_x = D^{-1} \sum_{y \sim x} (1 - \theta(\bar{u}_y^K)^2)$, and Proposition 5.6(b) gives $\mathbb{E}|(S\kappa)_x - \bar{\kappa}| \rightarrow 0$. Since $0 \leq \kappa_x \leq 1$, we have

$$\left| \frac{1}{n} \mathbb{E} \langle \kappa, S\kappa \rangle - \bar{\kappa} \frac{1}{n} \mathbb{E} \sum_x \kappa_x \right| \leq \frac{1}{n} \sum_x \mathbb{E}|(S\kappa)_x - \bar{\kappa}| \rightarrow 0,$$

and $n^{-1} \mathbb{E} \sum_x \kappa_x \rightarrow \bar{\kappa}$. This proves (a).

(b) Let $\mathbf{F} = f_K(\bar{U}^K) = \hat{F}_0 + \sum_{j < K_2} P_j \Delta_{j+1}$, a function of $(Y, \Delta_1, \dots, \Delta_{K_2})$, whose law is $\mathbb{N}$ by Lemma 5.22(b). We compare $\mathbb{E}[\mathbf{F}U_{K+1}]$, given by Lemma 5.22(c), with V . The errors have three sources.

- *Root message.* By Stein's identity, we have $q_0 \mathbb{E}p'(Y) = \mathbb{E}[Yp(Y)]$, and therefore $|\mathbb{E}p'(Y)| \leq \|Y\|_2 \|p(Y)\|_2 / q_0 = 1$. Using this bound and Lemma 5.24, we obtain

$$|q_{K_1} \mathbb{E}p'(Y) - \mathbb{E}[Yf(Y)]| \leq (q_0 - q_{K_1}) |\mathbb{E}p'(Y)| + |\mathbb{E}[Y(p - f)(Y)]| \leq \eta + \frac{2}{\ell}.$$

- *First innovation* ($K_2 \geq 1$). The algorithm weighs $\mathbb{E}P_0$ by $(\tau\tau_1)^{1/2}$ instead of τ . Since $\tau_1 \leq \mathbb{E}U_{K_1+1}^2 = q_0 \leq 1$ and $|\mathbb{E}P_0|, |\mathbb{E}G_0| \leq 1$, we have $|(\tau\tau_1)^{1/2} \mathbb{E}P_0 - \tau \mathbb{E}G_0| \leq \sqrt{\tau} + \tau \leq 2\sqrt{\tau}$.
- *Coefficients.* By Lemma 5.24 and $K_2\tau \leq 1$, $\tau \sum_{1 \leq j < K_2} |\mathbb{E}P_j - \mathbb{E}G_j| \leq \tau \sum_j \|P_j - G_j\|_2 \leq 2/\ell$.

Combining the three bounds, we obtain $|\mathbb{E}[\mathbf{F}U_{K+1}] - V| \leq 2\sqrt{\tau} \mathbf{1}_{\{K_2 \geq 1\}} + \eta + 4/\ell$. Next, $\mathbf{F} - F = (\hat{F}_0 - F_0) + \sum_j (P_j - G_j) \Delta_{j+1}$. The terms of the sum are orthogonal, with norms $\sqrt{\tau} \|P_j - G_j\|_2$. By Lemma 5.24 and $K_2\tau \leq 1$, we have $\|\mathbf{F} - F\|_2 \leq 4/\ell$ and $\|\Theta - \mathbf{F}\|_2 \leq \vartheta + 4/\ell$. By Lemma 5.20, with $\|\Theta\|_2 \leq 1$ and $\|\mathbf{F}\|_2 \leq 1$ by Lemma 5.22(c),

$$\limsup_{d \rightarrow \infty} \left| \frac{1}{n} \mathbb{E}H_{G_d}(m) - \mathbb{E}[\mathbf{F}U_{K+1}] \right| \leq M_W \left(\vartheta + \frac{4}{\ell} \right).$$

Combining the two bounds, we obtain (b). $\square$

5.7 The Parisi control

In this subsection, $\beta > 0$ and $h > 0$. By Theorem 2.2, we have $\text{supp } \nu = [q_-, q_+]$ with $0 < q_- \leq q_+ < 1$. We use the following facts, from (P1), (P2), and Lemmas 2.5 and 2.6. We write $\nu(t) = \nu([0, t])$.

- $|\partial_x \Phi| \leq 1$, $0 < \partial_x^2 \Phi \leq 1$, and $|\partial_x^3 \Phi| \leq 4$; hence $|v| \leq \beta^2$, $0 \leq \partial_x v \leq \beta^2$, $0 < a \leq \beta$, and $|\partial_x a| \leq 4\beta$.
- For $s, t \in [0, 1]$ and $x \in \mathbb{R}$, $|\partial_x \Phi(t, x) - \partial_x \Phi(s, x)| \leq 3\beta^2|t - s|$ and $|a(t, x) - a(s, x)| \leq 4\sqrt{3}\beta^2|t - s|^{1/2}$.
- For $0 \leq s \leq t \leq 1$, we have $M_t - M_s = \int_s^t a(r, X_r) dB_r$ and

$$\partial_x^2 \Phi(t, X_t) - \partial_x^2 \Phi(s, X_s) = - \int_s^t \beta^2 v(r) (\partial_x^2 \Phi(r, X_r))^2 dr + \int_s^t \beta \partial_x^3 \Phi(r, X_r) dB_r.$$

- $\mathbb{E}M_t^2 = t$ for $t \in [q_-, q_+]$; if $q_- < q_+$, then $\mathbb{E}a(t, X_t)^2 = 1$ for $t \in [q_-, q_+]$; and $M_{q_+} = \tanh X_{q_+}$.
- The function $f(y) = \partial_x \Phi(q_-, h + \beta y)$ is a root map with root variance q_- in the sense of Definition 5.23(a), and $f(B_{q_-}) = M_{q_-}$. Indeed, the function ϕ_f defined before Lemma 2.6 is the function ϕ_p of Section 5.6 with $p = f$, and Lemma 2.6(a) gives the required properties.

Since $v(t) = 0$ for $t < q_-$, the drift v vanishes on $[0, q_-) \times \mathbb{R}$. Then $X_t = h + \beta B_t$ for $t \leq q_-$, and by the second martingale identity, $\partial_x^2 \Phi(t, X_t)$ is a martingale on $[0, q_-]$. In particular, $\mathbb{E}a(t, X_t)$ is constant on $[0, q_-]$. Recall that $\mathcal{E}(\beta, h) = \int_0^{q_+} \mathbb{E}a(t, X_t) dt$.

The control. If $q_- < q_+$, let $K_2 \geq 1$ be an integer, and let $\tau = (q_+ - q_-)/K_2$ and $t_j = q_- + j\tau$ for $0 \leq j \leq K_2$, so that $t_{K_2} = q_+$. We realize the law N with $q_0 = q_-$ by $Y = B_{q_-}$ and $\Delta_j = B_{t_j} - B_{t_{j-1}}$. We define the Euler scheme for X driven by these variables, the normalized coefficients, and the output by

$$\begin{aligned} \hat{X}_0 &= h + \beta Y, & \hat{X}_{j+1} &= \hat{X}_j + v(t_j, \hat{X}_j) \tau + \beta \Delta_{j+1}, \\ \lambda_j &= \|a(t_j, \hat{X}_j)\|_2, & G_j &= \frac{a(t_j, \hat{X}_j)}{\lambda_j}, & \Theta &= \tanh \hat{X}_{K_2}, \end{aligned} \quad (5.11)$$

with the root map $f(y) = \partial_x \Phi(q_-, h + \beta y)$. As in [103, (2.37)], the coefficients are normalized to have unit second moment under the law of the Euler scheme. If $q_- = q_+$, we take $K_2 = 0$, the root map f , and the output $\Theta(y) = f(y)$.

Proposition 5.26. *Let $\beta > 0$ and $h > 0$. There exists $C < \infty$, depending only on β and h , such that the following hold.*

- If $q_- = q_+$, the data above form a control with root variance q_- and length 0, and $V = \mathcal{E}(\beta, h)$, $\vartheta = 0$, and $\mu_\Theta = \mu_+$.
- If $q_- < q_+$, then for every integer $K_2 \geq 1$ the data (5.11) form a control with root variance q_- , step $\tau = (q_+ - q_-)/K_2$, and length K_2 , and

$$|V - \mathcal{E}(\beta, h)| \leq C\sqrt{\tau}, \quad \vartheta \leq C\sqrt{\tau}, \quad \|\Theta - M_{q_+}\|_2 \leq C\sqrt{\tau}.$$

In particular, $W_1(\mu_\Theta, \mu_+) \leq C\sqrt{\tau}$.

Proof. (a) Since $q_- = q_+$, we have $\partial_x \Phi(q_-, x) = \tanh x$ and $f(y) = \tanh(h + \beta y)$, and $\Theta = f$ is continuous with values in $(-1, 1)$. The data therefore form a control, with $F = F_0 = f(Y) = \Theta$ and $\vartheta = 0$. Further, $\Theta = f(B_{q_-}) = M_{q_-} = M_{q_+}$. In particular, $\mu_\Theta = \mu_+$. By Stein's identity, $V = \mathbb{E}[Yf(Y)] = q_- \mathbb{E}f'(Y) = q_- \mathbb{E}a(q_-, X_{q_-})$, and since $\mathbb{E}a(t, X_t)$ is constant on $[0, q_-]$, this equals $\mathcal{E}(\beta, h)$.

(b) Let C denote constants depending only on β and h . Let $\hat{e}_j = \hat{X}_j - X_{t_j}$. By the facts listed above, for $r \in [t_j, t_{j+1}]$ and $x \in \mathbb{R}$, we have

$$|a(t_j, x) - a(r, x)| \leq 4\sqrt{3}\beta^2\sqrt{\tau}, \quad |v(t_j, x) - v(r, x)| \leq 3\beta^4\tau + \beta^2(v(t_{j+1}) - v(t_j)),$$

the second because $v = \beta^2\nu\partial_x\Phi$, ν is nondecreasing, and $|\partial_x\Phi| \leq 1$. Also $\|X_r - X_{t_j}\|_2 \leq \beta^2\tau + \beta\sqrt{\tau}$, since $|v| \leq \beta^2$. The proof has five steps.

Euler error. We have $\hat{e}_0 = 0$, since $X_{q_-} = h + \beta B_{q_-}$. The Brownian parts of $\hat{X}$ and X cancel, which gives $\hat{e}_{j+1} = \hat{e}_j + \int_{t_j}^{t_{j+1}} (v(t_j, \hat{X}_j) - v(r, X_r)) dr$. We split the integrand into

$$[v(t_j, \hat{X}_j) - v(t_j, X_{t_j})] + [v(t_j, X_{t_j}) - v(t_j, X_r)] + [v(t_j, X_r) - v(r, X_r)].$$

The first two brackets are at most $\beta^2|\hat{e}_j|$ and $\beta^2|X_{t_j} - X_r|$ in absolute value. The third is at most $3\beta^4\tau + \beta^2(v(t_{j+1}) - v(t_j))$ in absolute value, and the sum over j of its ν -parts telescopes: $\sum_j \tau(v(t_{j+1}) - v(t_j)) \leq \tau$. This uses only the monotonicity of ν , and atoms of ν cause no difficulty. Taking L^2 norms, summing over $j < k$, and using $K_2\tau \leq 1$ and $\tau \leq 1$, we obtain $\|\hat{e}_k\|_2 \leq \beta^2\tau \sum_{j < k} \|\hat{e}_j\|_2 + C\sqrt{\tau}$, and the discrete Grönwall inequality gives $\max_{k \leq K_2} \|\hat{e}_k\|_2 \leq C\sqrt{\tau}(1 + \beta^2\tau)^{K_2} \leq C\sqrt{\tau}$.

Normalization. For $j < K_2$, we have $t_j \in [q_-, q_+)$ and $\mathbb{E}a(t_j, X_{t_j})^2 = 1$. Then

$$|\lambda_j^2 - 1| \leq \mathbb{E}|a(t_j, \hat{X}_j)^2 - a(t_j, X_{t_j})^2| \leq 2\beta \cdot 4\beta \|\hat{e}_j\|_1 \leq C\sqrt{\tau},$$

and $|\lambda_j - 1| = |\lambda_j^2 - 1|/(\lambda_j + 1) \leq C\sqrt{\tau}$. Since $a > 0$, we have $\lambda_j > 0$. Then G_j is well defined and $\mathbb{E}G_j^2 = 1$.

Admissibility. By the last of the listed facts, f is a root map. Each G_j is a continuous function of $(Y, \Delta_1, \dots, \Delta_j)$, since $x \mapsto x + v(t_j, x)\tau$ is Lipschitz. The output Θ is a continuous function of $(Y, \Delta_1, \dots, \Delta_{K_2})$ with values in $(-1, 1)$, and $q_- + K_2\tau = q_+ \leq 1$. This shows that (5.11) is a control.

Coefficient error. For $r \in [t_j, t_{j+1}]$, splitting as before and using $\|G_j - a(t_j, \hat{X}_j)\|_2 = |1 - \lambda_j|$, we have

$$\|G_j - a(r, X_r)\|_2 \leq |1 - \lambda_j| + 4\beta\|\hat{e}_j\|_2 + 4\beta\|X_{t_j} - X_r\|_2 + 4\sqrt{3}\beta^2\sqrt{\tau} \leq C\sqrt{\tau}.$$

Defect, output, and value. By the martingale identity, $M_{q_+} = M_{q_-} + \sum_{j < K_2} \int_{t_j}^{t_{j+1}} a(r, X_r) dB_r$. Since G_j is measurable with respect to $(B_s)_{s \leq t_j}$, we have

$G_j \Delta_{j+1} = \int_{t_j}^{t_{j+1}} G_j dB_r$. Since the first term of F is $F_0 = f(Y) = M_{q_-}$, we obtain

$$F - M_{q_+} = \sum_{j < K_2} \int_{t_j}^{t_{j+1}} (G_j - a(r, X_r)) dB_r,$$

and by the Itô isometry, the coefficient error, and $K_2 \tau \leq 1$, we have $\|F - M_{q_+}\|_2 \leq C\sqrt{\tau}$. Since $M_{q_+} = \tanh X_{q_+}$ and $\tanh$ is 1-Lipschitz, we have $\|\Theta - M_{q_+}\|_2 \leq \|\hat{e}_{K_2}\|_2 \leq C\sqrt{\tau}$. Combining the last two bounds, we obtain $\vartheta = \|F - \Theta\|_2 \leq C\sqrt{\tau}$. Since the pair (Θ, M_{q_+}) is a coupling of μ_Θ and μ_+ , we have $W_1(\mu_\Theta, \mu_+) \leq \|\Theta - M_{q_+}\|_2$. For the value, by Stein's identity, we have $\mathbb{E}[Y F_0] = \mathbb{E}[Y f(Y)] = q_- \mathbb{E} f'(Y) = q_- \mathbb{E} a(q_-, X_{q_-}) = \int_0^{q_-} \mathbb{E} a(t, X_t) dt$. Since $V = \mathbb{E}[Y F_0] + \tau \sum_{j < K_2} \mathbb{E} G_j$ and $t_{K_2} = q_+$,

$$|V - \mathcal{E}(\beta, h)| \leq \sum_{j < K_2} \int_{t_j}^{t_{j+1}} \|G_j - a(r, X_r)\|_1 dr \leq C\sqrt{\tau}. \quad \square$$

5.8 Homogenization

The law of the center is homogeneous only after averaging over the disorder. This subsection shows that its empirical behavior is nevertheless close to the limiting law, uniformly over the fields in $a\mathbf{1} + cS[0, B]^V$. The argument combines the finite range of the dependence of m_x on the couplings with the exponential decay of the number of eigenvalues of S away from zero on tori and hypercubes.

Lemma 5.27. *Let $G \in \mathcal{G}_D$ and run (5.3). For $1 \leq t \leq T$ and $x \in V$, the vector $\bar{u}_x^t$ is a measurable function of the couplings J_e of the edges e having an endpoint at graph distance at most $t - 1$ from x . Further, if $1 \leq K \leq T$, $\theta: \mathbb{R}^K \rightarrow \mathbb{R}$ is measurable, and $m = \theta(\bar{u}^K)$, then V can be partitioned into at most $(D + 1)^{2K-1}$ classes such that, within each class, the variables m_x are independent.*

Proof. The first claim holds for $t = 1$, since $u_x^1 = f_0 \sum_{y \sim x} W_{xy}$. If it holds at time t , then by (5.3), $u_x^{t+1} = \sum_{y \sim x} W_{xy} f_t(\bar{u}_y^t) - \sum_s b_{t,s} f_{s-1}(\bar{u}_x^{s-1})$ depends only on the edges xy , on edges with an endpoint at distance at most $t - 1$ from a neighbor of x , and on edges with an endpoint at distance at most $t - 2$ from x ; all of them have an endpoint at distance at most t from x . Here we used that the coefficients $b_{t,s}$ are real numbers. Now let E_x be the set of edges having an endpoint at distance at most $K - 1$ from x . If $E_x \cap E_{x'}$ contains an edge, then $\text{dist}(x, x') \leq 2K - 1$. We consider the graph on V in which $x \neq x'$ are joined if $\text{dist}(x, x') \leq 2K - 1$. Its maximal degree is at most $\sum_{j=1}^{2K-1} D^j \leq (D + 1)^{2K-1} - 1$, so it has a proper coloring with at most $(D + 1)^{2K-1}$ colors. Within a color class, the variables m_x are independent, since they are functions of the couplings in pairwise disjoint sets E_x . $\square$

If the Onsager coefficients were computed as empirical averages over all vertices, the recursion would generally introduce global dependence, and the preceding locality argument would no longer apply. The deterministic coefficients avoid this.

Lemma 5.28. *Let G_d be $\mathbb{T}_L^d$ with $L \geq 3$ or Q_d . For $0 < \alpha \leq 1$, let Π_α be the orthogonal projection onto the span of the eigenvectors of S with eigenvalues of absolute value at least α , and let $k_\alpha = \text{rank } \Pi_\alpha$. Then $k_\alpha \leq 2ne^{-d\alpha^2/2}$.*

Proof. For the torus, the characters $\chi_\omega(x) = e^{2\pi i \langle \omega, x \rangle / L}$, $\omega \in (\mathbb{Z}/L\mathbb{Z})^d$, form an orthogonal basis of $\mathbb{C}^V$. Since the neighbors of x are $x \pm e_j$, $1 \leq j \leq d$, we have $S\chi_\omega = \lambda(\omega)\chi_\omega$ with $\lambda(\omega) = d^{-1} \sum_{j=1}^d \cos(2\pi\omega_j/L)$. Since the multiplicity of an eigenvalue of the real symmetric matrix S equals the number of ω with $\lambda(\omega)$ equal to it, we have $k_\alpha = |\{\omega : |\lambda(\omega)| \geq \alpha\}|$. If ω is uniform on $(\mathbb{Z}/L\mathbb{Z})^d$, the variables $\cos(2\pi\omega_j/L)$ are independent, take values in $[-1, 1]$, and are centered, since $\sum_{k=0}^{L-1} e^{2\pi i k/L} = 0$. Hoeffding's inequality [84] gives $k_\alpha/n = \mathbb{P}(|\sum_j \cos(2\pi\omega_j/L)| \geq d\alpha) \leq 2e^{-d\alpha^2/2}$. For the hypercube, the characters $\chi_\omega(x) = (-1)^{\langle \omega, x \rangle}$, $\omega \in \{0, 1\}^d$, are eigenvectors of S with eigenvalues $d^{-1} \sum_j (-1)^{\omega_j}$, and the same argument applies with independent uniform signs. $\square$

The next lemma is a Hoeffding inequality for sums with a partition into independent classes; the argument is that of Janson [90].

Lemma 5.29. *Let $\Lambda > 0$, and let $(\xi_x)_{x \in V}$ be random variables with values in $[-\Lambda, \Lambda]$. Suppose that V is partitioned into χ classes such that, within each class, the variables ξ_x are independent. Then for every $b > 0$,*

$$\mathbb{P}\left(\left|\sum_{x \in V} (\xi_x - \mathbb{E}\xi_x)\right| \geq nb\right) \leq 2 \exp\left(-\frac{nb^2}{2\chi\Lambda^2}\right).$$

Proof. Let $V_1, \dots, V_\chi$ be the classes and $\lambda > 0$. By Hoeffding's lemma [84], $\mathbb{E}e^{s(\xi_x - \mathbb{E}\xi_x)} \leq e^{s^2\Lambda^2/2}$ for $s \in \mathbb{R}$. By Hölder's inequality with χ factors and independence within classes,

$$\mathbb{E} \exp\left(\lambda \sum_x (\xi_x - \mathbb{E}\xi_x)\right) \leq \prod_{i=1}^\chi \left(\prod_{x \in V_i} \mathbb{E} e^{\lambda(\xi_x - \mathbb{E}\xi_x)}\right)^{1/\chi} \leq \exp\left(\frac{\chi\lambda^2\Lambda^2n}{2}\right).$$

Using Markov's inequality with $\lambda = b/(\chi\Lambda^2)$, we bound the probability of the upper deviation by $e^{-nb^2/(2\chi\Lambda^2)}$. The lower deviation is treated in the same way. $\square$

Proposition 5.30. *Let (G_d) be the tori $\mathbb{T}_L^d$ for a fixed $L \geq 3$ or the hypercubes Q_d . Let μ be a probability measure on $[-1, 1]$, let $k \geq 1$ be an integer, and for each d let $m = m^{(d)}$ be a random vector in $[-1, 1]^V$ with the following two properties.*

- (i) The set V can be partitioned into at most $(D+1)^k$ classes such that, within each class, the variables m_x are independent.
- (ii) The law of m_x does not depend on $x \in V$, and it converges weakly to μ as $d \rightarrow \infty$.

Then the conclusion of Proposition 5.2(c) holds with μ in place of μ_ϵ .

Proof. Fix Y , L_Y , a , c , and B as in Proposition 5.2(c), and let $\mathcal{Z} = a\mathbf{1} + cS[0, B]^V$. Let $\bar{Y}(s) = \int Y(m', s) \mu(dm')$ for $s \geq 0$, and for $r \in [0, \infty)^V$ let

$$\mathcal{A}(r) = \frac{1}{n} \sum_x (Y(m_x, r_x) - \bar{Y}(r_x)).$$

Both $Y(m', \cdot)$ and $\bar{Y}$ are L_Y -Lipschitz. By the Cauchy–Schwarz inequality, we have, for all $r, r' \in [0, \infty)^V$,

$$|\mathcal{A}(r) - \mathcal{A}(r')| \leq \frac{2L_Y}{n} \sum_x |r_x - r'_x| \leq 2L_Y \|r - r'\|_n. \quad (5.12)$$

Let $M = a + cB$ and $\Lambda = 1 + \max_{[-1, 1] \times [0, M]} |Y|$, so that $|\mathcal{A}(r)| \leq 2\Lambda$ for $r \in [0, M]^V$. Fix auxiliary parameters $\alpha \in (0, 1]$, $\zeta > 0$, and $\xi > 0$, and let $\text{clip}: \mathbb{R}^V \rightarrow [0, M]^V$ be the coordinatewise projection onto $[0, M]$.

Step 1: reduction to a ball in the range of Π_α . Let $r \in \mathcal{Z}$, and write $r = a\mathbf{1} + cSu$ with $u \in [0, B]^V$. Since S is nonnegative with unit row sums, $r \in [0, M]^V$. Let Π_α and k_α be as in Lemma 5.28. Since $S\mathbf{1} = \mathbf{1}$, the vector $\mathbf{1}$ lies in the range of Π_α . Let $r_H = a\mathbf{1} + c\Pi_\alpha Su$. Then r_H belongs to the set

$$\mathcal{K}_\alpha = \{w \in \text{range } \Pi_\alpha : \|w\|_n \leq M\},$$

because $\|\Pi_\alpha Su\|_n \leq \|u\|_n \leq B$. Further, $\|r - r_H\|_n = c\|(I - \Pi_\alpha)Su\|_n \leq c\alpha B$, since every eigenvalue of S on the range of $I - \Pi_\alpha$ has absolute value less than α . Since clip is 1-Lipschitz in each coordinate and fixes r , we also have $\|r - \text{clip}(r_H)\|_n \leq c\alpha B$. By (5.12),

$$\sup_{r \in \mathcal{Z}} |\mathcal{A}(r)| \leq 2L_Y c\alpha B + \sup_{w \in \mathcal{K}_\alpha} |\mathcal{A}(\text{clip}(w))|. \quad (5.13)$$

Step 2: a net. The set $\mathcal{K}_\alpha$ is a Euclidean ball of radius M in the k_α -dimensional space $\text{range } \Pi_\alpha$, with the norm $\|\cdot\|_n$. A maximal ζ -separated subset $\mathcal{N} \subset \mathcal{K}_\alpha$ is a ζ -net, and the balls of radius $\zeta/2$ around its points are disjoint and contained in the ball of radius $M + \zeta/2$; comparing volumes gives $|\mathcal{N}| \leq (1 + 2M/\zeta)^{k_\alpha}$. For $w \in \mathcal{K}_\alpha$ and $w' \in \mathcal{N}$ with $\|w - w'\|_n \leq \zeta$, we have $\|\text{clip}(w) - \text{clip}(w')\|_n \leq \zeta$. Then (5.12) implies that

$$\sup_{w \in \mathcal{K}_\alpha} |\mathcal{A}(\text{clip}(w))| \leq 2L_Y \zeta + \max_{w' \in \mathcal{N}} |\mathcal{A}(\text{clip}(w'))|. \quad (5.14)$$

Step 3: a fixed field. Let $r \in [0, M]^V$ be deterministic. By (ii), the law λ_d of m_x does not depend on x . Then

$$|\mathbb{E}\mathcal{A}(r)| \leq \varepsilon_d, \quad \varepsilon_d = \sup_{s \in [0, M]} \left| \int Y(m', s) \lambda_d(dm') - \int Y(m', s) \mu(dm') \right|.$$

For each s , the difference inside the supremum tends to zero by (ii), since the function $Y(\cdot, s)$ is bounded and continuous on $[-1, 1]$. Both integrals are L_Y -Lipschitz in s , and a finite net of $[0, M]$ shows that $\varepsilon_d \rightarrow 0$. Next, by (i), the variables $Y(m_x, r_x) \in [-\Lambda, \Lambda]$ satisfy the hypothesis of Lemma 5.29 with $\chi \leq (D+1)^k$ classes, and we obtain

$$\mathbb{P}(|\mathcal{A}(r) - \mathbb{E}\mathcal{A}(r)| \geq \xi) \leq 2 \exp\left(-\frac{n\xi^2}{2(D+1)^k\Lambda^2}\right).$$

Step 4: union bound. By Step 3, the bound on $|\mathcal{N}|$, and Lemma 5.28,

$$\mathbb{P}\left(\max_{w \in \mathcal{N}} |\mathcal{A}(\text{clip}(w))| \geq \varepsilon_d + \xi\right) \leq 2 \exp\left(n\left[2e^{-d\alpha^2/2} \log\left(1 + \frac{2M}{\xi}\right) - \frac{\xi^2}{2(D+1)^k\Lambda^2}\right]\right).$$

Since $D \leq 2d$, we have $e^{-d\alpha^2/2}(D+1)^k \rightarrow 0$, and the bracket is at most $-\xi^2/(4(D+1)^k\Lambda^2)$ for large d . Since $n \geq 2^d$, we have $n/(D+1)^k \rightarrow \infty$, and the probability tends to zero. Since $|\mathcal{A}(\text{clip}(w))| \leq 2\Lambda$, we obtain

$$\limsup_{d \rightarrow \infty} \mathbb{E} \max_{w \in \mathcal{N}} |\mathcal{A}(\text{clip}(w))| \leq \limsup_{d \rightarrow \infty} \left(\varepsilon_d + \xi + 2\Lambda \mathbb{P}\left(\max_{w \in \mathcal{N}} |\mathcal{A}(\text{clip}(w))| \geq \varepsilon_d + \xi\right)\right) = \xi.$$

Combining this with (5.13) and (5.14), we find

$$\limsup_{d \rightarrow \infty} \mathbb{E} \sup_{r \in \mathcal{Z}} |\mathcal{A}(r)| \leq 2L_Y c \alpha B + 2L_Y \xi + \xi.$$

The left side does not depend on α , ζ , and ξ , which may be taken arbitrarily small. This completes the proof. $\square$

5.9 Proof of Proposition 5.2

Proof of Proposition 5.2. Fix $\beta > 0$, $h > 0$, and $\varepsilon > 0$, and let M_W be as in Lemma 5.20. By Lemma 2.6(a), (b), and (c), we have $\mathbb{E}M_{q_+}^2 = q_+$, $M_{q_+} = \tanh X_{q_+}$, and

$$\mathcal{P}(\beta, h) = \beta \mathcal{E}(\beta, h) + h \mathbb{E}M_{q_+} + \mathbb{E} \text{ent}(M_{q_+}) + \frac{\beta^2}{4}(1 - q_+)^2. \quad (5.15)$$

We choose the control of Proposition 5.26. If $q_- = q_+$, it is the control of part (a), and we set $\tau = 1$. If $q_- < q_+$, it is the control of part (b) with an integer $K_2 \geq 1$ that we now choose. Let Θ denote the output, with law μ_Θ . For a control of part (b), we define

$$e_1(\tau) = (\beta + \beta M_W + h + \beta^2)C\sqrt{\tau} + 2\beta\sqrt{\tau} + |\mathbb{E} \text{ent}(\Theta) - \mathbb{E} \text{ent}(M_{q_+})|,$$

where Θ and M_{q_+} are coupled as in (5.11), and we set $e_1 = 0$ for the control of part (a). In case (b), since ent is uniformly continuous on $[-1, 1]$ and $\|\Theta - M_{q_+}\|_2 \leq C\sqrt{\tau}$, we

have $e_1(\tau) \rightarrow 0$ as $K_2 \rightarrow \infty$. In this case, we choose K_2 so large that $e_1(\tau) \leq \epsilon/2$ and $C\sqrt{\tau} \leq \epsilon$. In either case, we set $\mu_\epsilon = \mu_\Theta$. This fixes the control. Next, we choose the accuracy parameters: $\eta \in (0, q_-)$ and an integer $\ell \geq 2$ such that

$$\beta\left(\eta + \frac{4}{\ell} + \frac{4M_W}{\ell}\right) \leq \frac{\epsilon}{2}.$$

This fixes p , K_1 , and the P_j , and with them the integer $K = K_1 + K_2$. Let $T = K + 1$, and let m be the center produced by the algorithm on G_d , with the query history of depth T of Step (iii) of the algorithm.

(b) By Proposition 5.26, $W_1(\mu_\epsilon, \mu_+) = 0$ in case (a) and $W_1(\mu_\epsilon, \mu_+) \leq C\sqrt{\tau} \leq \epsilon$ in case (b).

(a) By Proposition 5.25(a) and (b), the $\liminf$ in (a) is at least

$$\beta V - \beta\left(2\sqrt{\tau} \mathbf{1}_{\{K_2 \geq 1\}} + \eta + \frac{4}{\ell} + M_W \vartheta + \frac{4M_W}{\ell}\right) + h\mathbb{E}\Theta + \mathbb{E} \text{ent}(\Theta) + \frac{\beta^2}{4}(1 - \mathbb{E}\Theta^2)^2.$$

In case (a) of Proposition 5.26, we have $V = \mathcal{E}(\beta, h)$, $\vartheta = 0$, and $\Theta = M_{q_+}$, and by (5.15) and $\mathbb{E}M_{q_+}^2 = q_+$ the lower bound is at least $\mathcal{P}(\beta, h) - \epsilon/2$. In case (b), we have $V \geq \mathcal{E}(\beta, h) - C\sqrt{\tau}$ and $\vartheta \leq C\sqrt{\tau}$. Further, $|h\mathbb{E}\Theta - h\mathbb{E}M_{q_+}| \leq hC\sqrt{\tau}$, and since $|s^2 - s'^2| \leq 2|s - s'|$ for $s, s' \in [-1, 1]$,

$$|(1 - \mathbb{E}\Theta^2)^2 - (1 - \mathbb{E}M_{q_+}^2)^2| \leq 2\mathbb{E}|\Theta^2 - M_{q_+}^2| \leq 4\|\Theta - M_{q_+}\|_2 \leq 4C\sqrt{\tau}.$$

By (5.15) and $\mathbb{E}M_{q_+}^2 = q_+$, the lower bound is at least $\mathcal{P}(\beta, h) - e_1(\tau) - \epsilon/2 \geq \mathcal{P}(\beta, h) - \epsilon$.

(c) By Lemma 5.27, applied with θ as in Step (iii) of the algorithm, V can be partitioned into at most $(D + 1)^{2K-1}$ classes within which the variables m_x are independent. By Proposition 5.25(c), the law of m_x does not depend on x and converges weakly to $\mu_\Theta = \mu_\epsilon$. Then (c) follows from Proposition 5.30 with $k = 2K - 1$. $\square$

Gaussian conditioning and the Onsager term

The goal of this section is to prove Theorem 6.2, which states that on every regular graph of large degree, the pressure is at least the expected value of

$$\Theta_G(m) = \beta H_G(m) + h \sum_{x \in V} m_x + \sum_{x \in V} \text{ent}(m_x) + \frac{\beta^2}{4} \langle \kappa, S\kappa \rangle \quad (6.1)$$

per vertex, up to an error that tends to zero with the degree, provided that two conditions hold. The functional Θ_G is the analogue on G of the TAP free energy of Thouless, Anderson, and Palmer [135]. First, the center m must be computed from a bounded number of queries of the disorder, in the sense of Definition 5.1. Second, the planted product model of Definition 6.1 at m must have relative entropy $o(n)$ with respect to pure noise. Here and below, $\kappa_x = 1 - m_x^2$, and we suppress the dependence of Θ_G on β and h .

We first explain the structure of the bound. Let

$$Z_G = \sum_{\sigma \in \{-1,1\}^V} \exp \left(\beta H_G(\sigma) + h \sum_{x \in V} \sigma_x \right),$$

so that $p_G(\beta, h) = n^{-1} \mathbb{E} \log Z_G$. For every $m \in [-1, 1]^V$, even one that depends on the disorder, we have $\log Z_G \geq \Theta_G(m) - (\beta^2/4) \langle \kappa, S\kappa \rangle$ by Jensen's inequality with respect to π_m , because H_G has no diagonal terms. The last term in (6.1), the Onsager correction, requires more work. We write $\sigma = m + v$. After conditioning on the queries, the unobserved part of the disorder is an independent Gaussian vector z on the edge space, and it enters the exponent through $\beta \langle z, \Pi a(v) \rangle$, where Π is a projection and $a(v)$ is the edge feature of Definition 6.1. Averaging the exponential over z produces $\exp(\beta^2 \|\Pi a(v)\|^2 / 2)$, and for typical v this is approximately $\exp((\beta^2/4) \langle \kappa, S\kappa \rangle)$. However, the pressure involves the average of the logarithm, not the logarithm of the average. The difference between the two is a relative entropy $\text{KL}(P_0 \parallel Q)$, where Q is obtained from the planted product model by conditioning the planted configuration on an event of large probability and by projecting the observation. Using Talagrand's transport inequality for the Gaussian measure [132], we bound it by $4\beta(n\mathcal{D}(m))^{1/2}$, up to a negligible error. The remaining terms in the expansion of the partition function around m are not small for every v . We show that they are small on a set of configurations of large π_m -probability, and we restrict the expectation to this set.

Conditioning on the history of an iterative algorithm goes back to Bolthausen's construction of solutions of the TAP equations [25], and Bayati and Montanari used it to prove the state evolution of message passing [21]. Bolthausen [26] used this conditioning to compute the free energy of the SK model at high temperature. He showed that the quenched free energy equals the annealed free energy conditional on the iterates. Brennecke and Yau [29] extended his argument, which required β to be small, to all (β, h) with $h > 0$ and $\beta^2 \mathbb{E} \operatorname{sech}^2(\beta\sqrt{q}Z + h) \leq 1$, where $q = \mathbb{E} \tanh^2(\beta\sqrt{q}Z + h)$ and Z is a standard Gaussian variable. Fan and Wu adapted Bolthausen's argument to couplings with an orthogonally invariant law [64]. These works use a conditional second-moment method. Here the gap between the conditional quenched and annealed free energies is the relative entropy above, and we bound it through the planted model. For the SK model in zero field, the normalized partition function is likewise the likelihood ratio of a planted model with respect to the null model, and El Alaoui, Montanari, and Sellke used this together with contiguity at high temperature [61, Section 4.1]. Planting that preserves the typical properties of the null ensemble is called *quiet* [96].

Restricting the free energy to a thin band around the output of an algorithm was used by Fan, Li, and Sen [66] for spin glasses with orthogonally invariant couplings. Bands around a magnetization also appear in Subag's TAP representation of the free energy of spherical models [131], and they define the generalized TAP free energy of Chen, Panchenko, and Subag [41]. Huang and Sellke [85] proved the lower bound in the Parisi formula for spherical models by constructing an ultrametric tree of states, partly with an optimization algorithm, and bounding the free energy of bands around its leaves. For mixed p -spin models with Ising spins, in any field, Chen and Panchenko proved a TAP representation of the free energy [39, Theorem 1]. For the SK model, let q_+ be the largest point of the support of the Parisi measure, as in Section 2.3. Then their theorem states that $\mathcal{P}(\beta, h)$ is the limit, as $N \rightarrow \infty$ and then $\varepsilon \downarrow 0$, of the expected maximum of the TAP free energy over magnetizations with self-overlap at least $q_+ - \varepsilon$. In particular, the TAP free energy at a magnetization with self-overlap at least q_+ is asymptotically at most $\mathcal{P}(\beta, h)$, and Theorem 6.2 is an inequality of the same type on G , under different conditions on m .

Section 6.1 states Definition 6.1 and Theorem 6.2. Section 6.2 describes the law of the disorder given a query history. Section 6.3 expands the partition function around the center, and Section 6.4 restricts the expansion to a set of configurations of large product probability on which the error terms are small. Section 6.5 bounds the remaining Gaussian likelihood in terms of $\mathcal{D}(m)$, and Section 6.6 proves Theorem 6.2.

6.1 The planted model and the main result

We use the query histories of Definition 5.1. We recall that a query history of depth T on G consists of a random variable U independent of $(g_e)_{e \in E}$ and random vectors $x^1, \dots, x^T \in \mathbb{R}^V$ such that x^t is measurable with respect to the σ -algebra generated by U and $Wx^1, \dots, Wx^{t-1}$. The σ -algebra $\mathcal{H}$ is generated by U and $Wx^1, \dots, Wx^T$, and a center is an $\mathcal{H}$ -measurable vector $m \in (-1, 1)^V$ such that $m = x^t$ for some $t \leq T$. In particular, the response Wm of the disorder to the center is observed; see Lemma 6.4(a).

The following definition is also used in the proof of Theorem 7.5.

Definition 6.1. Let S be a symmetric $n \times n$ matrix with nonnegative entries, zero diagonal, and unit row sums, and let $E = \{\{x, y\} : S_{xy} > 0\}$. Fix $\beta \geq 0$ and $m \in [-1, 1]^n$. Let σ^* have law π_m , set $v^* = \sigma^* - m$, and let Z be a standard Gaussian vector in $\mathbb{R}^E$ independent of σ^* . The *planted product model* is the law Q_m of $Y = \beta a(v^*) + Z$, where $a(v)_{xy} = \sqrt{S_{xy}} v_x v_y$. We write P_0 for the standard Gaussian law on $\mathbb{R}^E$ and $\mathcal{D}(m) = \text{KL}(Q_m \| P_0)$.

For $G \in \mathcal{G}_D$ we always apply Definition 6.1 with $S = A_G/D$, and we identify $\{1, \dots, n\}$ with V . Then $E = E$, the planted model lives on the edge space $\mathbb{R}^E$ of the disorder, and $a(v)_{xy} = D^{-1/2} v_x v_y$ for $xy \in E$. We write $\mathbb{E}_{\pi_m}$ for the expectation over σ with law π_m , and in such expectations we always set $v = \sigma - m$. Since the Gaussian law on $\mathbb{R}^E$ with mean b and identity covariance has density $\exp(\langle y, b \rangle - \|b\|^2/2)$ with respect to P_0 , the law Q_m has density

$$L_m(y) = \mathbb{E}_{\pi_m} \exp \left(\beta \langle y, a(v) \rangle - \frac{\beta^2}{2} \|a(v)\|^2 \right) \quad (y \in \mathbb{R}^E) \quad (6.2)$$

with respect to P_0 . By Lemma 6.9 below, the function $m \mapsto \mathcal{D}(m)$ is continuous on $[-1, 1]^V$. In particular, $\mathcal{D}(m)$ is a bounded random variable whenever m is a random vector in $[-1, 1]^V$. If m is a random center, then $\mathcal{D}(m)$ denotes the value of this deterministic function at m . The planted configuration and the noise of Definition 6.1 are integrated out in the definition of $\mathcal{D}$, and they have no relation to the disorder g .

Theorem 6.2. Let $\beta \geq 0$, $h \in \mathbb{R}$, and $\delta \geq 0$, and let $T \geq 1$ be an integer. For each $k \geq 1$, let $D_k \geq 1$, let $G_k \in \mathcal{G}_{D_k}$ have n_k vertices, let $S_k = A_{G_k}/D_k$, and let a query history of depth T on G_k with center $m = m^{(k)}$ be given. Let $\mathcal{D}$ be as in Definition 6.1 with $S = S_k$. Suppose that $D_k \rightarrow \infty$ and

$$\limsup_{k \rightarrow \infty} \frac{1}{n_k} \mathbb{E} \mathcal{D}(m^{(k)}) \leq \delta. \quad (6.3)$$

Then

$$\liminf_{k \rightarrow \infty} p_{G_k}(\beta, h) \geq \liminf_{k \rightarrow \infty} \frac{1}{n_k} \mathbb{E} \left[\beta H_{G_k}(m) + h \sum_x m_x + \sum_x \text{ent}(m_x) + \frac{\beta^2}{4} \langle \kappa, S_k \kappa \rangle \right] - 4\beta\sqrt{\delta}, \quad (6.4)$$

where $m = m^{(k)}$ and $\kappa_x = 1 - m_x^2$.

Remark 6.3. Theorem 6.2 follows from the nonasymptotic bound of Proposition 6.12, which holds for every $G \in \mathcal{G}_D$ and every query history of depth T on G . The only property of the sequence (G_k) used in the proof is $D_k \rightarrow \infty$. No transitivity, no bound on the operator norm of W , and no relation between n_k and D_k is needed.

6.2 Gaussian conditioning

In this subsection, we fix $G \in \mathcal{G}_D$ and a query history of depth T on G , defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. We write $g = (g_e)_{e \in E}$ and $\mathcal{B}(\mathbb{R}^E)$ for the Borel σ -algebra of $\mathbb{R}^E$. For $0 \leq t \leq T$, let $\mathcal{F}_t$ be the σ -algebra generated by $\mathcal{U}$ and $Wx^1, \dots, Wx^t$, so that $\mathcal{F}_T = \mathcal{H}$.

For $b \in \mathbb{R}^V$, let $\Lambda_b: \mathbb{R}^E \rightarrow \mathbb{R}^V$ be the linear map defined by

$$(\Lambda_b w)_x = D^{-1/2} \sum_{y: xy \in E} w_{xy} b_y \quad (x \in V).$$

Then $Wb = \Lambda_b g$; each query reveals n linear functionals of g . For $0 \leq t \leq T$, let $\Lambda^{(t)}: \mathbb{R}^E \rightarrow (\mathbb{R}^V)^t$ be the map $w \mapsto (\Lambda_{x^1} w, \dots, \Lambda_{x^t} w)$, and let Π_t be the orthogonal projection of $\mathbb{R}^E$ onto $\ker \Lambda^{(t)}$; in particular, $\Pi_0 = I$. We set

$$\Pi = \Pi_T, \quad \Pi^\perp = I - \Pi, \quad \bar{g} = \Pi^\perp g,$$

and we let $\bar{W}$ be the symmetric $V \times V$ matrix with $\bar{W}_{xy} = D^{-1/2} \bar{g}_{xy}$ for $xy \in E$ and $\bar{W}_{xy} = 0$ otherwise. For a symmetric matrix K , we write $|K| = (K^2)^{1/2}$ and $\|K\|_{\text{tr}} = \text{tr } |K|$ for its trace norm.

The next lemma is a version of the conditioning lemmas of Bolthausen [25] and Bayati and Montanari [21, Lemmas 11 and 12], for linear functionals of the edge disorder that are chosen adaptively and may depend on the auxiliary randomness $\mathcal{U}$.

Lemma 6.4. *Let $G \in \mathcal{G}_D$, and let a query history of depth T on G be given. Then the following statements hold.*

- (a) *The projection Π and the vector $\bar{g}$ are $\mathcal{H}$ -measurable, and $\text{rank } \Pi^\perp \leq nT$. If m is a center of the history, then Wm is $\mathcal{H}$ -measurable.*
- (b) *For every $\mathcal{H} \otimes \mathcal{B}(\mathbb{R}^E)$ -measurable function $F: \Omega \times \mathbb{R}^E \rightarrow [0, \infty]$, we have*

$$\mathbb{E}[F(\cdot, g) \mid \mathcal{H}] = \int F(\cdot, \bar{g} + \Pi w) P_0(dw) \quad \text{almost surely.} \quad (6.5)$$

- (c) *We have $\mathbb{E}\|\bar{g}\|^2 = \mathbb{E} \text{rank } \Pi^\perp \leq nT$ and $\mathbb{E}\|\bar{W}\|_{\text{tr}} \leq n(2T/D)^{1/2}$.*

Part (b) states that, conditional on $\mathcal{H}$, the disorder has the law of $\bar{g} + \Pi z$, where z is a standard Gaussian vector on $\mathbb{R}^E$ independent of $\mathcal{H}$. The function F may depend on $\mathcal{H}$ -measurable quantities, such as the center and the projection Π ; this is how the lemma is used below. In particular, $\mathbb{E}[g | \mathcal{H}] = \bar{g}$ and $\bar{W} = \mathbb{E}[W | \mathcal{H}]$ is the conditional mean of the coupling matrix.

Proof. For a linear map M between Euclidean spaces, let M^+ denote its Moore–Penrose pseudoinverse. Then M^+M is the orthogonal projection onto $(\ker M)^\perp$, and $M \mapsto M^+$ is Borel measurable, being the pointwise limit of $(M^\top M + \epsilon I)^{-1}M^\top$ as $\epsilon \downarrow 0$. In particular, $\Pi_t^\perp = I - \Pi_t = (\Lambda^{(t)})^+ \Lambda^{(t)}$.

(a) Fix $1 \leq t \leq T$. Since x^s is $\mathcal{F}_{s-1}$ -measurable and $b \mapsto \Lambda_b$ is linear, the map $\Lambda^{(t)}$ is $\mathcal{F}_{t-1}$ -measurable, and so is Π_t . Further, $\Lambda^{(t)}g = (Wx^1, \dots, Wx^t)$ is $\mathcal{F}_t$ -measurable, and so is $\Pi_t^\perp g = (\Lambda^{(t)})^+ \Lambda^{(t)}g$. For $t = T$ this shows that Π and $\bar{g}$ are $\mathcal{H}$ -measurable. The rank of $\Pi^\perp$ equals that of $\Lambda^{(T)}$, which is at most nT , since $\Lambda^{(T)}$ takes values in a space of dimension nT . If m is a center, let τ be the least t with $m = x^t$. The events $\{\tau = t\}$ belong to $\mathcal{H}$, because m and $x^1, \dots, x^T$ are $\mathcal{H}$ -measurable, and $Wm = \sum_{t \leq T} \mathbf{1}_{\{\tau=t\}} Wx^t$ is $\mathcal{H}$ -measurable.

(b) For $0 \leq t \leq T$, let (b_t) denote the statement (b) with $\mathcal{F}_t$ and Π_t in place of $\mathcal{H}$ and Π . We prove (b_t) by induction on t . In (b_t) , the right side of (6.5) is $\mathcal{F}_t$ -measurable by Tonelli's theorem, since $\Pi_t^\perp g$ and Π_t are $\mathcal{F}_t$ -measurable by the proof of (a). The bounded $\mathcal{F}_t \otimes \mathcal{B}(\mathbb{R}^E)$ -measurable functions F that satisfy (6.5) form a vector space that contains the constants and is closed under bounded monotone limits, and nonnegative F are monotone limits of bounded ones. By the monotone-class theorem, it therefore suffices to prove (b_t) for $F(\omega, w) = \mathbf{1}_A(\omega)f(w)$ with $A \in \mathcal{F}_t$ and $f: \mathbb{R}^E \rightarrow [0, \infty)$ bounded and Borel. Since $\mathbf{1}_A$ is $\mathcal{F}_t$ -measurable, this amounts to

$$\mathbb{E}[f(g) | \mathcal{F}_t] = \int f(\Pi_t^\perp g + \Pi_t w) P_0(dw). \quad (6.6)$$

For $t = 0$, (6.6) holds because $\Pi_0 = I$ and g is independent of $\mathcal{F}_0$, which is generated by U .

Let $1 \leq t \leq T$, and assume (b_{t-1}) . We abbreviate $\Pi' = \Pi_{t-1}$ and $\Lambda = \Lambda_{x^t}$. Since $\ker \Lambda^{(t)} \subset \ker \Lambda^{(t-1)}$, the operator $\Xi = \Pi' - \Pi_t$ is the orthogonal projection onto the orthogonal complement of the range of Π_t in the range of Π' . The operators Π' , Π_t , Ξ , and Λ are $\mathcal{F}_{t-1}$ -measurable, and

$$\Lambda \Pi_t = 0, \quad \Lambda \Pi' = \Lambda \Xi, \quad \Pi_t^\perp \Pi'^\perp = \Pi'^\perp, \quad \Pi_t^\perp \Pi' = \Xi. \quad (6.7)$$

The events $A' \cap \{Wx^t \in C\}$, with $A' \in \mathcal{F}_{t-1}$ and $C \subset \mathbb{R}^V$ Borel, form a π -system that generates $\mathcal{F}_t$. By Dynkin's π - λ theorem and the measurability of the right side of (6.6), it suffices to show that

$$\mathbb{E}[\mathbf{1}_{A'} \mathbf{1}_C(\Lambda g) f(g)] = \mathbb{E}[\Psi(\cdot, g)], \quad (6.8)$$

where

$$\Psi(\omega, w) = \mathbf{1}_{A'}(\omega) \mathbf{1}_C(\Lambda w) \int f(\Pi_t^\perp w + \Pi_t w') P_0(dw'),$$

and where we used $Wx^t = \Lambda g$. The function Ψ is $\mathcal{F}_{t-1} \otimes \mathcal{B}(\mathbb{R}^E)$ -measurable. By (b_{t-1}) and (6.7), we have

$$\begin{aligned} \mathbb{E}[\Psi(\cdot, g)] &= \mathbb{E} \int \Psi(\cdot, \Pi'^\perp g + \Pi' w) P_0(dw) \\ &= \mathbb{E} \left[\mathbf{1}_{A'} \iint \mathbf{1}_C(\Lambda \Pi'^\perp g + \Lambda \Xi w) f(\Pi'^\perp g + \Xi w + \Pi_t w') P_0(dw) P_0(dw') \right]. \end{aligned}$$

For fixed ω , the Gaussian vectors Ξw and $\Pi_t w$ are independent under P_0 , because $\Xi \Pi_t = 0$. Then the pair $(\Xi w, \Xi w + \Pi_t w')$ under $P_0 \otimes P_0$ has the same law as $(\Xi w, \Pi' w)$ under P_0 . Using also $\Lambda \Xi w = \Lambda \Pi' w$, we find that the double integral equals

$$\int \mathbf{1}_C(\Lambda(\Pi'^\perp g + \Pi' w)) f(\Pi'^\perp g + \Pi' w) P_0(dw).$$

By (b_{t-1}) , applied to $(\omega, w) \mapsto \mathbf{1}_{A'}(\omega) \mathbf{1}_C(\Lambda(\omega)w) f(w)$, the expectation of $\mathbf{1}_{A'}$ times this integral is the left side of (6.8). This proves (b_t) , and (b) is the case $t = T$.

(c) By (b) with $F(\omega, w) = \|w\|^2$, and since $\bar{g}$ is orthogonal to the range of Π , we have

$$\mathbb{E}[\|g\|^2 | \mathcal{H}] = \int \|\bar{g} + \Pi w\|^2 P_0(dw) = \|\bar{g}\|^2 + \text{tr } \Pi = \|\bar{g}\|^2 + |E| - \text{rank } \Pi^\perp.$$

Taking expectations and using $\mathbb{E}\|g\|^2 = |E|$, we find $\mathbb{E}\|\bar{g}\|^2 = \mathbb{E} \text{rank } \Pi^\perp \leq nT$. Next, the squared Hilbert–Schmidt norm of $\bar{W}$ is $\sum_{x,y} \bar{W}_{xy}^2 = 2D^{-1} \|\bar{g}\|^2$. Since $\bar{W}$ has n eigenvalues, the Cauchy–Schwarz inequality implies that $\|\bar{W}\|_{\text{tr}} \leq (2n/D)^{1/2} \|\bar{g}\|$. Then

$$\mathbb{E}\|\bar{W}\|_{\text{tr}} \leq \left(\frac{2n}{D}\right)^{1/2} (\mathbb{E}\|\bar{g}\|^2)^{1/2} \leq n \left(\frac{2T}{D}\right)^{1/2}.$$

This completes the proof. $\square$

6.3 Expansion around the center

The following identity is deterministic. It holds for every realization of the disorder and every orthogonal projection Π of $\mathbb{R}^E$, with $\bar{g} = \Pi^\perp g$ and $\bar{W}$ defined from $\bar{g}$ as in Section 6.2. For $m \in (-1, 1)^V$, we define $r \in \mathbb{R}^V$ by

$$r_x = \beta(Wm)_x + h - \text{atanh}(m_x) - \beta^2 m_x (S\kappa)_x \quad (x \in V). \quad (6.9)$$

Lemma 6.5. *Let $G \in \mathcal{G}_D$, $\beta \geq 0$, $h \in \mathbb{R}$, and $m \in (-1, 1)^V$, and let Π be an orthogonal projection of $\mathbb{R}^E$. Then*

$$\begin{aligned} Z_G &= e^{\Theta_G(m)} \mathbb{E}_{\pi_m} \exp \left(\langle r, v \rangle + \frac{\beta}{2} \langle v, \bar{W} v \rangle + \beta^2 \langle m \odot v, S(m \odot v) \rangle - \frac{\beta^2}{2} \|\Pi^\perp a(v)\|^2 \right. \\ &\quad \left. + \beta \langle g, \Pi a(v) \rangle - \frac{\beta^2}{2} \|\Pi a(v)\|^2 \right). \end{aligned} \quad (6.10)$$

Proof. Let $y_x = \operatorname{atanh}(m_x)$. Then $\pi_m(\sigma) = \prod_x e^{y_x \sigma_x} / (2 \cosh y_x)$ is positive for all σ , and for every function f on $\{-1, 1\}^V$ we have

$$\sum_{\sigma} f(\sigma) = \mathbb{E}_{\pi_m} \left[f(\sigma) \prod_{x \in V} 2 \cosh(y_x) e^{-y_x \sigma_x} \right].$$

By (2.3), we have $\log 2 \cosh y_x - y_x m_x = \operatorname{ent}(m_x)$. With $\sigma = m + v$, it follows that the product inside the expectation equals $\exp(\sum_x \operatorname{ent}(m_x) - \langle y, v \rangle)$. Since W is symmetric, $H_G(m + v) = H_G(m) + \langle Wm, v \rangle + H_G(v)$. Applying the previous display to $f(\sigma) = \exp(\beta H_G(\sigma) + h \sum_x \sigma_x)$, we obtain

$$Z_G = \exp \left(\beta H_G(m) + h \sum_x m_x + \sum_x \operatorname{ent}(m_x) \right) \mathbb{E}_{\pi_m} \exp \left(\langle \beta Wm + h \mathbf{1} - y, v \rangle + \beta H_G(v) \right). \quad (6.11)$$

Next, $H_G(v) = \sum_{xy \in E} D^{-1/2} g_{xy} v_x v_y = \langle g, a(v) \rangle$. We split $g = \bar{g} + \Pi g$. Since $\langle \bar{g}, a(v) \rangle = \sum_{xy \in E} D^{-1/2} \bar{g}_{xy} v_x v_y = \langle v, \bar{W}v \rangle / 2$ and $\langle \Pi g, a(v) \rangle = \langle g, \Pi a(v) \rangle$, we have

$$\beta H_G(v) = \frac{\beta}{2} \langle v, \bar{W}v \rangle + \beta \langle g, \Pi a(v) \rangle. \quad (6.12)$$

Finally, we use the identity $v_x^2 = \kappa_x - 2m_x v_x$, which holds because $\sigma_x^2 = 1$. Since S is symmetric with zero diagonal, a sum over edges of a symmetric expression is half the corresponding sum over ordered pairs, and we find

$$\begin{aligned} \|a(v)\|^2 &= \sum_{xy \in E} S_{xy} (\kappa_x - 2m_x v_x) (\kappa_y - 2m_y v_y) \\ &= \frac{1}{2} \langle \kappa, S\kappa \rangle - 2 \langle m \odot S\kappa, v \rangle + 2 \langle m \odot v, S(m \odot v) \rangle. \end{aligned}$$

Since $\|\Pi a(v)\|^2 = \|a(v)\|^2 - \|\Pi^\perp a(v)\|^2$, this gives

$$\frac{\beta^2}{2} \|\Pi a(v)\|^2 = \frac{\beta^2}{4} \langle \kappa, S\kappa \rangle - \beta^2 \langle m \odot S\kappa, v \rangle + \beta^2 \langle m \odot v, S(m \odot v) \rangle - \frac{\beta^2}{2} \|\Pi^\perp a(v)\|^2. \quad (6.13)$$

We insert (6.12) into (6.11), and we add and subtract $(\beta^2/2) \|\Pi a(v)\|^2$ in the exponent. Using (6.13) for the added term, moving the constant $(\beta^2/4) \langle \kappa, S\kappa \rangle$ out of the expectation, and recalling (6.9), we obtain (6.10). $\square$

6.4 A band of typical configurations

We begin with a moment bound for the disorder. It holds on every $G \in \mathcal{G}_D$ and uses no bound on the operator norm of W . The corresponding bound for the Hamiltonian is (F6).

Lemma 6.6. *Let $G \in \mathcal{G}_D$. Then*

$$\mathbb{E} \max_{\sigma} \|W\sigma\|^2 \leq 8n \left(\log 2 + \frac{1}{4} \right) \leq 8n.$$

Further, for every realization of the disorder and every $w \in [-1, 1]^V$, we have $\|Ww\| \leq \max_{\sigma} \|W\sigma\|$.

Proof. Fix $\sigma \in \{-1, 1\}^V$. Then $W\sigma$ is a centered Gaussian vector. Its coordinate $(W\sigma)_x = D^{-1/2} \sum_{y: xy \in E} g_{xy} \sigma_y$ has variance 1, and for $x \neq y$ the coordinates $(W\sigma)_x$ and $(W\sigma)_y$ share only the variable g_{xy} , which is present if $xy \in E$. The covariance matrix of $W\sigma$ is $C_{\sigma} = I + \Delta_{\sigma} S \Delta_{\sigma}$, where Δ_{σ} is the diagonal matrix with entries σ_x . By (2.1), the eigenvalues $c_1, \dots, c_n$ of C_{σ} lie in $[0, 2]$, and $\sum_j c_j = \text{tr } C_{\sigma} = n$. Diagonalizing C_{σ} , we obtain

$$\mathbb{E} \exp \left(\frac{1}{8} \|W\sigma\|^2 \right) = \prod_{j=1}^n \left(1 - \frac{c_j}{4} \right)^{-1/2} \leq \exp \left(\sum_{j=1}^n \frac{c_j}{4} \right) = e^{n/4},$$

where we used that $-\log(1-s) \leq 2s$ for $s \in [0, 1/2]$. By Jensen's inequality,

$$\mathbb{E} \max_{\sigma} \|W\sigma\|^2 \leq 8 \log \mathbb{E} \max_{\sigma} e^{\|W\sigma\|^2/8} \leq 8 \log \sum_{\sigma} \mathbb{E} e^{\|W\sigma\|^2/8} \leq 8 \left(n \log 2 + \frac{n}{4} \right).$$

For the final claim, let $w \in [-1, 1]^V$. Since W is linear, we have $Ww = \mathbb{E} \pi_w W\sigma$, and by Jensen's inequality, $\|Ww\| \leq \mathbb{E} \pi_w \|W\sigma\| \leq \max_{\sigma} \|W\sigma\|$. $\square$

We now define the band. For the rest of this section, we write

$$C_{\beta, h} = 4(1 + h^2 + 8\beta^2 + \beta^4), \quad \varepsilon_D(\eta) = \frac{C_{\beta, h}}{\eta^2 D} + \frac{1 + (2T)^{1/2}}{\eta D^{1/2}} + \frac{T}{\eta D} \quad (\eta > 0), \quad (6.14)$$

and we note that $\varepsilon_D(\eta)$ depends only on β, h, T, D , and η .

Lemma 6.7. *Let $G \in \mathcal{G}_D$, $\beta \geq 0$, and $h \in \mathbb{R}$, and let a query history of depth T on G with center m be given. Let Π and $\bar{W}$ be as in Section 6.2, and let r be as in (6.9). For $\eta > 0$, let B_{η} be the set of $\sigma \in \{-1, 1\}^V$ such that $v = \sigma - m$ satisfies*

$$|\langle r, v \rangle| \leq \eta n, \quad \langle v, |\bar{W}|v \rangle \leq \eta n, \quad \langle m \odot v, |S|(m \odot v) \rangle \leq \eta n, \quad \|\Pi^{\perp} a(v)\|^2 \leq \eta n. \quad (6.15)$$

Then $\{\sigma \in B_{\eta}\} \in \mathcal{H}$ for every σ , and $p_{\eta} = \pi_m(B_{\eta})$ satisfies $\mathbb{E}(1 - p_{\eta}) \leq \varepsilon_D(\eta)$.

Proof. By Lemma 6.4(a), the vectors m, Wm , and r ; the projection Π ; and the matrix $\bar{W}$ are $\mathcal{H}$ -measurable, and $|\bar{W}|$ is a continuous function of $\bar{W}$. This proves the measurability claim, and it shows that p_{η} is $\mathcal{H}$ -measurable.

Fix a realization. Under π_m , the coordinates of v are independent and centered, and v_x has variance κ_x . We bound the probability that each condition in (6.15) fails. First,

$\mathbb{E}_{\pi_m} \langle r, v \rangle^2 = \sum_x \kappa_x r_x^2$, and Chebyshev's inequality bounds the probability of the first failure by $\sum_x \kappa_x r_x^2 / (\eta n)^2$. Second, since $|\bar{W}|$ is positive semidefinite,

$$\mathbb{E}_{\pi_m} \langle v, |\bar{W}| v \rangle = \sum_x |\bar{W}|_{xx} \kappa_x \leq \text{tr } |\bar{W}| = \|\bar{W}\|_{\text{tr}},$$

and Markov's inequality bounds the probability of the second failure by $\|\bar{W}\|_{\text{tr}} / (\eta n)$. Third, in the same way,

$$\mathbb{E}_{\pi_m} \langle m \odot v, |S|(m \odot v) \rangle = \sum_x |S|_{xx} m_x^2 \kappa_x \leq \text{tr } |S| \leq (n \text{tr } S^2)^{1/2} = \frac{n}{D^{1/2}},$$

where we used the Cauchy–Schwarz inequality for the eigenvalues of S and (2.1). Fourth, the entries of $a(v)$ are centered and uncorrelated under π_m . Indeed, for distinct edges e, e' , some vertex x belongs to exactly one of them, and the factor v_x in $a(v)_e a(v)_{e'}$ is centered and independent of the other factors. Further, $\mathbb{E}_{\pi_m} a(v)_{xy}^2 = S_{xy} \kappa_x \kappa_y \leq D^{-1}$. In other words, the covariance matrix of $a(v)$ is diagonal with entries at most D^{-1} , and

$$\mathbb{E}_{\pi_m} \|\Pi^\perp a(v)\|^2 = \sum_{e \in E} (\Pi^\perp)_{ee} \mathbb{E}_{\pi_m} a(v)_e^2 \leq \frac{\text{tr } \Pi^\perp}{D} = \frac{\text{rank } \Pi^\perp}{D}.$$

By the union bound, we conclude that

$$1 - p_\eta \leq \frac{\sum_x \kappa_x r_x^2}{\eta^2 n^2} + \frac{\|\bar{W}\|_{\text{tr}}}{\eta n} + \frac{1}{\eta D^{1/2}} + \frac{\text{rank } \Pi^\perp}{\eta n D}. \quad (6.16)$$

It remains to bound the expectation of the first term. Let $s_x = \text{atanh}(m_x)$. Then $\kappa_x = \text{sech}^2 s_x \leq 4e^{-2|s_x|}$ and $\kappa_x s_x^2 \leq 4 \sup_{s \geq 0} s^2 e^{-2s} = 4e^{-2} \leq 1$. This bound is uniform in $m_x \in (-1, 1)$. Using $(c_1 + c_2 + c_3 + c_4)^2 \leq 4(c_1^2 + c_2^2 + c_3^2 + c_4^2)$, $\kappa_x \leq 1$, $|m_x| \leq 1$, and $0 \leq (S\kappa)_x \leq 1$, we obtain from (6.9) that

$$\kappa_x r_x^2 \leq 4(\beta^2 (Wm)_x^2 + h^2 + \kappa_x s_x^2 + \beta^4) \leq 4(\beta^2 (Wm)_x^2 + h^2 + 1 + \beta^4).$$

Summing over x and using Lemma 6.6, which gives $\mathbb{E} \|Wm\|^2 \leq \mathbb{E} \max_\sigma \|W\sigma\|^2 \leq 8n$, we find

$$\mathbb{E} \sum_x \kappa_x r_x^2 \leq 32\beta^2 n + 4n(h^2 + 1 + \beta^4) = C_{\beta, h} n.$$

We take expectations in (6.16), and we use this bound together with Lemma 6.4(c). Since every D -regular simple graph has $n \geq D + 1$ vertices, we obtain $\mathbb{E}(1 - p_\eta) \leq \varepsilon_D(\eta)$. $\square$

On the band, the error terms in Lemma 6.5 cost at most a multiple of ηn . To state this, we set

$$c_\beta = 1 + \frac{\beta}{2} + \frac{3\beta^2}{2}, \quad (6.17)$$

and on the event $\{p_\eta > 0\}$ we define the function

$$L_\eta(y) = \mathbb{E}_{\pi_m} \left[\exp \left(\beta \langle y, \Pi a(v) \rangle - \frac{\beta^2}{2} \|\Pi a(v)\|^2 \right) \middle| B_\eta \right] \quad (y \in \mathbb{R}^E). \quad (6.18)$$

On $\{p_\eta = 0\}$, we set $L_\eta(y) = 1$ for all y . It depends on y only through Πy , and $(\omega, y) \mapsto L_\eta(y)$ is $\mathcal{H} \otimes \mathcal{B}(\mathbb{R}^E)$ -measurable on $\{p_\eta > 0\} \times \mathbb{R}^E$, because it is a finite sum of continuous functions of (m, Π, y) multiplied by $\mathcal{H}$ -measurable indicators.

Lemma 6.8. *Under the assumptions of Lemma 6.7, for every $\eta > 0$ we have, on the event $\{p_\eta > 0\}$,*

$$\log Z_G \geq \Theta_G(m) - c_\beta \eta n + \log p_\eta + \log L_\eta(g).$$

Proof. We apply Lemma 6.5 with the projection Π of Section 6.2, and we restrict the expectation in (6.10) to $\sigma \in B_\eta$; this decreases it, since the integrand is positive. Since a symmetric matrix K satisfies $-|K| \leq K \leq |K|$ in the order of quadratic forms, we have, for $\sigma \in B_\eta$,

$$\begin{aligned} \langle r, v \rangle &\geq -\eta n, & \frac{\beta}{2} \langle v, \bar{W}v \rangle &\geq -\frac{\beta}{2} \eta n, \\ \beta^2 \langle m \odot v, S(m \odot v) \rangle &\geq -\beta^2 \eta n, & -\frac{\beta^2}{2} \|\Pi^\perp a(v)\|^2 &\geq -\frac{\beta^2}{2} \eta n. \end{aligned}$$

The sum of the right sides is $-c_\beta \eta n$. Therefore

$$Z_G \geq e^{\Theta_G(m) - c_\beta \eta n} \mathbb{E}_{\pi_m} \left[\mathbf{1}_{B_\eta}(\sigma) \exp \left(\beta \langle g, \Pi a(v) \rangle - \frac{\beta^2}{2} \|\Pi a(v)\|^2 \right) \right] = e^{\Theta_G(m) - c_\beta \eta n} p_\eta L_\eta(g).$$

Taking logarithms completes the proof. $\square$

6.5 Relative entropy of the restricted and projected model

In this subsection, S , E , β , and n are as in Definition 6.1. Since S has zero diagonal and unit row sums, $\sum_{\{x,y\} \in E} S_{xy} = n/2$. Since $|v_x| \leq 1 + |m_x| \leq 2$ for $v = \sigma - m$, we have

$$\|a(v)\|^2 = \sum_{\{x,y\} \in E} S_{xy} v_x^2 v_y^2 \leq 16 \sum_{\{x,y\} \in E} S_{xy} = 8n \quad (6.19)$$

for every $m \in [-1, 1]^n$ and $\sigma \in \{-1, 1\}^n$.

Lemma 6.9. *The function $m \mapsto \mathcal{D}(m)$ is continuous on $[-1, 1]^n$. Further, let $m \in [-1, 1]^n$, let Π be an orthogonal projection of $\mathbb{R}^E$, and let $B \subset \{-1, 1\}^n$ satisfy $p = \pi_m(B) > 0$. Define*

$$L_{B,\Pi}(y) = \mathbb{E}_{\pi_m} \left[\exp \left(\beta \langle y, \Pi a(v) \rangle - \frac{\beta^2}{2} \|\Pi a(v)\|^2 \right) \middle| B \right] \quad (y \in \mathbb{R}^E).$$

Then $Q_{B,\Pi} = L_{B,\Pi}P_0$ is the law of $\beta\Pi a(v^*) + Z$, where σ^* has law $\pi_m(\cdot \mid B)$, $v^* = \sigma^* - m$, and Z is a standard Gaussian vector in $\mathbb{R}^E$ independent of σ^* . Further,

$$p \text{KL}(Q_{B,\Pi} \parallel P_0) \leq \mathcal{D}(m) - p \log p - (1-p) \log(1-p) \leq \mathcal{D}(m) + \log 2, \quad (6.20)$$

with the convention $0 \log 0 = 0$.

Proof. The identification of $Q_{B,\Pi}$ follows from the formula for the density of a Gaussian law with identity covariance, as in (6.2). By (6.19), we have

$$\exp(-\beta(8n)^{1/2}\|y\| - 4\beta^2 n) \leq L_{B,\Pi}(y) \leq \exp(\beta(8n)^{1/2}\|y\|), \quad (6.21)$$

and in particular all relative entropies in this proof are finite. For the continuity claim, the density L_m of Q_m , given by (6.2), is continuous in (m, y) , and by (6.21) with $B = \{-1, 1\}^n$ and $\Pi = I$,

$$|L_m(y) \log L_m(y)| \leq (\beta(8n)^{1/2}\|y\| + 4\beta^2 n) e^{\beta(8n)^{1/2}\|y\|}.$$

The right side does not depend on m and is integrable under P_0 . By dominated convergence, $\mathcal{D}(m) = \int L_m \log L_m dP_0$ is continuous in m .

We now prove (6.20), first for $\Pi = I$. We write $L_B = L_{B,I}$ and $Q_B = Q_{B,I}$. If $p = 1$, then $Q_B = Q_m$ and there is nothing to prove. If $p < 1$, then $L_m = pL_B + (1-p)L_{B^c}$, where L_{B^c} is defined in the same way. Since $L_m \geq pL_B$, we have $\log L_B \leq \log L_m - \log p$ and

$$\text{KL}(Q_B \parallel P_0) = \int L_B \log L_B dP_0 \leq \int L_B \log L_m dP_0 - \log p.$$

Since $L_m \geq (1-p)L_{B^c}$, we also have

$$\int L_{B^c} \log L_m dP_0 \geq \log(1-p) + \int L_{B^c} \log L_{B^c} dP_0 \geq \log(1-p),$$

because the last integral is a relative entropy and is nonnegative. Then

$$\begin{aligned} \mathcal{D}(m) &= \int L_m \log L_m dP_0 = p \int L_B \log L_m dP_0 + (1-p) \int L_{B^c} \log L_m dP_0 \\ &\geq p \int L_B \log L_m dP_0 + (1-p) \log(1-p). \end{aligned}$$

Combining the last three displays, we obtain (6.20) for $\Pi = I$. The second inequality in (6.20) holds because the binary entropy is at most $\log 2$.

For a general Π , it suffices to show $\text{KL}(Q_{B,\Pi} \parallel P_0) \leq \text{KL}(Q_B \parallel P_0)$. This is the data-processing inequality for the Markov kernel $y \mapsto \Pi y + \Pi^\perp z'$, with z' an independent standard Gaussian vector, which maps P_0 to itself and Q_B to $Q_{B,\Pi}$; we give the short direct argument. Under P_0 , the vectors Πy and $\Pi^\perp y$ are independent, and $\langle \Pi^\perp y, a \rangle$ is a centered Gaussian variable with variance $\|\Pi^\perp a\|^2$ for each $a \in \mathbb{R}^E$. Since $\langle y, a \rangle =$

$\langle \Pi y, \Pi a \rangle + \langle \Pi^\perp y, \Pi^\perp a \rangle$ and $\|a\|^2 = \|\Pi a\|^2 + \|\Pi^\perp a\|^2$, integrating out $\Pi^\perp y$ in each term of the finite average defining L_B gives

$$\mathbb{E}_{P_0}[L_B(y) \mid \Pi y] = L_{B, \Pi}(y).$$

By the convexity of $s \mapsto s \log s$ on $[0, \infty)$ and the conditional Jensen inequality, we have

$$\text{KL}(Q_{B, \Pi} \parallel P_0) = \mathbb{E}_{P_0}[L_{B, \Pi} \log L_{B, \Pi}] \leq \mathbb{E}_{P_0}[L_B \log L_B] = \text{KL}(Q_B \parallel P_0).$$

This completes the proof. $\square$

The next lemma converts the relative entropy of a Gaussian mixture with respect to the Gaussian law into the relative entropy in the reverse direction. We recall (F5): for every $k \geq 1$ and every probability measure ρ on $\mathbb{R}^k$, we have $W_2(\rho, \gamma_k)^2 \leq 2 \text{KL}(\rho \parallel \gamma_k)$, where γ_k is the standard Gaussian measure on $\mathbb{R}^k$ and W_2 is the quadratic Wasserstein distance.

Lemma 6.10. *Let $k \geq 1$ and $\ell \geq 0$, and let ν be a probability measure on $\{b \in \mathbb{R}^k : \|b\| \leq \ell\}$. Define*

$$L(y) = \int \exp\left(\langle y, b \rangle - \frac{1}{2}\|b\|^2\right) \nu(db) \quad (y \in \mathbb{R}^k),$$

and let $Q = L\gamma_k$. Then

$$0 \leq - \int \log L d\gamma_k = \text{KL}(\gamma_k \parallel Q) \leq \ell(2 \text{KL}(Q \parallel \gamma_k))^{1/2}.$$

Proof. The function L is positive and smooth, and $\nabla \log L(y) = \int b \nu_y(db)$, where ν_y is the probability measure proportional to $\exp(\langle y, b \rangle - \|b\|^2/2) \nu(db)$. Since ν_y is supported in the ball of radius ℓ , we have $\|\nabla \log L\| \leq \ell$, and $\log L$ is ℓ -Lipschitz. The measure Q is the law of $b + Z$ with $b \sim \nu$ and $Z \sim \gamma_k$ independent, so Q and γ_k have finite second moments, and $\log L$ is integrable under both. By definition, $\text{KL}(Q \parallel \gamma_k) = \int \log L dQ$ and $\text{KL}(\gamma_k \parallel Q) = - \int \log L d\gamma_k \geq 0$. Let (Y, Y') be a coupling of Q and γ_k . Then

$$\text{KL}(Q \parallel \gamma_k) + \text{KL}(\gamma_k \parallel Q) = \mathbb{E}[\log L(Y) - \log L(Y')] \leq \ell \mathbb{E}\|Y - Y'\| \leq \ell(\mathbb{E}\|Y - Y'\|^2)^{1/2}.$$

Taking the infimum over couplings and using (F5), we find that the left side is at most $\ell W_2(Q, \gamma_k) \leq \ell(2 \text{KL}(Q \parallel \gamma_k))^{1/2}$. Since $\text{KL}(Q \parallel \gamma_k) \geq 0$, the claim follows. $\square$

Corollary 6.11. *Under the assumptions of Lemma 6.9, we have*

$$\int \log L_{B, \Pi} dP_0 \geq -4\beta(n \text{KL}(Q_{B, \Pi} \parallel P_0))^{1/2} \geq -4\beta\left(\frac{n(\mathcal{D}(m) + \log 2)}{p}\right)^{1/2}.$$

Proof. By Lemma 6.9, $L_{B, \Pi}$ has the form of Lemma 6.10 with $k = |E|$ and ν the law of $\beta \Pi a(v^*)$. By (6.19) and $\|\Pi a\| \leq \|a\|$, we may take $\ell = \beta(8n)^{1/2}$. Then Lemma 6.10 implies the first inequality, since $\beta(8n)^{1/2} \cdot 2^{1/2} = 4\beta n^{1/2}$. The second inequality is (6.20). $\square$

6.6 Proof of the main result

We first prove a nonasymptotic version of Theorem 6.2. Recall the constants $\varepsilon_D(\eta)$ and c_β from (6.14) and (6.17).

Proposition 6.12. *Let $\beta \geq 0$, $h \in \mathbb{R}$, $D \geq 1$, and $T \geq 1$. Let $G \in \mathcal{G}_D$ have n vertices, let $\mathcal{D}$ be as in Definition 6.1 with $S = A_G/D$, and let a query history of depth T on G with center m be given. Then for every $\eta > 0$ and $\varepsilon \in (0, 1)$,*

$$p_G(\beta, h) \geq \frac{1}{n} \mathbb{E} \Theta_G(m) - c_\beta \eta - C'_{\beta, h} \left(\frac{\varepsilon_D(\eta)}{\varepsilon} \right)^{1/2} - 4\beta \left(\frac{n^{-1} \mathbb{E} \mathcal{D}(m) + n^{-1} \log 2}{1 - \varepsilon} \right)^{1/2} + \frac{\log(1 - \varepsilon)}{n}, \quad (6.22)$$

where $C'_{\beta, h} = \beta C_2^{1/2} + |h| + \log 2 + \beta^2/4$ and C_2 is the constant of (F6) with $p = 2$.

Proof. We have $|\log Z_G| \leq n \log 2 + \beta \max_\sigma |H_G(\sigma)| + |h|n$, which is integrable by (F6). Further, $|\text{ent}(m_x)| \leq \log 2$ and $0 \leq \langle \kappa, S\kappa \rangle \leq \sum_{x,y} S_{xy} = n$, since $\kappa \in [0, 1]^V$. Together with (F6), this gives the pathwise bound

$$|\Theta_G(m)| \leq \beta \max_\sigma |H_G(\sigma)| + n \left(|h| + \log 2 + \frac{\beta^2}{4} \right),$$

and, by the Minkowski inequality,

$$(\mathbb{E} \Theta_G(m)^2)^{1/2} \leq \beta (\mathbb{E} \max_\sigma H_G(\sigma)^2)^{1/2} + n \left(|h| + \log 2 + \frac{\beta^2}{4} \right) \leq C'_{\beta, h} n. \quad (6.23)$$

Let $\mathcal{G} = \{p_\eta \geq 1 - \varepsilon\}$, where $p_\eta = \pi_m(B_\eta)$ is as in Lemma 6.7. Then $\mathcal{G} \in \mathcal{H}$, and by Markov's inequality and Lemma 6.7, we have

$$\mathbb{P}(\mathcal{G}^c) = \mathbb{P}(1 - p_\eta > \varepsilon) \leq \frac{\varepsilon_D(\eta)}{\varepsilon}. \quad (6.24)$$

On $\mathcal{G}$, Lemma 6.8 gives

$$\mathbf{1}_{\mathcal{G}} \log Z_G \geq \mathbf{1}_{\mathcal{G}} (\Theta_G(m) - c_\beta \eta n + \log(1 - \varepsilon)) + \mathbf{1}_{\mathcal{G}} \log L_\eta(g). \quad (6.25)$$

We compute the expectation of the last term by conditioning on $\mathcal{H}$. Let $F(\omega, y) = \mathbf{1}_{\mathcal{G}}(\omega) \log L_\eta(y)$. By the discussion after (6.18), F is $\mathcal{H} \otimes \mathcal{B}(\mathbb{R}^E)$ -measurable, and by (6.21) we have $|F(\omega, y)| \leq \beta(8n)^{1/2} \|y\| + 4\beta^2 n$. In particular, $F(\cdot, g)$ is integrable. Since $\Pi \bar{g} = 0$ and $L_\eta(y)$ depends on y only through Πy , we have $L_\eta(\bar{g} + \Pi w) = L_\eta(w)$. Applying Lemma 6.4(b) to the positive and negative parts of F , we obtain

$$\mathbb{E}[\mathbf{1}_{\mathcal{G}} \log L_\eta(g)] = \mathbb{E}\left[\mathbf{1}_{\mathcal{G}} \int \log L_\eta \, d\mathbb{P}_0\right].$$

For each fixed $\omega \in \mathcal{G}$, the function L_η is the function $L_{B,\Pi}$ of Lemma 6.9, with the deterministic data $m(\omega)$, $\Pi(\omega)$, and $B = B_\eta(\omega)$, and with $p = p_\eta(\omega) \geq 1 - \varepsilon$. Then Corollary 6.11, applied with $\mathcal{D}(m)$ the value of the deterministic function $\mathcal{D}$ at $m(\omega)$, implies that

$$\mathbf{1}_{\mathcal{G}} \int \log L_\eta d\mathbb{P}_0 \geq -4\beta \left(\frac{n(\mathcal{D}(m) + \log 2)}{1 - \varepsilon} \right)^{1/2}.$$

Taking expectations and using Jensen's inequality for the square root, we find

$$\mathbb{E}[\mathbf{1}_{\mathcal{G}} \log L_\eta(g)] \geq -4\beta \left(\frac{n(\mathbb{E} \mathcal{D}(m) + \log 2)}{1 - \varepsilon} \right)^{1/2}. \quad (6.26)$$

On $\mathcal{G}^c$ we use a cruder bound. Since W has zero diagonal, the average of $\beta H_G(\sigma) + h \sum_x \sigma_x$ under the uniform law on $\{-1, 1\}^V$ is zero. By Jensen's inequality, we have $\log Z_G \geq n \log 2 \geq 0$. Combining this with (6.25) and (6.26), and using $\mathbb{P}(\mathcal{G}) \leq 1$ and $\log(1 - \varepsilon) < 0$, we obtain

$$\mathbb{E} \log Z_G \geq \mathbb{E}[\mathbf{1}_{\mathcal{G}} \Theta_G(m)] - c_\beta \eta n + \log(1 - \varepsilon) - 4\beta \left(\frac{n(\mathbb{E} \mathcal{D}(m) + \log 2)}{1 - \varepsilon} \right)^{1/2}.$$

Finally, by the Cauchy–Schwarz inequality, (6.23), and (6.24),

$$\mathbb{E}[\mathbf{1}_{\mathcal{G}} \Theta_G(m)] \geq \mathbb{E} \Theta_G(m) - \mathbb{P}(\mathcal{G}^c)^{1/2} (\mathbb{E} \Theta_G(m)^2)^{1/2} \geq \mathbb{E} \Theta_G(m) - C'_{\beta, h} n \left(\frac{\varepsilon_D(\eta)}{\varepsilon} \right)^{1/2}.$$

Dividing by n gives (6.22). $\square$

Proof of Theorem 6.2. Fix $\eta > 0$ and $\varepsilon \in (0, 1)$, and let b_k denote the sum of the four error terms in (6.22) for $G = G_k$, so that $p_{G_k}(\beta, h) \geq n_k^{-1} \mathbb{E} \Theta_{G_k}(m^{(k)}) - b_k$ by Proposition 6.12. As $k \rightarrow \infty$, we have $\varepsilon_{D_k}(\eta) \rightarrow 0$, because $D_k \rightarrow \infty$ and β, h, T , and η are fixed. Further, the terms involving $n_k^{-1} \log 2$ and $n_k^{-1} \log(1 - \varepsilon)$ tend to zero, since $n_k \geq D_k + 1 \rightarrow \infty$. Then (6.3) and the continuity of the square root imply that $\limsup_k b_k \leq c_\beta \eta + 4\beta(\delta/(1 - \varepsilon))^{1/2}$. By (6.23), the sequence $n_k^{-1} \mathbb{E} \Theta_{G_k}(m^{(k)})$ is bounded. Since $\liminf_k (a_k - b_k) \geq \liminf_k a_k - \limsup_k b_k$ whenever the right side is defined, we obtain

$$\liminf_{k \rightarrow \infty} p_{G_k}(\beta, h) \geq \liminf_{k \rightarrow \infty} \frac{1}{n_k} \mathbb{E} \Theta_{G_k}(m^{(k)}) - c_\beta \eta - 4\beta \left(\frac{\delta}{1 - \varepsilon} \right)^{1/2}.$$

Letting $\eta \downarrow 0$ and then $\varepsilon \downarrow 0$ gives (6.4). $\square$

Remark 6.13. The requirement that the center is itself a query is used through Lemma 6.4(a): the vector Wm appears in the linear coefficient r , which must be $\mathcal{H}$ -measurable for the band B_η and the function L_η to be $\mathcal{H}$ -measurable. If Wm were not observed, then $\langle \beta Wm, v \rangle$ would contain a part that is linear in the unobserved Gaussian vector, and Lemma 6.4(b) could not be applied with $\mathcal{H}$ -measurable L_η .

Relative entropy of the planted product model

The goal of this section is to prove Theorem 7.5, which states that the relative entropy $\mathcal{D}(m)$ of the planted product model of Definition 6.1 is small compared to n , provided that two conditions hold. First, the scalar transforms ψ_{m_x} of the coordinates of the center must be close to the transform ψ_μ of a single law μ , on average over the vertices and uniformly over a compact set of site profiles. Second, μ must satisfy $\Gamma_\mu \leq 0$ on $[0, 1]$. The bound holds for every deterministic center, and since its error terms depend on the center only through the discrepancy between the two transforms, it can be averaged over random centers. We also prove Proposition 7.4, which states that the law μ_+ of the Parisi center satisfies $\Gamma_{\mu_+}(s) \leq 0$ for all $s \geq 0$. Its proof uses only the optimality of the Parisi measure. Corollary 7.6 combines the two results.

The planted product model is a spiked Wigner model with a sparse variance profile and a prior that depends on the vertex. For the spiked Wigner model with independent and identically distributed noise, the limit of the relative entropy, or equivalently of the mutual information between the spike and the observation, was computed in [98], where earlier work is discussed, and for bounded centered unit-variance priors, El Alaoui, Krzakala, and Jordan determined the region in which the spiked and the unspiked models are contiguous [59]. Guionnet, Ko, Krzakala, and Zdeborová computed this limit for inhomogeneous noise whose inverse variances form a positive-semidefinite profile with a block structure or a regular limit [78]. Behne and Reeves computed it when the signal-to-noise ratio is constant on each of finitely many blocks, for every nonnegative matrix of block ratios [23]. On the Nishimori line, multispecies spin glasses are spiked models with binary signals and a block variance profile. Alberici, Camilli, Contucci, and Mingione computed their free energy when the interaction is elliptic [4], and for the deep Boltzmann machine, whose interaction matrix is not positive semidefinite [3]. For general variance profiles and certain priors with independent and identically distributed symmetric coordinates of variance at most one, De and Kunisky proved that strong detection is impossible below a spectral threshold; their assumptions include a 1-moment-subgaussian condition on the product of two independent samples of the prior and, in our setting, require D to be of order n [52, Theorem 1.11].

Interpolations of Guerra type [76] between a matrix channel and scalar channels were used by Korada and Macris [94] for p -spin models on the Nishimori line and by Krzakala, Xu, and Zdeborová [97] for spiked models. Barbier and Macris made the interpolation

path adaptive [14]. Our proof of Theorem 7.5 follows the Hamilton–Jacobi approach of Mourrat [104, 105], in which the relative entropy of a family of interpolating channels is compared with the solution of a Hamilton–Jacobi equation given by a Hopf–Lax formula. For inference of matrix tensor products, Chen and Xia showed that the solution of the corresponding Hamilton–Jacobi equation is an upper bound on the limit of the free energy for every interaction matrix, also when the nonlinearity is not convex [35]. We need only an upper bound, the entries of S are of order at most $1/D$, and S need not be positive semidefinite. For the SK model, the function Γ_μ coincides with a function introduced by Chen, Panchenko, and Subag in their study of the TAP free energy [41]; see the discussion after (7.8).

Section 7.1 collects identities for Gaussian channels. Section 7.2 defines the scalar transform and proves Lemma 7.2. Section 7.3 proves Proposition 7.4, and Section 7.4 states Theorem 7.5 and Corollary 7.6. The remaining subsections prove Theorem 7.5. In Section 7.5, we interpolate between independent scalar channels at the vertices and the planted edge channel. We compare the normalized relative entropy $F(t, q)$ of the interpolating channel with a scalar Hopf–Lax formula, which is treated in Section 7.9. The comparison uses only the behavior of F at points where it is touched from above by a function with bounded Hessian. At such points, the time derivative of F is close to a quadratic form in the gradient of F (Proposition 7.13). The proof of this estimate rests on an exact covariance identity for the derivative of the quenched relative entropy (Section 7.6) and on concentration over a family of perturbations of small metric entropy (Section 7.7). Section 7.10 completes the proofs.

Throughout this section, S is a symmetric $n \times n$ matrix with nonnegative entries, zero diagonal, and unit row sums, as in Definition 6.1. We index its rows and columns by a set V with $|V| = n$, and we set $E = \{\{x, y\} : S_{xy} > 0\}$. We do not assume that S is positive semidefinite; see Remark 7.7. For $w \in \mathbb{R}^V$ we write $|w|$ for the Euclidean norm, $\|w\|_n = n^{-1/2}|w|$ for the normalized Euclidean norm, and $\text{dist}_n(w, \mathcal{A}) = \inf_{z \in \mathcal{A}} \|w - z\|_n$ for $\mathcal{A} \subset \mathbb{R}^V$. We write $\mathbf{1}$ for the vector with all coordinates equal to one, $w \cdot w'$ for the Euclidean inner product, $w \odot w'$ for the coordinatewise product, and $w^{\odot 2} = w \odot w$. For a set $\mathcal{A} \subset \mathbb{R}^V$ and $c \in \mathbb{R}$ we write $c\mathcal{A} = \{cw : w \in \mathcal{A}\}$ and $S\mathcal{A} = \{Sw : w \in \mathcal{A}\}$. If $\max_{x,y} S_{xy} \leq 1/D$ for some $D \geq 1$, then

$$\|S\| \leq 1, \quad \text{tr } S^2 \leq \frac{n}{D}, \quad \max_{x \in V} (S^2)_{xx} \leq \frac{1}{D}, \quad n \geq D + 1. \quad (7.1)$$

Indeed, the bound on the operator norm follows from Jensen’s inequality as in (2.1), and $(S^2)_{xx} = \sum_y S_{xy}^2 \leq D^{-1} \sum_y S_{xy} = D^{-1}$; summing over x gives the bound on the trace. Since each row of S has unit sum and entries at most $1/D$, it has at least D nonzero entries, none of them on the diagonal.

7.1 Gaussian channels

We begin with identities relating relative entropy and posterior moments for Gaussian channels. The first derivative formula in Lemma 7.1 is the relation of Guo, Shamai, and Verdú [79] between mutual information and the minimum mean-square error, written for relative entropy. The second derivatives in Lemma 7.1(c) are a special case of the Hessian formula of Payaró and Palomar [120, Theorem 5]. We include the short proofs for completeness.

Let $N \geq 1$, let θ be a random vector in $\mathbb{R}^N$ that takes finitely many values, and let Z be a standard Gaussian vector in $\mathbb{R}^N$ independent of θ . For $\lambda \in [0, \infty)^N$, we observe $Y^\lambda = (\sqrt{\lambda_j} \theta_j + Z_j)_{j \leq N}$. Let θ' be an independent copy of θ , and let $\mathbb{E}'$ denote the expectation over θ' only. Then the law of Y^λ has density

$$L_\lambda(y) = \mathbb{E}' \exp \left(\sum_{j \leq N} (\sqrt{\lambda_j} y_j \theta'_j - \lambda_j (\theta'_j)^2 / 2) \right) \quad (y \in \mathbb{R}^N) \quad (7.2)$$

with respect to the standard Gaussian law $N(0, I_N)$, and we define

$$I(\lambda) = \text{KL}(\text{Law}(Y^\lambda) \parallel N(0, I_N)) = \mathbb{E} \log L_\lambda(Y^\lambda).$$

By Bayes's rule, the conditional law of θ given $Y^\lambda = y$ is the law of θ' reweighted by the exponential in (7.2). Given Y^λ , let $\theta^1, \theta^2, \dots$ be conditionally independent samples from this law, and write $\langle \cdot \rangle_\lambda$ for the average over them; for instance, $\langle \theta_j^1 \rangle_\lambda = \mathbb{E}[\theta_j \mid Y^\lambda]$. Since θ and θ^1 have the same conditional law given Y^λ , we have the Bayesian form of the Nishimori identity [110]

$$\mathbb{E} \langle f(\theta, \theta^1, \dots, \theta^k, Y^\lambda) \rangle_\lambda = \mathbb{E} \langle f(\theta^{k+1}, \theta^1, \dots, \theta^k, Y^\lambda) \rangle_\lambda \quad (7.3)$$

for all $k \geq 0$ and all bounded measurable f . We use (7.3) only for deterministic λ . Substituting $Y^\lambda = (\sqrt{\lambda_j} \theta_j + Z_j)_j$, we obtain

$$I(\lambda) = \mathbb{E} \log \mathbb{E}' \exp \left(\sum_{j \leq N} (\sqrt{\lambda_j} Z_j \theta'_j + \lambda_j \theta_j \theta'_j - \lambda_j (\theta'_j)^2 / 2) \right), \quad (7.4)$$

and $\langle \cdot \rangle_\lambda$ is the average with weights proportional to the exponential in (7.4), with θ' replaced by the replica. We regard it as a function of (θ, Z) . Differentiating these weights, we find that for a function f of one replica,

$$\begin{aligned} \partial_{Z_j} \langle f(\theta^1) \rangle_\lambda &= \sqrt{\lambda_j} (\langle f(\theta^1) \theta_j^1 \rangle_\lambda - \langle f(\theta^1) \rangle_\lambda \langle \theta_j^1 \rangle_\lambda), \\ \partial_{\lambda_j} \langle f(\theta^1) \rangle_\lambda &= \langle f(\theta^1) A_j(\theta^1) \rangle_\lambda - \langle f(\theta^1) \rangle_\lambda \langle A_j(\theta^1) \rangle_\lambda, \end{aligned} \quad (7.5)$$

where the second identity holds for $\lambda_j > 0$ and

$$A_j(\theta^1) = \frac{Z_j \theta_j^1}{2\sqrt{\lambda_j}} + \theta_j \theta_j^1 - \frac{(\theta_j^1)^2}{2}.$$

Lemma 7.1. Let N , θ , and I be as above, and let $c_\theta = \max_j \|\theta_j\|_\infty$.

(a) The function I is nonnegative and continuous on $[0, \infty)^N$, $I(0) = 0$, and I is nondecreasing in each coordinate.

(b) Let $j \leq N$. At every $\lambda \in [0, \infty)^N$ with $\lambda_j > 0$, the partial derivative $\partial_{\lambda_j} I(\lambda)$ exists and

$$\partial_{\lambda_j} I(\lambda) = \frac{1}{2} \mathbb{E} \langle \theta_j^1 \rangle_\lambda^2 \in \left[0, \frac{1}{2} \mathbb{E} \theta_j^2\right].$$

The function $\lambda \mapsto \mathbb{E} \langle \theta_j^1 \rangle_\lambda^2$ is continuous on $[0, \infty)^N$.

(c) Let $j, k \leq N$. At every $\lambda \in [0, \infty)^N$ with $\lambda_j > 0$ and $\lambda_k > 0$,

$$\partial_{\lambda_k} \partial_{\lambda_j} I(\lambda) = \frac{1}{2} \mathbb{E} (\langle \theta_j^1 \theta_k^1 \rangle_\lambda - \langle \theta_j^1 \rangle_\lambda \langle \theta_k^1 \rangle_\lambda)^2,$$

and the right side is continuous on $[0, \infty)^N$.

Proof. (a) The exponent in (7.4) is bounded in absolute value by $\sum_j (c_\theta \sqrt{\lambda_j} |Z_j| + 3c_\theta^2 \lambda_j / 2)$, and so is the logarithm of its average over θ' . Since this bound is integrable and locally uniform in λ , and the integrand is continuous in λ , the function I is continuous by dominated convergence. Further, $I \geq 0$ because it is a relative entropy, and $I(0) = 0$. Monotonicity follows from (b) and continuity.

(b) Let $\lambda_j > 0$. For λ' in a neighborhood of λ on which $\lambda'_j \geq \lambda_j/2$, the derivative in λ_j of the integrand in (7.4) is $\langle A_j(\theta^1) \rangle_{\lambda'}$, which is bounded by $C(1 + |Z_j|)$ with C depending on c_θ and λ_j . We may therefore differentiate under the expectation, and $\partial_{\lambda_j} I = \mathbb{E} \langle A_j(\theta^1) \rangle_\lambda$. By (7.5) and Gaussian integration by parts (F1) in Z_j , which is independent of θ and of $(Z_k)_{k \neq j}$,

$$\mathbb{E} [Z_j \langle \theta_j^1 \rangle_\lambda] = \sqrt{\lambda_j} \mathbb{E} [\langle (\theta_j^1)^2 \rangle_\lambda - \langle \theta_j^1 \rangle_\lambda^2].$$

Inserting this identity into the expression for the derivative above, we obtain

$$\partial_{\lambda_j} I = \frac{1}{2} \mathbb{E} \langle (\theta_j^1)^2 \rangle_\lambda - \frac{1}{2} \mathbb{E} \langle \theta_j^1 \rangle_\lambda^2 + \mathbb{E} [\theta_j \langle \theta_j^1 \rangle_\lambda] - \frac{1}{2} \mathbb{E} \langle (\theta_j^1)^2 \rangle_\lambda = \mathbb{E} [\theta_j \langle \theta_j^1 \rangle_\lambda] - \frac{1}{2} \mathbb{E} \langle \theta_j^1 \rangle_\lambda^2.$$

By (7.3), $\mathbb{E} [\theta_j \langle \theta_j^1 \rangle_\lambda] = \mathbb{E} \langle \theta_j^2 \theta_j^1 \rangle_\lambda = \mathbb{E} \langle \theta_j^1 \rangle_\lambda^2$, which gives the formula. Since $\langle \theta_j^1 \rangle_\lambda = \mathbb{E} [\theta_j | Y^\lambda]$, we have $\mathbb{E} \langle \theta_j^1 \rangle_\lambda^2 \leq \mathbb{E} \theta_j^2$ by Jensen's inequality. For fixed (θ, Z) , the average $\langle \theta_j^1 \rangle_\lambda$ is a continuous function of $\lambda \in [0, \infty)^N$ bounded by c_θ , and the continuity follows by dominated convergence.

(c) Let $\lambda_j, \lambda_k > 0$. By (b) and (7.3), $2\partial_{\lambda_j} I(\lambda') = \mathbb{E} [\theta_j \langle \theta_j^1 \rangle_{\lambda'}]$ for all λ' near λ . We differentiate the right side in λ_k , which is justified as in (b). For brevity, we write

$b_l = \langle \theta_l^1 \rangle_\lambda$, $c_{jk} = \langle \theta_j^1 \theta_k^1 \rangle_\lambda$, and $c_{jkk} = \langle \theta_j^1 (\theta_k^1)^2 \rangle_\lambda$. By (7.5), we have $2\partial_{\lambda_k} \partial_{\lambda_j} I = T_1 + T_2 + T_3$, where

$$T_1 = \frac{1}{2\sqrt{\lambda_k}} \mathbb{E}[\theta_j Z_k (c_{jk} - b_j b_k)], \quad T_2 = \mathbb{E}[\theta_j \theta_k (c_{jk} - b_j b_k)],$$

$$T_3 = -\frac{1}{2} \mathbb{E}[\theta_j (c_{jkk} - b_j c_{kk})].$$

By (7.5), we have $\partial_{Z_k} (c_{jk} - b_j b_k) = \sqrt{\lambda_k} (c_{jkk} - b_j c_{kk} - 2c_{jk} b_k + 2b_j b_k^2)$, and Gaussian integration by parts in Z_k gives

$$T_1 = \frac{1}{2} \mathbb{E}[\theta_j (c_{jkk} - b_j c_{kk})] - \mathbb{E}[\theta_j (c_{jk} b_k - b_j b_k^2)].$$

The quantities b_l , c_{jk} , and c_{jkk} are functions of Y^λ , and $\mathbb{E}[\theta_j | Y^\lambda] = b_j$ and $\mathbb{E}[\theta_j \theta_k | Y^\lambda] = c_{jk}$. Conditioning on the observation, we find

$$T_1 + T_3 = -\mathbb{E}[b_j c_{jk} b_k - b_j^2 b_k^2], \quad T_2 = \mathbb{E}[c_{jk}^2 - c_{jk} b_j b_k],$$

and $T_1 + T_2 + T_3 = \mathbb{E}(c_{jk} - b_j b_k)^2$. The continuity follows as in (b). $\square$

7.2 The scalar transform

For $m \in [-1, 1]$ let ν_m be the law of $\sigma - m$, where $\sigma \in \{-1, 1\}$ has mean m . The law ν_m gives mass $(1+m)/2$ to $1-m$ and mass $(1-m)/2$ to $-1-m$; it has mean zero and variance $\kappa(m) = 1 - m^2$, and it is the point mass at zero if $m = \pm 1$. For $r \geq 0$ the scalar transform is

$$\psi_m(r) = \mathbb{E} \log \mathbb{E}_{\tau \sim \nu_m} \exp \left(\sqrt{r} Z \tau + r \tau \tau^* - \frac{r}{2} \tau^2 \right), \quad (7.6)$$

where the outer expectation is over independent $\tau^* \sim \nu_m$ and $Z \sim N(0, 1)$. By (7.4), $\psi_m(r)$ is the function I for $N = 1$ and $\theta = \tau^*$; that is, $\psi_m(r) = \text{KL}(\text{Law}(\sqrt{r} \tau^* + Z) \| N(0, 1))$. For a probability measure μ on $[-1, 1]$ we write

$$\psi_\mu(r) = \int \psi_m(r) \mu(dm), \quad \Gamma_\mu(s) = 2 \left[\psi_\mu(\beta^2 s) - \frac{\beta^2 s^2}{4} \right] \quad (r, s \geq 0).$$

Lemma 7.2. *Let $m \in [-1, 1]$, and for $r \geq 0$ let $b_{m,r} = \mathbb{E}[\tau^* | \sqrt{r} \tau^* + Z]$, with τ^* and Z as in (7.6).*

(a) *We have $\psi_m \geq 0$ and $\psi_m(0) = 0$.*

(b) *The function ψ_m is continuously differentiable on $[0, \infty)$, with a right derivative at zero, and*

$$\psi'_m(r) = \frac{1}{2} \mathbb{E} b_{m,r}^2 \in \left[0, \frac{\kappa(m)}{2} \right].$$

(c) The function ψ_m is nondecreasing and $1/2$ -Lipschitz.

(d) The function $(m, r) \mapsto \psi_m(r)$ is continuous on $[-1, 1] \times [0, \infty)$.

Proof. Parts (a) and (b) are Lemma 7.1(a,b) with $N = 1$ and $\theta = \tau^*$, since $\mathbb{E}(\tau^*)^2 = \kappa(m)$; at $r = 0$, the right derivative exists and equals $\lim_{r \downarrow 0} \psi'_m(r)$ by the mean value theorem. Part (c) follows from (b), since $\kappa(m) \leq 1$.

For (d), let $r \geq 0$ and

$$c_{\pm} = \pm\sqrt{r}(1 \mp m), \quad \ell_{m,r}(w) = \frac{1+m}{2} e^{c_+ w - c_+^2/2} + \frac{1-m}{2} e^{c_- w - c_-^2/2} \quad (w \in \mathbb{R}).$$

By (7.2), the function $\ell_{m,r}$ is the density of the law of $\sqrt{r}\tau^* + Z$ with respect to $N(0, 1)$, and $\psi_m(r) = \mathbb{E}[\ell_{m,r}(Z) \log \ell_{m,r}(Z)]$. The function $(m, r, w) \mapsto \ell_{m,r}(w)$ is positive and continuous. Since $\ell_{m,r}(w)$ is an average of the numbers $e^{c_{\pm} w - c_{\pm}^2/2}$ and $|c_{\pm}| \leq 2\sqrt{r}$, we have $|\log \ell_{m,r}(w)| \leq 2\sqrt{r}|w| + 2r$ and

$$|\ell_{m,r}(w) \log \ell_{m,r}(w)| \leq (2\sqrt{r}|w| + 2r) e^{2\sqrt{r}|w|}.$$

For r in a bounded interval, the right side is bounded by a function of w that does not depend on (m, r) and is integrable under $N(0, 1)$. Then (d) follows by dominated convergence. $\square$

By Lemma 7.2, for each probability measure μ on $[-1, 1]$, the function ψ_{μ} is non-negative and nondecreasing, $\psi_{\mu}(0) = 0$, and ψ_{μ} is continuously differentiable on $[0, \infty)$ with

$$\psi'_{\mu}(r) = \int \psi'_m(r) \mu(dm) \in \left[0, \frac{1}{2} \int \kappa(m) \mu(dm)\right] \subset \left[0, \frac{1}{2}\right]. \quad (7.7)$$

Indeed, $m \mapsto \psi'_m(u)$ is measurable, being a limit of difference quotients of continuous functions of m , and $\psi_{\mu}(r) = \int_0^r \int \psi'_m(u) \mu(dm) du$ by Fubini's theorem, where the inner integral is continuous in u by dominated convergence. Since $\Gamma_{\mu}(0) = 0$, we obtain

$$\Gamma_{\mu}(s) = \beta^2 \int_0^s (2\psi'_{\mu}(\beta^2 u) - u) du \quad (s \geq 0). \quad (7.8)$$

Since $\psi_{-m} = \psi_m$, the function Γ_{μ} depends only on the law of $|m|$ under μ . For the SK model, it is the function Γ_{μ} defined by Chen, Panchenko, and Subag in [41, (3.43)–(3.44)], applied to this law; a short computation with (7.11) below identifies the two definitions. Let $q = \int m^2 \mu(dm)$, and suppose that $q < 1$. Chen, Panchenko, and Subag showed that $\Gamma_{\mu} \leq 0$ on $[0, 1 - q]$ if and only if the variational formula for their generalized TAP correction at the law of $|m|$ is minimized by the point mass at zero [41, Proposition 13]. In this case, the generalized TAP correction is replica symmetric and equals the classical TAP correction, and Plefka's condition $\beta^2 \int (1 - m^2)^2 \mu(dm) \leq 1$ holds [41, Propositions 13 and 14].

7.3 The scalar criterion for the Parisi center

Let $\beta > 0$ and $h \geq 0$. We use the notation of Section 2.3: $\nu = \nu_{\beta,h}$ is the Parisi measure, $\nu(t) = \nu([0, t])$, $\Phi = \Phi_\nu$, $q_+ = \max \text{supp } \nu$, X is the Parisi diffusion (2.8), $M_t = \partial_x \Phi(t, X_t)$, μ_+ is the law of M_{q_+} , and Ψ is the function (2.11). By (P4), $\partial_x \Phi(t, x) = \tanh x$ for $t \in [q_+, 1]$. We have $\nu(t) = 1$ for $t \geq q_+$, and (2.8) becomes

$$X_t = X_{q_+} + \beta^2 \int_{q_+}^t \tanh(X_u) du + \beta(B_t - B_{q_+}), \quad M_t = \tanh X_t \quad (t \in [q_+, 1]). \quad (7.9)$$

In particular, μ_+ is the law of $\tanh X_{q_+}$, and $\mu_+((-1, 1)) = 1$. By Corollary 2.4, we have $q_+ < 1$ and $\mathbb{E}M_{q_+}^2 = q_+$, which implies that

$$\int \kappa(m) \mu_+(dm) = 1 - \mathbb{E}M_{q_+}^2 = 1 - q_+. \quad (7.10)$$

The next lemma identifies the law of the Parisi diffusion above q_+ . On $[q_+, 1]$ the drift is $\beta^2 \tanh x$, and the diffusion is the Doob transform of Brownian motion with variance β^2 per unit time by the function $\cosh$.

Lemma 7.3. *Let $\beta > 0$ and $h \geq 0$, let $s \in [0, 1 - q_+]$, and let $g: \mathbb{R} \rightarrow \mathbb{R}$ be bounded and measurable. For $y \in \mathbb{R}$ let*

$$k_s(y) = \sum_{\sigma \in \{-1, 1\}} \frac{e^{\sigma y}}{2 \cosh y} \mathbb{E}g(y + \beta^2 s \sigma + \beta \sqrt{s} Z),$$

where $Z \sim N(0, 1)$. Then $\mathbb{E}g(X_{q_++s}) = \mathbb{E}k_s(X_{q_+})$.

Proof. For $s = 0$ there is nothing to prove. Let $s > 0$ and $T' = q_+ + s \leq 1$, and let $(\mathfrak{F}_t)$ be the filtration generated by B , to which X is adapted. For $t \in [q_+, T']$ let

$$\Lambda_t = \exp \left(-\beta \int_{q_+}^t \tanh(X_u) dB_u - \frac{\beta^2}{2} \int_{q_+}^t \tanh^2(X_u) du \right).$$

Since the integrand $\beta \tanh(X_u)$ is bounded, Novikov's condition holds (see [92, Chapter 3] for this condition and for Girsanov's theorem), and Λ is a martingale on $[q_+, T']$ with $\Lambda_{q_+} = 1$. Let $\mathbb{R}$ be the probability measure with density $\Lambda_{T'}$ with respect to $\mathbb{P}$. Since $\mathbb{E}[\Lambda_{T'} \mid \mathfrak{F}_{q_+}] = 1$, the measures $\mathbb{R}$ and $\mathbb{P}$ agree on $\mathfrak{F}_{q_+}$. By Girsanov's theorem, the process $\tilde{B}_t = B_t - B_{q_+} + \beta \int_{q_+}^t \tanh(X_u) du$, $t \in [q_+, T']$, is a Brownian motion under $\mathbb{R}$, independent of $\mathfrak{F}_{q_+}$. By (7.9), we have $X_t = X_{q_+} + \beta \tilde{B}_t$ for $t \in [q_+, T']$.

Next, (7.9) and Itô's formula give

$$\begin{aligned} \log \cosh X_{T'} - \log \cosh X_{q_+} &= \beta \int_{q_+}^{T'} \tanh(X_u) dB_u + \frac{\beta^2}{2} \int_{q_+}^{T'} (1 + \tanh^2(X_u)) du \\ &= -\log \Lambda_{T'} + \frac{\beta^2 s}{2}. \end{aligned}$$

Since $\Lambda_{T'} > 0$, we obtain

$$\mathbb{E}g(X_{T'}) = \mathbb{E}_{\mathbb{R}}[g(X_{T'})\Lambda_{T'}^{-1}] = e^{-\beta^2 s/2} \mathbb{E}_{\mathbb{R}}\left[\frac{(g \cosh)(X_{q_+} + \beta \tilde{B}_{T'})}{\cosh X_{q_+}}\right].$$

Under $\mathbb{R}$, the variable $\tilde{B}_{T'}$ has law $N(0, s)$ and is independent of X_{q_+} , whose law under $\mathbb{R}$ is its law under $\mathbb{P}$. Since $\cosh(x + z) \leq e^{|z|} \cosh x$, the expression inside the last expectation is bounded by $\|g\|_{\infty} e^{\beta|\tilde{B}_{T'}|}$. Then $\mathbb{E}g(X_{T'}) = \mathbb{E}k(X_{q_+})$, where

$$k(y) = \frac{e^{-\beta^2 s/2}}{\cosh y} \mathbb{E}[(g \cosh)(y + \beta \sqrt{s} Z)].$$

Finally, we write $\cosh$ as the average of $e^{\pm x}$ and use $\mathbb{E}[f(Z)e^{cZ}] = e^{c^2/2} \mathbb{E}f(Z + c)$ with $c = \pm\beta\sqrt{s}$, and we find $k = k_s$. $\square$

Proposition 7.4. *Let $\beta > 0$ and $h \geq 0$, and let μ_+ be the law of the Parisi center $M_{q_+} = \tanh X_{q_+}$. Then $\Gamma_{\mu_+}(s) \leq 0$ for every $s \geq 0$.*

Proof. Let $m = \tanh y \in (-1, 1)$ and $r \geq 0$, and let σ^* be independent of $Z \sim N(0, 1)$, with $\mathbb{P}(\sigma^* = \pm 1) = (1 \pm m)/2 = e^{\pm y}/(2 \cosh y)$. We claim that

$$2\psi'_m(r) = \mathbb{E} \tanh^2(y + \sqrt{r} Z + r\sigma^*) - m^2. \quad (7.11)$$

Indeed, $\tau^* = \sigma^* - m$ has law ν_m . Let $W = \sqrt{r}\tau^* + Z$. For $\sigma \in \{-1, 1\}$ and $\tau = \sigma - m$, we have

$$\sqrt{r}W\tau - \frac{r}{2}\tau^2 = \sigma(\sqrt{r}W + rm) - \sqrt{r}Wm - \frac{r}{2}(1 + m^2),$$

where we used $\sigma^2 = 1$. Since the last two terms do not depend on σ , Bayes's rule gives $\mathbb{P}(\sigma^* = \sigma \mid W) \propto e^{\sigma(y + \sqrt{r}W + rm)}$, and $\mathbb{E}[\sigma^* \mid W] = \tanh(y + \sqrt{r}W + rm) = \tanh(y + \sqrt{r}Z + r\sigma^*)$. Since $b_{m,r} = \mathbb{E}[\sigma^* \mid W] - m$ and $\mathbb{E}\mathbb{E}[\sigma^* \mid W] = m$, we obtain $\mathbb{E}b_{m,r}^2 = \mathbb{E} \tanh^2(y + \sqrt{r}Z + r\sigma^*) - m^2$, and Lemma 7.2(b) gives (7.11).

Let $s \in [0, 1 - q_+]$. We apply Lemma 7.3 with $g = \tanh^2$. By (7.11) with $r = \beta^2 s$, the function k_s of Lemma 7.3 is $k_s(y) = 2\psi'_{\tanh y}(\beta^2 s) + \tanh^2 y$. By (7.9) and (7.7), we obtain

$$\mathbb{E}M_{q_+ + s}^2 = \mathbb{E} \tanh^2 X_{q_+ + s} = \mathbb{E}[2\psi'_{M_{q_+}}(\beta^2 s)] + \mathbb{E}M_{q_+}^2 = 2\psi'_{\mu_+}(\beta^2 s) + q_+.$$

Inserting this into (7.8), we find for $s \in [0, 1 - q_+]$ that

$$\begin{aligned} \Gamma_{\mu_+}(s) &= \beta^2 \int_0^s (\mathbb{E}M_{q_+ + u}^2 - q_+ - u) du = \beta^2 \int_{q_+}^{q_+ + s} (\mathbb{E}M_t^2 - t) dt \\ &= 2(\Psi(q_+) - \Psi(q_+ + s)) \leq 0, \end{aligned}$$

where the inequality holds by Theorem 2.3, since $q_+ \in \text{supp } \nu$. Finally, let $s \geq 1 - q_+$. By (7.8), (7.7), and (7.10),

$$\Gamma'_{\mu_+}(s) = \beta^2(2\psi'_{\mu_+}(\beta^2 s) - s) \leq \beta^2(1 - q_+ - s) \leq 0,$$

which gives $\Gamma_{\mu_+}(s) \leq \Gamma_{\mu_+}(1 - q_+) \leq 0$. $\square$

Since $\int m^2 \mu_+(dm) = q_+ < 1$, Proposition 7.4 and [41, Proposition 13] imply that, for the SK model, the generalized TAP correction of Chen, Panchenko, and Subag at the law of $|M_{q_+}|$ is replica symmetric and equals the classical TAP correction. For mixed p -spin models, they noted that the correction is approximately classical at the generalized TAP states whose self-overlap is the largest point of the support of the Parisi measure [40, Remark 6]. By [41, Remark 15], for some β and $h > 0$, the correction is not replica symmetric at some other laws whose second moment is q_+ .

7.4 The relative-entropy bound

We recall the planted product model of Definition 6.1. For $m \in [-1, 1]^V$, the planted configuration $v^* = \sigma^* - m$ has independent coordinates, and v_x^* has law ν_{m_x} . The law Q_m is the law of $Y = \beta a(v^*) + Z$ on $\mathbb{R}^E$, where $a(v)_{xy} = \sqrt{S_{xy}} v_x v_y$ and Z is a standard Gaussian vector, and $\mathcal{D}(m) = \text{KL}(Q_m \parallel P_0)$. For $a \geq 0$ we define the compact set

$$\mathcal{A}_S(a) = a\mathbf{1} + 2\beta^2 S[0, 1]^V = \{a\mathbf{1} + 2\beta^2 S y : y \in [0, 1]^V\}.$$

Since S has nonnegative entries and unit row sums, $\mathcal{A}_S(a) \subset [a, a + 2\beta^2]^V$. For a probability measure μ on $[-1, 1]$ and $m \in [-1, 1]^V$, we define

$$\Delta_{S,a}(m; \mu) = \sup_{r \in \mathcal{A}_S(a)} \frac{1}{n} \sum_{x \in V} (\psi_{m_x}(r_x) - \psi_\mu(r_x)). \quad (7.12)$$

Since the function inside the supremum is continuous in (m, r) by Lemma 7.2(d), and $\mathcal{A}_S(a)$ is compact, the map $\Delta_{S,a}(\cdot; \mu)$ is continuous on $[-1, 1]^V$, and the same holds with the absolute value of the function inside the supremum.

Theorem 7.5. *Let $\beta \geq 0$.*

- (a) *For every $a > 0$ there exists a nonincreasing function $e_{\beta,a} : [1, \infty) \rightarrow [0, \infty)$, depending only on β and a , with $\lim_{D \rightarrow \infty} e_{\beta,a}(D) = 0$, such that the following holds. Let $D \geq 1$, let S be as in Definition 6.1 with $\max_{x,y} S_{xy} \leq 1/D$, let $m \in [-1, 1]^V$, and let μ be a probability measure on $[-1, 1]$. Then*

$$\frac{1}{n} \mathcal{D}(m) \leq \frac{1}{2} \sup_{0 \leq s \leq 1} \Gamma_\mu(s) + 2a + \Delta_{S,a}(m; \mu) + e_{\beta,a}(D).$$

- (b) *For each $k \geq 1$, let $D_k \geq 1$, let S_k be an $n_k \times n_k$ matrix as in Definition 6.1 with $\max_{x,y} (S_k)_{xy} \leq 1/D_k$, and let $m^{(k)}$ be a random vector in $[-1, 1]^{n_k}$. Let μ be a probability measure on $[-1, 1]$. Suppose that $D_k \rightarrow \infty$ and that for every $a > 0$,*

$$\lim_{k \rightarrow \infty} \mathbb{E} \sup_{r \in \mathcal{A}_{S_k}(a)} \left| \frac{1}{n_k} \sum_x \psi_{m_x^{(k)}}(r_x) - \frac{1}{n_k} \sum_x \psi_\mu(r_x) \right| = 0. \quad (7.13)$$

Then

$$\limsup_{k \rightarrow \infty} \frac{1}{n_k} \mathbb{E} \mathcal{D}(m^{(k)}) \leq \frac{1}{2} \sup_{0 \leq s \leq 1} \Gamma_\mu(s).$$

Corollary 7.6. *Let $\beta > 0$ and $h > 0$, and let (μ_k) be probability measures on $[-1, 1]$ that converge weakly to μ_+ . Then $\limsup_{k \rightarrow \infty} \sup_{0 \leq s \leq 1} \Gamma_{\mu_k}(s) \leq 0$.*

Remark 7.7. The matrix S need not be positive semidefinite. For example, if $G \in \mathcal{G}_D$ is bipartite, as are the tori $\mathbb{T}_L^d$ with L even and the hypercubes Q_d , then -1 is an eigenvalue of A_G/D . Besides symmetry, the proof of Theorem 7.5 uses three properties of S . First, $S \preceq I$, which is used for the separable comparison function in Section 7.10. Second, S has nonnegative entries and unit row sums, which gives $S[0, 1]^V \subset [0, 1]^V$; this controls the set of velocities in Lemma 7.16. Third, $\|S\| \leq 1$, which enters the proof of Proposition 7.13 through the bound $|w \cdot Sw| \leq |w| \|Sw\|$; there, the quadratic form $w \cdot Sw$ is not assumed to have a sign.

7.5 Interpolating channels

In this subsection and the next three, we fix $\beta > 0$ and $m \in [-1, 1]^V$. Let $v^* = (v_x^*)_{x \in V}$ have independent coordinates with laws ν_{m_x} , and let $Z = (Z_e)_{e \in E}$ and $\zeta = (\zeta_x)_{x \in V}$ be standard Gaussian vectors, with v^* , Z , and ζ independent. For $t \geq 0$ and $q \in [0, \infty)^V$, we observe

$$Y_e = \beta(tS_{xy})^{1/2} v_x^* v_y^* + Z_e \quad (e = \{x, y\} \in E), \quad Y'_x = q_x^{1/2} v_x^* + \zeta_x \quad (x \in V). \quad (7.14)$$

This is the setting of Section 7.1 with $N = |E| + n$, with $\theta = ((v_x^* v_y^*)_{\{x, y\} \in E}, (v_x^*)_{x \in V})$, and with $\lambda = \lambda(t, q) = ((\beta^2 t S_e)_{e \in E}, (q_x)_{x \in V})$, where $S_e = S_{xy}$ for $e = \{x, y\}$. We define

$$F(t, q) = \frac{1}{n} I(\lambda(t, q)).$$

Let $\nu_m = \bigotimes_x \nu_{m_x}$ be the law of v^* . For v in its support, let

$$\begin{aligned} \mathcal{E}_{t, q}(v) = & \sum_{\{x, y\} \in E} \left(\beta(tS_{xy})^{1/2} Z_{xy} v_x v_y + \beta^2 t S_{xy} v_x^* v_y^* v_x v_y - \frac{\beta^2 t S_{xy}}{2} v_x^2 v_y^2 \right) \\ & + \sum_{x \in V} \left(q_x^{1/2} \zeta_x v_x + q_x v_x^* v_x - \frac{q_x}{2} v_x^2 \right), \end{aligned}$$

and let

$$\widehat{F}(t, q) = \frac{1}{n} \log \sum_v \nu_m(v) e^{\mathcal{E}_{t, q}(v)}.$$

By (7.4), we have $F(t, q) = \mathbb{E} \widehat{F}(t, q)$, where $\mathbb{E}$ is the expectation over (v^*, Z, ζ) . The posterior average $\langle \cdot \rangle_{t, q}$ of Section 7.1 is the average over independent samples $v^1, v^2, \dots$

from the probability measure proportional to $\nu_m(v)e^{\mathcal{E}_{t,q}(v)}$; inside $\langle \cdot \rangle_{t,q}$ we write v for v^1 . All profiles (t, q) are coupled through the same (v^*, Z, ζ) . We write

$$b_x = \langle v_x \rangle_{t,q}, \quad \Sigma_{xy} = \langle v_x v_y \rangle_{t,q} - b_x b_y, \quad p_x = \mathbb{E} b_x^2.$$

These depend on (t, q) , which will be clear from the context. Since $|v_x| \leq 1 + |m_x| \leq 2$ and $|v_x^*| \leq 2$, we have $|b_x| \leq 2$ and $|\Sigma_{xy}| \leq 8$.

Lemma 7.8. *Let $\beta > 0$ and $m \in [-1, 1]^V$.*

- (a) *We have $F(1, 0) = n^{-1} \mathcal{D}(m)$ and $F(0, q) = n^{-1} \sum_x \psi_{m_x}(q_x)$ for $q \in [0, \infty)^V$.*
- (b) *The function F is continuous on $[0, \infty) \times [0, \infty)^V$, nondecreasing in t and in each coordinate q_x , and $|F(t, q) - F(t, q')| \leq \|q - q'\|_n / 2$.*
- (c) *Let $q \in (0, \infty)^V$. For every $t \geq 0$, the function $F(t, \cdot)$ is twice continuously differentiable in a neighborhood of q , and*

$$2n \partial_{q_x} F(t, q) = p_x \in [0, \kappa_x], \quad 2n \partial_{q_x} \partial_{q_y} F(t, q) = \mathbb{E} \Sigma_{xy}^2.$$

For every $t > 0$, the function $F(\cdot, q)$ is differentiable at t , and

$$\partial_t F(t, q) = \frac{\beta^2}{4n} \sum_{x,y \in V} S_{xy} \mathbb{E} \langle v_x v_y \rangle_{t,q}^2.$$

Proof. (a) At $q = 0$ the site observations are independent of everything else, and since $\beta S_{xy}^{1/2} v_x^* v_y^* = \beta a(v^*)_{xy}$, the law of the observation (7.14) at $(1, 0)$ is $Q_m \otimes N(0, I_V)$. Its relative entropy with respect to $P_0 \otimes N(0, I_V)$ is $\mathcal{D}(m)$. At $t = 0$ the edge observations are pure noise, and the site observations are independent over x ; the law of the observation is then $N(0, I_E) \otimes \bigotimes_x \text{Law}(q_x^{1/2} v_x^* + \zeta_x)$. Relative entropy is additive over products, and the relative entropy of $\text{Law}(q_x^{1/2} v_x^* + \zeta_x)$ with respect to $N(0, 1)$ is $\psi_{m_x}(q_x)$.

(b) and (c) The map $(t, q) \mapsto \lambda(t, q)$ is linear and has nonnegative coefficients. Continuity and monotonicity follow from Lemma 7.1(a). If $t > 0$ and $q \in (0, \infty)^V$, then all coordinates of $\lambda(t, q)$ are positive, and by Lemma 7.1(b,c) the function I has continuous first and second partial derivatives near $\lambda(t, q)$; since the coordinates of λ indexed by E play no role in the derivatives in q , the same holds in q for $t = 0$. The formulas follow from Lemma 7.1(b,c) and the chain rule, together with $\mathbb{E}(v_x^*)^2 = \kappa_x$ and

$$\sum_{e \in E} \beta^2 S_e \cdot \frac{1}{2} \mathbb{E} \langle v_x v_y \rangle_{t,q}^2 = \frac{\beta^2}{4} \sum_{x,y \in V} S_{xy} \mathbb{E} \langle v_x v_y \rangle_{t,q}^2,$$

which holds because S is symmetric with zero diagonal and $S_{xy} = 0$ if $\{x, y\} \notin E$. Finally, since $0 \leq \partial_{q_x} F \leq 1/(2n)$ on $(0, \infty)^V$, we have $|F(t, q) - F(t, q')| \leq (2n)^{-1} \sum_x |q_x - q'_x| \leq \|q - q'\|_n / 2$ for $q, q' \in (0, \infty)^V$, and by continuity for $q, q' \in [0, \infty)^V$. $\square$

7.6 The score covariance identity

Fix $t \geq 0$ and $q \in (0, \infty)^V$. Differentiating $\widehat{F}(t, \cdot)$ as in (7.5), we find that $\widehat{F}(t, \cdot)$ is smooth near q and $2n\nabla_q \widehat{F}(t, q) = \chi$, where

$$\chi_x = \frac{\zeta_x b_x}{\sqrt{q_x}} + 2v_x^* b_x - \langle v_x^2 \rangle_{t,q}. \quad (7.15)$$

Differentiating once more, we find

$$n \partial_{q_x} \partial_{q_y} \widehat{F}(t, q) = \langle A_x A_y \rangle_{t,q} - \langle A_x \rangle_{t,q} \langle A_y \rangle_{t,q} - \mathbf{1}_{x=y} \frac{\zeta_x b_x}{4q_x^{3/2}}, \quad A_x(v) = \frac{\zeta_x v_x}{2\sqrt{q_x}} + v_x^* v_x - \frac{v_x^2}{2}. \quad (7.16)$$

The first two terms form a covariance matrix, which is positive semidefinite. Since $|b_x| \leq 2$, it follows that

$$\nabla_q^2 \widehat{F}(t, q) \succeq -\frac{1}{n} \text{diag} \left(\frac{|\zeta_x|}{2q_x^{3/2}} \right)_{x \in V}. \quad (7.17)$$

Lemma 7.9. *Let $t \geq 0$ and $q \in (0, \infty)^V$, and let $U = \chi - b^{\odot 2}$. Then $\mathbb{E}\chi = p$, $\mathbb{E}U = 0$, and*

$$\mathbb{E}[U_x U_y] = \mathbf{1}_{x=y} \frac{p_x}{q_x} + \mathbb{E}\Sigma_{xy}^2 \quad (x, y \in V).$$

Proof. By Lemma 7.8(c), $\mathbb{E}\chi_x = 2n\partial_{q_x} F(t, q) = p_x$, and $\mathbb{E}U = 0$. Since $\langle v_x^2 \rangle_{t,q} = \Sigma_{xx} + b_x^2$, we have $U = U^1 + U^2$ with

$$U_x^1 = 2b_x(v_x^* - b_x), \quad U_x^2 = \frac{\zeta_x b_x}{\sqrt{q_x}} - \Sigma_{xx}.$$

We first compute $\mathbb{E}[U_x^2 U_y^2]$, conditionally on (v^*, Z) . Then b is a smooth function of ζ , and by (7.5), $\partial_{\zeta_y} b_x = \sqrt{q_y} \Sigma_{xy}$. Let $f_x = b_x / \sqrt{q_x}$ and write $\partial_y = \partial_{\zeta_y}$. Then $\partial_y f_x = (q_y / q_x)^{1/2} \Sigma_{xy}$ and $U_x^2 = \zeta_x f_x - \partial_x f_x$. Since the functions f_x have bounded derivatives of all orders in ζ , Gaussian integration by parts (F1) in ζ gives

$$\begin{aligned} \mathbb{E}[\zeta_x f_x \zeta_y f_y] &= \mathbf{1}_{x=y} \mathbb{E}f_x^2 + \mathbb{E}[\partial_y \partial_x f_x f_y + \partial_x f_x \partial_y f_y + \partial_y f_x \partial_x f_y + f_x \partial_x \partial_y f_y], \\ \mathbb{E}[\zeta_x f_x \partial_y f_y] &= \mathbb{E}[\partial_x f_x \partial_y f_y + f_x \partial_x \partial_y f_y], \quad \mathbb{E}[\partial_x f_x \zeta_y f_y] = \mathbb{E}[\partial_y \partial_x f_x f_y + \partial_x f_x \partial_y f_y], \end{aligned}$$

where the expectations are conditional on (v^*, Z) . Expanding the product $U_x^2 U_y^2$ and combining these identities, we obtain

$$\mathbb{E}[U_x^2 U_y^2] = \mathbf{1}_{x=y} \mathbb{E}f_x^2 + \mathbb{E}[\partial_y f_x \partial_x f_y] = \mathbf{1}_{x=y} \frac{\mathbb{E}b_x^2}{q_x} + \mathbb{E}\Sigma_{xy}^2,$$

first conditionally on (v^*, Z) and then after taking expectations.

Next, let $\mathcal{Y}$ be the σ -algebra generated by the observations (7.14). The quantities b and Σ are $\mathcal{Y}$ -measurable, and by Bayes's rule the conditional law of v^* given $\mathcal{Y}$ is the posterior. In particular, $\mathbb{E}[v_x^* - b_x \mid \mathcal{Y}] = 0$ and $\mathbb{E}[(v_x^* - b_x)(v_y^* - b_y) \mid \mathcal{Y}] = \Sigma_{xy}$, which gives $\mathbb{E}[U_x^1 U_y^1] = 4\mathbb{E}[b_x b_y \Sigma_{xy}]$. Since $\zeta_y = Y'_y - \sqrt{q_y} v_y^*$, we have

$$U_y^2 = \frac{Y'_y b_y}{\sqrt{q_y}} - b_y^2 - \Sigma_{yy} - b_y(v_y^* - b_y),$$

and the first three terms are $\mathcal{Y}$ -measurable. Hence $\mathbb{E}[U_x^1 U_y^2] = -2\mathbb{E}[b_x b_y \Sigma_{xy}]$, and by symmetry $\mathbb{E}[U_x^2 U_y^1] = -2\mathbb{E}[b_x b_y \Sigma_{xy}]$. When we sum the four contributions, the terms involving $b_x b_y \Sigma_{xy}$ cancel, and $\mathbb{E}[U_x U_y] = \mathbb{E}[U_x^2 U_y^2]$. $\square$

7.7 Concentration

We use the bounded-differences inequality of McDiarmid [101]: if $\xi_1, \dots, \xi_n$ are independent and f changes by at most c_x when the x th argument is changed, then $\mathbb{P}(|f(\xi) - \mathbb{E}f(\xi)| \geq \epsilon) \leq 2\exp(-2\epsilon^2 / \sum_x c_x^2)$ for all $\epsilon \geq 0$.

Lemma 7.10. *Let $\beta > 0$, $\Lambda > 0$, $m \in [-1, 1]^V$, $t \in [0, 1]$, and $r \in [0, \Lambda]^V$. Then for every $\epsilon > 0$,*

$$\mathbb{P}(|\hat{F}(t, r) - F(t, r)| > \epsilon) \leq 4\exp\left(-\frac{n\epsilon^2}{C_\Lambda}\right), \quad C_\Lambda = 32(4\beta^2 + \Lambda + 1)^2.$$

Proof. Fix v^* . By (7.5), $\partial_{Z_e} \hat{F} = n^{-1}\beta(tS_{xy})^{1/2}\langle v_x v_y \rangle_{t,r}$ for $e = \{x, y\}$, and $\partial_{\zeta_x} \hat{F} = n^{-1}r_x^{1/2}b_x$. Since $|\langle v_x v_y \rangle| \leq 4$, $|b_x| \leq 2$, and $\sum_{e \in E} S_e = n/2$, the squared norm of the gradient of $\hat{F}$ in (Z, ζ) is at most $n^{-2}(8\beta^2 n + 4\Lambda n)$. Therefore, $\hat{F}$ is a Lipschitz function of (Z, ζ) with constant $((8\beta^2 + 4\Lambda)/n)^{1/2}$, and (F4) gives

$$\mathbb{P}(|\hat{F} - \Phi_0(v^*)| > \epsilon/2 \mid v^*) \leq 2\exp\left(-\frac{n\epsilon^2}{8(8\beta^2 + 4\Lambda)}\right), \quad \Phi_0(v^*) = \mathbb{E}[\hat{F} \mid v^*].$$

Next, changing the coordinate v_x^* to the other point of the support of ν_{m_x} changes it by at most 2 and changes $\mathcal{E}_{t,r}(v)$ by at most $2|v_x|(\beta^2 t \sum_y S_{xy} |v_y^* v_y| + r_x) \leq 4(4\beta^2 + \Lambda)$, for every v . It follows that $\hat{F}$ changes by at most $4(4\beta^2 + \Lambda)/n$, and so does Φ_0 . Since the coordinates of v^* are independent and $\mathbb{E}\Phi_0(v^*) = F(t, r)$, McDiarmid's inequality implies that

$$\mathbb{P}(|\Phi_0(v^*) - F(t, r)| > \epsilon/2) \leq 2\exp\left(-\frac{n\epsilon^2}{32(4\beta^2 + \Lambda)^2}\right).$$

Since $8(8\beta^2 + 4\Lambda) \leq C_\Lambda$ and $32(4\beta^2 + \Lambda)^2 \leq C_\Lambda$, the claim follows from the two bounds. $\square$

The next lemma controls $\widehat{F} - F$ uniformly over a family of profiles of the form $q + \eta S^2 y$. This family has small metric entropy, because most eigenvalues of S are small: since $\text{tr } S^2 \leq n/D$, at most $n/(D\alpha^2)$ eigenvalues of S have absolute value larger than α .

Lemma 7.11. *Let $\beta > 0$ and $0 < \vartheta \leq \Lambda$. There exists C_0 , depending only on β , ϑ , and Λ , such that the following holds. Let $D \geq 1$, let S be as in Definition 6.1 with $\max_{x,y} S_{xy} \leq 1/D$, and let $m \in [-1, 1]^V$, $t \in [0, 1]$, and $q \in [\vartheta, \Lambda]^V$. Let $A \geq 1$, $\eta \in (0, \vartheta/(2A)]$, and $\alpha \in (0, 1)$, and let $\mathcal{M} = \{q + \eta S^2 y : y \in [-A, A]^V\}$. Then*

$$\mathbb{E} \sup_{r \in \mathcal{M}} |\widehat{F}(t, r) - F(t, r)| \leq C_0 \left(\left(\frac{\log(3/\alpha^2)}{D\alpha^2} \right)^{1/2} + \eta A \alpha^2 \right).$$

Proof. Since S^2 has nonnegative entries and unit row sums, $S^2 y \in [-A, A]^V$ for $y \in [-A, A]^V$, and $\mathcal{M}$ is a compact convex subset of $[\vartheta/2, \Lambda + \vartheta/2]^V \subset [\vartheta/2, 2\Lambda]^V$. The supremum is measurable, since it may be taken over a countable dense subset of $\mathcal{M}$ by the continuity of $\widehat{F}(t, \cdot)$ and $F(t, \cdot)$.

We first construct a net. Let Π_α be the orthogonal projection onto the span of the eigenvectors of S whose eigenvalues have absolute value larger than α , and let k_α be its rank. By (7.1), $k_\alpha \alpha^2 \leq \text{tr } S^2 \leq n/D$, and $|(I - \Pi_\alpha)S^2 w| \leq \alpha^2 |w|$ for $w \in \mathbb{R}^V$. Let $\varrho = \alpha^2 A \sqrt{n}$. For $y \in [-A, A]^V$, the vector $\Pi_\alpha S^2 y$ lies in the ball of radius $A\sqrt{n}$ in the range of Π_α , since $\|S\| \leq 1$. A maximal ϱ -separated subset $\mathcal{C}$ of this ball satisfies $|\mathcal{C}| \leq (1 + 2A\sqrt{n}/\varrho)^{k_\alpha} \leq (3/\alpha^2)^{k_\alpha}$, by comparing the volumes of the disjoint balls of radius $\varrho/2$ around its points with the ball of radius $A\sqrt{n} + \varrho/2$; and every point of the ball lies within distance ϱ of $\mathcal{C}$. For each $c \in \mathcal{C}$ for which there exists $y \in [-A, A]^V$ with $|\Pi_\alpha S^2 y - c| \leq \varrho$, we choose one such y_c , and we let $\mathcal{N}$ be the set of the corresponding profiles $q + \eta S^2 y_c \in \mathcal{M}$. Then $\mathcal{N}$ is deterministic and $|\mathcal{N}| \leq (3/\alpha^2)^{k_\alpha}$. Let $y \in [-A, A]^V$, and choose $c \in \mathcal{C}$ with $|\Pi_\alpha S^2 y - c| \leq \varrho$. Then

$$|S^2(y - y_c)| \leq |\Pi_\alpha S^2(y - y_c)| + |(I - \Pi_\alpha)S^2(y - y_c)| \leq 2\varrho + \alpha^2 |y - y_c| \leq 4\varrho.$$

In particular, every $r \in \mathcal{M}$ lies within normalized distance $4\eta\alpha^2 A$ of $\mathcal{N}$.

Next, by Lemma 7.10 with Λ replaced by 2Λ and a union bound, $\mathbb{P}(\max_{r \in \mathcal{N}} |\widehat{F}(t, r) - F(t, r)| > \epsilon) \leq 4|\mathcal{N}|e^{-n\epsilon^2/C}$, where $C = C_{2\Lambda}$. Let $\epsilon_0 = (C \log(4|\mathcal{N}|)/n)^{1/2}$. For $\epsilon \geq \epsilon_0$ we have $4|\mathcal{N}|e^{-n\epsilon^2/C} \leq e^{-n(\epsilon - \epsilon_0)^2/C}$, since $\epsilon^2 - \epsilon_0^2 \geq (\epsilon - \epsilon_0)^2$. Integrating the tail bound, we obtain

$$\mathbb{E} \max_{r \in \mathcal{N}} |\widehat{F}(t, r) - F(t, r)| \leq \epsilon_0 + \int_{\epsilon_0}^{\infty} e^{-n(\epsilon - \epsilon_0)^2/C} d\epsilon \leq \left(\frac{C \log(4|\mathcal{N}|)}{n} \right)^{1/2} + \left(\frac{C}{n} \right)^{1/2}.$$

Further, $\log(4|\mathcal{N}|) \leq \log 4 + n \log(3/\alpha^2)/(D\alpha^2)$. Since $n \geq D$, $\alpha < 1$, and $\log(3/\alpha^2) \geq 1$, both $\log 4/n$ and $1/n$ are at most $2 \log(3/\alpha^2)/(D\alpha^2)$, and the right side is at most $4C^{1/2}(\log(3/\alpha^2)/(D\alpha^2))^{1/2}$.

Finally, we compare $\mathcal{M}$ with $\mathcal{N}$. Let $r' \in \mathcal{M}$. Since the coordinates of r' are at least $\vartheta/2$, (7.15) gives $2n|\partial_{r_x}\widehat{F}(t, r')| \leq 2^{3/2}\vartheta^{-1/2}|\zeta_x| + 12$, where we used $|b_x| \leq 2$, $|v_x^*| \leq 2$, and $\langle v_x^2 \rangle \leq 4$. Since $\mathcal{M}$ is convex, for $r, r' \in \mathcal{M}$ we obtain

$$|\widehat{F}(t, r) - \widehat{F}(t, r')| \leq \frac{1}{2n} \sum_x (2^{3/2}\vartheta^{-1/2}|\zeta_x| + 12)|r_x - r'_x| \leq (2^{1/2}\vartheta^{-1/2}\|\zeta\|_n + 6)\|r - r'\|_n,$$

and $|F(t, r) - F(t, r')| \leq \|r - r'\|_n/2$ by Lemma 7.8(b). Using $\mathbb{E}\|\zeta\|_n \leq 1$, we conclude that

$$\mathbb{E} \sup_{r \in \mathcal{M}} |\widehat{F}(t, r) - F(t, r)| \leq 4C^{1/2} \left(\frac{\log(3/\alpha^2)}{D\alpha^2} \right)^{1/2} + 4\eta\alpha^2 A(2^{1/2}\vartheta^{-1/2} + 7).$$

This proves the lemma with $C_0 = \max(4C_{2\Lambda}^{1/2}, 4(2^{1/2}\vartheta^{-1/2} + 7))$. $\square$

7.8 The time derivative at contact points

We now show that at every point where $F(t, \cdot)$ is touched from above by a function with Hessian at most K/n , the time derivative of F is close to $\beta^2(4n)^{-1}p \cdot Sp$, where $p = 2n\nabla_q F$.

Definition 7.12. Let $\beta > 0$, $m \in [-1, 1]^V$, $0 < \vartheta \leq \Lambda$, and $K > 0$. A point $(t, q) \in (0, 1] \times [\vartheta, \Lambda]^V$ is a (ϑ, Λ, K) -contact point of F if

$$F(t, q + h) \leq F(t, q) + \nabla_q F(t, q) \cdot h + \frac{K}{2n}|h|^2 \quad \text{for every } h \in [-\vartheta/2, \vartheta/2]^V. \quad (7.18)$$

Proposition 7.13. Let $\beta > 0$, $0 < \vartheta \leq \Lambda$, and $K > 0$. There exists a nonincreasing function $\epsilon_{\beta, \vartheta, \Lambda, K}: [1, \infty) \rightarrow [0, \infty)$, depending only on β , ϑ , Λ , and K , with $\lim_{D \rightarrow \infty} \epsilon_{\beta, \vartheta, \Lambda, K}(D) = 0$, such that the following holds. Let $D \geq 1$, let S be as in Definition 6.1 with $\max_{x,y} S_{xy} \leq 1/D$, let $m \in [-1, 1]^V$, and let (t, q) be a (ϑ, Λ, K) -contact point of F . Then, with $p_x = \mathbb{E}b_x^2 = 2n\partial_{q_x}F(t, q)$,

$$\partial_t F(t, q) \leq \frac{\beta^2}{4n} p \cdot Sp + \epsilon_{\beta, \vartheta, \Lambda, K}(D).$$

Proof. All posterior quantities are taken at the deterministic point (t, q) , and all uses of (7.3), through Lemmas 7.8 and 7.9, are at this point. The perturbed profiles in Step 4 below are random, and they enter only through a pointwise Taylor bound and through Lemma 7.11.

Step 1: the Hessian. By Lemma 7.8(c), $F(t, \cdot)$ is twice continuously differentiable near q . For $h \in \mathbb{R}^V$ and small $s > 0$, Taylor's formula and (7.18) with sh in place of h give $s^2 h \cdot \nabla_q^2 F(t, q) h / 2 + o(s^2) \leq Ks^2|h|^2/(2n)$. This gives $\nabla_q^2 F(t, q) \preceq (K/n)I$, and by Lemma 7.8(c) the matrix $\Xi = (\mathbb{E}\Sigma_{xy}^2)_{x,y \in V}$ satisfies

$$\Xi \preceq 2KI, \quad \mathbb{E} \sum_{x,y} \Sigma_{xy}^2 = \mathbf{1} \cdot \Xi \mathbf{1} \leq 2Kn. \quad (7.19)$$

Step 2: reduction to the filtered fluctuation of $b^{\odot 2}$. Let $w = b^{\odot 2} - p$, so that $\mathbb{E}w = 0$ and $|w_x| \leq 4$. By Lemma 7.8(c) and $\langle v_x v_y \rangle = \Sigma_{xy} + b_x b_y$,

$$\frac{4n}{\beta^2} \partial_t F(t, q) - p \cdot Sp = \sum_{x,y} S_{xy} \mathbb{E} \Sigma_{xy}^2 + 2 \sum_{x,y} S_{xy} \mathbb{E} [\Sigma_{xy} b_x b_y] + \mathbb{E} [w \cdot Sw],$$

where we used $\mathbb{E}[b^{\odot 2} \cdot Sb^{\odot 2}] - p \cdot Sp = \mathbb{E}[w \cdot Sw]$. By (7.19) and $S_{xy} \leq 1/D$, the first term is at most $2Kn/D$. By the Cauchy–Schwarz inequality, $|b_x| \leq 2$, (7.1), and (7.19), the second term is at most

$$8 \left(\sum_{x,y} S_{xy}^2 \right)^{1/2} \left(\mathbb{E} \sum_{x,y} \Sigma_{xy}^2 \right)^{1/2} \leq 8 \left(\frac{n}{D} \right)^{1/2} (2Kn)^{1/2}.$$

For the third term we only use $\|S\| \leq 1$ and $|w| \leq 4\sqrt{n}$: we have $|w \cdot Sw| \leq |w| \|Sw\| \leq 4\sqrt{n} \|Sw\|$, and the third term is at most $4n(\mathbb{E} \|Sw\|_n^2)^{1/2}$. Combining the three bounds, we obtain

$$\partial_t F(t, q) - \frac{\beta^2}{4n} p \cdot Sp \leq \frac{\beta^2}{4} \left(\frac{2K}{D} + 8 \left(\frac{2K}{D} \right)^{1/2} + 4(\mathbb{E} \|Sw\|_n^2)^{1/2} \right). \quad (7.20)$$

Step 3: the term U . Let $\xi = \chi - p$, with χ as in (7.15), and let $U = \chi - b^{\odot 2}$ as in Lemma 7.9. Then $w = \xi - U$. By Lemma 7.9, (7.1), and (7.19),

$$\mathbb{E} |SU|^2 = \text{tr} (S \mathbb{E} [UU^\top] S) = \sum_x (S^2)_{xx} \frac{p_x}{q_x} + \text{tr} (S \Xi S) \leq \frac{n}{D\vartheta} + 2K \text{tr} S^2 \leq \left(\frac{1}{\vartheta} + 2K \right) \frac{n}{D},$$

where we used $p_x \leq 1$, $q_x \geq \vartheta$, and $S \Xi S \preceq 2KS^2$. This implies

$$\mathbb{E} \|Sw\|_n^2 \leq 2\mathbb{E} \|S\xi\|_n^2 + 2 \left(\frac{1}{\vartheta} + 2K \right) \frac{1}{D}. \quad (7.21)$$

Step 4: the term ξ . Fix $A \geq 1$, $\eta \in (0, \vartheta/(2A)]$, and $\alpha \in (0, 1)$. Let ξ^A be the vector with coordinates $\max(-A, \min(A, \xi_x))$, and let $h = \eta S^2 \xi^A$. Then $h \in [-\eta A, \eta A]^V \subset [-\vartheta/2, \vartheta/2]^V$, and $q + h$ lies in the set $\mathcal{M}$ of Lemma 7.11. The coordinates of the points of the segment from q to $q + h$ are at least $\vartheta/2$, and (7.17) gives $\nabla_q^2 \hat{F} \succeq -n^{-1} \text{diag}(2^{1/2} \vartheta^{-3/2} |\zeta_x|)$ on this segment. Taylor's formula with integral remainder and $2n \nabla_q \hat{F}(t, q) = \chi$ give

$$\hat{F}(t, q + h) - \hat{F}(t, q) \geq \frac{1}{2n} \chi \cdot h - \frac{\eta^2 A^2}{\vartheta^{3/2}} \cdot \frac{1}{n} \sum_x |\zeta_x|.$$

On the other hand, (7.18) and $2n \nabla_q F(t, q) = p$ give

$$F(t, q + h) - F(t, q) \leq \frac{1}{2n} p \cdot h + \frac{K}{2} \eta^2 A^2.$$

Subtracting the second inequality from the first, we have

$$\frac{\eta}{2n} \xi \cdot S^2 \xi^A \leq 2 \sup_{r \in \mathcal{M}} |\widehat{F}(t, r) - F(t, r)| + \eta^2 A^2 \left(\vartheta^{-3/2} \frac{1}{n} \sum_x |\xi_x| + \frac{K}{2} \right).$$

Taking expectations, using $\mathbb{E}|\xi_x| \leq 1$ and applying Lemma 7.11, we obtain

$$\mathbb{E} \left[\frac{1}{n} \xi \cdot S^2 \xi^A \right] \leq \frac{4C_0}{\eta} \left(\frac{\log(3/\alpha^2)}{D\alpha^2} \right)^{1/2} + 4C_0 A \alpha^2 + \eta A^2 (2\vartheta^{-3/2} + K), \quad (7.22)$$

where C_0 is the constant of Lemma 7.11. Next, by (7.15) and $q_x \geq \vartheta$, we have $|\xi_x| \leq \varphi(\xi_x)$ with $\varphi(z) = 2\vartheta^{-1/2}|z| + 13$. Let $G \sim N(0, 1)$ and $\omega(A) = \mathbb{E}[\varphi(G)^2 \mathbf{1}_{\varphi(G) > A}]$, which tends to zero as $A \rightarrow \infty$. Then $\mathbb{E}\|\xi\|_n^2 \leq \mathbb{E}\varphi(G)^2$ and $\mathbb{E}\|\xi - \xi^A\|_n^2 \leq \omega(A)$. Using $\|S\| \leq 1$ and the Cauchy–Schwarz inequality, we find

$$\mathbb{E}\|S\xi\|_n^2 = \mathbb{E} \left[\frac{1}{n} \xi \cdot S^2 \xi^A \right] + \mathbb{E} \left[\frac{1}{n} \xi \cdot S^2 (\xi - \xi^A) \right] \leq \mathbb{E} \left[\frac{1}{n} \xi \cdot S^2 \xi^A \right] + (\mathbb{E}\varphi(G)^2)^{1/2} \omega(A)^{1/2}.$$

Let $e(D)$ be the infimum over $A \geq 1$, $\eta \in (0, \vartheta/(2A)]$, and $\alpha \in (0, 1)$ of the sum of the right side of (7.22) and $(\mathbb{E}\varphi(G)^2)^{1/2} \omega(A)^{1/2}$. Then $\mathbb{E}\|S\xi\|_n^2 \leq e(D)$, and e depends only on β , ϑ , Λ , and K , and is nonincreasing in D . Further, $e(D) \rightarrow 0$ as $D \rightarrow \infty$. Indeed, given $\iota > 0$, we first choose A such that $(\mathbb{E}\varphi(G)^2)^{1/2} \omega(A)^{1/2} \leq \iota$, then $\eta \leq \vartheta/(2A)$ such that $\eta A^2 (2\vartheta^{-3/2} + K) \leq \iota$, then α such that $4C_0 A \alpha^2 \leq \iota$, and finally D_0 such that the first term in (7.22) is at most ι for $D \geq D_0$. Then $e(D) \leq 4\iota$ for $D \geq D_0$.

Step 5: conclusion. Combining (7.20), (7.21), and Step 4, we obtain the proposition with

$$\epsilon_{\beta, \vartheta, \Lambda, K}(D) = \frac{\beta^2}{4} \left(\frac{2K}{D} + 8 \left(\frac{2K}{D} \right)^{1/2} + 4 \left(2e(D) + 2 \left(\frac{1}{\vartheta} + 2K \right) \frac{1}{D} \right)^{1/2} \right).$$

This function is nonincreasing and tends to zero as $D \rightarrow \infty$. □

7.9 The scalar Hopf–Lax formula

Lemma 7.14. *Let $\beta \geq 0$, let μ be a probability measure on $[-1, 1]$, and for $t, r \geq 0$ let*

$$f_\mu(t, r) = \max_{0 \leq s \leq 1} \left(\psi_\mu(r + \beta^2 ts) - \frac{\beta^2 ts^2}{4} \right).$$

- (a) *We have $f_\mu(0, r) = \psi_\mu(r)$ and $f_\mu(1, 0) = \max_{0 \leq s \leq 1} \Gamma_\mu(s)/2$.*
- (b) *The function f_μ is nondecreasing in each variable, and $|f_\mu(t, r) - f_\mu(t', r')| \leq \beta^2 |t - t'|/4 + |r - r'|/2$.*
- (c) *At every point $(t, r) \in (0, \infty)^2$ at which f_μ is differentiable, $\partial_t f_\mu = \beta^2 (\partial_r f_\mu)^2$ and $0 \leq \partial_r f_\mu \leq 1/2$.*

Proof. We write $f = f_\mu$. The maximum exists because ψ_μ is continuous, and (a) follows from the definitions.

(b) For each s , the function $r \mapsto \psi_\mu(r + \beta^2 ts) - \beta^2 ts^2/4$ is nondecreasing and $1/2$ -Lipschitz by (7.7). Taking the maximum over s , we find that f is nondecreasing and $1/2$ -Lipschitz in r . Let $0 \leq t < t'$. If s is a maximizer for (t, r) , then $s' = st/t' \in [0, 1]$ satisfies $r + \beta^2 t's' = r + \beta^2 ts$ and $\beta^2 t's'^2/4 \leq \beta^2 ts^2/4$, so $f(t', r) \geq f(t, r)$. Further, (7.7) implies that for every $s \in [0, 1]$,

$$\psi_\mu(r + \beta^2 t's) - \frac{\beta^2 t's^2}{4} \leq \psi_\mu(r + \beta^2 ts) - \frac{\beta^2 ts^2}{4} + \beta^2 (t' - t) \left(\frac{s}{2} - \frac{s^2}{4} \right),$$

and $s/2 - s^2/4 \leq 1/4$, which gives $f(t', r) \leq f(t, r) + \beta^2 (t' - t)/4$.

(c) If $\beta = 0$, then $f(t, r) = \psi_\mu(r)$, and (c) follows from (7.7). Let $\beta > 0$, let $(t, r) \in (0, \infty)^2$ be a point of differentiability of f , and let s^* be a maximizer for (t, r) . The function $\tilde{f}(t', r') = \psi_\mu(r' + \beta^2 t's^*) - \beta^2 t's^{*2}/4$ satisfies $\tilde{f} \leq f$, with equality at (t, r) , and it is differentiable at (t, r) , since ψ_μ is continuously differentiable and $r + \beta^2 ts^* > 0$. Since $f - \tilde{f}$ has a minimum at (t, r) , we have $\nabla f(t, r) = \nabla \tilde{f}(t, r)$, that is,

$$\partial_r f(t, r) = \psi'_\mu(r + \beta^2 ts^*), \quad \partial_t f(t, r) = \beta^2 s^* \psi'_\mu(r + \beta^2 ts^*) - \frac{\beta^2 s^{*2}}{4}.$$

The derivative of $s \mapsto \psi_\mu(r + \beta^2 ts) - \beta^2 ts^2/4$ is $\beta^2 t(\psi'_\mu(r + \beta^2 ts) - s/2)$. If $s^* = 0$, it is nonpositive at $s = 0$, and then $\psi'_\mu(r) = 0$ by (7.7) and $\partial_r f = \partial_t f = 0$. If $s^* \in (0, 1)$, it vanishes at s^* . If $s^* = 1$, it is nonnegative at $s = 1$, and since $\psi'_\mu \leq 1/2$ by (7.7), we have $\psi'_\mu(r + \beta^2 t) = 1/2$. In the last two cases $\psi'_\mu(r + \beta^2 ts^*) = s^*/2$, and the formulas above give $\partial_r f = s^*/2$ and $\partial_t f = \beta^2 s^{*2}/4 = \beta^2 (\partial_r f)^2$. In all cases $\partial_r f \in [0, 1/2]$. $\square$

The function f_μ is only Lipschitz, and we smooth it by averaging over small shifts toward larger times and larger profiles. We fix once and for all a smooth probability density k on $\mathbb{R}^2$ whose support is a compact subset of $(1/2, 1)^2$, and we let $c_k = \int |\partial_2 k(u, v)| \, du \, dv$.

Lemma 7.15. *Let $\beta \geq 0$, let μ be a probability measure on $[-1, 1]$, and let $\gamma > 0$. For $(t, r) \in (-\gamma/2, \infty)^2$ let*

$$w(t, r) = \int_{\mathbb{R}^2} f_\mu(t + \gamma u, r + \gamma v) k(u, v) \, du \, dv.$$

Then w is smooth, and for all $t, r > -\gamma/2$:

- (a) $0 \leq \partial_r w(t, r) \leq 1/2$ and $|\partial_r^2 w(t, r)| \leq c_k/(2\gamma)$;
- (b) $\partial_t w(t, r) \geq \beta^2 (\partial_r w(t, r))^2$;
- (c) $w(0, r) \geq \psi_\mu(r)$ if $r \geq 0$;
- (d) $w(1, r) \leq f_\mu(1, 0) + r/2 + (\beta^2 + 2)\gamma/4$ if $r \geq 0$.

Proof. Since k is supported in $(1/2, 1)^2$, the arguments of f_μ in the definition of w lie in $(0, \infty)^2$. Writing $w(t, r) = \gamma^{-2} \int f_\mu(t', r') k((t' - t)/\gamma, (r' - r)/\gamma) dt' dr'$ and differentiating under the integral shows that w is smooth. By Lemma 7.14(b) and Rademacher's theorem, f_μ is differentiable almost everywhere in $(0, \infty)^2$, and its difference quotients are bounded. By dominated convergence,

$$\begin{aligned}\partial_t w(t, r) &= \int \partial_t f_\mu(t + \gamma u, r + \gamma v) k(u, v) du dv, \\ \partial_r w(t, r) &= \int \partial_r f_\mu(t + \gamma u, r + \gamma v) k(u, v) du dv.\end{aligned}$$

The second formula and Lemma 7.14(c) give $0 \leq \partial_r w \leq 1/2$. Writing the second formula as $\gamma^{-2} \int \partial_r f_\mu(t', r') k((t' - t)/\gamma, (r' - r)/\gamma) dt' dr'$ and differentiating in r , we obtain $|\partial_r^2 w| \leq (1/2)\gamma^{-1} c_k$, which proves (a). By Lemma 7.14(c) and Jensen's inequality,

$$\partial_t w = \beta^2 \int (\partial_r f_\mu)^2(t + \gamma u, r + \gamma v) k(u, v) du dv \geq \beta^2 (\partial_r w)^2,$$

which proves (b). By Lemma 7.14(a,b), $f_\mu(\gamma u, r + \gamma v) \geq f_\mu(0, r) = \psi_\mu(r)$ for $u, v \geq 0$, which gives (c). By Lemma 7.14(b), $f_\mu(1 + \gamma u, r + \gamma v) \leq f_\mu(1 + \gamma, r + \gamma) \leq f_\mu(1, 0) + \beta^2 \gamma/4 + (r + \gamma)/2$ for $u, v \leq 1$, which gives (d). $\square$

7.10 Proof of the relative-entropy bound

The comparison function in the proof of Theorem 7.5 contains a penalty for the distance from a moving convex set of profiles. Since only the initial profiles in this set are used, the initial discrepancy is needed only on the set $\mathcal{A}_S(a)$.

Lemma 7.16. *Let $\beta > 0$ and $a > 0$, let $\mathcal{V} = \beta^2 S[0, 1]^V$, and for $t \in [0, 1]$ let $\mathcal{A}_t = a\mathbf{1} + (2 - t)\mathcal{V}$. For $(t, q) \in [0, 1] \times \mathbb{R}^V$, let $z(t, q)$ be the point of $\mathcal{A}_t$ closest to q , and let*

$$\mathcal{B}(t, q) = \frac{1}{2a} \text{dist}_n(q, \mathcal{A}_t)^2 = \frac{1}{2na} |q - z(t, q)|^2.$$

- (a) *Each $\mathcal{A}_t$ is compact and convex, $a\mathbf{1} \in \mathcal{A}_t$, $\mathcal{A}_t \subset [a, a + 2\beta^2]^V$, and $\mathcal{A}_0 = \mathcal{A}_S(a)$.*
- (b) *For each t , the function $\mathcal{B}(t, \cdot)$ is differentiable, $\nabla_q \mathcal{B}(t, q) = (q - z(t, q))/(na)$, and*

$$\mathcal{B}(t, q + h) \leq \mathcal{B}(t, q) + \nabla_q \mathcal{B}(t, q) \cdot h + \frac{|h|^2}{2na} \quad (q, h \in \mathbb{R}^V).$$

- (c) *For each q , the function $\mathcal{B}(\cdot, q)$ is differentiable on $[0, 1]$, with one-sided derivatives at the endpoints, and*

$$\partial_t \mathcal{B}(t, q) = \max_{v \in \mathcal{V}} \nabla_q \mathcal{B}(t, q) \cdot v.$$

- (d) *The function $\mathcal{B}$ is continuous on $[0, 1] \times \mathbb{R}^V$.*

Proof. (a) The set $\mathcal{V}$ is compact and convex, since it is the image of a cube under a linear map. It contains 0 , and $\mathcal{V} \subset [0, \beta^2]^V$ because S has nonnegative entries and unit row sums. This gives (a).

(b) Fix t , and write $z = z(t, q)$ and $z' = z(t, q + h)$. Since $z \in \mathcal{A}_t$, we have $2na\mathcal{B}(t, q + h) \leq |q + h - z|^2 = |q - z|^2 + 2(q - z) \cdot h + |h|^2$, which is the claimed inequality. Since the projection onto a closed convex set is 1-Lipschitz, $|z - z'| \leq |h|$, and

$$2na\mathcal{B}(t, q + h) = |q - z'|^2 + 2(q - z') \cdot h + |h|^2 \geq |q - z|^2 + 2(q - z) \cdot h - 2|h|^2.$$

From the two bounds, we see that $\mathcal{B}(t, \cdot)$ is differentiable at q with the stated gradient.

(c) Fix q . For $s \in [1, 2]$, let $v_s \in \mathcal{V}$ be such that $z(2 - s, q) = a\mathbf{1} + sv_s$; it is unique since $s > 0$. Let $\phi(s) = |q - a\mathbf{1} - sv_s|^2 = \min_{v \in \mathcal{V}} |q - a\mathbf{1} - sv|^2$, so that $\mathcal{B}(t, q) = \phi(2 - t)/(2na)$. For $s, s' \in [1, 2]$,

$$\phi(s') \leq |q - a\mathbf{1} - s'v_s|^2 = \phi(s) - 2(s' - s)(q - a\mathbf{1} - sv_s) \cdot v_s + (s' - s)^2|v_s|^2,$$

and the same holds with s and s' exchanged. The map $s \mapsto v_s$ is continuous: if $s_j \rightarrow s$ and $\bar{v}$ is a limit point of (v_{s_j}) , then $\bar{v} \in \mathcal{V}$ and $|q - a\mathbf{1} - s\bar{v}|^2 = \lim_j \phi(s_j) \leq \lim_j |q - a\mathbf{1} - s_j v_{s_j}|^2 = \phi(s)$ along the corresponding subsequence, which forces $\bar{v} = v_s$. Combining this with the two inequalities, we see that ϕ is differentiable on $[1, 2]$, with one-sided derivatives at the endpoints, and $\phi'(s) = -2(q - z) \cdot v_s$, where $z = z(2 - s, q)$. Since $t = 2 - s$, this gives $\partial_t \mathcal{B}(t, q) = \nabla_q \mathcal{B}(t, q) \cdot v_s$. Finally, for $v \in \mathcal{V}$ the point $a\mathbf{1} + sv$ lies in $\mathcal{A}_t$, and the variational characterization of the projection gives $(q - z) \cdot (a\mathbf{1} + sv - z) \leq 0$, that is, $s(q - z) \cdot (v - v_s) \leq 0$. Hence $\nabla_q \mathcal{B}(t, q) \cdot v \leq \nabla_q \mathcal{B}(t, q) \cdot v_s$ for every $v \in \mathcal{V}$.

(d) Every point $a\mathbf{1} + (2 - t)v$ of $\mathcal{A}_t$, with $v \in \mathcal{V}$, is within normalized distance $|t - t'| \|v\|_n \leq \beta^2 |t - t'|$ of the point $a\mathbf{1} + (2 - t')v$ of $\mathcal{A}_{t'}$, and vice versa. This gives $|\text{dist}_n(q, \mathcal{A}_t) - \text{dist}_n(q', \mathcal{A}_{t'})| \leq \|q - q'\|_n + \beta^2 |t - t'|$. $\square$

Proof of Theorem 7.5. (a) If $\beta = 0$, then $Q_m = P_0$, $\mathcal{D}(m) = 0$, $\Gamma_\mu = 0$, $\mathcal{A}_S(a) = \{a\mathbf{1}\}$, and $\Delta_{S,a}(m; \mu) \geq -\psi_\mu(a) \geq -a/2$, so (a) holds with $e_{0,a} = 0$.

Let $\beta > 0$ and $a > 0$. Let $\gamma = a/(1 + \beta^2)$, $\vartheta = a/2$, $\Lambda = 3a/2 + 2\beta^2$, and $K = c_k/(2\gamma) + 1/a$, and let $e_{\beta,a} = e_{\beta,\vartheta,\Lambda,K}$ be the function of Proposition 7.13. Fix D, S, m , and μ as in the statement. Let w be the function of Lemma 7.15 for μ and γ , and let

$$\Omega(t, q) = \frac{1}{n} \sum_{x \in V} w(t, q_x) \quad (t \in [0, 1], q \in [0, \infty)^V).$$

Let $\mathcal{V}, \mathcal{A}_t, z(t, q)$, and $\mathcal{B}$ be as in Lemma 7.16. Fix $\rho > e_{\beta,a}(D)$, and let

$$\mathcal{L}(t, q) = F(t, q) - \Omega(t, q) - \mathcal{B}(t, q) - \rho t \quad ((t, q) \in [0, 1] \times [0, \infty)^V).$$

Step 1: $\mathcal{L}$ attains its maximum. By Lemma 7.8(b), $F(t, q) \leq F(1, 0) + \|q\|_n/2$. Further, $\Omega \geq 0$, because $f_\mu \geq \psi_\mu \geq 0$. By Lemma 7.16(a), $\text{dist}_n(q, \mathcal{A}_t) \geq \|q\|_n - a - 2\beta^2$. Combining these bounds, we see that $\mathcal{L}(t, q) \rightarrow -\infty$ as $\|q\|_n \rightarrow \infty$, uniformly in t . Since $\mathcal{L}$ is

continuous by Lemma 7.8(b), Lemma 7.15, and Lemma 7.16(d), it attains its maximum on $[0, 1] \times [0, \infty)^V$ at some point (t_0, q_0) .

Step 2: $q_0 \in (0, \infty)^V$. Suppose that $(q_0)_x = 0$ for some x , let e_x be the corresponding coordinate vector, and let $h \in (0, a)$. By Lemma 7.8(b), $F(t_0, q_0 + he_x) \geq F(t_0, q_0)$. By Lemma 7.15(a), $\Omega(t_0, q_0 + he_x) - \Omega(t_0, q_0) \leq h/(2n)$. Let $z = z(t_0, q_0)$. Then $z_x \geq a$ by Lemma 7.16(a), and Lemma 7.16(b) gives

$$\mathcal{B}(t_0, q_0 + he_x) - \mathcal{B}(t_0, q_0) \leq \frac{h((q_0)_x - z_x)}{na} + \frac{h^2}{2na} \leq -\frac{h}{n} + \frac{h^2}{2na}.$$

Hence $\mathcal{L}(t_0, q_0 + he_x) - \mathcal{L}(t_0, q_0) \geq (h/n)(1/2 - h/(2a)) > 0$, which contradicts the maximality of (t_0, q_0) .

Step 3: location of q_0 . By Step 2, Lemma 7.8(c), Lemma 7.15, and Lemma 7.16(b), the function $\mathcal{L}(t_0, \cdot)$ is differentiable at q_0 , and $\nabla_q \mathcal{L}(t_0, q_0) = 0$ by maximality. Let

$$p = 2n \nabla_q F(t_0, q_0), \quad p' = 2n \nabla_q \Omega(t_0, q_0) = (2 \partial_r w(t_0, (q_0)_x))_{x \in V}.$$

By Lemma 7.8(c) and Lemma 7.15(a), $p, p' \in [0, 1]^V$. Let $z = z(t_0, q_0)$. By Lemma 7.16(b),

$$\frac{q_0 - z}{a} = n \nabla_q \mathcal{B}(t_0, q_0) = \frac{p - p'}{2} \in \left[-\frac{1}{2}, \frac{1}{2}\right]^V.$$

Hence $|(q_0)_x - z_x| \leq a/2$ for all x , and Lemma 7.16(a) gives $q_0 \in [a/2, 3a/2 + 2\beta^2]^V = [\vartheta, \Lambda]^V$ and $\text{dist}_n(q_0, \mathcal{A}_{t_0}) \leq a/2$.

Step 4: $t_0 = 0$. Suppose that $t_0 > 0$. We first show that (t_0, q_0) is a (ϑ, Λ, K) -contact point of F . Let $h \in [-\vartheta/2, \vartheta/2]^V$. Then $q_0 + h \in [0, \infty)^V$, and $\mathcal{L}(t_0, q_0 + h) \leq \mathcal{L}(t_0, q_0)$. By Taylor's formula and Lemma 7.15(a), and by Lemma 7.16(b), we obtain

$$\begin{aligned} F(t_0, q_0 + h) - F(t_0, q_0) &\leq \Omega(t_0, q_0 + h) - \Omega(t_0, q_0) + \mathcal{B}(t_0, q_0 + h) - \mathcal{B}(t_0, q_0) \\ &\leq \nabla_q(\Omega + \mathcal{B})(t_0, q_0) \cdot h + \left(\frac{c_k}{4\gamma} + \frac{1}{2a}\right) \frac{|h|^2}{n}. \end{aligned}$$

Since $\nabla_q(\Omega + \mathcal{B})(t_0, q_0) = \nabla_q F(t_0, q_0)$ and $c_k/(4\gamma) + 1/(2a) = K/2$, this is (7.18). By Proposition 7.13,

$$\partial_t F(t_0, q_0) \leq \frac{\beta^2}{4n} p \cdot Sp + e_{\beta, a}(D).$$

Next, by Lemma 7.15(b) and $S \preceq I$,

$$\partial_t \Omega(t_0, q_0) = \frac{1}{n} \sum_x \partial_t w(t_0, (q_0)_x) \geq \frac{\beta^2}{n} \sum_x (\partial_r w(t_0, (q_0)_x))^2 = \frac{\beta^2}{4n} |p'|^2 \geq \frac{\beta^2}{4n} p' \cdot Sp'.$$

Further, since S is symmetric and $p - p' = 2n \nabla_q \mathcal{B}(t_0, q_0)$ by Step 3,

$$\frac{\beta^2}{4n} (p \cdot Sp - p' \cdot Sp') = \frac{\beta^2}{4n} S(p + p') \cdot (p - p') = \tilde{v} \cdot \nabla_q \mathcal{B}(t_0, q_0), \quad \tilde{v} = \beta^2 S \left(\frac{p + p'}{2} \right).$$

Since $(p + p')/2 \in [0, 1]^V$, we have $\tilde{v} \in \mathcal{V}$, and Lemma 7.16(c) gives $\tilde{v} \cdot \nabla_q \mathcal{B}(t_0, q_0) \leq \partial_t \mathcal{B}(t_0, q_0)$. Combining the last three displays, we find

$$\partial_t F(t_0, q_0) \leq \partial_t \Omega(t_0, q_0) + \partial_t \mathcal{B}(t_0, q_0) + e_{\beta, a}(D) < \partial_t \Omega(t_0, q_0) + \partial_t \mathcal{B}(t_0, q_0) + \rho,$$

that is, $\partial_t \mathcal{L}(t_0, q_0) < 0$, where at $t_0 = 1$ all time derivatives are left derivatives. However, $t \mapsto \mathcal{L}(t, q_0)$ is differentiable at t_0 by Lemma 7.8(c), Lemma 7.15, and Lemma 7.16(c), and since it attains its maximum over $[0, 1]$ at $t_0 \in (0, 1]$, we have $\partial_t \mathcal{L}(t_0, q_0) \geq 0$. This contradiction shows that $t_0 = 0$.

Step 5: conclusion. Since $a\mathbf{1} \in \mathcal{A}_1$, we have $\mathcal{B}(1, a\mathbf{1}) = 0$, and $\Omega(1, a\mathbf{1}) = w(1, a)$. Since $\mathcal{B} \geq 0$ and $t_0 = 0$, the maximality of $(0, q_0)$ gives

$$\begin{aligned} F(1, a\mathbf{1}) - w(1, a) - \rho &= \mathcal{L}(1, a\mathbf{1}) \leq \mathcal{L}(0, q_0) \leq F(0, q_0) - \Omega(0, q_0) \\ &\leq \frac{1}{n} \sum_x (\psi_{m_x}((q_0)_x) - \psi_\mu((q_0)_x)), \end{aligned}$$

where in the last step we used Lemma 7.8(a) and Lemma 7.15(c). By Step 3 and Lemma 7.16(a), there exists $z \in \mathcal{A}_0 = \mathcal{A}_S(a)$ with $\|q_0 - z\|_n \leq a/2$. Since ψ_{m_x} and ψ_μ are $1/2$ -Lipschitz, the right side is at most

$$\frac{1}{n} \sum_x (\psi_{m_x}(z_x) - \psi_\mu(z_x)) + \frac{1}{n} \sum_x |(q_0)_x - z_x| \leq \Delta_{S, a}(m; \mu) + \|q_0 - z\|_n \leq \Delta_{S, a}(m; \mu) + \frac{a}{2}.$$

By Lemma 7.8(a,b), Lemma 7.15(d), and Lemma 7.14(a), and since $(\beta^2 + 2)\gamma/4 \leq a/2$, we have

$$\begin{aligned} \frac{1}{n} \mathcal{D}(m) = F(1, 0) &\leq F(1, a\mathbf{1}) \leq w(1, a) + \rho + \Delta_{S, a}(m; \mu) + \frac{a}{2} \\ &\leq \frac{1}{2} \sup_{0 \leq s \leq 1} \Gamma_\mu(s) + \frac{3a}{2} + \Delta_{S, a}(m; \mu) + \rho. \end{aligned}$$

Letting $\rho \downarrow e_{\beta, a}(D)$ proves (a).

(b) Fix $a > 0$. For each k , we apply (a) with $S = S_k$ and $D = D_k$ to each value of $m^{(k)}$. The function $\mathcal{D}$ is continuous on $[-1, 1]^{n_k}$, by Lemma 7.8(a) and the dominated convergence theorem applied to (7.4) (see also Section 6), and $\Delta_{S_k, a}(\cdot; \mu)$ is continuous. Taking expectations, we obtain

$$\frac{1}{n_k} \mathbb{E} \mathcal{D}(m^{(k)}) \leq \frac{1}{2} \sup_{0 \leq s \leq 1} \Gamma_\mu(s) + 2a + \mathbb{E} \sup_{r \in \mathcal{A}_{S_k}(a)} \left| \frac{1}{n_k} \sum_x (\psi_{m_x^{(k)}}(r_x) - \psi_\mu(r_x)) \right| + e_{\beta, a}(D_k).$$

By (7.13) and $D_k \rightarrow \infty$, the limit superior of the right side as $k \rightarrow \infty$ is at most $\sup_{0 \leq s \leq 1} \Gamma_\mu(s)/2 + 2a$. Letting $a \downarrow 0$ proves (b). $\square$

Proof of Corollary 7.6. Let $r \geq 0$. By Lemma 7.2(d), the function $m \mapsto \psi_m(r)$ is bounded and continuous on $[-1, 1]$, and we have $\psi_{\mu_k}(r) \rightarrow \psi_{\mu_+}(r)$ by the weak convergence. By (7.7), the functions ψ_{μ_k} and ψ_{μ_+} are $1/2$ -Lipschitz, and a finite net of $[0, \beta^2]$ shows that the convergence is uniform on $[0, \beta^2]$. Then $\sup_{0 \leq s \leq 1} |\Gamma_{\mu_k}(s) - \Gamma_{\mu_+}(s)| = 2 \sup_{0 \leq r \leq \beta^2} |\psi_{\mu_k}(r) - \psi_{\mu_+}(r)| \rightarrow 0$, and Proposition 7.4 completes the proof. $\square$

Proofs of the main theorems

This section proves Theorems 1.1 and 1.2 and Corollary 1.3. Section 8.1 combines the results of Sections 5–7 into a lower bound for tori of a fixed side length and for hypercubes. Section 8.2 shows that the pressure of the torus $\mathbb{T}_L^d$ differs from a limit that does not depend on L by at most $\beta^2/(4L)$, uniformly in d . Section 8.3 completes the proofs, using Proposition 4.1 for the upper bound.

8.1 The lower bound at a fixed side length

Proposition 8.1. *Let $\beta \geq 0$ and $h \in \mathbb{R}$. For every integer $L \geq 3$,*

$$\liminf_{d \rightarrow \infty} p_{d,L}(\beta, h) \geq \mathcal{P}(\beta, h), \quad \text{and} \quad \liminf_{d \rightarrow \infty} p_{Q_d}(\beta, h) \geq \mathcal{P}(\beta, h).$$

Proof. At $\beta = 0$ both sides equal $\log 2 \cosh h$, by Lemma 2.1(a) for p_G and for p_N^{SK} . Since p_G and $\mathcal{P}$ are even in h , we may assume that $\beta > 0$ and $h \geq 0$. Let $(G_d)_{d \geq 1}$ be either the tori $\mathbb{T}_L^d$, with $D_d = 2d$, or the hypercubes Q_d , with $D_d = d$. In both cases $G_d \in \mathcal{G}_{D_d}$, $D_d \rightarrow \infty$, and $n = |V(G_d)| \rightarrow \infty$.

We first reduce to $h > 0$. By Lemma 2.1(d), which also holds for $\mathcal{P}$, we have $|p_{G_d}(\beta, 0) - p_{G_d}(\beta, \eta)| \leq \eta$ and $|\mathcal{P}(\beta, 0) - \mathcal{P}(\beta, \eta)| \leq \eta$ for every $\eta > 0$. If the claim holds at (β, η) , then

$$\liminf_{d \rightarrow \infty} p_{G_d}(\beta, 0) \geq \liminf_{d \rightarrow \infty} p_{G_d}(\beta, \eta) - \eta \geq \mathcal{P}(\beta, \eta) - \eta \geq \mathcal{P}(\beta, 0) - 2\eta,$$

and letting $\eta \rightarrow 0$ gives the claim at $(\beta, 0)$. It therefore suffices to treat $\beta > 0$ and $h > 0$. We write $S = A_{G_d}/D_d$, which satisfies the hypotheses of Definition 6.1 and $\max_{x,y} S_{xy} = 1/D_d$.

Fix $\varepsilon \in (0, 1)$. Let T , μ_ε , and the query histories of depth T on G_d with centers $m = m^{(d)}$ be given by Proposition 5.2. We first bound the relative entropy of the planted product model at m . For $m' \in [-1, 1]$ and $r \geq 0$ let $Y(m', r) = \psi_{m'}(r)$. By Lemma 7.2, Y is continuous on $[-1, 1] \times [0, \infty)$ and $1/2$ -Lipschitz in its second variable, and by the definition of ψ_μ , we have $\int Y(m', r) \mu_\varepsilon(dm') = \psi_{\mu_\varepsilon}(r)$. For every $a > 0$, Proposition 5.2(c), applied with $c = 2\beta^2$ and $B = 1$, is the hypothesis (7.13) of Theorem 7.5(b) with $\mu = \mu_\varepsilon$. Then Theorem 7.5(b) implies that

$$\limsup_{d \rightarrow \infty} \frac{1}{n} \mathbb{E} \mathcal{D}(m) \leq \omega(\varepsilon), \quad \omega(\varepsilon) = \frac{1}{2} \sup_{0 \leq s \leq 1} \Gamma_{\mu_\varepsilon}(s). \quad (8.1)$$

We have $\omega(\varepsilon) \geq 0$, because $\Gamma_{\mu_\varepsilon}(0) = 0$. Further, $\mu_\varepsilon \rightarrow \mu_+$ weakly as $\varepsilon \rightarrow 0$, since $W_1(\mu_\varepsilon, \mu_+) \leq \varepsilon$ by Proposition 5.2(b), and Corollary 7.6 gives $\limsup_{\varepsilon \rightarrow 0} \omega(\varepsilon) \leq 0$.

By (8.1), Theorem 6.2 applies with $\delta = \omega(\varepsilon)$. Combining it with Proposition 5.2(a), we obtain

$$\begin{aligned} \liminf_{d \rightarrow \infty} p_{G_d}(\beta, h) &\geq \liminf_{d \rightarrow \infty} \frac{1}{n} \mathbb{E} \left[\beta H_{G_d}(m) + h \sum_x m_x + \sum_x \text{ent}(m_x) + \frac{\beta^2}{4} \langle \kappa, S\kappa \rangle \right] - 4\beta \sqrt{\omega(\varepsilon)} \\ &\geq \mathcal{P}(\beta, h) - \varepsilon - 4\beta \sqrt{\omega(\varepsilon)}. \end{aligned}$$

The left side does not depend on ε , and letting $\varepsilon \rightarrow 0$ completes the proof. $\square$

8.2 The side length

The proof of Theorem 1.1 reduces the uniformity in L to the fixed side lengths of Proposition 8.1. For this we compare tori with boxes. It is convenient to allow graphs that are not regular. For a finite simple graph $\Gamma = (V_\Gamma, E_\Gamma)$ and a number $D \geq 1$, we define

$$p_\Gamma^D(\beta, h) = \frac{1}{|V_\Gamma|} \mathbb{E} \log \sum_{\sigma \in \{-1, 1\}^{V_\Gamma}} \exp \left(\frac{\beta}{\sqrt{D}} \sum_{xy \in E_\Gamma} g_{xy} \sigma_x \sigma_y + h \sum_{x \in V_\Gamma} \sigma_x \right),$$

where $(g_e)_{e \in E_\Gamma}$ are independent standard Gaussian variables. If $\Gamma \in \mathcal{G}_D$, then $p_\Gamma^D = p_\Gamma$.

The next lemma states that the pressure increases when Gaussian couplings are added, and bounds the increase. The monotonicity was proved by Contucci and Graffi for short-range Gaussian spin glasses, in the form of the superadditivity of the pressure [44, Theorem 2], using the interpolation of Guerra and Toninelli [75]. Contucci and Lebowitz extended it to couplings with symmetric distributions [46, Theorem 1]. We include the short proof, since we also need the upper bound.

Lemma 8.2. *Let $D \geq 1$, $\beta \geq 0$, and $h \in \mathbb{R}$, and let Γ and Γ' be finite simple graphs on the same set of n vertices such that $E_\Gamma \subset E_{\Gamma'}$ and $E_{\Gamma'} \setminus E_\Gamma$ has k elements. Then*

$$0 \leq p_{\Gamma'}^D(\beta, h) - p_\Gamma^D(\beta, h) \leq \frac{\beta^2 k}{2Dn}.$$

Further, $\log 2 \cosh h \leq p_\Gamma^D(\beta, h) \leq \log 2 \cosh h + \beta^2 |E_\Gamma| / (2Dn)$, and if Γ is the disjoint union of graphs $\Gamma_1, \dots, \Gamma_r$, then $p_\Gamma^D = \sum_{j=1}^r (|V_{\Gamma_j}|/n) p_{\Gamma_j}^D$.

Proof. Let $F = E_{\Gamma'} \setminus E_\Gamma$. For $s \in [0, 1]$ let $\langle \cdot \rangle_s$ be the Gibbs average for the exponent

$$\frac{\beta}{\sqrt{D}} \sum_{xy \in E_\Gamma} g_{xy} \sigma_x \sigma_y + \frac{\beta \sqrt{s}}{\sqrt{D}} \sum_{xy \in F} g_{xy} \sigma_x \sigma_y + h \sum_x \sigma_x,$$

and let $\varphi(s)$ be n^{-1} times the expected logarithm of the corresponding partition function. Then $\varphi(0) = p_\Gamma^D$ and $\varphi(1) = p_\Gamma^D$. For $s \in (0, 1]$, Gaussian integration by parts (F1) in g_{xy} , $xy \in F$, gives

$$\varphi'(s) = \frac{\beta}{2n\sqrt{sD}} \sum_{xy \in F} \mathbb{E}[g_{xy} \langle \sigma_x \sigma_y \rangle_s] = \frac{\beta^2}{2nD} \sum_{xy \in F} (1 - \mathbb{E}[\langle \sigma_x \sigma_y \rangle_s^2]) \in \left[0, \frac{\beta^2 k}{2nD}\right].$$

Since φ is continuous on $[0, 1]$, integrating gives the first claim. The bounds on p_Γ^D follow from the first claim applied to the graph with vertex set V_Γ and no edges, whose pressure is $\log 2 \cosh h$, and to Γ . For a disjoint union, the partition function is the product of the partition functions of the components, which are independent, and the last claim follows. $\square$

For integers $d \geq 1$ and $N \geq 1$ let $\Lambda_N^d = \{0, \dots, N-1\}^d$ be the box in which two vertices are adjacent if they differ by 1 in exactly one coordinate, and write $b_{d,N} = p_{\Lambda_N^d}^{2d}$ for its pressure with couplings of variance $1/(2d)$. More generally, a box is a product of d intervals of integers, with the same adjacency.

The existence of the thermodynamic limit of short-range spin glasses is classical [93, 137, 138]; see also [44, 46, 47]. Contucci and Graffi compared free and periodic boundary conditions for the EA model with the interpolation of Lemma 8.2, and they bounded the difference of the pressures by a multiple of the size of the boundary [45, Theorems 1 and 2]. For the torus, the existence of the limit and a bound of order $1/L$ on the difference with the free-boundary pressure were recently proved by Wang, Zhu, and Chueng under weak moment assumptions [139]. The next lemma is a form of these bounds that is uniform in d .

Lemma 8.3. *Let $d \geq 1$, $\beta \geq 0$, and $h \in \mathbb{R}$. The limit $p_d^\infty(\beta, h) = \lim_{N \rightarrow \infty} b_{d,N}(\beta, h)$ exists, and for every integer $L \geq 3$,*

$$|p_{d,L}(\beta, h) - p_d^\infty(\beta, h)| \leq \frac{\beta^2}{4L}.$$

Proof. We first compare the torus with the box. Let $L \geq 3$. The torus $\mathbb{T}_L^d$ is obtained from Λ_L^d by adding, for each coordinate $j \leq d$ and each choice of the other $d-1$ coordinates, the edge between the two vertices whose j th coordinates are $L-1$ and 0 . Since $L \geq 3$, these dL^{d-1} edges are not edges of Λ_L^d . Since $\mathbb{T}_L^d \in \mathcal{G}_{2d}$ has L^d vertices, Lemma 8.2 with $D = 2d$ gives

$$0 \leq p_{d,L} - b_{d,L} \leq \frac{\beta^2 d L^{d-1}}{2 \cdot 2d \cdot L^d} = \frac{\beta^2}{4L}. \quad (8.2)$$

Next, we compare boxes of different sizes. Fix integers $m \geq 1$ and $N \geq m$, and write $N = km + r$ with $k = \lfloor N/m \rfloor$ and $0 \leq r < m$. We partition $\{0, \dots, N-1\}$ into k

consecutive intervals of length m and, if $r > 0$, one interval of length r . Taking products, we obtain a partition of Λ_N^d into boxes, of which k^d are translates of Λ_m^d . Let Γ be the graph on Λ_N^d whose edges are the edges of Λ_N^d with both endpoints in the same box of the partition. For each of the d coordinate directions, there are N^{d-1} lines in Λ_N^d parallel to it, and on each of them at most k edges join different intervals. In total, at most $dkN^{d-1} \leq dN^d/m$ edges of Λ_N^d are not edges of Γ . By Lemma 8.2 with $D = 2d$, we have $0 \leq b_{d,N} - p_\Gamma^{2d} \leq \beta^2/(4m)$. Further, p_Γ^{2d} is the average of the pressures of the boxes of the partition, weighted by their numbers of vertices. Each translate of Λ_m^d has pressure $b_{d,m}$. Every box with n' vertices has at most dn' edges, and by Lemma 8.2 its pressure, like $b_{d,m}$, lies in $[\log 2 \cosh h, \log 2 \cosh h + \beta^2/4]$. Let $f_N = (km/N)^d$ be the fraction of vertices of Λ_N^d that lie in translates of Λ_m^d . We have shown that

$$b_{d,m} - (1 - f_N) \frac{\beta^2}{4} \leq b_{d,N} \leq b_{d,m} + (1 - f_N) \frac{\beta^2}{4} + \frac{\beta^2}{4m}.$$

Since $f_N = (1 - r/N)^d \rightarrow 1$ as $N \rightarrow \infty$, with d and m fixed, we obtain

$$b_{d,m} \leq \liminf_{N \rightarrow \infty} b_{d,N} \leq \limsup_{N \rightarrow \infty} b_{d,N} \leq b_{d,m} + \frac{\beta^2}{4m}. \quad (8.3)$$

Taking the limit inferior in m on the right side of (8.3), we obtain $\limsup_N b_{d,N} \leq \liminf_m b_{d,m}$, so the limit p_d^∞ exists. By (8.3), it satisfies $b_{d,m} \leq p_d^\infty \leq b_{d,m} + \beta^2/(4m)$ for every $m \geq 1$. For $m = L$, combining this with (8.2) shows that $p_{d,L}$ and p_d^∞ both lie in the interval $[b_{d,L}, b_{d,L} + \beta^2/(4L)]$, which completes the proof. $\square$

Remark 8.4. Lemma 8.3 is the only place where the geometry of the torus enters the passage from fixed to arbitrary side lengths. It shows that the limits $N \rightarrow \infty$ and $L \rightarrow \infty$ of the box and torus pressures in a fixed dimension coincide. Combined with Theorem 1.1, it gives $\lim_{d \rightarrow \infty} p_d^\infty(\beta, h) = \mathcal{P}(\beta, h)$, so the convergence also holds when the thermodynamic limit is taken first.

8.3 Proofs of the main results

Proof of Theorem 1.1. Fix $\beta \geq 0$ and $h \in \mathbb{R}$. Since $\mathbb{T}_L^d \in \mathcal{G}_{2d}$ for every $L \geq 3$, Proposition 4.1 implies that

$$\limsup_{d \rightarrow \infty} \sup_{L \geq 3} p_{d,L}(\beta, h) \leq \limsup_{D \rightarrow \infty} \sup_{G \in \mathcal{G}_D} p_G(\beta, h) \leq \mathcal{P}(\beta, h).$$

For the lower bound, fix an integer $m \geq 3$. For $L \geq m$, applying Lemma 8.3 at the side lengths L and m , we obtain

$$p_{d,L} \geq p_d^\infty - \frac{\beta^2}{4L} \geq p_{d,m} - \frac{\beta^2}{4m} - \frac{\beta^2}{4L} \geq p_{d,m} - \frac{\beta^2}{2m}.$$

Hence $\inf_{L \geq 3} p_{d,L} \geq \min_{3 \leq L \leq m} p_{d,L} - \beta^2/(2m)$. Since the minimum is over finitely many side lengths, Proposition 8.1 implies that

$$\liminf_{d \rightarrow \infty} \inf_{L \geq 3} p_{d,L}(\beta, h) \geq \mathcal{P}(\beta, h) - \frac{\beta^2}{2m}.$$

Letting $m \rightarrow \infty$ and combining with the upper bound, we obtain Theorem 1.1. $\square$

Proof of Theorem 1.2. Since $Q_d \in \mathcal{G}_d$, we have $\limsup_d p_{Q_d}(\beta, h) \leq \mathcal{P}(\beta, h)$ by Proposition 4.1, and Proposition 8.1 gives the matching lower bound. $\square$

Proof of Corollary 1.3. Let $G \in \mathcal{G}_D$, $h \in \mathbb{R}$, and $\beta > 0$. By the bound after (2.4), $e(G; h)$ lies in the interval $[\beta^{-1}p_G(\beta, \beta h) - \beta^{-1} \log 2, \beta^{-1}p_G(\beta, \beta h)]$, and by (2.5), $P_*(h)$ lies in $[\beta^{-1}\mathcal{P}(\beta, \beta h) - \beta^{-1} \log 2, \beta^{-1}\mathcal{P}(\beta, \beta h)]$. Comparing the two intervals, we obtain

$$|e(G; h) - P_*(h)| \leq \frac{|p_G(\beta, \beta h) - \mathcal{P}(\beta, \beta h)|}{\beta} + \frac{\log 2}{\beta}.$$

Taking the supremum over $L \geq 3$ for $G = \mathbb{T}_L^d$, letting $d \rightarrow \infty$, and using Theorem 1.1 at $(\beta, \beta h)$, we obtain $\limsup_d \sup_{L \geq 3} |e(\mathbb{T}_L^d; h) - P_*(h)| \leq (\log 2)/\beta$. Letting $\beta \rightarrow \infty$ proves the claim for tori, and the same argument with Theorem 1.2 proves it for hypercubes. $\square$

Finally, we justify Remark 1.4. By Lemma 2.1(d), for each $B \geq 0$ the functions p_G , $G \in \bigcup_D \mathcal{G}_D$, and p_N^{SK} , $N \geq 1$, satisfy $|f(\beta, h) - f(\beta', h')| \leq B|\beta - \beta'|/2 + |h - h'|$ for $\beta, \beta' \in [0, B]$ and $h, h' \in \mathbb{R}$, and so does their pointwise limit $\mathcal{P}$. Let $K \subset [0, B] \times \mathbb{R}$ be compact and $\eta > 0$, and let K' be a finite subset of K such that every point of K lies within distance η of K' in each coordinate. Then

$$\sup_{(\beta, h) \in K} \sup_{L \geq 3} |p_{d,L}(\beta, h) - \mathcal{P}(\beta, h)| \leq \max_{(\beta, h) \in K'} \sup_{L \geq 3} |p_{d,L}(\beta, h) - \mathcal{P}(\beta, h)| + (B + 2)\eta,$$

and the maximum tends to zero as $d \rightarrow \infty$ by Theorem 1.1. Since η is arbitrary, the convergence in Theorem 1.1 is uniform on K . The same argument applies to hypercubes, and to the ground-state energy, which is 1-Lipschitz in h by its definition. Its SK limit $P_*(h)$ is also 1-Lipschitz, as the pointwise limit of the 1-Lipschitz functions $m_N(h)$.

We also prove the limit in Remark 1.8. A similar interpolation, with independent SK models on the two parts, is in [16, Section 3.1]. Let $D \geq 1$ and $n = 2D$, let $G = K_{D,D}$ have parts V_1 and V_2 of size D , and let $v \in \{-1, 1\}^V$ be equal to 1 on V_1 and to -1 on V_2 . For $\sigma^1, \sigma^2 \in \{-1, 1\}^V$ let $u = \sigma^1 \odot \sigma^2$. Since $\sum_{x \in V_1} u_x$ and $\sum_{y \in V_2} u_y$ are $(\langle u, \mathbf{1} \rangle \pm \langle u, v \rangle)/2$, and every vertex of V_1 is adjacent to every vertex of V_2 , we have

$$\mathbb{E} H_G(\sigma^1) H_G(\sigma^2) = \frac{1}{D} \sum_{x \in V_1} \sum_{y \in V_2} u_x u_y = \frac{1}{2n} (\langle u, \mathbf{1} \rangle^2 - \langle u, v \rangle^2).$$

We identify V with $\{1, \dots, n\}$ and let H_n^{SK} be an SK Hamiltonian on $\{-1, 1\}^V$, independent of $(g_e)_{e \in E}$. Then $\mathbb{E} H_n^{\text{SK}}(\sigma^1) H_n^{\text{SK}}(\sigma^2) = (\langle u, \mathbf{1} \rangle^2 - n) / (2n)$, and the difference of the two covariances is

$$C(\sigma^1, \sigma^2) = \mathbb{E} H_n^{\text{SK}}(\sigma^1) H_n^{\text{SK}}(\sigma^2) - \mathbb{E} H_G(\sigma^1) H_G(\sigma^2) = \frac{\langle u, v \rangle^2 - n}{2n}.$$

In particular, $C(\sigma, \sigma) = -1/2$ for every σ , since $\langle \mathbf{1}, v \rangle = 0$. For $\theta \in [0, 1]$, let $\langle \cdot \rangle_\theta$ be the Gibbs average for the exponent $\beta(\sqrt{1-\theta} H_G(\sigma) + \sqrt{\theta} H_n^{\text{SK}}(\sigma)) + h \sum_x \sigma_x$, and let $\varphi(\theta)$ be n^{-1} times the expected logarithm of the corresponding partition function. Then $\varphi(0) = p_G(\beta, h)$ and $\varphi(1) = p_n^{\text{SK}}(\beta, h)$. For $\theta \in (0, 1)$, Gaussian integration by parts (F1) gives, as in [75],

$$\varphi'(\theta) = \frac{\beta^2}{2n} \mathbb{E} \langle C(\sigma^1, \sigma^1) - C(\sigma^1, \sigma^2) \rangle_\theta = -\frac{\beta^2}{4n^2} \mathbb{E} \langle \langle \sigma^1 \odot \sigma^2, v \rangle^2 \rangle_\theta \leq 0,$$

where σ^1 and σ^2 are independent samples from the Gibbs measure. Since φ is continuous on $[0, 1]$, we obtain $p_{K_{D,D}}(\beta, h) \geq p_{2D}^{\text{SK}}(\beta, h)$, and the right side tends to $\mathcal{P}(\beta, h)$ as $D \rightarrow \infty$. Since $K_{D,D} \in \mathcal{G}_D$, Proposition 4.1 gives the matching upper bound.

References

- [1] M. Aizenman, J. L. Lebowitz, and D. Ruelle. Some rigorous results on the Sherrington–Kirkpatrick spin glass model. *Comm. Math. Phys.*, 112(1):3–20, 1987.
- [2] M. Aizenman, R. Sims, and S. L. Starr. Extended variational principle for the Sherrington–Kirkpatrick spin-glass model. *Phys. Rev. B*, 68:214403, 2003.
- [3] Diego Alberici, Francesco Camilli, Pierluigi Contucci, and Emanuele Mingione. The solution of the deep Boltzmann machine on the Nishimori line. *Comm. Math. Phys.*, 387(2):1191–1214, 2021. arXiv:2012.13987.
- [4] Diego Alberici, Francesco Camilli, Pierluigi Contucci, and Emanuele Mingione. The multi-species mean-field spin-glass on the Nishimori line. *J. Stat. Phys.*, 182(1): Paper No. 2, 2021. arXiv:2007.08891.
- [5] Antonio Auffinger and Wei-Kuo Chen. Free energy and complexity of spherical bipartite models. *J. Stat. Phys.*, 157(1):40–59, 2014. arXiv:1405.2321.
- [6] Antonio Auffinger and Wei-Kuo Chen. On properties of Parisi measures. *Probab. Theory Related Fields*, 161(3-4):817–850, 2015. numbered citations refer to arXiv:1303.3573v1.
- [7] Antonio Auffinger and Wei-Kuo Chen. The Parisi formula has a unique minimizer. *Comm. Math. Phys.*, 335(3):1429–1444, 2015. numbered citations refer to arXiv:1402.5132v3.
- [8] Antonio Auffinger and Wei-Kuo Chen. The Legendre structure of the Parisi formula. *Comm. Math. Phys.*, 348(3):751–770, 2016. numbered citations refer to arXiv:1510.03414v2.
- [9] Antonio Auffinger and Aukosh Jagannath. Thouless–Anderson–Palmer equations for generic p -spin glasses. *Ann. Probab.*, 47(4):2230–2256, 2019. arXiv:1612.06359.
- [10] Antonio Auffinger, Wei-Kuo Chen, and Qiang Zeng. The SK model is infinite step replica symmetry breaking at zero temperature. *Comm. Pure Appl. Math.*, 73(5): 921–943, 2020.
- [11] Jinho Baik and Ji Oon Lee. Free energy of bipartite spherical Sherrington–Kirkpatrick model. *Ann. Inst. Henri Poincaré Probab. Stat.*, 56(4):2897–2934, 2020. arXiv:1711.06364.
- [12] Afonso S. Bandeira and Ramon van Handel. Sharp nonasymptotic bounds on the

- norm of random matrices with independent entries. *Ann. Probab.*, 44(4):2479–2506, 2016. arXiv:1408.6185.
- [13] Zhigang Bao, Qiyang Han, and Xiaocong Xu. A leave-one-out approach to approximate message passing. *Ann. Appl. Probab.*, 35(4):2716–2766, 2025. arXiv:2312.05911.
- [14] Jean Barbier and Nicolas Macris. The adaptive interpolation method: a simple scheme to prove replica formulas in Bayesian inference. *Probab. Theory Related Fields*, 174(3-4):1133–1185, 2019. arXiv:1705.02780.
- [15] Adriano Barra, Aldo Di Biasio, and Francesco Guerra. Replica symmetry breaking in mean-field spin glasses through the Hamilton–Jacobi technique. *J. Stat. Mech. Theory Exp.*, page P09006, 2010. arXiv:1003.5226.
- [16] Adriano Barra, Giuseppe Genovese, and Francesco Guerra. Equilibrium statistical mechanics of bipartite spin systems. *J. Phys. A*, 44(24):245002, 2011. arXiv:1012.1261.
- [17] Adriano Barra, Pierluigi Contucci, Emanuele Mingione, and Daniele Tantari. Multi-species mean field spin glasses. Rigorous results. *Ann. Henri Poincaré*, 16(3):691–708, 2015. arXiv:1307.5154.
- [18] Anirban Basak and Sumit Mukherjee. Universality of the mean-field for the Potts model. *Probab. Theory Related Fields*, 168(3-4):557–600, 2017. arXiv:1508.03949.
- [19] Erik Bates and Youngtak Sohn. Balanced multi-species spin glasses, 2025. Preprint, arXiv:2507.06522v2.
- [20] Joshua Batson, Daniel A. Spielman, and Nikhil Srivastava. Twice-Ramanujan sparsifiers. *SIAM J. Comput.*, 41(6):1704–1721, 2012. arXiv:0808.0163.
- [21] Mohsen Bayati and Andrea Montanari. The dynamics of message passing on dense graphs, with applications to compressed sensing. *IEEE Trans. Inform. Theory*, 57(2):764–785, 2011. numbered citations refer to arXiv:1001.3448v4.
- [22] Mohsen Bayati, Marc Lelarge, and Andrea Montanari. Universality in polytope phase transitions and message passing algorithms. *Ann. Appl. Probab.*, 25(2):753–822, 2015.
- [23] Joshua K. Behne and Galen Reeves. Fundamental limits for rank-one matrix estimation with groupwise heteroskedasticity. In *Proceedings of the 25th International Conference on Artificial Intelligence and Statistics*, volume 151 of *Proc. Mach. Learn. Res.*, pages 8650–8672, 2022. arXiv:2106.11950.
- [24] Patrick Billingsley. *Probability and measure*. John Wiley & Sons, New York, third edition, 1995.
- [25] Erwin Bolthausen. An iterative construction of solutions of the TAP equations for the Sherrington–Kirkpatrick model. *Comm. Math. Phys.*, 325(1):333–366, 2014. numbered citations refer to arXiv:1201.2891v1.

- [26] Erwin Bolthausen. A Morita type proof of the replica-symmetric formula for SK. In Véronique Gayrard, Louis-Pierre Arguin, Nicola Kistler, and Irina Kourkova, editors, *Statistical mechanics of classical and disordered systems*, volume 293 of *Springer Proc. Math. Stat.*, pages 63–93. Springer, Cham, 2019. arXiv:1809.07972.
- [27] Erwin Bolthausen and Alain-Sol Sznitman. On Ruelle’s probability cascades and an abstract cavity method. *Comm. Math. Phys.*, 197(2):247–276, 1998.
- [28] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration inequalities: a nonasymptotic theory of independence*. Oxford University Press, Oxford, 2013.
- [29] Christian Brennecke and Horng-Tzer Yau. The replica symmetric formula for the SK model revisited. *J. Math. Phys.*, 63(7):073302, 2022. numbered citations refer to arXiv:2109.07354v2.
- [30] Sourav Chatterjee. Spin glass phase at zero temperature in the Edwards–Anderson model, 2023. Preprint, revised May 2, 2026; arXiv:2301.04112v6.
- [31] Hong-Bin Chen. On free energy of non-convex multi-species spin glasses. *ALEA Lat. Am. J. Probab. Math. Stat.*, 23:429–473, 2026. doi: 10.30757/ALEA.v23-17. arXiv:2411.13342.
- [32] Hong-Bin Chen. FRSB in the SK spin glass: convergence to full-interval support at zero temperature, 2026. Preprint, arXiv:2607.18032v1.
- [33] Hong-Bin Chen and Jean-Christophe Mourrat. On the free energy of vector spin glasses with non-convex interactions. *Probab. Math. Phys.*, 6:1–80, 2025. arXiv:2311.08980.
- [34] Hong-Bin Chen and Jean-Christophe Mourrat. Free energy of non-convex multi-species spherical spin glasses, 2026. Preprint, arXiv:2609.32710v1.
- [35] Hong-Bin Chen and Jiaming Xia. Hamilton–Jacobi equations for inference of matrix tensor products. *Ann. Inst. Henri Poincaré Probab. Stat.*, 58(2):755–793, 2022. arXiv:2009.01678.
- [36] Hong-Bin Chen and Jiaming Xia. Hamilton–Jacobi equations with monotone nonlinearities on convex cones. *Ann. Fac. Sci. Toulouse Math. (6)*, 34(1):1–45, 2025. arXiv:2206.12537; theorem numbers refer to arXiv version 4.
- [37] Hong-Bin Chen, Victor Issa, and Jean-Christophe Mourrat. Free energy of non-convex multi-species spin glasses with centered Ising spins, 2026. Preprint, arXiv:2606.16636v2.
- [38] Wei-Kuo Chen. Variational representations for the Parisi functional and the two-dimensional Guerra–Talagrand bound. *Ann. Probab.*, 45(6A):3929–3966, 2017. numbered citations refer to arXiv:1501.06635v2.
- [39] Wei-Kuo Chen and Dmitry Panchenko. On the TAP free energy in the mixed p -spin models. *Comm. Math. Phys.*, 362(1):219–252, 2018. numbered citations refer to arXiv:1709.03468v2.

- [40] Wei-Kuo Chen, Dmitry Panchenko, and Eliran Subag. The generalized TAP free energy II. *Comm. Math. Phys.*, 381(1):257–291, 2021. numbered citations refer to arXiv:1903.01030v1.
- [41] Wei-Kuo Chen, Dmitry Panchenko, and Eliran Subag. The generalized TAP free energy. *Comm. Pure Appl. Math.*, 76(7):1329–1415, 2023. numbered citations refer to arXiv:1812.05066v3.
- [42] Michael B. Cohen, Jelani Nelson, and David P. Woodruff. Optimal approximate matrix product in terms of stable rank. In *43rd International Colloquium on Automata, Languages, and Programming (ICALP 2016)*, volume 55 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 11:1–11:14, 2016. arXiv:1507.02268; theorem numbers refer to arXiv version 3.
- [43] Pierluigi Contucci and Cristian Giardinà. The Ghirlanda–Guerra identities. *J. Stat. Phys.*, 126(4-5):917–931, 2007. arXiv:math-ph/0505055.
- [44] Pierluigi Contucci and Sandro Graffi. Monotonicity and thermodynamic limit for short range disordered models. *J. Stat. Phys.*, 115:581–589, 2004. numbered citations refer to arXiv:math-ph/0302013v2.
- [45] Pierluigi Contucci and Sandro Graffi. On the surface pressure for the Edwards–Anderson model. *Comm. Math. Phys.*, 248:207–216, 2004. numbered citations refer to arXiv:math-ph/0306006v3.
- [46] Pierluigi Contucci and Joel L. Lebowitz. Correlation inequalities for spin glasses. *Ann. Henri Poincaré*, 8:1461–1467, 2007. numbered citations refer to arXiv:cond-mat/0612371v1.
- [47] Pierluigi Contucci, Cristian Giardinà, and Joseph Pulé. The thermodynamic limit for finite dimensional classical and quantum disordered systems. *Rev. Math. Phys.*, 16(5):629–637, 2004. arXiv:cond-mat/0401211.
- [48] Pierluigi Contucci, Emanuele Mingione, and Shannon Starr. Factorization properties in d -dimensional spin glasses. Rigorous results and some perspectives. *J. Stat. Phys.*, 151(5):809–829, 2013. arXiv:1212.5281.
- [49] Michael G. Crandall and Pierre-Louis Lions. Viscosity solutions of Hamilton–Jacobi equations. *Trans. Amer. Math. Soc.*, 277(1):1–42, 1983.
- [50] Michael G. Crandall, Lawrence C. Evans, and Pierre-Louis Lions. Some properties of viscosity solutions of Hamilton–Jacobi equations. *Trans. Amer. Math. Soc.*, 282(2):487–502, 1984.
- [51] A. Crisanti and T. Rizzo. Analysis of the ∞ -replica symmetry breaking solution of the Sherrington–Kirkpatrick model. *Phys. Rev. E*, 65:046137, 2002. arXiv:cond-mat/0111037.

- [52] Debsurya De and Dmitriy Kunisky. Computational and statistical lower bounds for low-rank estimation under general inhomogeneous noise, 2025. Preprint, arXiv:2510.08541v1.
- [53] J. R. L. de Almeida and D. J. Thouless. Stability of the Sherrington–Kirkpatrick solution of a spin glass model. *J. Phys. A*, 11(5):983–990, 1978.
- [54] Amir Dembo, Andrea Montanari, and Subhabrata Sen. Extremal cuts of sparse random graphs. *Ann. Probab.*, 45(2):1190–1217, 2017.
- [55] DLMF. NIST digital library of mathematical functions. <https://dlmf.nist.gov/>, Release 1.2.8 of 2026-09-15, 2026. F. W. J. Olver, A. B. Olde Daalhuis, D. W. Lozier, B. I. Schneider, R. F. Boisvert, C. W. Clark, B. R. Miller, B. V. Saunders, H. S. Cohl, and M. A. McClain, eds.
- [56] Tomas Dominguez and Jean-Christophe Mourrat. *Statistical mechanics of mean-field disordered systems: a Hamilton–Jacobi approach*. Zurich Lectures in Advanced Mathematics. EMS Press, 2024. arXiv:2311.08976; theorem numbers refer to the arXiv version.
- [57] David L. Donoho, Arian Maleki, and Andrea Montanari. Message-passing algorithms for compressed sensing. *Proc. Natl. Acad. Sci. USA*, 106(45):18914–18919, 2009. arXiv:0907.3574.
- [58] S. F. Edwards and P. W. Anderson. Theory of spin glasses. *J. Phys. F: Met. Phys.*, 5(5):965–974, 1975.
- [59] Ahmed El Alaoui, Florent Krzakala, and Michael I. Jordan. Fundamental limits of detection in the spiked Wigner model. *Ann. Statist.*, 48(2):863–885, 2020. arXiv:1806.09588.
- [60] Ahmed El Alaoui, Andrea Montanari, and Mark Sellke. Optimization of mean-field spin glasses. *Ann. Probab.*, 49(6):2922–2960, 2021. numbered citations refer to arXiv:2001.00904v1.
- [61] 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 (FOCS)*, pages 323–334. IEEE, 2022. numbered citations refer to arXiv:2203.05093v2.
- [62] Ahmed El Alaoui, Andrea Montanari, and Mark Sellke. Local algorithms for maximum cut and minimum bisection on locally treelike regular graphs of large degree. *Random Structures Algorithms*, 63(3):689–715, 2023. arXiv:2111.06813.
- [63] Houssam El Cheairi and David Gamarnik. Algorithmic universality, low-degree polynomials, and Max-Cut in sparse random graphs, 2024. Preprint, arXiv:2412.18014.

- [64] Zhou Fan and Yihong Wu. The replica-symmetric free energy for Ising spin glasses with orthogonally invariant couplings. *Probab. Theory Related Fields*, 190:1–77, 2024. arXiv:2105.02797.
- [65] Zhou Fan, Song Mei, and Andrea Montanari. TAP free energy, spin glasses and variational inference. *Ann. Probab.*, 49(1):1–45, 2021. arXiv:1808.07890.
- [66] Zhou Fan, Yufan Li, and Subhabrata Sen. TAP equations for orthogonally invariant spin glasses at high temperature, 2022. arXiv:2202.09325v2; to appear in *Ann. Inst. H. Poincaré Probab. Statist.*
- [67] L. A. Fernández, V. Martín-Mayor, G. Parisi, and B. Seoane. Spin glasses on the hypercube. *Phys. Rev. B*, 81:134403, 2010. arXiv:0911.4667.
- [68] Daniel S. Fisher and David A. Huse. Equilibrium behavior of the spin-glass ordered phase. *Phys. Rev. B*, 38(1):386–411, 1988.
- [69] Silvio Franz and Fabio Lucio Toninelli. Kac limit for finite-range spin glasses. *Phys. Rev. Lett.*, 92(3):030602, 2004.
- [70] Silvio Franz and Fabio Lucio Toninelli. Finite-range spin glasses in the Kac limit: free energy and local observables. *J. Phys. A*, 37(30):7433–7446, 2004.
- [71] J. Fröhlich and B. Zegarlinski. Some comments on the Sherrington–Kirkpatrick model of spin glasses. *Comm. Math. Phys.*, 112(4):553–566, 1987. doi: 10.1007/BF01225372.
- [72] A. Georges, M. Mézard, and J. S. Yedidia. Low-temperature phase of the Ising spin glass on a hypercubic lattice. *Phys. Rev. Lett.*, 64(24):2937–2940, 1990.
- [73] Antoine Georges and Jonathan S. Yedidia. How to expand around mean-field theory using high-temperature expansions. *J. Phys. A*, 24(9):2173–2192, 1991.
- [74] Stefano Ghirlanda and Francesco Guerra. General properties of overlap probability distributions in disordered spin systems. Towards Parisi ultrametricity. *J. Phys. A*, 31(46):9149–9155, 1998. arXiv:cond-mat/9807333.
- [75] F. Guerra and F. L. Toninelli. The thermodynamic limit in mean field spin glass models. *Comm. Math. Phys.*, 230(1):71–79, 2002.
- [76] Francesco Guerra. Broken replica symmetry bounds in the mean field spin glass model. *Comm. Math. Phys.*, 233(1):1–12, 2003.
- [77] Francesco Guerra and Fabio Lucio Toninelli. Some comments on the connection between disordered long range spin glass models and their mean field version. *J. Phys. A*, 36(43):10987–10995, 2003.
- [78] Alice Guionnet, Justin Ko, Florent Krzakala, and Lenka Zdeborová. Low-rank matrix estimation with inhomogeneous noise. *Inf. Inference*, 14(2):iaaf010, 2025. arXiv:2208.05918.

- [79] Dongning Guo, Shlomo Shamai, and Sergio Verdú. Mutual information and minimum mean-square error in Gaussian channels. *IEEE Trans. Inform. Theory*, 51(4): 1261–1282, 2005. arXiv:cs/0412108.
- [80] Walid Hachem. Approximate message passing for sparse matrices with application to the equilibria of large ecological Lotka–Volterra systems. *Stochastic Process. Appl.*, 170:104276, 2024. arXiv:2302.09847v4.
- [81] Walid Hachem. The SK model with a sparse variance profile: free energy and AMP algorithm for TAP equations at high temperature, 2026. Preprint, arXiv:2604.25535v1.
- [82] A. B. Harris, T. C. Lubensky, and J.-H. Chen. Critical properties of spin-glasses. *Phys. Rev. Lett.*, 36(8):415–418, 1976.
- [83] Fu-Hsuan Ho. A concavity theorem for the Parisi PDE, 2026. Preprint, arXiv:2606.15432v1.
- [84] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.*, 58(301):13–30, 1963.
- [85] Brice Huang and Mark Sellke. A constructive proof of the spherical Parisi formula, 2023. Preprint, arXiv:2311.15495.
- [86] Victor Issa. Existence and uniqueness of permutation-invariant optimizers for Parisi formula. *ALEA Lat. Am. J. Probab. Math. Stat.*, 23:761–800, 2026. doi: 10.30757/ALEA.v23-30. arXiv:2407.13846; numbered citations refer to the published version.
- [87] Aukosh Jagannath and Ian Tobasco. A dynamic programming approach to the Parisi functional. *Proc. Amer. Math. Soc.*, 144(7):3135–3150, 2016. numbered citations refer to arXiv:1502.04398v4.
- [88] Aukosh Jagannath and Ian Tobasco. Some properties of the phase diagram for mixed p -spin glasses. *Probab. Theory Related Fields*, 167:615–672, 2017. numbered citations refer to arXiv:1504.02731v2.
- [89] Aukosh Jagannath, Justin Ko, and Subhabrata Sen. Max κ -cut and the inhomogeneous Potts spin glass. *Ann. Appl. Probab.*, 28(3):1536–1572, 2018. arXiv:1703.03455.
- [90] Svante Janson. Large deviations for sums of partly dependent random variables. *Random Structures Algorithms*, 24(3):234–248, 2004.
- [91] Adel Javanmard and Andrea Montanari. State evolution for general approximate message passing algorithms, with applications to spatial coupling. *Inf. Inference*, 2(2):115–144, 2013. numbered citations refer to arXiv:1211.5164v2.
- [92] Ioannis Karatzas and Steven E. Shreve. *Brownian motion and stochastic calculus*, volume 113 of *Graduate Texts in Mathematics*. Springer, New York, second edition, 1991.

- [93] K. M. Khanin and Ya. G. Sinai. Existence of free energy for models with long-range random Hamiltonians. *J. Stat. Phys.*, 20:573–584, 1979.
- [94] Satish Babu Korada and Nicolas Macris. Exact solution of the gauge symmetric p -spin glass model on a complete graph. *J. Stat. Phys.*, 136(2):205–230, 2009.
- [95] F. Krzakala and O. C. Martin. Spin and link overlaps in three-dimensional spin glasses. *Phys. Rev. Lett.*, 85(14):3013–3016, 2000.
- [96] Florent Krzakala and Lenka Zdeborová. Hiding quiet solutions in random constraint satisfaction problems. *Phys. Rev. Lett.*, 102:238701, 2009. arXiv:0901.2130.
- [97] Florent Krzakala, Jiaming Xu, and Lenka Zdeborová. Mutual information in rank-one matrix estimation. In *2016 IEEE Information Theory Workshop (ITW)*, pages 71–75. IEEE, 2016. arXiv:1603.08447.
- [98] Marc Lelarge and Léo Miolane. Fundamental limits of symmetric low-rank matrix estimation. *Probab. Theory Related Fields*, 173:859–929, 2019. arXiv:1611.03888.
- [99] Patrick Lopatto. Parisi measure of the Sherrington–Kirkpatrick model with external field, 2026. Preprint, arXiv:2610.00127v1.
- [100] Patrick Lopatto. Full replica symmetry breaking in the Sherrington–Kirkpatrick model, 2026. Preprint, arXiv:2607.11756v3.
- [101] Colin McDiarmid. On the method of bounded differences. In *Surveys in Combinatorics, 1989 (Norwich, 1989)*, volume 141 of *London Math. Soc. Lecture Note Ser.*, pages 148–188. Cambridge Univ. Press, Cambridge, 1989.
- [102] Marc Mézard, Giorgio Parisi, and Miguel Angel Virasoro. *Spin glass theory and beyond: An introduction to the replica method and its applications*, volume 9 of *World Scientific Lecture Notes in Physics*. World Scientific, 1987.
- [103] Andrea Montanari. Optimization of the Sherrington–Kirkpatrick Hamiltonian. *SIAM J. Comput.*, 54(4):FOCS19–1–FOCS19–38, 2025. Published online 2021; numbered citations refer to arXiv:1812.10897v2.
- [104] Jean-Christophe Mourrat. Hamilton–Jacobi equations for finite-rank matrix inference. *Ann. Appl. Probab.*, 30(5):2234–2260, 2020. arXiv:1904.05294.
- [105] Jean-Christophe Mourrat. Hamilton–Jacobi equations for mean-field disordered systems. *Ann. H. Lebesgue*, 4:453–484, 2021. arXiv:1811.01432.
- [106] Jean-Christophe Mourrat. Nonconvex interactions in mean-field spin glasses. *Probab. Math. Phys.*, 2(2):281–339, 2021. numbered citations refer to arXiv:2004.01679v8.
- [107] Jean-Christophe Mourrat. The Parisi formula is a Hamilton–Jacobi equation in Wasserstein space. *Canad. J. Math.*, 74(3):607–629, 2022. arXiv:1906.08471.
- [108] Jean-Christophe Mourrat. Free energy upper bound for mean-field vector spin glasses. *Ann. Inst. Henri Poincaré Probab. Stat.*, 59(3):1143–1182, 2023. numbered citations refer to arXiv:2010.09114v2.

- [109] Jean-Christophe Mourrat and Dmitry Panchenko. Extending the Parisi formula along a Hamilton–Jacobi equation. *Electron. J. Probab.*, 25:Paper No. 23, 1–17, 2020. arXiv:1910.14482.
- [110] Hidetoshi Nishimori. Internal energy, specific heat and correlation function of the bond-random Ising model. *Progr. Theoret. Phys.*, 66(4):1169–1181, 1981.
- [111] Matteo Palassini and A. P. Young. Nature of the spin glass state. *Phys. Rev. Lett.*, 85(14):3017–3020, 2000.
- [112] Dmitry Panchenko. A question about the Parisi functional. *Electron. Comm. Probab.*, 10:155–166, 2005. arXiv:math/0412463.
- [113] Dmitry Panchenko. The Parisi ultrametricity conjecture. *Ann. of Math. (2)*, 177(1):383–393, 2013. arXiv:1112.1003.
- [114] Dmitry Panchenko. *The Sherrington–Kirkpatrick model*. Springer Monographs in Mathematics. Springer, New York, 2013.
- [115] Dmitry Panchenko. The free energy in a multi-species Sherrington–Kirkpatrick model. *Ann. Probab.*, 43(6):3494–3513, 2015. arXiv:1310.6679.
- [116] Dmitry Panchenko. Free energy in the Potts spin glass. *Ann. Probab.*, 46(2):829–864, 2018. arXiv:1512.00370.
- [117] Dmitry Panchenko. Free energy in the mixed p -spin models with vector spins. *Ann. Probab.*, 46(2):865–896, 2018. arXiv:1512.04441.
- [118] Giorgio Parisi. Infinite number of order parameters for spin-glasses. *Phys. Rev. Lett.*, 43:1754–1756, 1979.
- [119] Giorgio Parisi. A sequence of approximated solutions to the S-K model for spin glasses. *J. Phys. A*, 13(4):L115–L121, 1980.
- [120] Miquel Payaró and Daniel P. Palomar. Hessian and concavity of mutual information, differential entropy, and entropy power in linear vector Gaussian channels. *IEEE Trans. Inform. Theory*, 55(8):3613–3628, 2009. numbered citations refer to arXiv:0903.1945v1.
- [121] T. Plefka. Convergence condition of the TAP equation for the infinite-ranged Ising spin glass model. *J. Phys. A*, 15(6):1971–1978, 1982.
- [122] Tim Robertson, F. T. Wright, and R. L. Dykstra. *Order restricted statistical inference*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Chichester, 1988.
- [123] David Ruelle. A mathematical reformulation of Derrida’s REM and GREM. *Comm. Math. Phys.*, 108(2):225–239, 1987.
- [124] René L. Schilling, Renming Song, and Zoran Vondraček. *Bernstein functions: theory and applications*, volume 37 of *De Gruyter Studies in Mathematics*. Walter de Gruyter, Berlin, second edition, 2012.

- [125] Mark Sellke. Optimizing mean field spin glasses with external field. *Electron. J. Probab.*, 29(4):1–47, 2024. numbered citations refer to arXiv:2105.03506v6.
- [126] Subhabrata Sen. Optimization on sparse random hypergraphs and spin glasses. *Random Structures Algorithms*, 53(3):504–536, 2018. arXiv:1606.02365.
- [127] D. Sherrington and S. Kirkpatrick. Solvable model of a spin-glass. *Phys. Rev. Lett.*, 35(26):1792–1796, 1975.
- [128] Charles M. Stein. Estimation of the mean of a multivariate normal distribution. *Ann. Statist.*, 9(6):1135–1151, 1981.
- [129] Daniel L. Stein and Charles M. Newman. *Spin glasses and complexity*. Primers in Complex Systems. Princeton University Press, Princeton, NJ, 2013.
- [130] Eliran Subag. TAP approach for multispecies spherical spin glasses II: the free energy of the pure models. *Ann. Probab.*, 51(3):1004–1024, 2023. arXiv:2111.07134.
- [131] Eliran Subag. Free energy landscapes in spherical spin glasses. *Duke Math. J.*, 173(7):1291–1357, 2024. arXiv:1804.10576.
- [132] Michel Talagrand. Transportation cost for Gaussian and other product measures. *Geom. Funct. Anal.*, 6(3):587–600, 1996.
- [133] Michel Talagrand. *Spin glasses: a challenge for mathematicians. Cavity and mean field models*, volume 46 of *Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge*. Springer, Berlin, 2003.
- [134] Michel Talagrand. The Parisi formula. *Ann. of Math. (2)*, 163(1):221–263, 2006.
- [135] D. J. Thouless, P. W. Anderson, and R. G. Palmer. Solution of ‘Solvable model of a spin glass’. *Philos. Mag.*, 35(3):593–601, 1977.
- [136] F. L. Toninelli. About the Almeida–Thouless transition line in the Sherrington–Kirkpatrick mean-field spin glass model. *Europhys. Lett.*, 60(5):764–767, 2002.
- [137] A. C. D. van Enter and J. L. van Hemmen. The thermodynamic limit for long-range random systems. *J. Stat. Phys.*, 32:141–152, 1983.
- [138] P. A. Vuillermot. Thermodynamics of quenched random spin systems, and application to the problem of phase transitions in magnetic (spin) glasses. *J. Phys. A*, 10: 1319–1333, 1977.
- [139] Hexiang Wang, Keheng Zhu, and Mauris Chueng. The thermodynamic limit of short-range spin glasses with periodic boundary conditions, 2026. Preprint, arXiv:2609.19005.
- [140] Tianhao Wang, Xinyi Zhong, and Zhou Fan. Universality of approximate message passing algorithms and tensor networks. *Ann. Appl. Probab.*, 34(4):3943–3994, 2024. arXiv:2206.13037.

- [141] Yuxin Zhou. Existence of full replica symmetry breaking for the Sherrington–Kirkpatrick model at low temperature. *Comm. Pure Appl. Math.*, 79(7):1746–1770, 2026. arXiv:2504.00269.