HAMILTON–JACOBI EQUATIONS FOR FINITE-RANK MATRIX INFERENCE

BY J.-C. MOURRAT

DMA, Ecole normale supérieure, CNRS, PSL University, mourrat@dma.ens.fr

We compute the large-scale limit of the free energy associated with the problem of inference of a finite-rank matrix. The method follows the principle put forward in Mourrat (2018) which consists in identifying a suitable Hamilton–Jacobi equation satisfied by the limit free energy. We simplify the approach of Mourrat (2018) using a notion of weak solution of the Hamilton–Jacobi equation which is more convenient to work with and is applicable whenever the nonlinearity in the equation is convex.

1. Introduction. We fix an integer $K \in \{1, 2, \ldots\}$ once and for all, and let $(\bar{x}_{1,k})_{1 \le k \le K}, \ldots, (\bar{x}_{N,k})_{1 \le k \le K}$ be N independent and identically distributed random vectors taking values in $\mathbb{R}^K$. We denote the law of one of these vectors by P, and use the shorthand notation $P_N := P^{\otimes N}$ to denote their joint law. In the inference problem we consider, we observe the matrix

$$(1.1) \qquad Y := \sqrt{\frac{t}{N}} \bar{x} \bar{x}^{\mathsf{t}} + W \in \mathbb{R}^{N \times N},$$

where $t > 0$, $W = (W_{ij})_{1 \le i, j \le N}$ is an N-by-N matrix of independent standard Gaussians, and $\bar{x}^{\mathsf{t}}$ denotes the transpose of $\bar{x} = (\bar{x}_{i,k})_{1 \le i \le N, 1 \le k \le K} \in \mathbb{R}^{N \times K}$. The matrix W should be thought of as noise that perturbs the observation, and we aim to recover information about the matrix $\bar{x} \bar{x}^{\mathsf{t}}$, whose rank is at most K, given the observation of Y.

The estimation of a low-rank matrix from noisy observations is a basic model in statistical inference, with a variety of applications. One of the key properties of this problem, which can be recovered from the results presented here, is that there exists a threshold value $t_c > 0$ such that for $t < t_c$, asymptotically no information can be gained about $\bar{x} \bar{x}^{\mathsf{t}}$ from the observation of Y; while for $t > t_c$, the information gain is bounded away from 0 uniformly as $N \to \infty$ [12]. This problem, being "replica-symmetric", is also an excellent training ground for the development of robust mathematical techniques that one may hope to later extend to models with "replica symmetry breaking".

In order to understand this inference problem, it is of particular interest to study the conditional law of $\bar{x}$ given Y. This conditional law is the Gibbs measure associated with the quantity

$$(1.2) \qquad H_N^\circ(t, x) := \sqrt{\frac{t}{N}} x \cdot Wx + \frac{t}{N} |x^{\mathsf{t}} \bar{x}|^2 - \frac{t}{2N} |x^{\mathsf{t}} x|^2,$$

where $x \in \mathbb{R}^{N \times K}$, and where for any two matrices A and B of the same size, we write

$$(1.3) \qquad A \cdot B = \mathrm{tr}(A^{\mathsf{t}} B) \quad \text{and} \quad |A| = (A \cdot A)^{\frac{1}{2}} = \sqrt{\mathrm{tr}(A^{\mathsf{t}} A)}.$$

Denoting by $\mathbb{P}$ the joint law of $\bar{x}$ and W, this means that for any bounded measurable function $f : \mathbb{R}^{N \times K} \to \mathbb{R}$, we have

$$(1.4) \qquad \mathbb{E}[f(\bar{x})|Y] = \frac{\int_{\mathbb{R}^{N \times K}} f(x) \exp(H_N^\circ(t, x)) \, dP_N(x)}{\int_{\mathbb{R}^{N \times K}} \exp(H_N^\circ(t, x)) \, dP_N(x)}.$$

As in problems of statistical mechanics, it is highly informative to understand the large-N limit of the denominator in the expression above, which we may call the "partition function". Indeed, this quantity is essentially a moment-generating function. We aim to tackle this problem by proceeding in three steps: (1) we enrich the "energy" in (1.2) by adding a simpler term where the quadratic interaction term $x \cdot Wx$ is replaced by a linear term; (2) we find a relationship between the derivatives of the logarithm of the enriched partition function, up to error terms; (3) we show that the effect of the error terms becomes negligible in the large-N limit. We refer to the discussion of the Curie-Weiss model in [14], Section 1, for a more concrete illustration of this plan, and for further motivations.

As we enrich the energy in (1.2), it will be of fundamental importance for the analysis of the problem that we preserve the inference structure evidenced in (1.4). We thus define the enriched model indirectly by considering that we observe, in addition to Y in (1.1), the quantity

$$(1.5) \qquad Y' := \bar{x}\sqrt{h} + z \in \mathbb{R}^{N \times K},$$

where $z = (z_{i,k})_{1 \leq i \leq N, 1 \leq k \leq K}$ is an N-by-K matrix of independent standard Gaussian entries, independent of $(\bar{x}, W)$, and h is a fixed K-by-K symmetric positive semidefinite matrix. The conditional law of $\bar{x}$ given the observation of Y and Y' is the Gibbs measure associated with the quantity defined, for every $x \in \mathbb{R}^{N \times K}$, by

$$H_N(t, h, x) := H_N^\circ(t, x) + \sqrt{h} \cdot x^{\mathrm{t}}z + h \cdot x^{\mathrm{t}}\bar{x} - \frac{1}{2}h \cdot x^{\mathrm{t}}x.$$

The proof of this fact is recalled in the Appendix. As explained above, our goal is to study the large-N limit of the "free energy"

$$(1.6) \qquad F_N(t, h) := \frac{1}{N} \log\left(\int_{\mathbb{R}^{N \times K}} e^{H_N(t,h,x)} \, dP_N(x)\right),$$

or of its expectation (with respect to the variables $\bar{x}$, W and z)

$$(1.7) \qquad \bar{F}_N(t, h) := \mathbb{E}[F_N(t, h)].$$

To state the main result, we introduce some definitions. First, notice that $\bar{F}_N(0, h)$ does not depend on N; we denote it by

$$\psi(h) := \bar{F}_N(0, h).$$

We denote the set of K-by-K symmetric matrices by S^K, and write S_+^K and S_{++}^K for the subsets of positive semidefinite and positive definite matrices respectively. For any open set $U \subset S^K$ and smooth function $f : U \to \mathbb{R}$, we define the gradient of f, which we denote by $\nabla f : U \to S^K$, to be the unique mapping such that, for every $h \in U$ and $a \in S^K$, we have

$$\lim_{\varepsilon \to 0} \varepsilon^{-1}\big(f(h + \varepsilon a) - f(h)\big) = \nabla f(h) \cdot a.$$

This gradient has a formal adjoint in $L^2(S^K)$, which we denote by $-\nabla\cdot$, and we set $\Delta := \nabla \cdot \nabla$ to denote the corresponding Laplacian. The set S^K can be identified with $\mathbb{R}^{\frac{K(K+1)}{2}}$ after the choice of an orthonormal basis, and these differential operators on S^K then match the standard definitions on $\mathbb{R}^{\frac{K(K+1)}{2}}$. We will most of the time encounter functions defined on $\mathbb{R}_+ \times S_+^K$, in which case the operators ∇ and Δ are understood to act on the second variable only, keeping the first variable fixed, which we interpret as a "time" variable. Here is the main result of the paper.

THEOREM 1.1 (Convergence to HJ).   *Let $f : \mathbb{R}_+ \times S_+^K \to \mathbb{R}$ be the unique weak solution of the Hamilton–Jacobi equation*

$$\partial_t f - 2|\nabla f|^2 = 0 \quad \text{in } \mathbb{R}_+ \times S_+^K \tag{1.8}$$

with initial condition $f(0, h) = \psi(h)$. For every $M \geq 1$, there exists a constant $C < \infty$ such that for every $N \geq 2$ and $t \in [0, M]$,

$$\int_{|h| \leq M} |\bar{F}_N - f|(t, h) \, dh \leq \frac{C}{\log N}. \tag{1.9}$$

In (1.9), the variable of integration is implicitly understood to range in S_+^K; the notation dh stands for the $\frac{K(K+1)}{2}$-dimensional Lebesgue measure on this set. The convergence of $\bar{F}_N$ to f in (1.9) can easily be improved to, say, convergence in $L_{\text{loc}}^\infty(\mathbb{R}_+ \times S_+^K)$, using that the functions $\bar{F}_N$ are Lipschitz uniformly over N. Using the local semiconvexity of $\bar{F}_N$, see Definition 2.1 and (3.37) below, one can also obtain the convergence of the derivatives of $\bar{F}_N$ to those of f at every point of differentiability of f. In particular, calculating the large-N limit of $\partial_t \bar{F}_N(t, 0)$ allows to identify the asymptotic minimum mean-square error of the original inference problem, see [12]. Using Lemma 3.7, one can also obtain the convergence of the random free energy F_N to f, for instance in the sense that (1.9) can be replaced by

$$\mathbb{E}\left[\left(\int_{|h| \leq M} |F_N - f|(t, h) \, dh\right)^2\right]^{\frac{1}{2}} \leq \frac{C}{\log N}.$$

I do not know if the rate of convergence in (1.9) is sharp. In the special rank-one case $K = 1$, the proof given below simplifies in several ways, most importantly in relation with Remark 3.5 below, and yields an algebraic instead of logarithmic rate of convergence.

As the proof reveals, the constant C in (1.9) can be chosen to be a power of the rank K and the cutoff M. The result thus allows to let K diverge slowly with N. Also, the independence assumption on the rows of $\bar{x}$ can be relaxed: what we really need is that $\bar{F}_N(0, \cdot)$ converges in $L_{\text{loc}}^1(S_+^K)$.

Although it appears there in a different formulation, the qualitative convergence of $\bar{F}_N$ to f was already proved in [12]. Besides the fact that Theorem 1.1 gives a quantitative estimate, the main contribution of the present paper is to provide an alternative proof of this result, which I view as simpler and more "conceptual" than the original proof. The argument is simple enough that obtaining a rate of convergence essentially comes without additional effort. The driving idea is similar to that in [14], in that we first show that $\bar{F}_N$ satisfies the Hamilton–Jacobi equation (1.8) approximately, and then pass to the limit. However, compared with [14], one important difference is that we use here a notion of weak solution of the Hamilton–Jacobi equation which differs from the notion of viscosity solution used in [14]. The precise definition of weak solution we rely on is given below in Section 2. We also explain there why this notion of solution is more adapted to the purpose of proving Theorem 1.1.

The primary ingredient for proving Theorem 1.1 is the following result. We denote the condition number of a matrix $h \in S_+^K$ by

$$\kappa(h) := \begin{vmatrix} |h||h^{-1}| & \text{if } h \in S_{++}^K, \\ +\infty & \text{otherwise.} \end{vmatrix} \tag{1.10}$$

PROPOSITION 1.2 (Approximate HJ in finite volume).   *There exists $C < \infty$ (which depends only on K and on a bound on the support of P) such that for every $N \geq 1$ and uniformly over $\mathbb{R}_+ \times S_+^K$,*

$$0 \leq \partial_t \bar{F}_N - 2|\nabla \bar{F}_N|^2 \leq C\kappa(h)N^{-\frac{1}{4}}(\Delta \bar{F}_N + C|h^{-1}|)^{\frac{1}{4}} + C\mathbb{E}[|\nabla F_N - \nabla \bar{F}_N|^2]. \tag{1.11}$$

In (1.11), we understand that the notation $|h^{-1}|$ stands for the mapping $(t, h) \mapsto |h^{-1}|$, and that the right side is infinite whenever $h \in S_+^K \setminus S_{++}^K$. In (1.11) and throughout the paper, whenever a statement of the form "$\mathcal{P}(x^\alpha)$ holds" appears, with $\alpha \in (0, 1)$, the statement should be understood as "$x \geq 0$ and $\mathcal{P}(x^\alpha)$ holds". For instance, the fact that $\Delta \bar{F}_N + C|h^{-1}| \geq 0$ is implied by (1.11).

In order to prove Proposition 1.2, one needs to find an upper bound on the variance of the K-by-K matrix $x^t \bar{x}$, see Lemma 3.4 below. Compared with the rank-one case, an additional difficulty appears, since the generalization of the rank-one argument only gives information about the symmetric part of the matrix $x^t \bar{x}$. For general spin glass problems, this difficulty was resolved in [15, 16] using relatively involved combinatorial arguments of Ramsey type. In the simpler setting of inference problems, a more direct approach was discovered in [1], and we will essentially follow the same line of reasoning here.

Previous works on the problem, be it rank-one or more general, include [2–4, 6, 8, 12, 13]. We refer to [12] for a precise description of these results.

The rest of the paper is organized as follows. In Section 2, we define the notion of weak solution of (1.8) and show well-posedness of this equation. We next prove Proposition 1.2 in Section 3, and Theorem 1.1 in Section 4. In order to make the paper self-contained, we provide a proof of (1.4) and its generalization to the enriched model in an Appendix.

2. Weak solutions of Hamilton–Jacobi equations.

In this section, we define precisely the notion of weak solution appearing in Theorem 1.1, and prove the well-posedness of the Hamilton–Jacobi equation (1.8). We also discuss, in relation with the specific features of our problem, the advantages of this notion compared with that of viscosity solution.

2.1. *Definition of weak solution.* In order to make the structure of the equation more salient, we give ourselves a function $\mathsf{H} \in C(S_+^K; \mathbb{R})$, and consider equations of the form

$$(2.1) \qquad \partial_t f - \mathsf{H}(\nabla f) = 0 \quad \text{in } \mathbb{R}_+ \times S_+^K.$$

We will always assume that the function H is *convex*. In view of the statement of Theorem 1.1, we are mostly interested in the case when $\mathsf{H}(p) = 2|p|^2$. In order to state the definition of weak solution of (2.1), we introduce, for every $\delta > 0$,

$$(2.2) \qquad S_{+\delta}^K := \delta \mathsf{I}_K + S_+^K,$$

where I_K denotes the K-by-K identity matrix. In words, the set $S_{+\delta}^K$ is the set of symmetric matrices with spectrum in $[\delta, +\infty)$.

DEFINITION 2.1. We say that a Lipschitz function $f : \mathbb{R}_+ \times S_+^K \to \mathbb{R}$ is a *weak solution* of (2.1) if the following conditions hold:

- the relation (2.1) holds almost everywhere in $\mathbb{R}_+ \times S_+^K$;
- for every $t \geq 0$, the mapping $h \mapsto f(t, h)$ is nondecreasing;
- for every $\delta \in (0, 1]$, there exists $C_\delta < \infty$ such that for every $t \in [\delta, \delta^{-1}]$, the mapping $h \mapsto f(t, h) + C_\delta |h|^2$ is convex on the set $\{h \in S_{+\delta}^K : |h| \leq \delta^{-1}\}$.

We refer to the second and third conditions in Definition 2.1 as the *monotonicity* and *local semiconvexity* conditions, respectively. Let us clarify the meaning of the monotonicity condition. For every $A, B \in S^K$, we write $A \leq B$ if and only if $B - A \in S_+^K$. This defines a partial order on S^K. We say that a function $g : S_+^K \to \mathbb{R}$ is nondecreasing if, for every $h, h' \in S_+^K$, we have

$$(2.3) \qquad h \leq h' \implies g(h) \leq g(h').$$

Recall that by the Rademacher theorem, a Lipschitz function is differentiable almost everywhere. In view of the second part of the following elementary lemma, if f is Lipschitz and satisfies the monotonicity condition, it then makes sense to ask about the measure of the set of points where (2.1) holds.

LEMMA 2.2.   (1) *Let* $a \in S^K$. *We have*

$$(2.4) \qquad\qquad a \in S_+^K \quad \Longleftrightarrow \quad \forall b \in S_+^K, \quad a \cdot b \geq 0.$$

(2) *Let* $f : S_+^K \to \mathbb{R}$ *be a Lipschitz function. The function* f *is nondecreasing if and only if, for almost every* $a \in S_+^K$, *we have* $\nabla f(a) \in S_+^K$.

PROOF.   If $a, b \in S_+^K$, then $a \cdot b = |\sqrt{a}\sqrt{b}|^2 \geq 0$, and thus the direct implication in (2.4) holds. For the converse implication, without loss of generality, we can assume that a is a diagonal matrix, in which case the result is easily derived by considering diagonal matrices for b. For part (2), by approximation (see e.g., (2.7) and (2.9) below), we can assume that the function f is smooth. For every $\varepsilon > 0$ and $a, b \in S_+^K$, we have

$$0 \leq \varepsilon^{-1}\big(f(a + \varepsilon b) - f(a)\big),$$

and the right side tends to $\nabla f(a) \cdot b$ as ε tends to 0. Using also (2.4), we obtain the direct implication in part (2). The converse statement follows by writing, for every $a \leq b \in S_+^K$,

$$f(b) - f(a) = \int_0^1 (b - a) \cdot \nabla f\big(a + s(b - a)\big) \, \mathrm{d}s,$$

and using again (2.4) to conclude. $\square$

The monotonicity condition in Definition 2.1 is only really used in a neighborhood of the set $\mathbb{R}_+ \times (S_+^K \setminus S_{++}^K)$, and plays the role of a one-sided boundary condition of Neumann type. We will combine this with the additional assumption that H is nondecreasing to obtain the uniqueness of solutions. In the case of domains without boundary, say $h \in S^K$, then both conditions can be dropped (assuming then that we are given a convex H defined on the entirety of S^K, not just on S_+^K). Moreover, as will be seen in the proof of uniqueness of solutions given below, these conditions can be weakened significantly. Roughly speaking, we need that $\nabla \mathsf{H}(\nabla f) \cdot \mathbf{n}_{S_+^K} \leq 0$ almost everywhere on ∂S_+^K, where $\mathbf{n}_{S_+^K}$ is the unit outer normal to S_+^K (seen as a subset of S^K with Lipschitz boundary).

While this monotonicity condition on weak solutions can be weakened significantly, it cannot be dropped altogether without losing the uniqueness of weak solutions. For an example with $K = 1$, for any $p \in \mathbb{R}_+$, the function $(t, h) \mapsto (\mathsf{H}(-p)t - ph)_+$ satisfies (2.1) almost everywhere, is convex in h, and is constant equal to 0 at $t = 0$. Similarly, the local semiconvexity condition cannot be dropped without losing the uniqueness of weak solutions.

2.2. *Well-posedness and comparison with viscosity-solution approach.*   The next proposition shows that the notion of weak solution introduced in Definition 2.1 indeed ensures the uniqueness of solutions. The proof essentially follows the classical approach of [7, 10, 11], see also [5] and [9], Theorem I.3.3.7, that we adapt to our particular setting (in particular in relation with the boundary condition).

PROPOSITION 2.3 (Uniqueness of weak solutions).   *Let* $\mathsf{H} \in C^2(S_+^K, \mathbb{R})$ *be convex and nondecreasing. If* f *and* g *are two weak solutions to* (2.1) *such that* $f(0, \cdot) = g(0, \cdot)$, *then* $f = g$.

PROOF. We decompose the proof into three steps.

Step 1. We identify an equation satisfied by the difference $w := f - g$. The following identities hold almost everywhere in $\mathbb{R}_+ \times S_+^K$:

$$\partial_t w = \partial_t f - \partial_t g$$

$$= \mathsf{H}(\nabla f) - \mathsf{H}(\nabla g)$$

$$= \int_0^1 \partial_a \big(\mathsf{H}(a\nabla f + (1-a)\nabla g)\big)\, da$$

$$= \int_0^1 \nabla w \cdot \nabla \mathsf{H}(a\nabla f + (1-a)\nabla g)\, da.$$

Setting

$$\mathbf{b} := \int_0^1 \nabla \mathsf{H}(a\nabla f + (1-a)\nabla g)\, da,$$

we thus have that

$$(2.5) \qquad \partial_t w - \mathbf{b} \cdot \nabla w = 0 \quad \text{a.e. in } \mathbb{R}_+ \times S_+^K.$$

Let $\phi \in C^\infty(\mathbb{R})$ be a nonnegative smooth function that will be specified in the course of the argument. We set $v := \phi(w)$, and multiply (2.5) by $\phi'(w)$ to obtain that

$$\partial_t v - \mathbf{b} \cdot \nabla v = 0 \quad \text{a.e. in } \mathbb{R}_+ \times S_+^K.$$

Step 2. Roughly speaking, the idea of the proof is to observe that the "local mass" of v cannot increase much, using integration by parts and the fact that $\nabla \cdot \mathbf{b}$ is bounded below. Since $\nabla \cdot \mathbf{b}$ is not well-defined pointwise, we first regularize f and g.

Let $\zeta \in C_c^\infty(S^K)$ be a smooth function with compact support in $\{h \in S^K : |h| \leq 1\}$ and such that $\int_{S^K} \zeta = 1$. For every $\varepsilon > 0$, we set

$$(2.6) \qquad \zeta_\varepsilon := \varepsilon^{-\frac{K(K+1)}{2}} \zeta\left(\frac{\cdot}{\varepsilon}\right).$$

We define the mollified functions on $\mathbb{R}_+ \times S_{+\varepsilon}^K$

$$(2.7) \qquad f_\varepsilon := f * \zeta_\varepsilon \quad \text{and} \quad g_\varepsilon := g * \zeta_\varepsilon,$$

where $*$ denotes convolution in the h variable only. Explicitly, for every $t \geq 0$ and $h \in S_{+\varepsilon}^K$,

$$f_\varepsilon(t, h) := \int_{S^K} f(t, h - h')\zeta_\varepsilon(h')\, dh',$$

where on the right side, the notation dh' stands for the $\frac{K(K+1)}{2}$-dimensional Lebesgue measure on S^K. The definitions of f_ε and g_ε make sense since whenever $|h'| \leq \varepsilon$ and $h \in S_{+\varepsilon}^K$, we have $\varepsilon I_K - h' \in S_+^K$ and thus $h - h' \in S_+^K$. Moreover, in the case when $h \in S_{+\delta}^K$ for some $\delta \in [2\varepsilon, 1]$, we have $h - h' \in S_{+\frac{\delta}{2}}^K$. By the local semiconvexity assumption on f and g, there exists a constant $C_\delta < \infty$ such that for every $\varepsilon \leq \frac{\delta}{2}$ and $t \in [\delta, \delta^{-1}]$,

$$(2.8) \qquad \text{the mapping } \begin{cases} \{h \in S_{+\delta}^K : |h| \leq \delta^{-1}\} \to \mathbb{R} \\ h \mapsto f_\varepsilon(t, h) + C_\delta|h|^2 \end{cases} \text{ is convex,}$$

and the same property holds with f_ε replaced by g_ε. (Indeed, this property is preserved under convolution.) We also have that

$$(2.9) \qquad \lim_{\varepsilon \to 0} \nabla f_\varepsilon = \nabla f, \qquad \lim_{\varepsilon \to 0} \nabla g_\varepsilon = \nabla g \quad \text{a.e. in } \mathbb{R}_+ \times S_+^K$$

(see e.g. [9], Theorem C.7, for a proof), as well as

$$(2.10) \quad \|\nabla f_\varepsilon\|_{L^\infty(\mathbb{R}_+\times S^K_{+\varepsilon})} \leq \|\nabla f\|_{L^\infty(\mathbb{R}_+\times S^K_+)}, \qquad \|\nabla g_\varepsilon\|_{L^\infty(\mathbb{R}_+\times S^K_{+\varepsilon})} \leq \|\nabla g\|_{L^\infty(\mathbb{R}_+\times S^K_+)}.$$

For future reference, we introduce the shorthand notation

$$(2.11) \quad L := \max\big(\|\nabla f\|_{L^\infty(\mathbb{R}_+\times S^K_+)}, \|\nabla g\|_{L^\infty(\mathbb{R}_+\times S^K_+)}\big).$$

Throughout the proof, we enforce without further mention that $\delta > 0$ is sufficiently small that $\delta^{-1} \geq L$. Identifying $(S^K, |\cdot|)$ with the Euclidean space $\mathbb{R}^{\frac{K(K+1)}{2}}$, we denote the Hessian of a function $\widetilde{f}$ defined on an open subset of S^K by $\nabla^2\widetilde{f}$. For the purposes of this proof, we think of $\nabla^2\widetilde{f}$ as being a $\frac{K(K+1)}{2}$-by-$\frac{K(K+1)}{2}$ symmetric matrix. By (2.8), we have, for every $\varepsilon \leq \frac{\delta}{2}$, $t \in [\delta, \delta^{-1}]$ and $h \in S^K_{+\delta}$ with $|h| \leq \delta^{-1}$,

$$(2.12) \quad \nabla^2 f_\varepsilon(t, h) + 2C_\delta I_{\frac{K(K+1)}{2}} \in S^{\frac{K(K+1)}{2}}_+ \quad \text{and} \quad \nabla^2 g_\varepsilon(t, h) + 2C_\delta I_{\frac{K(K+1)}{2}} \in S^{\frac{K(K+1)}{2}}_+.$$

We set

$$\mathbf{b}_\varepsilon := \int_0^1 \nabla\mathsf{H}(a\nabla f_\varepsilon + (1-a)\nabla g_\varepsilon)\,\mathrm{d}a,$$

and observe that

$$\partial_t v - \mathbf{b}_\varepsilon \cdot \nabla v = (\mathbf{b} - \mathbf{b}_\varepsilon)\cdot\nabla v \quad \text{a.e. in } \mathbb{R}_+\times S^K_{+\varepsilon},$$

that is,

$$(2.13) \quad \partial_t v - \nabla\cdot(\mathbf{b}_\varepsilon v) + v\nabla\cdot\mathbf{b}_\varepsilon = (\mathbf{b}-\mathbf{b}_\varepsilon)\cdot\nabla v \quad \text{a.e. in } \mathbb{R}_+\times S^K_{+\varepsilon}.$$

Moreover,

$$\nabla\cdot\mathbf{b}_\varepsilon = \int_0^1 \big(a\nabla^2 f_\varepsilon + (1-a)\nabla^2 g_\varepsilon\big)\cdot\nabla^2\mathsf{H}(a\nabla f_\varepsilon + (1-a)\nabla g_\varepsilon)\,\mathrm{d}a,$$

where, as for $\nabla^2 f_\varepsilon$ and $\nabla^2 g_\varepsilon$, we think of $\nabla^2\mathsf{H}$ as being a $\frac{K(K+1)}{2}$-by-$\frac{K(K+1)}{2}$ symmetric matrix. Since H is convex, the $\frac{K(K+1)}{2}$-by-$\frac{K(K+1)}{2}$ matrix $\nabla^2\mathsf{H}(p)$ is positive semidefinite for every $p \in S^K_+$. It thus follows from (2.4) and (2.12) that for every $\varepsilon \leq \frac{\delta}{2}$, we have

$$\int_0^1 \big(a\nabla^2 f_\varepsilon + (1-a)\nabla^2 g_\varepsilon + 2C_\delta I_{\frac{K(K+1)}{2}}\big)\cdot\nabla^2\mathsf{H}(a\nabla f_\varepsilon + (1-a)\nabla g_\varepsilon)\,\mathrm{d}a \geq 0 \quad \text{on } [\delta, \delta^{-1}]\times S^K_{+\delta}.$$

Hence, by (2.10) and (2.11), we get that for every $\varepsilon \leq \frac{\delta}{2}$,

$$\nabla\cdot\mathbf{b}_\varepsilon \geq -2C_\delta\sup\{|\nabla^2\mathsf{H}(p)| : |p| \leq L\} \quad \text{on } [\delta, \delta^{-1}]\times S^K_{+\delta}.$$

Up to a redefinition of $C_\delta < \infty$, we may thus assume that for every $\varepsilon \leq \frac{\delta}{2}$,

$$(2.14) \quad \nabla\cdot\mathbf{b}_\varepsilon + C_\delta \geq 0 \quad \text{on } [\delta, \delta^{-1}]\times S^K_{+\delta}.$$

Step 3. We are now ready to implement the argument announced at the beginning of Step 2. Denote

$$R := 1 + \sup\{|\nabla\mathsf{H}(p)| : |p| \leq L\}.$$

We fix $T \geq 1$ and define, for every $t \in [0, \frac{T}{2}]$,

$$(2.15) \quad B_\delta(t) := \{h \in S^K_{+\delta} : |h| \leq R(T-t)\},$$

$$(2.16) \quad \partial_+ B_\delta(t) := \{h \in S^K_{+\delta} : |h| = R(T-t)\},$$

and

$$(2.17) \qquad \partial_0 B_\delta(t) := \{h \in \partial S_{+\delta}^K : |h| < R(T-t)\}.$$

We assume without further mention that $\delta > 0$ is sufficiently small that $\delta^{-1} \geq RT$. Notice that the boundary of $B_\delta(t)$ is the disjoint union of $\partial_+ B_\delta(t)$ and $\partial_0 B_\delta(t)$. We aim to obtain a Gronwall inequality for the quantity

$$(2.18) \qquad J_\delta(t) := \int_{B_\delta(t)} v(t, \cdot) = \int_{B_\delta(t)} v(t, h)\, dh \quad \left(t \in \left[0, \frac{T}{2}\right]\right).$$

The function J_δ is Lipschitz, and for almost every $t \in [0, \frac{T}{2}]$, we have

$$\partial_t J_\delta(t) = \int_{B_\delta(t)} \partial_t v(t, \cdot) - R \int_{\partial_+ B_\delta(t)} v(t, \cdot),$$

where the second integral is a boundary integral (with respect to the $\frac{K(K-1)}{2}$-dimensional Hausdorff measure on $\partial_+ B_\delta(t)$). Using (2.13), (2.14), and that $v \geq 0$, we get that for almost every $t \in [\delta, \frac{T}{2}]$,

$$\partial_t J_\delta(t) = \int_{B_\delta(t)} \left((\mathbf{b} - \mathbf{b}_\varepsilon) \cdot \nabla v - v \nabla \cdot \mathbf{b}_\varepsilon + \nabla \cdot (\mathbf{b}_\varepsilon v)\right)(t, \cdot) - R \int_{\partial_+ B_\delta(t)} v(t, \cdot)$$

$$\leq C_\delta J_\delta(t) + \int_{B_\delta(t)} \left((\mathbf{b} - \mathbf{b}_\varepsilon) \cdot \nabla v\right)(t, \cdot)$$

$$+ \int_{\partial_+ B_\delta(t)} \left((\mathbf{b}_\varepsilon \cdot \mathbf{n} - R) v\right)(t, \cdot) + \int_{\partial_0 B_\delta(t)} (v \mathbf{b}_\varepsilon \cdot \mathbf{n})(t, \cdot),$$

where we denote by $\mathbf{n}$ the unit outer normal to $B_\delta(t)$. Using (2.9) and the dominated convergence theorem, we see that the first integral on the right side above tends to 0 as ε tends to 0. The first boundary integral on the right side above is nonpositive, by the definitions of $\mathbf{b}_\varepsilon$ and R, and (2.10). We now show that

$$(2.19) \qquad \int_{\partial_0 B_\delta(t)} (v \mathbf{b}_\varepsilon \cdot \mathbf{n})(t, \cdot) \leq 0.$$

We first notice that $-\mathbf{n} \in S_+^K$ almost everywhere on $\partial_0 B_\delta(t)$. By (2.4), in order to show (2.19), it suffices to verify that $\mathbf{b}_\varepsilon \in S_+^K$. This follows from the assumption that H, f and g are nondecreasing and an application of Lemma 2.2. Summarizing, we have shown that for almost every $t \in [\delta, \frac{T}{2}]$,

$$(2.20) \qquad \partial_t J_\delta(t) \leq C_\delta J_\delta(t).$$

Finally, we observe that

$$w(\delta, h) = (f - g)(\delta, h) \leq \delta \|\partial_t (f - g)\|_{L^\infty(\mathbb{R}_+ \times S_+^K)}.$$

Selecting $\phi = \phi_\delta \in C^\infty(\mathbb{R})$ in such a way that, for every $x \in \mathbb{R}$,

$$\phi_\delta(x) > 0 \quad \Longleftrightarrow \quad |x| > \delta \|\partial_t (f - g)\|_{L^\infty(\mathbb{R}_+ \times S_+^K)}$$

ensures that $J_\delta(\delta) = 0$. Combined with (2.20) and the fact that $J_\delta \geq 0$, this yields that for every $t \in [\delta, \frac{T}{2}]$,

$$J_\delta(t) = 0,$$

and in particular, for every $t \in [\delta, \frac{T}{2}]$ and almost every $h \in B_\delta(t)$,

$$|(f - g)(t, h)| \leq \delta \|\partial_t (f - g)\|_{L^\infty(\mathbb{R}_+ \times S_+^K)}.$$

Letting δ tend to 0 yields the desired result. $\quad \square$

As long as the assumptions of monotonicity and convexity of H are satisfied, the notion of weak solution in Definition 2.1 turns out to be much more convenient to work with than the notion of viscosity solution employed in [14], as we now explain.

To start with, the monotonicity condition on the solution allows to circumvent the relatively cumbersome treatment of the viscosity-solution interpretation of the Neumann boundary condition used in [14], replacing it with a straightforward verification of the fact that $h \mapsto \bar{F}_N(t, h)$ is nondecreasing for every N.

More significantly, as was shown in [14], Section 4, the proper treatment of odd-degree tensor versions of this problem no longer involve one-sided estimates of the general form $0 \leq \partial_t \bar{F}_N - H(\bar{F}_N) \leq N^{-1} \Delta \bar{F}_N + \cdots$, but rather two-sided estimates of the form $|\partial_t \bar{F}_N - H(\bar{F}_N)| \leq N^{-1} \Delta \bar{F}_N + \cdots$ (whether or not the right-hand side is raised to a power $\alpha \in (0, 1)$ is irrelevant to this discussion, so we ignore it). In principle, this is a worrisome situation, since viscosity solutions are meant to "remember the sign of the Laplacian" in the vanishing viscosity limit. The reason why the proof could still be successfully carried out in spite of this is precisely by leveraging on the additional information that $\bar{F}_N$ is locally semiconvex. Arguments based on the notion of weak solution use this property in a much more transparent way.

Finally, arguments based on weak solutions are more adapted to the type of error terms that appear in the right-hand side of (1.11). The proof of convergence of $\bar{F}_N$ will be a more quantitative version of the argument in Proposition 2.3: instead of showing that two weak solutions must be equal, we will show that two "almost weak solutions" must be almost equal. This proof is most suited to accommodate for error terms that are estimated in an $L_t^\infty L_h^1$-type norm, see for instance the definition of J_δ in (2.18). The error terms on the right side of (1.11), in particular $|\nabla F_N - \nabla \bar{F}_N|^2$ (but also $N^{-1} \Delta \bar{F}_N$), can be estimated in such a norm in a straightforward way. On the other hand, the notion of viscosity solution handles most naturally errors that are estimated in $L_t^\infty L_h^\infty$. While this problem can be (and has been) circumvented by appealing to local convolutions, the approach based on weak solutions is more straightforward and easily yields quantitative estimates.

To conclude this section, we give the Hopf–Lax formula for weak solutions to (2.1), which in particular proves the existence of solutions to (2.1). Except for the treatment of the boundary condition, the argument is classical. We denote by H^* the convex dual of H: that is, for each $q \in S^K$, we set

$$(2.21) \qquad H^*(q) := \sup_{p \in S_+^K} \left(p \cdot q - H(p) \right).$$

PROPOSITION 2.4 (Hopf–Lax formula). *Let $H \in C(S_+^K; \mathbb{R})$ be a convex function such that $H(p)$ depends only on $|p|$. Let $\psi : S_+^K \to \mathbb{R}$ be a nondecreasing Lipschitz function satisfying the following local semiconvexity property: for every $\delta > 0$, there exists a constant $C_\delta < \infty$ such that*

$$(2.22) \qquad \textit{the mapping } \begin{cases} S_{+\delta}^K \to \mathbb{R} \\ h \mapsto \psi(h) + C_\delta |h|^2 \end{cases} \textit{ is convex.}$$

For each $t \geq 0$ and $h \in S_+^K$, we define

$$(2.23) \qquad f(t, h) := \sup_{h' \in S_+^K} \left(\psi(h') - t H^* \left(\frac{h' - h}{t} \right) \right),$$

with the understanding that $f(0, \cdot) = \psi$. The function f is a weak solution of (2.1).

REMARK 2.5. The local semiconvexity assumption (2.22) of the initial condition is used to show that the mapping $h \mapsto f(t, h)$ is locally semiconvex. This property can also be obtained by assuming instead that H is uniformly convex, see [9], Lemma 3.3.4. However, it is convenient to state Proposition 2.4 in this way because the explicit local semiconvexity property of f as stated in (2.33) below will be used later to obtain a quantitative rate of convergence in Theorem 1.1. The assumption that $H(p)$ depends only on $|p|$ simplifies the consideration of problems related to the presence of a boundary. It can certainly be weakened significantly, but I do not know whether it can be removed altogether.

PROOF OF PROPOSITION 2.4. We decompose the proof into five steps.

Step 1. In this first step, we use that ψ is nondecreasing and the symmetry assumption on H to assert that, for every $t \geq 0$ and $h \in S_+^K$,

$$(2.24) \qquad f(t, h) = \sup\left\{ \psi(h') - tH^*\left(\frac{h' - h}{t}\right) : h' \in S_+^K \text{ s.t. } h' \geq h \right\}$$

$$(2.25) \qquad = \sup\left\{ \psi(h + h') - tH^*\left(\frac{h'}{t}\right) : h' \in S_+^K \right\}.$$

For every $h \in S^K$, we write h_+ to denote the image of h under the mapping $x \mapsto \max(x, 0)$. In a basis where h is diagonal, this means that we replace the negative eigenvalues by zeros. We first show that, for every $h \in S^K$,

$$(2.26) \qquad H^*(h) = H^*(h_+).$$

This statement is equivalent to

$$(2.27) \qquad \sup_{p \in S_+^K} \left(p \cdot h - H(p)\right) = \sup_{p \in S_+^K} \left(p \cdot h_+ - H(p)\right).$$

Since $h \leq h_+$, the statement (2.27) with the inequality $\leq$ in place of the equality is clear by (2.4). Conversely, for every $p' \in S_+^K$, we may choose $p := \frac{h_+}{|h_+|}|p'| \in S_+^K$, so that

$$|p| = |p'| \quad \text{and} \quad p \cdot h = |p'|\frac{h \cdot h_+}{|h_+|} = |p'||h_+|.$$

Since we assume that $H(p)$ depends only on $|p|$, and since $p' \cdot h_+ \leq |p'||h_+|$, this proves the inequality $\geq$ in (2.27), and therefore (2.26).

Using (2.26), we get that for every $h, h' \in S_+^K$,

$$\psi(h') - tH^*\left(\frac{h' - h}{t}\right) = \psi(h') - tH^*\left(\frac{(h' - h)_+}{t}\right)$$

$$\leq \psi\left(h + (h' - h)_+\right) - tH^*\left(\frac{h + (h' - h)_+ - h}{t}\right),$$

where we also used that $h' - h \leq (h' - h)_+$ and that ψ is nondecreasing in the last step. The identities (2.24)–(2.25) thus follow. From (2.25), it is clear that the mapping $h \mapsto f(t, h)$ is nondecreasing.

Step 2. We prove the dynamic programming principle, that is, for every $s, t \geq 0$ and $h \in S_+^K$,

$$(2.28) \qquad f(t + s, h) = \sup_{h' \in S_+^K} \left(f(t, h') - sH^*\left(\frac{h' - h}{s}\right) \right).$$

Since H* is convex, we have, for every $s, t \geq 0$ and $h, h', h'' \in S_+^K$,

$$(2.29) \qquad \mathsf{H}^*\left(\frac{h'-h}{t+s}\right) \leq \frac{t}{t+s}\mathsf{H}^*\left(\frac{h'-h''}{t}\right) + \frac{s}{t+s}\mathsf{H}^*\left(\frac{h''-h}{s}\right),$$

and thus

$$(2.30) \qquad f(t+s,h) \geq \sup_{h',h'' \in S_+^K}\left(\psi(h') - t\mathsf{H}^*\left(\frac{h'-h''}{t}\right) - s\mathsf{H}^*\left(\frac{h''-h}{s}\right)\right)$$

$$(2.31) \qquad \qquad \geq \sup_{h'' \in S_+^K}\left(f(t,h'') - s\mathsf{H}^*\left(\frac{h''-h}{s}\right)\right).$$

Conversely, for each fixed $h' \in S_+^K$, we select $h'' := h + \frac{s}{t+s}(h'-h)$. Notice that h'' lies on the segment joining h and h', and since S_+^K is a convex set, we deduce that $h'' \in S_+^K$. Moreover, we have

$$\frac{h'-h}{t+s} = \frac{h'-h''}{t} = \frac{h''-h}{s},$$

and thus the general inequality in (2.29) becomes an equality. This implies that

$$\sup_{h'' \in S_+^K}\left(\psi(h') - t\mathsf{H}^*\left(\frac{h'-h''}{t}\right) - s\mathsf{H}^*\left(\frac{h''-h}{s}\right)\right) \geq \psi(h') - (t+s)\mathsf{H}^*\left(\frac{h'-h}{t+s}\right).$$

Notice also that the two terms on the right side of (2.30) and (2.31) are equal. Taking the supremum over $h' \in S_+^K$ in the display above thus completes the proof of (2.28).

Step 3. We show that f is Lipschitz continuous. It follows from (2.25) that for every $h, h_1 \in S_+^K$,

$$f(t,h) \leq f(t,h_1) + \|\nabla\psi\|_{L^\infty}|h-h_1|,$$

and by symmetry, that

$$(2.32) \qquad \left|f(t,h) - f(t,h_1)\right| \leq \|\nabla\psi\|_{L^\infty}|h-h_1|.$$

This shows that f is Lipschitz continuous in the h variable. For the regularity in time, we recall that (2.28) also holds with the additional restriction $h' \geq h$, and appeal to (2.32) to write, for every $s, t \geq 0$,

$$f(t,h) - s\mathsf{H}^*(0) \leq f(t+s,h) \leq \sup_{h' \geq h}\left(f(t,h) + \|\nabla\psi\|_{L^\infty}|h'-h| - s\mathsf{H}^*\left(\frac{h'-h}{s}\right)\right).$$

Recalling the definition of H* in (2.21), we see that $\mathsf{H}^*(0) \geq -\mathsf{H}(0)$. Moreover, testing the supremum in this definition with $p = \lambda\frac{h'-h}{|h'-h|} \in S_+^K$ for some $\lambda > 0$ to be determined, we obtain that for every $h' \geq h \in S_+^K$,

$$s\mathsf{H}^*\left(\frac{h'-h}{s}\right) \geq \lambda|h'-h| - s\mathsf{H}\left(\lambda\frac{h'-h}{|h'-h|}\right).$$

Selecting $\lambda = \|\nabla\psi\|_{L^\infty}$, we conclude that

$$0 \leq f(t+s,h) - f(t,h) \leq s\sup\{\mathsf{H}(p) : p \in S_+^K, |p| \leq \|\nabla\psi\|_{L^\infty}\}.$$

This shows in particular that the function f is Lipschitz continuous in the t variable.

Step 4. We show that for every $t \geq 0$ and $\delta > 0$,

$$(2.33) \qquad \text{the mapping } \begin{cases} S_{+2\delta}^K \to \mathbb{R} \\ h \mapsto f(t,h) + C_\delta|h|^2 \end{cases} \text{ is convex,}$$

where C_δ is the constant appearing in (2.22). Reproducing the argument in the previous step with $\lambda = 1 + \|\nabla \psi\|_{L^\infty}$, we see that the supremum in (2.24) is attained. Fix $t > 0$, $h \in S^K_{+2\delta}$, and denote by $h' \in S^K_+$ a point realizing the supremum in (2.25):

$$f(t, h) = \psi(h + h') - tH^* \left(\frac{h'}{t} \right)$$

for every $h'' \in S^K$ sufficiently small (in terms of δ) to guarantee that $h - h''$ and $h + h''$ belong to $S^K_{+\delta}$, we have

$$f(t, h + h'') + f(t, h - h'') - 2f(t, h) \geq \psi(h + h' + h'') + \psi(h + h' - h'') - 2\psi(h + h').$$

(This follows by writing (2.25) for $f(t, h + h'')$ and $f(t, h - h'')$ and testing the supremum at h'.) Recalling that $h' \in S^K_+$ and using (2.22), we deduce that

$$f(t, h + h'') + f(t, h - h'') - 2f(t, h) \geq -2C_\delta |h''|^2.$$

This proves (2.33).

Step 5. Since f is Lipschitz continuous, it is differentiable almost everywhere. In this final step, we show that for every $t > 0$ and $h \in S^K_{++}$, if (t, h) is a point of differentiability of f, then the equation (2.1) is satisfied at (t, h). We fix such (t, h), and for every $q \in S^K$ and $s > 0$ sufficiently small, we use (2.28) to write

$$f(t + s, h) \geq f(t, h + sq) - sH^*(q).$$

Letting s tend to 0, we infer that

$$(2.34) \qquad \partial_t f(t, h) - q \cdot \nabla f(t, h) + H^*(q) \geq 0.$$

We may view the function H as defined on S^K, by setting $H(p) = +\infty$ whenever $p \notin S^K_+$. This function is convex and lower semicontinuous, and thus, for every $p \in S^K$,

$$(2.35) \qquad H^{**}(p) := \sup_{q \in S^K} (p \cdot q - H^*(q)) = H(p).$$

Taking the infimum over q in (2.34), we thus conclude that

$$(2.36) \qquad (\partial_t f - H(\nabla f))(t, h) \geq 0.$$

(Since f is nondecreasing, we already knew by Lemma 2.2 that $\nabla f(t, h) \in S^K_+$.) There remains to show the converse inequality to (2.36). Let $h' \in S^K_+$ be such that

$$f(t, h) = \psi(h') - tH^* \left(\frac{h' - h}{t} \right).$$

For every $s > 0$ sufficiently small, we have

$$f \left(t - s, h + \frac{s}{t}(h' - h) \right) \geq \psi(h') - (t - s)H^* \left(\frac{h' - h - \frac{s}{t}(h' - h)}{t - s} \right)$$

$$\geq f(t, h) + sH^* \left(\frac{h' - h}{t} \right).$$

Letting s tend to 0, we obtain that

$$\partial_t f(t, h) - \frac{h' - h}{t} \cdot \nabla f(t, h) + H^* \left(\frac{h' - h}{t} \right) \leq 0.$$

Together with (2.35), this yields the converse inequality to (2.36) and thus completes the proof. $\square$

3. Approximate Hamilton–Jacobi equation and basic estimates. The main goal of this section is to prove Proposition 1.2. We will also record basic derivative and concentration estimates that will be useful in the next section.

3.1. *Nishimori identity.* We start by presenting the Nishimori identity, a simple but crucial property which allows to avoid facing a never-ending cascade of new replicas as we differentiate the free energy. We denote by $\langle \cdot \rangle$ the Gibbs measure associated with the energy $H_N(t, h, \cdot)$. That is, for each bounded measurable function $f : \mathbb{R}^{N \times K} \to \mathbb{R}$, we set

$$(3.1) \qquad \langle f(x) \rangle := \frac{1}{Z_N(t, h)} \int_{\mathbb{R}^{N \times K}} f(x) e^{H_N(t, h, x)} \, dP_N(x),$$

where

$$Z_N(t, h) := \int_{\mathbb{R}^{N \times K}} e^{H_N(t, h, x)} \, dP_N(x).$$

Note that although the notation does not display it, this random probability measure depends on t, h, as well as on the realization of the random variables $\bar{x}$, W and z. We also consider "replicated" (or tensorized) versions of this measure, and write x, x', x'', etc. for the canonical "replicated" random variables. Conditionally on $\bar{x}$, W and z, these random variables are independent and each is distributed according to the Gibbs measure $\langle \cdot \rangle$. With a slight abuse of notation, we still denote this tensorized measure by $\langle \cdot \rangle$. That is, for every bounded measurable function $f : (\mathbb{R}^{N \times K})^2 \to \mathbb{R}$, we denote

$$\langle f(x, x') \rangle = \frac{1}{Z_N^2(t, h)} \int_{(\mathbb{R}^{N \times K})^2} f(x, x') e^{H_N(t, h, x) + H_N(t, h, x')} \, dP_N(x) \, dP_N(x'),$$

and so on for more than two replicas.

As discussed in the Introduction and shown in the Appendix, the measure $\langle \cdot \rangle$ is the conditional measure of $\bar{x}$ given $\mathcal{Y} := (Y, Y')$. That is, for every bounded measurable function $f : \mathbb{R}^{N \times K} \to \mathbb{R}$,

$$(3.2) \qquad \langle f(x) \rangle = \mathbb{E}[f(\bar{x}) | \mathcal{Y}].$$

In particular, we have

$$\mathbb{E}\langle f(x) \rangle = \mathbb{E}[f(\bar{x})].$$

Using the conditional independence between replicas, we also have, for any bounded measurable functions $f_1, f_2 : \mathbb{R}^{N \times K} \to \mathbb{R}$,

$$\mathbb{E}\langle f_1(x) f_2(x') \rangle = \mathbb{E}[\langle f_1(x) \rangle \langle f_2(x) \rangle]$$
$$= \mathbb{E}[\langle f_1(x) \rangle f_2(\bar{x})].$$

By the monotone class lemma, this implies that for any bounded measurable function $f : (\mathbb{R}^{N \times K})^2 \to \mathbb{R}$,

$$(3.3) \qquad \mathbb{E}\langle f(x, x') \rangle = \mathbb{E}\langle f(x, \bar{x}) \rangle.$$

This identity can be generalized to more than two replicas: for instance,

$$\mathbb{E}\langle f(x, x', x'') \rangle = \mathbb{E}\langle f(x, x', \bar{x}) \rangle.$$

This relation, regardless of the number of replicas involved, is often called the *Nishimori identity*. This property will be the crucial ingredient allowing us to "close the equation" and show that the system stays in a replica-symmetric phase. We will repeatedly use it without further mention. Notice that we can also incorporate dependencies in $\mathcal{Y}$, that is:

$$\mathbb{E}\langle f(\mathcal{Y}, x, x') \rangle = \mathbb{E}\langle f(\mathcal{Y}, x, \bar{x}) \rangle,$$

and so on with more replicas.

3.2. *Matrix square root.*   We record here some elementary properties of the matrix square root that will be useful in the sequel. Denote by $D_{\sqrt{h}}$ the differential of the square-root function, seen as a mapping from S_{++}^K to S^K, at the point $h \in S_{++}^K$. This differential is a linear mapping on S^K which satisfies, for every $h \in S_{++}^K$ and $a \in S_K$,

$$D_{\sqrt{h}}(a) = \lim_{\varepsilon \to 0} \varepsilon^{-1}(\sqrt{h + \varepsilon a} - \sqrt{h}).$$

Notice that

$$h + \varepsilon a = (\sqrt{h + \varepsilon a})^2 = (\sqrt{h} + \varepsilon D_{\sqrt{h}}(a) + o(\varepsilon))^2$$
$$= h + \varepsilon(\sqrt{h} D_{\sqrt{h}}(a) + D_{\sqrt{h}}(a)\sqrt{h}) + o(\varepsilon),$$

and thus

$$(3.4) \qquad \sqrt{h} D_{\sqrt{h}}(a) + D_{\sqrt{h}}(a)\sqrt{h} = a.$$

Although we will not need this fact, it is interesting to note that this property characterizes $D_{\sqrt{h}}$. This tells us that any identity involving $D_{\sqrt{h}}$ is provable using only (3.4) as a definition of $D_{\sqrt{h}}$.

LEMMA 3.1 (Unique Jordan inverse).   *Let* $A \in S_{++}^K$ *and* $B, C \in \mathbb{R}^{K \times K}$ *be such that*

$$(3.5) \qquad AB + BA = AC + CA.$$

Then $B = C$.

PROOF.   Without loss of generality, we can assume that $C = 0$ and that A is a diagonal matrix, with positive eigenvalues $\lambda_1, \dots, \lambda_K$. The condition (3.5) then reads, for $B = (B_{kl})_{1 \le k, l \le K}$,

$$\forall 1 \le k, l \le K, \quad B_{kl}(\lambda_k + \lambda_l) = 0.$$

Since $\lambda_k + \lambda_l > 0$, this implies that $B = 0$. $\square$

It follows from (3.4) that whenever $b \in S^K$, we have

$$(3.6) \qquad (D_{\sqrt{h}}(a)\sqrt{h}) \cdot b = (\sqrt{h} D_{\sqrt{h}}(a)) \cdot b = \frac{a \cdot b}{2}.$$

When $b \in \mathbb{R}^{K \times K}$ is not assumed to be symmetric, we can estimate the error in this relation in terms of the size of the antisymmetric part of b. Recall from (1.10) that we denote by $\kappa(h)$ the condition number of a matrix $h \in S_+^K$.

LEMMA 3.2.   *There exists* $C < \infty$ *such that for every* $h \in S_{++}^K$, $a \in S^K$ *and* $b \in \mathbb{R}^{K \times K}$,

$$\left| D_{\sqrt{h}}(a)\sqrt{h} \cdot b - \frac{a \cdot b}{2} \right| + \left| \sqrt{h} D_{\sqrt{h}}(a) \cdot b - \frac{a \cdot b}{2} \right| \le C\sqrt{\kappa(h)}|a||b - b^t|.$$

PROOF.   We first show that

$$(3.7) \qquad \left| D_{\sqrt{h}}(a) \right| \le C|a||h^{-1}|^{\frac{1}{2}}.$$

The proof is similar to that of Lemma 3.1. Without loss of generality, we assume that h is diagonal, with positive eigenvalues $0 < \lambda_1 \le \cdots \le \lambda_K$. Denoting $(d_{kl})_{1 \le k, l \le K}$ the entries of the matrix $D_{\sqrt{h}}(a)$, and $a = (a_{kl})_{1 \le k, l \le K}$, the relation (3.4) reads

$$d_{kl}(\sqrt{\lambda_k} + \sqrt{\lambda_l}) = a_{kl}.$$

We deduce that

$$\max_{1 \le k,l \le K} |\mathsf{d}_{kl}| \le \lambda_1^{-\frac{1}{2}} \max_{1 \le k,l \le K} |a_{kl}| \le \left(\sum_{k=1}^K \lambda_k^{-2} \right)^{\frac{1}{4}} \max_{1 \le k,l \le K} |a_{kl}|,$$

and thus that (3.7) holds, by equivalence of norms. The conclusion of the lemma then follows using the decomposition

$$2D_{\sqrt{h}}(a)\sqrt{h} \cdot b = D_{\sqrt{h}}(a)\sqrt{h} \cdot (b + b^{\mathsf{t}}) + D_{\sqrt{h}}(a)\sqrt{h} \cdot (b - b^{\mathsf{t}}),$$

and that by (3.6), the first term on the right side is $a \cdot \frac{b+b^{\mathsf{t}}}{2} = a \cdot b$, since $a \in S^K$. $\square$

3.3. *Proof of Proposition* 1.2.   For convenience, we now record some identities that follow from Gaussian integration by parts. Recall that $z = (z_{i,k})_{1 \le i \le N, 1 \le k \le K}$ is an N-by-K matrix of independent standard Gaussian entries, independent of $(\bar{x}, W)$.

LEMMA 3.3 (Gaussian integration by parts).   *For every bounded measurable function* $F : (\mathbb{R}^{N \times K})^2 \to \mathbb{R}^{N \times K}$, *we have*

(3.8)
$$\mathbb{E}\langle z \cdot F(x, \bar{x}) \rangle = \mathbb{E}\langle (x - x')\sqrt{h} \cdot F(x, \bar{x}) \rangle$$

and

(3.9) $$\mathbb{E}\langle (z \cdot F(x, \bar{x}))^2 \rangle = \mathbb{E}\langle ((x - x')\sqrt{h} \cdot F(x, \bar{x}))(z \cdot F(x, \bar{x})) \rangle + \mathbb{E}\langle |F(x, \bar{x})|^2 \rangle,$$

while for $F : (\mathbb{R}^{N \times K})^3 \to \mathbb{R}^{N \times K}$,

(3.10)
$$\mathbb{E}\langle z \cdot F(x, x', \bar{x}) \rangle = \mathbb{E}\langle (x + x' - 2x'')\sqrt{h} \cdot F(x, x', \bar{x}) \rangle.$$

Finally,

(3.11)
$$\mathbb{E}\langle x \cdot Wx \rangle = \sqrt{\frac{t}{N}} \mathbb{E}\langle |x^{\mathsf{t}}x|^2 - |x^{\mathsf{t}}\bar{x}|^2 \rangle.$$

PROOF.   We will use repeatedly the following simple "Gaussian integration by parts": for every $g \in C^{0,1}(\mathbb{R}; \mathbb{R})$,

$$\int_{\mathbb{R}} g(s)s e^{-\frac{s^2}{2}} \, ds = \int_{\mathbb{R}} g'(s) e^{-\frac{s^2}{2}} \, ds.$$

We write $F = (F_{ik})_{1 \le i \le N, 1 \le k \le K}$ and notice that, by the observation above,

$$\mathbb{E}[z_{ik}\langle F_{ik}(x, \bar{x}) \rangle] = \mathbb{E}[\partial_{z_{ik}}\langle F_{ik}(x, \bar{x}) \rangle]$$
$$= \mathbb{E}\langle F_{ik}(x, \bar{x})((x - x')\sqrt{h})_{ik} \rangle.$$

Summing over (i, k), we obtain (3.8). The proof of (3.10) is similar: we write

$$\mathbb{E}[z_{ik}\langle F_{ik}(x, x', \bar{x}) \rangle] = \mathbb{E}[\partial_{z_{ik}}\langle F_{ik}(x, x', \bar{x}) \rangle]$$
$$= \mathbb{E}\langle F_{ik}(x, x', \bar{x})((x + x' - 2x'')\sqrt{h})_{ik} \rangle,$$

and then sum over (i, k). For (3.9), we calculate first, for any bounded measurable $f : (\mathbb{R}^{N \times K})^2 \to \mathbb{R}$ and any $(i, k), (j, l) \in \{1, \ldots, N\} \times \{1, \ldots, K\}$ satisfying $(i, k) \ne (j, l)$,

$$\mathbb{E}[z_{ik}z_{jl}\langle f(x, \bar{x}) \rangle] = \mathbb{E}[z_{jl}\partial_{z_{ik}}\langle f(x, \bar{x}) \rangle]$$
$$= \mathbb{E}[z_{jl}\langle f(x, \bar{x})((x - x')\sqrt{h})_{ik} \rangle].$$

When $(i, k) = (j, l)$ this relation becomes

$$\mathbb{E}[z_{ik}^2 \langle f(x, \bar{x}) \rangle] = \mathbb{E}[z_{ik} \langle f(x, \bar{x})((x - x')\sqrt{h})_{ik} \rangle] + \mathbb{E}\langle f(x, \bar{x}) \rangle.$$

Replacing $f(x, \bar{x})$ by $F_{ik}(x, \bar{x})F_{jl}(x, \bar{x})$ and summing over all indices, we obtain (3.9). Similarly, we observe that

$$\mathbb{E}\langle x \cdot Wx \rangle = \mathbb{E}\langle W \cdot xx^{\mathrm{t}} \rangle = \sqrt{\frac{t}{N}} \mathbb{E}[\langle |xx^{\mathrm{t}}|^2 \rangle - |\langle xx^{\mathrm{t}} \rangle|^2],$$

and since

$$|\langle xx^{\mathrm{t}} \rangle|^2 = \langle xx^{\mathrm{t}} \cdot \bar{x}\bar{x}^{\mathrm{t}} \rangle = \langle |x^{\mathrm{t}}\bar{x}|^2 \rangle,$$

we obtain (3.11). □

We next present an intermediate result towards the proof of Proposition 1.2, which is interesting on its own in that it displays the relevance of the question of assessing the concentration of the matrix $x^{\mathrm{t}}\bar{x} \in \mathbb{R}^{K \times K}$.

LEMMA 3.4. *We have*

$$(3.12) \qquad \partial_t \bar{F}_N - 2|\nabla \bar{F}_N|^2 = \frac{1}{2N^2} \mathbb{E}\langle |x^{\mathrm{t}}\bar{x} - \mathbb{E}\langle x^{\mathrm{t}}\bar{x} \rangle|^2 \rangle.$$

Moreover, for every $t \geq 0$, the mapping $h \mapsto \bar{F}_N(t, h)$ is nondecreasing.

PROOF. Starting with the derivative with respect to t, we have

$$(3.13) \qquad \partial_t F_N(t, h) = \frac{1}{N} \left\langle \frac{1}{2\sqrt{tN}} x \cdot Wx + \frac{1}{N} |x^{\mathrm{t}}\bar{x}|^2 - \frac{1}{2N} |x^{\mathrm{t}}x|^2 \right\rangle.$$

By (3.11), we deduce that

$$(3.14) \qquad \partial_t \bar{F}_N(t, h) = \frac{1}{2N^2} \mathbb{E}\langle |x^{\mathrm{t}}\bar{x}|^2 \rangle.$$

We also have, for every $h \in S_{++}^K$ and $a \in S^K$,

$$(3.15) \qquad a \cdot \nabla F_N(t, h) = \frac{1}{N} \left\langle D_{\sqrt{h}}(a) \cdot x^{\mathrm{t}}z + a \cdot x^{\mathrm{t}}\bar{x} - \frac{1}{2} a \cdot x^{\mathrm{t}}x \right\rangle.$$

By (3.8), we have

$$\mathbb{E}\langle D_{\sqrt{h}}(a) \cdot x^{\mathrm{t}}z \rangle = \mathbb{E}\langle D_{\sqrt{h}}(a) \cdot (x^{\mathrm{t}}(x - \bar{x})\sqrt{h}) \rangle$$

$$= D_{\sqrt{h}}(a)\sqrt{h} \cdot \mathbb{E}\langle x^{\mathrm{t}}(x - \bar{x}) \rangle.$$

Since $\mathbb{E}\langle x^{\mathrm{t}}(x - \bar{x}) \rangle = \mathbb{E}[\langle x^{\mathrm{t}}x \rangle - \langle x \rangle^{\mathrm{t}}\langle x \rangle]$ is a symmetric matrix, we can use (3.6) to infer that

$$\mathbb{E}\langle D_{\sqrt{h}}(a) \cdot x^{\mathrm{t}}z \rangle = \frac{1}{2} \mathbb{E}\langle a \cdot (x^{\mathrm{t}}(x - \bar{x})) \rangle,$$

and therefore

$$(3.16) \qquad a \cdot \nabla \bar{F}_N = \frac{1}{2N} \mathbb{E}\langle a \cdot x^{\mathrm{t}}\bar{x} \rangle.$$

Since

$$\mathbb{E}\langle x^{\mathrm{t}}\bar{x} \rangle = \mathbb{E}[\langle x \rangle^{\mathrm{t}}\langle x \rangle] \in S_+^K,$$

we obtain that

$$(3.17) \qquad \nabla \bar{F}_N = \frac{1}{2N} \mathbb{E} \langle x^{\mathrm{t}} \bar{x} \rangle,$$

and, by Lemma 2.2, that the mapping $h \mapsto \bar{F}_N(t, h)$ is nondecreasing. Combining (3.14) and (3.17) yields (3.12). □

We are now ready to complete the proof of Proposition 1.2.

PROOF OF PROPOSITION 1.2.   In view of Lemma 3.4, we aim to show that

$$(3.18) \qquad \begin{aligned} &\frac{1}{2N^2} \mathbb{E} \langle |x^{\mathrm{t}} \bar{x} - \mathbb{E} \langle x^{\mathrm{t}} \bar{x} \rangle|^2 \rangle \\ &\qquad \leq C \kappa(h) N^{-\frac{1}{4}} (\Delta \bar{F}_N + C|h^{-1}|)^{\frac{1}{4}} + C \mathbb{E}[|\nabla F_N - \nabla \bar{F}_N|^2]. \end{aligned}$$

We decompose the proof of this fact into five steps.

Step 1. For every $a \in S^K$, we denote

$$H_N'(a, h, x) := D_{\sqrt{h}}(a) \cdot x^{\mathrm{t}} z + a \cdot x^{\mathrm{t}} \bar{x} - \frac{1}{2} a \cdot x^{\mathrm{t}} x.$$

In this step, we show that, for every $a \in S^K$ and $h \in S^K_{++}$,

$$(3.19) \qquad \begin{aligned} &\mathbb{E} \langle (H_N'(a, h, x) - \mathbb{E} \langle H_N'(a, h, x) \rangle)^2 \rangle \\ &\qquad \leq N a \cdot \nabla (a \cdot \nabla \bar{F}_N(t, h)) + N^2 \mathbb{E}[(a \cdot \nabla F_N(t, h) - a \cdot \nabla \bar{F}_N(t, h))^2] \\ &\qquad\quad + C N |a|^2 |h^{-1}|. \end{aligned}$$

In the expression on the right side, the quantity $\nabla (a \cdot \nabla \bar{F}_N(t, h))$ is the gradient of the mapping $h \mapsto a \cdot \nabla \bar{F}_N(t, h)$, evaluated at h. In particular, $\nabla (a \cdot \nabla \bar{F}_N(t, h)) \in S^K$. To show (3.19), we start from the variance decomposition

$$\begin{aligned} &\mathbb{E} \langle (H_N'(a, h, x) - \mathbb{E} \langle H_N'(a, h, x) \rangle)^2 \rangle \\ &\quad = \mathbb{E} \langle (H_N'(a, h, x) - \langle H_N'(a, h, x) \rangle)^2 \rangle + \mathbb{E}[(\langle H_N'(a, h, x) \rangle - \mathbb{E} \langle H_N'(a, h, x) \rangle)^2]. \end{aligned}$$

By (3.15), we have

$$\mathbb{E}[(\langle H_N'(a, h, x) \rangle - \mathbb{E} \langle H_N'(a, h, x) \rangle)^2] = N^2 \mathbb{E}[(a \cdot \nabla F_N(t, h) - a \cdot \nabla \bar{F}_N(t, h))^2].$$

For every $h \in S^K_{++}$ and $a, b \in S^K$, we write

$$D^2_{\sqrt{h}}(a, b) := \lim_{\varepsilon \to 0} \varepsilon^{-1} (D_{\sqrt{h+\varepsilon b}}(a) - D_{\sqrt{h}}(a)),$$

so that, by (3.15),

$$(3.20) \qquad \begin{aligned} &a \cdot \nabla (a \cdot \nabla F_N(t, h)) \\ &\quad = \frac{1}{N} (\langle (H_N'(a, h, x))^2 \rangle - \langle H_N'(a, h, x) \rangle^2) + \frac{1}{N} \langle D^2_{\sqrt{h}}(a, a) \cdot x^{\mathrm{t}} z \rangle. \end{aligned}$$

To see that (3.20) holds, one can start by noticing that for an arbitrary probability measure μ over $\mathbb{R}$, and provided all the integrals are well defined,

$$\begin{aligned} \partial_\lambda^2 \left(\log \int \exp(\lambda s) \, d\mu(s) \right) &= \partial_\lambda \left(\frac{\int s \exp(\lambda s) \, d\mu(s)}{\int \exp(\lambda s) \, d\mu(s)} \right) \\ &= \frac{\int s^2 \exp(\lambda s) \, d\mu(s)}{\int \exp(\lambda s) \, d\mu(s)} - \left(\frac{\int s \exp(\lambda s) \, d\mu(s)}{\int \exp(\lambda s) \, d\mu(s)} \right)^2. \end{aligned}$$

This corresponds to the variance part on the right side of (3.20). Contrary to the "model" calculation above, the quantity H'_N depends itself on h, and this is the reason for the presence of the last term in (3.20). Differentiating the identity (3.4), we find that

$$(3.21) \qquad 2(D_{\sqrt{h}}(a))^2 + \sqrt{h} D^2_{\sqrt{h}}(a, a) + D^2_{\sqrt{h}}(a, a)\sqrt{h} = 0.$$

By Lemma 3.3, we also have that

$$\mathbb{E}\langle D^2_{\sqrt{h}}(a, a) \cdot x^\mathrm{t} z \rangle = D^2_{\sqrt{h}}(a, a)\sqrt{h} \cdot \mathbb{E}\langle x^\mathrm{t}(x - x') \rangle.$$

Combining the two previous displays with the fact that the matrix $\mathbb{E}\langle x^\mathrm{t}(x - x') \rangle$ is symmetric, we obtain that

$$a \cdot \nabla (a \cdot \nabla \bar{F}_N(t, h))$$

$$(3.22) \qquad = \frac{1}{N}\mathbb{E}\langle (H'_N(a, h, x) - \langle H'_N(a, h, x) \rangle)^2 \rangle - \frac{1}{N}\mathbb{E}\langle (D_{\sqrt{h}}(a))^2 \cdot x^\mathrm{t}(x - x') \rangle.$$

By (3.7), this completes the proof of (3.19).

   Step 2. We now aim to show that the variance of $x^\mathrm{t}\bar{x}$ is controlled by a finite sum over a of variances of $H'_N(a, h, x)$ (or equivalently, by the supremum over $|a| \le 1$ of these variances, since $a \mapsto H'_N(a, h, x)$ is linear). In this step, we show that there exists a constant $C < \infty$ such that for every $a \in S^K$,

$$(3.23) \qquad \begin{aligned} &\mathbb{E}\langle (a \cdot x^\mathrm{t}\bar{x} - \mathbb{E}\langle a \cdot x^\mathrm{t}\bar{x} \rangle)^2 \rangle \\ &\le 4\mathbb{E}\langle (H'_N(a, h, x) - \mathbb{E}\langle H'_N(a, h, x) \rangle)^2 \rangle + C|a|^2\kappa(h)\mathsf{Skew}, \end{aligned}$$

where Skew is a quantity measuring the skewness of the matrix $x^\mathrm{t}\bar{x}$:

$$(3.24) \qquad \mathsf{Skew} := N(\mathbb{E}\langle |x^\mathrm{t}\bar{x} - \bar{x}^\mathrm{t}x|^2 \rangle)^{\frac{1}{2}}.$$

Since

$$\mathbb{E}\langle a \cdot x^\mathrm{t}\bar{x} \rangle = 4\mathbb{E}\langle H'_N(a, h, x) \rangle,$$

it suffices to show that

$$(3.25) \qquad \mathbb{E}\langle (a \cdot x^\mathrm{t}\bar{x})^2 \rangle \le 4\mathbb{E}\langle H'_N(a, h, x)^2 \rangle + C|a|^2\kappa(h)\mathsf{Skew}.$$

By Lemmas 3.3 and 3.2 and (3.7), we have

$$\begin{aligned} &\mathbb{E}\langle (D_{\sqrt{h}}(a) \cdot x^\mathrm{t} z)^2 \rangle \\ &\quad = \mathbb{E}\langle (D_{\sqrt{h}}(a)\sqrt{h} \cdot x^\mathrm{t}(x - x'))(D_{\sqrt{h}}(a) \cdot x^\mathrm{t} z) \rangle + \mathbb{E}\langle |D_{\sqrt{h}}(a)x|^2 \rangle \\ &\quad = \mathbb{E}\langle (D_{\sqrt{h}}(a)\sqrt{h} \cdot x^\mathrm{t}(x - x'))(D_{\sqrt{h}}(a)\sqrt{h} \cdot x^\mathrm{t}(x + x' - 2\bar{x})) \rangle + \mathbb{E}\langle |D_{\sqrt{h}}(a)x|^2 \rangle \\ &\quad \ge \frac{1}{4}\mathbb{E}\langle (a \cdot x^\mathrm{t}(x - x'))(a \cdot x^\mathrm{t}(x + x' - 2\bar{x})) \rangle - C|a|^2\kappa(h)\mathsf{Skew}. \end{aligned}$$

The first term on the right side of the previous display can be rewritten as

$$\frac{1}{4}\mathbb{E}\langle (a \cdot x^\mathrm{t}x)^2 \rangle - \frac{1}{2}\mathbb{E}\langle (a \cdot x^\mathrm{t}x)(a \cdot x^\mathrm{t}\bar{x}) \rangle - \frac{1}{4}\mathbb{E}\langle (a \cdot x^\mathrm{t}\bar{x})^2 \rangle + \frac{1}{2}\mathbb{E}\langle (a \cdot x^\mathrm{t}\bar{x})(a \cdot x^\mathrm{t}x') \rangle.$$

Appealing again to Lemmas 3.3 and 3.2, we can also write

$$\begin{aligned} &\mathbb{E}\Big\langle 2(D_{\sqrt{h}}(a) \cdot x^\mathrm{t} z)\Big(a \cdot x^\mathrm{t}\bar{x} - \frac{1}{2}a \cdot x^\mathrm{t}x\Big) \Big\rangle \\ &\quad = \mathbb{E}\Big\langle 2(D_{\sqrt{h}}(a)\sqrt{h} \cdot x^\mathrm{t}(x - x'))\Big(a \cdot x^\mathrm{t}\bar{x} - \frac{1}{2}a \cdot x^\mathrm{t}x\Big) \Big\rangle \\ &\quad \ge \mathbb{E}\Big\langle (a \cdot x^\mathrm{t}(x - x'))\Big(a \cdot x^\mathrm{t}\bar{x} - \frac{1}{2}a \cdot x^\mathrm{t}x\Big) \Big\rangle - C|a|^2\sqrt{\kappa(h)}\mathsf{Skew}, \end{aligned}$$

and the first term on the right side is equal to

$$-\frac{1}{2}\mathbb{E}\langle (a \cdot x^{\mathrm{t}}x)^2 \rangle + \frac{3}{2}\mathbb{E}\langle (a \cdot x^{\mathrm{t}}x)(a \cdot x^{\mathrm{t}}\bar{x}) \rangle - \mathbb{E}\langle (a \cdot x^{\mathrm{t}}\bar{x})(a \cdot x^{\mathrm{t}}x') \rangle.$$

Finally,

$$\mathbb{E}\left\langle \left(a \cdot x^{\mathrm{t}}\bar{x} - \frac{1}{2}a \cdot x^{\mathrm{t}}x \right)^2 \right\rangle = \frac{1}{4}\mathbb{E}\langle (a \cdot x^{\mathrm{t}}x)^2 \rangle - \mathbb{E}\langle (a \cdot x^{\mathrm{t}}x)(a \cdot x^{\mathrm{t}}\bar{x}) \rangle + \mathbb{E}\langle (a \cdot x^{\mathrm{t}}\bar{x})^2 \rangle.$$

Summing the previous displays, we obtain that

$$(3.26) \qquad \mathbb{E}\langle H'_N(a, h, x)^2 \rangle \geq \frac{3}{4}\mathbb{E}\langle (a \cdot x^{\mathrm{t}}\bar{x})^2 \rangle - \frac{1}{2}\mathbb{E}\langle (a \cdot x^{\mathrm{t}}\bar{x})(a \cdot x^{\mathrm{t}}x') \rangle - C|a|^2\kappa(h)\mathsf{Skew}.$$

By the Cauchy–Schwarz inequality,

$$|\mathbb{E}\langle (a \cdot x^{\mathrm{t}}\bar{x})(a \cdot x^{\mathrm{t}}x') \rangle| \leq \mathbb{E}\langle (a \cdot x^{\mathrm{t}}\bar{x})^2 \rangle.$$

Combining the two previous displays yields (3.25), and therefore also (3.23).

Step 3. There remains to control the skew-symmetric part of $x^{\mathrm{t}}\bar{x}$. Following the approach of [1], we decompose the argument into two steps. In this step, we find a convenient expression for the second derivative of $\bar{F}_N$, namely,

$$(3.27) \qquad \begin{aligned} &a \cdot \nabla(a \cdot \nabla \bar{F}_N) \\ &\quad = \frac{1}{2N}\left(\mathbb{E}\langle (a \cdot x^{\mathrm{t}}\bar{x})^2 \rangle - 2\mathbb{E}\langle (a \cdot x^{\mathrm{t}}\bar{x})(a \cdot x^{\mathrm{t}}x') \rangle + \mathbb{E}[\langle a \cdot x^{\mathrm{t}}x' \rangle^2] \right). \end{aligned}$$

Differentiating (3.16) gives

$$(3.28) \qquad a \cdot \nabla(a \cdot \nabla \bar{F}_N) = \frac{1}{2N}\mathbb{E}\langle (a \cdot x^{\mathrm{t}}\bar{x})(H'_N(a, h, x) - H'_N(a, h, x')) \rangle.$$

Moreover, by Lemma 3.3,

$$\mathbb{E}\langle (a \cdot x^{\mathrm{t}}\bar{x})D_{\sqrt{h}}(a) \cdot x^{\mathrm{t}}z \rangle = \mathbb{E}\langle (a \cdot x^{\mathrm{t}}\bar{x})D_{\sqrt{h}}(a)\sqrt{h} \cdot x^{\mathrm{t}}(x - x') \rangle,$$

while

$$\mathbb{E}\langle (a \cdot x^{\mathrm{t}}\bar{x})D_{\sqrt{h}}(a) \cdot (x')^{\mathrm{t}}z \rangle = \mathbb{E}\langle (a \cdot x^{\mathrm{t}}\bar{x})D_{\sqrt{h}}(a)\sqrt{h} \cdot (x')^{\mathrm{t}}(x + x' - 2x'') \rangle.$$

Observe that

$$x^{\mathrm{t}}x - (x^{\mathrm{t}}x' + (x')^{\mathrm{t}}x) - (x')^{\mathrm{t}}x' + (x')^{\mathrm{t}}x'' + (x'')^{\mathrm{t}}x'$$

is a symmetric matrix. Using the symmetry between the replicas x' and x'', (3.6), and then that a is a symmetric matrix, we thus obtain that

$$\begin{aligned} &\mathbb{E}\langle (a \cdot x^{\mathrm{t}}\bar{x})D_{\sqrt{h}}(a) \cdot (x - x')^{\mathrm{t}}z \rangle \\ &\quad = \frac{1}{2}\mathbb{E}\langle (a \cdot x^{\mathrm{t}}\bar{x})a \cdot (x^{\mathrm{t}}x - x^{\mathrm{t}}x' - (x')^{\mathrm{t}}x - (x')^{\mathrm{t}}x' + (x')^{\mathrm{t}}x'' + (x'')^{\mathrm{t}}x') \rangle \\ &\quad = \frac{1}{2}\mathbb{E}\langle (a \cdot x^{\mathrm{t}}\bar{x})a \cdot (x^{\mathrm{t}}x - (x')^{\mathrm{t}}x' - 2x^{\mathrm{t}}x' + 2(x')^{\mathrm{t}}x'') \rangle. \end{aligned}$$

Combining this with (3.28) yields

$$a \cdot \nabla(a \cdot \nabla \bar{F}_N) = \frac{1}{2N}\mathbb{E}\langle (a \cdot x^{\mathrm{t}}\bar{x})a \cdot (-x^{\mathrm{t}}x' + (x')^{\mathrm{t}}x'' + x^{\mathrm{t}}\bar{x} - (x')^{\mathrm{t}}\bar{x}) \rangle,$$

which is (3.27).

Step 4. In this step, we show that

$$\text{(3.29)} \qquad \mathsf{Skew} \le C N^{\frac{7}{4}} \Big(\sup_{\substack{v \in \mathbb{R}^{K \times 1} \\ |v| \le 1}} (v v^{\mathsf{t}}) \cdot \nabla \big((v v^{\mathsf{t}}) \cdot \nabla \bar{F}_N \big) \Big)^{\frac{1}{4}}.$$

For each pair of vectors $v, w \in \mathbb{R}^{K \times 1}$, we look for an upper bound on the quantity

$$\text{(3.30)} \qquad \mathbb{E}\langle (v w^{\mathsf{t}} \cdot (x^{\mathsf{t}} \bar{x} - \langle x^{\mathsf{t}} x' \rangle))^2 \rangle = \mathbb{E}\langle (v w^{\mathsf{t}} \cdot (x^{\mathsf{t}} x' - \langle x^{\mathsf{t}} x' \rangle))^2 \rangle.$$

Since

$$\text{Span}(\{ v w^{\mathsf{t}} : v, w \in \mathbb{R}^{K \times 1} \}) = \mathbb{R}^{K \times K},$$

and since the matrix $\langle x^{\mathsf{t}} x' \rangle = \langle x \rangle^{\mathsf{t}} \langle x \rangle$ is symmetric, knowing that the quantity in (3.30) is small would indeed tell us that the antisymmetric part of $x^{\mathsf{t}} \bar{x}$ concentrates around 0. We will make use of the fact that for every $x, x' \in \mathbb{R}^{N \times K}$ and $v, w \in \mathbb{R}^{K \times 1}$,

$$\text{(3.31)} \qquad (v w^{\mathsf{t}} \cdot x^{\mathsf{t}} x')^2 = (x v v^{\mathsf{t}} x^{\mathsf{t}}) \cdot (x' w w^{\mathsf{t}} (x')^{\mathsf{t}}).$$

This follows from

$$(v w^{\mathsf{t}} \cdot x^{\mathsf{t}} x')^2 = (x v \cdot x' w)^2 = (v^{\mathsf{t}} x^{\mathsf{t}} x' w)^2$$

$$= (v^{\mathsf{t}} x^{\mathsf{t}} x' w) \cdot (v^{\mathsf{t}} x^{\mathsf{t}} x' w) = (x v v^{\mathsf{t}} x^{\mathsf{t}}) \cdot (x' w w^{\mathsf{t}} (x')^{\mathsf{t}}).$$

In particular,

$$\langle (v w^{\mathsf{t}} \cdot x^{\mathsf{t}} x')^2 \rangle = \langle x v v^{\mathsf{t}} x^{\mathsf{t}} \rangle \cdot \langle x w w^{\mathsf{t}} x^{\mathsf{t}} \rangle$$

and

$$(v w^{\mathsf{t}} \cdot \langle x^{\mathsf{t}} x' \rangle)^2 = (\langle x \rangle v v^{\mathsf{t}} \langle x \rangle^{\mathsf{t}}) \cdot (\langle x \rangle w w^{\mathsf{t}} \langle x \rangle^{\mathsf{t}}).$$

We can thus rewrite the quantity in (3.30) as

$$\mathbb{E}[\langle (v w^{\mathsf{t}} \cdot x^{\mathsf{t}} x')^2 \rangle - (v w^{\mathsf{t}} \cdot \langle x^{\mathsf{t}} x' \rangle)^2]$$

$$= \mathbb{E}[\langle x v v^{\mathsf{t}} x^{\mathsf{t}} \rangle \cdot \langle x w w^{\mathsf{t}} x^{\mathsf{t}} \rangle - (\langle x \rangle v v^{\mathsf{t}} \langle x \rangle^{\mathsf{t}}) \cdot (\langle x \rangle w w^{\mathsf{t}} \langle x \rangle^{\mathsf{t}})]$$

$$= \mathbb{E}[(\langle x v v^{\mathsf{t}} x^{\mathsf{t}} \rangle - \langle x \rangle v v^{\mathsf{t}} \langle x \rangle^{\mathsf{t}}) \cdot \langle x w w^{\mathsf{t}} x^{\mathsf{t}} \rangle + \langle x \rangle v v^{\mathsf{t}} \langle x \rangle^{\mathsf{t}} \cdot (\langle x w w^{\mathsf{t}} x^{\mathsf{t}} \rangle - \langle x \rangle w w^{\mathsf{t}} \langle x \rangle^{\mathsf{t}})].$$

By the triangle and the Cauchy–Schwarz inequalities, we thus see that

$$\mathsf{Skew} = N \big(\mathbb{E}\langle |x^{\mathsf{t}} \bar{x} - \bar{x}^{\mathsf{t}} x|^2 \rangle \big)^{\frac{1}{2}}$$

$$\le 2 N \big(\mathbb{E}\langle |x^{\mathsf{t}} \bar{x} - \langle x^{\mathsf{t}} x' \rangle|^2 \rangle \big)^{\frac{1}{2}}$$

$$\le C N \Big(\sup_{\substack{v, w \in \mathbb{R}^{K \times 1} \\ |v|, |w| \le 1}} \mathbb{E}\langle (v w^{\mathsf{t}} \cdot (x^{\mathsf{t}} \bar{x} - \langle x^{\mathsf{t}} x' \rangle))^2 \rangle \Big)^{\frac{1}{2}}$$

$$\le C N^{\frac{3}{2}} \Big(\sup_{\substack{v \in \mathbb{R}^{K \times 1} \\ |v| \le 1}} \mathbb{E}[|\langle x v v^{\mathsf{t}} x^{\mathsf{t}} \rangle - \langle x \rangle v v^{\mathsf{t}} \langle x \rangle^{\mathsf{t}}|^2] \Big)^{\frac{1}{4}}.$$

(Recall that the constant C is allowed to depend on K. Instead of the supremum over v and w, it may be more natural to write first a sum over $v, w \in \mathcal{V}$ with $\mathcal{V}$ a finite set such that $\{ v w^{\mathsf{t}} : v, w \in \mathcal{V} \}$ spans $\mathbb{R}^{K \times K}$.) Appealing again to (3.31), we see that

$$\mathbb{E}[|\langle x v v^{\mathsf{t}} x^{\mathsf{t}} \rangle|^2] = \mathbb{E}\langle (v v^{\mathsf{t}} \cdot x^{\mathsf{t}} x')^2 \rangle,$$

$$\mathbb{E}[|\langle x \rangle v v^{\mathsf{t}} \langle x \rangle^{\mathsf{t}}|^2] = \mathbb{E}[(v v^{\mathsf{t}} \cdot x^{\mathsf{t}} x')^2].$$

A minor variant of (3.31) gives that

$$(vv^t \cdot x^t\bar{x})(vv^t \cdot x^t x') = (xvv^t x^t) \cdot (\bar{x}vv^t(x')^t),$$

and thus

$$\mathbb{E}[\langle xvv^t x^t\rangle \cdot \langle x\rangle vv^t\langle x\rangle^t] = \mathbb{E}\langle(xvv^t x^t) \cdot (\bar{x}vv^t(x')^t)\rangle$$
$$= \mathbb{E}\langle(vv^t \cdot x^t\bar{x})(vv^t \cdot x^t x')\rangle.$$

Combining these identities yields that

$$\mathbb{E}[|\langle xvv^t x^t\rangle - \langle x\rangle vv^t\langle x\rangle^t|^2]$$
$$= \mathbb{E}\langle(vv^t \cdot x^t x')^2\rangle - 2\mathbb{E}\langle(vv^t \cdot x^t\bar{x})(vv^t \cdot x^t x')\rangle + \mathbb{E}[\langle vv^t \cdot x^t x'\rangle^2].$$

Using also (3.27) with $a = vv^t$ completes the proof of (3.29).

Step 5. We combine the results of the previous steps and complete the proof. Notice first that, since $\mathsf{Skew} \leq CN^2$,

$$\mathbb{E}\langle|x^t\bar{x} - \mathbb{E}\langle x^t\bar{x}\rangle|^2\rangle \leq C \sup_{\substack{a\in S^K \\ |a|\leq 1}} \mathbb{E}\langle(a \cdot x^t\bar{x} - \mathbb{E}\langle a \cdot x^t\bar{x}\rangle)^2\rangle + CN^{-2}(\mathsf{Skew})^2$$

(3.32)
$$\leq C \sup_{\substack{a\in S^K \\ |a|\leq 1}} \mathbb{E}\langle(a \cdot x^t\bar{x} - \mathbb{E}\langle a \cdot x^t\bar{x}\rangle)^2\rangle + C\mathsf{Skew}.$$

From (3.22) and (3.7), we see that, for every $a \in S^K$,

(3.33)
$$a \cdot \nabla(a \cdot \nabla\bar{F}_N(t, h)) \geq -C|h^{-1}||a|^2.$$

Writing the Laplacian explicitly as a sum of derivatives in an orthogonal basis that contains a vector colinear to $a \in S^K$, we deduce that

$$a \cdot \nabla(a \cdot \nabla\bar{F}_N(t, h)) \leq |a|^2(\Delta\bar{F}_N(t, h) + C|h^{-1}|).$$

In particular, we can replace $a \cdot \nabla(a \cdot \nabla\bar{F}_N(t, h))$ by $|a|^2\Delta\bar{F}_N(t, h)$ on the right side of (3.19), up to a modification of the constant C, and rewrite (3.29) as

$$\mathsf{Skew} \leq CN^{\frac{7}{4}}(\Delta\bar{F}_N + C|h^{-1}|)^{\frac{1}{4}}.$$

Combining these estimates with (3.23) and (3.32), we obtain that

$$\mathbb{E}\langle|x^t\bar{x} - \mathbb{E}\langle x^t\bar{x}\rangle|^2\rangle \leq C(N\Delta_N\bar{F}_N + N^2\mathbb{E}[|\nabla F_N - \nabla\bar{F}_N|^2] + CN|h^{-1}|)$$
$$+ C\kappa(h)N^{\frac{7}{4}}(\Delta\bar{F}_N + C|h^{-1}|)^{\frac{1}{4}}.$$

Since the left side of this inequality is bounded, we can simplify this into (3.18), thereby completing the proof. □

REMARK 3.5. Contrary to the rank-one case, I do not know whether the mapping $h \mapsto \bar{F}_N(t, h)$ is convex. However, this mapping does satisfy a partial convexity property in the direction of positive semidefinite matrices. (We will not make use of this fact.) To make this point explicit, notice first that if y, y', y'' are three i.i.d. random vectors under $\langle\cdot\rangle$, then

(3.34)
$$\langle(y \cdot y')^2\rangle - 2\langle(y \cdot y')(y \cdot y'')\rangle + \langle y \cdot y'\rangle^2 \geq 0.$$

Indeed, this follows from the fact that the left side of (3.34) can be rewritten as

$$|\langle yy^t\rangle - \langle y\rangle\langle y\rangle^t|^2.$$

Recalling (3.27), we see that, for every $a \in S_+^K$,

$$a \cdot \nabla(a \cdot \nabla \bar{F}_N)$$

$$= \frac{1}{2N} \mathbb{E}[\langle(a \cdot x^{\mathsf{t}}\bar{x})^2\rangle - 2\langle(a \cdot x^{\mathsf{t}}\bar{x})(a \cdot x^{\mathsf{t}}x')\rangle + \langle a \cdot x^{\mathsf{t}}x'\rangle^2]$$

$$= \frac{1}{2N} \mathbb{E}[\langle(x\sqrt{a} \cdot \bar{x}\sqrt{a})^2\rangle - 2\langle(x\sqrt{a} \cdot \bar{x}\sqrt{a})(x\sqrt{a} \cdot x'\sqrt{a})\rangle + \langle x\sqrt{a} \cdot x'\sqrt{a}\rangle^2].$$

Viewing $x\sqrt{a}$ as a vector with NK entries and applying (3.34), we obtain that this quantity is nonnegative. However, since this reasoning requires that we take the square root of a, it only applies to the situation when $a \in S_+^K$. (By symmetry, the case when $-a \in S_+^K$ is of course also covered.)

3.4. *Derivative and concentration estimates.*

We next record simple derivative and concentration estimates. Recall that $W = (W_{ij})_{1 \le i, j \le N}$ is an N-by-N matrix of independent standard Gaussians. We denote

$$\|W\|_{\ell^2 \to \ell^2} := \sup\{|Wx| : x \in \mathbb{R}^{N \times 1}, |x| \le 1\}.$$

LEMMA 3.6 (Derivative estimates). *There exists a constant $C < \infty$ such that the following estimates hold uniformly over $\mathbb{R}_+ \times S_+^K$:*

$$(3.35) \qquad |\partial_t \bar{F}_N| + |\nabla \bar{F}_N| \le C,$$

$$(3.36) \qquad |\partial_t F_N| \le C + \frac{C\|W\|_{\ell^2 \to \ell^2}}{\sqrt{Nt}} \quad and \quad |\nabla F_N| \le C + \frac{C|z||h^{-1}|^{\frac{1}{2}}}{\sqrt{N}}.$$

Moreover, for every $a \in S^K$,

$$(3.37) \qquad a \cdot \nabla(a \cdot \nabla \bar{F}_N) \ge -C|a|^2|h^{-1}| \quad and \quad a \cdot \nabla(a \cdot \nabla F_N) \ge -\frac{C|a|^2|z||h^{-1}|^{\frac{3}{2}}}{\sqrt{N}}.$$

PROOF. The estimates in (3.35) and (3.36) follow from (3.14), (3.17), (3.13), (3.15) and (3.7). (Recall that the constants are allowed to depend on K.) The first part of (3.37) is a consequence of (3.22) and (3.7). To obtain the second part of (3.37), we see from (3.20) that it suffices to establish that

$$(3.38) \qquad \left|D_{\sqrt{h}}^2(a, a)\right| \le C|a|^2|h^{-1}|^{\frac{3}{2}}.$$

Up to a change of basis, we may assume that the matrix h is diagonal, with eigenvalues $0 < \lambda_1 \le \cdots \le \lambda_K$. Denoting by $(\mathsf{d}_{kl})_{1 \le k, l \le K}$ and $(\mathsf{d}'_{kl})_{1 \le k, l \le K}$ the entries of the matrices $D_{\sqrt{h}}(a)$ and $D_{\sqrt{h}}^2(a)$ respectively, we see from (3.21) that for every $k, l \in \{1, \ldots, K\}$,

$$(\sqrt{\lambda_k} + \sqrt{\lambda_l})\mathsf{d}'_{kl} = -2\sum_{m=1}^{K} \mathsf{d}_{km}\mathsf{d}_{ml}.$$

We thus obtain (3.38) using (3.7). $\square$

We now turn to a concentration estimate. We simply state an L^2 estimate with a suboptimal exponent, since this is sufficient for our purposes, but point out that it is classical to improve upon this.

LEMMA 3.7 (Concentration of free energy). *There exists $\alpha > 0$ and, for every compact set $V \subset \mathbb{R}_+ \times S_+^K$, a constant $C < \infty$ such that for every $N \in \mathbb{N}$,*

$$\mathbb{E}\big[\|F_N - \bar{F}_N\|_{L^\infty(V)}^2\big] \leq CN^{-\alpha}.$$

PROOF. The proof is essentially the same as that of [14], Lemma 3.2, so we only briefly sketch the argument. First, using the Efron–Stein and the Gaussian Poincaré inequalities, we verify that for every $M \geq 1$, there exists $C < \infty$ such that for every $t \leq M$ and $|h| \leq M$,

$$\mathbb{E}\big[(F_N - \bar{F}_N)^2(t, h)\big] \leq CN^{-1}.$$

We next use (3.35) and (3.36) to assert that $F_N - \bar{F}_N$ is $\frac{1}{2}$-Hölder continuous, with a random Hölder seminorm that has finite moments of every order. We then write, for every $\varepsilon \in (0, 1]$,

$$\mathbb{E}\Big[\sup_{t \leq M, |h| \leq M} (F_N - \bar{F}_N)^2(t, h)\Big] \leq C\sqrt{\varepsilon} + \mathbb{E}\Big[\sup_{(t,h)\in A_\varepsilon} (F_N - \bar{F}_N)^2(t, h)\Big],$$

where A_ε is an ε-net of the set $\{(t, h) \in \mathbb{R}_+ \times S_+^K : t \leq M \text{ and } |h| \leq M\}$. We can choose A_ε in such a way that $|A_\varepsilon| \leq C\varepsilon^{-1-\frac{K(K+1)}{2}}$, and thus, by a union bound,

$$\mathbb{E}\Big[\sup_{t \leq M, |h| \leq M} (F_N - \bar{F}_N)^2(t, h)\Big] \leq C\sqrt{\varepsilon} + C\varepsilon^{-1-\frac{K(K+1)}{2}} N^{-1}.$$

Optimizing over ε leads to the desired result. □

4. Convergence to the weak solution. We now show how Proposition 1.2, together with the concentration estimate in Lemma 3.7, implies Theorem 1.1. The argument is an adaptation of the proof of uniqueness of weak solutions of (1.8), see Proposition 2.3. One minor simplification comes from the fact that we can assert the local semiconvexity property uniformly in a neighborhood of the region $t = 0$. Indeed, the local semiconvexity of $\bar{F}_N$ given by (3.37) does not degenerate as $t \to 0$, and the corresponding property for the limit solution f is provided by (2.33).

PROOF OF THEOREM 1.1. We decompose the proof into four steps.

Step 1. We set up the argument, find an approximate equation for the difference between $\bar{F}_N$ and the candidate limit, and state elementary bounds, paralleling Steps 1 and 2 of the proof of Proposition 2.3. Denote by f the weak solution of (1.8) with initial condition $f(0, \cdot) = \psi = \bar{F}_1(0, \cdot)$. We set

$$w_N := \bar{F}_N - f \quad \text{and} \quad \mathsf{Err}_N := \partial_t \bar{F}_N - 2|\nabla \bar{F}_N|^2.$$

The following holds almost everywhere in $\mathbb{R}_+ \times S_+^K$:

$$\partial_t w_N = 2|\nabla \bar{F}_N|^2 - 2|\nabla f|^2 + \mathsf{Err}_N$$

$$= 2(\nabla \bar{F}_N + \nabla f) \cdot \nabla w_N + \mathsf{Err}_N.$$

We denote

$$\mathbf{b}_N := 2(\nabla \bar{F}_N + \nabla f).$$

Let $\phi \in C^\infty(\mathbb{R})$ be a nonnegative smooth function satisfying $|\phi'| \leq 1$ and $\phi(0) = 0$, and set $v_N := \phi(w_N)$. We have

$$\partial_t v_N - \mathbf{b}_N \cdot \nabla v_N = \phi'(w_N)\mathsf{Err}_N \quad \text{a.e. in } \mathbb{R}_+ \times S_+^K.$$

We define ζ_ε as in (2.6), f_ε as in (2.7), and

$$\mathbf{b}_{N,\varepsilon} := 2(\nabla \bar{F}_N + \nabla f_\varepsilon),$$

so that

$$\partial_t v_N - \mathbf{b}_{N,\varepsilon} \cdot \nabla v_N = (\mathbf{b}_N - \mathbf{b}_{N,\varepsilon}) \cdot \nabla v_N + \phi'(w_N)\mathsf{Err}_N \quad \text{a.e. in } \mathbb{R}_+ \times S_+^K.$$

By (3.37) with $N = 1$ and (2.33), there exists $C < \infty$ such that for every $\delta \in (0, 1]$ and $\varepsilon \in [0, \frac{\delta}{2}]$,

$$\Delta f_\varepsilon + C\delta^{-1} \geq 0 \quad \text{on } \mathbb{R}_+ \times S_{+\delta}^K.$$

By (3.37), there exists $C < \infty$ such that for every $N \in \mathbb{N}$ and $\delta \in (0, 1]$,

$$\Delta \bar{F}_N + C\delta^{-1} \geq 0 \quad \text{on } \mathbb{R}_+ \times S_{+\delta}^K.$$

We thus obtain that for every $N \in \mathbb{N}$, $\delta \in (0, 1]$ and $\varepsilon \in [0, \frac{\delta}{2}]$,

$$(4.1) \qquad \nabla \cdot \mathbf{b}_{N,\varepsilon} + C\delta^{-1} \geq 0 \quad \text{on } \mathbb{R}_+ \times S_{+\delta}^K.$$

Step 2. We essentially reproduce the arguments in Step 3 of the proof of Proposition 2.3, temporarily leaving aside the new error term Err_N. We denote

$$R := 1 + 2\|\nabla f\|_{L^\infty(\mathbb{R}_+ \times S_+^K)} + 2 \sup_{N \in \mathbb{N}} (\|\nabla \bar{F}_N\|_{L^\infty(\mathbb{R}_+ \times S_+^K)}),$$

which is finite by (3.35). We fix $T \geq 1$ and define, for every $t \in [0, \frac{T}{2}]$, the sets $B_\delta(t)$, $\partial_+ B_\delta(t)$, and $\partial_0 B_\delta(t)$ displayed in (2.15)–(2.17), as well as

$$J_{\delta,N}(t) := \int_{B_\delta(t)} v_N(t, \cdot).$$

The function $J_{\delta,N}$ is Lipschitz, and for almost every $t \in [0, \frac{T}{2}]$,

$$\partial_t J_{\delta,N}(t)$$
$$= \int_{B_\delta(t)} \partial_t v_N(t, \cdot) - R \int_{\partial_+ B_\delta(t)} v_N(t, \cdot)$$
$$= \int_{B_\delta(t)} (\mathbf{b}_{N,\varepsilon} \cdot \nabla v_N + (\mathbf{b}_N - \mathbf{b}_{N,\varepsilon}) \cdot \nabla v_N + \phi'(w_N)\mathsf{Err}_N)(t, \cdot) - R \int_{\partial_+ B_\delta(t)} v_N(t, \cdot).$$

By (2.9) and the dominated convergence theorem, we have

$$\lim_{\varepsilon \to 0} \int_{B_\delta(t)} ((\mathbf{b}_N - \mathbf{b}_{N,\varepsilon}) \cdot \nabla v_N)(t, \cdot) = 0.$$

Integrating by parts, we see that

$$\int_{B_\delta(t)} (\mathbf{b}_{N,\varepsilon} \cdot \nabla v_N)(t, \cdot) = -\int_{B_\delta(t)} (v_N \nabla \cdot \mathbf{b}_{N,\varepsilon})(t, \cdot) + \int_{\partial B_\delta(t)} (v_N \mathbf{b}_N \cdot \mathbf{n})(t, \cdot),$$

where $\mathbf{n}$ is the unit outer normal to $B_\delta(t)$. By (4.1), we have

$$-\int_{B_\delta(t)} (v_N \nabla \cdot \mathbf{b}_{N,\varepsilon})(t, \cdot) \leq C\delta^{-1} J_{\delta,N}(t).$$

Recalling (2.15)–(2.17), we decompose the boundary integral into

$$\int_{\partial B_\delta(t)} (v_N \mathbf{b}_N \cdot \mathbf{n})(t, \cdot) = \int_{\partial_+ B_\delta(t)} (v_N \mathbf{b}_N \cdot \mathbf{n})(t, \cdot) + \int_{\partial_0 B_\delta(t)} (v_N \mathbf{b}_N \cdot \mathbf{n})(t, \cdot).$$

By the definition of R, we have

$$\int_{\partial_+ B_\delta(t)} (v_N \mathbf{b}_N \cdot \mathbf{n})(t, \cdot) \leq R \int_{\partial_+ B_\delta(t)} v_N(t, \cdot).$$

Since both $\bar{F}_N$ and f are nondecreasing in h, we infer from (2.4) that

$$\int_{\partial_0 B_\delta(t)} (v_N \mathbf{b}_N \cdot \mathbf{n})(t, \cdot) \leq 0.$$

Step 3. There remains to estimate the contribution of the error term Err_N. By Proposition 1.2, and since $|\phi'| \leq 1$, we have

$$\begin{aligned}
(4.2) \qquad \int_{B_\delta(t)} (\phi'(w_N)\mathsf{Err}_N)(t, \cdot) &\leq C\delta^{-1} N^{-\frac{1}{4}} \int_{B_\delta(t)} (\Delta \bar{F}_N + C\delta^{-1})^{\frac{1}{4}}(t, \cdot) \\
&\quad + C \int_{B_\delta(t)} \mathbb{E}[|\nabla F_N - \nabla \bar{F}_N|^2](t, \cdot),
\end{aligned}$$

where we allow the multiplicative constant to depend also on R and T. We estimate each of these two integrals in turn. By Jensen's inequality,

$$\int_{B_\delta(t)} (\Delta \bar{F}_N + C\delta^{-1})^{\frac{1}{4}}(t, \cdot) \leq \left(C\delta^{-1} + \int_{B_\delta(t)} \Delta \bar{F}_N(t, \cdot) \right)^{\frac{1}{4}},$$

and moreover, by integration by parts and (3.35),

$$\int_{B_\delta(t)} \Delta \bar{F}_N(t, \cdot) \leq C.$$

Turning to the second integral on the right side of (4.2), we introduce the notation $V := \{(t, h) : t \leq T, |h| \leq RT\}$, and integrate by parts and use (3.35) again to get

$$\begin{aligned}
\int_{B_\delta(t)} &|\nabla F_N - \nabla \bar{F}_N|^2(t, \cdot) \\
&= \int_{\partial B_\delta(t)} ((F_N - \bar{F}_N)\nabla(F_N - \bar{F}_N) \cdot \mathbf{n})(t, \cdot) - \int_{B_\delta(t)} ((F_N - \bar{F}_N)\Delta(F_N - \bar{F}_N))(t, \cdot) \\
&\leq \|F_N - \bar{F}_N\|_{L^\infty(V)} \left(C + \int_{B_\delta(t)} |\Delta(F_N - \bar{F}_N)|(t, \cdot) \right).
\end{aligned}$$

We can then write

$$\begin{aligned}
\int_{B_\delta(t)} &|\Delta(F_N - \bar{F}_N)|(t, \cdot) \\
&\leq C\delta^{-\frac{3}{2}}\left(1 + \frac{|z|}{\sqrt{N}} \right) + \int_{B_\delta(t)} \left| \Delta(F_N - \bar{F}_N) + C\delta^{-\frac{3}{2}}\left(1 + \frac{|z|}{\sqrt{N}} \right) \right|(t, \cdot).
\end{aligned}$$

We next observe that for $C < \infty$ sufficiently large, the quantity between absolute values above is nonnegative, by (3.37). Integrating by parts and using (3.36), we obtain that

$$\int_{B_\delta(t)} |\Delta(F_N - \bar{F}_N)|(t, \cdot) \leq C\delta^{-\frac{3}{2}}\left(1 + \frac{|z|}{\sqrt{N}} \right).$$

Summarizing, and using the Cauchy–Schwarz inequality, we conclude that

$$\int_{B_\delta(t)} \mathbb{E}[|\nabla F_N - \nabla \bar{F}_N|^2](t, \cdot) \leq C\delta^{-\frac{3}{2}} \mathbb{E}[\|F_N - \bar{F}_N\|_{L^\infty(V)}^2]^{\frac{1}{2}},$$

and thus, by Lemma 3.7, that there exists an exponent $\alpha \in (0, \frac{1}{4}]$ such that

$$\int_{B_\delta(t)} \mathbb{E}[|\nabla F_N - \nabla \bar{F}_N|^2](t, \cdot) \leq C\delta^{-\frac{3}{2}} N^{-\alpha}.$$

Step 4. We conclude the proof. Combining the results of the two previous steps, we obtain that almost everywhere in $[0, \frac{T}{2}]$, we have

$$\partial_t J_{N,\delta} \le C\delta^{-1} J_{\delta,N} + C\delta^{-\frac{3}{2}} N^{-\alpha},$$

that is,

$$\partial_t (\exp(-C\delta^{-1}t) J_{\delta,N}) \le C \exp(-C\delta^{-1}t) \delta^{-\frac{3}{2}} N^{-\alpha}.$$

Since $\phi(0) = 0$ and $w_N(0, \cdot) = 0$, this implies that for every $t \in [0, \frac{T}{2}]$,

$$\int_{B_\delta(t)} \phi(\bar{F}_N - f)(t, \cdot) \le C \exp(C\delta^{-1}) \delta^{-\frac{3}{2}} N^{-\alpha}.$$

(Recall that we allow the constant C to depend on T.) We may as well absorb the term $\delta^{-\frac{3}{2}}$ into the exponential. Since this estimate is valid uniformly over nonnegative $\phi \in C^\infty(\mathbb{R})$ satisfying $\phi(0) = 0$ and $|\phi'| \le 1$, we deduce that for every $t \in [0, \frac{T}{2}]$,

$$\int_{B_\delta(t)} |\bar{F}_N - f|(t, \cdot) \le C \exp(C\delta^{-1}) N^{-\alpha}.$$

Since the functions $\bar{F}_N$ and f are locally bounded, uniformly over N, and the measure of the set $B_0(t) \setminus B_\delta(t)$ is bounded by $C\delta$, this implies that for every $t \in [0, \frac{T}{2}]$,

$$\int_{B_0(t)} |\bar{F}_N - f|(t, \cdot) \le C\delta + C \exp(C\delta^{-1}) N^{-\alpha}.$$

We select $\delta := C \log^{-1} N$, for a sufficiently large constant C, so that for every $t \in [0, \frac{T}{2}]$,

$$\int_{B_0(t)} |\bar{F}_N - f|(t, \cdot) \le \frac{C}{\log N}.$$

This completes the proof of Theorem 1.1. $\square$

APPENDIX: COMPUTATION OF THE CONDITIONAL LAW

We denote

$$(\text{A.1}) \qquad \mathcal{Y} = (Y, Y') = \left(\sqrt{\frac{t}{N}} \bar{x}\bar{x}^{\mathrm{t}} + W, \bar{x}\sqrt{h} + z \right).$$

In this appendix, we verify that the conditional law of $\bar{x}$ given $\mathcal{Y}$ is given by

$$(\text{A.2}) \qquad \frac{e^{H_N(t,h,x)} \, dP_N(x)}{\int_{\mathbb{R}^{N \times K}} e^{H_N(t,h,x')} \, dP_N(x')}.$$

(See also (3.2) for an equivalent statement.) For every bounded measurable functions f and g, we can write $\mathbb{E}[f(\bar{x})g(\mathcal{Y})]$, up to a normalization constant that depends neither on f nor on g, as

$$\int f(x)g\left(\sqrt{\frac{t}{N}} xx^{\mathrm{t}} + W, x\sqrt{h} + z \right) \exp\left(-\frac{|W|^2}{2} - \frac{|z|^2}{2} \right) dW \, dz \, dP_N(x),$$

with the shorthand notation $dW := \prod_{i,j} dW_{ij}$ and $dz := \prod_{i,k} dz_{ik}$. A change of variables allows to rewrite the expression above as

$$\int f(x)g(\mathcal{Y}) \exp\left(-\frac{1}{2} \left| Y - \sqrt{\frac{t}{N}} xx^{\mathrm{t}} \right|^2 - \frac{1}{2} |Y' - x\sqrt{h}|^2 \right) d\mathcal{Y} \, dP_N(x).$$

Denoting the exponential factor above by $\mathcal{E}(x, \mathcal{Y})$, we thus obtain that the law of $\mathcal{Y}$ is the law with density given, up to a normalization constant, by

$$\bar{\mathcal{E}}(\mathcal{Y}) := \int \mathcal{E}(x, \mathcal{Y}) \, dP_N(x),$$

and that, denoting by c the normalization constant,

(A.3) $$\mathbb{E}[f(\bar{x})g(\mathcal{Y})] = c \int f(x) \frac{\mathcal{E}(x, \mathcal{Y})}{\bar{\mathcal{E}}(\mathcal{Y})} \, dP_N(x) g(\mathcal{Y}) \bar{\mathcal{E}}(\mathcal{Y}) \, d\mathcal{Y}.$$

The conditional law of $\bar{x}$ given $\mathcal{Y}$ is thus the probability measure given by

$$\frac{\mathcal{E}(x, \mathcal{Y})}{\bar{\mathcal{E}}(\mathcal{Y})} \, dP_N(x),$$

and this quantity can indeed be rewritten in the form of (A.1).

REFERENCES

[1] BARBIER, J. (2019). Overlap matrix concentration in optimal Bayesian inference. Preprint. Available at arXiv:1904.02808.

[2] BARBIER, J., DIA, M., MACRIS, N., KRZAKALA, F., LESIEUR, T. and ZDEBOROVÁ, L. (2016). Mutual information for symmetric rank-one matrix estimation: A proof of the replica formula. In *Advances in Neural Information Processing Systems* 29 424–432.

[3] BARBIER, J. and MACRIS, N. (2019). The adaptive interpolation method: A simple scheme to prove replica formulas in Bayesian inference. *Probab. Theory Related Fields* **174** 1133–1185. MR3980313 https://doi.org/10.1007/s00440-018-0879-0

[4] BARBIER, J., MACRIS, N. and MIOLANE, L. (2017). The layered structure of tensor estimation and its mutual information. In 55*th Annual Allerton Conference on Communication, Control, and Computing* 1056–1063. IEEE, Piscataway, NJ.

[5] BENTON, S. H. JR. (1977). *The Hamilton–Jacobi Equation*: *A Global Approach. Mathematics in Science and Engineering* **131** Academic Press, New York. MR0442431

[6] DESHPANDE, Y. and MONTANARI, A. (2014). Information-theoretically optimal sparse PCA. In *IEEE International Symposium on Information Theory* 2197–2201.

[7] DOUGLIS, A. (1965). Solutions in the large for multi-dimensional, non-linear partial differential equations of first order. *Ann. Inst. Fourier* (*Grenoble*) **15** 1–35. MR0199542

[8] EL ALAOUI, A. and KRZAKALA, F. (2018). Estimation in the spiked Wigner model: A short proof of the replica formula. Preprint. Available at arXiv:1801.01593.

[9] EVANS, L. C. (2010). *Partial Differential Equations*, 2nd ed. *Graduate Studies in Mathematics* **19**. Amer. Math. Soc., Providence, RI. MR2597943 https://doi.org/10.1090/gsm/019

[10] KRUŽKOV, S. N. (1966). Generalized solutions of nonlinear equations of the first order with several variables. I. *Mat. Sb.* (*N.S.*) **70 (112)** 394–415. MR0199543

[11] KRUŽKOV, S. N. (1967). Generalized solutions of nonlinear equations of the first order with several independent variables. II. *Mat. Sb.* (*N.S.*) **72 (114)** 108–134. MR0204847

[12] LELARGE, M. and MIOLANE, L. (2019). Fundamental limits of symmetric low-rank matrix estimation. *Probab. Theory Related Fields* **173** 859–929. MR3936148 https://doi.org/10.1007/s00440-018-0845-x

[13] LESIEUR, T., KRZAKALA, F. and ZDEBOROVÁ, L. (2015). Phase transitions in sparse PCA. In *IEEE International Symposium on Information Theory* 1635–1639.

[14] MOURRAT, J.-C. (2018). Hamilton–Jacobi equations for mean-field disordered systems. Preprint. Available at arXiv:1811.01432.

[15] PANCHENKO, D. (2018). Free energy in the Potts spin glass. *Ann. Probab.* **46** 829–864. MR3773375 https://doi.org/10.1214/17-AOP1193

[16] PANCHENKO, D. (2018). Free energy in the mixed *p*-spin models with vector spins. *Ann. Probab.* **46** 865–896. MR3773376 https://doi.org/10.1214/17-AOP1194