Statistical inference of finite-rank tensors

Hongbin Chen, Jean-Christophe Mourrat, Jiaming Xia

STATISTICAL INFERENCE OF FINITE-RANK TENSORS

HONG-BIN CHEN, JEAN-CHRISTOPHE MOURRAT, AND JIAMING XIA

Abstract. We consider a general statistical inference model of finite-rank tensor products. For any interaction structure and any order of tensor products, we identify the limit free energy of the model in terms of a variational formula. Our approach consists of showing first that the limit free energy must be the viscosity solution to a certain Hamilton-Jacobi equation.

1. Introduction

1.1. **Setting.** Let $K, L, \mathsf{p} \in \mathbb{N}$ and $A \in \mathbb{R}^{K^{\mathsf{p}} \times L}$, which will be kept fixed throughout the paper. For every $N \in \mathbb{N}$, $t \geqslant 0$ and a random matrix $X \in \mathbb{R}^{N \times K}$, we consider the inference task of recovering X from the observation of

$$Y := \sqrt{\frac{2t}{N^{\mathsf{p}-1}}} X^{\otimes \mathsf{p}} A + W \ \in \mathbb{R}^{N^{\mathsf{p}} \times L}, \tag{1.1}$$

where $\otimes$ denotes the tensor product of matrices, and $W \in \mathbb{R}^{N^{\mathsf{p}} \times L}$ consists of independent standard Gaussian entries (we view $X^{\otimes \mathsf{p}}$ as an N^{p}-by-K^{p} matrix). Throughout, the dot product between two vectors or matrices of the same size is the entry-wise inner product. The associated norm is denoted by $|\cdot|$. For convenience of analysis, we assume that the random matrix X almost surely satisfies

$$|X| \leqslant \sqrt{NK}. \tag{1.2}$$

We denote the law of X by P_N^X. Using Bayes' rule, the law of X conditioned on observing Y is the measure proportional to $e^{H_N^\circ(t,x)} \, \mathrm{d}P_N^X(x)$, where the Hamiltonian H_N° is

$$H_N^\circ(t,x) := \sqrt{\frac{2t}{N^{\mathsf{p}-1}}} (x^{\otimes \mathsf{p}} A) \cdot Y - \frac{t}{N^{\mathsf{p}-1}} |x^{\otimes \mathsf{p}} A|^2.$$

The associated free energy is given by

$$F_N^\circ(t) := \frac{1}{N} \log \int_{\mathbb{R}^{N \times K}} e^{H_N^\circ(t,x)} \, \mathrm{d}P_N^X(x).$$

The mutual information $I(X,Y)$ between X and Y is an important information-theoretical quantity, which is equal to $\mathbb{E}F_N^\circ(t)$ up to a simple additive term. Computing the limit of the mutual information as $N \to \infty$ allows one to determine the critical value of t below which the inference task is theoretically impossible. Therefore, the limit of $\mathbb{E}F_N^\circ(t)$ is the central object of investigation in many inference models. For more details, we refer to the discussion in [1].

In order to analyze this model, we start by enriching the system by adding an additional observation $\overline{Y} = X\sqrt{2h} + Z$ for $h \in \mathbf{S}_+^K$, where $\mathbf{S}_+^K$ is the set of $K \times K$ symmetric positive semi-definite matrices, and Z consists of i.i.d. standard Gaussian entries. Then, the law of X conditioned on observing Y and $\overline{Y}$ is a Gibbs measure with Hamiltonian

$$H_N(t,h,x) := H_N^\circ(t,x) + \sqrt{2h} \cdot (x^\top \overline{Y}) - h \cdot (x^\top x).$$

Date: April 14, 2021.

The corresponding free energy is

$$F_N(t,h) := \frac{1}{N}\log\int_{\mathbb{R}^{N\times K}} e^{H_N(t,h,x)}\,\mathrm{d}P_N^X(x). \tag{1.3}$$

We also set $\overline{F}_N = \mathbb{E}F_N$. Note that the initial free energy satisfies $F_N^\circ(t) = F_N(t,0)$. We let $\mathsf{H}:\mathbf{S}_+^K\to\mathbb{R}$ be the mapping such that, for every $q\in\mathbf{S}_+^K$,

$$\mathsf{H}(q) := (AA^\intercal)\cdot q^{\otimes \mathsf{p}}. \tag{1.4}$$

Our main result is the identification of the limit free energy, for any given choice of interaction matrix A and $\mathsf{p}\in\mathbb{N}$.

Theorem 1.1. *In addition to (1.2), suppose that*

- $\big(\overline{F}_N(0,\cdot)\big)_{N\in\mathbb{N}}$ *converges pointwise to some C^1 function $\psi:\mathbf{S}_+^K\to\mathbb{R}$;*
- $\lim_{N\to\infty}\mathbb{E}\|F_N-\overline{F}_N\|_{L^\infty(D)}^2 = 0$ *for every compact $D\subseteq[0,\infty)\times\mathbf{S}_+^K$.*

Then, for every $(t,h)\in[0,\infty)\times\mathbf{S}_+^K$, we have

$$\lim_{N\to\infty}\overline{F}_N(t,h) = \sup_{h''\in\mathbf{S}_+^K}\inf_{h'\in\mathbf{S}_+^K}\big\{h''\cdot(h-h')+\psi(h')+t\mathsf{H}(h'')\big\}. \tag{1.5}$$

Remark 1.2. The above convergence can be improved into convergence in the local uniform topology by using that $\overline{F}_N$ is Lipschitz uniformly over N (see Lemma 2.1).

We briefly comment on the hypotheses of the theorem. One can see that $F_N(0,\cdot)$ is the free energy associated with a decoupled system where the only observation $\overline{Y}$ is linear in X. Therefore, in many cases, the limit of $\overline{F}_N(0,\cdot)$ can be computed straightforwardly. In particular, if P_N^X is the N-fold tensor product of a fixed probability measure on $\mathbb{R}^K$, then $\overline{F}_N(0,\cdot)$ in fact does not depend on N, and is C^1. The next assumption can be rephrased as local uniform concentration of F_N. Again, this condition is straightforward to verify in many models, with standard tools available: see for instance [8, Lemma C.1] for the case when the rows of X are i.i.d. and bounded.

Among our assumptions, perhaps the only surprising one is the requirement that ψ be of class C^1. For certain choices of the nonlinearity H, such as when H is convex, this assumption is not necessary (see for instance [8]). However, when considering arbitrary choices of A and p as we do here, this assumption may be required. In a simpler setting, we illustrate the usefulness of this assumption in Remark A.3.

1.2. **Related works.** Many inference models can be viewed as special cases of (1.1). Indeed, one could argue that essentially any "fully-connected" inference problem will have the form of (1.1) for some suitable choice of A and p. Among them, the models where the limit free energy has been studied include the spiked Wigner model [1, 12, 2, 18, 19], the spiked Wishart model [17, 4, 11, 15, 7], the stochastic block model (or community detection problem) [12, 16, 25], the inference of second order matrix tensor products [24], and the inference of higher order vector tensor products [13, 2, 18]. The model closest to (1.1) is the inference of finite-rank even-order tensor products studied in [14]. The case of tensors of odd order was left open there, see [14, Section 7]. In Section 5.2, we apply our main result to this model, for tensor products of arbitrary order ($\mathsf{p}\in\mathbb{N}$). For a more detailed discussion on these models, we refer to the introduction in [8].

Many of the results mentioned above were obtained by the powerful method of adaptive interpolation introduced in [2, 3] and refined in subsequent works. In [24], a novel extension using interpolation paths parameterized by order-preserving positive semi-definite matrices was employed to completely describe the limit in the general second

order tensor products model. The order-preserving property ([24, Proposition 4]) has a similar counterpart that plays a crucial role in this work (Lemma 2.2 and Proposition 4.7).

The approach taken up in the present paper is based instead on identifying the limit free energy as the viscosity solution to a certain Hamilton-Jacobi equation. This alternative approach was introduced in [18, 19], and can also inform the analysis of spin glass models [20, 23, 21, 22]; related considerations also appeared in the physics literature [9, 10, 6, 5].

The setting of the present paper is identical to that of [8], in which partial results were obtained. There, for general interaction matrix A and order p, only an upper bound on the limit free energy could be proved; a complete identification of this limit could only be obtained for particular choices of A and p. Here, we close this gap and cover all cases in a unified approach.

Compared with [8], the main novelty of the present paper is that we will rely on a different method for the identification of the viscosity solution. This method relies crucially on the fact that the functions under consideration are *convex*. We explain this new uniqueness criterion in the simpler context of Hamilton-Jacobi equations on $[0,\infty)\times\mathbb{R}^d$ in the appendix. The gist of our work is then to extend this criterion to Hamilton-Jacobi equations posed on $[0,\infty)\times\mathbf{S}^K_+$, and then to verify that any possible limit of the free energy does satisfy this criterion.

The rest of the paper is organized as follows. In Section 2, we present basic properties of $\overline{F}_N$. In particular, we record that $\overline{F}_N$ is convex, nondecreasing, and has nondecreasing gradients. In Section 3, we recall basic facts of convex analysis and prove some useful results in preparation for the study of the Hamilton-Jacobi equation. Using these, we prove a convenient criterion for identifying viscosity solutions in Section 4. Lastly, Section 5 contains the proof of Theorem 1.1 and an application to the model (5.7).

Acknowledgements. We would like to warmly thank Stefano Bianchini for providing us with the idea for the proof of Proposition A.2. JCM was partially supported by the NSF grant DMS-1954357.

2. Properties of the free energy

In this section, we study basic properties of $\overline{F}_N$. We start by introducing notation.

For any measurable $g:\mathbb{R}^{N\times K}\to\mathbb{R}$, we denote by $\langle g(x)\rangle$ the expectation of g with respect to the Gibbs measure proportional to $e^{H_N(t,h,x)}\,\mathrm{d}P_N^X(x)$, which can also be written as $\langle g(x)\rangle = \mathbb{E}[g(X)|Y,\overline{Y}]$ for Y and $\overline{Y}$ introduced in the previous section. Note that the dependence of $\langle\cdot\rangle$ on t,h is suppressed from the notation when there is no confusion. Within the bracket $\langle\cdot\rangle$, we denote by x',x'',x''' independent copies of x, which are called replicas of x. The transpose operator on matrices is denoted by superscript $\intercal$.

In addition to $\mathbf{S}^K_+$, we denote by $\mathbf{S}^K$ and $\mathbf{S}^K_{++}$, the set of $K\times K$ symmetric matrices, and symmetric positive definite matrices, respectively. We view $\mathbf{S}^K$ as an ambient linear space for $\mathbf{S}^K_+$ and $\mathbf{S}^K_{++}$. By choosing an orthonormal basis with respect to the entry-wise dot product, we can identify $\mathbf{S}^K$ with $\mathbb{R}^{K(K+1)/2}$ isometrically. Therefore, differentiation makes sense on $\mathbf{S}^K$ as the usual one on Euclidean spaces. Naturally, we also identify the dual space of $\mathbf{S}^K$ with itself. For a function $g:[0,\infty)\times\mathbf{S}^K_+\to\mathbb{R}$ which is differentiable at (t',h'), we denote by $\partial_t g(t',h')\in\mathbb{R}$ its derivative with respect to the first variable, and by $\nabla g(t',h')\in\mathbf{S}^K$ the gradient with respect to the second variable.

Using the expression (1.3), we can compute that

$$\partial_t\overline{F}_N = \frac{1}{N^{\mathsf{p}}}\mathbb{E}\left\langle x^{\otimes\mathsf{p}}A\cdot x'^{\otimes\mathsf{p}}A\right\rangle,$$

and

$$\nabla \overline{F}_N = \frac{1}{N}\mathbb{E}\left\langle x^\top x' \right\rangle.$$

This computation involves the Nishimori identity and Gaussian integration by parts. For details, we refer to [8, (3.5)-(3.6)]. Recalling the definition of H in (1.4), we obtain that $\overline{F}_N$ satisfies

$$\partial_t \overline{F}_N - \mathsf{H}\big(\nabla \overline{F}_N\big) = \frac{1}{N^{\mathsf{p}}}\bigg(\mathbb{E}\big\langle \mathsf{H}(x^\top x')\big\rangle - \mathsf{H}\big(\mathbb{E}\langle x^\top x'\rangle\big)\bigg),$$

and the right-hand side is expected to be small when N is large. Hence, $\overline{F}_N$ can be viewed to approximately satisfy the Hamilton-Jacobi equation

$$\partial_t f - \mathsf{H}(\nabla f) = 0 \quad \text{in } [0,\infty)\times \mathbf{S}^K_+. \tag{2.1}$$

This is the key insight for the Hamilton-Jacobi equation approach. Later, we will show that indeed $\overline{F}_N$ converges to the unique solution to (2.1); and then that this solution admits the variational representation appearing on the right side of (1.5).

In the remaining two subsections, we collect useful properties of derivatives of $\overline{F}_N$ and prove that $\overline{F}_N$ is convex.

2.1. **Derivatives of free energy.** From [8, Lemma 3.2], we recall the following result.

Lemma 2.1. *For each $N \in \mathbb{N}$, the function $\overline{F}_N$ is C^1 and the following holds:*

$$\sup_{N\in\mathbb{N},\,(t,h)\in[0,\infty)\times\mathbf{S}^K_+} \big|(\partial_t,\nabla)\overline{F}_N\big|(t,h) < \infty;$$

$$(\partial_t,\nabla)\overline{F}_N(t,h) \in [0,\infty)\times\mathbf{S}^K_+, \quad \forall N\in\mathbb{N},\ (t,h)\in[0,\infty)\times\mathbf{S}^K_+.$$

The first display ensures that $\overline{F}_N$ is Lipschitz uniformly in N. The second display indicates that $(\partial_t,\nabla)\overline{F}_N$ is "nonnegative" in the sense of the following partial orders. On $\mathbf{S}^K$ and on $\mathbb{R}\times\mathbf{S}^K$, we declare

$$h_1 \leqslant h_2 \quad\Longleftrightarrow\quad h_2 - h_1 \in \mathbf{S}^K_+; \tag{2.2}$$

$$(t_1,h_1) \leqslant (t_2,h_2) \quad\Longleftrightarrow\quad (t_2,h_2)-(t_1,h_1)\in[0,\infty)\times\mathbf{S}^K_+. \tag{2.3}$$

The next result shows that $(\partial_t,\nabla)\overline{F}_N$ is "nondecreasing".

Lemma 2.2. *For each $N\in\mathbb{N}$, for every $(t_1,h_1)\leqslant(t_2,h_2)$, it holds that*

$$(\partial_t,\nabla)\overline{F}_N(t_1,h_1) \leqslant (\partial_t,\nabla)\overline{F}_N(t_2,h_2).$$

Proof. For $k=1,2$, we set

$$Y_k := \left(\sqrt{\frac{2t_k}{N^{\mathsf{p}-1}}}X^{\otimes\mathsf{p}}A + W_k\,,\ X\sqrt{2h_k} + Z_k\right)$$

where W_k and Z_k consist of i.i.d. standard Gaussian random variables. To display the dependence of $\langle\cdot\rangle$ on (t_k,h_k), we write

$$\langle g(x)\rangle_{t_k,h_k} = \mathbb{E}[g(X)\,|\,Y_k]$$

for any measurable function g. Let $\widehat{X} := \operatorname{diag}(X^{\otimes\mathsf{p}}A,\,X)$. For any matrix y, we write $\mathbf{c}(y) := y^\top y$. Note that $\mathbf{c}(X^{\otimes\mathsf{p}}A)\in\mathbb{R}^{L\times L}$ and $\mathbf{c}(X)\in\mathbb{R}^{K\times K}$. Let $D=\operatorname{diag}(I_L,0_K)$ where I_m is the $m\times m$ identity matrix and 0_m is the $m\times m$ zero matrix. Let $P\in\mathbb{R}^{L+K,K}$ be the matrix projecting $\mathbb{R}^{L+K}$ onto the last K coordinates. Then, we have

$$(\partial_t,\nabla)\overline{F}(t_k,h_k) = \left(\frac{1}{N^{\mathsf{p}}}D\cdot\mathbb{E}\,\mathbf{c}\big(\mathbb{E}[\widehat{X}\,|\,Y_k]\big),\ \frac{1}{N}P^\top\Big(\mathbb{E}\,\mathbf{c}\big(\mathbb{E}[\widehat{X}\,|\,Y_k]\big)\Big)P\right),$$

Hence, it suffices to show that

$$\mathbb{E}\,\mathbf{c}\big(\mathbb{E}[\widehat{X}\,|\,Y_1]\big) \leqslant \mathbb{E}\,\mathbf{c}\big(\mathbb{E}[\widehat{X}\,|\,Y_2]\big). \tag{2.4}$$

To compare these quantities, we introduce

$$Y' := \left(\sqrt{\frac{2t_2-2t_1}{N^{p-1}}}X^{\otimes p}A + W',\ X\sqrt{2h_2-2h_1}+Z'\right),$$

where W' and Z' have i.i.d. standard Gaussian entries. We claim that, for every nonnegative measurable function f, we have

$$\mathbb{E}[f(X)\,|\,Y_2] \overset{\mathrm{d}}{=} \mathbb{E}[f(X)\,|\,Y_1,Y'], \tag{2.5}$$

where the equality holds in the sense of distributions. Temporarily assuming this, and using that $\mathbb{E}[\widehat{X}\,|\,Y_1] = \mathbb{E}\big[\mathbb{E}[\widehat{X}\,|\,Y_1,Y']\,\big|\,Y_1\big]$, we can verify, analogously to a bias-variance decomposition, that

$$\mathbb{E}\,\mathbf{c}\big(\mathbb{E}[\widehat{X}\,|\,Y_1,Y']\big) = \mathbb{E}\,\mathbf{c}\big(\mathbb{E}[\widehat{X}\,|\,Y_1,Y'] - \mathbb{E}[\widehat{X}\,|\,Y_1]\big) + \mathbb{E}\,\mathbf{c}\big(\mathbb{E}[\widehat{X}\,|\,Y_1]\big).$$

Since the first term on the right is a positive semi-definite matrix, we get that

$$\mathbb{E}\,\mathbf{c}\big(\mathbb{E}[\widehat{X}\,|\,Y_1,Y']\big) \geqslant \mathbb{E}\,\mathbf{c}\big(\mathbb{E}[\widehat{X}\,|\,Y_1]\big).$$

In view of (2.5), this yields (2.4) and thus the desired result.

It remains to prove (2.5). The quantities on both sides can be written as integrations of f with respect to Gibbs measures with a common reference measure P_N^X (the law of X). Hence, it suffices to compare the Hamiltonians. The Hamiltonian for the left-hand side can be computed to be

$$\begin{aligned}&\frac{2t_2}{N^{p-1}}(x^{\otimes p}A)\cdot(X^{\otimes p}A) + \frac{1}{\sqrt{N^{p-1}}}(x^{\otimes p}A)\cdot\sqrt{2t_2}W_2 - \frac{t_2}{N^{p-1}}|x^{\otimes p}A|^2\\ &\qquad +2h_2\cdot(x^\top X) + (Z_2\sqrt{2h_2})\cdot x - 2h_2\cdot(x^\top x),\end{aligned}$$

while the Hamiltonian for the right-hand side is

$$\begin{aligned}&\frac{2t_2}{N^{p-1}}(x^{\otimes p}A)\cdot(X^{\otimes p}A) + \frac{1}{\sqrt{N^{p-1}}}(x^{\otimes p}A)\cdot\Big(\sqrt{2t_1}W_1 + \sqrt{2t_2-2t_1}W'\Big) - \frac{t_2}{N^{p-1}}|x^{\otimes p}A|^2\\ &\qquad +2h_2\cdot(x^\top X) + \Big(Z_1\sqrt{2h_1} + Z'\sqrt{2h_2-2h_1}\Big)\cdot x - 2h_2\cdot(x^\top x).\end{aligned}$$

Since $W_1, W_2, W', Z_1, Z_2, Z'$ all consist of i.i.d. standard Gaussian entries, we can conclude that the two Hamiltonians have the same distribution, which implies (2.5). □

2.2. **Convexity.** In this subsection, we show the following.

Lemma 2.3. *For each $N \in \mathbb{N}$, the function $\overline{F}_N : [0,\infty)\times\mathbf{S}_+^K \to \mathbb{R}$ is convex.*

Proof. We want to show that for every $(s,a)\in\mathbb{R}\times\mathbf{S}^K$ and every $(t,h)\in[0,\infty)\times\mathbf{S}_+^K$,

$$\big(s\partial_t + a\cdot\nabla\big)^2\overline{F}_N(t,h) \geqslant 0.$$

For brevity, we set $y = \sqrt{\frac{2}{N^{p-1}}}x^{\otimes p}A$ and similarly for replicas of x. We can compute that

$$\begin{aligned}s^2\partial_t^2\overline{F}_N(t,h) &= \frac{2s^2}{N}\mathbb{E}\left\langle (y\cdot y')(y\cdot y' - 2y\cdot y'' + y''\cdot y''')\right\rangle,\\ s\partial_t\big(a\cdot\nabla\overline{F}_N(t,h)\big) &= \frac{2s}{N}\mathbb{E}\left\langle (a\cdot x^\top x')(y\cdot y' - 2y\cdot y'' + y''\cdot y''')\right\rangle,\\ (a\cdot\nabla)^2\overline{F}_N(t,h) &= \frac{2}{N}\mathbb{E}\left\langle (a\cdot x^\top x')(a\cdot x^\top x' - 2a\cdot x^\top x'' + a\cdot x''^\top x''')\right\rangle.\end{aligned}$$

Again, this computation uses the Nishimori identity and the Gaussian integration parts. Details for deriving the third identity above can be seen in the derivation of [19, (3.27)]. The two other identities can be computed by following the same procedure. Let I be the identity matrix of the same size as $y^\intercal y'$. Setting $b = \mathrm{diag}(a, sI)$, $z = \mathrm{diag}(x, y)$ and similarly for replicas, we have $b \cdot z^\intercal z' = sy \cdot y' + a \cdot x^\intercal x'$ (where the matrix product is carried out prior to the dot product). In this notation, adding the above identities together and using the symmetry between replicas, we have

$$\begin{aligned}\big(s\partial_t + a \cdot \nabla\big)^2 \overline{F}_N(t,h) &= \frac{2}{N}\mathbb{E}\left\langle (b \cdot z^\intercal z')^2 - 2(b \cdot z^\intercal z')(b \cdot z^\intercal z'') + (b \cdot z^\intercal z')(b \cdot z'^\intercal z''')\right\rangle \\ &= \frac{2}{N}\mathbb{E}\left\langle (b \otimes b) \cdot \Big(z^\intercal z' \otimes z^\intercal z' - 2z^\intercal \langle z'\rangle \otimes z^\intercal \langle z'\rangle + \langle z\rangle^\intercal \langle z'\rangle \otimes \langle z\rangle^\intercal \langle z'\rangle\Big)\right\rangle.\end{aligned}$$

Writing $\overline{z} = z - \langle z\rangle$ and similarly for replicas, we obtain that the above is equal to

$$\frac{2}{N}\mathbb{E}\left\langle (b \otimes b) \cdot \Big(z^\intercal \overline{z}' \otimes z^\intercal \overline{z}' - \overline{z}^\intercal \langle z'\rangle \otimes \overline{z}^\intercal \langle z'\rangle\Big)\right\rangle.$$

Since b is symmetric, we can see that

$$(b \otimes b) \cdot (\overline{z}^\intercal \langle z'\rangle \otimes \overline{z}^\intercal \langle z'\rangle) = (b \otimes b) \cdot (\langle z'\rangle^\intercal \overline{z} \otimes \langle z'\rangle^\intercal \overline{z}).$$

Using the symmetry between replicas, we conclude from the above three displays that

$$\big(s\partial_t + a \cdot \nabla\big)^2 \overline{F}_N(t,h) = \frac{2}{N}\mathbb{E}\left\langle (b \otimes b) \cdot \Big(\overline{z}^\intercal \overline{z}' \otimes \overline{z}^\intercal \overline{z}'\Big)\right\rangle \geqslant 0. \qquad \square$$

3. Some results of convex analysis

As mentioned above, our approach to proving Theorem 1.1 relies on the identification of the limit of $\overline{F}_N$ as the unique viscosity solution to (2.1). The uniqueness criterion we will use for this purpose is inspired by that described in Appendix A. Compared with the setting explored there, equation (2.1) poses additional difficulties that are caused by the fact that the domain $\mathbf{S}^K_+$ of the "space" variable has a boundary. This is compounded by the fact that the relevant order on $\mathbf{S}^K_+$ is not total. The main purpose of this section is to demonstrate Proposition 3.9, which states that, despite this, the subgradient of a nondecreasing convex function with nondecreasing gradients always has a maximal element (and this maximal element has further good properties). This proposition will be particularly handy in Section 4.

3.1. Preliminaries.

We start by recalling basic definitions and results from convex analysis. Since we need results for both functions defined on $\mathbf{S}^K_+$ and functions on $[0,\infty) \times \mathbf{S}^K_+$, we consider a slightly more general setting in this subsection and specialize into these two spaces when needed.

Let $\mathscr{H}$ be a finite-dimensional Hilbert space. The associated inner product is denoted by a dot product, and the norm by $|\cdot|$. Since $\mathscr{H}$ can be isometrically identified with a Euclidean space, the usual notion of differentiability for any function $u : \mathscr{H} \to \mathbb{R}$ still makes sense. If u is differentiable at $x \in \mathscr{H}$, we denote by $Du(x)$ its differential at x. We also identify $\mathscr{H}$ with its dual and thus $Du(x) \in \mathscr{H}$. For the purpose of this work, the space $\mathscr{H}$ will be taken to be either $\mathbb{R} \times \mathbf{S}^K$ or $\mathbf{S}^K$, and, correspondingly, D will be taken to be either (∂_t, ∇) or ∇.

Let $u : \mathscr{H} \to \mathbb{R} \cup \{\infty\}$ be a convex function. We define its subdifferential at $x \in \mathscr{H}$ by

$$\partial u(x) := \Big\{ y \in \mathscr{H} : \ u(x') \geqslant u(x) + y \cdot (x' - x), \ \forall x' \in \mathscr{H} \Big\}. \tag{3.1}$$

The effective domain of u is

$$\mathsf{dom}\, u := \{x \in \mathscr{H} : u(x) < \infty\}.$$

The function u is called *proper* if $\mathsf{dom}\, u \neq \varnothing$. The outer normal cone to a subset $\mathscr{S} \subseteq \mathscr{H}$ at $x \in \mathscr{H}$ is given by

$$\mathbf{n}_{\mathscr{S}}(x) := \big\{ y \in \mathscr{H} : \ y \cdot (x' - x) \leqslant 0, \ \forall x' \in \mathscr{S} \big\}. \tag{3.2}$$

The following result characterizes the subdifferential as the sum of the outer normal cone and the set of accumulation points of differentials at nearby differentiable points; we refer to [26, Theorem 25.6] for a proof.

Lemma 3.1. *Let $u : \mathscr{H} \to \mathbb{R} \cup \{\infty\}$ be a proper lower semi-continuous convex function such that $\mathsf{dom}\, u$ has nonempty interior. Then, for every $x \in \mathscr{H}$,*

$$\partial u(x) = \mathsf{cl}(\mathsf{conv}\, S) + \mathbf{n}_{\mathsf{dom}\, u}(x).$$

where S is the set of all limits of sequences of the form $\big(Du(x_i)\big)_{i\in\mathbb{N}}$ such that u is differentiable at x_i and $\lim_{i\to\infty} x_i = x$.

We also record two classical results which, while not relevant to the proof of Proposition 3.9, will be useful later on. The first one characterizes the subdifferential of the sum of two convex functions, assuming that one of them is differentiable for simplicity. The second one states a correspondence between elements of the subdifferential at a point and smooth functions that "touch the convex function from below".

Lemma 3.2. *Let $u : \mathscr{H} \to \mathbb{R} \cup \{\infty\}$ be a proper lower semi-continuous convex function such that $\mathsf{dom}\, u$ has nonempty interior. Let $v : \mathscr{H} \to \mathbb{R}$ be convex and differentiable everywhere. Set $u' = u + v$. Then, $\mathsf{dom}\, u = \mathsf{dom}\, u'$ and, for every $x \in \mathsf{dom}\, u$, it holds that*

$$\partial u'(x) = \partial u(x) + \big\{ Dv(x) \big\}.$$

Proof. The first claim is obvious due to the finiteness of v. To see the second claim, we start by noting that due to $\mathsf{dom}\, u = \mathsf{dom}\, u'$, the outer normal cone to $\mathsf{dom}\, u$ is the same as the outer normal cone to $\mathsf{dom}\, u'$ at every point. The differentiability of v implies that u' is differentiable at some point x' if and only if u is also differentiable at x'. Hence, the second claim follows from Lemma 3.1. □

Lemma 3.3. *Let $u : \mathscr{H} \to \mathbb{R} \cup \{\infty\}$ be convex. Then, $p \in \partial u(x)$ for some x if and only if there exists a smooth function $\phi : \mathscr{H} \to \mathbb{R}$ such that $u - \phi$ achieves its minimum at x and $D\phi(x) = p$.*

Proof. Assuming $p \in \partial u(x)$, we can deduce from the definition of subdifferential that $u - \phi$ achieves its minimum at x for $\phi : y \mapsto p \cdot y$. Now, let us assume the converse. The convexity of u implies that

$$u(x') - u(x) \geqslant \frac{1}{\lambda}\Big(u\big(x + \lambda(x' - x)\big) - u(x) \Big), \quad \forall x', \ \forall \lambda \in (0, 1].$$

Using the minimality of $u - \phi$ at x and the differentiability of ϕ at x, we can obtain $D\phi(x) \in \partial u(x)$ by sending $\lambda \to 0$. □

To apply these results to the study of solutions to (2.1), we make the following remark.

Remark 3.4. Any convex function $f : [0, \infty) \times \mathbf{S}^K_+ \to \mathbb{R}$ can be extended in a standard way to a convex function $\overline{f} : \mathbb{R} \times \mathbf{S}^K \to \mathbb{R} \cup \{\infty\}$ by setting $\overline{f} = f$ on $[0, \infty) \times \mathbf{S}^K_+$ and $\overline{f} = \infty$ elsewhere. Note that $\overline{f}$ is proper and its effective domain is $[0, \infty) \times \mathbf{S}^K_+$ which has nonempty interior. If f is continuous, then $\overline{f}$ is lower semi-continuous. In the following, we do not distinguish between f and its standard extension. Then, the notions and results discussed above can be applied to f by setting $\mathscr{H} = \mathbb{R} \times \mathbf{S}^K$ and $D = (\partial_t, \nabla)$. Similar treatments can be taken for any convex function $\psi : \mathbf{S}^K_+ \to \mathbb{R}$.

Finally, since we will work with functions defined on $\mathbf{S}^K_+$ and $[0,\infty)\times\mathbf{S}^K_+$, we record these two simple lemmas.

Lemma 3.5. *For every $a\in\mathbf{S}^K$, we have $a\in\mathbf{S}^K_+$ if and only if $a\cdot b\geqslant 0$ for all $b\in\mathbf{S}^K_+$.*

Lemma 3.6. *For every $t\geqslant 0$ and $x\in\mathbf{S}^K_+$, we have $\mathbf{n}_{\mathbf{S}^K_+}(x)\subseteq -\mathbf{S}^K_+$ and $\mathbf{n}_{[0,\infty)\times\mathbf{S}^K_+}(t,x)\subseteq -([0,\infty)\times\mathbf{S}^K_+)$.*

The first lemma is an application of the diagonalizability of real symmetric matrices (see e.g. [19, Lemma 2.2]), and the second lemma is a consequence of the definition of outer normal cones in (3.2).

3.2. **Nondecreasing gradients.** The key result of this subsection is Proposition 3.9. To state it, it is convenient to introduce the following definitions. Recall the partial orders defined in (2.2) and (2.3).

Definition 3.7 (Nondecreasingness)**.** A real-valued function g defined on $\mathbf{S}^K_+$ or $[0,\infty)\times\mathbf{S}^K_+$ is said to be nondecreasing if $g(y_1)\leqslant g(y_2)$ whenever $y_1\leqslant y_2$.

Definition 3.8 (Nondecreasing gradients)**.** A Lipschitz function $f:[0,\infty)\times\mathbf{S}^K_+\to\mathbb{R}$ is said to have nondecreasing gradients if, for every (t_1,x_1) and (t_2,x_2) that are differentiable points of f and satisfy $(t_1,x_1)\leqslant(t_2,x_2)$, it holds that

$$(\partial_t,\nabla)f(t_1,x_1)\leqslant(\partial_t,\nabla)f(t_2,x_2). \tag{3.3}$$

Recall that, by Rademacher's theorem, a Lipschitz function is differentiable almost everywhere. Here is the main result of this section.

Proposition 3.9. *Suppose that $f:[0,\infty)\times\mathbf{S}^K_+\to\mathbb{R}$ is nondecreasing, Lipschitz, convex, and has nondecreasing gradients. Then, for every $(t,x)\in[0,\infty)\times\mathbf{S}^K_+$, there exists $(b,q)\in\partial f(t,x)\cap[0,\infty)\times\mathbf{S}^K_+$ such that $|(b,q)|\leqslant\|f\|_{\mathrm{Lip}}$ and*

$$\text{for every } (a,p)\in\partial f(t,x),\quad (a,p)\leqslant(b,q). \tag{3.4}$$

In addition, if f satisfies (2.1) on a dense set, then (b,q) can be chosen to satisfy $b-\mathsf{H}(q)=0$.

Remark 3.10. In the statement of Proposition 3.9, the precise interpretation of the phrase that f satisfies (2.1) on a dense set is that the set

$$\left\{(t,x)\in(0,\infty)\times\mathbf{S}^K_{++}\ :\ f\text{ is differentiable at }(t,x)\text{ and }\left(\partial_t f-\mathsf{H}(\nabla f)\right)(t,x)=0\right\}$$

is dense in $[0,\infty)\times\mathbf{S}^K_+$. We point out that one could equivalently replace this condition by the condition that f satisfies (2.1) at every point of differentiability in $(0,\infty)\times\mathbf{S}^K_{++}$. Indeed, one direction of this equivalence is immediate, since every Lipschitz function is differentiable almost everywhere. Conversely, if $(t,x)\in(0,\infty)\times\mathbf{S}^K_{++}$ is a point of differentiability of f, one can find a sequence of points (t_n,x_n) that converge to (t,x) and such that (2.1) is satisfied at (t_n,x_n). Then every subsequential limit of $(\partial_t f,\nabla f)(t_n,x_n)$, say $(a,p)\in\mathbb{R}\times\mathbf{S}^K$, satisfies $a-\mathsf{H}(p)=0$, and one can check that $(a,p)\in\partial f(t,x)$. But since f is differentiable at (t,x) and (t,x) is in the interior (implying that the outer normal cone is $\{0\}$), the subdifferential $\partial f(t,x)$ is the singleton $\{(\partial_t,\nabla)f(t,x)\}$.

Proof of Proposition 3.9. Let $(t,x)\in[0,\infty)\times\mathbf{S}^K_+$. We start by fixing some $(s_0,y_0)\in(0,\infty)\times\mathbf{S}^K_{++}$ such that $|(s_0,y_0)|=1$. Note that

$$(t,x)+\lambda(s_0,y_0)\in[0,\infty)\times\mathbf{S}^K_+,\quad\forall\lambda\geqslant 0.$$

Since f is differentiable a.e. on $[0,\infty)\times\mathbf{S}_+^K$, we can find a sequence $(t_{0,j},x_{0,j})_{j\in\mathbb{N}}$ of differentiable points such that

$$\left|(t_{0,j},x_{0,j})-\left((t,x)+j^{-1}(s_0,y_0)\right)\right|\leqslant j^{-2},\quad \forall j\in\mathbb{N}. \tag{3.5}$$

If, in addition, f satisfies (2.1) on a dense set, then clearly we can choose $(t_{0,j},x_{0,j})_{j\in\mathbb{N}}$ from that set. Since f is Lipschitz, by passing to a subsequence, we may assume that $\lim_{j\to\infty}(\partial_t,\nabla)f(t_{0,j},x_{0,j})$ exists. Denote this limit by (b,q). By Lemma 3.1, we know that $(b,q)\in\partial f(t,x)$. It is clear that $|(b,q)|\leqslant\|f\|_{\mathrm{Lip}}$. Since f is nondecreasing, we also have that $(b,q)\in[0,\infty)\times\mathbf{S}_+^K$. Continuity of H implies that $b-\mathsf{H}(q)=0$ if f satisfies (2.1) on a dense set. It remains to show (3.4).

We apply Lemma 3.1 to the standard extension of f. Note that $\mathsf{dom}\,f=[0,\infty)\times\mathbf{S}_+^K$. Let S be the corresponding set at (t,x) in this lemma. Then, due to this and Lemma 3.6, for each $(a,p)\in\partial f(t,x)$, there is $(a',p')\in\mathsf{cl}(\mathsf{conv}\,S)$ such that $(a,p)\leqslant(a',p')$. Therefore, it suffices to prove (3.4) for $(a,p)\in\mathsf{cl}(\mathsf{conv}\,S)$. In fact, since the condition on (a,p) in (3.4) is stable under convex combinations and passage to the limit, it suffices to show (3.4) for every $(a,p)\in S$.

Let $(a,p)\in S$. By definition of S, there exists a sequence $((t_i,x_i))_{i\in\mathbb{N}}$ converging to (t,x) such that

$$\lim_{i\to\infty}(\partial_t,\nabla)f(t_i,x_i)=(a,p). \tag{3.6}$$

Due to our choice of (s_0,y_0), we can see that for sufficiently large j there is $i(j)\in\mathbb{N}$ such that

$$(t_i,x_i)\leqslant(t_{0,j},x_{0,j}),\quad \forall i\geqslant i(j). \tag{3.7}$$

Indeed, since (s_0,y_0) is strictly positive, there is $C>0$ such that

$$C^{-1}|(a',p')|\leqslant(a',p')\cdot(s_0,y_0)\leqslant C|(a',p')|,\quad \forall(a',p')\in[0,\infty)\times\mathbf{S}_+^K.$$

By this and (3.5), we have that, for every $\mathbf{a}\in[0,\infty)\times\mathbf{S}_+^K$,

$$\mathbf{a}\cdot\left((t_{0,j},x_{0,j})-(t,x)-\frac{1}{2j}(s_0,y_0)\right)\geqslant\frac{1}{2j}\mathbf{a}\cdot(s_0,y_0)-j^{-2}|\mathbf{a}|\geqslant|\mathbf{a}|\left(\frac{1}{2Cj}-\frac{1}{j^2}\right).$$

The right-hand side is nonnegative for sufficiently large j. Lemma 3.5 thus implies that

$$(t_{0,j},x_{0,j})-(t,x)\geqslant\frac{1}{2j}(s_0,y_0).$$

On the other hand, similar arguments yield that, for sufficiently large i (in terms of j),

$$(t_i,x_i)-(t,x)\leqslant\frac{1}{2j}(s_0,y_0).$$

The two previous displays justify (3.7). Using (3.6), (3.7) and the property (3.3), by first sending $i\to\infty$ and then $j\to\infty$, we obtain that

$$(a,p)\leqslant\lim_{j\to\infty}(\partial_t,\nabla)f(t_{0,j},x_{0,j})=(b,q),$$

as desired. □

4. Viscosity solutions

In this section, we study the Hamilton-Jacobi equation (2.1). First, we give the precise definition of viscosity solutions. Then, we recall the uniqueness and existence of viscosity solutions ensured by the comparison principle and the fact that the Hopf formula gives a viscosity solution. We next turn to the main goal of this section, which is to prove Proposition 4.7. This proposition provides us with a convenient sufficient condition for checking whether a function is the unique viscosity solution. This is instrumental in our proof of the convergence of the free energy in Section 5.

Recall that the notion of nondecreasing functions was introduced in Definition 3.7.

Definition 4.1 (Viscosity solutions)**.**

(1) A nondecreasing Lipschitz function $f : [0,\infty) \times \mathbf{S}^K_+ \to \mathbb{R}$ is a viscosity subsolution to (2.1) if for every $(t,h) \in (0,\infty) \times \mathbf{S}^K_+$ and every smooth $\phi : (0,\infty) \times \mathbf{S}^K_+ \to \mathbb{R}$ such that $f - \phi$ has a local maximum at (t,h), we have

$$\begin{cases} \big(\partial_t \phi - \mathsf{H}(\nabla \phi)\big)(t,h) \leqslant 0, & \text{if } h \in \mathbf{S}^K_{++}, \\ \nabla \phi(t,h) \in \mathbf{S}^K_+, & \text{if } h \in \mathbf{S}^K_+ \smallsetminus \mathbf{S}^K_{++}. \end{cases}$$

(2) A nondecreasing Lipschitz function $f : [0,\infty) \times \mathbf{S}^K_+ \to \mathbb{R}$ is a viscosity supersolution to (2.1) if for every $(t,h) \in (0,\infty) \times \mathbf{S}^K_+$ and every smooth $\phi : (0,\infty) \times \mathbf{S}^K_+ \to \mathbb{R}$ such that $f - \phi$ has a local minimum at (t,h), we have

$$\begin{cases} \big(\partial_t \phi - \mathsf{H}(\nabla \phi)\big)(t,h) \geqslant 0, & \text{if } h \in \mathbf{S}^K_{++}, \\ \partial_t \phi(t,h) - \inf \mathsf{H}(q) \geqslant 0, \quad , & \text{if } h \in \mathbf{S}^K_+ \smallsetminus \mathbf{S}^K_{++}, \end{cases}$$

where the infimum is taken over all $q \in \big(\nabla \phi(t,h) + \mathbf{S}^K_+\big) \cap \mathbf{S}^K_+$ and $|q| \leqslant \|f\|_{\mathrm{Lip}}$.

(3) A nondecreasing Lipschitz function $f : [0,\infty) \times \mathbf{S}^K_+ \to \mathbb{R}$ is a viscosity solution to (2.1) if f is both a viscosity subsolution and supersolution.

Remarks 4.3 and 4.4 below aim to provide a somewhat more intuitive understanding of Definition 4.1. Before doing so, we record the following observation.

Lemma 4.2. *The function $\mathsf{H} : \mathbf{S}^K_+ \to \mathbb{R}$ given in (1.4) is nondecreasing.*

Proof. Let $a, b \in \mathbf{S}^K_+$ be such that $a \leqslant b$. Recalling that the tensor product of two positive semidefinite matrices is positive definite, see for instance [27, Theorem 7.20], one can show by induction on p that $a^{\otimes \mathsf{p}} \leqslant b^{\otimes \mathsf{p}}$. Since $AA^\mathsf{T} \in \mathbf{S}^{K^\mathsf{p}}_+$, we can use Lemma 3.5 to obtain that $\mathsf{H}(a) \leqslant \mathsf{H}(b)$, as desired. □

Remark 4.3. Given a nondecreasing Lipschitz function f, define the extension of H by

$$\overline{\mathsf{H}}(p) := \inf \Big\{ \mathsf{H}(q) : q \geqslant p,\ q \in \mathbf{S}^K_+,\ |q| \leqslant \|f\|_{\mathrm{Lip}} \Big\}, \quad \forall p \in \mathbf{S}^K. \tag{4.1}$$

As usual, the infimum over an empty set is understood to be ∞. Note that $\overline{\mathsf{H}} : \mathbf{S}^K \to \mathbb{R} \cup \{\infty\}$ is lower semi-continuous and agrees with H on $\mathbf{S}^K_+$ due to Lemma 4.2. Then, Definition 4.1 (2) can be reformulated as follows: f is viscosity supersolution if for every $(t,h) \in (0,\infty) \times \mathbf{S}^K_+$ and every smooth $\phi : (0,\infty) \times \mathbf{S}^K_+$ such that $f - \phi$ has a local minimum at (t,h), we have

$$\big(\partial_t \phi - \overline{\mathsf{H}}(\nabla \phi)\big)(t,h) \geqslant 0.$$

Note that, in this formulation, we do not need to distinguish between $h \in \mathbf{S}^K_{++}$ and $h \in \mathbf{S}^K_+ \smallsetminus \mathbf{S}^K_{++}$.

Remark 4.4. Equation 2.1 is probably best thought of as the vanishing-viscosity limit of the equation with the nonlinearity H replaced by $\overline{\mathsf{H}}$ and with a Neumann boundary condition. For such equations, it would be more standard to write the condition in the definition of subsolution as

$$\begin{cases} \big(\partial_t\phi - \overline{\mathsf{H}}(\nabla\phi)\big)(t,h) \leqslant 0, & \text{if } h \in \mathbf{S}^K_{++}, \\ \min\left(\big(\partial_t\phi - \overline{\mathsf{H}}(\nabla\phi)\big)(t,h), \inf\limits_{\nu\in\mathbf{n}_{\mathbf{S}^K_+}(h)} \nu\cdot\nabla\phi(t,h)\right) \leqslant 0, & \text{if } h \in \mathbf{S}^K_+ \setminus \mathbf{S}^K_{++}, \end{cases}$$

and the condition in the definition of supersolution as

$$\begin{cases} \big(\partial_t\phi - \overline{\mathsf{H}}(\nabla\phi)\big)(t,h) \geqslant 0, & \text{if } h \in \mathbf{S}^K_{++}, \\ \max\left(\big(\partial_t\phi - \overline{\mathsf{H}}(\nabla\phi)\big)(t,h), \sup\limits_{\nu\in\mathbf{n}_{\mathbf{S}^K_+}(h)} \nu\cdot\nabla\phi(t,h)\right) \geqslant 0, & \text{if } h \in \mathbf{S}^K_+ \setminus \mathbf{S}^K_{++}. \end{cases}$$

Using Lemma 3.6, one can verify that Definition 4.1 is more stringent than the one alluded to in this remark. In principle, adopting a more stringent definition should complicate the argument for the existence of solutions. But in our context, we can rely on the Hopf formula to obtain without much additional difficulties that the solution to (2.1) indeed satisfies this more stringent (and somewhat simpler) definition.

We turn to the well-posedness of equation (2.1). We first state a comparison principle, which ensures in particular that there is at most one viscosity solution with a given initial condition.

Proposition 4.5 (Comparison principle)**.** *If u is a supersolution and v is a subsolution to (2.1), then*

$$\sup_{[0,\infty)\times\mathbf{S}^K_+}(u-v) = \sup_{\{0\}\times\mathbf{S}^K_+}(u-v).$$

For suitable initial conditions, the viscosity solution admits the following variational representation.

Proposition 4.6 (Hopf formula)**.** *Let $\psi : \mathbf{S}^K_+ \to \mathbb{R}$ be convex, Lipschitz and nondecreasing, and let f be given by*

$$f(t,h) := \sup_{h''\in\mathbf{S}^K_+}\inf_{h'\in\mathbf{S}^K_+}\big\{h''\cdot(h-h') + \psi(h') + t\mathsf{H}(h'')\big\}, \quad \forall (t,h)\in[0,\infty)\times\mathbf{S}^K_+.$$

Then, the function f is a viscosity solution to (2.1) with initial condition $f(0,\cdot) = \psi$.

For the proofs of these two propositions, we refer to [8, Section 6].

In the remaining of this section, for convenience, we will use x, y as spatial variables in place of h, which should not be confused with the notation for random variables under the Gibbs measure $\langle\cdot\rangle$ in Section 2.

4.1. **Identification criterion.** The following result gives a convenient criterion for a function to be a viscosity solution.

Proposition 4.7. *Let $f : [0,\infty)\times\mathbf{S}^K_+ \to \mathbb{R}$ be nondecreasing, Lipschitz, convex, and have nondecreasing gradients. Suppose that $\psi = f(0,\cdot)$ is C^1 and that f satisfies (2.1) on a dense subset. Then, f is a viscosity solution to (2.1) with initial condition ψ.*

For the reader's convenience, the idea for the proof of this proposition is also presented in the simpler setting of Hamilton-Jacobi equations on $[0,\infty)\times\mathbb{R}^d$ in Appendix A. Two essential ingredients for this argument are the C^1 assumption of the initial condition

and the convexity of f. At least in the simpler context explored in Appendix A, both assumptions are necessary; see in particular Remark A.3 there.

Compared with the Euclidean setting discussed in Appendix A, the existence of the boundary of $\mathbf{S}^K_+$ complicates the arguments. Indeed, in view of Lemma 3.1, on the boundary, the subdifferential contains an additional component from the outer normal cone. Therefore, if $p \in \partial\psi(y)$ for a boundary point y, we cannot identify p with $\nabla\psi(y)$. The identity $p = \nabla\psi(y)$ is important in Step 2 of the proof of Proposition A.2. It turns out that for Proposition 4.7, a work-around is available by exploiting the assumption that the function f has nondecreasing gradients.

As preparation for this, we use Proposition 3.9 to prove the following lemma. This lemma can be interpreted as stating that we can always "lift" a subdifferential $p \in \partial\psi(y)$ to a subdifferential $(b,p) \in \partial f(0,y)$ which is dominated by some $(b,p') \in \partial f(0,y)$ satisfying the Hamilton-Jacobi equation. This lemma is needed due to the presence of boundary. Indeed, on $[0,\infty)\times\mathbb{R}^d$, the existence of such a "lift" is automatic, which can be seen in Step 2 of the proof of Proposition A.2.

Lemma 4.8. *Under the assumptions in Proposition 4.7, for every $y \in \mathbf{S}^K_+$ and every $p \in \partial\psi(y)$, there is $(b,p') \in [0,\infty)\times\mathbf{S}^K_+$ such that $(b,p) \in \partial f(0,y)$, $p' \geqslant p$, $|(b,p')| \leqslant \|f\|_{\mathrm{Lip}}$ and $b - \mathsf{H}(p') = 0$.*

Proof. Since $\psi : \mathbf{S}^K_+ \to \mathbb{R}$ is C^1, by Lemma 3.1 and setting $p' = \nabla\psi(y)$, we have

$$\partial\psi(y) = \{p'\} + \mathbf{n}_{\mathbf{S}^K_+}(y).$$

This implies that

$$p = p' + n \tag{4.2}$$

for some

$$n \in \mathbf{n}_{\mathbf{S}^K_+}(y). \tag{4.3}$$

Due to Lemma 3.6, we have $-n \in \mathbf{S}^K_+$, that is,

$$p \leqslant p'. \tag{4.4}$$

The same argument also yields that,

$$\text{for every } q' \in \partial\psi(y), \quad q' \leqslant p'. \tag{4.5}$$

Since f is nondecreasing, we have that, for all $(t',x') \in [0,\infty)\times\mathbf{S}^K_+$,

$$f(t',x') - f(0,y) \geqslant f(0,x') - f(0,y) = \psi(x') - \psi(y),$$

which implies that $(0,p') \in \partial f(0,y)$. Let

$$(b,q) \in \partial f(0,y) \tag{4.6}$$

be as described in Proposition 3.9, for f at the point $(0,y)$. Then, the following properties hold

$$(0,p') \leqslant (b,q), \tag{4.7}$$

$$|(b,q)| \leqslant \|f\|_{\mathrm{Lip}}, \quad b - \mathsf{H}(q) = 0. \tag{4.8}$$

Since $f(0,\cdot) = \psi$, we must have $q \in \partial\psi(y)$. Combining (4.5) and (4.7), we see that

$$p' = q. \tag{4.9}$$

We are now ready to conclude. By (4.3) and the definition of the outer normal in (3.2), we can verify that

$$(0,n) \in \mathbf{n}_{[0,\infty)\times\mathbf{S}^K_+}(0,y).$$

This along with Lemma 3.1, (4.6) and (4.9) implies

$$(b, p' + n) \in \partial f(0, y).$$

The lemma then follows from this display, (4.2), (4.4), (4.8) and (4.9). □

We are now ready to prove our criterion for the identification of solutions.

Proof of Proposition 4.7. We check that f must be a subsolution to (2.1). Let $\phi \in C^\infty((0,\infty) \times \mathbf{S}^K_+)$, and $(t,x) \in (0,\infty) \times \mathbf{S}^K_+$ be such that $f - \phi$ has a local maximum at (t,x). If $x \in \mathbf{S}^K_+ \setminus \mathbf{S}^K_{++}$, since, for each $a \in \mathbf{S}^K_+$ and sufficiently small $\varepsilon > 0$,

$$0 \leqslant f(t, x + \varepsilon a) - f(t,x) \leqslant \phi(t, x + \varepsilon a) - \phi(t,x),$$

we must have $a \cdot \nabla\phi(t,x) \geqslant 0$ for all $a \in \mathbf{S}^K_+$. By Lemma 3.5, this implies that $\nabla\phi(t,x) \in \mathbf{S}^K_+$. If $x \in \mathbf{S}^K_{++}$, then we have,

$$f(t',x') - f(t,x) \leqslant (t'-t)\partial_t\phi(t,x) + (x'-x)\cdot\nabla\phi(t,x) + o(|t'-t| + |x'-x|).$$

This implies that the subdifferential $\partial f(t,x)$ is the singleton $\{(\partial_t, \nabla)\phi(t,x)\}$, and thus that f is differentiable at (t,x), with $(\partial_t,\nabla)f(t,x) = (\partial_t,\nabla)\phi(t,x)$. Using also Remark 3.10, we deduce that

$$\big(\partial_t\phi - \mathsf{H}(\nabla\phi)\big)(t,x) = \big(\partial_t f - \mathsf{H}(\nabla f)\big)(t,x) = 0,$$

as desired.

Now we want to show that f is a supersolution to (2.1). Fix any (t,x), and any

$$(a,p) \in \partial f(t,x). \tag{4.10}$$

Recall Remark 4.3 and the extension $\overline{\mathsf{H}}$ defined there. By Lemma 3.3, it suffices to show

$$a - \overline{\mathsf{H}}(p) \geqslant 0. \tag{4.11}$$

We proceed in four steps.

Step 1. We claim that, for every $\varepsilon > 0$, the following infimum

$$\inf_{y \in \mathbf{S}^K_+} \big(f_\varepsilon(0,y) - y\cdot p\big) \tag{4.12}$$

is achieved, where, for every $(s,y) \in [0,\infty) \times \mathbf{S}^K_+$, we have set

$$f_\varepsilon(s,y) := f(s,y) + \varepsilon\sqrt{1+|y|^2}.$$

One can verify that $y \mapsto \sqrt{1+|y|^2}$ is convex, and thus so is f_ε. By the definition of subdifferentials, we have

$$f(0,y) - f(t,x) \geqslant (a,p)\cdot(-t, y-x), \quad \forall y \in \mathbf{S}^K_+,$$

which implies that

$$f_\varepsilon(0,y) - y\cdot p \geqslant \varepsilon\sqrt{1+|y|^2} + f(t,x) - (a,p)\cdot(t,x), \quad \forall y \in \mathbf{S}^K_+.$$

Hence, the left-hand side of the inequality above is bounded below and tends to infinity as $|y|$ tends to infinity. Therefore, a minimizer exists and we denote it by $y_\varepsilon \in \mathbf{S}^K_+$.

Step 2. We show

$$\lim_{\varepsilon\to 0} \varepsilon\sqrt{1+|y_\varepsilon|^2} = 0. \tag{4.13}$$

We first observe that

$$\limsup_{\varepsilon\to 0} \inf_{y\in\mathbf{S}^K_+} \Big(f(0,y) + \varepsilon\sqrt{1+|y|^2} - y\cdot p\Big) = \inf_{y\in\mathbf{S}^K_+} \big(f(0,y) - y\cdot p\big). \tag{4.14}$$

Indeed, for any $\delta > 0$, there is $\overline{y} \in \mathbf{S}^K_+$ such that

$$f(0,\overline{y}) - \overline{y}\cdot p \leqslant \inf(f(0,y) - y\cdot p) + \delta/2,$$

and we can choose $\overline{\varepsilon} > 0$ small enough such that, for every $\varepsilon \in (0,\overline{\varepsilon})$,

$$f(0,\overline{y}) + \varepsilon\sqrt{1+|\overline{y}|^2} - \overline{y}\cdot p \leqslant \inf(f(0,y) - y\cdot p) + \delta.$$

This implies that

$$\limsup_{\varepsilon\to 0} \inf_{y\in\mathbf{S}^K_+} \left(f(0,y) + \varepsilon\sqrt{1+|y|^2} - y\cdot p\right) \leqslant \inf_{y\in\mathbf{S}^K_+} \left(f(0,y) - y\cdot p\right),$$

and the other direction of the inequality in (4.14) is obvious. Since y_ε achieves the infimum on the left-hand side of (4.14) and also satisfies

$$f(0,y_\varepsilon) - y_\varepsilon\cdot p \geqslant \inf_{y\in\mathbf{S}^K_+} \left(f(0,y) - y\cdot p\right),$$

we conclude that (4.13) holds.

Step 3. Let $\psi_\varepsilon := f_\varepsilon(0,\cdot)$, so that $\psi_\varepsilon = \psi + \varepsilon\sqrt{1+|\cdot|^2}$. Since y_ε achieves the infimum in (4.12), we have that $p \in \partial\psi_\varepsilon(y_\varepsilon)$. Lemma 3.2 implies that

$$p = p_\varepsilon + \frac{\varepsilon y_\varepsilon}{\sqrt{1+|y_\varepsilon|^2}}$$

for some $p_\varepsilon \in \partial\psi(y_\varepsilon)$. In particular, we have

$$|p - p_\varepsilon| \leqslant \varepsilon. \tag{4.15}$$

By Lemma 4.8 applied to p_ε, there exists $(b_\varepsilon, p'_\varepsilon) \in [0,\infty)\times\mathbf{S}^K_+$ such that

$$(b_\varepsilon, p_\varepsilon) \in \partial f(0,y), \tag{4.16}$$

$$p_\varepsilon \leqslant p'_\varepsilon, \quad p'_\varepsilon \in \mathbf{S}^K_+, \quad |p'_\varepsilon| \leqslant \|f\|_{\mathrm{Lip}} \tag{4.17}$$

$$b_\varepsilon - \mathsf{H}(p'_\varepsilon) = 0. \tag{4.18}$$

Step 4. We are now ready to prove (4.11). Define $h : \lambda \mapsto f\left(\lambda(t,x) + (1-\lambda)(0,y_\varepsilon)\right)$ on $[0,1]$. Clearly, h is convex. By (4.16), the right derivative of h at 0 satisfies

$$h'_+(0) \geqslant b_\varepsilon t + p_\varepsilon\cdot(x - y_\varepsilon).$$

On the other hand, due to (4.10), the left derivative at 1 satisfies

$$h'_-(1) \leqslant at + p\cdot(x - y_\varepsilon).$$

By convexity of h, we must have $h'_+(0) \leqslant h'_-(1)$. This along with (4.15) and (4.13) implies that, as ε tends to zero,

$$a \geqslant b_\varepsilon + o(1).$$

By (4.18), the definition of $\overline{\mathsf{H}}$ in (4.1), and (4.17), we have that

$$b_\varepsilon = \mathsf{H}(p'_\varepsilon) \geqslant \overline{\mathsf{H}}(p_\varepsilon).$$

Using that $\overline{\mathsf{H}}$ is lower semi-continuous and (4.15) together with the two previous displays yields that

$$a \geqslant \overline{\mathsf{H}}(p) + o(1),$$

and (4.11) follows by letting ε tend to zero. □

In the corollary below, we rephrase our criterion for identifying solutions in the following way: instead of asking for the equation to be valid on a dense subset, we ask that it be valid at any point at which the candidate function can be touched from above by a smooth function. As will be seen in the next section, the main advantage to this formulation is that, by convexity, we automatically benefit from a control on the Hessian of the candidate function at the contact point.

Corollary 4.9. *Let $f : [0,\infty) \times \mathbf{S}_+^K \to \mathbb{R}$ be nondecreasing, Lipschitz, convex, and have nondecreasing gradients. Suppose that $\psi = f(0,\cdot)$ is C^1, and that the following property holds: for every $\phi \in C^\infty((0,\infty)\times \mathbf{S}_+^K)$ and $(t,x) \in (0,\infty)\times \mathbf{S}_{++}^K$ such that $f-\phi$ achieves a strict local maximum at (t,x), we have*

$$(\partial_t \phi - \mathsf{H}(\nabla \phi))(t,x) = 0.$$

Then f is a viscosity solution to (2.1).

Proof. Let ϕ and (t,x) be as in the statement of the corollary. Since f is convex, we have that, for any $(a,p) \in \partial f(t,x)$ and $(t',x') \in (0,\infty)\times \mathbf{S}_+^K$,

$$\begin{aligned} a(t'-t) + p\cdot(x'-x) &\leqslant f(t',x') - f(t,x) \\ &\leqslant \partial_t\phi(t,x)(t'-t) + \nabla\phi(t,x)\cdot(x'-x) + o(|t'-t|+|x'-x|). \end{aligned}$$

It then follows that f is differentiable at (t,x) and the derivatives of f at (t,x) coincide with those of ϕ. By Proposition 4.7, it suffices to show that the set

(4.19)
$$\Big\{(t,x) \in (0,\infty)\times \mathbf{S}_{++}^K : \exists\, \phi \in C^\infty((0,\infty)\times \mathbf{S}_{++}^K) \text{ s.t. } (t,x) \text{ is a local maximum of } f-\phi\Big\}$$

is dense. (The additional restriction that the local maximum be strict is easily addressed a posteriori.) We fix any $(t,x) \in (0,\infty)\times \mathbf{S}_{++}^K$, and for every $\alpha \geqslant 1$, we define

$$\phi_\alpha :\ (t',x') \ \mapsto\ \frac{\alpha}{2}(t'-t)^2 + \frac{\alpha}{2}|x'-x|^2.$$

Since f is Lipschitz, we can verify that $f-\phi_\alpha$ achieves a global maximum at some point (t_α, x_α). Using the Lipschitzness of f and that $(f-\phi_\alpha)(t_\alpha,x_\alpha) \geqslant (f-\phi_\alpha)(t,x)$, we can show that there is a constant $C<\infty$ such that for every $\alpha \geqslant 1$,

$$|t_\alpha - t| + |x_\alpha - x| \leqslant \frac{C}{\alpha}.$$

This implies that $\lim_{\alpha\to\infty}(t_\alpha,x_\alpha) = (t,x)$. Also, since $(t,x) \in (0,\infty)\times \mathbf{S}_{++}^K$, we have that $(t_\alpha,x_\alpha) \in (0,\infty)\times \mathbf{S}_{++}^K$ for every sufficiently large α. Hence (t_α,x_α) belongs to the set in (4.19), and we conclude that the set in (4.19) is a dense subset of $[0,\infty)\times \mathbf{S}_+^K$. □

5. Convergence and application

The main goal of this section is to prove Theorem 1.1, using the tools developed in the previous section. For illustration, we also apply the theorem to a specific model.

5.1. Convergence.

In view of Proposition 4.6, Theorem 1.1 follows from the next theorem.

Theorem 5.1. *Under the conditions of Theorem 1.1, the function $\overline{F}_N$ converges pointwise to the unique viscosity solution to (2.1) with initial condition ψ.*

In order to prove this result, we start by recalling from [8, Proposition 3.1] (cf. also [19, Proposition 1.2]) that the function $\overline{F}_N$ satisfies an approximate form of the equation. In (5.1), we implicitly understand that the relevant functions are evaluated at $(t,h) \in [0,\infty)\times \mathbf{S}_+^K$.

Proposition 5.2 (Approximate Hamilton-Jacobi equation)**.** *There exists $C < \infty$ such that for every $N \geqslant 1$ and uniformly over $[0,\infty) \times \mathbf{S}^K_+$,*

$$\left|\partial_t \overline{F}_N - \mathsf{H}(\nabla \overline{F}_N)\right|^2 \leqslant C\kappa(h) N^{-\frac{1}{4}} \left(\Delta \overline{F}_N + |h^{-1}|\right)^{\frac{1}{4}} + C\mathbb{E}\left[\left|\nabla F_N - \nabla \overline{F}_N\right|^2\right], \tag{5.1}$$

where κ is the condition number of $h \in \mathbf{S}^K_+$ given by

$$\kappa(h) := \begin{cases} |h||h^{-1}|, & \text{if } h \in \mathbf{S}^K_{++}, \\ +\infty & \text{otherwise.} \end{cases}$$

Proof of Theorem 5.1. Since $\overline{F}_N$ is Lipschitz uniformly in N by Lemma 2.1, the Arzelá-Ascoli theorem implies that, for every subsequence of $(\overline{F}_N)_{n\in\mathbb{N}}$, there is a further subsequence converging to some function f in the local uniform topology. It suffices to show that f is a viscosity solution to (2.1) and the uniqueness is ensured by Proposition 4.5. For convenience, we assume that the whole sequence $(\overline{F}_N)_{N\in\mathbb{N}}$ converges to f.

Lemmas 2.1 and 2.3 ensure that f is nondecreasing, Lipschitz and convex. Since $\overline{F}_N$ and f are convex, we have

$$\lim_{N\to\infty} (\partial_t, \nabla)\overline{F}_N(t,h) = (\partial_t, \nabla) f(t,h)$$

at every differentiable point (t,h) of f (indeed, any limit point of $(\partial_t, \nabla)\overline{F}_N(t,h)$ must belong to the subdifferential of f at (t,h), which is a singleton if f is differentiable at (t,h)). This along with Lemma 2.2 yields that f has nondecreasing gradients. Let $(t,h) \in (0,\infty) \times \mathbf{S}^K_{++}$ and $\phi \in C^\infty((0,\infty) \times \mathbf{S}^K_+)$ be such that $f - \phi$ has a strict local maximum at (t,h). By Corollary 4.9, it suffices to show that

$$(\partial_t \phi - \mathsf{H}(\nabla \phi))(t,h) = 0. \tag{5.2}$$

Since $\overline{F}_N$ converges locally uniformly to f, there exists $(t_N, h_N) \in [0,\infty) \times \mathbf{S}^K_+$ such that $\overline{F}_N - \phi$ has a local maximum at (t_N, h_N), and (t_N, h_N) converges to (t,h) as N tends to infinity. Since $(t,h) \in (0,\infty) \times \mathbf{S}^K_{++}$, each (t_N, h_N) also ultimately belongs to $(0,\infty) \times \mathbf{S}^K_{++}$, and without loss of generality, we can assume that every (t_N, h_N) remains a positive distance away from the boundary of $[0,\infty) \times \mathbf{S}^K_+$, uniformly over N. Notice that

$$(\partial_t \overline{F}_N - \partial_t \phi)(t_N, h_N) = 0 \quad \text{and} \quad (\nabla \overline{F}_N - \nabla \phi)(t_N, h_N) = 0. \tag{5.3}$$

Throughout the rest of the proof, we use the letter $C < \infty$ to denote a constant whose value may change from one occurrence to the next, and is allowed to depend on (t,h) and ϕ. We decompose the argument into three steps.

Step 1. We show that for every $h' \in \mathbf{S}^K_+$ with $|h'| \leqslant C^{-1}$, we have

$$0 \leqslant \overline{F}_N(t_N, h_N + h') - \overline{F}_N(t_N, h_N) - h' \cdot \nabla \overline{F}_N(t_N, h_N) \leqslant C|h'|^2. \tag{5.4}$$

The first inequality follows from the convexity of $\overline{F}_N$. To derive the second inequality, we start by writing Taylor's formula:

$$\begin{aligned} &\overline{F}_N(t_N, h_N + h') - \overline{F}_N(t_N, h_N) \\ &\qquad = h' \cdot \nabla \overline{F}_N(t_N, h_N) + \int_0^1 (1-s) h' \cdot \nabla \left(h' \cdot \nabla \overline{F}_N\right)(t_N, h_N + sh')\, \mathrm{d}s. \end{aligned} \tag{5.5}$$

The same formula also holds if we substitute $\overline{F}_N$ by ϕ throughout. Since $\overline{F}_N - \phi$ has a local maximum at (t_N, h_N), we have for every $|h'| \leqslant C^{-1}$ that

$$\overline{F}_N(t_N, h_N + h') - \overline{F}_N(t_N, h_N) \leqslant \phi(t_N, h_N + h') - \phi(t_N, h_N).$$

Using also (5.3), we obtain that

$$\int_0^1 (1-s) h' \cdot \nabla \left(h' \cdot \nabla \overline{F}_N\right)(t_N, h_N + sh')\, \mathrm{d}s \leqslant \int_0^1 (1-s) h' \cdot \nabla \left(h' \cdot \nabla \phi\right)(t_N, h_N + sh')\, \mathrm{d}s.$$

Since the function ϕ is smooth, the right side of this inequality is bounded by $C|h'|^2$. Using (5.5) once more, we obtain (5.4).

Step 2. Let

$$D := \left\{(t', h') \in [0,\infty) \times \mathbf{S}_+^K \ : \ |t' - t| \leqslant C^{-1} \text{ and } |h' - h| \leqslant C^{-1}\right\}.$$

In this step, we show that

$$\mathbb{E}\left[|\nabla F_N - \nabla \overline{F}_N|^2(t_N, h_N)\right] \leqslant C\left(\mathbb{E}\left[\sup_D |F_N - \overline{F}_N|^2\right]\right)^{\frac{1}{2}}. \tag{5.6}$$

We recall from [8, (3.13)] that, for every $a \in \mathbf{S}^K$ and $(t', h') \in [0,\infty) \times \mathbf{S}_+^K$ such that $|h' - h| \leqslant C^{-1}$, we have

$$a \cdot \nabla(a \cdot \nabla F_N)(t', h') \geqslant -C|a|^2 \frac{|Z|}{\sqrt{N}},$$

and that Z is an N-by-K matrix of independent standard Gaussians. As a consequence, for every $|h'| \leqslant C^{-1}$, we have

$$F_N(t_N, h_N + h') \geqslant F_N(t_N, h_N) + h' \cdot \nabla F_N(t_N, h_N) - C|h'|^2 \frac{|Z|}{\sqrt{N}}.$$

Combining this with (5.4), we obtain that, for every $|h'| \leqslant C^{-1}$,

$$h' \cdot \left(\nabla F_N - \nabla \overline{F}_N\right)(t_N, h_N) \leqslant 2 \sup_D |F_N - \overline{F}_N| + C|h'|^2 \left(1 + \frac{|Z|}{\sqrt{N}}\right).$$

For some deterministic $\lambda \in [0, C^{-1}]$ to be determined, we fix

$$h' := \lambda \frac{\left(\nabla F_N - \nabla \overline{F}_N\right)(t_N, h_N)}{\left|\left(\nabla F_N - \nabla \overline{F}_N\right)(t_N, h_N)\right|},$$

so that

$$\lambda |\nabla F_N - \nabla \overline{F}_N|(t_N, h_N) \leqslant 2 \sup_D |F_N - \overline{F}_N| + C\lambda^2 \left(1 + \frac{|Z|}{\sqrt{N}}\right).$$

Squaring this expression and taking the expectation yields

$$\lambda^2 \mathbb{E}\left[|\nabla F_N - \nabla \overline{F}_N|^2(t_N, h_N)\right] \leqslant 8\mathbb{E}\left[\sup_D |F_N - \overline{F}_N|^2\right] + C\lambda^4 \mathbb{E}\left[\left(1 + \frac{|Z|}{\sqrt{N}}\right)^2\right].$$

Since $\mathbb{E}[|Z|^2] = NK$, choosing $\lambda^4 = \mathbb{E}\left[\sup_D |F_N - \overline{F}_N|^2\right]$ yields (5.6).

Step 3. Recall that we assume that $\mathbb{E}\left[\sup_D |F_N - \overline{F}_N|^2\right]$ tends to zero as N tends to infinity. By Proposition 5.2, (5.4), and (5.6), we obtain that

$$\lim_{N \to \infty} \left(\partial_t \overline{F}_N - \mathsf{H}(\nabla \overline{F}_N)\right)(t_N, h_N) = 0.$$

Using also (5.3) and the fact that the function ϕ is smooth, this yields (5.2), and thus completes the proof. □

5.2. **Application.** We study the model considered in [14], which corresponds to (1.1) with $L = 1$, $\mathsf{p} \in \mathbb{N}$, and $A \in \mathbb{R}^{K^{\mathsf{p}} \times 1}$ given by $A_{\mathbf{j}} = 1$ if $j_1 = j_2 = \cdots = j_{\mathsf{p}}$ and zero otherwise. Here, we used the multi-index notation $\mathbf{j} = (j_1, j_2, \ldots, j_{\mathsf{p}}) \in \{1, \ldots, K\}^{\mathsf{p}}$. Explicitly, this model can be expressed as

$$Y_{\mathbf{i}} = \sqrt{\frac{2t}{N^{\mathsf{p}-1}}} \sum_{j=1}^{K} \prod_{n=1}^{\mathsf{p}} X_{i_n, j} + W_{\mathbf{i}}, \quad \mathbf{i} \in \{1, \cdots, N\}^{\mathsf{p}}, \tag{5.7}$$

where $X \in \mathbb{R}^{N\times K}$ is assumed to have i.i.d. row vectors. For even p, the limit of the free energy associated with this model has been proved to satisfy a variational formula in [14]. When p is odd, the situation is more difficult; in [8], it was only proven that the limit is bounded above by a variational formula. Here, we will apply Theorem 1.1 to treat both even and odd values of p.

Recall the definition of H in (1.4). In this case, the nonlinearity H is given by

$$\mathsf{H}(q) = \sum_{k,k'=1}^{K} (q_{k,k'})^{\mathsf{p}}, \quad \forall q \in \mathbf{S}_+^K. \tag{5.8}$$

Since row vectors of X are i.i.d., we have $\overline{F}_N(0,\cdot) = \overline{F}_1(0,\cdot)$ for all $N \in \mathbb{N}$. Setting $\psi := \overline{F}_1(0,\cdot)$ and using the formula for F_N in (1.3), we have

$$\psi(h) = \mathbb{E}\log \int_{\mathbb{R}^{1\times K}} \exp\left(2h\cdot(x^\top X_{1,\cdot}) + \sqrt{2h}\cdot(x^\top Z) - h\cdot(x^\top x)\right) \mathrm{d}P(x), \quad \forall h \in \mathbf{S}_+^K, \tag{5.9}$$

where P is the law of the first row vector $X_{1,\cdot} = (X_{1,k})_{1\leqslant k\leqslant K}$. By Lemma 2.1, ψ is C^1. The concentration condition $\lim_{N\to\infty} \mathbb{E}\|F_N - \overline{F}_N\|_{L^\infty(D)}^2 = 0$ for each compact $D \subseteq [0,\infty)\times\mathbf{S}_+^K$ is proved in [8, Lemma C.1]. Hence, the next result follows from Theorem 1.1.

Corollary 5.3. *Under the assumption (1.2), in the model described above with $\mathsf{p} \in \mathbb{N}$, it holds that, for every $(t,h) \in [0,\infty)\times\mathbf{S}_+^K$,*

$$\lim_{N\to\infty} \overline{F}_N(t,h) = \sup_{h''\in\mathbf{S}_+^K} \inf_{h'\in\mathbf{S}_+^K} \left\{h''\cdot(h-h') + \psi(h') + t\mathsf{H}(h'')\right\},$$

for H and ψ given in (5.8) and (5.9), respectively.

Appendix A. On convex viscosity solutions

The goal of this section is to demonstrate the workings of a convenient uniqueness criterion for Hamilton-Jacobi equations, in the simpler context of equations posed on $[0,\infty)\times\mathbb{R}^d$. This criterion states that, if the function under consideration is *convex*, then we can assert that it is the viscosity solution of some Hamilton-Jacobi equation as soon as it satisfies the equation on a dense subset and the initial condition is of class C^1. This criterion is generalized to equations posed on $[0,\infty)\times\mathbf{S}_+^K$ in Proposition 4.7.

Let $\mathsf{H}:\mathbb{R}^d\to\mathbb{R}$ be a smooth function. We start by recalling the notion of viscosity solutions to

$$\partial_t f - \mathsf{H}(\nabla f) = 0 \qquad \text{on } [0,\infty)\times\mathbb{R}^d. \tag{A.1}$$

Definition A.1.

(1) A continuous function $f:[0,\infty)\times\mathbb{R}^d\to\mathbb{R}$ is a viscosity subsolution to (2.1) if for every $(t,h)\in(0,\infty)\times\mathbb{R}^d$ and every smooth $\phi:(0,\infty)\times\mathbb{R}^d\to\mathbb{R}$ such that $f-\phi$ has a local maximum at (t,h), we have

$$\big(\partial_t\phi - \mathsf{H}(\nabla\phi)\big)(t,h) \leqslant 0.$$

(2) A continuous function $f:[0,\infty)\times\mathbb{R}^d\to\mathbb{R}$ is a viscosity supersolution to (2.1) if for every $(t,h)\in(0,\infty)\times\mathbb{R}^d$ and every smooth $\phi:(0,\infty)\times\mathbb{R}^d\to\mathbb{R}$ such that $f-\phi$ has a local minimum at (t,h), we have

$$\big(\partial_t\phi - \mathsf{H}(\nabla\phi)\big)(t,h) \geqslant 0.$$

(3) A continuous function $f:[0,\infty)\times\mathbb{R}^d\to\mathbb{R}$ is a viscosity solution to (2.1) if f is both a viscosity subsolution and supersolution.

The main goal of this section is to prove the following proposition.

Proposition A.2. *Let $f : [0,\infty) \times \mathbb{R}^d \to \mathbb{R}$ be Lipschitz and convex. Suppose that f satisfies (A.1) on a dense subset of $(0,\infty) \times \mathbb{R}^d$, and that the initial condition $f(0,\cdot)$ is C^1. Under these conditions, the function f is a viscosity solution to (A.1) with initial condition $f(0,\cdot)$.*

Remark A.3. In Proposition A.2, the assumption that $f(0,\cdot)$ be C^1 is necessary. Indeed, notice for instance that

$$f(t,x) := |x| - t$$

is convex and satisfies

$$\partial_t f + |\nabla f|^2 = 0$$

at every point of differentiability of f. However, since the null function is clearly a solution, the statement that f is also a solution would contradict the maximum principle. Instead, the viscosity solution to this equation with same initial condition is given by the Hopf-Lax formula

$$(t,x) \mapsto \inf_{y\in\mathbb{R}} \left(|y| + \frac{|y-x|^2}{4t} \right) = \begin{cases} \frac{|x|^2}{4t} & \text{if } |x| \leqslant 2t, \\ |x| - t & \text{if } |x| > 2t. \end{cases}$$

Proof of Proposition A.2. Recall the definition of subdifferential in (3.1). We decompose the proof into three steps.

Step 1. We check that f must be a subsolution to (A.1). Let $\phi \in C^\infty((0,\infty)\times\mathbb{R}^d)$, and $(t,x) \in (0,\infty)\times\mathbb{R}^d$ be such that $f - \phi$ has a local maximum at (t,x). We then have

$$f(t',x') - f(t,x) \leqslant (t'-t)\partial_t\phi(t,x) + (x'-x)\cdot\nabla\phi(t,x) + o(|t'-t| + |x'-x|).$$

This implies that the subdifferential $\partial f(t,x)$ is the singleton $\{(\partial_t\phi, \nabla\phi)(t,x)\}$, and thus that f is differentiable at (t,x), with $(\partial_t f, \nabla f)(t,x) = (\partial_t\phi, \nabla\phi)(t,x)$. Using similar arguments as in Remark 3.10, we deduce that

$$\left(\partial_t\phi - \mathsf{H}(\nabla\phi)\right)(t,x) = \left(\partial_t f - \mathsf{H}(\nabla f)\right)(t,x) = 0,$$

as desired.

Step 2. In the next two steps, we show that f is a supersolution to (A.1). Let $(a,p) \in \partial f(t,x)$. In view of Lemma 3.3, it is sufficient to show that

$$a - \mathsf{H}(p) \geqslant 0. \tag{A.2}$$

Since $(a,p) \in \partial f(t,x)$, we have for every $(t',x') \in [0,\infty)\times\mathbb{R}^d$ that

$$f(t',x') \geqslant f(t,x) + (t'-t)a + (x'-x)\cdot p.$$

In particular, the mapping $y \mapsto f(0,y) - y\cdot p$ is bounded from below. In this step, we assume that the infimum

$$\inf_{y\in\mathbb{R}^d} \left(f(0,y) - y\cdot p\right) \tag{A.3}$$

is achieved, and we denote by y a point realizing the infimum. By arguing as in Remark 3.10, we see that there exists $(b,p') \in \partial f(0,y)$ such that $b - \mathsf{H}(p') = 0$. Since $f(0,\cdot)$ is C^1 and y realizes (A.3), we must have $p' = \partial_y f(0,y) = p$, and thus $(b,p) \in \partial f(0,y)$ with $b - \mathsf{H}(p) = 0$.

We next observe that the mapping $g : \lambda \mapsto f\left(\lambda(t,x) + (1-\lambda)(0,y)\right)$ is convex over $[0,1]$. Since $(b,p) \in \partial f(0,y)$, the right derivative of g at 0 satisfies

$$g'_+(0) \geqslant bt + p\cdot(x-y),$$

while the left derivative at 1 satisfies

$$g'_-(1) \leqslant at + p \cdot (x - y).$$

By convexity of g, we must have $g'_+(0) \leqslant g'_-(1)$, and thus $a \geqslant b$. Recalling that $b - \mathsf{H}(p) = 0$, we obtain (A.2), as desired.

Step 3. To conclude, there remains to consider the case when the infimum in (A.3) is not achieved. For every $\varepsilon > 0$, we consider

$$\inf_{y \in \mathbb{R}^d} \left(f(0,y) + \varepsilon |y| - y \cdot p \right).$$

This infimum is achieved at a point $y_\varepsilon \in \mathbb{R}^d$, and

$$|\nabla f(0, y_\varepsilon) - p| \leqslant \varepsilon. \tag{A.4}$$

Moreover,

$$\limsup_{\varepsilon \to 0} \inf_{y \in \mathbb{R}^d} \left(f(0,y) + \varepsilon |y| - y \cdot p \right) = \inf_{y \in \mathbb{R}^d} \left(f(0,y) - y \cdot p \right),$$

and

$$f(0, y_\varepsilon) - y_\varepsilon \cdot p \geqslant \inf_{y \in \mathbb{R}^d} \left(f(0,y) - y \cdot p \right),$$

so that

$$\lim_{\varepsilon \to 0} \varepsilon |y_\varepsilon| = 0. \tag{A.5}$$

Following the argument in Step 2, we can find $b_\varepsilon \in \mathbb{R}$ such that $(b_\varepsilon, \nabla f(0, y_\varepsilon)) \in \partial f(0, y_\varepsilon)$ and $b_\varepsilon - \mathsf{H}(\nabla f(0, y_\varepsilon)) = 0$. Continuing as in Step 2, we then obtain that

$$b_\varepsilon t + \nabla f(0, y_\varepsilon) \cdot (x - y_\varepsilon) \leqslant at + p \cdot (x - y_\varepsilon).$$

Using (A.4) and (A.5), we deduce that, as ε tends to zero,

$$a \geqslant b_\varepsilon + o(1).$$

Recalling that $b_\varepsilon - \mathsf{H}(\nabla f(0, y_\varepsilon)) = 0$, and using again (A.4) and the continuity of H, we obtain (A.2). □

References

[1] J. Barbier, M. Dia, N. Macris, F. Krzakala, T. Lesieur, and L. Zdeborová. Mutual information for symmetric rank-one matrix estimation: A proof of the replica formula. In *Advances in Neural Information Processing Systems (NIPS)*, volume 29, pages 424–432, 2016.

[2] J. Barbier and N. Macris. The adaptive interpolation method: a simple scheme to prove replica formulas in bayesian inference. *Probability Theory and Related Fields*, 174(3-4):1133–1185, 2019.

[3] J. Barbier and N. Macris. The adaptive interpolation method for proving replica formulas. Applications to the Curie-Weiss and Wigner spike models. *Journal of Physics A: Mathematical and Theoretical*, 52(29):294002, 2019.

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

[5] A. Barra, G. Dal Ferraro, and D. Tantari. Mean field spin glasses treated with PDE techniques. *The European Physical Journal B*, 86(7):1–10, 2013.

[6] A. Barra, A. Di Biasio, and F. Guerra. Replica symmetry breaking in mean-field spin glasses through the Hamilton–Jacobi technique. *Journal of Statistical Mechanics: Theory and Experiment*, 2010(09):P09006, 2010.

[7] H.-B. Chen. Hamilton-Jacobi equations for nonsymmetric matrix inference. *arXiv preprint arXiv:2006.05328*, 2020.

[8] H.-B. Chen and J. Xia. Hamilton-Jacobi equations for inference of matrix tensor products. *Ann. Inst. Henri Poincaré Probab. Stat.*, to appear.

[9] G. Genovese and A. Barra. A mechanical approach to mean field spin models. *Journal of Mathematical Physics*, 50(5):053303, 2009.

[10] F. Guerra. Sum rules for the free energy in the mean field spin glass model. *Fields Institute Communications*, 30(11), 2001.
[11] J. Kadmon and S. Ganguli. Statistical mechanics of low-rank tensor decomposition. In *Advances in Neural Information Processing Systems*, pages 8201–8212, 2018.
[12] M. Lelarge and L. Miolane. Fundamental limits of symmetric low-rank matrix estimation. *Probability Theory and Related Fields*, 173(3-4):859–929, 2019.
[13] T. Lesieur, L. Miolane, M. Lelarge, F. Krzakala, and L. Zdeborová. Statistical and computational phase transitions in spiked tensor estimation. In *2017 IEEE International Symposium on Information Theory (ISIT)*, pages 511–515. IEEE, 2017.
[14] C. Luneau, J. Barbier, and N. Macris. Mutual information for low-rank even-order symmetric tensor estimation. *Information and Inference: A Journal of the IMA*, iaaa022, 2020.
[15] C. Luneau, N. Macris, and J. Barbier. High-dimensional rank-one nonsymmetric matrix decomposition: the spherical case. *arXiv preprint arXiv:2004.06975*, 2020.
[16] V. Mayya and G. Reeves. Mutual information in community detection with covariate information and correlated networks. In *2019 57th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 602–607. IEEE, 2019.
[17] L. Miolane. Fundamental limits of low-rank matrix estimation: the non-symmetric case. *arXiv preprint arXiv:1702.00473*, 2017.
[18] J.-C. Mourrat. Hamilton-Jacobi equations for mean-field disordered systems. *Annales Henri Lebesgue*, to appear.
[19] J.-C. Mourrat. Hamilton-Jacobi equations for finite-rank matrix inference. *Annals of Applied Probabaility*, to appear.
[20] J.-C. Mourrat. Parisi's formula is a Hamilton-Jacobi equation in Wasserstein space. *Canadian Journal of Mathematics*, to appear.
[21] J.-C. Mourrat. Nonconvex interactions in mean-field spin glasses. *Probability and Mathematical Physics*, to appear.
[22] J.-C. Mourrat. Free energy upper bound for mean-field vector spin glasses. *arXiv preprint arXiv:2010.09114*, 2020.
[23] J.-C. Mourrat and D. Panchenko. Extending the Parisi formula along a Hamilton-Jacobi equation. *Electronic Journal of Probability*, 25:Paper No. 23, 17, 2020.
[24] G. Reeves. Information-theoretic limits for the matrix tensor product. *IEEE Journal on Selected Areas in Information Theory*, 1(3):777–798, 2020.
[25] G. Reeves, V. Mayya, and A. Volfovsky. The geometry of community detection via the mmse matrix. In *2019 IEEE International Symposium on Information Theory (ISIT)*, pages 400–404. IEEE, 2019.
[26] R. T. Rockafellar. *Convex Analysis*, volume 36. Princeton university press, 1970.
[27] F. Zhang. *Matrix theory*. Universitext. Springer, New York, second edition, 2011.

(H.-B. Chen) Courant Institute of Mathematical Sciences, New York University, New York, New York, USA

E-mail address: hbchen@cims.nyu.edu

(J.-C. Mourrat) Courant Institute of Mathematical Sciences, New York University, New York, New York, USA; CNRS, France

E-mail address: jcm777@nyu.edu

(J. Xia) Department of Mathematics, University of Pennsylvania, Philadelphia, Pennsylvania, USA

E-mail address: xiajiam@sas.upenn.edu