ASYMPTOTIC PROPERTIES OF DISORDERED SYSTEMS

Jiaming Xia

A DISSERTATION

in

Mathematics

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2022

Supervisor of Dissertation

Jian Ding

Professor of Mathematics

Co-Supervisor of Dissertation

Robin Pemantle

Professor of Mathematics

Graduate Group Chairperson

Ron Donagi, Professor of Mathematics

Dissertation Committee

Xin Sun, Assistant Professor of Mathematics

ABSTRACT

ASYMPTOTIC PROPERTIES OF DISORDERED SYSTEMS

Jiaming Xia

Jian Ding

Robin Pemantle

This thesis considers asymptotic behaviors of high-dimensional disordered systems, including Ising model and mean-field spin glass models. We first study the decay rate of correlations in the two-dimensional random field Ising model (RFIM). Second, we study the limit free energy of disordered systems.

For RFIM, we are interested in the two-dimensional case where the external field is of i.i.d centered Gaussian variables. We show that under nonnegative temperature, the effect of boundary conditions on the magnetization in a finite box decays exponentially in the side length of the box.

On the side of mean-field models, we use the Hamilton-Jacobi equation (HJE) approach, initiated by Jean-Christophe Mourrat, to characterize limiting free energy in many models from statistical inference problems and mean-field spin glass models. We now investigate infinite-dimensional models including many spin glass models and inference problems where the rank of the signal matrix increases as n is sent to infinity. We give an intrinsic meaning to the Hamilton–Jacobi equation arising from mean-field spin glass models in the viscosity sense, and establish the corresponding well-posedness. This will shed more light on the mysterious Parisi formula as the limit of free energy in the Sherrington–Kirkpatrick model.

ACKNOWLEDGEMENT

I would like to express my gratitude to my primary advisor, Professor Jian Ding, who guided me through the RFIM project and selflessly offered deep insights into the study. I am also very grateful to my advisor, Professor Robin Pemantle, who has been generous with his valuable professional advice.

I am fortunate to have a supportive husband, Hong-Bin Chen, who is also a wonderful collaborator. His impeccable work ethic always inspires me and his extremely dry sense of humor keeps me a moderately happy person most of the time.

Special thanks to my mother, Mrs. Lin Wang, and my father, Mr. Bosong Xia, without whom I could not have achieved anything. Both of my parents have been my best role models since I was a child. One big motivation in my life is to make my family proud.

Finally, I greatly appreciate the companionship of my cats, Baoyu and Yenhsi, who adopted me in 2020 and 2021, respectively. I believe they unconditionally love me.

To all of my dear human friends, I equally thank you for your support throughout these years. The list is endless and I doubt you will ever read my thesis, so I might just skip that. Please do not feel betrayed. I will text you sooner or later. I love you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

CHAPTER 1

Introduction

1.1. Random Field Ising Model

Ising model is a mathematical model of ferromagnetism, the two-dimensional version of which can be described as follows. For some integer N, let Λ_N be a centered box with side length $2N$. On each vertex $v \in \Lambda_N$ of the square lattice is placed a magnetic spin σ_v which takes value $+1$ or -1. The collection of states $\{\sigma_v\}_{v\in\Lambda_N}$ is called a configuration. Additionally, we are given a collection $\{h_v\}_{v\in\Lambda_N}$ of independent Gaussian random variables with zero mean and variance ε^2 at each vertex, serving as a random external field imposing on each vertex. Here, we are dealing with Gaussian disorder $\mathcal{N}(0,\varepsilon^2)$ where $\varepsilon > 0$ can be arbitrarily small so that disorder is arbitrarily weak.

We are interested in the effect of spins in the boundary $\partial\Lambda_N$ on the spin σ_o at the center as N increases. For each configuration σ, its energy, also called the RFIM Hamiltonian, with external field $\{h_v\}_{v\in\Lambda_N}$ and plus (respectively, minus) boundary condition, where spins on $\partial\Lambda_N$ are all $+1$ (respectively, -1), is given by

$$H^{\pm}(\sigma) = -\left(\sum_{u\sim v;\ u,v\in\Lambda_N} \sigma_u\sigma_v \pm \sum_{u\sim v;\ u\in\Lambda_N;\ v\in\partial\Lambda_N} \sigma_u + \sum_{u\in\Lambda_N}\sigma_u h_u\right),$$

where $u \sim v$ means u and v are neighboring. Note that the first sum accounts for the neighboring interaction: neighboring spins with the same sign contribute lower energy. The second sum is the effect from the boundary and third the external field. The Ising measure, quenched on the external field $\{h_v\}_{v\in\Lambda_N}$ with plus (respectively, minus) boundary condition, is defined such that for all $\sigma \in \{-1,1\}^{\Lambda_N}$,

$$\mu^{\pm}(\sigma) = \frac{e^{-\beta H^{\pm}(\sigma)}}{Z^{\pm}}, \text{ where } Z^{\pm} = \Sigma_{\sigma'\in\{-1,1\}^{\Lambda_N}} e^{-\beta H^{\pm}(\sigma')}.$$

There are two sources of randomness here. To clarify, we use notations $\mathbb{P}$ and $\mathbb{E}$ for the randomness with respect to the external field $\{h_v\}_{v\in\Lambda_N}$; and we denote the Ising measures and their expectations by $\mu^{\pm}$ and $\langle\cdot\rangle_{\mu^{\pm}}$.

Under zero temperature, the Ising measure is supported on the unique configuration, also known as the *ground state*, that minimizes the corresponding RFIM Hamiltonian. Denote by $\sigma^{\Lambda_N,\pm}$ the ground state with plus (respectively, minus) boundary condition under zero temperature. Therefore, in this case the only randomness comes from the external field. Our zero temperature result is the following theorem.

Theorem 1.1.1. *For any $\varepsilon > 0$, there is $c_\varepsilon > 0$ such that $\mathbb{P}\{\sigma_o^{\Lambda_N,+} \neq \sigma_o^{\Lambda_N,-}\} \leqslant c_\varepsilon^{-1} e^{-c_\varepsilon N}$ for all $N \geqslant 1$.*

By first proving in this simplified case, some of the key ideas, including the crucial application of [1], can be more transparent. To give a brief sketch of the proof, we reformulate Theorem 1.1.1. For $v \in \Lambda_N$, we define

$$\xi_v^{\Lambda_N} = \begin{cases} +, & \text{if } \sigma_v^{\Lambda_N,+} = \sigma_v^{\Lambda_N,-} = 1\,, \\ -, & \text{if } \sigma_v^{\Lambda_N,+} = \sigma_v^{\Lambda_N,-} = -1\,, \\ \mathsf{0}, & \text{if } \sigma_v^{\Lambda_N,+} = 1 \text{ and } \sigma_v^{\Lambda_N,-} = -1\,. \end{cases}$$

By monotonicity (c.f. [3, Section 2.2]), $\xi_v^{\Lambda_N}$ is well-defined for all $v \in \Lambda_N$. Theorem 1.1.1 can be restated as

$$m_N \leqslant c^{-1} e^{-cN} \text{ for } c = c(\varepsilon) > 0, \text{ where } m_N \triangleq \mathbb{P}(\xi_o^{\Lambda_N} = \mathsf{0})\,. \tag{1.1.1}$$

For any $A \subseteq \mathbb{Z}^2$, we can analogously define ξ^A by replacing Λ_N with A. Let $\mathcal{C}^A = \{v \in A : \xi_v^A = \mathsf{0}\}$, meaning that $\mathcal{C}^A$ is the collection of *disagreements*.

In particular, m_N is decreasing in N, so we only consider $N = 2^n$ for $n \geqslant 1$. Clearly, for any $v \in \mathcal{C}^A$, there exists a path in $\mathcal{C}^A$ joining v and ∂A, suggesting percolation properties of $\mathcal{C}^A$. Indeed, a key step in our proof is the following proposition on the lower bound on the length exponent for geodesics in $\mathcal{C}^{\Lambda_N}$. For any $A \subseteq \mathbb{Z}^2$, we denote by $d_A(\cdot,\cdot)$ the intrinsic distance on A. Let $d_A(A_1, A_2) = \min_{x\in A_1\cap A, y\in A_2\cap A} d_A(x,y)$.

Proposition. *There exist $\alpha = \alpha(\varepsilon) > 1$, $\kappa = \kappa(\varepsilon) > 0$ such that for all $N \geqslant 1$*

$$\mathbb{P}(d_{\mathcal{C}^{\Lambda_N}}(\partial\Lambda_{N/4}, \partial\Lambda_{N/2}) \leqslant N^\alpha) \leqslant \kappa^{-1} e^{-N^\kappa}.$$

The proof of this proposition relies on [1], which requires the next lemma. For any rectangle $A \subseteq \mathbb{R}^2$, let ℓ_A be the length of the longer side and let A^{Large} be the square box concentric with A, of side length $32\ell_A$ and with sides parallel to axes. For a set $\mathcal{C} \subseteq \mathbb{Z}^2$, we use $\text{Cross}(A, \mathcal{C})$ to denote the event that there exists a path $v_0, \ldots, v_k \in A \cap \mathcal{C}$ connecting the two shorter sides of A (that is, v_0, v_k are of ℓ_∞-distances less than 1 respectively from the two shorter sides of A).

Lemma. *There exists ℓ_0 and $\delta > 0$ such that the following holds for any $N \geqslant 1$. For any $k \geqslant 1$ and any rectangles $A_1, \ldots, A_k \subseteq \{v \in \mathbb{R}^2 : |v|_\infty \leqslant N/2\}$, each with the ratio between the lengths of the longer and shorter sides at least 100, such that*

- *$\ell_0 \leqslant \ell_{A_i} \leqslant N/32$ for all $1 \leqslant i \leqslant k$ and*
- *$A_1^{\text{Large}}, \ldots, A_k^{\text{Large}}$ are disjoint,*

we have $\mathbb{P}(\cap_{i=1}^k \text{Cross}(A_i, \mathcal{C}^{\Lambda_N})) \leqslant (1-\delta)^k$.

Although the authors of [1] treated random curves in $\mathbb{R}^2$, the main capacity analysis can be copied in the discrete case, and the connection between the capacity and the box-counting dimension is straightforward (c.f. [50, Lemma 2.3]). With the lemma above, we can apply [1, Theorem 1.3] to deduce that for some $\alpha = \alpha(\varepsilon) > 1$,

$$\mathbb{P}(d_{\mathcal{C}^{\Lambda_N}}(\partial\Lambda_{N/4}, \partial\Lambda_{N/2}) \leqslant N^\alpha) \to 0 \text{ as } N \to \infty.$$

Then by a standard percolation argument to be shown later, we can enhance the probability decay to the exponential decay, proving (1.1.1).

In the case of positive temperatures, our proof follows the framework as in the zero temperature case. However, major obstacles emerge due to the randomness of Ising measures under positive temperatures. In order to overcome these obstacles and to avoid further complications, we need new ideas to delicately treat the couplings of Ising measures. The following theorem is our result for all nonnegative temperatures.

Theorem. *For any* $\varepsilon > 0$*,* $T \in [0, \infty)$*, there exists* $c = c(\varepsilon, T) > 0$ *such that*

$$\mathbb{E}(\langle \sigma_o \rangle_{\Lambda_N,+} - \langle \sigma_o \rangle_{\Lambda_N,-}) \leqslant c^{-1} e^{-cN} \quad \text{for all } N \geqslant 1,$$

where $\langle \cdot \rangle_{\Lambda_N,\pm}$ *denotes the expectation with respect to the Ising measures.*

In other words, the effect on the center spin from the boundary N away decays *exponentially* as $N \to \infty$. The result can be understood in the context of the general Imry-Ma [78] phenomenon stating that introducing arbitrarily weak disorder rounds off any first order phase transition in two-dimensional systems. It was previously known [4, 5] that such effect decays. The decay rate was previously proven to be polynomial [35, 3], and exponential for large disorder ε [26, 64, 31]. Our result rigorously settled the long-standing debate [70, 29, 51] on whether the decay rate could be polynomial when ε is small.

In spite of that our proof is seemingly related to the Mandelbrot percolation analogy presented in [3, Appendix B], our proof method is different from all of [5, 35, 3, 26, 64, 31]. The works [4, 5] treated a wide class of distributions for disorder, while [35, 3] and our work deal with Gaussian disorder. The main features of Gaussian distributions used in our work include the simple formula for the change of measure and linear decompositions for Gaussian process. In addition, the analysis in [5, 3] extends to the case with finite-range interactions. Though we expect our framework to be useful in the finite-range case, the lack of planar duality presents non-trivial obstacles in extending the framework.

During submission of our work, a paper [2] proving the same result was completed. Its proof was inspired by our zero temperature proof, using [1] as a crucial tool. In terms of the basic intuition, both proofs harness the fluctuation of the sum of the random field in a box to compare to the influence of the boundary condition, which could trace back to [4, 5], and both apply [1] to disagreement percolation in a crucial manner. However, the two approaches are also very different in at least two important aspects: (1) we employ first moment analysis by exploiting perturbations of the random field, while [2], similar to [3], relies on concentration/anti-concentration type of analysis, which uses second-moment computations; (2) at positive temperatures, we employ a certain monotone coupling between Ising measures with different boundary conditions, whereas inspired by [107, 115], the authors of [2] extend the Ising model continuously into the metric graph allowing them to study spin correlations via disagreement percolation for two *independent* samples.

For 2D RFIM, one can further determine the *correlation length*, which is the critical size of the box where the influence of random field is comparable to that of the boundary condition. The correlation length has been found in [53] at zero temperature and an upper bound at positive temperatures. A recent work [55] found the lower bound correlation length at low temperatures that scales as in [53] for 2D RFIM. Future research interests are in studying other spin glass models and their correlation length in dimension two and higher.

1.2. Free energy of mean-field disordered systems

Unlike the Ising model where the interaction is between neighboring spins, a mean-field model averages over interactions with every spin. Many models in spin glasses and statistical inference fall into this category.

A simple mean-field model is the statistical inference problem of rank-one symmetric matrix estimation. Let $\mathbb{R}^n$-valued random vector X be an unknown signal and $n \times n$ matrix W be some additional noise. For $t > 0$ interpreted as the signal-to-noise ratio (SNR), we observe

$$Y = \sqrt{\frac{t}{n}} X X^\intercal + W \tag{1.2.1}$$

where $\frac{1}{\sqrt{n}}$ is a proper scaling. The inference task is to recover X from the noisy observation Y.

We assume that W consists of i.i.d. standard Gaussian entries. By Bayes' law, the distribution of X conditioned on Y is a Gibbs measure proportional to $e^{H_n(t,x;Y)}P_X(\mathrm{d}x)$ where P_X is the law of X and $H_n(t,x;Y)$ is called the Hamiltonian. We are interested in the limit, as $n \to \infty$, of the average of the free energy

$$F_n(t) = \frac{1}{n}\log \int e^{H(t,x;Y)}P_X(\mathrm{d}x), \tag{1.2.2}$$

and we use the notation $\overline{F}_n(t) = \mathbb{E}F_n(t)$. The Hamilton–Jacobi equation (HJE) approach initiated in [95] starts by enriching the model via adding an simple linear observation parameterized by $h \geqslant 0$. The associated enriched free energy $\overline{F}_n(t,h)$ is related to the original one by $\overline{F}_n(t,0) = \overline{F}_n(t)$.

One main reason of studying this topic is because the limit of the free energy is related, via a simple additive relation, to an important information-theoretical quantity: the mutual information $I(X;Y)$, heuristically measuring the dependence between the unknown signal X and the observation Y. Computing the limit of $\frac{1}{n}I(X;Y)$ as $n \to \infty$ allows one to determine the critical value t_c of SNR beyond which the inference task is theoretically impossible. The definition of the critical value, also known as the information theoretic threshold, is given by

$$t_c = \inf\left\{t > 0 : \lim_{n\to\infty}\frac{I(X;Y)}{n} \text{ is analytic in } (t,\infty)\right\}.$$

To see the relation between the free energy and the mutual information, we first define

$$I(X;Y) = \mathbb{E}\left[\log\frac{P_{X,Y}(X,Y)}{P_X(X)P_Y(Y)}\right],$$

where $P_{X,Y}(X,Y)$, $P_X(X)$ and $P_Y(Y)$ are the joint law of (X,Y), the law of X and the law of Y, respectively. We also assume that $P_{X,Y}(X,Y)$, $P_X(X)$ and $P_Y(Y)$ are absolutely continuous with respect to the Lebesgue measure and can be identified with their densities. By setting $P_{Y|X}(y|x) = P_{X,Y}(x,y)/P_X(x)$, we have $P_Y(y) = \int P_{Y|X}(y|x)P_X(x)\mathrm{d}x$ and

$$\frac{I(X;Y)}{n} = \frac{1}{n}\mathbb{E}\left[\log P_{Y|X}(Y|X)\right] - \frac{1}{n}\mathbb{E}\left[\log \int P_{Y|X}(Y|x)P_X(x)dx\right]. \tag{1.2.3}$$

Recall that W consists of i.i.d. standard Gaussian entries, so we can compute that

$$P_{Y|X}(y|x) = \frac{1}{(2\pi)^{\frac{n^2}{2}}} e^{-\frac{1}{2}|\sqrt{\frac{t}{n}}xx^\intercal - y|^2}.$$

Then, the first term on the right hand side of (1.2.3) is $-\frac{n}{2}(1+\log 2\pi)$ and the second term is

$$-\mathbb{E}F_n + \frac{1}{2n}\mathbb{E}|Y|^2 + \frac{n}{2}\log 2\pi = -\mathbb{E}F_n + \frac{t}{2n^2}\mathbb{E}|XX^\intercal|^2 + \frac{n}{2}(1+\log 2\pi),$$

where F_n is the free energy as in (1.2.2) with Hamiltonian

$$H_n(t,x;Y) = Y \cdot \left(\sqrt{\frac{t}{n}}xx^\intercal\right) - \frac{t}{2n}|xx^\intercal|^2.$$

We therefore have the relation

$$\frac{I(X;Y)}{n} = -\mathbb{E}F_n + \frac{t}{2n^2}\mathbb{E}|XX^\intercal|,$$

implying that it suffices to identify the limit of $\mathbb{E}F_n$ as $n \to \infty$ to understand the asymptotics of $I(X:Y)$.

We start with the high-dimensional limit of the free energy of finite-rank matrix tensor products. Fix $K, p \in \mathbb{N}$ and let P_n^X be the law of $X \in \mathbb{R}^{n\times K}$ where $n \in \mathbb{N}$. For any fixed

$L \in \mathbb{N}$, we observe

$$Y = \sqrt{\frac{2t}{n^{p-1}}} X^{\otimes p} A + W \in \mathbb{R}^{n^p \times L}, \tag{1.2.4}$$

where $t \geqslant 0$ is the SNR; $\otimes$ is the Kronecker product; $A \in \mathbb{R}^{K^p \times L}$; and $W \in \mathbb{R}^{n^p \times L}$ is the noise matrix consisting of i.i.d. standard Gaussian entries.

We briefly discuss the generality of this model and how it relates to other inference matrix product models. The models of the second order products are widely studied. The spiked Wishart model is given by $Y = \sqrt{\frac{2t}{N}} X_1 X_2^\intercal + W$, which is investigated in works including [90, 14, 12, 79, 86, 36]. When $X_1 = X_2$, this becomes the spiked Wigner model, studied in [82, 52, 95, 94]. A generalization of these spiked matrix models can be seen in the study of community detection problems and the stochastic block models. The community detection problem in certain settings is asymptotically equivalent to $Y = \sqrt{\frac{2t}{N}} X B X^\intercal + W$ where B is the community interaction matrix (see [104]). More generally, the community detection with several correlated networks is asymptotically equivalent to the multiview spiked matrix model $Y_l = \sqrt{\frac{2t}{N}} X B_l X^\intercal + W_l$ for $l = 1, 2, \ldots, L$ where each B_l models one network (see [87, 88]). All the examples of second order models can be represented in $Y = \sqrt{\frac{2t}{N}} X^{\otimes 2} \sqrt{S} + W$ where S is positive semidefinite. They can be seen as special cases of (1.2.4) for $p = 2$.

By Bayes' rule, we can compute in a straightforward fashion that the original Hamiltonian H_n^o and original free energy F_n^o of (1.2.4). By introducing an additional variable h, we enrich the Hamiltonian and its corresponding free energy to be $H_n(t,h)$ and $F_n(t,h)$, respectively. The goal is to compare $\lim_{n\to\infty} \mathbb{E} F_n(t,h)$ with the solution of the HJE

$$\left(\partial_t f - \mathsf{H}(\nabla f)\right)(t,h) = 0, \quad \forall (t,h) \in \mathbb{R}_+ \times \mathbb{S}_+^K,$$

where

$$\mathsf{H}(q) := (AA^\intercal) \cdot q^{\otimes p}, \quad \forall q \in \mathbb{S}^K_+,$$

and $\mathbb{S}^K_+$ is the set of $K \times K$ positive semi-definite matrices.

In [39], we bound the limit from above by the unique solution to the HJE displayed above. If in addition we assume H is convex by choosing particular A and p, then we can identify the limit with the solution.

Theorem. *Let $p \in \mathbb{N}$. Suppose that $F_n(0,\cdot)$ pointwise converges to some function ψ and assume concentration of F_n Then for any H of the form above, there is a unique Lipschitz viscosity solution f to the HJE with $f(0,\cdot) = \psi$, and*

$$\limsup_{N\to\infty} \overline{F}_N(t,h) \leqslant f(t,h), \quad \forall (t,h) \in \mathbb{R}_+ \times \mathbb{S}^K_+.$$

If H is convex, then a corresponding lower bound holds and thus we have the following identity

$$\lim_{n\to\infty} F_n(t,h) = f(t,h), \quad \forall (t,h) \in \mathbb{R}_+ \times \mathbb{S}^K_+.$$

Later in [37], we improve the result that for any nonlinearity and any order p, we can identify the limit with a variational formula. To be more precise, our result is the following.

Theorem. *Suppose that $\overline{F}_n(0,\cdot)$ pointwise converges to some C^1 function ψ and assume local uniform concentration of F_n. Then for every $(t,h) \in [0,\infty) \times \mathbb{S}^K_+$, the limit of the free energy can be written in terms of a variational formula, namely*

$$\lim_{n\to\infty} \overline{F}_n(t,h) = \sup_{h'' \in \mathbb{S}^K_+} \inf_{h' \in \mathbb{S}^K_+} \left\{ h'' \cdot (h - h') + \psi(h') + t\mathsf{H}(h'') \right\}.$$

Note that the assumption that ψ is of class C^1 can be omitted for certain nonlinearity H, such as convex H. However, this assumption may be required if we consider arbitrary A and p.

Another line of research using the classical interpolation method and the new adaptive interpolation method introduced in [12, 13] includes [10, 82, 12, 90, 14, 79, 86, 82, 87, 104, 103, 83, 85, 65]. Our approach appears to be more versatile, as we are able to treat the most general setting in the inference of matrix tensor products [37], and even models with multiple layers [40].

More specifically, there are two major advantages of the HJE approach. The first advantage is that it only requires one side bound, while the interpolation method requires convexity of the nonlinearity and concentration property. Another advantage is that the HJE approach works as a black box, meaning that if we feed in the convergence and concentration of $\overline{F}_N(0,\cdot)$, it spits out $\overline{F}_N$ converging to the unique solution to the HJE. This feature of black box enables us to apply the HJE approach to multi-layer generalized linear models.

In order to study HJE on more general spaces, we need a convex analysis result on Fenchel–Moreau identities [38]. On the Hilbert space H with inner product $\langle\cdot,\cdot\rangle$, the classical Fenchel–Moreau identity is $f = f^{**}$ for convex $f : \mathsf{H} \to (-\infty,\infty]$ satisfying a few additional regularity conditions. The convex conjugate is given by

$$f^*(x) = \sup_{y\in\mathsf{H}}\{\langle y, x\rangle - f(y)\}, \quad \forall x \in \mathsf{H},$$

where it is worth noting that the supremum is taken over the entire space H.

On the other hand, it is well-known (c.f. [105, Theorem 12.4]) that on the cone $[0,\infty)^d$ in $\mathbb{R}^d$, if $f : [0,\infty)^d \to (-\infty,\infty]$ is convex with extra assumptions and nondecreasing with respect to the partial order induced by the cone, namely

$$f(x) \geqslant f(y), \quad \text{if } x - y \in [0,\infty)^d,$$

then we have $f = f^{**}$. The monotone conjugate is defined by

$$f^*(x) = \sup_{y\in[0,\infty)^d} \{\langle y, x\rangle - f(y)\}, \quad \forall x \in [0,\infty)^d,$$

where the inner product is the standard one in $\mathbb{R}^d$. Compared with the convex conjugate, the supremum in this monotone conjugate is taken over the cone. In [36], a version of the Fenchel–Moreau identity on $\mathbb{S}^n_+$ is needed to verify that the unique solution to a certain HJE with spatial variables in $\mathbb{S}^n_+$ admits a variational formula. On $\mathbb{S}^n_+$, [36, Proposition B.1] proves that $f = f^{**}$ holds if $f : \mathbb{S}^n_+ \to (-\infty, \infty]$ is convex with some usual regularity assumptions and is nondecreasing in the sense that

$$f(x) \geqslant f(y), \quad \text{if } x - y \in \mathbb{S}^n_+.$$

Accordingly, here $*$ stands for the monotone conjugate with respect to $\mathbb{S}^n_+$ given by

$$f^*(x) = \sup_{y\in\mathbb{S}^n_+} \{\langle y, x\rangle - f(y)\}, \quad \forall x \in \mathbb{S}^n_+.$$

In this case, the inner product is the Frobenius inner product for matrices and $\mathbb{S}^n_+$ can be viewed as a cone in $\mathbb{S}^n$.

It is thus natural to pursue a generalization to an arbitrary cone $\mathcal{C}$ in a Hilbert space H. In other words, we want to show for proper, lower semicontinuous and convex $f : \mathcal{C} \to (-\infty, \infty]$ which is also nondecreasing in the sense that

$$f(x) \geqslant f(y), \quad \text{if } x - y \in \mathcal{C},$$

the identity $f = f^{**}$ holds. Here $*$ stands for

$$f^*(y) = \sup_{z\in\mathcal{C}}\{\langle z, y\rangle - f(z)\}, \quad \forall y \in \mathcal{C}^\vee, \quad \text{and}$$
$$f^{**}(x) = \sup_{y\in\mathcal{C}^\vee}\{\langle y, x\rangle - f^*(y)\}, \quad \forall x \in \mathcal{C},$$

where $\mathcal{C}^\vee$ is the dual cone of $\mathcal{C}$.

The generality pursued in our work is motivated by the study of HJE arising in mean-field disordered systems [95, 94, 92, 96, 93, 36], where the solution is defined on a set, which can often be identified with a cone, and is expected to be nondecreasing with respect to the cone.

With this useful tool in hand, we go back to apply the HJE approach to more general spaces. Now let us briefly describe the multi-layer generalized linear model. For $n \in \mathbb{N}$, fix any $L \in \mathbb{N}$ as the number of layers. For each $l \in \{0, 1, 2, \ldots, L\}$, let $n_l = n_l(n) \in \mathbb{N}$ be the dimension of the signal at the l-th layer. We assume that $n_0 = n$ and $\lim_{n\to\infty} \frac{n_l}{n} = \alpha_l > 0$, for some $\alpha_l > 0$.

Starting with $X^{(0)} = X$ where $X \in \mathbb{R}^n$ is the original signal with law P_X, we iteratively define, for each $l \in \{1, \ldots, L\}$,

$$X_j^{(l)} = \varphi_l\left(\frac{1}{\sqrt{n_{l-1}}}\sum_{k=1}^{n_{l-1}} \Phi_{jk}^{(l)} X_k^{(l-1)},\ A_j^{(l)}\right), \quad \forall 1 \leqslant j \leqslant n_l,$$

where each φ_l is measurable for some fixed $k_l \in \mathbb{N}$ independent of n; $(A_j^{(l)})_{1\leqslant j\leqslant n_l}$ is a sequence of independent $\mathbb{R}^{k_l}$-valued random vectors with law $P_{A^{(l)}}$; and each $\Phi^{(l)}$ is a random matrix consisting of independent standard Gaussians.

For $\beta \geqslant 0$, the observable is given by

$$Y^\circ = \sqrt{\beta} X^{(L)} + Z,$$

where Z is an n_L-dimensional standard Gaussian vector. The inference task is to recover X based on the knowledge of Y°, $(\varphi_l)_{1\leqslant l\leqslant L}$ and $(\Phi^{(l)})_{1\leqslant l\leqslant L}$. Once again we can compute its free energy $F^o_{\beta,L,n}$ by Bayes' rule straightforwardly.

To state the main result of this model, we further need to define that for every $l \in \{0,1,\ldots,L\}$ and $n \in \mathbb{N}$,

$$\rho_{l,n} = \frac{1}{n_l(n)}\mathbb{E}\left|X^{(l)}\right|^2.$$

It can be shown that the following limit exists $\lim_{n\to\infty}\rho_{l,n} = \rho_l$ for some $\rho_l > 0$. Set

$$\Psi_0(r) = \mathbb{E}\log\int_{\mathbb{R}} e^{rX_1x_1+\sqrt{r}Z_1'x_1-\frac{r}{2}|x_1|^2}\mathrm{d}P_{X_1}(x_1), \quad \forall r\in\mathbb{R}_+,$$

where Z_1' is standard Gaussian. Then for every $l\in\{1,\ldots,L\}$, $\rho \geqslant 0$ and $h=(h_1,h_2)\in[0,\rho]\times\mathbb{R}_+$, set

$$\begin{aligned}
&\Psi_l(h;\rho)\\
&=\mathbb{E}\log\\
&\quad\int\widetilde{\mathcal{P}}_{h_2,l}\left(\sqrt{h_2}\varphi_l\left(\sqrt{h_1}V_1+\sqrt{\rho-h_1}W_1,A_1^{(l)}\right)+Z_1\middle|\sqrt{h_1}V_1+\sqrt{\rho-h_1}w\right)\mathrm{d}P_{W_1}(w),
\end{aligned}$$

where V_1, W_1, Z_1 are independent standard Gaussians and

$$\widetilde{\mathcal{P}}_{h_2,l}(y|z) = \int_{\mathbb{R}^{k_l}} e^{-\frac{1}{2}|y-\sqrt{h_2}\varphi_l(z,a_1^{(l)})|^2}\mathrm{d}P_{A_1^{(l)}}\left(a_1^{(l)}\right), \quad \forall y,z\in\mathbb{R}.$$

We can now state our main result, in which we identify the limit of the free energy with a variational formula.

Theorem. *Under mild assumptions, it holds that*

$$\lim_{n\to\infty}\mathbb{E}F^\circ_{\beta,L,n} = \sup_{z^{(L)}}\inf_{y^{(L)}}\sup_{z^{(L-1)}}\inf_{y^{(L-1)}}\cdots\sup_{z^{(1)}}\inf_{y^{(1)}}\phi_L\left(\beta;y^{(1)},\cdots,y^{(L)};z^{(1)},\cdots,z^{(L)}\right)$$

where $\sup_{z^{(l)}}$ *is taken over* $z^{(l)} \in \mathbb{R}_+ \times [0, \frac{\alpha_{l-1}\rho_{l-1}}{2}]$, $\inf_{y^{(l)}}$ *is taken over* $y^{(l)} \in [0, \rho_{l-1}] \times \mathbb{R}_+$, *and*

$$\begin{aligned}
&\phi_L\left(\beta; y^{(1)}, \cdots, y^{(L)}; z^{(1)}, \cdots, z^{(L)}\right) \\
&\qquad = \alpha_L \Psi_L\left(y_1^{(L)}, \beta; \rho_{L-1}\right) + \sum_{l=1}^{L-1} \alpha_l \Psi_l\left(y_1^{(l)}, y_2^{(l+1)}; \rho_{l-1}\right) + \Psi_0\left(y_2^{(1)}\right) \\
&\qquad\quad + \sum_{l=1}^{L}\left(-y^{(l)} \cdot z^{(l)} + \frac{2}{\alpha_{l-1}} z_1^{(l)} z_2^{(l)}\right) + \sum_{l=2}^{L} \frac{\alpha_{l-1}}{2}\left(1 + \rho_{l-1} y_2^{(l)}\right).
\end{aligned}$$

Now, we study the limit free energy of mean-field spin glass models. The goals are first giving it an intrinsic meaning and then establishing the well-posedness. We interpret the inverse temperature t as the temporal variable and enrich the model by introducing a random magnetic field with a parameter ϱ as the spacial variable. As before, we want to compare the enriched free energy with solutions to a certain Cauchy problem of a HJE.

The equations in question are originally defined on the set of monotone probability measures. A probability measure μ on $\mathbb{S}_+^K$, which is now seen as a cone, is said to be *monotone*, if

$$\mathbb{P}\left\{a \cdot X < a \cdot X' \text{ and } b \cdot X > b \cdot X'\right\} = 0, \quad \forall a, b \in \mathbb{S}_+^K,$$

where X and X' are i.i.d. with the law μ. Note that in the case $K = 1$, every probability measure on $\mathbb{R}_+$ is monotone. For fixed $K \in \mathbb{N}$, let $\mathcal{P}^{\nearrow}$ be the set of monotone probability measures on $\mathbb{S}_+^K$.

When ξ is convex, the solution is defined by a version of the Hopf–Lax formula in [92, 98], which proves it equivalent to the Parisi's formula proposed in [101] and rigorously verified in [72, 111] (see also [100, 99, 112, 113]). When the solution is defined as limits of finite-dimensional approximations as in [96, 93], the solution is shown to be an upper bound for the limiting free energy of models in a wide class.

The first notion of solutions has a *ad hoc* nature, whereas the second notion manifests its *extrinsic* nature. This motivates us to seek an *intrinsic* definition of solutions.

Formally, the equation is of the following form:

$$\left(\partial_t f - \int \xi(\partial_\varrho f)\mathrm{d}\varrho\right)(t,\varrho) = 0, \quad \text{on } \mathbb{R}_+ \times \mathcal{P}^{\nearrow},$$

where ξ is $\mathbb{R}$-valued on $\mathbb{R}^{K\times K}$. Our result can be formally stated as

Theorem. *Under certain admissible assumptions on ξ and the initial condition ψ in mean-field spin glass models, there is a unique viscosity solution f of the Cauchy problem of the HJE displayed above. Moreover,*

1. *f is the limit of viscosity solutions of finite-dimensional approximations of the HJE;*
2. *f is given by a Hopf–Lax formula if ξ is convex on $\mathbb{S}^K_+$;*
3. *f is given by a Hopf formula if ψ is convex.*

We start with making sense of $\partial_\varrho f$. We restrict $\mathcal{P}^{\nearrow}$ to $\mathcal{P}^{\nearrow}_2$, the set of monotone measures with finite second moments, and equipped with the 2-Wasserstein metric. Heuristically, $\partial_\varrho f(t,\varrho)$ describes the asymptotic behavior of $f(t,\vartheta) - f(t,\varrho)$ as ϑ tends to ϱ in the Wasserstein metric. Fortunately, $\mathcal{P}^{\nearrow}_2$ can be isometrically embedded onto a closed convex cone in an L^2 space that has empty interior but generates the L^2 space. Via this isometry, $\partial_\varrho f$ can be understood in the sense of the Fréchet derivative. Therefore, we can interpret the HJE above as a special case of the following HJE

$$\partial_t f - \mathsf{H}(\nabla f) = 0, \quad \text{on } \mathbb{R}_+ \times \mathcal{C},$$

where $\mathcal{C}$ is a closed convex cone in a separable Hilbert space $\mathcal{H}$, and H is a general nonlinearity.

One obstacle comes from the lack of local compactness in infinite dimensions. Another important issue is to figure out a suitable boundary condition. To treat these problems, we exploit the fact that H is nondecreasing along the dual cone of $\mathcal{C}$ in the spin glass setting. Given this condition, we do not need to prescribe any additional condition (e.g. Neumann or Dirichlet) on the boundary, and thus only need the HJE to be satisfied.

CHAPTER 2

Exponential decay of correlations in the two-dimensional random field Ising model

This chapter is essentially borrowed from [54], joint with Jian Ding.

Abstract. We study the random field Ising model on $\mathbb{Z}^2$ where the external field is given by i.i.d. Gaussian variables with mean zero and positive variance. We show that the effect of boundary conditions on the magnetization in a finite box decays exponentially in the distance to the boundary.

2.1. Introduction

For $v \in \mathbb{Z}^2$, let h_v be i.i.d. Gaussian variables with mean zero and variance $\varepsilon^2 > 0$. We consider the random field Ising model (RFIM) with external field $\{h_v : v \in \mathbb{Z}^2\}$ at temperature $T = 1/\beta \in [0, \infty)$. For $N \geqslant 1$, let $\Lambda_N = \{v \in \mathbb{Z}^2 : |v|_\infty \leqslant N\}$ be a box in $\mathbb{Z}^2$ centered at the origin o and of side length $2N$. For any set $A \subseteq \mathbb{Z}^2$, define $\partial A = \{v \in \mathbb{Z}^2 \setminus A : u \sim v \text{ for some } u \in A\}$ (where $u \sim v$ if $|u - v|_1 = 1$). The RFIM Hamiltonian $H^{\Lambda_N, \pm}$ on the configuration space $\{-1, 1\}^{\Lambda_N}$ with plus (respectively, minus) boundary condition and external field $\{h_v : v \in \Lambda_N\}$ is defined to be

$$H^{\Lambda_N,\pm}(\sigma) = -\Big(\sum_{u \sim v, u, v \in \Lambda_N} \sigma_u \sigma_v \pm \sum_{u \sim v, u \in \Lambda_N, v \in \partial \Lambda_N} \sigma_u + \sum_{u \in \Lambda_N} \sigma_u h_u \Big) \text{ for } \sigma \in \{-1, 1\}^{\Lambda_N}. \tag{2.1.1}$$

(In the preceding summation, each unordered pair $u \sim v$ only appears once.) Quenched on the external field $\{h_v\}$, the Ising measure with plus boundary condition (respectively minus boundary condition) is defined such that for all $\sigma \in \{-1, 1\}^{\Lambda_N}$ (throughout the paper the temperature is fixed, and thus we suppress the dependence on β in all notations)

$$\mu^{\Lambda_N,\pm}(\sigma) = \frac{e^{-\beta H^{\Lambda_N,\pm}(\sigma)}}{Z^{\Lambda_N,\pm}}, \text{ where } Z^{\Lambda_N,\pm} = \sum_{\sigma' \in \{-1,1\}^{\Lambda_N}} e^{-\beta H^{\Lambda_N,\pm}(\sigma')}. \tag{2.1.2}$$

Note that $\mu^{\Lambda_N,\pm}$ is a random measure which itself depends on $\{h_v\}$. To be clear of the two different sources of randomness, we use $\mathbb{P}$ and $\mathbb{E}$ to refer to the probability measure with respect to the external field $\{h_v\}$; and we use $\mu^{\Lambda_N,\pm}$ for the Ising measures and use $\langle\cdot\rangle_{\mu^{\Lambda_N,\pm}}$ to denote the expectations with respect to the Ising measures.

Theorem 2.1.1. *For any $\varepsilon > 0, T \in [0,\infty)$, there exists $c = c(\varepsilon, T) > 0$ such that*

$$\mathbb{E}(\langle\sigma_o\rangle_{\mu^{\Lambda_N,+}} - \langle\sigma_o\rangle_{\mu^{\Lambda_N,-}}) \leqslant c^{-1}e^{-cN} \text{ for all } N \geqslant 1\,.$$

This result lies under the umbrella of the general Imry–Ma [78] phenomenon, which states that in two-dimensional systems any first order transition is rounded off upon the introduction of arbitrarily weak static, or quenched, disorder in the parameter conjugate to the corresponding extensive quantity. In the particular case of the RFIM, it was shown in [4, 5] that the effect of the boundary conditions on magnetization at distance N decays to 0 as $N \to \infty$ for all non-negative temperatures and arbitrarily weak quenched disorder (this also implies the uniqueness of the Gibbs state). The decay rate was then improved to $1/\sqrt{\log\log N}$ in [35] and to $1/N^\gamma$ (for some $\gamma > 0$) in [3]. In the presence of strong disorder it has been shown that there is an exponential decay [26, 64, 31] (see also [3, Appendix A]). The main remaining challenge is to decide whether the decay rate is exponential when the disorder is weak. In fact, there have been debates even among physicists as to whether there exists a regime where the decay rate is polynomial, and weak supporting arguments have been made in both directions [70, 29, 51]—in particular in [51] an argument was made for polynomial decay *at zero temperature* for a certain choice of disorder. Theorem 2.1.1 provides a complete answer to this question when the random field consists of i.i.d. Gaussian variables.

The two-dimensional behavior of the RFIM is drastically different from that for dimensions three and higher: it was shown in [77] that at zero temperature the effect on the local quenched magnetization of the boundary conditions at distance N does not vanish in N

in the presence of weak disorder, and later an analogous result was proved in [30] at low temperatures. A heuristic explanation behind the different behaviors is as follows: in d dimensions the fluctuation of the random field in a box of side length N is of order $N^{d/2}$, whereas boundary condition effect is of order N^{d-1} (thus, in two dimensions the fluctuation of the random field in a box is of the same order as the size of the boundary, while in three dimensions and above the fluctuation of the random field is substantially smaller than the size of the boundary).

Our proof method is different from all of [5, 35, 3] (and different from [26, 64, 31]), except that in the heuristic level our proof seems to be related to the Mandelbrot percolation analogy presented in [3, Appendix B]. The works [4, 5] treated a wide class of distributions for disorder, while [35, 3] and this paper work with Gaussian disorder. The main features of Gaussian distributions used in this paper are the simple formula for the change of measure (see (2.2.12)) and linear decompositions for Gaussian process (see (2.2.21)). In addition, we remark that the analysis in [5, 3] extends to the case with finite-range interactions. While we expect our framework to be useful in analyzing the finite-range case, the lack of planar duality seems to present some non-trivial obstacle (see Remark 2.2.3).

The rest of the paper consists of two sections. In Section 2.2, we prove Theorem 2.1.1 in the special case of $T = 0$. In our opinion, this is a significant simplification of the general case but still captures the core challenge of the problem. We hope that some of the key ideas (e.g., the crucial application of [1]) can be more transparent by first presenting the proof in this simplified case. In Section 2.3, we then present the proof for the case of $T > 0$. While the proof naturally shares the key insights with the case for $T = 0$, it seems to us that there are significant additional obstacles. As a result, the proof is not presented as an extension of the zero-temperature case. Instead, we present an almost self-contained proof, but omit details at times when they are merely adaption of arguments in Section 2.2.

Our (shared) notations in Sections 2.2 and 2.3 are consistent with each other, and a few notations in Section 2.3 are natural extensions of those in Section 2.2. However, for clarity of exposition, we will recall or re-explain all notations in Section 2.3.

Concurrent work. During the submission of this paper, a paper [2] which proved the same result was completed. The proof of [2] was inspired by the proof at zero temperature in this paper (for the crucial application of [1]). Both proofs share the basic intuition of "using the fluctuation of the sum of the random field in a box to fight the influence of the boundary condition" (which went back to [4, 5]) and both apply [1] to disagreement percolation in a crucial manner. However, the two approaches seem to be rather different in at least the following two important aspects: (1) This paper employs first moment analysis via various perturbations of the random field, and the paper [2] (similar to [3]) relies on concentration/anti-concentration type of analysis (which in particular uses second-moment computations); (2) At positive temperatures, this paper employs a certain monotone coupling (adaptive admissible coupling as in Definition 2.3.9) between Ising measures with different boundary conditions, and the paper [2] considers a continuous extension of the Ising model into the metric graph which allows to study spin correlations via disagreement percolation for two *independent* samples (inspired by [107, 115]).

2.2. Exponential decay at zero temperature

At zero temperature, $\mu^{\Lambda_N,+}$ (and respectively $\mu^{\Lambda_N,-}$) is supported on the minimizer of (2.1.1), which is known as the *ground state* and is unique with probability 1. We denote by $\sigma^{\Lambda_N,+}$ the ground state with respect to the plus-boundary condition and by $\sigma^{\Lambda_N,-}$ the ground state with respect to the minus-boundary condition. Therefore, for $T = 0$ we have the simplification that the only randomness is from the $\mathbb{P}$-measure. Thus, Theorem 2.1.1 for $T = 0$ can then be simplified as follows.

Theorem 2.2.1. *For any $\varepsilon > 0$, there exists $c = c(\varepsilon) > 0$ such that $\mathbb{P}(\sigma_o^{\Lambda_N,+} \neq \sigma_o^{\Lambda_N,-}) \leqslant c^{-1}e^{-cN}$ for all $N \geqslant 1$.*

2.2.1. Outline of the proof

We first reformulate Theorem 2.2.1. For $v \in \Lambda_N$, we define

$$\xi_v^{\Lambda_N} = \begin{cases} +, & \text{if } \sigma_v^{\Lambda_N,+} = \sigma_v^{\Lambda_N,-} = 1\,, \\ -, & \text{if } \sigma_v^{\Lambda_N,+} = \sigma_v^{\Lambda_N,-} = -1\,, \\ \mathsf{0}, & \text{if } \sigma_v^{\Lambda_N,+} = 1 \text{ and } \sigma_v^{\Lambda_N,-} = -1\,. \end{cases} \tag{2.2.1}$$

By monotonicity (c.f. [3, Section 2.2]), the case of $\sigma_v^{\Lambda_N,+} = -1$ and $\sigma_v^{\Lambda_N,-} = 1$ cannot occur, so $\xi_v^{\Lambda_N}$ is well-defined for all $v \in \Lambda_N$. Theorem 2.2.1 can be restated as

$$m_N \leqslant c^{-1} e^{-cN} \text{ for } c = c(\varepsilon) > 0, \text{ where } m_N \triangleq \mathbb{P}(\xi_o^{\Lambda_N} = \mathsf{0})\,. \tag{2.2.2}$$

For any $A \subseteq \mathbb{Z}^2$, we can analogously define ξ^A by replacing Λ_N with A in (2.1.1) and (2.2.1). Let $\mathcal{C}^A = \{v \in A : \xi_v^A = \mathsf{0}\}$ (that is, $\mathcal{C}^A$ is the collection of *disagreements*). Monotonicity (see [3, (2.7)]) implies that

$$\mathcal{C}^B \cap B' \subseteq \mathcal{C}^{B'} \text{ provided that } B' \subseteq B\,. \tag{2.2.3}$$

In particular, this implies that m_N is decreasing in N, so we need only consider $N = 2^n$ for $n \geqslant 1$. Clearly, for any $v \in \mathcal{C}^A$, there exists a path in $\mathcal{C}^A$ joining v and ∂A. This suggests consideration of percolation properties of $\mathcal{C}^A$. Indeed, a key step in our proof for (2.2.2) is the following proposition on the lower bound on the length exponent for geodesics (i.e., shortest paths) in $\mathcal{C}^{\Lambda_N}$. For any $A \subseteq \mathbb{Z}^2$, we denote by $d_A(\cdot,\cdot)$ the intrinsic distance on A, i.e., the graph distance on the induced subgraph on A. Let $d_A(A_1, A_2) = \min_{x \in A_1 \cap A, y \in A_2 \cap A} d_A(x, y)$ (with the convention that $\min \emptyset = \infty$).

Proposition 2.2.2. *There exist $\alpha = \alpha(\varepsilon) > 1$, $\kappa = \kappa(\varepsilon) > 0$ such that for all $N \geqslant 1$*

$$\mathbb{P}(d_{\mathcal{C}^{\Lambda_N}}(\partial \Lambda_{N/4}, \partial \Lambda_{N/2}) \leqslant N^{\alpha}) \leqslant \kappa^{-1} e^{-N^{\kappa}}\,. \tag{2.2.4}$$

Remark 2.2.3. The "only" place where our proof breaks in extending to the finite range case is to verify Proposition 2.2.2 (and its analogue at positive temperatures, Proposition 2.3.1). The exact points where the extension of the proof encounters issues depend somewhat on exact formulations for sub-lemmas. For instance, at zero temperature one can try to prove a version of Lemma 2.2.8 sticking to nearest neighbor crossings, then for lack of planar duality there are issues both in the proof of Lemma 2.2.8 (more specifically in Case 1) and in the proof of (2.2.6) which applies Lemma 2.2.8. Of course one can also try to prove a stronger version of Lemma 2.2.8 (which suffices to prove (2.2.6)), but this may be hard.

The proof of Proposition 2.2.2 will rely on [1], which takes the next lemma as input. For any rectangle $A \subseteq \mathbb{R}^2$ (whose sides are not necessarily parallel to the axes), let ℓ_A be the length of the longer side and let A^{Large} be (the lattice points of) the square box concentric with A, of side length $32\ell_A$ and with sides parallel to axes. In addition, define the aspect ratio of A to be the ratio between the lengths of the longer and shorter sides. For a (random) set $\mathcal{C} \subseteq \mathbb{Z}^2$, we use $\mathrm{Cross}(A, \mathcal{C})$ to denote the event that there exists a path $v_0, \ldots, v_k \in A \cap \mathcal{C}$ connecting the two shorter sides of A (that is, v_0, v_k are of ℓ_∞-distances less than 1 respectively from the two shorter sides of A).

Lemma 2.2.4. *Write $a = 100$. There exists $\ell_0 = \ell_0(\varepsilon)$ and $\delta = \delta(\varepsilon) > 0$ such that the following holds for any $N \geqslant 1$. For any $k \geqslant 1$ and any rectangles $A_1, \ldots, A_k \subseteq \{v \in \mathbb{R}^2 : |v|_\infty \leqslant N/2\}$ with aspect ratios at least a such that (a) $\ell_0 \leqslant \ell_{A_i} \leqslant N/32$ for all $1 \leqslant i \leqslant k$ and (b) $A_1^{\mathrm{Large}}, \ldots, A_k^{\mathrm{Large}}$ are disjoint, we have*

$$\mathbb{P}(\cap_{i=1}^k \mathrm{Cross}(A_i, \mathcal{C}^{\Lambda_N})) \leqslant (1-\delta)^k .$$

(Actually, the authors of [1] treated random curves in $\mathbb{R}^2$. However, the main capacity analysis can be copied in the discrete case, and the connection between the capacity and the box-counting dimension is straightforward (c.f. [50, Lemma 2.3]).) Armed with Lemma 2.2.4, we can apply [1, Theorem 1.3] to deduce that for some $\alpha = \alpha(\varepsilon) > 1$,

$$\mathbb{P}(d_{\mathcal{C}^{\Lambda_N}}(\partial\Lambda_{N/4}, \partial\Lambda_{N/2}) \leqslant N^{\alpha}) \to 0 \text{ as } N \to \infty\,. \tag{2.2.5}$$

By a standard percolation argument (Lemma 2.2.10) which we will explain later, we can enhance the probability decay in (2.2.5) and prove (2.2.4).

By (2.2.3), the random set $\mathcal{C}^{\Lambda_N} \cap A$ is stochastically dominated by $\mathcal{C}^{A^{\text{Large}}} \cap A$ as long as $A^{\text{Large}} \subseteq \Lambda_N$. Moreover, it is obvious that $\mathcal{C}^{A_i^{\text{Large}}}$ for $1 \leqslant i \leqslant k$ are mutually independent, as long as the sets A_i^{Large} for $1 \leqslant i \leqslant k$ are disjoint. Therefore, in order to prove Lemma 2.2.4, it suffices to show that for any rectangle A with aspect ratio at least $a = 100$ we have

$$\mathbb{P}(\text{Cross}(A, \mathcal{C}^{A^{\text{Large}}})) \leqslant 1 - \delta \text{ where } \delta = \delta(\varepsilon) > 0\,. \tag{2.2.6}$$

Both the proof of (2.2.6) and the application of (2.2.4) rely on a perturbative analysis, which is another key feature of our proof. Roughly speaking, the logic is as follows:

- We first consider the perturbation by increasing the field by an amount of order $1/N$, and use this to show that the probability for a $\mathtt{0}$-valued contour surrounding an annulus is strictly bounded away from 1.
- Based on this property, we prove (2.2.6), which then implies (2.2.4).
- Given (2.2.4), we then show that increasing the field by an amount of order $1/N^{\alpha}$ (recall that $\alpha > 1$ is from Proposition 2.2.2 and thus the perturbation here is $1/N^{\alpha} \ll 1/N$) will most likely change the $\mathtt{0}$'s to $+$'s. Based on this, we prove polynomial decay for m_N with large power, which can then be enhanced to exponential decay.

For compactness of exposition, the actual implementation will differ slightly from the above plan:

- We first prove a general perturbation result (Lemma 2.2.5) in Section 2.2.2, where the size of perturbation is related to the intrinsic distance on $\mathcal{C}^{\Lambda_N}$.

- In Section 2.2.3, we apply Lemma 2.2.5 by bounding $d_{\mathcal{C}^{\Lambda_N}}$ from below by the ℓ_1-distance and correspondingly setting the perturbation amount to $1/N$, thereby proving Lemma 2.2.8. As a consequence, we verify (2.2.6).

- In Section 2.2.4, we apply Lemma 2.2.5 again by applying a lower bound on $d_{\mathcal{C}^{\Lambda_N}}$ from Proposition 2.2.2. This allows us to derive Lemma 2.2.11. As a consequence, we prove in Lemma 2.2.14 polynomial decay for m_N with large power, which is then enhanced to exponential decay by a standard argument.

2.2.2. A perturbative analysis

We first introduce some notation. For $A \subseteq \mathbb{Z}^2$, we set $h_A = \sum_{v \in A} h_v$. For $A, B \subseteq \mathbb{Z}^2$, we denote by $E(A,B) = \{\langle u,v \rangle : u \sim v, u \in A, v \in B\}$. Note that we treat $\langle u,v \rangle$ as an ordered edge. For simplicity, we will only consider $N = 2^n$ for $n \geqslant 10$. Let $\mathcal{A}_N = \Lambda_N \setminus \Lambda_{N/2}$ be an annulus. Define $\{\widetilde{h}_v^{(N)} : v \in \Lambda_N\}$ to be a perturbation of the original field parameterized by $\Delta > 0$, as follows:

$$\widetilde{h}_v^{(N)} = h_v + \Delta \text{ for } v \in \Lambda_N \,. \tag{2.2.7}$$

We will use $\widetilde{H}^{\Lambda_N,\pm}(\sigma)$, $\widetilde{\sigma}^{\Lambda_N,\pm}$, $\widetilde{\xi}^{\Lambda_N}$, $\widetilde{\mathcal{C}}^{\Lambda_N}$ to denote the corresponding tilde versions of $H^{\Lambda_N,\pm}(\sigma)$, $\sigma^{\Lambda_N,\pm}$, ξ^{Λ_N}, $\mathcal{C}^{\Lambda_N}$ respectively, i.e., defined analogously but with respect to the field $\{\widetilde{h}_v^{(N)}\}$. In addition, define $\mathcal{C}_*^{\Lambda_N} = \widetilde{\mathcal{C}}^{\Lambda_N} \cap \mathcal{C}^{\Lambda_N}$ (so $\mathcal{C}_*^{\Lambda_N}$ is the intersection of disagreements with respect to the original and the perturbed field; in informal discussions we will refer to vertices in $\mathcal{C}_*^{\Lambda_N}$ as disagreements too).

Lemma 2.2.5. *Consider $K, \Delta > 0$. Define $\{\widetilde{h}_v^{(N)} : v \in \Lambda_N\}$ as in (2.2.7). The following two conditions cannot hold simultaneously:*

(a) $d_{\mathcal{C}_*^{\Lambda_N}}(\partial\Lambda_{N/4}, \partial\Lambda_{N/2}) \geqslant K$*;*

(b) $|\mathcal{C}_*^{\Lambda_N} \cap \Lambda_{N/4}| \cdot \Delta > \frac{8}{K}|\mathcal{C}_*^{\Lambda_N} \cap \mathcal{A}_{N/2}|$.

Proof. Suppose otherwise both (a) and (b) hold. Let $B_k = \{v \in \mathcal{A}_{N/2} : d_{\mathcal{C}_*^{\Lambda_N}}(\partial\Lambda_{N/4}, v) = k\}$, for $k = 1, \ldots, K$. Note that $B_k \subseteq \mathcal{C}_*^{\Lambda_N} \cap \mathcal{A}_{N/2}$ for all $1 \leqslant k \leqslant K$ by (a). It is obvious that the B_k's are disjoint from each other, and thus there exists a minimal value k_* such that

$$|B_{k_*}| \leqslant K^{-1}|\mathcal{C}_*^{\Lambda_N} \cap \mathcal{A}_{N/2}|\,. \tag{2.2.8}$$

Let

$$S = (\mathcal{C}_*^{\Lambda_N} \cap \Lambda_{N/4}) \cup \cup_{k=1}^{k^*-1} B_k\,,$$

and for $\tau \in \{-, \mathsf{0}, +\}$, define

$$g(S,\tau) = \{\langle u,v\rangle \in E(S,S^c) : \xi_v^{\Lambda_N} = \tau\} \text{ and } \widetilde{g}(S,\tau) = \{\langle u,v\rangle \in E(S,S^c) : \widetilde{\xi}_v^{\Lambda_N} = \tau\}\,. \tag{2.2.9}$$

Note that for any $v \in \Lambda_N$ with $\xi_v^{\Lambda_N} = \mathsf{0}$ we have $\sigma_v^{\Lambda_N,+} = 1$. Since $\xi_v^{\Lambda_N} = \mathsf{0}$ for $v \in S$ (which implies that $\sigma_v^{\Lambda_N,+} = 1$ for $v \in S$),

$$h_S + |g(S,+)| - |g(S,-)| + |g(S,\mathsf{0})| \geqslant 0\,, \tag{2.2.10}$$

because if (2.2.10) does not hold, then $H^{\Lambda_N,+}(\sigma') < H^{\Lambda_N,+}(\sigma^{\Lambda_N,+})$ where σ' is obtained from $\sigma^{\Lambda_N,+}$ by flipping its value on S, thus contradicting the minimality of $H^{\Lambda_N,+}(\sigma^{\Lambda_N,+})$. In addition, by monotonicity (with respect to the external field), we have $g(S,\mathsf{0}) \subseteq \widetilde{g}(S,\mathsf{0}) \cup \widetilde{g}(S,+)$, $g(S,+) \subseteq \widetilde{g}(S,+)$, and thus

$$|\widetilde{g}(S,+)| - |g(S,+)| \geqslant |g(S,\mathsf{0}) \setminus \widetilde{g}(S,\mathsf{0})|\,.$$

Similarly, we have $\widetilde{g}(S,-) \subseteq g(S,-)$ and $\widetilde{g}(S,\mathsf{0}) \subseteq g(S,-) \cup g(S,\mathsf{0})$, and thus

$$|g(S,-)| - |\widetilde{g}(S,-)| \geqslant |\widetilde{g}(S,\mathsf{0}) \setminus g(S,\mathsf{0})|\,.$$

By our definition of B_k's, we see that $\widetilde{g}(S,\mathsf{0}) \cap g(S,\mathsf{0}) = E(S, B_{k_*})$. Therefore, (2.2.10) and the preceding two displays imply that

$$\begin{aligned}\widetilde{h}_S^{(N)} + |\widetilde{g}(S,+)| - |\widetilde{g}(S,-)| - |\widetilde{g}(S,0)| &\geqslant \widetilde{h}_S^{(N)} + |g(S,+)| - |g(S,-)| \\ &\quad + |g(S,0)| - 2|E(S,B_{k_*})| \\ &\geqslant |S|\Delta - 8|B_{k_*}| > 0\,,\end{aligned}$$

where the last inequality follows from (b) and (2.2.8). The preceding inequality implies $\widetilde{H}^{\Lambda_N,-}(\sigma') < \widetilde{H}^{\Lambda_N,-}(\widetilde{\sigma}^{\Lambda_N,-})$ where σ' is obtained from $\widetilde{\sigma}^{\Lambda_N,-}$ by flipping its value on S. This contradicts the minimality of $\widetilde{H}^{\Lambda_N,-}(\widetilde{\sigma}^{\Lambda_N,-})$, completing the proof of the lemma. □

Lemma 2.2.6. *For any $x_v \geqslant 0$ for $v \in \Lambda_N$, let $\check{h}_v^{(N)} = h_v + x_v$ for $v \in \Lambda_N$ (we will use $\check{H}^{\Lambda_N,\pm}(\sigma)$, $\check{\sigma}^{\Lambda_N,\pm}$, $\check{\xi}^{\Lambda_N}$, $\check{\mathcal{C}}^{\Lambda_N}$ to denote the corresponding ˇ versions of $H^{\Lambda_N,\pm}(\sigma)$, $\sigma^{\Lambda_N,\pm}$, ξ^{Λ_N}, $\mathcal{C}^{\Lambda_N}$). Then with probability 1, for any $v \in \mathcal{C}^{\Lambda_N} \cap \check{\mathcal{C}}^{\Lambda_N}$ there is a path in $\mathcal{C}^{\Lambda_N} \cap \check{\mathcal{C}}^{\Lambda_N}$ joining v and $\partial\Lambda_N$.*

Proof. The proof is similar to that of Lemma 2.2.5, and in a way it is the case of $K = \infty$ there.

Suppose that the claim is not true. Then take $v \in \mathcal{C}^{\Lambda_N} \cap \check{\mathcal{C}}^{\Lambda_N}$ (for which the claim fails), and let S be the connected component in $\mathcal{C}^{\Lambda_N} \cap \check{\mathcal{C}}^{\Lambda_N}$ that contains v (thus S is not neighboring $\partial\Lambda_N$). Define $g(S,\tau)$ as in (2.2.9) and define $\check{g}(S,\tau) = \{\langle u,v\rangle \in E(S,S^c) : \check{\xi}_v^{\Lambda_N} = \tau\}$. Similar to (2.2.10), we have that

$$h_S + |g(S,+)| - |g(S,-)| + |g(S,0)| \geqslant 0\,.$$

In our case, $g(S,0) \cup g(S,+) \subseteq \check{g}(S,+)$ and $\check{g}(S,0) \cup \check{g}(S,-) \subseteq g(S,-)$. Therefore,

$$\check{h}_S^{(N)} + |\check{g}(S,+)| - |\check{g}(S,-)| - |\check{g}(S,0)| \geqslant h_S + |g(S,+)| - |g(S,-)| + |g(S,0)| \geqslant 0\,.$$

The preceding inequality implies that $\check{H}^{\Lambda_N,-}(\sigma') \leqslant \check{H}^{\Lambda_N,-}(\check{\sigma}^{\Lambda_N,-})$ where σ' is obtained from $\check{\sigma}^{\Lambda_N,-}$ by flipping its value on S. This happens with probability 0 since the ground state is unique with probability 1. □

2.2.3. Proof of Proposition 2.2.2

In this section, we will set $K = K(N) = N/4$, and $\Delta = \Delta(N) = \gamma/N$ for an absolute constant $\gamma > 0$ to be selected, and we consider $\widetilde{h}^{(N)}$ as in (2.2.7). In this case Condition (a) in Lemma 2.2.5 holds trivially. For convenience, we use $\mathbb{P}_N$ to denote the probability measure with respect to the field $\{h_v : v \in \Lambda_N\}$ and use $\widetilde{\mathbb{P}}_N$ to denote the probability measure with respect to $\{\widetilde{h}_v^{(N)} : v \in \Lambda_N\}$.

Lemma 2.2.7. *Recall that ε is the variance parameter for the field $\{h_v\}$. For any $p > 0$, there exists $c = c(\varepsilon, p, \gamma) > 0$ such that for any event E_N with $\widetilde{\mathbb{P}}_N(E_N) \geqslant p$, we have that*

$$\mathbb{P}_N(E_N) \geqslant c\,.$$

Proof. There exists a constant $C > 0$ such that $\widetilde{\mathbb{P}}_N(|\widetilde{h}^{(N)}_{\Lambda_N} - \Delta|\Lambda_N|| \geqslant C\varepsilon N) \leqslant p/2$. Thus we have

$$\widetilde{\mathbb{P}}_N(E_N; |\widetilde{h}^{(N)}_{\Lambda_N} - \Delta|\Lambda_N|| \leqslant C\varepsilon N) \geqslant p/2\,. \tag{2.2.11}$$

Also, by a straightforward Gaussian computation, we see that

$$\frac{d\mathbb{P}_N}{d\widetilde{\mathbb{P}}_N} = \exp\Big\{-\frac{\Delta(\widetilde{h}^{(N)}_{\Lambda_N} - \Delta|\Lambda_N|)}{\varepsilon^2}\Big\}\exp\Big\{\frac{-\Delta^2|\Lambda_N|}{2\varepsilon^2}\Big\} \tag{2.2.12}$$

and thus there exists $\iota = \iota(\varepsilon) > 0$ such that

$$\frac{d\mathbb{P}_N}{d\widetilde{\mathbb{P}}_N} \geqslant \iota \text{ provided that } |\widetilde{h}^{(N)}_{\Lambda_N} - \Delta|\Lambda_N|| \leqslant C\varepsilon N\,.$$

Combined with (2.2.11), this completes the proof of the lemma. □

For any annulus $\mathcal{A}$, we denote by $\mathrm{Cross}_{\mathrm{hard}}(\mathcal{A}, \mathcal{C})$ the event that there is a contour in $\mathcal{C}$ which separates the inner and outer boundaries of $\mathcal{A}$, and by $\mathrm{Cross}_{\mathrm{easy}}(\mathcal{A}, \mathcal{C})$ the event that there is a path in $\mathcal{C}$ which connects the inner and outer boundaries of $\mathcal{A}$.

Lemma 2.2.8. *There exists $\delta = \delta(\varepsilon) > 0$ such that*

$$\min\{\mathbb{P}(\mathrm{Cross}_{\mathrm{hard}}(\Lambda_{N/8} \setminus \Lambda_{N/32}, \mathcal{C}^{\Lambda_N})), \mathbb{P}(\mathrm{Cross}_{\mathrm{easy}}(\Lambda_{N/8} \setminus \Lambda_{N/32}, \mathcal{C}^{\Lambda_N}))\} \leqslant 1 - \delta \text{ for all } N \geqslant 32.$$

Proof. We first provide a brief discussion on the outline of the proof. We refer to the disagreements on $\Lambda_{N/32}$ with plus/minus boundary conditions posed on $\partial\Lambda_{N/8}$ as the "enhanced" disagreements (the word enhanced is chosen since the enhanced disagreements stochastically dominate the disagreements with boundary conditions on $\partial\Lambda_N$ by monotonicity of the Ising model). Note that the set of disagreements in $\mathcal{A}_{N/2}$ is stochastically dominated by the union of a constant number of copies of enhanced disagreements, which are independent of the enhanced disagreements in $\Lambda_{N/32}$. Therefore, with positive probability the number of enhanced disagreements in $\Lambda_{N/32}$ is larger than (up to a constant factor) the number of disagreements in $\mathcal{A}_{N/2}$ (see (2.2.14)). On this event, (modulo a caveat) by Lemma 2.2.5 at least one of the enhanced disagreements is not a disagreement when considering boundary conditions on $\partial\Lambda_N$ — this yields the desired statement as incorporated in **Case 1** below. In **Case 2**, we tighten the argument by addressing the caveat which is the scenario that the enhanced disagreement is empty (this is relatively simple).

We are now ready to carry out the formal proof. We can write $\mathcal{A}_{N/2} = \cup_{i=1}^{r} A_i$ where each A_i is a box of side length $N/16$ (so a copy of $\Lambda_{N/32}$) and $r \geqslant 16$ is a fixed integer (while it is conventional to choose A_i's as disjoint boxes, the disjointness is not used in the proof). For a box A, denoting by A^{Big} as the concentric box of A whose side length is $4\ell_A$. We have that (see Figure 2.1)

$$A_i^{\mathrm{Big}} \cap \Lambda_{N/8} = \emptyset \text{ and } A_i^{\mathrm{Big}} \subseteq \Lambda_N \text{ for all } 1 \leqslant i \leqslant r. \tag{2.2.13}$$

For any $A \subseteq \Lambda_N$, let $\overline{\mathcal{C}}^A$ be defined as $\mathcal{C}^A$ but replacing $\{h_v : v \in A\}$ by $\{\widetilde{h}_v^{(N)} : v \in A\}$

(note that $\overline{\mathcal{C}}^{\Lambda_{N/2}}$ is different from $\widetilde{\mathcal{C}}^{\Lambda_{N/2}}$, which is defined with respect to $\widetilde{h}^{(N/2)}$). Write $\mathcal{C}_\diamond^A = \mathcal{C}^A \cap \overline{\mathcal{C}}^A$. Write $X_i = |\mathcal{C}_\diamond^{A_i^{\mathrm{Big}}} \cap A_i|$ and $X = |\mathcal{C}_\diamond^{\Lambda_{N/8}} \cap \Lambda_{N/32}|$. Clearly, X_i's and X are identically distributed and by (2.2.13) X_i's are independent of X (but X_i's are not mutually independent). Let $\theta = \inf\{x : \mathbb{P}(X \leqslant x) \geqslant 1 - 1/2r\}$. Thus,

$$\mathbb{P}(X \geqslant \max_{1\leqslant i\leqslant r} X_i, X \geqslant \theta) \geqslant \mathbb{P}(X \geqslant \theta)\mathbb{P}(\max_{1\leqslant i\leqslant r} X_i \leqslant \theta) \geqslant 1/4r\,. \tag{2.2.14}$$

The rest of the proof divides into two cases.

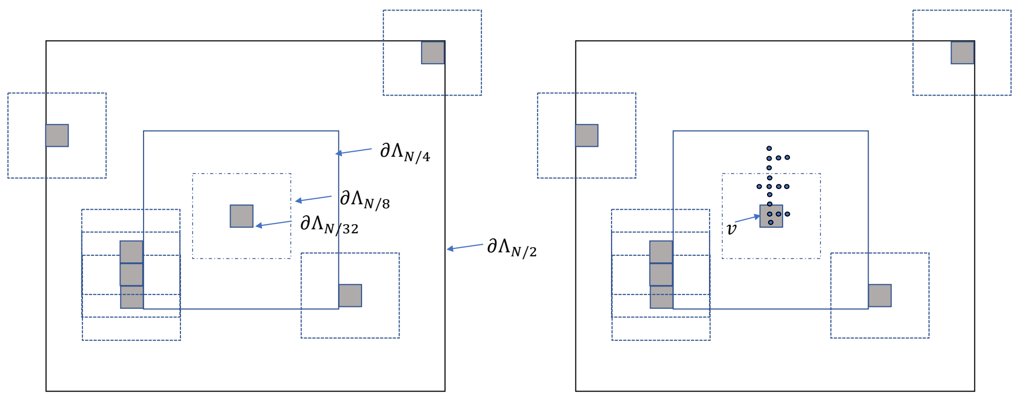

Figure 2.1: Illustration for the geometric setup of the proof for Lemma 2.7. In the picture on the left we cover $\mathcal{A}_{N/2}$ by a collection of translated copies of $\Lambda_{N/32}$ (the grey boxes) — we only draw out a few copies for an illustration. Note that the (4-times) enlargements of translated copies (while overlapping among themselves) are all disjoint with $\Lambda_{N/8}$. The picture on the right illustrates the scenario in Case 1: for some $v \in \mathcal{C}_\diamond^{\Lambda_{N/8}} \setminus \mathcal{C}^{\Lambda_N}$, we draw its component with the same ξ^{Λ_N}-value and this component necessarily goes out of $\Lambda_{N/8}$.

Case 1: $\theta > 0$. Let $\mathcal{E} = \{|\mathcal{C}_\diamond^{\Lambda_{N/8}} \cap \Lambda_{N/32}| \geqslant r^{-1}|\mathcal{C}_*^{\Lambda_N} \cap \mathcal{A}_{N/2}|\} \cap \{|\mathcal{C}_\diamond^{\Lambda_{N/8}} \cap \Lambda_{N/32}| > 0\}$. By (2.2.3) and (2.2.13), we have $|\mathcal{C}_*^{\Lambda_N} \cap \mathcal{A}_{N/2}| \leqslant \sum_{i=1}^r X_i$. Combined with (2.2.14), it gives that $\mathbb{P}(\mathcal{E}) \geqslant 1/4r$. Setting $\gamma = 100r$, we get that $|\mathcal{C}_\diamond^{\Lambda_{N/8}} \cap \Lambda_{N/32}| \cdot \Delta > 16K^{-1}|\mathcal{C}_*^{\Lambda_N} \cap \mathcal{A}_{N/2}|$ on $\mathcal{E}$. By Lemma 2.2.5, on $\mathcal{E}$ there is at least one vertex $v \in \mathcal{C}_\diamond^{\Lambda_{N/8}} \cap \Lambda_{N/32}$ but $v \notin \mathcal{C}_*^{\Lambda_N}$. So either $v \notin \mathcal{C}^{\Lambda_N}$ or $v \notin \widetilde{\mathcal{C}}^{\Lambda_N}$ on $\mathcal{E}$. Assume that $v \notin \mathcal{C}^{\Lambda_N}$ and the other case can be treated similarly.

We will use the following property: for any connected set $\mathcal{A}$, $u \notin \mathcal{C}^{\mathcal{A}}$ if and only if there exists a connected set $A \subseteq \mathcal{A}$ with $u \in A$ such that $\xi_w^{\mathcal{A}} = +$ for all $w \in A$ or $\xi_w^{\mathcal{A}} = -$ for all $w \in A$. The "if" direction of the property follows from (2.2.3). For the "only if" direction, we assume without loss that $\xi_u^{\mathcal{A}} = +$ and let A be the connected component containing u where the $\xi^{\mathcal{A}}$-value is $+$. Note $\sigma_w^{\mathcal{A},-} = -1$ for all $w \in \partial A$ and $\sigma_w^{\mathcal{A},-} = 1$ for all $w \in A$. This implies that $\xi_w^{A} = +$ for all $w \in A$.

By the preceding property, there exists a connected set $A \subseteq \Lambda_N$ with $v \in A$ such that $\xi_w^{A} = +$ for all $w \in A$ or $\xi_w^{A} = -$ for all $w \in A$ (see Figure 2.1 for an illustration). In addition, A cannot be contained in $\Lambda_{N/8}$ since otherwise it contradicts $v \in \mathcal{C}^{\Lambda_{N/8}}$. By planar duality, this implies that on $\mathcal{E}$, either $\mathrm{Cross}_{\mathrm{hard}}(\Lambda_{N/8} \setminus \Lambda_{N/32}, \mathcal{C}^{\Lambda_N})$ or $\mathrm{Cross}_{\mathrm{hard}}(\Lambda_{N/8} \setminus \Lambda_{N/32}, \widetilde{\mathcal{C}}^{\Lambda_N})$ does not occur (the second case corresponds to the case when $v \notin \widetilde{\mathcal{C}}^{\Lambda_N}$). Therefore,

$$\mathbb{P}((\mathrm{Cross}_{\mathrm{hard}}(\Lambda_{N/8} \setminus \Lambda_{N/32}, \mathcal{C}^{\Lambda_N}))^c) + \mathbb{P}((\mathrm{Cross}_{\mathrm{hard}}(\Lambda_{N/8} \setminus \Lambda_{N/32}, \widetilde{\mathcal{C}}^{\Lambda_N}))^c) \geqslant \mathbb{P}(\mathcal{E}) \geqslant 1/4r\,.$$

Combined with Lemma 2.2.7, this completes the proof of the lemma.

Case 2: $\theta = 0$. Applying a simple union bound (by using 16 copies of $\Lambda_{N/32}$ to cover $\Lambda_{N/8}$, and a derivation similar to $|\mathcal{C}_*^{\Lambda_N} \cap \mathcal{A}_{N/2}| \leqslant \sum_{i=1}^r X_i$) we get that $\mathbb{P}(\mathcal{C}_*^{\Lambda_N} \cap \Lambda_{N/8} = \emptyset) \geqslant 1/2$. We assume without loss that $\mathbb{P}(\mathrm{Cross}_{\mathrm{easy}}(\Lambda_{N/8} \setminus \Lambda_{N/32}, \mathcal{C}^{\Lambda_N})) \geqslant 3/4$ (otherwise there is nothing further to prove), and thus

$$\mathbb{P}(\mathrm{Cross}_{\mathrm{easy}}(\Lambda_{N/8} \setminus \Lambda_{N/32}, \mathcal{C}^{\Lambda_N}) \text{ and } \mathcal{C}_*^{\Lambda_N} \cap \Lambda_{N/8} = \emptyset) \geqslant 1/4\,.$$

On the event $\mathrm{Cross}_{\mathrm{easy}}(\Lambda_{N/8} \setminus \Lambda_{N/32}, \mathcal{C}^{\Lambda_N})$ and $\mathcal{C}_*^{\Lambda_N} \cap \Lambda_{N/8} = \emptyset$, the easy crossing (joining two boundaries of $\Lambda_{N/8} \setminus \Lambda_{N/32}$) in $\mathcal{C}^{\Lambda_N}$ becomes an easy crossing with $\widetilde{\xi}^{\Lambda_N}$-values $+$. Thus, by planar duality, it prevents existence of a contour surrounding $\Lambda_{N/32}$ in $(\Lambda_{N/8} \setminus \Lambda_{N/32}) \cap \widetilde{\mathcal{C}}^{\Lambda_N}$. Therefore,

$$\mathbb{P}((\mathrm{Cross}_{\mathrm{hard}}(\Lambda_{N/8} \setminus \Lambda_{N/32}, \widetilde{\mathcal{C}}^{\Lambda_N}))^c) \geqslant 1/4\,.$$

Combined with Lemma 2.2.7, this completes the proof of the lemma. □

Proof of (2.2.6). Let $N = \min\{2^n : 2^{n+2} \geqslant \ell_A\}$. By our assumption on A, it is clear that we can position four copies A_1, A_2, A_3, A_4 of A by translation or rotation by 90 degrees so that (see the left of Figure 2.2)

- $A_1, A_2, A_3, A_4 \subseteq \Lambda_{N/8} \setminus \Lambda_{N/32}$.
- The union of any crossings through A_1, A_2, A_3, A_4 in their longer directions surrounds $\Lambda_{N/32}$.
- $\Lambda_N \subseteq A_i^{\text{Large}}$ for $1 \leqslant i \leqslant 4$.

Set $p = \mathbb{P}(\text{Cross}(A, \mathcal{C}^{A^{\text{Large}}}))$ (note that p depends on the dimension of A and also the orientation of A). By rotation symmetry and (2.2.3) we see that $\mathbb{P}(\text{Cross}(A_i, \mathcal{C}^{\Lambda_N})) \geqslant \mathbb{P}(\text{Cross}(A_i, \mathcal{C}^{A_i^{\text{Large}}})) = p$. In what follows, we denote $\mathcal{A} = \Lambda_{N/8} \setminus \Lambda_{N/32}$. Then, by $\mathbb{P}(\text{Cross}(A_i, \mathcal{C}^{\Lambda_N})) \geqslant p$ and a simple union bound, we get that

$$\mathbb{P}(\text{Cross}_{\text{hard}}(\mathcal{A}, \mathcal{C}^{\Lambda_N})) \geqslant \mathbb{P}(\cap_{i=1}^4 \text{Cross}(A_i, \mathcal{C}^{\Lambda_N})) \geqslant 1 - 4(1-p)\,. \tag{2.2.15}$$

Similarly, we can arrange two copies A_a, A_b of A obtained by translation and rotation by 90 degrees such that $\Lambda_N \subseteq A_a^{\text{Large}}, A_b^{\text{Large}}$ and that the union of any two crossings through $A_a^{\text{Large}}, A_b^{\text{Large}}$ in the longer direction connects the two boundaries of $\mathcal{A}$ (see the right of Figure 2.2). This implies that

$$\mathbb{P}(\text{Cross}_{\text{easy}}(\mathcal{A}, \mathcal{C}^{\Lambda_N})) \geqslant \mathbb{P}(\text{Cross}(A_a, \mathcal{C}^{\Lambda_N}) \cap \text{Cross}(A_b, \mathcal{C}^{\Lambda_N})) \geqslant 1 - 2(1-p)\,. \tag{2.2.16}$$

Combined with (2.2.15) and Lemma 2.2.8, it yields that $p \leqslant 1 - \delta$ for some $\delta = \delta(\varepsilon) > 0$ as required. $\square$

The following standard lemma will be applied several times below. Before presenting the lemma, we first provide a definition.

Definition 2.2.9. Divide Λ_N into disjoint boxes of side lengths $N' \leqslant N$ where $N' = 2^{n'}$ for some $n' \geqslant 1$, and denote by $\mathcal{B}(N, N')$ the collection of such boxes. Consider a percolation process on $\mathcal{B}(N, N')$, where each box $B \in \mathcal{B}(N, N')$ is regarded open or closed randomly. For $C, p > 0$, we say that the percolation process satisfies the (N, N', C, p)-condition if for each $B \in \mathcal{B}(N, N')$, there exists an event E_B such that

- On E_B^c, B is closed.
- $\mathbb{P}(E_B) \leqslant p$ for each B.
- If $\min_{x \in B_i, y \in B_j} |x - y|_\infty \geqslant CN'$ for all $1 \leqslant i < j \leqslant k$, then the events $E_{B_1}, \ldots, E_{B_k}$ are mutually independent.

Furthermore, we say two boxes B_1, B_2 are adjacent if $\min_{x_1 \in B_1, x_2 \in B_2} |x_1 - x_2|_\infty \leqslant 1$, and we say a collection of boxes is a lattice animal if these boxes form a connected graph.

Lemma 2.2.10. *For any $C > 0$, there exists $p > 0$ such that for all N and $N' \leqslant N$ and any percolation process on $\mathcal{B}(N, N', C, p)$ satisfying the (N, N', C, p)-condition, we have*

$$\mathbb{P}(\textit{there exists a lattice animal of open boxes on } \mathcal{B}(N, N') \textit{ of size at least } k) \leqslant (\tfrac{N}{N'})^2 2^{-k}\,.$$

Proof. On the one hand, the number of lattice animals of size exactly k is bounded by $(\frac{N}{N'})^2 8^{2k}$ (the bound comes from first choosing a starting box, and then encoding the lattice animal by a surrounding contour on $\mathcal{B}(N, N')$ of length $2k$). On the other hand, for any k such boxes, we can extract a sub-collection of ck boxes (here $c > 0$ is a constant that depends only on C) such that the pairwise distances of boxes in this sub-collection are at least CN'; hence the probability that all these k boxes are open is at most p^{ck}. The proof of the lemma is then completed by a simple union bound, employing the (N, N', C, p)-condition. $\square$

Proof of Proposition 2.2.2. Let $N' = N^{1-(\frac{\alpha-1}{10} \wedge \frac{1}{10})}$, where α is as in (2.2.5). For each $B \in \mathcal{B}(N, N')$, we say B is open if $d_{\mathcal{C}^{B^{\text{Large}}}}(\partial B, \partial B^{\text{large}}) \leqslant (N')^\alpha$, where B^{large} is the box concentric with B of doubled side length and B^{Large} (as we recall) is a concentric box of B with

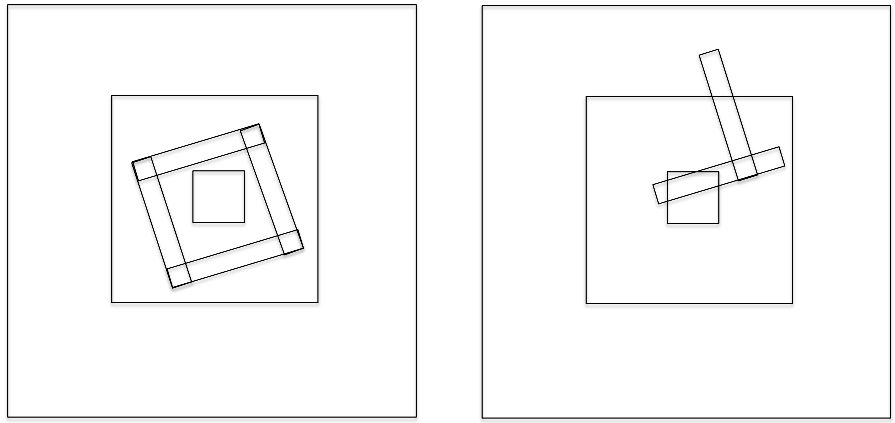

Figure 2.2: On both left and right, the three concentric square boxes are Λ_N, $\Lambda_{N/8}$ and $\Lambda_{N/32}$ respectively. On the left, the four rectangles are A_1, A_2, A_3, A_4 and on the right the two rectangles are A_a, A_b.

side length $32\ell_B$. By (2.2.5), we see that this percolation process satisfies the $(N, N', 64, p)$-condition where $p \to 0$ as $N \to \infty$. Now, in order that $d_{\mathcal{C}^{\Lambda_N}}(\partial\Lambda_{N/4}, \partial\Lambda_{N/2}) \leqslant (N')^\alpha$, there must exist an open lattice animal on $\mathcal{B}(N, N')$ of size at least $\frac{N}{16N'}$. Applying Lemma 2.2.10 completes the proof of Proposition 2.2.2 (since $(\alpha(1 - (\frac{\alpha-1}{10} \wedge \frac{1}{10})) > 1)$. □

2.2.4. Proof of Theorem 2.2.1

In this subsection, we will show that the probability for $\{o \in \mathcal{C}^{\Lambda_N}\}$ has a polynomial decay with large power (Lemma 2.2.14), which then yields Theorem 2.2.1 by a standard application of Lemma 2.2.10. In order to prove Lemma 2.2.14, we first provide a bound on the probability for $\{o \in \mathcal{C}_*^{\Lambda_N}\}$ (Lemma 2.2.11), whose proof crucially relies on Proposition 2.2.2.

Let $\alpha > 1$ be as in Proposition 2.2.2 (note that we can assume without loss that $\alpha \leqslant 2$). Let $\sqrt{1/\alpha} < \alpha' < 1$ (and thus we have $\alpha(\alpha')^2 > 1$).

Lemma 2.2.11. *For $N^\diamond \geqslant 16$, set $\Delta = (N^\diamond)^{-\alpha(\alpha')^2}$ and let $\widetilde{h}^{(N)}$ be defined as in* (2.2.7) *for $N \leqslant N^\diamond$. Write $m_N^\diamond = m_N^\diamond(N^\diamond) = \mathbb{P}(o \in \mathcal{C}_*^{\Lambda_N})$. Then there exists $C = C(\varepsilon) > 0$ such that $m_{N^\diamond}^\diamond \leqslant C(N^\diamond)^{-6}$.*

Remark 2.2.12. (1) In this lemma, regardless of the size of the box under consideration, the amount of perturbation Δ in our field $\widetilde{h}^{(N)}$ only depends on $N^\diamond$. This is crucial for (2.2.18) below. (2) Since $\alpha(\alpha')^2 > 1$, we have that $\Delta \ll 1/N^\diamond$ (this is crucial for getting a large power in the polynomial bound as in Lemma 2.2.14). (3) Since our perturbation $\Delta = (N^\diamond)^{-\alpha(\alpha')^2}$ applies to all $N \leqslant N^\diamond$, when N is very small in comparison of $N^\diamond$ the perturbation is possibly too mild and thus we may not have a good control on $\mathcal{C}_*^{\Lambda_N}$. However, this is not a problem because in the proof below we will only consider $N \geqslant (N^\diamond)^{\alpha'}$ (for which the perturbation is still significant).

Proof. Write $K = (N^\diamond)^{\alpha\alpha'}$. We claim it suffices to show that there exists $N_0 = N_0(\varepsilon)$ such that for $N^\diamond \geqslant N_0$

$$m^\diamond_{2N} \leqslant K^{-\frac{1-\alpha'}{2}} m^\diamond_{N/2} \text{ for } (N^\diamond)^{\alpha'} \leqslant N \leqslant N^\diamond\,. \tag{2.2.17}$$

Indeed, since $K = (N^\diamond)^{\alpha\alpha'}$, we can deduce from (2.2.17) by recursion that $m^\diamond_{N^\diamond} \leqslant e^{-c(\log N^\diamond)^2}$ for some constant $c > 0$, which yields the claimed bound in the lemma (with room to spare).

We now turn to the proof of (2.2.17). Suppose that (2.2.17) fails for some $(N^\diamond)^{\alpha'} \leqslant N \leqslant N^\diamond$. Since $\Lambda_N \subseteq v + \Lambda_{2N}$ for all $v \in \Lambda_{N/4}$ and $v + \Lambda_{N/2} \subseteq \Lambda_N$ for all $v \in \mathcal{A}_{N/2}$, by (2.2.3) we see

$$\mathbb{E}|\mathcal{C}_*^{\Lambda_N} \cap \Lambda_{N/4}| \geqslant \frac{N^2}{32} m^\diamond_{2N} \text{ and } \mathbb{E}|\mathcal{C}_*^{\Lambda_N} \cap \mathcal{A}_{N/2}| \leqslant N^2 m^\diamond_{N/2}\,. \tag{2.2.18}$$

Together with the assumption that (2.2.17) fails, this yields that

$$\mathbb{E}|\mathcal{C}_*^{\Lambda_N} \cap \Lambda_{N/4}| > 32^{-1} K^{-\frac{1-\alpha'}{2}} \mathbb{E}|\mathcal{C}_*^{\Lambda_N} \cap \mathcal{A}_{N/2}|\,.$$

Since $|\mathcal{C}_*^{\Lambda_N} \cap \Lambda_{N/4}|$ and $|\mathcal{C}_*^{\Lambda_N} \cap \mathcal{A}_{N/2}|$ are integer-valued and are at most N^2, the preceding inequality implies that (recall that $\alpha' > 1/\sqrt{\alpha} \geqslant 1/\sqrt{2}$)

$$\mathbb{P}(|\mathcal{C}_*^{\Lambda_N} \cap \Lambda_{N/4}| > 64^{-1} K^{-\frac{1-\alpha'}{2}} |\mathcal{C}_*^{\Lambda_N} \cap \mathcal{A}_{N/2}|) \geqslant \frac{1}{32N^3}\,.$$

Now, set $N_0 = N_0(\varepsilon)$ sufficiently large so that

$$\frac{1}{10^6 N^3} > \kappa^{-1} e^{-N^\kappa} \text{and } 64^{-1} K^{-\frac{1-\alpha'}{2}} > \frac{8}{K\Delta} \text{ for all } N \geqslant (N_0)^{\alpha'} . \tag{2.2.19}$$

Therefore, by Proposition 2.2.2, there is a positive probability such that

$$|\mathcal{C}_*^{\Lambda_N} \cap \Lambda_{N/4}| > 64^{-1} K^{-\frac{1-\alpha'}{2}} |\mathcal{C}_*^{\Lambda_N} \cap \mathcal{A}_{N/2}| \text{ and } d_{\mathcal{C}_*^{\Lambda_N}}(\partial\Lambda_{N/4}, \partial\Lambda_{N/2}) \geqslant K .$$

In particular, there exists at least one instance for the two events in the preceding display to occur simultaneously. This contradicts Lemma 2.2.5, thus completing the proof of the lemma. □

In the proof of Lemma 2.2.14 below, it is important for us to have independence between different scales. To this end, it is useful to consider a perturbation which only occurs in an annulus. In order to make a difference in notation from the previous perturbation (which occurs in a whole box), for $\Delta(N) > 0$ we define (we emphasize the dependence of Δ on N in the notation here since later in Lemma 2.2.14 we will consider perturbations for different N's simultaneously)

$$\widehat{h}_v^{(N)} = \begin{cases} h_v + \Delta(N) & \text{for } v \in \Lambda_N \setminus \Lambda_{N/4} , \\ h_v & \text{for } v \in \Lambda_{N/4} . \end{cases} \tag{2.2.20}$$

We then define $\widehat{\mathcal{C}}^{\Lambda_N}$ similar to $\mathcal{C}^{\Lambda_N}$ but with respect to the field $\{\widehat{h}_v^N : v \in \Lambda_N\}$. Further, let $\mathcal{C}_\star^{\Lambda_N} = \mathcal{C}^{\Lambda_N} \cap \widehat{\mathcal{C}}^{\Lambda_N}$ (so $\mathcal{C}_\star^{\Lambda_N}$ is a version of $\mathcal{C}_*^{\Lambda_N}$, but it replaces $\widetilde{\mathcal{C}}^{\Lambda_N}$ with $\widehat{\mathcal{C}}^{\Lambda_N}$ in its definition).

Lemma 2.2.13. *Let $\Delta(N) = (N/4)^{-\alpha(\alpha')^2}$ and define $\{\widehat{h}_v^{(N)} : v \in \Lambda_N\}$ as in (2.2.20). Then there exists $C = C(\varepsilon) > 0$ such that $\mathbb{P}(o \in \mathcal{C}_\star^{\Lambda_N}) \leqslant CN^{-5}$.*

Proof. For $v \in \partial\Lambda_{N/2}$, let B_v be a translated copy of $\Lambda_{N/4}$ centered at v. Thus, for all

$u \in B_v$ we have $\widehat{h}^{(N)}_u = h_u + (N/4)^{-\alpha(\alpha')^2}$. Recall $m^{\diamond}_{N/4}(N/4)$ as in Lemma 2.2.11. By (2.2.3) and Lemma 2.2.11,

$$\mathbb{P}(v \in \mathcal{C}_\star^{\Lambda_N}) \leqslant m^{\diamond}_{N/4}(N/4) \leqslant CN^{-6}\,.$$

Hence, $\mathbb{P}(\partial\Lambda_{N/2} \cap \mathcal{C}_\star^{\Lambda_N} \neq \emptyset) \leqslant CN^{-5}$ by a simple union bound. Combined with Lemma 2.2.6 (and the simple observation that o cannot be connected to $\partial\Lambda_N$ by a path in $\mathcal{C}_\star^{\Lambda_N}$ if $\partial\Lambda_{N/2} \cap \mathcal{C}_\star^{\Lambda_N} = \emptyset$), this completes the proof of the lemma. □

Lemma 2.2.14. *There exists $C = C(\varepsilon) > 0$ such that $m_N \leqslant CN^{-3}$.*

Proof. A rough intuition behind the proof is as follows: the random field in each dyadic annulus has probability close to 1 to stop the event $\{o \in \mathcal{C}^{\Lambda_N}\}$ from occurring and thus altogether we get a polynomial upper bound with large power. In order to formalize the proof, we will apply Lemma 2.2.13 and employ a careful analysis to justify the "independence" among different scales.

Without loss of generality, let us only consider $N = 4^n$ for some $n \geqslant 1$. For each such N, define $\{\widehat{h}^{(N)}_v : v \in \Lambda_N\}$ as in (2.2.20) with $\Delta(N) = (N/4)^{-\alpha(\alpha')^2}$. Let $E_\ell = \{o \notin \mathcal{C}_\star^{\Lambda_{4^\ell}}\}$ and $E = \cap_{0.9n \leqslant \ell \leqslant n} E_\ell$. (Note that there is no containment relation among the events E_ℓ's, since each event depends on a different perturbation.) By Lemma 2.2.13, we see that $\mathbb{P}(E^c) \leqslant CN^{-3}$ for some $C = C(\varepsilon) > 0$ (whose value may be adjusted later in the proof). Write $\mathfrak{A}_\ell = \Lambda_{4^\ell} \setminus \Lambda_{4^{\ell-1}}$. For $0.9n \leqslant \ell \leqslant n$, let $\mathcal{F}_\ell = \sigma(h_v : v \in \Lambda_{4^\ell})$ and write

$$h_v = (|\mathfrak{A}_\ell|)^{-1} h_{\mathfrak{A}_\ell} + g_v \text{ for } v \in \mathfrak{A}_\ell\,, \tag{2.2.21}$$

where $\{g_v : v \in \mathfrak{A}_\ell\}$ is a mean-zero Gaussian process independent of $h_{\mathfrak{A}_\ell}$ and $\{g_v : v \in \mathfrak{A}_\ell\}$ for $0.9n \leqslant \ell \leqslant n$ are mutually independent (note that g_v's are linear combinations of a Gaussian process and their means and covariances can be easily computed). Let $\mathcal{F}'_\ell$ be the σ-field which contains every event in $\mathcal{F}_\ell$ that is independent of $h_{\mathfrak{A}_\ell}$ (so in particular

$\mathcal{F}_\ell \subseteq \mathcal{F}'_{\ell+1} \subseteq \mathcal{F}_{\ell+1}$). By monotonicity, there exists an interval I_ℓ measurable with respect to $\mathcal{F}'_\ell$ such that conditioned on $\mathcal{F}'_\ell$ we have $o \in \mathcal{C}^{\Lambda_{4^\ell}}$ if and only if $h_{\mathfrak{A}_\ell} \in I_\ell$. Let I'_ℓ be the maximal sub-interval of I_ℓ which shares the upper endpoint and with length $|I'_\ell| \leqslant \frac{|\mathfrak{A}_\ell| \cdot 4^{\alpha(\alpha')^2}}{4^{\alpha(\alpha')^2 \ell}}$. By our definition of E_ℓ, we see from (2.2.21) that conditioned on $\mathcal{F}'_\ell$ we have $\{o \in \mathcal{C}^{\Lambda_{4^\ell}}\} \cap E_\ell$ only if $h_{\mathfrak{A}_\ell} \in I'_\ell$. Thus, for $0.9n \leqslant \ell \leqslant n$,

$$\mathbb{P}(\{o \in \mathcal{C}^{\Lambda_{4^\ell}}\} \cap E_\ell \mid \mathcal{F}'_\ell) \leqslant \mathbb{P}(h_{\mathfrak{A}_\ell} \in I'_\ell)\,.$$

Combined with the fact that $\mathbb{V}\mathrm{ar}(h_{\mathfrak{A}_\ell}) = \varepsilon^2 |\mathfrak{A}_\ell|$, this gives that

$$\mathbb{P}(\{o \in \mathcal{C}^{\Lambda_{4^\ell}}\} \cap E_\ell \mid \mathcal{F}'_\ell) \leqslant \frac{C}{4^{\ell(\alpha(\alpha')^2 - 1)}}\,.$$

Since $\{o \in \mathcal{C}^{\Lambda_{4^n}}\} \cap E = \cap_{\ell=0.9n}^{n} (\{o \in \mathcal{C}^{\Lambda_{4^\ell}}\} \cap E_\ell)$ and since $\{o \in \mathcal{C}^{\Lambda_{4^\ell}}\} \cap E_\ell$ is $\mathcal{F}_\ell$-measurable (and thus is $\mathcal{F}'_{\ell+1}$-measurable), we deduce that $\mathbb{P}(\{o \in \mathcal{C}^{\Lambda_N}\} \cap E) \leqslant CN^{-3}$. Combined with the fact that $\mathbb{P}(E^c) \leqslant CN^{-3}$, it completes the proof of the lemma. □

Proof of Theorem 2.2.1. Let $N_0 = N_0(\varepsilon)$ be chosen later. For $B \in \mathcal{B}(N, N_0)$, we say B is open if $\mathcal{C}^{B^{\mathrm{large}}} \cap B \neq \emptyset$. Clearly, this percolation process satisfies the $(N, N_0, 4, p)$-condition where

$$p = \mathbb{P}(\mathcal{C}^{B^{\mathrm{large}}} \cap B \neq \emptyset) \leqslant N_0^2 m_{N_0/2} \leqslant C N_0^{-1} \text{ for } C = C(\varepsilon) > 0\,. \tag{2.2.22}$$

(The last transition above follows from Lemma 2.2.14.) In addition, we note that in order for $o \in \mathcal{C}^{\Lambda_N}$, it is necessary that there exists an open lattice animal on $B \in \mathcal{B}(N, N_0)$ with size at least $\frac{N}{10N_0}$. Now, choosing N_0 sufficiently large (so that p is sufficiently small, by (2.2.22)) and applying Lemma 2.2.10 completes the proof. □

2.3. Exponential decay at positive temperatures

In this section, we prove Theorem 2.1.1 for the case of $T > 0$. Our proof method follows the basic framework presented in Section 2.2 for the case of $T = 0$, which applies the result in

[1] in a crucial way. However, there seem to be significant additional obstacles due to the randomness of Ising measures at positive temperatures. For $T = 0$, it suffices to consider the ground state which is unique with probability 1, and thus ground states with different boundary conditions and external fields are naturally coupled together. In the case of $T > 0$, on the one hand we try to carry out our analysis with validity for all reasonable (e.g., for all monotone couplings) couplings of Ising measures whenever possible (see Section 2.3.1); on the other hand it seems necessary to construct a coupling with some desirable properties in order to apply [1] (see Section 2.3.2). Both of these require some new ideas as well as some delicate treatment.

Organization for the rest of this section is as follows. In Section 2.3.1, we verify the hypothesis in [1] via a perturbation argument and thereby prove that under any monotone coupling for Ising spins with plus/minus boundary conditions, the intrinsic distance for the induced graph on vertices with disagreements has dimension strictly larger than 1. The proof method is inspired by that of Proposition 2.2.2, but the implementation is largely different with new tricks involved. In Section 2.3.2, we introduce the notion of adaptive admissible coupling and a multi-scale construction of an adaptive admissible coupling is then given in Section 2.3.3. In Section 2.3.3, we then introduce another perturbation argument, using which we analyze our adaptive admissible coupling in Section 2.3.3 and prove a crucial estimate in Lemma 2.3.17. In Section 2.3.4, we provide the proof of Theorem 2.1.1 for $T > 0$, which requires to employ an admissible coupling such that the disagreement percolates to the boundary.

2.3.1. Intrinsic distance on disagreements via a perturbation argument

For any $A \subseteq \mathbb{Z}^2$, we continue to denote by $d_A(\cdot, \cdot)$ the intrinsic distance on A, i.e., the graph distance on the induced subgraph on A. Let $\sigma^{\Lambda_N, \pm}$ be spins sampled according to $\mu^{\Lambda_N, \pm}$. We will continue to use repeatedly the standard monotonicity properties of the Ising model with respect to external fields and boundary conditions (c.f. [3, Section 2.2] for detailed discussions). Let π be a monotone coupling of $\mu^{\Lambda_N, \pm}$ (that is, under π we have

$\sigma^{\Lambda_N,+} \geqslant \sigma^{\Lambda_N,-}$) and let

$$\mathcal{C}^{\Lambda_N} = \mathcal{C}^{\Lambda_N,\pi} = \{v \in \Lambda_N : \sigma_v^{\Lambda_N,+} > \sigma_v^{\Lambda_N,-}\}\,. \tag{2.3.1}$$

(Note that π depends on the random field h.) In addition, denote by $\mathbb{P} \otimes \pi$ the joint measure of the external fields and the spin configurations (similar notations also apply below). The following proposition is the major goal of this section.

Proposition 2.3.1. *There exist $\alpha = \alpha(\varepsilon, \beta) > 1$, $\kappa = \kappa(\varepsilon, \beta) > 0$ such that the following holds. For all $0 < c \leqslant 1$, there exists $N_0 = N_0(\varepsilon, \beta, c)$ such that for all $N \geqslant N_0$ and $1 \leqslant N_1 \leqslant N_2 \leqslant N/2$ with $N_2 - N_1 \geqslant N^c$ the following holds for all monotone coupling π of $\mu^{\Lambda_N,\pm}$:*

$$\mathbb{P} \otimes \pi(d_{\mathcal{C}^{\Lambda_N}}(\partial\Lambda_{N_1}, \partial\Lambda_{N_2}) \leqslant (N_2 - N_1)^{\alpha}) \leqslant \kappa^{-1} e^{-N^{\kappa c}}\,. \tag{2.3.2}$$

Remark 2.3.2. (1) The preceding proposition is analogous to Proposition 2.2.2. In the present case, it is crucial that the result holds for all monotone couplings (note that the intrinsic distance may depend on the coupling), so that we can apply it to couplings which we construct later.

(2) In Proposition 2.3.1, we introduce parameters N_1, N_2 (as opposed to $N_1 = N/4$ and $N_2 = N/2$ in Proposition 2.2.2) for convenience of later applications. The condition that $N_2 - N_1 \geqslant N^c$ is just to ensure that the decay in probability absorbs the number of choices for starting and ending points of the shortest path. This slight extension does not introduce complication to the proof.

The proof of Proposition 2.3.1 again crucially relies on the result of [1]. In order to apply [1], the following lemma (analogous to Lemma 2.2.8) is a key ingredient. For any annulus $\mathcal{A}$ and $\mathcal{C} \subseteq \mathbb{Z}^2$, we continue to denote by $\mathrm{Cross}_{\mathrm{hard}}(\mathcal{A}, \mathcal{C})$ the event that there is a contour in $\mathcal{C}$ which separates the inner and outer boundaries of $\mathcal{A}$. Let

$$\mathcal{E}^{\pm} = \mathcal{E}^{\pm}_N = \mathrm{Cross}_{\mathrm{hard}}(\Lambda_{N/8} \setminus \Lambda_{N/32}, \{v \in \Lambda_N : \sigma_v^{\Lambda_N, \pm} = \pm 1\})\,. \tag{2.3.3}$$

Lemma 2.3.3. *There exists $\delta = \delta(\varepsilon, \beta) > 0$ such that for all $N \geqslant 32$*

$$\min\{\mathbb{P} \otimes \mu^{\Lambda_N,+}(\mathcal{E}^+), \mathbb{P}(\textstyle\sum_{v \in \Lambda_{N/8}}(\langle \sigma_v^{\Lambda_N,+}\rangle_{\mu^{\Lambda_N,+}} - \langle \sigma_v^{\Lambda_N,-}\rangle_{\mu^{\Lambda_N,-}}) > 10^{-3}N)\} \leqslant 1 - \delta\,.$$

In particular, $\mathbb{P} \otimes \pi(\mathrm{Cross}_{\mathrm{hard}}(\Lambda_{N/8} \setminus \Lambda_{N/32}, \mathcal{C}^{\Lambda_N})) \leqslant 1 - \delta$ for all monotone coupling π of $\mu^{\Lambda_N,\pm}$.

Remark 2.3.4. By Lemma 2.3.3, either of the following holds: (i) with positive probability the plus-spins with respect to the plus boundary condition does not separate the boundaries of an annulus (this is a stronger than what was proved in Case 1 in the proof of Lemma 2.2.8); (ii) with positive probability the expected number of disagreements (averaged over the Ising measures) is small (this corresponds to Case 2 in the proof of Lemma 2.2.8). Assuming either property, we are able to derive a uniform bound on crossing probabilities for disagreements under any monotone coupling.

After establishing exponential decay, then it is clear that Property (ii) holds. In addition, we know that with overwhelming probability away from the boundary the spin configurations with plus and minus boundary conditions agree with each other. Therefore, by symmetry and planar duality we see that Property (i) also holds.

A perturbative analysis

Before proving Lemma 2.3.3, we need some preparational work on a certain perturbative analysis. This is analogous to Lemma 2.2.5, which has been applied twice in the case of $T = 0$: in the proof of Lemma 2.2.8 and the proof of Lemma 2.2.11. For $T > 0$, it is more complicated and thus we provide two separate versions of perturbative analysis, both of which are proved via keeping track of the free energy. The first version is presented in

Lemma 2.3.5 in the present section (for the application in Lemma 2.3.3), and the second version is presented in Section 2.3.3 (for the application in Lemma 2.3.17).

For any set $\Lambda \subseteq \mathbb{Z}^2$ and a configuration $\tau \in \{-1,1\}^{\partial\Lambda}$, analogous to (2.1.1) we can define the Hamiltonian on Λ with boundary condition τ and external field $\{h_v\}$ by:

$$H^{\Lambda,\tau}(\sigma) = -\Big(\sum_{u\sim v, u,v\in\Lambda} \sigma_u\sigma_v + \sum_{u\sim v, u\in\Lambda, v\in\partial\Lambda} \sigma_u\tau_v + \sum_{u\in\Lambda} \sigma_u h_u\Big) \text{ for } \sigma \in \{-1,1\}^{\Lambda}\,. \quad (2.3.4)$$

We can then analogously define the Ising measure $\mu^{\Lambda,\tau}$ by assigning probability to $\sigma \in \{-1,1\}^{\Lambda}$ proportional to $e^{-\beta H^{\Lambda,\tau}(\sigma)}$. In addition, we define the corresponding log-partition-function (it is the negative of the free energy; in our analysis, it seems cleaner to work with the log-partition-function so not to be confused by the negative sign)

$$F^{\Lambda,\tau} = \frac{1}{\beta}\log\Big(\sum_{\sigma\in\{-1,1\}^{\Lambda}} e^{-\beta H^{\Lambda,\tau}(\sigma)}\Big). \quad (2.3.5)$$

For simplicity, we will only consider $N = 2^n$ for $n \geqslant 10$. For $\Delta > 0$, $\Delta' \geqslant 0$ and $0 \leqslant t \leqslant 1$, we will consider the following perturbed field in this section (which is increasing in t):

$$h_v^{(t)} = h_v^{(t,N)} = \begin{cases} h_v + \Delta', & \text{for } v \in \Lambda_N \setminus \Lambda_{N/8}\,, \\ h_v + t\Delta, & \text{for } v \in \Lambda_{N/8}\,. \end{cases} \quad (2.3.6)$$

(We draw the reader's attention to that t appeared in the definition of $h_v^{(t)}$ only for $v \in \Lambda_{N/8}$, and that $h^{(0)} \neq h$ if $\Delta' > 0$. The perturbation in (2.3.6) is more subtle than that in (2.2.7), for the reason that we wish to take advantage of (2.3.17) below later with a judicious choice of Δ'.) Let $\mu^{\Lambda_N,\pm,t}$ be Ising measures with plus/minus boundary conditions and external field $\{h_v^{(t)} : v \in \Lambda_N\}$. In addition, let $H^{\Lambda_N,\pm,t}$ be the corresponding Hamiltonians, let $F^{\Lambda_N,\pm,t}$ be the corresponding log-partition-functions, and let $\sigma^{\Lambda_N,\pm,t}$ be spin configurations sampled according to $\mu^{\Lambda_N,\pm,t}$.

For notation convenience, for any set $\Gamma \subseteq \mathbb{Z}^2$, let S_Γ be the collection of vertices which are not in Γ and are separated by Γ from ∞ on $\mathbb{Z}^2$ (i.e., the collection of vertices that are enclosed by Γ).

Let $S \subseteq \Lambda_N$ be a subset which contains $\Lambda_{N/8}$ and let $\Gamma = \partial S$ (thus we have $S \subseteq S_\Gamma$). For any $\tau \in \{-1,1\}^\Gamma$, we denote by $\mu^{S,\tau,t}$ the Ising measure on S with boundary condition τ and external field $\{h_v^{(t)} : v \in S\}$. In addition, let $H^{S,\tau,t}$ be the Hamiltonian for the corresponding Ising spin, and let $F^{S,\tau,t}$ be the corresponding log-partition-function. Also, we let $\sigma^{S,\tau,t}$ be the spin configuration sampled according to $\mu^{S,\tau,t}$. For later applications, it would be useful to consider the log-partition-function restricted to a subset of configurations. To this end, we define

$$F_\Omega^{S,\tau,t} = \frac{1}{\beta}\log\big(\sum_{\sigma\in\Omega} e^{-\beta H^{S,\tau,t}(\sigma)}\big) \text{ for } \Omega \subseteq \{-1,1\}^S. \tag{2.3.7}$$

In addition, for any measure $\mu^{S,\tau,t}$, we define $\mu_\Omega^{S,\tau,t}$ to be a measure such that

$$\mu_\Omega^{S,\tau,t}(\sigma) = (\mu^{S,\tau,t}(\Omega))^{-1}\mu^{S,\tau,t}(\sigma) \text{ for } \sigma \in \Omega\,.$$

(We draw readers' attention to that $\mu^{S,\tau,t}(\Omega)$ is the total measure of Ω under $\mu^{S,\tau,t}$ and thus is a number, and that $\mu_\Omega^{S,\tau,t}$ is the measure $\mu^{S,\tau,t}$ conditioned on the occurrence of Ω.) For convenience, we let $\sigma_\Omega^{S,\tau,t}$ be the spin configuration sampled according to $\mu_\Omega^{S,\tau,t}$. Further, define (note that below we sum over $v \in \Lambda_{N/32}$ as opposed to $v \in S$)

$$m_\Omega^{S,\tau,t} = \sum_{v\in\Lambda_{N/32}} \langle \sigma_{\Omega,v}^{S,\tau,t}\rangle_{\mu_\Omega^{S,\tau,t}}\,. \tag{2.3.8}$$

For notation convenience, we write $m^{S,\tau,t} = m_\Omega^{S,\tau,t}$ if $\Omega = \{-1,1\}^S$. We say $\Omega \subseteq \{-1,1\}^S$ is an increasing set if $\sigma \in \Omega$ implies that $\sigma' \in \Omega$ provided $\sigma' \geqslant \sigma$, and we say Ω is a decreasing set if Ω^c is an increasing set. In what follows, we consider $\tau^+, \tau^- \in \{-1,1\}^\Gamma$ such that $\tau^+ \geqslant \tau^-$.

Lemma 2.3.5. *Quench on the external field $\{h_v\}$. We have that for any increasing set*

$\Omega^+ \subseteq \{-1,1\}^S$ *and any decreasing set* $\Omega^- \subseteq \{-1,1\}^S$

$$\Delta \int_0^1 (m_{\Omega^+}^{S,\tau^+,t} - m_{\Omega^-}^{S,\tau^-,t}) dt \leqslant 8 \sum_{v\in\Gamma} (\tau_v^+ - \tau_v^-) - \frac{1}{\beta}\big(\log \mu^{S,\tau^+,0}(\Omega^+) + \log \mu^{S,\tau^-,1}(\Omega^-)\big)\,.$$

Proof. The proof is done via keeping track of the change on the difference of log-partition-functions with respect to different boundary conditions when we perturb the external field. In **Step 1**, we bound such difference from above by the number of disagreements on boundary conditions; in **Step 2** we bound such difference from below by the expected number of disagreements, with a caveat that we use the notion of "restricted" log-partition-functions as in (2.3.7); in **Step 3**, we address the caveat by linking the two notions of log-partition-functions.

Step 1. We will prove (below the equality is obvious since $\tau^+ \geqslant \tau^-$)

$$(F^{S,\tau^+,1} - F^{S,\tau^-,1}) - (F^{S,\tau^+,0} - F^{S,\tau^-,0}) \leqslant 16 \cdot \#\{v \in \Gamma : \tau_v^+ \neq \tau_v^-\} = 8 \sum_{v\in\Gamma} (\tau_v^+ - \tau_v^-)\,. \quad (2.3.9)$$

(Here we use $\#A$ to denote the cardinality of A for a finite set A. We switch from the more compact notation $|A|$ to $\#A$ in this section, as we wish to avoid somewhat awkward notation when $|$ is followed by another $|$ which means "conditioned on".) Since each vertex has 4 neighbors in $\mathbb{Z}^2$, a straightforward computation gives that

$$F^{S,\tau^+,1} - F^{S,\tau^-,1} = \frac{1}{\beta} \log \frac{\sum_\sigma e^{-\beta H^{S,\tau^+,1}(\sigma)}}{\sum_\sigma e^{-\beta H^{S,\tau^-,1}(\sigma)}} \leqslant \frac{1}{\beta} \log e^{8\beta \cdot \#\{v\in\Gamma : \tau_v^+ \neq \tau_v^-\}}$$
$$\leqslant 8 \cdot \#\{v \in \Gamma : \tau_v^+ \neq \tau_v^-\}\,.$$

Similarly, we have that $F^{S,\tau^+,0} - F^{S,\tau^-,0} \geqslant -8 \cdot \#\{v \in \Gamma : \tau_v^+ \neq \tau_v^-\}$. This proves (2.3.9).

Step 2. We will prove

$$(F_{\Omega^+}^{S,\tau^+,1} - F_{\Omega^-}^{S,\tau^-,1}) - (F_{\Omega^+}^{S,\tau^+,0} - F_{\Omega^-}^{S,\tau^-,0}) \geqslant \Delta \int_0^1 (m_{\Omega^+}^{S,\tau^+,t} - m_{\Omega^-}^{S,\tau^-,t}) dt\,. \quad (2.3.10)$$

We write

$$(F^{S,\tau^+,1}_{\Omega^+} - F^{S,\tau^-,1}_{\Omega^-}) - (F^{S,\tau^+,0}_{\Omega^+} - F^{S,\tau^-,0}_{\Omega^-}) = (F^{S,\tau^+,1}_{\Omega^+} - F^{S,\tau^+,0}_{\Omega^+}) - (F^{S,\tau^-,1}_{\Omega^-} - F^{S,\tau^-,0}_{\Omega^-})\,. \quad (2.3.11)$$

Thus, we get that

$$F^{S,\tau^+,1}_{\Omega^+} - F^{S,\tau^+,0}_{\Omega^+} = \int_0^1 \frac{dF^{S,\tau^+,t}_{\Omega^+}}{dt} dt, \qquad F^{S,\tau^-,1}_{\Omega^-} - F^{S,\tau^-,0}_{\Omega^-} = \int_0^1 \frac{dF^{S,\tau^-,t}_{\Omega^-}}{dt} dt\,. \quad (2.3.12)$$

Since $\frac{dF^{S,\tau^+,t}_{\Omega^+}}{dt} = \sum_{v\in\Lambda_{N/8}} \Delta\langle\sigma^{S,\tau^+,t}_{\Omega^+,v}\rangle_{\mu^{S,\tau^+,t}_{\Omega^+}}$ and $\frac{dF^{S,\tau^-,t}_{\Omega^-}}{dt} = \sum_{v\in\Lambda_{N/8}} \Delta\langle\sigma^{S,\tau^-,t}_{\Omega^-,v}\rangle_{\mu^{S,\tau^-,t}_{\Omega^-}}$, we see

$$\frac{dF^{S,\tau^+,t}_{\Omega^+}}{dt} - \frac{dF^{S,\tau^-,t}_{\Omega^-}}{dt} \geqslant \sum_{v\in\Lambda_{N/32}} \Delta(\langle\sigma^{S,\tau^+,t}_{\Omega^+,v}\rangle_{\mu^{S,\tau^+,t}_{\Omega^+}} - \langle\sigma^{S,\tau^-,t}_{\Omega^-,v}\rangle_{\mu^{S,\tau^-,t}_{\Omega^-}}) = \Delta m^{S,\tau^+,t}_{\Omega^+} - \Delta m^{S,\tau^-,t}_{\Omega^-}\,,$$

where the inequality follows from the fact that

$$\langle\sigma^{S,\tau^+,t}_{\Omega^+,v}\rangle_{\mu^{S,\tau^+,t}_{\Omega^+}} \geqslant \langle\sigma^{S,\tau^+,t}_{v}\rangle_{\mu^{S,\tau^+,t}} \geqslant \langle\sigma^{S,\tau^-,t}_{v}\rangle_{\mu^{S,\tau^-,t}} \geqslant \langle\sigma^{S,\tau^-,t}_{\Omega^-,v}\rangle_{\mu^{S,\tau^-,t}_{\Omega^-}} \text{ for all } v \in S\,.$$

In the preceding display, the first and the third inequalities follow from FKG inequality [63] and the second inequality follows from monotonicity. Combined with (2.3.12) and (2.3.11), it yields (2.3.10).

Step 3. From definitions as in (2.3.5) and (2.3.7), we see that

$$F^{S,\tau^+,1} - F^{S,\tau^+,1}_{\Omega^+} = -\frac{1}{\beta}\log\mu^{S,\tau^+,1}(\Omega^+)\,, \quad (2.3.13)$$

and similar equalities hold for other combinations of boundary conditions, external fields and $\Omega^{\pm}$.

Combining (2.3.9), (2.3.10) and (2.3.13), we complete the proof of the lemma. □

A lower bound on the intrinsic distance

Denote by $\mathcal{V}^{\sigma,\pm} = \{v \in S : \sigma_v = \pm 1\}$ for $S \subseteq \Lambda_N$ and $\sigma \in \{-1,1\}^S$. For any $S \supset \Lambda_{N/8}$, define

$$\Omega^{\pm} = \Omega^{\pm}(S) = \{\sigma \in \{-1,1\}^S : \mathrm{Cross}_{\mathrm{hard}}(\Lambda_{N/8} \setminus \Lambda_{N/32}, \mathcal{V}^{\sigma,\pm}) \text{ occurs }\}\,. \tag{2.3.14}$$

We see that Ω^+ is an increasing set and Ω^- is a decreasing set. For $A \subseteq \Lambda \subseteq \mathbb{Z}^2$ and $\sigma \in \{-1,1\}^\Lambda$, we denote by σ_A the restriction of σ on A. Let $r > 0$ be a constant chosen later. Recall (2.3.6). Let $\Delta = \frac{10^{10} r^8}{N(\beta \wedge 1)}$ and $\Delta' = t^*\Delta$ for $0 \leqslant t^* \leqslant 1$ to be chosen.

Lemma 2.3.6. *For any $p, r > 0$, there exists $c = c(\varepsilon, p, r, \beta) > 0$ such that for any event E_N with $\mathbb{P}(\{h_v^{(t)} : v \in \Lambda_N\} \in E_N) \geqslant p$ for some $0 \leqslant t, t^* \leqslant 1$, we have that $\mathbb{P}(\{h_v : v \in \Lambda_N\} \in E_N) \geqslant c$.*

Proof. The proof is an adaption of Lemma 2.2.7 except for minimal notation change, and thus we omit further details. □

Proof of Lemma 2.3.3. The proof shares similarity with that of Lemma 2.2.8, but the present proof is substantially more involved. We first provide a heuristic outline of the proof, and we will not be precise on notations or unimportant constants in this informal description. The statement will follow immediately if the probability for existence of a plus contour with respect to plus boundary condition is strictly less than 1, and thus we suppose otherwise (formally, we suppose (2.3.15) below). We wish to compare the number of disagreements in $\Lambda_{N/32}$ with that in $\mathcal{A}_{N/2}$. To this end, it will be useful to consider the "enhanced" disagreements in $\Lambda_{N/32}$ (that is, when we pose plus and minus boundary conditions on $\partial\Lambda_{N/8}$ instead of $\partial\Lambda_N$; the word "enhanced" is chosen because by monotonicity the enhanced disagreements stochastically dominate the original disagreements). We now compare the enhanced disagreements in $\Lambda_{N/32}$ and disagreements in $\mathcal{A}_{N/2}$ in both directions.

- The "$\leqslant$" direction (**Step 1** below): This is where plus (minus) contours come into play. Conditioned on existence of plus and minus contours, the disagreements in $\Lambda_{N/32}$ stochastically dominate the enhanced disagreements. In addition, by Lemma 2.3.5, the number of disagreements in $\Lambda_{N/32}$ is upper bounded by that in $\mathcal{A}_{N/2}$ (up to an additive term that is related to the probability of existence of plus/minus contours, which we will address later). Altogether, we get that the number of enhanced disagreements in $\Lambda_{N/32}$ is upper bounded by the number of disagreements in $\mathcal{A}_{N/2}$ (see (2.3.25)).

- The "$\geqslant$" direction (**Step 2** below): The set of disagreements in $\mathcal{A}_{N/2}$ is dominated by a union of constant copies of enhanced disagreements in $\Lambda_{N/32}$, where the number of disagreements in all these copies are independent of the enhanced disagreements in $\Lambda_{N/32}$ (but not of each other). This implies that with positive probability, the number of enhanced disagreements in $\Lambda_{N/32}$ is larger (up to a constant factor) than the number of disagreements in $\mathcal{A}_{N/2}$ (see (2.3.29)).

Now, if we choose the constants appropriately, we will see that the preceding two scenarios will occur simultaneously with positive probability, which yield bounds in two directions that "almost" contradict each other. These events can only happen concurrently if the logarithmic term we ignored earlier (which becomes $\frac{N}{2\beta}$ in (2.3.25)) plays a significant role. But this can happen only when the typical number of enhanced disagreements is at most of order N, in which case an application of Markov's inequality (see (2.3.21)) yields the desired lemma.

We next carry out the proof formally, where we slightly shuffle the order of arguments: we first show that if the typical number of enhanced disagreements is at most of order N (see (2.3.18)), then the lemma holds. Next, we prove (2.3.18) (which is the main challenge) by contradiction, via the aforementioned two directional comparisons.

For convenience of notation, write

$$\mathcal{E}^{\pm,t} = \mathrm{Cross}_{\mathrm{hard}}(\Lambda_{N/8} \setminus \Lambda_{N/32}, \mathcal{V}^{\sigma^{\Lambda_N,\pm,t},\pm})\,.$$

We suppose that

$$\min_{0\leqslant t\leqslant 1}\{\mathbb{P}\otimes\mu^{\Lambda_N,+,t}(\mathcal{E}^{+,t}),\mathbb{P}\otimes\mu^{\Lambda_N,-,t}(\mathcal{E}^{-,t})\}\geqslant 1-r^{-4}10^{-10}\,. \tag{2.3.15}$$

Otherwise Lemma 2.3.3 follows from Lemma 2.3.6 (since under any monotone coupling we have $\mathrm{Cross}_{\mathrm{hard}}(\Lambda_{N/8}\setminus\Lambda_{N/32},\mathcal{C}^{\Lambda_N})\subseteq\mathcal{E}_N^{+}\cap\mathcal{E}_N^{-}$, where $\mathcal{E}_N^{\pm}$ is defined in (2.3.3)). We remark that by monotonicity the preceding inequality is equivalent to $\min\{\mathbb{P}\otimes\mu^{\Lambda_N,+,0}(\mathcal{E}^{+,0}),\mathbb{P}\otimes\mu^{\Lambda_N,-,1}(\mathcal{E}^{-,1})\}\geqslant 1-r^{-4}10^{-10}$.

Let $\mathcal{E}^\star=\{\mu^{\Lambda_N,+,0}(\mathcal{E}^{+,0})\geqslant 99/100\}\cap\{\mu^{\Lambda_N,-,1}(\mathcal{E}^{-,1})\geqslant 99/100\}$ be an event measurable with respect to the Gaussian field. By (2.3.15), we see that

$$\mathbb{P}(\mathcal{E}^\star)\geqslant 1-10^{-2}r^{-4}\,. \tag{2.3.16}$$

Let $t^*\in[0,1]$ be such that

$$\inf\{\theta:\mathbb{P}(m^{\Lambda_{N/8},+,t^*}-m^{\Lambda_{N/8},-,t^*}\geqslant\theta)\leqslant 1/2r\}=\theta^*\,, \tag{2.3.17}$$

where $\theta^*=\min_{0\leqslant t\leqslant 1}\inf\{\theta:\mathbb{P}(m^{\Lambda_{N/8},+,t}-m^{\Lambda_{N/8},-,t}\geqslant\theta)\leqslant 1/2r\}$. We claim that

$$\theta^*\leqslant 10^{-3}r^{-1}N\,. \tag{2.3.18}$$

We first show that (2.3.18) implies the lemma. For any box A, let A^{Big} be the concentric box of A with side length 4 times that of A. Let r be a large enough constant so that we can write $\Lambda_{N/8}=\cup_{i=1}^r A_i$, where A_i is a copy of $\Lambda_{N/32}$ and A_i's are disjoint such that $A_i^{\mathrm{Big}}\subseteq\Lambda_N$ for $1\leqslant i\leqslant r$. By monotonicity, we see that for each $1\leqslant i\leqslant r$

$$\begin{aligned}&\mathbb{P}\Big(\sum_{v\in A_i}(\langle\sigma_v^{\Lambda_N,+,t^*}\rangle_{\mu^{\Lambda_N,+,t^*}}-\langle\sigma_v^{\Lambda_N,-,t^*}\rangle_{\mu^{\Lambda_N,-,t^*}})>\theta^*\Big)\\&\leqslant\mathbb{P}\Big(\sum_{v\in A_i}(\langle\sigma_v^{A_i^{\mathrm{Big}},+,t^*}\rangle_{\mu^{A_i^{\mathrm{Big}},+,t^*}}-\langle\sigma_v^{A_i^{\mathrm{Big}},-,t^*}\rangle_{\mu^{A_i^{\mathrm{Big}},-,t^*}})>\theta^*\Big)\leqslant(2r)^{-1}\,,\end{aligned}$$

where the last inequality holds due to our choice of t^* as in (2.3.17) and $\Delta' = t^*\Delta$ (thus $h_v^{(t^*)} = h_v + \Delta'$ for $v \in \Lambda_N$). Hence, a simple union bound gives that

$$\mathbb{P}\Big(\sum_{v\in\Lambda_{N/8}} (\langle\sigma_v^{\Lambda_N,+,t^*}\rangle_{\mu^{\Lambda_N,+,t^*}} - \langle\sigma_v^{\Lambda_N,-,t^*}\rangle_{\mu^{\Lambda_N,-,t^*}}) \leqslant r\theta^*\Big) \geqslant \frac{1}{2}\,. \tag{2.3.19}$$

By Lemma 2.3.6, we get that

$$\mathbb{P}\Big(\sum_{v\in\Lambda_{N/8}} (\langle\sigma_v^{\Lambda_N,+}\rangle_{\mu^{\Lambda_N,+}} - \langle\sigma_v^{\Lambda_N,-}\rangle_{\mu^{\Lambda_N,-}}) > r\theta^*\Big) \leqslant 1-\delta \text{ for } \delta = \delta(\varepsilon,\beta,r) > 0\,. \tag{2.3.20}$$

Note that $2\langle\#(\mathcal{C}^{\Lambda_N} \cap \Lambda_{N/8})\rangle_\pi = \sum_{v\in\Lambda_{N/8}}(\langle\sigma_v^{\Lambda_N,+}\rangle_{\mu^{\Lambda_N,+}} - \langle\sigma_v^{\Lambda_N,-}\rangle_{\mu^{\Lambda_N,-}})$ on each instance of the Gaussian field for any monotone coupling π of $\mu^{\Lambda_N,\pm}$. Therefore, on each instance of Gaussian field (which occurs with probability at least δ) such that $\sum_{v\in\Lambda_{N/8}}(\langle\sigma_v^{\Lambda_N,+}\rangle_{\mu^{\Lambda_N,+}} - \langle\sigma_v^{\Lambda_N,-}\rangle_{\mu^{\Lambda_N,-}}) \leqslant r\theta^*$, we apply Markov's inequality and get that

$$\pi(\mathrm{Cross}_{\mathrm{hard}}(\Lambda_{N/8}\setminus\Lambda_{N/32}, \mathcal{C}^{\Lambda_N})) \leqslant \pi(\#(\mathcal{C}^{\Lambda_N}\cap\Lambda_{N/8}) \geqslant \tfrac{N}{32}) \leqslant \frac{\theta^* r}{N/32} \leqslant \frac{1}{2}\,, \tag{2.3.21}$$

where the last inequality follows from (2.3.18). This implies that $\mathbb{P}\otimes\pi(\mathrm{Cross}_{\mathrm{hard}}(\Lambda_{N/8}\setminus\Lambda_{N/32}, \mathcal{C}^{\Lambda_N})) \leqslant 1-\delta/2$, completing the proof of Lemma 2.3.3 (combined with (2.3.20)).

It remains to prove (2.3.18). Suppose that (2.3.18) does not hold. We will derive a contradiction, using the following two steps.

Step 1. We refer to Figure 2.3 for an illustration of geometric setup in this step. Fix $N/4 \leqslant k \leqslant N/2$. Write $S = \Lambda_k$ and $\Gamma = \partial S$. We first quench on the Gaussian field and also condition on

$$(\sigma^{\Lambda_N,+,1})_\Gamma = \tau^+ \text{ and } (\sigma^{\Lambda_N,-,0})_\Gamma = \tau^- \text{ where } \tau^\pm \in \{-1,1\}^\Gamma \text{ and } \tau^+ \geqslant \tau^-\,. \tag{2.3.22}$$

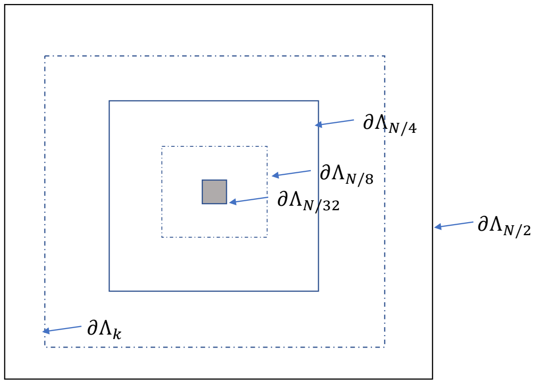

Figure 2.3: Illustrations for geometric setup in Step 1 of Lemma 2.3.3. The picture on the left illustrates the setup for derivation of (2.3.23), where we bound disagreements in the grey square by disagreements on $\partial\Lambda_k$ (the larger dot-line boundary). The picture on the right illustrates the setup for derivation of (2.3.24): by FKG conditioned on plus (respectively minus) contour (drawn in dots in the picture) the magnetization on the grey box is pushed up (respectively down); this allows us to compare the disagreements and enhanced disagreements.

Applying Lemma 2.3.5, we get that (recall $\Omega^\pm = \Omega^\pm(S)$ as in (2.3.14))

$$\Delta \int_0^1 (m_{\Omega^+}^{S,\tau^+,t} - m_{\Omega^-}^{S,\tau^-,t})dt \leqslant 8\sum_{v\in\Gamma}(\tau_v^+ - \tau_v^-) - \frac{1}{\beta}\big(\log \mu^{S,\tau^+,0}(\Omega^+) + \log \mu^{S,\tau^-,1}(\Omega^-)\big)\,. \tag{2.3.23}$$

Conditioned on $\sigma^{S,\tau^+,t} \in \Omega^+$, let $\mathfrak{C} \subseteq \mathcal{V}^{\sigma^{S,\tau^+,t},+} \cap (\Lambda_{N/8} \setminus \Lambda_{N/32})$ be the outmost contour which surrounds $\Lambda_{N/32}$. Note that $\mathfrak{C} = \Gamma'$ is measurable with respect to $\{\sigma_v^{S,\tau^+,t} : v \in S_{\Gamma'}^c\}$. Thus, by monotonicity of Ising model we see that $(\sigma^{S,\tau^+,t})_{\Lambda_{N/32}}$ conditioned on $\mathfrak{C} = \Gamma'$ stochastically dominates $(\sigma^{\Lambda_{N/8},+,t})_{\Lambda_{N/32}}$. A similar analysis applies to $(\sigma^{S,\tau^-,t})_{\Lambda_{N/32}}$. Combined with (2.3.23), it yields that

$$\Delta \int_0^1 (m^{\Lambda_{N/8},+,t} - m^{\Lambda_{N/8},-,t})dt \leqslant 8\sum_{v\in\Gamma}(\tau_v^+ - \tau_v^-) - \frac{1}{\beta}\big(\log \mu^{S,\tau^+,0}(\Omega^+) + \log \mu^{S,\tau^-,1}(\Omega^-)\big)\,. \tag{2.3.24}$$

Define $\mathcal{E}_{\Gamma,+} = \{\tau^+ : \mu^{S,\tau^+,0}(\Omega^+) \geqslant 3/4\}$ and $\mathcal{E}_{\Gamma,-} = \{\tau^- : \mu^{S,\tau^-,1}(\Omega^-) \geqslant 3/4\}$. Thus,

$$\begin{aligned}\mu^{\Lambda_N,+,0}(\mathcal{E}^{+,0}) &= \mu^{\Lambda_N,+,0}(\mathcal{E}^{+,0} \mid (\sigma^{\Lambda_N,+,0})_\Gamma \in \mathcal{E}_{\Gamma,+})\mu^{\Lambda_N,+,0}((\sigma^{\Lambda_N,+,0})_\Gamma \in \mathcal{E}_{\Gamma,+}) \\ &\quad + \mu^{\Lambda_N,+,0}(\mathcal{E}^{+,0} \mid (\sigma^{\Lambda_N,+,0})_\Gamma \notin \mathcal{E}_{\Gamma,+})\mu^{\Lambda_N,+,0}((\sigma^{\Lambda_N,+,0})_\Gamma \notin \mathcal{E}_{\Gamma,+}) \\ &\leqslant \mu^{\Lambda_N,+,0}((\sigma^{\Lambda_N,+,0})_\Gamma \in \mathcal{E}_{\Gamma,+}) + \tfrac{3}{4}\mu^{\Lambda_N,+,0}((\sigma^{\Lambda_N,+,0})_\Gamma \notin \mathcal{E}_{\Gamma,+})\,.\end{aligned}$$

Since $\mu^{\Lambda_N,+,0}(\mathcal{E}^{+,0}) \geqslant 99/100$ on $\mathcal{E}^\star$, it gives that $\mu^{\Lambda_N,+,0}((\sigma^{\Lambda_N,+,0})_\Gamma \in \mathcal{E}_{\Gamma,+}) \geqslant 3/4$ and thus by monotonicity $\mu^{\Lambda_N,+,1}((\sigma^{\Lambda_N,+,1})_\Gamma \in \mathcal{E}_{\Gamma,+}) \geqslant 3/4$ (note $\mathcal{E}_{\Gamma,+}$ is an increasing set). Similarly, we get $\mu^{\Lambda_N,-,0}((\sigma^{\Lambda_N,-,0})_\Gamma \in \mathcal{E}_{\Gamma,-}) \geqslant 3/4$ on $\mathcal{E}^\star$. Consider an arbitrary monotone coupling π_Γ of $\mu^{\Lambda_N,+,1}$ and $\mu^{\Lambda_N,-,0}$ restricted to Γ. Then we see that on $\mathcal{E}^\star$

$$\pi_\Gamma(\mathcal{E}_{\Gamma,+,-}) \geqslant \tfrac{3}{4} + \tfrac{3}{4} - 1 = \tfrac{1}{2} \text{ where } \mathcal{E}_{\Gamma,+,-} = \{(\sigma^{\Lambda_N,+,1})_\Gamma \in \mathcal{E}_{\Gamma,+}, (\sigma^{\Lambda_N,-,0})_\Gamma \in \mathcal{E}_{\Gamma,-}\}\,.$$

Averaging (2.3.24) over the conditioning of (2.3.22) but restricted to the event $\mathcal{E}_{\Gamma,+,-}$, we get that on $\mathcal{E}^\star$

$$\frac{\Delta}{2}\int_0^1 (m^{\Lambda_{N/8},+,t} - m^{\Lambda_{N/8},-,t})dt \leqslant 8\sum_{v\in\Gamma}\langle(\sigma_v^{\Lambda_N,+,1} - \sigma_v^{\Lambda_N,-,0})\mathbb{1}_{\mathcal{E}_{\Gamma,+,-}}\rangle_{\pi_\Gamma} + 2/\beta\,.$$

Since π_Γ is a monotone coupling, we thus obtain that on $\mathcal{E}^\star$

$$\begin{aligned}\frac{\Delta}{2}\int_0^1 (m^{\Lambda_{N/8},+,t} - m^{\Lambda_{N/8},-,t})dt &\leqslant 8\sum_{v\in\Gamma}\langle\sigma_v^{\Lambda_N,+,1} - \sigma_v^{\Lambda_N,-,0}\rangle_{\pi_\Gamma} + 2/\beta \\ &= 8\sum_{v\in\Gamma}(\langle\sigma_v^{\Lambda_N,+,1}\rangle_{\mu^{\Lambda_N,+,1}} - \langle\sigma_v^{\Lambda_N,-,0}\rangle_{\mu^{\Lambda_N,-,0}}) + 2/\beta\,.\end{aligned}$$

Summing over $N/4 \leqslant k \leqslant N/2$, we deduce that on $\mathcal{E}^\star$

$$8\sum_{v\in\mathcal{A}_{N/2}}(\langle\sigma_v^{\Lambda_N,+,1}\rangle_{\mu^{\Lambda_N,+,1}} - \langle\sigma_v^{\Lambda_N,-,0}\rangle_{\mu^{\Lambda_N,-,0}}) + \frac{N}{2\beta} \quad \geqslant \frac{N\Delta}{8}\int_0^1 (m^{\Lambda_{N/8},+,t} - m^{\Lambda_{N/8},-,t})dt\,. \tag{2.3.25}$$

Step 2. For $N \geqslant 2$, recall that $\mathcal{A}_N = \Lambda_N \setminus \Lambda_{N/2}$ is an annulus. Adjust the value of r if necessary so that we can write $\mathcal{A}_{N/2} = \cup_{i=1}^r A_i$, where A_i is a copy of $\Lambda_{N/32}$ and A_i's are disjoint such that

$$A_i^{\mathrm{Big}} \subseteq \Lambda_N \setminus \Lambda_{N/8} \text{ for all } 1 \leqslant i \leqslant r\,. \tag{2.3.26}$$

(The geometric setup here is similar to that in the proof of Lemma 2.2.8; see the left picture of Figure 2.1 for an illustration.) By monotonicity, we see that for each $1 \leqslant i \leqslant r$

$$\begin{aligned}
&\mathbb{P}\Big(\sum_{v\in A_i}(\langle\sigma_v^{\Lambda_N,+,1}\rangle_{\mu^{\Lambda_N,+,1}} - \langle\sigma_v^{\Lambda_N,-,0}\rangle_{\mu^{\Lambda_N,-,0}}) > \theta^*\Big)\\
&\quad\leqslant \mathbb{P}\Big(\sum_{v\in A_i}(\langle\sigma_v^{A_i^{\mathrm{Big}},+,1}\rangle_{\mu^{A_i^{\mathrm{Big}},+,1}} - \langle\sigma_v^{A_i^{\mathrm{Big}},-,0}\rangle_{\mu^{A_i^{\mathrm{Big}},-,0}}) > \theta^*\Big)\\
&\quad= \mathbb{P}(m^{\Lambda_{N/8},+,t^*} - m^{\Lambda_{N/8},-,t^*} > \theta^*) \leqslant 1/2r\,,
\end{aligned}$$

where the equality holds due to (2.3.26) and $\Delta' = t^*\Delta$ (note that $h_v^{(t)} = h_v + \Delta'$ for $v \in \Lambda_N \setminus \Lambda_{N/8}$ and for *all* $0 \leqslant t \leqslant 1$), and in addition the last inequality holds due to (2.3.17). Thus, a simple union bound gives that the event $\{\sum_{v\in\mathcal{A}_{N/2}}(\langle\sigma_v^{\Lambda_N,+,1}\rangle_{\mu^{\Lambda_N,+,1}} - \langle\sigma_v^{\Lambda_N,-,0}\rangle_{\mu^{\Lambda_N,-,0}}) \leqslant r\theta^*\}$ contains an event $\mathcal{E}_{\mathcal{A}_{N/2}}$ which is measurable with respect to $\{h_v : v \notin \Lambda_{N/8}\}$ such that

$$\mathbb{P}(\mathcal{E}_{\mathcal{A}_{N/2}}) \geqslant 1/2\,. \tag{2.3.27}$$

Furthermore, let $\mathcal{T} = \{1 \leqslant t \leqslant 1 : m^{\Lambda_{N/8},+,t} - m^{\Lambda_{N/8},-,t} \geqslant \theta^*\}$. By (2.3.17) we have $\mathbb{E}|\mathcal{T}| \geqslant 1/2r$ where $|\mathcal{T}|$ is the Lebesgue measure of $\mathcal{T}$. Since $|\mathcal{T}| \leqslant 1$, we have $\mathbb{P}(|\mathcal{T}| \geqslant 1/4r) \geqslant 1/4r$. Therefore,

$$\mathbb{P}\Big(\int_0^1 (m^{\Lambda_{N/8},+,t} - m^{\Lambda_{N/8},-,t})dt \geqslant \theta^*/4r\Big) \geqslant 1/4r\,. \tag{2.3.28}$$

Combined with (2.3.27), this yields that

$$\mathbb{P}(\mathcal{E}^\diamond) \geqslant 1/8r \tag{2.3.29}$$

where $\mathcal{E}^\diamond$ is the event such that

$$\int_0^1 (m^{\Lambda_{N/8},+,t} - m^{\Lambda_{N/8},-,t})dt \geqslant \frac{\theta^*}{4r} \geqslant (4r^2)^{-1} \sum_{v\in\mathcal{A}_{N/2}} (\langle\sigma_v^{\Lambda_N,+,1}\rangle_{\mu^{\Lambda_N,+,1}} - \langle\sigma_v^{\Lambda_N,-,0}\rangle_{\mu^{\Lambda_N,-,0}})\,.$$

Suppose (2.3.18) does not hold. Then by (2.3.25) and the preceding display, the events $\mathcal{E}^\star$ and $\mathcal{E}^\diamond$ are mutually exclusive. But by (2.3.16) and (2.3.29), we have $\mathbb{P}(\mathcal{E}^\star) + \mathbb{P}(\mathcal{E}^\diamond) > 1$, arriving at a contradiction. □

Proof of Proposition 2.3.1. The proof of Proposition 2.3.1 at this point is highly similar to that of Proposition 2.2.2. As a result, we only provide a sketch emphasizing the additional subtleties.

Let π be an arbitrary monotone coupling of $\mu^{\Lambda_N,\pm}$ and let $\mathcal{C}^{\Lambda_N} = \mathcal{C}^{\Lambda_N,\pi}$ be defined as in (2.3.1).

For any rectangle $A \subseteq \mathbb{R}^2$ (whose sides are not necessarily parallel to the axes), recall that ℓ_A is the length of the longer side and A^{Large} is the square box concentric with A and of side length $32\ell_A$. In addition, the aspect ratio of A is the ratio between the lengths of the longer and shorter sides. Consider an arbitrary rectangle A with aspect ratio at least $a = 100$. For a (random) set $\mathcal{C} \subseteq \mathbb{Z}^2$, we continue to use $\mathrm{Cross}(A, \mathcal{C})$ to denote the event that there exists a path $v_0, \ldots, v_k \in A \cap \mathcal{C}$ connecting the two shorter sides of A. For *any* monotone coupling $\pi^{A^{\mathrm{Large}}}$ of $\mu^{A^{\mathrm{Large}},\pm}$ (below we denote $\mathcal{C}^{A^{\mathrm{Large}}} = \{v \in A^{\mathrm{Large}} : \sigma^{A^{\mathrm{Large}},+} > \sigma^{A^{\mathrm{Large}},-}\}$ under $\pi^{A^{\mathrm{Large}}}$), we can adapt the proof of (2.2.15) and deduce that (write $N' = \min\{2^n : 2^{n+2} \geqslant \ell_A\}$, and recall $\mathcal{E}^+$ as in Lemma 2.3.3)

$$\begin{aligned}\mathbb{P} \otimes \mu^{\Lambda_{N'},+}(\mathcal{E}^+_{N'}) &\geqslant 1 - 4(1 - \mathbb{P} \otimes \pi^{A^{\mathrm{Large}}}(\mathrm{Cross}(A, \mathcal{V}^{\sigma^{A^{\mathrm{Large}},+},+})))\\ &\geqslant 1 - 4(1 - \mathbb{P} \otimes \pi^{A^{\mathrm{Large}}}(\mathrm{Cross}(A, \mathcal{C}^{A^{\mathrm{Large}}})))\,,\end{aligned}$$

where the second inequality follows from the fact $\mathrm{Cross}(A, \mathcal{C}^{A^{\mathrm{Large}}}) \subseteq \mathrm{Cross}(A, \mathcal{V}^{\sigma^{A^{\mathrm{Large}},+},+})$.

In addition, by a similar derivation of (2.3.21),

$$\begin{aligned}\mathbb{P}\otimes\pi^{A^{\text{Large}}}(\text{Cross}(A,\mathcal{C}^{A^{\text{Large}}})) &\leqslant \mathbb{P}\otimes\pi^{A^{\text{Large}}}(\#(\mathcal{C}^{A^{\text{Large}}}\cap A)\geqslant \ell_A/2)\\ &\leqslant \frac{1}{2}(1+\mathbb{P}(\textstyle\sum_{v\in\Lambda_{N'/8}}(\langle\sigma_v^{\Lambda_{N'},+}\rangle_{\mu^{\Lambda_{N'},+}}-\langle\sigma_v^{\Lambda_{N'},-}\rangle_{\mu^{\Lambda_{N'},-}})>10^{-3}N'))\,.\end{aligned}$$

Therefore, by Lemma 2.3.3,

$$\mathbb{P}\otimes\pi^{A^{\text{Large}}}(\text{Cross}(A,\mathcal{C}^{A^{\text{Large}}}))\leqslant 1-\delta \text{ where } \delta=\delta(\varepsilon,\beta)>0\,. \tag{2.3.30}$$

It is crucial that (2.3.30) holds uniformly for all possible monotone couplings $\pi^{A^{\text{Large}}}$. Note that the probability for $\text{Cross}(A,\mathcal{C}^{\Lambda_N,\pi})$ could potentially depend on the location of A, either due to different influences from the boundary at different locations or different coupling mechanisms chosen at different location. However, thanks to (2.3.30), all these probabilities have a uniform upper bound which is strictly less than 1. In addition, by monotonicity of the Ising model, for a collection of rectangles that are well-separated, the corresponding crossing events can be dominated by independent events which have probabilities strictly less than 1. Next, we complete the proof of Proposition 2.3.1 by utilizing this intuition. For any $k\geqslant 1$ and any rectangles $A_1,\ldots,A_k\subseteq\{v\in\mathbb{R}^2:|v|_\infty\leqslant N/2\}$ with aspect ratios at least a such that (a) $\ell_0\leqslant\ell_{A_i}\leqslant N/32$ for all $1\leqslant i\leqslant k$ and (b) $A_1^{\text{Large}},\ldots,A_k^{\text{Large}}$ are disjoint, we see that under any coupling π of $\mu^{\Lambda_N,\pm}$, there exist sets $\mathcal{C}^{A_i^{\text{Large}}}$ such that

- $\mathcal{C}^{A_i^{\text{Large}}}$ is sampled according to *some* monotone coupling of $\mu^{A_i^{\text{Large}},\pm}$.
- $\mathcal{C}^{\Lambda_N,\pi}\cap A_i\subseteq\mathcal{C}^{A_i^{\text{Large}}}\cap A_i$ (by monotonicity of Ising model with respect to boundary conditions).
- $\mu^{A_i^{\text{Large}},\pm}$'s are mutually independent (as they only depend on $\{h_v:v\in A_i^{\text{Large}}\}$ respectively).

Therefore, by (2.3.30),

$$\mathbb{P} \otimes \pi(\cap_{i=1}^{k} \mathrm{Cross}(A_i, \mathcal{C}^{A_i^{\mathrm{Large}}})) \leqslant (1-\delta)^k \,.$$

This proves an analogue of Lemma 2.2.4, which verifies the hypothesis required in order to apply [1]. The remaining proof is merely an adaption of Proposition 2.2.2 and thus we omit further details. □

2.3.2. Admissible coupling and adaptive admissible coupling

In Sections 2.3.2 and 2.3.3, we wish to prove an analogue of Lemma 2.2.11. In the case for $T > 0$, it seems quite a bit more challenging as the choice of the coupling for various Ising measures plays a role, which seems to be subtle in light of Remark 2.3.8 below. To address the issue, we consider a general class of couplings for various Ising measures (i.e., adaptive admissible couplings) in this section. In Section 2.3.3, we describe a particular construction of adaptive admissible coupling, which is suited for the multi-scale analysis (the multi-scale analysis is a more complicated version of the proof for Lemma 2.2.11) presented in Section 2.3.3.

For $k \geqslant 1$, we consider *deterministic* boundary conditions and external fields $(\tau^{(i)}, \{h_v^{(i)} : v \in \Lambda\})$ where $\tau^{(i)} \in \{-1,1\}^{\partial\Lambda}$ for $1 \leqslant i \leqslant k$ (these will be fixed throughout this section). We define the partial order $\prec$ by

$$i \prec j \text{ if } \tau^{(i)} \leqslant \tau^{(j)} \text{ and } h^{(i)} \leqslant h^{(j)} \,. \tag{2.3.31}$$

We say that $(\sigma^{(1)}, \ldots, \sigma^{(k)})$ (for $\sigma^{(1)}, \ldots, \sigma^{(k)} \in \{-1,1\}^{\Lambda}$) is an admissible configuration if $\sigma^{(i)} \leqslant \sigma^{(j)}$ for all $i \prec j$. Denote by Σ_k the collection of all admissible configurations. For $A \subseteq \Lambda$, write $(\sigma^{(1)}, \ldots, \sigma^{(k)})_A$ for the restriction of $(\sigma^{(1)}, \ldots, \sigma^{(k)})$ on A.

Definition 2.3.7. For each $1 \leqslant i \leqslant k$, let $\mu^{(i)}$ be the Ising measure on Λ with boundary condition $\tau^{(i)}$ and external field $h^{(i)}$. We say that a measure π is an admissible coupling of

$\mu^{(1)}, \ldots, \mu^{(k)}$ if π is supported on Σ_k and its marginal distributions agree with $\mu^{(i)}$'s.

Remark 2.3.8. Ideally, it would be great if there would exist an admissible coupling π which satisfies the Markov field property. Or, it would also be great if there would exist an admissible coupling π which satisfies a weak version of Markov field property, such that for any $\Gamma \subseteq \Lambda$ the measure $\pi(\sigma^{(i)}_{S_\Gamma} \in \cdot \mid (\sigma^{(1)}, \ldots, \sigma^{(k)})_\Gamma)$ is the Ising measure on S_Γ with boundary condition $\sigma^{(i)}_{\partial S_\Gamma}$ and external field $\{h^{(i)}_v : v \in S_\Gamma\}$. However, such coupling does not exist as we can see from the following simple example. Let us consider Ising measures on a line segment with no external field and plus/minus boundary conditions on one end (denoted as u). Suppose that there exists an admissible coupling π (in this case a monotone coupling) with weak Markov field property. Then conditioned on the event that the two spins disagree at the other end of the line (denoted as v), we claim that the spins from the two Ising measures have to disagree on every vertex on the line, thereby violating the weak Markov property. In order to verify the claim, we suppose the claim fails and let w be the first vertex (from u) where the two spins agree with each other. Conditioned on spins from u to w, the two marginals at v are the same (by the weak Markov property) and thus have to agree in a monotone coupling.

In light of Remark 2.3.8, we will seek for admissible couplings with a desirable property even weaker than the weak Markov field property. To this end, we will explore the spins using certain "adaptive" algorithm and then we will argue that the marginal measures on the unexplored region remain to be Ising measures. This motivates us to consider the *adaptive admissible coupling* (see Definition 2.3.9 below). Let $\Xi_k = \{(\sigma^{(1)}, \ldots, \sigma^{(k)}) \in \{-1, 1\}^k : \sigma^{(i)} \leqslant \sigma^{(j)} \text{ for all } i \prec j\}$. For $\theta_1, \ldots, \theta_k$ which are measures on $\{-1, 1\}$, we say that $\theta_1, \ldots, \theta_k$ are admissible if $\theta_i(1) \leqslant \theta_j(1)$ for all $i \prec j$. In this case, let θ be the monotone coupling of $\theta_1, \ldots, \theta_k$. That is, θ is the joint measure of $(\sigma_1, \ldots, \sigma_k)$, which is defined in terms of a uniform variable U on $[0, 1]$ such that

$$\sigma_i = -1 \text{ if and only if } U \leqslant 1 - \theta_i(1)\,.$$

Clearly, θ is supported on Ξ_k and its marginals are $\theta_1, \ldots, \theta_k$. In addition, θ is consistent, i.e.,

$$\text{The projection of } \theta \text{ onto the first } (k-1) \text{ spins is the monotone coupling for } \theta_1, \ldots, \theta_{k-1}\,. \tag{2.3.32}$$

In order to define adaptive admissible couplings, we make use of exploration procedures. An exploration procedure can be encoded by a family of deterministic maps $\{f_V : V \subseteq \Lambda, V \neq \Lambda\}$ where f_V is a mapping that maps an admissible configuration on V to a vertex in $\Lambda \setminus V$. That is to say, if we have explored a set $V \subseteq \Lambda$ and the spin configuration on V is given by $(\sigma^{(1)}, \ldots, \sigma^{(k)})_V$, then the next vertex we will explore is $f_V((\sigma^{(1)}, \ldots, \sigma^{(k)})_V)$.

Definition 2.3.9. For each exploration procedure $\{f_V\}$, we associate an admissible coupling in the following manner. Let $\mathcal{V}_0 = \emptyset$. For $t \geqslant 1$, let $v_t = f_{\mathcal{V}_{t-1}}((\sigma^{(1)}, \ldots, \sigma^{(k)})_{\mathcal{V}_{t-1}})$. Let $\mathcal{V}_t = \mathcal{V}_{t-1} \cup \{v_t\}$. Quenched on the realization of $\{\mathcal{V}_{t-1}, (\sigma^{(1)}, \ldots, \sigma^{(k)})_{\mathcal{V}_{t-1}}\}$, for $1 \leqslant i \leqslant k$ let $\theta_i^{(t)}(\pm 1) = \mu^{(i)}(\sigma_{v_t}^{(i)} = \pm 1 \mid \sigma_{\mathcal{V}_{t-1}}^{(i)})$. Let $\theta^{(t)}$ be the monotone coupling of $\theta_1^{(t)}, \ldots, \theta_k^{(t)}$, and we sample $(\sigma^{(1)}, \ldots, \sigma^{(k)})_{v_t}$ according to $\theta^{(t)}$. We repeat this procedure until $t = \#\Lambda$. We let π be the measure on $(\sigma^{(1)}, \ldots, \sigma^{(k)})$ at the end of the procedure. In addition, we say that a random set $\mathcal{V}$ is a *stopping set* if $\{\mathcal{V} = \mathcal{V}_t = V_t\}$ (for any deterministic $V_t \subseteq \Lambda$) is measurable with respect to $\{(\sigma^{(1)}, \ldots, \sigma^{(k)})_{V_t})\}$.

Remark 2.3.10. In the study of spin models, it is common to use an exploration procedure to discover certain observables (such as interfaces) associated with spin configurations. Often times, an instance of spin configurations is sampled a priori (which is usually sampled according to a Gibbs measure) and then the exploration procedure is performed on this instance. That being said, it is not uncommon to construct a measure as the exploration process evolves. Definition 2.3.9 is one example of such constructions, where the spin configuration is sampled as the exploration procedure evolves and more importantly the measure on spin configurations depends on the exploration procedure.

Lemma 2.3.11. *For each exploration procedure, the measure π given in Definition 2.3.9 is* *a well-defined admissible coupling. In addition, for any stopping set $\mathcal{V}$, given the realization of $\mathcal{V}$ and $(\sigma^{(1)}, \ldots, \sigma^{(k)})_{\mathcal{V}}$, the conditional measure of π restricted on $\mathcal{V}^c$ has marginals corresponding to Ising measures on $\mathcal{V}^c$ with boundary condition $\sigma^{(i)}_{\partial \mathcal{V}^c}$ and external field $\{h^{(i)}_v : v \in \mathcal{V}^c\}$.*

Proof. The measure π is well-defined since we can inductively verify that for $t = 0, 1, 2, \ldots$, the sequence $\theta^{(t)}_1, \ldots, \theta^{(t)}_k$ is admissible and thus $(\sigma^{(1)}, \ldots, \sigma^{(k)})_{\mathcal{V}_{t+1}}$ is admissible. To prove the second part of the statement, it suffices to show that for each $1 \leqslant i \leqslant k$ and $1 \leqslant t \leqslant \#\Lambda$,

$$\pi(\sigma^{(i)}_{\Lambda \setminus V_{t-1}} \in \cdot \mid (\sigma^{(1)}, \ldots, \sigma^{(k)})_{\mathcal{V}_{t-1}}, \mathcal{V}_{t-1} = V_{t-1}) = \mu^{(i)}(\sigma^{(i)}_{\Lambda \setminus V_{t-1}} \in \cdot \mid \sigma^{(i)}_{V_{t-1}}) \,. \tag{2.3.33}$$

We prove (2.3.33) by induction for $t = \#\Lambda, \ldots, 1$. It is obvious from Definition 2.3.9 that (2.3.33) holds for $t = \#\Lambda$. Suppose (2.3.33) holds for t, we then deduce for $t-1$ that

$$\begin{aligned}
&\pi(\sigma^{(i)}_{\Lambda \setminus (V_{t-2} \cup \{v_{t-1}\})} \in \cdot, \sigma^{(i)}_{v_{t-1}} = \pm 1 \mid (\sigma^{(1)}, \ldots, \sigma^{(k)})_{\mathcal{V}_{t-2}}, \mathcal{V}_{t-2} = V_{t-2}) \\
&= \mu^{(i)}(\sigma^{(i)}_{v_{t-1}} = \pm 1 \mid \sigma^{(i)}_{V_{t-2}}) \times \mu^{(i)}(\sigma^{(i)}_{\Lambda \setminus (V_{t-2} \cup \{v_{t-1}\})} \in \cdot \mid \sigma^{(i)}_{V_{t-2}}, \sigma^{(i)}_{v_{t-1}} = \pm 1) \,.
\end{aligned}$$

This implies that $\pi(\sigma^{(i)}_{\Lambda \setminus V_{t-2}} \in \cdot \mid (\sigma^{(1)}, \ldots, \sigma^{(k)})_{\mathcal{V}_{t-2}}, \mathcal{V}_{t-2} = V_{t-2}) = \mu^{(i)}(\sigma^{(i)}_{\Lambda \setminus V_{t-2}} \in \cdot \mid \sigma^{(i)}_{V_{t-2}})$, thereby completing the proof by induction. □

In what follows, we refer to π as in Definition 2.3.9 as an adaptive admissible coupling. In addition, we will always define adaptive admissible couplings by presenting an exploration procedure and then consider the associated admissible coupling given in Definition 2.3.9. For convenience of exposition, we usually describe an exploration procedure in words rather than specifying the maps $\{f_V\}$.

2.3.3. A multi-scale analysis via another perturbation argument

Let $\alpha > 1$ be as in Proposition 2.3.1. Let $\sqrt{1/\alpha} < \alpha' < 1$. Let $N_0 = N_0(\varepsilon, \beta)$ be a large number to be chosen. For each $N \geqslant N_0$ (of the form 4^n), set $\Delta = \Delta(N) = N^{-\alpha(\alpha')^2}$. In the rest of the paper, we consider the following perturbation:

$$\widetilde{h}_v^{(N)} = \begin{cases} h_v + \Delta, & \text{for } v \in \Lambda_N \setminus \Lambda_{N/4}\,, \\ h_v, & \text{for } v \in \Lambda_{N/4}\,. \end{cases} \tag{2.3.34}$$

We denote by $\widetilde{\mu}^{\Lambda_N,\pm}$ the Ising measures on Λ_N with respect to plus/minus boundary conditions and external field $\{\widetilde{h}_v^{(N)} : v \in \Lambda_N\}$, and denote by $\widetilde{\sigma}^{\Lambda_N,\pm}$ the spins sampled according to $\widetilde{\mu}^{\Lambda_N,\pm}$. In this whole section except in (2.3.45) and (2.3.46), we will quench on the realization of $\{h_v\}$ and thus the external field is viewed as deterministic.

A construction of an adaptive admissible coupling

We will define the following adaptive admissible coupling π_{Λ_N} for $\mu^{\Lambda_N,\pm}$ and $\widetilde{\mu}^{\Lambda_N,\pm}$. According to Definition 2.3.9, in order to specify π_{Λ_N}, we only need to specify the exploration procedure (i.e., the order of vertices in which we sample the spins), as described as follows. Throughout the procedure, we let $\mathcal{C}_*^{\Lambda_N}$ be the collection of vertices v which have been sampled such that $\sigma_v^{\Lambda_N,+} > \sigma_v^{\Lambda_N,-}$ and $\widetilde{\sigma}_v^{\Lambda_N,+} > \widetilde{\sigma}_v^{\Lambda_N,-}$. We first sample spins at vertices on $\partial\Lambda_k$ for $k = N-1, N-2, \ldots, \frac{N}{2}$. For vertices on $\partial\Lambda_k$, for concreteness we sample in clockwise order starting from the right top corner. Next, let $K = \lfloor N^{\alpha'\alpha} \rfloor$ and $\ell = \lfloor \frac{1}{4} N^{1-\alpha'} \rfloor$. A comment on the order of the scales chosen: the exploration procedure below contains ℓ phases, and in every phase we consider an annulus where the inner and outer boundaries have Euclidean distance $N^{\alpha'}$ and thus by Proposition 2.3.1 typically have intrinsic distance $\geqslant K \gg N$. This is why we can hope to gain a contraction when comparing the number of disagreements on an annulus to that on its neighboring (larger) annulus (see (2.3.47) below).

We now turn to the description of the exploration procedure. For each $1 \leqslant j \leqslant \ell$ our construction employs the following procedure which we refer to as Phase j (see Figure 2.4 for an illustration). Let $N' = \frac{N}{2} - (j-1)N^{\alpha'}$.

- We set $A_{j,0} = \partial\Lambda_{N'} \cap \mathcal{C}_*^{\Lambda_N}$, $V_{j,0} = \Lambda_N \setminus \Lambda_{N'}$, and for $k = 0, 1, \ldots, K$, we inductively employ the following procedure (which we refer to as stage). At the beginning of Stage

$k+1$, we first set $A_{j,k+1} = \emptyset$ and $V_{j,k+1} = V_{j,k}$.

– If $A_{j,k} = \emptyset$ (which we denote as event $\mathcal{E}_{j,k,\emptyset}$), we sample the unexplored vertices in Λ_N in a prefixed order (which can be arbitrary) and stop our procedure. Otherwise, we explore all the neighbors of $A_{j,k}$ (in a certain prefixed order, which can be arbitrary) which are in $\Lambda_{N'} \setminus V_{j,k}$ (that is, vertices which have not been explored) and sample the spins at these vertices. We also put these vertices into $V_{j,k+1}$.

– If a newly sampled vertex is in $\partial\Lambda_{N'-N^{\alpha'}}$ (we denote this as event $\mathcal{E}_{j,k,d}$, where the subscript d suggests an event related to the intrinsic distance), we sample the unexplored vertices in Λ_N in a prefixed order (which can be arbitrary) and stop our procedure. Otherwise, if a newly sampled vertex ends up in $\mathcal{C}_*^{\Lambda_N}$ then we add it to $A_{j,k+1}$. (For $k \geqslant 1$, it is clear that $A_{j,k}$ records all the vertices in $\Lambda_{N'}$ that are of $d_{\mathcal{C}_*^{\Lambda_N}}$-distance k to $\partial\Lambda_{N'}$ and $V_{j,k}$ records all the explored vertices up to Stage k.)

- Sample the unexplored vertices in $\Lambda_{N'} \setminus \Lambda_{N'-N^{\alpha'}}$ in a prefixed order (which can be arbitrary).

Finally, if the procedure is not yet stopped after ℓ phases, we sample the unexplored vertices in Λ_N in a prefixed order (which can be arbitrary).

Remark 2.3.12. (1) Later in the analysis, when we refer to sets such as $A_{j,k}$, $V_{j,k}$ we mean to use their values at the end of our procedure. (2) Note that in the preceding procedure, unless some event of the form $\mathcal{E}_{j,k,\emptyset}$ or $\mathcal{E}_{j,k,d}$ occurred, the exploration in all the ℓ phases is within $\Lambda_N \setminus \Lambda_{N/4}$.

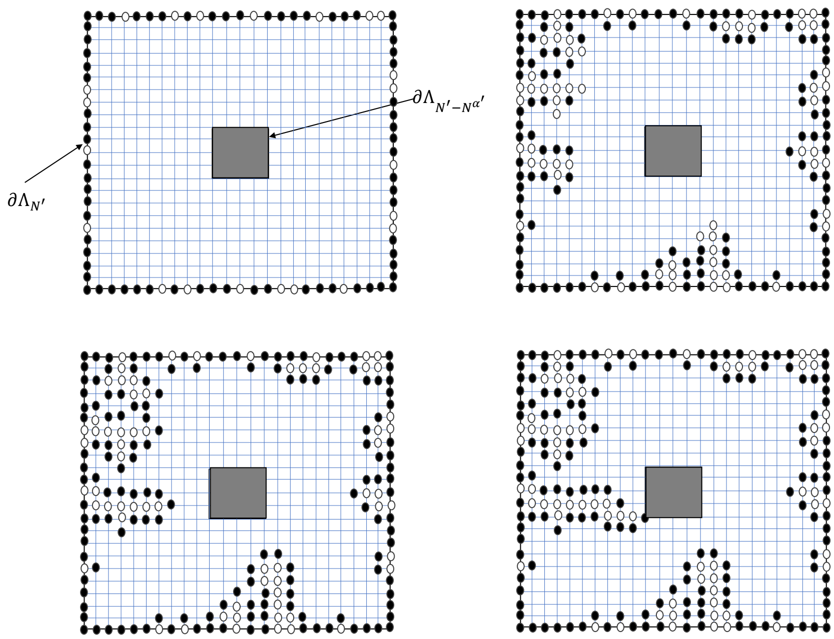

Figure 2.4: Illustration for Phase j of the construction in Section 2.3.3. The inside square is $\Lambda_{N'-N^{\alpha'}}$, whose size has been reduced in the picture for better demonstration. On lattice points, empty indicates an unexplored vertex, an open circle indicates a vertex in $\mathcal{C}_*^{\Lambda_N}$, and a solid disk indicates a vertex not in $\mathcal{C}_*^{\Lambda_N}$. The top-left illustrates the beginning of Phase j, where vertices on $\partial\Lambda_{N'}$ have been explored (vertices outside have been explored too but we did not draw); the top-right illustrates the middle of Phase j (here $k = 5$); the bottom-left picture illustrates event $\mathcal{E}_{j,k,\emptyset}$ (here $k = 8$); the bottom right event illustrates $\mathcal{E}_{j,k,d}$ (here $k = 12$).

Another perturbation argument

We use $\widetilde{H}^{\Lambda_N,\pm}$, $\widetilde{F}^{\Lambda_N,\pm}$, $\widetilde{\sigma}^{\Lambda_N,\pm}$ to denote tilde versions of $H^{\Lambda_N,\pm}$, $F^{\Lambda_N,\pm}$, $\sigma^{\Lambda_N,\pm}$, i.e., defined analogously but with respect to the field $\{\widetilde{h}_v^{(N)}\}$ defined as in (2.3.34). Without further notice, we will always consider measures where we couple all these Ising spins together. Thus, in particular, $\mathcal{C}^{\Lambda_N}$ and $\widetilde{\mathcal{C}}^{\Lambda_N}$ are defined in the same probability space and we can then define $\mathcal{C}_*^{\Lambda_N} = \widetilde{\mathcal{C}}^{\Lambda_N} \cap \mathcal{C}^{\Lambda_N}$.

We need some preparation before presenting our perturbative analysis. Suppose that $\mathcal{V}$ is a stopping set (see Definition 2.3.9) obtained when constructing π_{Λ_N} described in Section 2.3.3. Let $\pi'_{\mathcal{V}^c}$ be the restriction of π_{Λ_N} to $\mathcal{V}^c$. (We use prime in the notation $\pi'_{\mathcal{V}^c}$ as we wish to save $\pi_{\mathcal{V}^c}$ for later use.) By Lemma 2.3.11 and our definition of π_{Λ_N}, we see that $\pi'_{\mathcal{V}^c}$ depends on $(\sigma^{\Lambda_N,\pm})_{\mathcal{V}}, (\widetilde{\sigma}^{\Lambda_N,\pm})_{\mathcal{V}}$ only through $(\sigma^{\Lambda_N,\pm})_{\partial\mathcal{V}^c}, (\widetilde{\sigma}^{\Lambda_N,\pm})_{\partial\mathcal{V}^c}$. Thus, we may denote by $(\sigma^{\mathcal{V}^c,(\sigma^{\Lambda_N,\pm})_{\partial\mathcal{V}^c}}, \widetilde{\sigma}^{\mathcal{V}^c,(\widetilde{\sigma}^{\Lambda_N,\pm})_{\partial\mathcal{V}^c}})$ the spin configurations sampled according to $\pi'_{\mathcal{V}^c}$ with corresponding boundary conditions on $\partial\mathcal{V}^c$. Thus,

$$((\sigma_{\mathcal{V}}^{\Lambda_N,\pm}, \sigma^{\mathcal{V}^c,(\sigma^{\Lambda_N,\pm})_{\partial\mathcal{V}^c}}), (\widetilde{\sigma}_{\mathcal{V}}^{\Lambda_N,\pm}, \widetilde{\sigma}^{\mathcal{V}^c,(\widetilde{\sigma}^{\Lambda_N,\pm})_{\partial\mathcal{V}^c}})) \text{ has law } \pi_{\Lambda_N}\,. \tag{2.3.35}$$

In what follows, we will mainly consider the measure $\pi'_{\mathcal{V}^c}$. For clarity of exposition, we quench on the realization of $\mathcal{V} = V$. Let $S = V^c$ and $\Gamma = \partial S$ (thus we have $S \subseteq S_\Gamma$). Further, we quench on the values of $(\sigma^{\Lambda_N,\pm})_\Gamma, (\widetilde{\sigma}^{\Lambda_N,\pm})_\Gamma$ by

$$(\sigma^{\Lambda_N,\pm})_\Gamma = \tau^\pm, (\widetilde{\sigma}^{\Lambda_N,\pm})_\Gamma = \widetilde{\tau}^\pm\,, \text{ where } \tau^\pm, \widetilde{\tau}^\pm \in \{-1,1\}^\Gamma\,. \tag{2.3.36}$$

For $v \in \Gamma$ (in fact, any $v \in \Lambda_N$), by admissibility there are only six possible values for $(\tau_v^+, \tau_v^-, \widetilde{\tau}_v^+, \widetilde{\tau}_v^-)$ as shown in Table 2.1. For each such possible spin value, we will define a "hat" version $(\widehat{\tau}_v^+, \widehat{\tau}_v^-, \widehat{\widetilde{\tau}}_v^+, \widehat{\widetilde{\tau}}_v^-)$, where the definition is given in Table 2.2. Note that the hat version is a modification of the original spin value, and we emphasize the change in Table 2.2 by circling out the modifications. We will explain why we introduced the hat version of the spin on Γ after a number of definitions. From Tables 2.1 and 2.2, we see that

$$\widehat{\tau}^+ \geqslant \widehat{\tau}^- \geqslant \tau^-, \widetilde{\tau}^+ \geqslant \widehat{\widetilde{\tau}}^+ \geqslant \widehat{\widetilde{\tau}}^-, \widehat{\widetilde{\tau}}^+ = \widehat{\tau}^+ \geqslant \tau^+, \widehat{\widetilde{\tau}}^- = \widehat{\tau}^- = \widetilde{\tau}^-\,. \tag{2.3.37}$$

From a notation point of view, despite the fact that $\widehat{\tau}^\pm = \widehat{\widetilde{\tau}}^\pm$, we still differentiate these two notations because our mental picture is that the boundary conditions $\widehat{\tau}^\pm$ are matched to external field $\{h_v\}$ and the boundary conditions $\widehat{\widetilde{\tau}}^\pm$ are matched to external field $\{\widetilde{h}_v^{(N)}\}$.

Recall that π'_S is the admissible coupling for Ising measures with boundary conditions and external fields $((\tau^\pm)_\Gamma, \{h_v\})$, $((\widetilde{\tau}^\pm)_\Gamma, \{\widetilde{h}_v^{(N)}\})$, where the order of sampling vertex is given by that of π_{Λ_N} conditioned on spin configurations on the stopping set $\mathcal{V} = V$.

Table 2.1: Original spins on Γ

type	τ_v^+	τ_v^-	$\widetilde{\tau}_v^+$	$\widetilde{\tau}_v^-$
a.	-1	-1	-1	-1
b.	-1	-1	$+1$	-1
c.	-1	-1	$+1$	$+1$
d.	$+1$	$+1$	$+1$	$+1$
e.	$+1$	-1	$+1$	$+1$
f.	$+1$	-1	$+1$	-1

Table 2.2: The hat version of the spins on Γ

type	$\widehat{\tau}_v^+$	$\widehat{\tau}_v^-$	$\widehat{\widetilde{\tau}}_v^+$	$\widehat{\widetilde{\tau}}_v^-$
a.	-1	-1	-1	-1
$\bullet b.$	-1	-1	-1	-1
$\bullet c.$	$+1$	$+1$	$+1$	$+1$
d.	$+1$	$+1$	$+1$	$+1$
$\bullet e.$	$+1$	$+1$	$+1$	$+1$
f.	$+1$	-1	$+1$	-1

In addition, we can extend π'_S to an adaptive admissible coupling π_S for Ising measures with boundary conditions and external fields $((\tau^\pm)_\Gamma, \{h_v\})$, $((\widetilde{\tau}^\pm)_\Gamma, \{\widetilde{h}_v^{(N)}\})$, $((\widehat{\tau}^\pm)_\Gamma, \{h_v\})$, $((\widehat{\widetilde{\tau}}^\pm)_\Gamma, \{\widetilde{h}_v^{(N)}\})$, where the order of sampling vertices is determined by the coupling π'_S. Let $(\sigma^{S,\tau^\pm}, \widetilde{\sigma}^{S,\widetilde{\tau}^\pm}, \sigma^{S,\widehat{\tau}^\pm}, \widetilde{\sigma}^{S,\widehat{\widetilde{\tau}}^\pm})$ be the spin configuration sampled according to π_S (note that we use the tilde symbol on σ to emphasize the dependence on the external field $\{\widetilde{h}_v^{(N)}\}$; similarly for H and F below). By (2.3.32), we see that the projection of π_S onto $(\sigma^{S,\tau^\pm}, \widetilde{\sigma}^{S,\widetilde{\tau}^\pm})$ has measure π'_S. As a result, we will simply use π_S in what follows. We also let $H^{S,\tau^\pm}, \widetilde{H}^{S,\widetilde{\tau}^\pm}, H^{S,\widehat{\tau}^\pm}, \widetilde{H}^{S,\widehat{\widetilde{\tau}}^\pm}$ denote Hamiltonians for corresponding Ising spins. Similarly, we denote by $F^{S,\tau^\pm}$, $\widetilde{F}^{S,\widetilde{\tau}^\pm}$, $F^{S,\widehat{\tau}^\pm}$, $\widetilde{F}^{S,\widehat{\widetilde{\tau}}^\pm}$ the log-partition-functions of corresponding Ising measures. Define

$$\mathcal{C}^{S,\tau^\pm} = \{v \in S : \sigma_v^{S,\tau^+} = 1, \sigma_v^{S,\tau^-} = -1\}$$

and similarly define $\widetilde{\mathcal{C}}^{S,\widetilde{\tau}^\pm}, \mathcal{C}^{S,\widehat{\tau}^\pm}, \widetilde{\mathcal{C}}^{S,\widehat{\widetilde{\tau}}^\pm}$. Define $\mathcal{C}_*^{S,\tau^\pm,\widetilde{\tau}^\pm} = \mathcal{C}^{S,\tau^\pm} \cap \widetilde{\mathcal{C}}^{S,\widetilde{\tau}^\pm}$ and $\mathcal{C}_*^{S,\widehat{\tau}^\pm,\widehat{\widetilde{\tau}}^\pm} = \mathcal{C}^{S,\widehat{\tau}^\pm} \cap \widetilde{\mathcal{C}}^{S,\widehat{\widetilde{\tau}}^\pm}$.

Now we have necessary notations to explain the reason for introducing the hat version of the spins on Γ. We wish to bound $\#(\mathcal{C}_*^{\Lambda_N} \cap S \cap (\Lambda_N \setminus \Lambda_{N/4}))$ in terms of $\#(\mathcal{C}_*^{\Lambda_N} \cap \Gamma)$. One way to achieve this is to track the increment for the difference between the log-partition-functions with plus and minus boundary conditions when the external field is perturbed. We see that on the one hand, the increment for the difference between log-partition-functions can be bounded from below in terms of $\#(\mathcal{C}_*^{\Lambda_N} \cap S \cap (\Lambda_N \setminus \Lambda_{N/4}))$ (see Lemma 2.3.15); and on the other hand such increment can be bounded from above by the number of disagreements for spins on Γ with respect to the plus and minus boundary conditions. However, when approaching the upper bound, the spin values of Type b, c, e as in Table 2.1 will also contribute to the upper bound despite the fact that they do not belong to $\mathcal{C}_*^{\Lambda_N} \cap \Gamma$. To address this, we introduce the hat version of the spins, which are in agreement except on $\mathcal{C}_*^{\Lambda_N} \cap \Gamma$. A crucial feature as we will show in Lemma 2.3.13, is that under the admissible coupling π_S we have $\mathcal{C}_*^{S,\tau^\pm,\widetilde{\tau}^\pm} \subseteq \mathcal{C}_*^{S,\widehat{\tau}^\pm,\widehat{\widetilde{\tau}}^\pm}$. Therefore, the intended lower bound on the increment for the difference between log-partition-functions is still valid for the hat version. Another crucial feature of the hat version of the spin is that

$$\begin{aligned} &\{v \in \Gamma : \tau_v^+ = \widetilde{\tau}_v^+ = 1, \tau_v^- = \widetilde{\tau}_v^- = -1\} = \{v \in \Gamma : \widehat{\tau}_v^+ = \widehat{\widetilde{\tau}}_v^+ = 1, \widehat{\tau}_v^- = \widehat{\widetilde{\tau}}_v^- = -1\} \\ &= \{v \in \Gamma : \widehat{\tau}_v^+ = 1, \widehat{\tau}_v^- = -1\} = \{v \in \Gamma : \widehat{\widetilde{\tau}}_v^+ = 1, \widehat{\widetilde{\tau}}_v^- = -1\}. \end{aligned} \tag{2.3.38}$$

Lemma 2.3.13. *Under the admissible coupling π_S, we have $\mathcal{C}_*^{S,\tau^\pm,\widetilde{\tau}^\pm} \subseteq \mathcal{C}_*^{S,\widehat{\tau}^\pm,\widehat{\widetilde{\tau}}^\pm}$.*

Proof. For $u \in \mathcal{C}_*^{S,\tau^\pm,\widetilde{\tau}^\pm}$, we have $\sigma_u^{S,\tau^+} = \widetilde{\sigma}_u^{S,\widetilde{\tau}^+} = 1$ and $\sigma_u^{S,\tau^-} = \widetilde{\sigma}_u^{S,\widetilde{\tau}^-} = -1$. By (2.3.37) and the admissible coupling, we see that $\sigma_u^{S,\widehat{\tau}^+} \geqslant \sigma_u^{S,\tau^+} = 1$; similarly, $\sigma_u^{S,\widehat{\tau}^-} \leqslant \widetilde{\sigma}_u^{S,\widetilde{\tau}^-} = -1$. So $u \in \mathcal{C}^{S,\widehat{\tau}^\pm}$. In addition, by (2.3.37) and the admissible coupling, we see that $\widetilde{\sigma}_u^{S,\widehat{\widetilde{\tau}}^+} \geqslant \sigma_u^{S,\tau^+} = 1$; similarly, $\widetilde{\sigma}_u^{S,\widehat{\widetilde{\tau}}^-} = \widetilde{\sigma}_u^{S,\widetilde{\tau}^-} = -1$. So $u \in \widetilde{\mathcal{C}}^{S,\widehat{\widetilde{\tau}}^\pm}$. Thus, $u \in \mathcal{C}_*^{S,\widehat{\tau}^\pm,\widehat{\widetilde{\tau}}^\pm}$ as required. □

Corollary 2.3.14. *Under the admissible coupling π_S, we have $o \notin \mathcal{C}_*^{S,\tau^\pm,\widetilde{\tau}^\pm}$ provided that $\mathcal{C}_*^{\Lambda_N} \cap \Gamma = \emptyset$.*

Proof. If $\mathcal{C}_*^{\Lambda_N} \cap \Gamma = \emptyset$, we have $\widehat{\tau}^+ = \widehat{\tau}^- = \widehat{\widetilde{\tau}}^+ = \widehat{\widetilde{\tau}}^-$, in which case we have $\mathcal{C}_*^{S,\widehat{\tau}^\pm,\widehat{\widetilde{\tau}}^\pm} = \emptyset$ and in particular $o \notin \mathcal{C}_*^{S,\widehat{\tau}^\pm,\widehat{\widetilde{\tau}}^\pm}$. Combined with Lemma 2.3.13, this completes the proof of the corollary. □

Lemma 2.3.15. *We have that*

$$2\Delta\langle \#(\mathcal{C}_*^{S,\widehat{\tau}^\pm,\widehat{\widetilde{\tau}}^\pm} \cap (\Lambda_N \setminus \Lambda_{N/4}))\rangle_{\pi_S} \leqslant (\widetilde{F}^{S,\widehat{\widetilde{\tau}}^+} - \widetilde{F}^{S,\widehat{\widetilde{\tau}}^-}) - (F^{S,\widehat{\tau}^+} - F^{S,\widehat{\tau}^-}) \tag{2.3.39}$$

$$\leqslant 16\#\{v \in \Gamma : \widehat{\tau}_v^+ = \widehat{\widetilde{\tau}}_v^+ = 1, \widehat{\tau}_v^- = \widehat{\widetilde{\tau}}_v^- = -1\}\,. \tag{2.3.40}$$

Proof. The proof of the lemma shares some similarity to that of Lemma 2.3.5. However, we give a self-contained proof here in order for clarity of exposition.

We first prove (2.3.40). A straightforward computation gives that

$$\widetilde{F}^{S,\widehat{\widetilde{\tau}}^+} - \widetilde{F}^{S,\widehat{\widetilde{\tau}}^-} = \frac{1}{\beta}\log\frac{\sum_\sigma e^{-\beta \widetilde{H}^{S,\widehat{\widetilde{\tau}}^+}(\sigma)}}{\sum_\sigma e^{-\beta \widetilde{H}^{S,\widehat{\widetilde{\tau}}^-}(\sigma)}} \leqslant \frac{1}{\beta}\log e^{8\beta\cdot\#\{v\in\Gamma:\widehat{\widetilde{\tau}}_v^+\neq\widehat{\widetilde{\tau}}_v^-\}}$$
$$\leqslant 8\cdot\#\{v\in\Gamma : \widehat{\widetilde{\tau}}_v^+ \neq \widehat{\widetilde{\tau}}_v^-\}\,.$$

Similarly, $F^{S,\widehat{\tau}^+} - F^{S,\widehat{\tau}^-} \geqslant -8\cdot\#\{v\in\Gamma : \widehat{\tau}_v^+ \neq \widehat{\tau}_v^-\}$. Combined with (2.3.38), this proves (2.3.40).

Now we turn to prove (2.3.39). We write

$$(\widetilde{F}^{S,\widehat{\widetilde{\tau}}^+} - \widetilde{F}^{S,\widehat{\widetilde{\tau}}^-}) - (F^{S,\widehat{\tau}^+} - F^{S,\widehat{\tau}^-}) = (\widetilde{F}^{S,\widehat{\widetilde{\tau}}^+} - F^{S,\widehat{\tau}^+}) - (\widetilde{F}^{S,\widehat{\widetilde{\tau}}^-} - F^{S,\widehat{\tau}^-}). \tag{2.3.41}$$

For $0 \leqslant t \leqslant 1$, define

$$\widetilde{h}_v^{(t)} = \begin{cases} h_v + t\Delta, & \text{for } v \in \Lambda_N \setminus \Lambda_{N/4}\,,\\ h_v, & \text{for } v \in \Lambda_{N/4}\,. \end{cases} \tag{2.3.42}$$

Let $F^{S,\widehat{\tau}^+,t}$ be the log-partition-function on S with boundary condition $\widehat{\tau}^+$ (note that $\widehat{\widetilde{\tau}}^+ = \widehat{\tau}^+$ by (2.3.37)) and external field $\{\widetilde{h}_v^{(t)}\}$. In particular, $F^{S,\widehat{\tau}^+,0} = F^{S,\widehat{\tau}^+}$ and $F^{S,\widehat{\tau}^+,1} = \widetilde{F}^{S,\widehat{\widetilde{\tau}}^+}$. Similar notations apply for $F^{S,\widehat{\tau}^-,t}$. Thus, we get that

$$\widetilde{F}^{S,\widehat{\widetilde{\tau}}^+} - F^{S,\widehat{\tau}^+} = \int_0^1 \frac{dF^{S,\widehat{\tau}^+,t}}{dt}dt, \qquad \widetilde{F}^{S,\widehat{\widetilde{\tau}}^-} - F^{S,\widehat{\tau}^-} = \int_0^1 \frac{dF^{S,\widehat{\tau}^-,t}}{dt}dt\,. \tag{2.3.43}$$

Denote by $\sigma^{S,\widehat{\tau}^{\pm},t}$ spins sampled according to Ising measures with boundary conditions $\widehat{\tau}^{\pm}$ and external field $\{\widetilde{h}^{(t)}\}$. In addition, for any fixed t, we let $\pi_{S,t}$ be the admissible coupling extended from π_S by also incorporating the spins $\sigma^{S,\widehat{\tau}^{\pm},t}$ (again, the order of sampling vertex is given by that of π_S). Therefore, we see

$$\frac{dF^{S,\widehat{\tau}^{+},t}}{dt} = \Delta \sum_{v\in S\cap(\Lambda_N\setminus\Lambda_{N/4})} \langle \sigma_v^{S,\widehat{\tau}^{+},t}\rangle_{\pi_{S,t}} \text{ and } \frac{dF^{S,\widehat{\tau}^{-},t}}{dt} = \Delta \sum_{v\in S\cap(\Lambda_N\setminus\Lambda_{N/4})} \langle \sigma_v^{S,\widehat{\tau}^{-},t}\rangle_{\pi_{S,t}}\,.$$

Combined with (2.3.43) and (2.3.41), it yields that

$$(F^{S,\widehat{\widetilde{\tau}}^{+}} - F^{S,\widehat{\widetilde{\tau}}^{-}}) - (F^{S,\widehat{\tau}^{+}} - F^{S,\widehat{\tau}^{-}}) = 2\int_0^1 \Delta\langle \#\{v\in S\cap(\Lambda_N\setminus\Lambda_{N/4}) : \sigma_v^{S,\widehat{\tau}^{+},t} \neq \sigma_v^{S,\widehat{\tau}^{-},t}\}\rangle_{\pi_{S,t}} dt\,. \tag{2.3.44}$$

For any $v\in S$ and $t\in(0,1)$, by admissible coupling we have $\sigma_v^{S,\widehat{\tau}^{+}} \leqslant \sigma_v^{S,\widehat{\tau}^{+},t} \leqslant \widetilde{\sigma}_v^{S,\widehat{\widetilde{\tau}}^{+}}$ and $\sigma_v^{S,\widehat{\tau}^{-}} \leqslant \sigma_v^{S,\widehat{\tau}^{-},t} \leqslant \widetilde{\sigma}_v^{S,\widehat{\widetilde{\tau}}^{-}}$. Therefore, $\{v\in S\cap(\Lambda_N\setminus\Lambda_{N/4}) : \sigma_v^{S,\widehat{\tau}^{+},t} \neq \sigma_v^{S,\widehat{\tau}^{-},t}\} \supset \mathcal{C}_*^{S,\widehat{\tau}^{\pm},\widehat{\widetilde{\tau}}^{\pm}} \cap (\Lambda_N\setminus\Lambda_{N/4})$. Combined with (2.3.44), this completes the proof of (2.3.39). $\square$

Corollary 2.3.16. *Conditioned on the realization of the stopping set $\mathcal{V}=V$, let $S=V^c$ and $\Gamma=\partial S$. Then we have*

$$\Delta\langle \#(\mathcal{C}_*^{\Lambda_N}\cap S\cap(\Lambda_N\setminus\Lambda_{N/4})) \mid (\sigma^{\Lambda_N,\pm}, \widetilde{\sigma}^{\Lambda_N,\pm})_V\rangle_{\pi_{\Lambda_N}} \leqslant 8\#\{\Gamma\cap\mathcal{C}_*^{\Lambda_N}\}\,.$$

Proof. Quench on the realization of $(\sigma^{\Lambda_N,\pm}, \widetilde{\sigma}^{\Lambda_N,\pm})_\Gamma$ as in (2.3.36). By Lemmas 2.3.13 and 2.3.15,

$$\begin{aligned}\Delta\langle \#(\mathcal{C}_*^{S,\tau^{\pm},\widetilde{\tau}^{\pm}}\cap(\Lambda_N\setminus\Lambda_{N/4}))\rangle_{\pi_S} &\leqslant 8\#\{v\in\Gamma : \widehat{\tau}_v^{+} = \widehat{\widetilde{\tau}}_v^{+} = 1, \widehat{\tau}_v^{-} = \widehat{\widetilde{\tau}}_v^{-} = -1\}\\ &= 8\#\{v\in\Gamma : \tau_v^{+} = \widetilde{\tau}_v^{+} = 1, \tau_v^{-} = \widetilde{\tau}_v^{-} = -1\}\,,\end{aligned}$$

where the equality follows from (2.3.38). Combined with (2.3.35), this completes the proof of the corollary. $\square$

Analysis of the adaptive admissible coupling

We now analyze the adaptive admissible coupling π_{Λ_N}. Recall that $\ell = \lfloor \frac{1}{4} N^{1-\alpha'} \rfloor$ and $K = \lfloor N^{\alpha\alpha'} \rfloor$, and define $\mathcal{D}_N$ to be the event (measurable with respect to the Gaussian field) by

$$\mathcal{D}_N = \{\pi_{\Lambda_N}(\min_{1\leqslant j\leqslant \ell} d_{\mathcal{C}^{\Lambda_N}}(\partial\Lambda_{N/2-jN^{\alpha'}}, \partial\Lambda_{N/2-(j-1)N^{\alpha'}}) \leqslant K) \geqslant N^{-20}\}\,. \tag{2.3.45}$$

By Proposition 2.3.1 and a simple Markov's inequality, we see that for $C = C(\varepsilon,\beta) > 0$

$$\mathbb{P}(\mathcal{D}_N) \leqslant CN^{-20}\,. \tag{2.3.46}$$

In what follows, we quench on the Gaussian field at which $\mathcal{D}_N$ does not occur.

Lemma 2.3.17. *We have that $\pi_{\Lambda_N}(o \in \mathcal{C}_*^{\Lambda_N}) \leqslant CN^{-10}$ on $\mathcal{D}_N^c$, for $C = C(\varepsilon,\beta) > 0$.*

Proof. For $1 \leqslant j \leqslant \ell$, $1 \leqslant k \leqslant K$, let $\mathcal{E}_{j,k,\emptyset}, \mathcal{E}_{j,k,d}, V_{j,k}, A_{j,k}$ be defined as in Section 2.3.3. For each $1 \leqslant j \leqslant \ell$, let $\mathcal{E}_{j,\emptyset} = \cup_{i=1}^{j} \cup_{k=1}^{K} \mathcal{E}_{i,k,\emptyset}$ and define

$$m_j^* = \langle \#(\mathcal{C}_*^{\Lambda_N} \cap (\Lambda_{N/2-(j-1)N^{\alpha'}} \setminus \Lambda_{N/2-jN^{\alpha'}}))\mathbb{1}_{\mathcal{E}_{j-1,\emptyset}^c}\rangle_{\pi_{\Lambda_N}}\,.$$

By Corollary 2.3.14, it suffices to prove that $m_\ell^* \leqslant 2N^{-10}$. To this end, it suffices to prove that for $N \geqslant N_0 = N_0(\varepsilon,\beta)$ (where N_0 is to be selected)

$$m_{j+1}^* \leqslant 10^{-3} m_j^* + N^{-10} \text{ for all } 1 \leqslant j \leqslant \ell - 1\,. \tag{2.3.47}$$

Let $\mathcal{E}_{j,d} = \cup_{i=1}^{j} \cup_{k=1}^{K} \mathcal{E}_{i,k,d}$. Since $\pi_{\Lambda_N}(\mathcal{E}_{j,d}) \leqslant CN^{-20}$ on $\mathcal{D}_N^c$, it suffices to show that

$$\langle \#(\mathcal{C}_*^{\Lambda_N} \cap (\Lambda_{N/2-jN^{\alpha'}} \setminus \Lambda_{N/2-(j+1)N^{\alpha'}}))\mathbb{1}_{\mathcal{E}_{j,\emptyset}^c}\mathbb{1}_{\mathcal{E}_{j,d}^c}\rangle_{\pi_{\Lambda_N}} \leqslant 10^{-3} m_j^*\,. \tag{2.3.48}$$

Fix $1 \leqslant j \leqslant \ell$. For $1 \leqslant k \leqslant K$, write $\mathcal{E}_{j,\leqslant k,\emptyset} = \mathcal{E}_{j-1,\emptyset} \cup \cup_{i=1}^{k} \mathcal{E}_{j,i,\emptyset}$ and $\mathcal{E}_{j,\leqslant k,d} = \mathcal{E}_{j-1,d} \cup$

$\cup_{i=1}^{k}\mathcal{E}_{j,i,d}$. Thus, we can deduce that

$$\begin{aligned}
&\Delta\langle \#(\mathcal{C}_*^{\Lambda_N} \cap (\Lambda_{N/2-jN^{\alpha'}} \setminus \Lambda_{N/2-(j+1)N^{\alpha'}}))\mathbb{1}_{\mathcal{E}^c_{j,\leqslant k,\emptyset}}\mathbb{1}_{\mathcal{E}^c_{j,\leqslant k,d}} \mid (\sigma^{\Lambda_N,\pm}, \widetilde{\sigma}^{\Lambda_N,\pm})_{V_{j,k}}\rangle_{\pi_{\Lambda_N}} \\
&= \mathbb{1}_{\mathcal{E}^c_{j,\leqslant k,\emptyset}}\mathbb{1}_{\mathcal{E}^c_{j,\leqslant k,d}}\Delta\langle \#(\mathcal{C}_*^{\Lambda_N} \cap (\Lambda_{N/2-jN^{\alpha'}} \setminus \Lambda_{N/2-(j+1)N^{\alpha'}})) \mid (\sigma^{\Lambda_N,\pm}, \widetilde{\sigma}^{\Lambda_N,\pm})_{V_{j,k}}\rangle_{\pi_{\Lambda_N}} \\
&\leqslant 8\#A_{j,k} \cdot \mathbb{1}_{\mathcal{E}^c_{j,\leqslant k,\emptyset}}\mathbb{1}_{\mathcal{E}^c_{j,\leqslant k,d}}\,,
\end{aligned}$$

where the equality holds since $\mathcal{E}_{j,\leqslant k,\emptyset}$ and $\mathcal{E}_{j,\leqslant k,d}$ are measurable with respect to $(\sigma^{\Lambda_N,\pm}, \widetilde{\sigma}^{\Lambda_N,\pm})_{V_{j,k}}$, and the inequality is obtained by applying Corollary 2.3.16 with $V = V_{j,k}$ (note that $\Lambda_{N/2-jN^{\alpha'}} \cap V_{j,k} = \emptyset$ on the event $\mathcal{E}^c_{j,\leqslant k,d}$). Averaging over the conditioning in the preceding display and recalling that $\mathcal{E}_{j-1,\emptyset} \subseteq \mathcal{E}_{j,\leqslant k,\emptyset} \subseteq \mathcal{E}_{j,\emptyset}$ and $\mathcal{E}_{j,\leqslant k,d} \subseteq \mathcal{E}_{j,d}$, we deduce that

$$\Delta\langle \#(\mathcal{C}_*^{\Lambda_N} \cap (\Lambda_{N/2-jN^{\alpha'}} \setminus \Lambda_{N/2-(j+1)N^{\alpha'}}))\mathbb{1}_{\mathcal{E}^c_{j,\emptyset}}\mathbb{1}_{\mathcal{E}^c_{j,d}}\rangle_{\pi_{\Lambda_N}} \leqslant \langle 8\#A_{j,k} \cdot \mathbb{1}_{\mathcal{E}^c_{j-1,\emptyset}}\mathbb{1}_{\mathcal{E}^c_{j,\leqslant k,d}}\rangle_{\pi_{\Lambda_N}}\,.$$

Since $\sum_{k=1}^{K} \#A_{j,k} \cdot \mathbb{1}_{\mathcal{E}^c_{j,\leqslant k,d}} \leqslant \#(\mathcal{C}_*^{\Lambda_N} \cap (\Lambda_{N/2-(j-1)N^{\alpha'}} \setminus \Lambda_{N/2-jN^{\alpha'}}))$, summing the preceding display over $1 \leqslant k \leqslant K$ yields (2.3.48) (recall that $\Delta K = N^{-\alpha(\alpha')^2}\lfloor N^{\alpha\alpha'}\rfloor \geqslant 10^5$ if $N \geqslant N_0$ for large enough N_0). This completes the proof of the lemma. □

2.3.4. Proof of Theorem 2.1.1 for positive temperature

We continue to consider $\widetilde{h}^{(N)}$ defined as in (2.3.34), and let $\mu^{\Lambda_N,\pm}, \widetilde{\mu}^{\Lambda_N,\pm}, \pi_{\Lambda_N}$ be defined as in Section 2.3.3. For $\delta > 0$, let $Q_\delta \subseteq [-1,1]$ be the collection of multiples of δ, and for $q \in Q_\delta$ define $\mathcal{E}^*_{o,N,q}$ to be an event measurable with respect to the Gaussian field by (the tilde symbol only applies on the minus version below)

$$\mathcal{E}^*_{o,N,q} = \{\langle \sigma_o^{\Lambda_N,+}\rangle_{\mu^{\Lambda_N,+}} \geqslant q+\delta, \langle \widetilde{\sigma}_o^{\Lambda_N,-}\rangle_{\widetilde{\mu}^{\Lambda_N,-}} \leqslant q-\delta\}\,. \tag{2.3.49}$$

By admissibility, on the event $\mathcal{E}^*_{o,N,q}$ we have $\pi_{\Lambda_N}(o \in \mathcal{C}_*^{\Lambda_N}) \geqslant \delta$. Combined with Lemma 2.3.17 and (2.3.46), it yields that

$$\mathbb{P}(\mathcal{E}^*_{o,N,q}) = O(N^{-10}/\delta)\,. \tag{2.3.50}$$

(Throughout, $O(1)$ hides a constant that may depend on (ε, β).) Next, we define

$$\mathcal{E}_{o,N,q} = \{\langle \sigma_o^{\Lambda_N,+}\rangle_{\mu^{\Lambda_N,+}} \geqslant q+\delta, \langle \sigma_o^{\Lambda_N,-}\rangle_{\mu^{\Lambda_N,-}} \leqslant q-\delta\}\,. \tag{2.3.51}$$

By monotonicity, we thus have

$$\mathcal{E}_{o,N,q} \subseteq \mathcal{E}_{o,N',q} \text{ and } \mathcal{E}^*_{o,N,q} \subseteq \mathcal{E}^*_{o,N',q} \text{ for all } N' \leqslant N\,. \tag{2.3.52}$$

Lemma 2.3.18. *Let $\delta = N^{-3}/3$. There exists $C = C(\varepsilon, \beta) > 0$ such that $\mathbb{P}(\mathcal{E}_{o,N,q}) \leqslant CN^{-6}$ for all $q \in Q_\delta$.*

Proof. While the proof of the lemma is similar to that of Lemma 2.2.14, we nevertheless provide a self-contained proof for clarity of exposition.

For $A \subseteq \mathbb{Z}^2$, we set $h_A = \sum_{v\in A} h_v$. Without loss of generality, let us only consider $N = 4^n$ for some $n \geqslant 1$, and for $1 \leqslant \ell \leqslant n$, we define $\{\widetilde{h}_v^{(4^\ell)} : v \in \Lambda_{4^\ell}\}$ as in (2.3.34). Write $\mathfrak{A}_\ell = \Lambda_{4^\ell} \setminus \Lambda_{4^{\ell-1}}$. For $0.9n \leqslant \ell \leqslant n$, let $\mathcal{F}_\ell = \sigma(h_v : v \in \Lambda_{4^\ell})$ and write

$$h_v = (\#\mathfrak{A}_\ell)^{-1} h_{\mathfrak{A}_\ell} + g_v \text{ for } v \in \mathfrak{A}_\ell\,, \tag{2.3.53}$$

where $\{g_v : v \in \mathfrak{A}_\ell\}$ is a mean-zero Gaussian process independent of $h_{\mathfrak{A}_\ell}$ and $\{g_v : v \in \mathfrak{A}_\ell\}$ for $0.9n \leqslant \ell \leqslant n$ are mutually independent. Let $\mathcal{F}'_\ell$ be the σ-field which contains every event in $\mathcal{F}_\ell$ that is independent of $h_{\mathfrak{A}_\ell}$ (so in particular $\mathcal{F}_\ell \subseteq \mathcal{F}'_{\ell+1} \subseteq \mathcal{F}_{\ell+1}$). Write $\mathcal{E}_* = \cup_{0.9n\leqslant\ell\leqslant n}\mathcal{E}^*_{o,4^\ell,q}$. By monotonicity of $\langle \sigma_o^{\Lambda_N,+}\rangle_{\mu^{\Lambda_N,+}}$ and $\langle \sigma_o^{\Lambda_N,-}\rangle_{\mu^{\Lambda_N,-}}$ with respect to the external field, there exists an interval I_ℓ measurable with respect to $\mathcal{F}'_\ell$ such that conditioned on $\mathcal{F}'_\ell$ we have $\mathcal{E}_{o,4^\ell,q}$ occurs if and only if $h_{\mathfrak{A}_\ell} \in I_\ell$. Let I'_ℓ be the maximal sub-interval of I_ℓ which shares the upper endpoint and $|I'_\ell| \leqslant \frac{\#\mathfrak{A}_\ell}{4^{\alpha(\alpha')^2\ell}}$ (here $|I'_\ell|$ denotes the length of the interval I'_ℓ). By definition in (2.3.49) and (2.3.34), we see from (2.3.53) that conditioned on $\mathcal{F}'_\ell$ we have that $\mathcal{E}_{o,4^\ell,q} \cap (\mathcal{E}^*_{o,4^\ell,q})^c$ occurs only if $h_{\mathfrak{A}_\ell} \in I'_\ell$. Thus,

$$\mathbb{P}(\mathcal{E}_{o,4^\ell,q} \cap (\mathcal{E}^*_{o,4^\ell,q})^c \mid \mathcal{F}'_\ell) \leqslant \mathbb{P}(h_{\mathfrak{A}_\ell} \in I'_\ell)\,, \text{ for } 0.9n \leqslant \ell \leqslant n\,.$$

Combined with the fact that $\mathbb{V}\mathrm{ar}(h_{\mathfrak{A}_\ell}) = \varepsilon^2 \# \mathfrak{A}_\ell$, this gives that for $C = C(\varepsilon, \beta) > 0$ (whose value may be adjusted below)

$$\mathbb{P}(\mathcal{E}_{o,4^\ell,q} \cap (\mathcal{E}^*_{o,4^\ell,q})^c \mid \mathcal{F}'_\ell) \leqslant \frac{C}{4^{\ell(\alpha(\alpha')^2-1)}}\,.$$

By (2.3.52), we have $\mathcal{E}_{o,N,q} \cap \mathcal{E}^c_* = \cap_{\ell=0.9n}^n (\mathcal{E}_{o,4^\ell,t} \cap (\mathcal{E}^*_{o,4^\ell,q})^c)$. Since $(\mathcal{E}_{o,4^\ell,t} \cap (\mathcal{E}^*_{o,4^\ell,q})^c)$ is $\mathcal{F}_\ell$-measurable (and thus is $\mathcal{F}'_{\ell+1}$-measurable), we deduce that (recalling $\alpha(\alpha')^2 > 1$)

$$\mathbb{P}(\mathcal{E}_{o,N,q} \cap \mathcal{E}^c_*) \leqslant CN^{-6}\,.$$

By (2.3.50), we have $\mathbb{P}(\mathcal{E}_*) \leqslant CN^{-6}$. Combined with the preceding display, this completes the proof of the lemma. □

Define $\mathcal{E}_{o,N}$ to be an event measurable with respect to the Gaussian field by

$$\mathcal{E}_{o,N} = \{\langle \sigma_o^{\Lambda_N,+} \rangle_{\mu^{\Lambda_N,+}} - \langle \sigma_o^{\Lambda_N,-} \rangle_{\mu^{\Lambda_N,-}} \geqslant N^{-3}\}\,. \tag{2.3.54}$$

Since $\mathcal{E}_{o,N} \subseteq \cup_{q \in Q_\delta} \mathcal{E}_{o,N,q}$ with $\delta = N^{-3}/3$, we get from Lemma 2.3.18 that $\mathbb{P}(\mathcal{E}_{o,N}) = O(N^{-3})$. Thus,

$$\begin{aligned}\mathbb{E}(\langle \sigma_o^{\Lambda_N,+} \rangle_{\mu^{\Lambda_N,+}} - \langle \sigma_o^{\Lambda_N,-} \rangle_{\mu^{\Lambda_N,-}}) &\leqslant 2\mathbb{P}(\mathcal{E}_{o,N}) + \mathbb{E}(\mathbb{1}_{\mathcal{E}^c_{o,N}}(\langle \sigma_o^{\Lambda_N,+} \rangle_{\mu^{\Lambda_N,+}} - \langle \sigma_o^{\Lambda_N,-} \rangle_{\mu^{\Lambda_N,-}})) \\ &= O(N^{-3})\,. \end{aligned} \tag{2.3.55}$$

Remark 2.3.19. In Lemma 2.3.18, we work with $\mathcal{E}_{o,N,q}$ other than $\mathcal{E}_{o,N}$, for the reason that we do not have the property that $\mathcal{E}_{o,N}$ occurs if and only if $h_{\mathfrak{A}_{\ell+1}}$ is in a certain interval (but the property holds for $\mathcal{E}_{o,N,q}$).

In order to prove Theorem 2.1.1, we will consider a monotone coupling of $\mu^{\Lambda_N,\pm}$ and consider $\mathcal{C}^{\Lambda_N} = \{v \in \Lambda_N : \sigma_v^{\Lambda_N,+} > \sigma_v^{\Lambda_N,-}\}$. We wish to have that $\{o \in \mathcal{C}^{\Lambda_N}\}$ occurs only if o is connected to $\partial\Lambda_N$ in $\mathcal{C}^{\Lambda_N}$. However, as we have seen in Remark 2.3.8, this property does not hold for all monotone couplings of $\mu^{\Lambda_N,\pm}$ (For instance if we build an adaptive admissible coupling by first sampling the spin at o and then the rest of the spins, then it is possible to get a configuration where the spin disagrees at o but there exists a contour surrounding o where all spins agree on this contour). In order to address this issue, we will construct an adaptive admissible coupling $\overline{\pi}_{\Lambda_N}$ such that this percolation property holds. Our construction is similar to that in Section 2.3.3 in a way that we explore $\mathcal{C}^{\Lambda_N}$ in a breadth first search order. But our construction now is much simpler as we no longer need to consider multiple phases.

By Definition 2.3.9, in order to define $\overline{\pi}_{\Lambda_N}$ we only need to specify the order of vertices in which we sample the spins, as described as follows. Throughout the procedure, we let $\mathcal{C}^{\Lambda_N}$ be the collection of vertices v which have been sampled and satisfy $\sigma_v^{\Lambda_N,+} > \sigma_v^{\Lambda_N,-}$. We set $A_0 = \partial\Lambda_N$ and for $k = 0, 1, 2, \ldots$, we inductively employ the following procedure (which we refer to as stage).

- At stage $k+1$, first set $A_{k+1} = \emptyset$. If $A_k = \emptyset$, we sample the unexplored vertices in Λ_N in an (arbitrary) prefixed order and stop our procedure. Otherwise, we explore all the unexplored neighbors of A_k (in a certain arbitrary prefixed order) and sample the spins at these vertices.
- For each newly sampled vertex, if it is in $\mathcal{C}^{\Lambda_N}$ then we add it to A_{k+1}.

Lemma 2.3.20. *Under the coupling $\overline{\pi}_{\Lambda_N}$, $o \in \mathcal{C}^{\Lambda_N}$ only if o is connected to $\partial\Lambda_N$ in $\mathcal{C}^{\Lambda_N}$.*

Proof. Let k_* be the first k such that $A_k = \emptyset$. If o has been explored by the end of Stage $(k_* - 1)$, we see that o is connected to $\partial\Lambda_N$ in $\mathcal{C}^{\Lambda_N}$. Otherwise, denote V_{k_*} the collection of explored vertices at the end of Stage (k_*). If o was explored in Stage k_*, then $o \notin \mathcal{C}^{\Lambda_N}$ (since $A_{k_*} = \emptyset$). If o was not explored by the end of Stage k^*, we see that $\sigma^{\Lambda_N,+}$ and $\sigma^{\Lambda_N,-}$ agree on $\partial V_{k_*}^c$, and thus they will have to agree with each other on $V_{k_*}^c$ by

Lemma 2.3.11 (this is because $\sigma_v^{\Lambda_N,+}$ and $\sigma^{\Lambda_N,-}$ have the same conditional marginal for all $v \in V_{k_*}^c$ and thus have to agree with each other in an admissible coupling). This in particular implies that $o \notin \mathcal{C}^{\Lambda_N}$, completing the proof of the lemma. $\square$

Proof of Theorem 2.1.1: $T > 0$. Consider the adaptive admissible coupling $\overline{\pi}_{\Lambda_N}$. We will use the fact that $\mathbb{P} \otimes \overline{\pi}_{\Lambda_N}(v \in \mathcal{C}^{\Lambda_N}) = \frac{1}{2}\mathbb{E}(\langle \sigma_v^{\Lambda_N,+}\rangle_{\mu^{\Lambda_N,+}} - \langle \sigma_v^{\Lambda_N,-}\rangle_{\mu^{\Lambda_N,-}})$ for all $v \in \Lambda_N$. Let $N_0 = N_0(\varepsilon, \beta)$ be chosen later. For any box B, recall that B^{large} is the box concentric with B of doubled side length. For $B \in \mathcal{B}(N, N_0)$, we say B is open if $\mathcal{C}^{\Lambda_N} \cap B \neq \emptyset$. In order to analyze this percolation process, we say a box B is exceptional if $\sum_{v\in B}(\langle \sigma_v^{B^{\mathrm{large}},+}\rangle_{\mu^{B^{\mathrm{large}},+}} - \langle \sigma_v^{B^{\mathrm{large}},-}\rangle_{\mu^{B^{\mathrm{large}},-}}) \geqslant N_0^{-1/2}$ (so exceptional is a property measurable with respect to $\{h_v : v \in B^{\mathrm{large}}\}$). By (2.3.55) and monotonicity,

$$\mathbb{P}(B \text{ is exceptional}) \leqslant N_0^{1/2} \sum_{v\in B} \mathbb{E}(\langle \sigma_v^{B^{\mathrm{large}},+}\rangle_{\mu^{B^{\mathrm{large}},+}} - \langle \sigma_v^{B^{\mathrm{large}},-}\rangle_{\mu^{B^{\mathrm{large}},-}}) = O(N_0^{-1/2})\,.$$

Recall Definition 2.2.9. We see that the exceptional boxes on $\mathcal{B}(N, N_0)$ form a percolation process which satisfies the $(N, N_0, 4, p)$-condition with $p = O(N_0^{-1/2})$. In addition, for any box B which is not exceptional, denoting by $\mathcal{F}_B$ the σ-field generated by spin configurations outside B^{large}, we see from monotonicity that

$$\overline{\pi}_{\Lambda_N}(B \text{ is open } \mid \mathcal{F}_B) \leqslant \sum_{v\in B}(\langle \sigma_v^{B^{\mathrm{large}},+}\rangle_{\mu^{B^{\mathrm{large}},+}} - \langle \sigma_v^{B^{\mathrm{large}},-}\rangle_{\mu^{B^{\mathrm{large}},-}}) = O(N_0^{-1/2})\,.$$

Altogether, this implies that the collection of open boxes forms a percolation process which also satisfies the $(N, N_0, 4, p)$-condition with $p = O(N_0^{-1/2})$. By Lemma 2.3.20, in order for $o \in \mathcal{C}^{\Lambda_N}$, it is necessary that there exists an open lattice animal on $B \in \mathcal{B}(N, N_0)$ with size at least $\frac{N}{10N_0}$. Now, choosing N_0 sufficiently large (so that p is sufficiently small) and applying Lemma 2.2.10 yields that

$$\mathbb{P} \otimes \overline{\pi}_{\Lambda_N}(o \in \mathcal{C}^{\Lambda_N}) \leqslant c^{-1}e^{-cN} \text{ for } c = c(\varepsilon, \beta) > 0\,,$$

completing the proof of the theorem. □

Acknowledgement. We thank Tom Spencer for introducing the problem to us, thank Steve Lalley for many interesting discussions and thank Subhajit Goswami, Steve Lalley for a careful reading of an earlier version of the manuscript. We also thank Michael Aizenman and Ron Peled for helpful conversations. We thank two anonymous referees for many helpful suggestions on exposition.

CHAPTER 3

Hamilton–Jacobi equations for inference of matrix tensor products

This chapter is essentially borrowed from [39], joint with Hong-Bin Chen.

Abstract. We study the high-dimensional limit of the free energy associated with the inference problem of finite-rank matrix tensor products. In general, we bound the limit from above by the unique solution to a certain Hamilton–Jacobi equation. Under additional assumptions on the nonlinearity in the equation which is determined explicitly by the model, we identify the limit with the solution. Two notions of solutions, weak solutions and viscosity solutions, are considered, each of which has its own advantages and requires different treatments. For concreteness, we apply our results to a model with i.i.d. entries and symmetric interactions. In particular, for the first order and even order tensor products, we identify the limit and obtain estimates on convergence rates; for other odd orders, upper bounds are obtained.

3.1. Introduction

Tensor factorizations or tensor decompositions play important roles in numerous applications. In this work, we study the inference problem of estimating tensor products of matrices. Let us first describe the model we are concerned with. Fix $K \in \mathbb{N}$ and let P_N^X be the law of $X \in \mathbb{R}^{N\times K}$, where $N \in \mathbb{N}$ will be sent to ∞. For a fixed $L \in \mathbb{N}$, we observe

$$Y = \sqrt{\frac{2t}{N^{p-1}}} X^{\otimes p} A + W \in \mathbb{R}^{N^p \times L}. \tag{3.1.1}$$

where $t \geqslant 0$ is interpreted as the signal-to-noise ratio; $\otimes$ is the Kronecker product (hence $X^{\otimes p} \in \mathbb{R}^{N^p \times K^p}$); $A \in \mathbb{R}^{K^p \times L}$ is a deterministic matrix; and $W \in \mathbb{R}^{N^p \times L}$ consists of independent standard Gaussian entries.

The inference task is to recover the information of X based on the observation of Y. Hence, we investigate the law of X conditioned on observing Y. Bayes' rule gives that, for any bounded measurable $g : \mathbb{R}^{N\times K} \to \mathbb{R}$, we have

$$\mathbb{E}\big[g(X)\big|Y\big] = \frac{\int_{\mathbb{R}^{N\times K}} g(x)e^{H_N^\circ(t,x)} P_N^X(\mathrm{d}x)}{\int_{\mathbb{R}^{N\times K}} e^{H_N^\circ(t,x)} P_N^X(\mathrm{d}x)}.$$

Here the Hamiltonian associated with this model is given by

$$H_N^\circ(t,x) = \sqrt{\frac{2t}{N^{p-1}}}(x^{\otimes p}A)\cdot Y - \frac{t}{N^{p-1}}|x^{\otimes p}A|^2. \tag{3.1.2}$$

Throughout this paper, the dot product between two tensors, matrices or vectors of the same size is the entry-wise inner product. We denote by $|\cdot|$ the associated norm. The goal is to understand the high-dimensional limit as $N\to\infty$ of the free energy

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

We briefly discuss the generality of the model (3.1.1) and its relation to other models involving the inference of matrix products. Among the ones widely studied are the models concerning the second order products. The inference problem of nonsymmetric matrices (or the spiked Wishart model) is given by $Y = \sqrt{\frac{2t}{N}}X_1X_2^\intercal + W$. Works investigating this model include [90, 14, 12, 79, 86, 36]. When $X_1 = X_2$, this becomes the inference problem of symmetric matrices (or the spiked Wigner model), which is studied in [82, 52, 95, 94]. A generalization of these spiked matrix models can be seen in the study of community detection problems and the stochastic block models. In certain settings, the community detection problem is asymptotically equivalent to $Y = \sqrt{\frac{2t}{N}}XBX^\intercal + W$ where B is deterministic and models the community interactions (see [104]). More generally, the community detection with several correlated networks is asymptotically equivalent to the multiview spiked matrix model $Y_l = \sqrt{\frac{2t}{N}}XB_lX^\intercal + W_l$ for $l = 1, 2, \ldots, L$ where each B_l reflects one network (see [87, 88]). All of these second order models can be represented in the form of

$Y = \sqrt{\frac{2t}{N}} X^{\otimes 2}\sqrt{S} + W$ where S is a positive semidefinite matrix. This model is studied in [103], and its equivalence to the models above is discussed in more details therein. Hence, the models so far mentioned can be seen as special cases of (3.1.1) for $p = 2$. In Appendix 3.8, we will demonstrate the representation of the nonsymmetric matrix inference problem into the form of (3.1.1). Higher order cases ($p \geqslant 2$) include $Y = \sqrt{\frac{2t}{N^{p-1}}} X^{\otimes p} + W$ with vector $X \in \mathbb{R}^N$ in [12, 95], and $Y = \sqrt{\frac{2t}{N^{p-1}}} \sum_{k=1}^{r} X_k^{\otimes p} + W$ with each vector $X_k \in \mathbb{R}^N$ in [83]. The model (3.2.16) studied in [85] and considered in Section 3.2.3 as a special case also belongs to this class. Again, they can be viewed as special cases of (3.1.1).

Recently, the powerful method of adaptive interpolations was introduced in [12]. This technique and its improvements have been employed in works including [11, 86, 103]. In this work, we follow the approach via Hamilton–Jacobi equations set forth in [95, 94, 92, 98, 96, 93]. Let $F_N(t, h)$ be the free energy corresponding to an enriched version of the Hamiltonian (3.1.2). Here h is an additional variable and the original free energy satisfies $F_N^\circ(t) = F_N(t, 0)$. We seek to compare the limit of $\mathbb{E}F_N(t, h)$ as $N \to \infty$ with the solution of the following Hamilton–Jacobi equation

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

Here the nonlinearity H is given by a simple formula (3.2.6) in terms of the interaction matrix A in (3.1.1). To make sense of solutions of this equation and the convergence, two notions have been explored. The notion of viscosity solutions of Hamilton–Jacobi equations was initially adopted to study convergence of free energies in [95] and later the notion of weak solutions was taken in [94]. Viscosity solutions are in general heavier to handle. Bounds from two sides require different treatments, and often one side is much easier than the other and requires weaker assumptions. The convergence happens in the local $L_t^\infty L_h^\infty$ topology while it takes considerable effort to obtain convergence rates. On the other hand, weak solutions are simpler and it is easier to obtain estimates on convergence rates, although the convergence takes place in local $L_t^\infty L_h^1$. It can be upgraded to estimates in $L_t^\infty L_h^\infty$ by giving up some powers (see Remark 3.2.4). A more detailed comparison of these two notions of solutions can be found in [94, Section 2].

We utilize both notions in this work. For any interaction matrix A (equivalently, for any H of the form (3.2.6)), we obtain an upper bound on the limit of the free energy in Theorem 3.2.2 via viscosity solutions. This theorem also gives the corresponding lower bound under an additional assumption that H is convex. Employing weak solutions as in Theorem 3.2.1, we obtain convergence and estimates on convergence rates under an assumption on H which is weaker than convexity.

We emphasize that, different from the usual approach in statistical mechanics, the existence of a variational formula for the limit of free energies is not *a priori* needed in our approach. Instead, the existence of solutions to the Hamilton–Jacobi equation is sufficient. In the weak solution approach, we prove the existence in a straightforward manner by verifying that the free energies form a Cauchy sequence. For viscosity solutions, there are classical tools to ensure existence. Here, we prove that the Hopf formula is a viscosity solution as a useful fact (see Remark 3.2.5), and simply use this to furnish the existence for convenience.

The rest of the paper is organized as follows. We describe the setting and state main results in Section 3.2. We apply these results to a special case where X has i.i.d. entries and the interaction is symmetric in Section 3.2.3. In Section 3.3, we show that the free energy satisfies an approximate Hamilton–Jacobi equation and collect some basic results of the derivatives of the free energy. Section 3.4 gives the precise definition of weak solutions and the uniqueness of solutions. In Section 3.5, we show the convergence of the free energy to a weak solution, and finish the proof of Theorem 3.2.1. The definition of viscosity solutions and the corresponding well-posedness results are in Section 3.6. The ensuing Section 3.7 studies the convergence of the free energy to the viscosity solution and proves Theorem 3.2.2. A special version of the Fenchel–Moreau biconjugation theorem on the set of positive semidefinite matrices is needed to analyze the Hopf formula. It is stated and proved in Appendix 3.9.

Acknowledgement

We warmly thank Jean–Christophe Mourrat for many helpful comments and discussions.

3.2. Setting and Main results

3.2.1. Setting

We assume that the random matrix $X \in \mathbb{R}^{N\times K}$ in (3.1.1) satisfies

$$|X| \leqslant \sqrt{NK}. \tag{3.2.1}$$

For convenience, we use the shorthand notation

$$\widetilde{x} = x^{\otimes p} A, \quad \forall x \in \mathbb{R}^{N\times K}. \tag{3.2.2}$$

We enrich the Hamiltonian (3.1.2) by introducing

$$\begin{aligned} H_N(t,h,x) = \sqrt{\frac{2t}{N^{p-1}}}\widetilde{x}\cdot Y - \frac{t}{N^{p-1}}|\widetilde{x}|^2 \\ + \sqrt{2h}\cdot (x^\intercal \overline{Y}) - h\cdot (x^\intercal x). \end{aligned} \tag{3.2.3}$$

Here $\overline{Y} = X\sqrt{2h} + Z$, where $h \in \mathbb{S}^K_+$, the set of $K\times K$ (symmetric) positive semi-definite matrices, and entries of $Z \in \mathbb{R}^{N\times K}$ are independent standard Gaussian variables. This Hamiltonian H_N is associated with the law of X conditioned on observing both Y and $\overline{Y}$. The corresponding free energy is given by

$$F_N(t,h) = \frac{1}{N}\log\int_{\mathbb{R}^{N\times K}} e^{H_N(t,h,x)} P^X_N(\mathrm{d}x). \tag{3.2.4}$$

Let $\overline{F}_N(t,h) = \mathbb{E}F_N(t,h)$ be its expectation.

Set $\mathbb{R}_+ = [0,\infty)$. We consider the Hamilton–Jacobi equation

$$\partial_t f - \mathsf{H}(\nabla f) = 0, \quad \text{in } \mathbb{R}_+\times\mathbb{S}^K_+ \tag{3.2.5}$$

where $\mathsf{H} : \mathbb{S}^K_+ \to \mathbb{R}$ is given by

$$\mathsf{H}(q) = \big(AA^\intercal\big) \cdot q^{\otimes p}, \quad \forall q \in \mathbb{S}^K_+. \tag{3.2.6}$$

3.2.2. Main results

To state the results, we need more notation. Let us introduce

$$\mathbb{S}^K_{+,M} = \big\{h \in \mathbb{S}^K_+ : \ |h| \leqslant M\big\}. \tag{3.2.7}$$

We also denote the set of $K \times K$ symmetric matrices by $\mathbb{S}^K$, and the set of $K \times K$ symmetric positive definite matrices by $\mathbb{S}^K_{++}$. For $N \in \mathbb{N}$ and $M > 0$, we define

$$\mathcal{K}_{M,N} = \left(\mathbb{E} \sup_{(t,h)\in[0,M]\times\mathbb{S}^K_{+,M}} \big|F_N - \overline{F}_N\big|^2\right)^{\frac{1}{2}}, \tag{3.2.8}$$

and for any function $\psi : \mathbb{S}^K_+ \to \mathbb{R}$,

$$\mathcal{L}_{\psi,M,N} = \sup_{h\in\mathbb{S}^K_{+,M}} \big|\overline{F}_N(0,h) - \psi(h)\big|. \tag{3.2.9}$$

The quantity $\mathcal{K}_{M,N}$ measures the concentration of F_N. Many tools are available to estimate this. In view of (3.2.3) and (3.2.4), we can recast $\overline{F}_N(0, h)$ as the free energy corresponding to a decoupled system (inference of X based on the observation of $\overline{Y}$ with $\overline{Y}$ in (3.2.3)). Hence, $\mathcal{L}_{\psi,M,N}$ is also a relatively simple object to analyze.

Throughout, the gradient ∇ is taken in the space variable $h \in \mathbb{S}^K_+$ (sometimes written as $x \in \mathbb{S}^K_+$). To avoid confusion when multiple ∇ are present, we specifically denote the differential of H by $\mathcal{D}\mathsf{H}$. We identify $\mathbb{S}^K$ with $\mathbb{R}^{K(K+1)/2}$ in an isometric way (see (3.4.1)) and endow it with the Lebesgue measure. Let $\mathcal{A}$ be the set of real-valued nondecreasing, Lipschitz and convex functions on $\mathbb{S}^K_+$. Here a function $u : \mathbb{S}^K_+ \to \mathbb{R}$ is said to be nondecreasing provided

$$u(a) \geqslant u(b), \quad \text{if } a - b \in \mathbb{S}^K_+. \tag{3.2.10}$$

We define

$$\mathcal{A}_{\mathsf{H}} = \Big\{\phi \in \mathcal{A} : \ \nabla \cdot \big(\mathcal{D}\mathsf{H}(\nabla\phi)\big) \geqslant 0\Big\}, \tag{3.2.11}$$

where the inequality is understood in the sense of distribution, namely $\int \mathcal{D}\mathsf{H}(\nabla\phi) \cdot \nabla\eta \leqslant 0$, for all nonnegative smooth function η compactly supported on $\mathbb{S}^K_{++}$.

Before stating the theorems, we comment that the assumptions imposed in them are threefold. The first part is on the concentration, namely, the quantity $\mathcal{K}_{M,N}$. The second part is on $\overline{F}_N(0,\cdot)$ or $\mathcal{L}_{\psi,M,N}$, which is about the convergence of the free energy in the aforementioned decoupled system. The third part is on H (equivalently on A due to (3.2.6)) or, further, on $\mathcal{A}_{\mathsf{H}}$.

Theorem 3.2.1. *Let $p \in \mathbb{N}$. Suppose*

- $\sup_{M\geqslant 1, N\in\mathbb{N}}(\mathcal{K}_{M,N}/M^\beta) < \infty$ *for some* $\beta > 0$, *and* $\lim_{N\to\infty}\mathcal{K}_{M,N} = 0$ *for each* $M \geqslant 1$;
- *there is a function* $\psi : \mathbb{S}^K_+ \to \mathbb{R}$ *such that* $\lim_{N\to\infty}\mathcal{L}_{\psi,M,N} = 0$ *for each* $M \geqslant 1$;
- $\mathcal{A}_{\mathsf{H}}$ *is convex and* $\overline{F}_N(t,\cdot) \in \mathcal{A}_{\mathsf{H}}$ *for all* $t \geqslant 0$ *and* $N \in \mathbb{N}$.

Then there is a unique weak solution f to (3.2.5) with $f(0,\cdot) = \psi$, and there is a constant $C > 0$ such that the following holds for all $M \geqslant 1$ and all $N \in \mathbb{N}$:

$$\sup_{t\in[0,M]}\int_{\mathbb{S}^K_{+,M}} \big|\overline{F}_N(t,h) - f(t,h)\big|\mathrm{d}h \leqslant CM^\alpha\Big(\mathcal{L}_{\psi,CM,N} + N^{-\frac{1}{14}} + \big(\mathcal{K}_{CM,N}/M^\beta\big)^{\frac{2}{7}}\Big), \tag{3.2.12}$$

where $\alpha = \frac{K(K+1)}{2} + \frac{\beta \vee 1}{2} + 1$.

Theorem 3.2.2. *Let* $p \in \mathbb{N}$. *Suppose that there is* $\psi : \mathbb{S}^K_+ \to \mathbb{R}$ *such that* $\overline{F}_N(0, \cdot)$ *converges to* ψ *pointwise, and that for each* $M > 0$ *we have*

$$\lim_{N\to\infty} \mathcal{K}_{M,N} = 0. \tag{3.2.13}$$

Then, for any H *of the form* (3.2.6)*, there is a unique Lipschitz viscosity solution* f *to* (3.2.5) *with* $f(0, \cdot) = \psi$*, and*

$$\limsup_{N\to\infty} \overline{F}_N(t, h) \leqslant f(t, h), \quad \forall (t, h) \in \mathbb{R}_+ \times \mathbb{S}^K_+.$$

If, in addition, H *is convex, then a corresponding lower bound holds and thus*

$$\lim_{N\to\infty} \overline{F}_N(t, h) = f(t, h), \quad \forall (t, h) \in \mathbb{R}_+ \times \mathbb{S}^K_+.$$

The proofs of Theorem 3.2.1 and Theorem 3.2.2 are in Section 3.5 and Section 3.6, respectively.

Remark 3.2.3 (Conditions on $\mathcal{A}_{\mathsf{H}}$). When ϕ is smooth, we can compute that $\nabla \cdot \big(\mathcal{D}\mathsf{H}(\nabla\phi)\big) = \mathcal{D}^2\mathsf{H}(\nabla\phi) \cdot \nabla^2\phi$ where $\mathcal{D}^2\mathsf{H}$ is the Hessian of H. Lemma 3.4.5 will show that if H is convex, then the conditions on $\mathcal{A}_{\mathsf{H}}$ in Theorem 3.2.1, namely, the convexity of $\mathcal{A}_{\mathsf{H}}$ and $\overline{F}_N(t, \cdot) \in \mathcal{A}_{\mathsf{H}}$, are satisfied.

Note that when $p \leqslant 2$, $\mathcal{D}^2\mathsf{H}$ is constant and in this case $\mathcal{A}_{\mathsf{H}}$ is always convex. Hence, the only condition to check is that $\overline{F}_N(t, \cdot) \in \mathcal{A}_{\mathsf{H}}$. In Appendix 3.8, we demonstrate a special model of (3.1.1) with $p = 2$ where this condition is satisfied but H is not convex. This model is equivalent to the nonsymmetric matrix inference problem considered in [90, 14, 12, 79, 86, 36].

It seems that the conditions on $\mathcal{A}_{\mathsf{H}}$ are not satisfied by the model (3.2.16) for odd $p \geqslant 1$.

The explicit expression of $\mathcal{D}^2\mathsf{H}$ in this model is computed in (3.4.8). We believe that this issue is closely related to a similar difficulty in the adaptive interpolation approach to the same model with odd p, which is discussed in [85, Section 7].

Remark 3.2.4 (Local uniform convergence). The local $L^\infty_t L^1_x$ convergence in Theorem 3.2.1 can be upgraded to local $L^\infty_t L^\infty_x$. Let ξ be a smooth function supported on $-\mathbb{S}^K_{+,1}$, and satisfy $0 \leqslant \xi \leqslant 1$ and $\int \xi > 0$. For $\varepsilon \in (0,1)$, let $\xi_\varepsilon(x) = \varepsilon^{-K(K+1)/2}\xi(\varepsilon^{-1}x)$. Then, for every Lipschitz $g : \mathbb{S}^K_+ \to \mathbb{R}$, we have

$$\begin{aligned}\|g\|_{L^\infty(\mathbb{S}^K_{+,M})} &\leqslant \|g * \xi_\varepsilon\|_{L^\infty(\mathbb{S}^K_{+,M})} + \|g - g * \xi_\varepsilon\|_{L^\infty(\mathbb{S}^K_{+,M})} \\ &\leqslant C\varepsilon^{-K(K+1)/2}\|g\|_{L^1(\mathbb{S}^K_{+,M+1})} + C\varepsilon\|g\|_{\mathrm{Lip}}.\end{aligned}$$

By (3.3.8), we know $\overline{F}_N(t,\cdot)$ is Lipschitz uniformly in N and t, and thus $f(t,\cdot)$ is also Lipschitz. Replace g in the above by $\overline{F}_N(t,\cdot) - f(t,\cdot)$, apply Theorem 3.2.1 and optimize the above display over ε to see convergence in local $L^\infty_t L^\infty_x$.

Remark 3.2.5 (Variational formulae). Under the assumptions on ψ in the two theorems, we can show that ψ is Lipschitz, convex and nondecreasing in the sense that $\nabla\psi \in \mathbb{S}^K_+$. By the pointwise convergence $\overline{F}_N(0,\cdot) \to \psi$ and (3.3.8), (3.3.10), (3.3.12), and the pointwise convergence $\overline{F}_N(0,\cdot) \to \psi$, we can see that ψ is Lipschitz in the two theorems above. Proposition 3.6.6 will show that f in Theorem 3.2.2 can be represented by the following variational formula

$$f(t,x) = \sup_{z\in\mathbb{S}^K_+} \inf_{y\in\mathbb{S}^K_+} \left\{z \cdot (x-y) + \psi(y) + t\mathsf{H}(z)\right\}, \quad \forall (t,x) \in \mathbb{R}_+ \times \mathbb{S}^K_+. \tag{3.2.14}$$

When H is convex, comparing Theorem 3.2.1 with Theorem 3.2.2 in view of Remark 3.2.4, we can see that the unique weak solution f coincides with the viscosity solution pointwise, and thus also admits the representation (3.2.14). For general H, we believe weak solutions are still of the form (3.2.14). The relatively difficult part is to verify that (3.2.14) satisfies

(2) of Definition 3.4.1.

Remark 3.2.6 (Possibility for weaker assumptions on H). Let us point out key inequalities, where the assumptions on H are used. If these inequalities still hold in certain models, then our results should still be valid there.

The conditions on $\mathcal{A}_{\mathsf{H}}$ in Theorem 3.2.1 are used to obtain the inequality (3.4.4) in the proof of Lemma 3.4.3, which is further used to prove the uniqueness of weak solutions (Proposition 3.4.2), and the convergence to the unique weak solution (Proposition 3.5.1 and Proposition 3.5.2). In fact, uniqueness and convergence are still valid if the right-hand side of (3.4.4) is replaced by a negative constant depending locally on the temporal and spacial variables. However, the convergence rate can be much worse (logarithmic in N), because the absolute value of this constant will appear in the exponential factor of Gronwall's lemma.

The convexity assumption in the second assertion of Theorem 3.2.2 is only used to apply Jensen's inequality to derive (3.7.28) in the proof of that the limit of $\overline{F}_N$ is a viscosity supersolution.

3.2.3. Special case

We apply Theorem 3.2.1 and Theorem 3.2.2 to an i.i.d. case. Let $\mathcal{P}$ be a probability distribution in $\mathbb{R}^K$ supported on $\{z \in \mathbb{R}^K : |z| \leqslant \sqrt{K}\}$. For each $N \in \mathbb{N}$, let the row vectors of X, namely $X_{1,\cdot}\,, X_{2,\cdot}\,, \ldots,\ X_{N,\cdot}$, be i.i.d. with law $\mathcal{P}$. Set $L = 1$ and consider $A \in \mathbb{R}^{K^p \times 1}$ given by

$$A_{\mathbf{j}} = \begin{cases} 1, & \text{if } j_1 = j_2 = \cdots = j_p, \\ 0, & \text{otherwise.} \end{cases}$$

Here, we used the multi-index notation

$$\mathbf{j} = (j_1, j_2, \ldots, j_p) \in \{1, \cdots, K\}^p. \tag{3.2.15}$$

Explicitly, (3.1.1) now becomes

$$Y_{\mathbf{i}} = \sqrt{\frac{2t}{N^{p-1}}} \sum_{j=1}^{K} \prod_{n=1}^{p} X_{i_n,j} + W_{\mathbf{i}}, \quad \mathbf{i} \in \{1, \cdots, N\}^p, \tag{3.2.16}$$

and (3.2.6) becomes

$$\mathsf{H}(q) = \sum_{j,j'=1}^{K} \big(q_{j,j'}\big)^p, \quad q \in \mathbb{S}_+^K. \tag{3.2.17}$$

Using (3.2.4) and the fact that rows of X are i.i.d., we can see $\overline{F}_N(0,\cdot) = \overline{F}_1(0,\cdot)$, for all $N \in \mathbb{N}$. Setting $\psi = \overline{F}_1(0,\cdot)$, we clearly have $\mathcal{L}_{\psi,M,N} = 0$ for all M and N. Estimate on $\mathcal{K}_{M,N}$ is given in Lemma 3.10.1. When $p = 1$ or p is even, Lemma 3.4.5 shows that the assumptions on $\mathcal{A}_{\mathsf{H}}$ in Theorem 3.2.1 are satisfied. Applying the main results, we have the following corollary.

Corollary 3.2.7. *In the special case described above, let f be given by* (3.2.14) *with $\psi = \overline{F}_1(0,\cdot)$. Then for all $p \in \mathbb{N}$, we have*

$$\limsup_{N\to\infty} \overline{F}_N(t,h) \leqslant f(t,h), \quad \forall (t,h) \in \mathbb{R}_+ \times \mathbb{S}_+^K.$$

If p is even or $p = 1$, then there is $C > 0$ such that, for all $M \geqslant 1$ and $N \in \mathbb{N}$,

$$\sup_{t\in[0,M]} \int_{\mathbb{S}_{+,M}^K} \big|\overline{F}_N(t,h) - f(t,h)\big| \mathrm{d}h \leqslant CM^{\frac{K(K+1)+3}{2}} N^{-\frac{1}{14}}.$$

This model (3.2.16) has also been investigated in [85] and similar convergence results for even orders were established. Although the convergence for odd orders remains open, we are able to obtain an upper bound for the limit of free energy.

3.3. Approximate Hamilton–Jacobi equations

The goal of this section is to show that $\overline{F}_N$ satisfies an approximate Hamilton–Jacobi equation, as summarized in Proposition 3.3.1 below. There is a considerable overlap between results in this section and [94, Section 3], which follows the approach of [9]. To simplify our presentation, whenever similar arguments are available in [94, Section 3], we shall only demonstrate key steps and refer to [94, Section 3] for more detailed computations.

Proposition 3.3.1 (Approximate Hamilton–Jacobi equations)**.** *There exists $C > 0$ such that for every $N \geqslant 1$ and uniformly over $\mathbb{R}_+ \times \mathbb{S}^K_+$,*

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

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

$$\kappa(h) := \begin{cases} |h||h^{-1}|, & \textit{if } h \in \mathbb{S}^K_{++}, \\ +\infty & \textit{otherwise.} \end{cases} \tag{3.3.1}$$

3.3.1. Proof of Proposition 3.3.1

We start by proving the following identity

$$\partial_t \overline{F}_N - \mathsf{H}\big(\nabla \overline{F}_N\big) = \frac{1}{N^p}\Big(\mathbb{E}\left\langle \mathsf{H}(x^\intercal x')\right\rangle - \mathsf{H}\big(\mathbb{E}\left\langle x^\intercal x'\right\rangle\big)\Big). \tag{3.3.2}$$

Proof of (3.3.2)*.* Let us first compute $\partial_t \overline{F}_N$ and $\nabla \overline{F}_N$. Indeed, from (3.2.4), we can compute

$$\partial_t F_N(t,h) = \frac{1}{N}\left\langle \frac{2}{N^{p-1}}\widetilde{x}\cdot\widetilde{X} + \sqrt{\frac{1}{2N^{p-1}t}}\widetilde{x}\cdot W - \frac{|\widetilde{x}|^2}{N^{p-1}}\right\rangle, \tag{3.3.3}$$

and, for $a \in \mathbb{S}^K$,

$$a \cdot \nabla F_N(t,h) = \frac{1}{N}\left\langle 2a\cdot(x^\intercal X) + \sqrt{2}D_{\sqrt{h}}(a)\cdot(x^\intercal Z) - a\cdot(x^\intercal x)\right\rangle. \tag{3.3.4}$$

Here $D_{\sqrt{h}}$ is the differential of the square-root function at $h \in \mathbb{S}^K_{++}$. More precisely, for $h \in \mathbb{S}^K_{++}$ and $a \in \mathbb{S}_K$, we have

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

Using the Gaussian integration by parts (c.f. [94, Lemma 3.3]) and the Nishimori identity (c.f. [94, Section 3.1]), we can get from (3.3.3) that

$$\partial_t \overline{F}_N = \frac{1}{N^p} \mathbb{E} \left\langle \widetilde{x} \cdot \widetilde{x}' \right\rangle . \tag{3.3.5}$$

Here x' is an independent copy (or replica) of x with respect to the Gibbs measure $\langle \cdot \rangle$.

To compute $\nabla \overline{F}_N$, we refer to the derivation of [94, (3.17)]. The object $\overline{x}$ therein is X in our notation, and our $F_N(t,h)$ corresponds to $F_N(t,2h)$ there. Hence [94, (3.17)] is equivalent to $\nabla \overline{F}_N = \frac{1}{N} \mathbb{E} \langle x^\intercal X \rangle$. A further application of the Nishimori identity yields

$$\nabla \overline{F}_N = \frac{1}{N} \mathbb{E} \left\langle x^\intercal x' \right\rangle . \tag{3.3.6}$$

By (3.2.2) and (3.2.6), we have $\widetilde{x} \cdot \widetilde{x}' = \mathsf{H}(x^\intercal x')$. This along with (3.3.5), (3.3.6) and (3.2.6) implies (3.3.2). $\square$

Now, to prove Proposition 3.3.1, we only need to estimate the right hand side of (3.3.2). Using (3.2.6) and (3.2.1), we get

$$\begin{aligned} \left| \mathbb{E} \left\langle \mathsf{H}(x^\intercal x') \right\rangle - \mathsf{H}\big(\mathbb{E} \left\langle x^\intercal x' \right\rangle\big) \right| &\leqslant C \mathbb{E} \left\langle \left| \left(x^\intercal x'\right)^{\otimes p} - \left(\mathbb{E} \left\langle x^\intercal x' \right\rangle\right)^{\otimes p} \right| \right\rangle \\ &\leqslant C N^{p-1} \mathbb{E} \left\langle \left| x^\intercal x' - \mathbb{E} \left\langle x^\intercal x' \right\rangle \right| \right\rangle , \end{aligned}$$

Jensen's inequality gives

$$\left|\mathbb{E}\left\langle \mathsf{H}(x^\intercal x')\right\rangle - \mathsf{H}\big(\mathbb{E}\left\langle x^\intercal x'\right\rangle\big)\right|^2 \leqslant CN^{2p-2}\mathbb{E}\left\langle \left|x^\intercal x' - \mathbb{E}\left\langle x^\intercal x'\right\rangle\right|^2\right\rangle.$$

We need the following estimate

$$\frac{1}{N^2}\mathbb{E}\left\langle \left|x^\intercal x' - \mathbb{E}\left\langle x^\intercal x'\right\rangle\right|^2\right\rangle \leqslant C\kappa(h)N^{-\frac{1}{4}}\big(\Delta\overline{F}_N + |h^{-1}|\big)^{\frac{1}{4}} + C\mathbb{E}\big|\nabla F_N - \nabla\overline{F}_N\big|^2.$$

This is exactly [94, (3.18)], and we shall omit the derivation here. The above two displays and (3.3.2) gives the desired result.

3.3.2. Estimates of derivatives

We finish this section by collecting useful results in Lemma 3.3.2 and (7.5.2). Recall $A \in \mathbb{R}^{K^p\times L}$ and $W \in \mathbb{R}^{N^p\times L}$. We define

$$\|WA^\intercal\| = \sup_{y_1,\, y_2,\, \ldots,\, y_p\in\mathbb{S}^{NK-1}}\left\{(WA^\intercal)\cdot(y_1\otimes y_2\otimes\cdots\otimes y_p)\right\} \tag{3.3.7}$$

where $\mathbb{S}^{NK-1}$ denotes the unit sphere in $\mathbb{R}^{NK}$.

Lemma 3.3.2. *There exists a constant $C > 0$ such that the following estimates hold uniformly over $\mathbb{R}_+ \times \mathbb{S}^K_+$ for every $N \in \mathbb{N}$:*

$$|\partial_t\overline{F}_N| + |\nabla\overline{F}_N| \leqslant C, \tag{3.3.8}$$

$$|\partial_t F_N| \leqslant C\left(1 + \frac{\|WA^\intercal\|}{\sqrt{Nt}}\right), \quad \textit{and} \quad |\nabla F_N| \leqslant C\left(1 + \frac{|Z||h^{-1}|^{\frac{1}{2}}}{\sqrt{N}}\right). \tag{3.3.9}$$

Everywhere in $\mathbb{R}_+ \times \mathbb{S}^K_+$, we have

$$\partial_t\overline{F}_N \geqslant 0, \qquad \nabla\overline{F}_N \in \mathbb{S}^K_+, \tag{3.3.10}$$

$$\partial_t^2\overline{F}_N \geqslant 0. \tag{3.3.11}$$

Moreover, for every $a \in \mathbb{S}^K$, we have

$$a \cdot \nabla(a \cdot \nabla \overline{F}_N) \geqslant 0, \tag{3.3.12}$$

$$a \cdot \nabla(a \cdot \nabla F_N) \geqslant -\frac{C|a|^2|Z||h^{-1}|^{\frac{3}{2}}}{\sqrt{N}}. \tag{3.3.13}$$

Proof of (3.3.8). It follows easily from (3.2.1), (3.3.5) and (3.3.6). □

Proof of (3.3.9). In view of (3.3.7), we have

$$\left|\left(x^{\otimes p} A\right) \cdot W\right| = \left|\left(W A^{\intercal}\right) \cdot \left(x^{\otimes p}\right)\right| \leqslant \|W A^{\intercal}\||x|^p.$$

In addition, it can be seen from (3.2.2) that $|\widetilde{x}| \leqslant C|x|^p$. Using these, (3.3.3) and (3.2.1), we have

$$\begin{aligned}\left|\partial_t F_N(t,h)\right| &\leqslant \left\langle \frac{2}{N^p}|\widetilde{x}||\widetilde{X}| + \sqrt{\frac{1}{2tN^{p+1}}}\left|\left(x^{\otimes p} A\right) \cdot W\right| + \frac{1}{N^p}|\widetilde{x}|^2 \right\rangle \\ &\leqslant C + \frac{C\|W A^{\intercal}\|}{\sqrt{Nt}} + C = C\left(1 + \frac{\|W A^{\intercal}\|}{\sqrt{Nt}}\right).\end{aligned}$$

For the second estimate in (3.3.9), we need the following estimate

$$|D_{\sqrt{h}}(a)| \leqslant C|a||h^{-1}|^{\frac{1}{2}}. \tag{3.3.14}$$

Its proof can be seen from the derivation of [94, (3.7)]. Insert $a = \frac{\nabla F_N}{|\nabla F_N|} \in \mathbb{S}^K$ into (3.3.4) and then use (3.3.14) to see

$$|\nabla F_N| \leqslant \left\langle \frac{2}{N}\left|x^{\intercal} X\right| + \frac{C|h^{-1}|^{\frac{1}{2}}}{N}\left|x^{\intercal} Z\right| + \frac{1}{N}|x^{\intercal} x| \right\rangle \leqslant C\left(1 + \frac{|Z||h^{-1}|^{\frac{1}{2}}}{\sqrt{N}}\right).$$

□

Proof of (3.3.11). Recall (3.3.5). Using (3.2.3), we differentiate $\partial_t \overline{F}_N$ one more time in t to see

$$N^p\partial_t^2\overline{F}_N = \mathbb{E}\Big\langle (\widetilde{x}\cdot\widetilde{x}')\Big(\frac{2}{N^{p-1}}\big(\widetilde{x}+\widetilde{x}'-2\widetilde{x}''\big)\cdot\widetilde{X} - \frac{1}{N^{p-1}}\big(|\widetilde{x}|^2+|\widetilde{x}'|^2-2|\widetilde{x}''|^2\big)$$
$$+\frac{1}{\sqrt{2N^{p-1}t}}\big(\widetilde{x}+\widetilde{x}'-2\widetilde{x}''\big)\cdot W\Big)\Big\rangle.$$

Using the symmetry between replicas, the Nishimori identity and the Gaussian integration by parts, we can compute

$$\begin{aligned} N^{2p-1}\partial_t^2\overline{F}_N &= 2\mathbb{E}\left\langle (\widetilde{x}\cdot\widetilde{x}')\big(\widetilde{x}\cdot\widetilde{x}'-2\widetilde{x}\cdot\widetilde{x}''+\widetilde{x}''\cdot\widetilde{x}'''\big)\right\rangle \\ &= 2\mathbb{E}\sum_{\mathbf{i},\mathbf{j},k,l}\Big(\left\langle \widetilde{x}_{\mathbf{i},k}\widetilde{x}_{\mathbf{j},l}\right\rangle^2 - 2\left\langle \widetilde{x}_{\mathbf{i},k}\widetilde{x}_{\mathbf{j},l}\right\rangle\left\langle \widetilde{x}_{\mathbf{i},k}\right\rangle\left\langle \widetilde{x}_{\mathbf{j},l}\right\rangle + \left\langle \widetilde{x}_{\mathbf{i},k}\right\rangle^2\left\langle \widetilde{x}_{\mathbf{j},l}\right\rangle^2\Big) \geqslant 0. \end{aligned}$$

This gives (3.3.11). □

Proof of (3.3.10). By the independence of the replica x' from x, we can rewrite (3.3.5) as $\partial_t\overline{F}_N = \frac{1}{N^p}\mathbb{E}(\langle\widetilde{x}\rangle\cdot\langle\widetilde{x}\rangle)$ and rewrite (3.3.6) as $\nabla\overline{F}_N = \frac{1}{N}\mathbb{E}\left\langle x\right\rangle^\intercal\left\langle x\right\rangle$. Then, (3.3.10) clearly follows. □

Proof of (3.3.12). For $a\in\mathbb{S}^K$, we can compute

$$Na\cdot\nabla(a\cdot\nabla\overline{F}_N) = \mathbb{E}\left\langle \big(a\cdot x^\intercal x'\big)^2\right\rangle - 2\mathbb{E}\left\langle \big(a\cdot x^\intercal x'\big)\big(a\cdot x^\intercal x''\big)\right\rangle + \mathbb{E}\left\langle a\cdot x^\intercal x'\right\rangle^2.$$

The details of this computation can be seen from the derivation of [94, (3.27)]. Expand the right hand side of the above display to get

$$\mathbb{E}\sum_{i,j,k,m,n,l} a_{ij}a_{mn}\left\langle x_{ki}x'_{kj}x_{lm}x'_{ln} - 2x_{ki}x'_{kj}x_{lm}x''_{ln} + x_{ki}x'_{kj}x''_{lm}x'''_{ln}\right\rangle$$

where x', x'', x''' are replicas of x with respect to the measure $\langle\cdot\rangle$. Then, (3.3.12) follows if we can show the above is nonnegative. Use the independence and write $\widehat{x} = x - \langle x\rangle$ to see that the above display is equal to

$$
\begin{aligned}
&\mathbb{E} \sum_{i,j,k,m,n,l} a_{ij} a_{mn} \Big(\langle x_{ki} x_{lm} \rangle \langle x_{kj} x_{ln} \rangle - 2 \langle x_{ki} x_{lm} \rangle \langle x_{kj} \rangle \langle x_{ln} \rangle + \langle x_{ki} \rangle \langle x_{lm} \rangle \langle x_{kj} \rangle \langle x_{ln} \rangle \Big) \\
&= \mathbb{E} \sum_{i,j,k,m,n,l} a_{ij} a_{mn} \Big(\langle x_{ki} x_{lm} \rangle \langle \widehat{x}_{kj} \widehat{x}_{ln} \rangle - \langle \widehat{x}_{ki} \widehat{x}_{lm} \rangle \langle x_{kj} \rangle \langle x_{ln} \rangle \Big).
\end{aligned}
$$

Notice that since $a \in \mathbb{S}^K$, we can replace i and m by j and n, respectively, in the second term inside the last pair of parentheses. So the above becomes

$$
\begin{aligned}
&\mathbb{E} \sum_{i,j,k,m,n,l} a_{ij} a_{mn} \Big(\langle x_{ki} x_{lm} \rangle \langle \widehat{x}_{kj} \widehat{x}_{ln} \rangle - \langle \widehat{x}_{ki} \widehat{x}_{lm} \rangle \langle x_{ki} \rangle \langle x_{lm} \rangle \Big) \\
&= \mathbb{E} \sum_{i,j,k,m,n,l} a_{ij} a_{mn} \langle \widehat{x}_{ki} \widehat{x}_{lm} \rangle \langle \widehat{x}_{kj} \widehat{x}_{ln} \rangle = \mathbb{E} \sum_{i,j,k,m,n,l} a_{ij} a_{mn} \big\langle (\widehat{x}^\intercal \widehat{x}')_{ij} (\widehat{x}^\intercal \widehat{x}')_{mn} \big\rangle \\
&= \mathbb{E} \big\langle (a \cdot \widehat{x}^\intercal \widehat{x}')^2 \big\rangle \geqslant 0.
\end{aligned}
$$

□

Proof of (3.3.13). By (3.3.4), we can compute

$$
\begin{aligned}
&a \cdot \nabla (a \cdot \nabla F_N(t,h)) \\
&= \frac{1}{N} \Big(\Big\langle \big(H'_N(a,h,x) \big)^2 \Big\rangle - \big\langle H'_N(a,h,x) \big\rangle^2 \Big) + \frac{1}{N} \Big\langle \sqrt{2} D^2_{\sqrt{h}}(a,a) \cdot x^\intercal Z \Big\rangle, \qquad (3.3.15)
\end{aligned}
$$

where

$$
H'_N(a,h,x) = \sqrt{2} D_{\sqrt{h}}(a) \cdot x^\intercal Z + 2a \cdot x^\intercal X - a \cdot x^\intercal x,
$$

and, for every $h \in \mathbb{S}^K_{++}$ and $a, b \in \mathbb{S}^K$,

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

Recognizing a variance term in (3.3.15) and using (3.2.1), we have

$$a \cdot \nabla(a \cdot \nabla F_N(t,h)) \geqslant -C\left|D^2_{\sqrt{h}}(a,a)\right|\frac{|Z|}{\sqrt{N}}.$$

The display [94, (3.38)] states

$$\left|D^2_{\sqrt{h}}(a,a)\right| \leqslant C|a|^2|h^{-1}|^{\frac{3}{2}}.$$

Combining this with the previous display, we obtain (3.3.13). □

Lastly, we state an elementary lemma characterizing $\mathbb{S}^K_+$.

Lemma 3.3.3. *Let $a \in \mathbb{S}^K$, Then, $a \in \mathbb{S}^K_+$ if and only if $a \cdot b \geqslant 0$ for every $b \in \mathbb{S}^K_+$.*

Proof. If $a \in \mathbb{S}^K_+$, then for any $b \in \mathbb{S}^K_+$ we have $a \cdot b = \mathrm{tr}(\sqrt{a}\sqrt{b}\sqrt{b}\sqrt{a}) \geqslant 0$. For the other direction, by choosing an orthonormal basis, we may assume a is diagonal. Testing by $b \in \mathbb{S}^K_+$, we can show that all diagonal entries in a are nonnegative and thus $a \in \mathbb{S}^K_+$. □

3.4. Weak solutions of Hamilton–Jacobi equations

In this section, we study the Hamilton–Jacobi equation (3.2.5) through the perspective of weak solutions. Precise definitions of weak solutions will be stated and uniqueness of solutions is given in Proposition 3.4.2.

We identify $\mathbb{S}^K$ isometrically with $\mathbb{R}^{K(K+1)/2}$ via the orthonormal basis $\{e^{ij}\}_{1\leqslant i\leqslant j\leqslant K}$ given by, for $m,n \in \{1,2,\dots,K\}$,

$$(e^{ij})_{mn} = \left(\mathbb{1}_{i=j} + \frac{\sqrt{2}}{2}\mathbb{1}_{i\neq j}\right)\mathbb{1}_{\{m,n\}=\{i,j\}}. \tag{3.4.1}$$

Here $\mathbb{1}$ stands for the indicator function. Naturally, we endow $\mathbb{S}^K$ with the Lebesgue measure on $\mathbb{R}^{K(K+1)/2}$. Recall the definition of $\mathcal{A}_{\mathsf{H}}$ in (3.2.11).

Definition 3.4.1. A function $f : \mathbb{R}_+ \times \mathbb{S}^K_+ \to \mathbb{R}$ is a weak solution to (3.2.5) if

1. f is Lipschitz and satisfies (3.2.5) almost everywhere;

2. $f(t,\cdot) \in \mathcal{A}_{\mathsf{H}}$ for all $t \geqslant 0$.

Proposition 3.4.2 (Uniqueness)**.** *Under the assumption that $\mathcal{A}_{\mathsf{H}}$ is convex, there is at most one weak solution to* (3.2.5)*.*

3.4.1. Proof of Proposition 3.4.2

The idea of proof is classical and can be seen in [56, 80, 81]. See also [25] and [60, Section 3.3.3]. The following lemma will also be used later. Recall the definitions of $\mathbb{S}^K_{+,M}$ in (3.2.7).

Lemma 3.4.3. *Assume that $\mathcal{A}_{\mathsf{H}}$ is convex. For $M > 0, T \geqslant 1, \eta \in (0,1)$, define*

$$D_t = \mathbb{S}^K_{+,R(T-t)} \cap (\eta I + \mathbb{S}^K_+), \quad \forall t \in [0,T] \tag{3.4.2}$$

with $R = \sup\left\{|\mathcal{D}\mathsf{H}(p)| : p \in \mathbb{S}^K_{+,M}\right\}$. Let $\phi : \mathbb{R} \to \mathbb{R}_+$ be any smooth function. Then, the following holds for all choices of M, T, η, ϕ, and for every pair $f, g \in \mathcal{A}_{\mathsf{H}}$ satisfying $\|f\|_{\mathrm{Lip}}, \|g\|_{\mathrm{Lip}} \leqslant M$:

$$\frac{\mathrm{d}}{\mathrm{d}t} J(t) \leqslant \int_{D_t} \Big(\phi'(f-g)|r|\Big)(t,h)\mathrm{d}h, \quad \forall t \in [0,T]$$

where

$$J(t) = \int_{D_t} \phi(f-g)(t,h)\mathrm{d}h,$$
$$r = \Big(\partial_t f - \mathsf{H}(\nabla f)\Big) - \Big(\partial_t g - \mathsf{H}(\nabla g)\Big).$$

Proof. Let us set $w = f - g$ and $v = \phi(w)$. We proceed in steps.

Step 1. We study the relations which w and v satisfy. Since f and g are weak solutions, we have

$$\partial_t w = \mathsf{H}(\nabla f) - \mathsf{H}(\nabla g) + r = b \cdot \nabla w + r$$

where the function b is given by

$$b = \int_0^1 \mathcal{D}\mathsf{H}\big(s\nabla f + (1-s)\nabla g\big)\mathrm{d}s.$$

Here $\mathcal{D}\mathsf{H}$ is the gradient of H while ∇ is taking derivatives in the spacial variable h. Then, we also have

$$\partial_t v = b \cdot \nabla v + \phi'(w) r. \tag{3.4.3}$$

Step 2. We introduce a family of mollifiers. Let $\xi : \mathbb{R}^{K(K+1)/2} \to \mathbb{R}_+$ be smooth, be supported on $-\mathbb{S}^K_{+,1}$, and satisfy $\int \xi = 1$. For $\varepsilon \in (0,1)$, set

$$\xi_\varepsilon = \varepsilon^{-K(K+1)/2}\xi\Big(\frac{\cdot}{\varepsilon}\Big).$$

Define b_ε by the convolution

$$b_\varepsilon(t,h) = \big(b(t,\cdot) * \xi_\varepsilon\big)(h) = \int b(t,h-h')\xi_\varepsilon(h')\mathrm{d}h'.$$

Recall the definition of $\mathcal{A}_\mathsf{H}$ in (3.2.11). Since $\mathcal{A}_\mathsf{H}$ is assumed to be convex and $f, g \in \mathcal{A}_\mathsf{H}$ are weak solutions, by the definition of b, we must have $\nabla \cdot b \geqslant 0$ in the distribution sense. Then, it is easy to see that

$$\nabla \cdot b_\varepsilon \geqslant 0 \tag{3.4.4}$$

holds pointwise everywhere. We finish this step by proving

$$b_\varepsilon \in \mathbb{S}^K_+. \tag{3.4.5}$$

This follows from the next lemma, which will also be used later.

Lemma 3.4.4. *For* H *given in* (3.2.6)*, its differential* $\mathcal{D}\mathsf{H} \in \mathbb{S}^K_+$ *everywhere.*

Proof. For simplicity, we write $S = AA^\intercal \in \mathbb{S}^{K^p}_+$. Let $a, q \in \mathbb{S}^K$, then we can compute that

$$a \cdot \mathcal{D}\mathsf{H}(q) = pS \cdot \mathsf{sym}(a \otimes q^{\otimes p-1}).$$

Here sym denotes the symmetrization of tensors given by

$$\mathsf{sym}\Big(b_1 \otimes b_2 \otimes \cdots \otimes b_p\Big) = \frac{1}{p!}\sum_\sigma b_{\sigma(1)} \otimes b_{\sigma(2)} \otimes \cdots \otimes b_{\sigma(p)},$$

where the summation is taken over all permutations. Since $S \in \mathbb{S}^{K^p}_+$, to show $a \cdot \mathcal{D}\mathsf{H}(q) \geqslant 0$ it suffices to show $a \otimes q^{\otimes p-1} \in \mathbb{S}^{K^p}_+$. We only need to check

$$u^\intercal\big(a \otimes q^{\otimes p-1}\big)u \geqslant 0, \quad \forall u \in \mathbb{R}^{K^p}.$$

Index $u \in \mathbb{R}^{K^p}$ as $(u_{\mathbf{i}})_{\mathbf{i}}$ with $\mathbf{i}$ in the form of (3.2.15). Writing $\widehat{\mathbf{i}} = (i_2, i_3, \ldots, i_p)$, let us compute

$$\begin{aligned} u^\intercal\big(a \otimes q^{\otimes p-1}\big)u &= \sum_{\mathbf{i},\mathbf{j}} u_{\mathbf{i}}\big(a \otimes q^{\otimes p-1}\big)_{\mathbf{i},\mathbf{j}} u_{\mathbf{j}} = \sum_{\mathbf{i},\mathbf{j}} u_{i_1,\widehat{\mathbf{i}}}\, a_{i_1,j_1} (q^{\otimes p-1})_{\widehat{\mathbf{i}},\widehat{\mathbf{j}}}\, u_{j_1,\widehat{\mathbf{j}}} \\ &= \operatorname{tr}\big(u^\intercal a u q^{\otimes p-1}\big) = \operatorname{tr}\big(\sqrt{a} u q^{\otimes p-1} u^\intercal \sqrt{a}\big) \geqslant 0. \end{aligned}$$

Here, we used the fact that $q^{\otimes p-1}$ is positive semi-definite, which can be proved by iterating the above arguments. Therefore, we can conclude that $a \cdot \mathcal{D}\mathsf{H} \geqslant 0$ for every $a \in \mathbb{S}^K_+$, which by Lemma 3.3.3 implies $\mathcal{D}\mathsf{H} \in \mathbb{S}^K_+$. $\square$

Step 3. We study $J(t)$ which can be written as $J(t) = \int_{D_t} v(t,\cdot)$. On $\mathbb{R}_+ \times \mathbb{S}^K_+$, the equation (3.4.3) can be expressed as

$$\partial_t v = \mathsf{div}(vb_\varepsilon) - v\nabla\cdot b_\varepsilon + (b - b_\varepsilon)\cdot\nabla v + \phi'(w)r. \tag{3.4.6}$$

In addition to D_t, we set

$$\Gamma_t = \partial D_t \cap \{|x| = R(T-t)\}.$$

Using (3.4.6) and integration by parts, we can compute

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}J(t) &= \int_{D_t}\partial_t v - R\int_{\Gamma_t} v \\ &= \int_{\Gamma_t}(\mathbf{n}\cdot b_\varepsilon - R)v + \int_{\partial D_t\setminus\Gamma_t}(\mathbf{n}\cdot b_\varepsilon)v + \int_{D_t} v(-\nabla\cdot b_\varepsilon) + \int_{D_t}(b-b_\varepsilon)\cdot\nabla v \\ &\qquad + \int_{D_t}\phi'(w)r,\end{aligned} \tag{3.4.7}$$

where $\mathbf{n}$ stands for the outer normal vector, and the integrations are only carried out in the spacial variable. We treat the integrals in (3.4.7) individually. By the definitions of b_ε and ξ_ε, we can see $|b_\varepsilon| \leqslant R$. Hence, the first integral is nonpositive. Due to (3.4.5) and the fact that $-\mathbf{n} \in \mathbb{S}^K_+$ on $\partial D_t \setminus \Gamma_t$, the second integral is also nonpositive. In view of (3.4.4), the third integral is again nonpositive, while the last one is $o_\varepsilon(1)$. Therefore, taking $\varepsilon \to 0$, we conclude that $\frac{\mathrm{d}}{\mathrm{d}t}J(t) \leqslant \int_{D_t}\phi'(w)|v|$ as desired.

□

Proof of Proposition 3.4.2. Let f and g be two weak solutions to (3.2.5) with $f(0,\cdot) = g(0,\cdot)$. Let $M = \|f\|_{\mathrm{Lip}} \vee \|g\|_{\mathrm{Lip}}$. For each $\delta > 0$, we have $\|f(\delta,\cdot) - g(\delta,\cdot)\|_\infty \leqslant M\delta$. Let $\phi : \mathbb{R} \to \mathbb{R}_+$

be a smooth function and satisfy

$$\begin{cases} \phi(z) = 0, & \text{if } |z| \leqslant M\delta, \\ \phi(z) > 0, & \text{otherwise.} \end{cases}$$

Applying Lemma 3.4.3 to f, g, M, ϕ described above, and any choice of T, η, we have $J(t) \leqslant J(\delta)$ for $t \in [\delta, T]$. But our choice of ϕ implies that

$$J(\delta) = \int_{D_\delta} \phi(f - g)(\delta, h)\mathrm{d}h = 0.$$

Since $J(t)$ is nonnegative, we must have $J(t) = 0$ for all $t \in [\delta, T]$. This together with the definition of ϕ guarantees that

$$|f(t, h) - g(t, h)| \leqslant M\delta, \quad \forall h \in D_t, \ \forall t \in [\delta, T].$$

Recall the definition of D_t in (3.4.2) which depends on T and η. Sending $\delta \to 0$, $\eta \to 0$ and $T \to \infty$, we conclude that $f = g$.

□

3.4.2. Assumptions on $\mathcal{A}_{\mathsf{H}}$

Lastly, we show that assumptions on $\mathcal{A}_{\mathsf{H}}$ in Theorem 3.2.1 are satisfied when H is convex and in the special case considered in Section 3.2.3 for $p = 1$ or p even.

Lemma 3.4.5. *If H is convex, then $\mathcal{A}_{\mathsf{H}}$ is convex and contains $\overline{F}_N(t, \cdot)$ for all t and N. In the special case where H is given in (3.2.17) and $p = 1$ or p is even, we have that H is convex.*

Proof. Note that, if $\phi : \mathbb{S}^K_+ \to \mathbb{R}$ is smooth, then we have

$$\nabla \cdot \big(\mathcal{D}\mathsf{H}(\nabla\phi)\big) = \mathcal{D}^2\mathsf{H}(\nabla\phi) \cdot \nabla^2\phi.$$

If H is convex, a sufficient condition for the above to be nonnegative is the convexity of ϕ. Recall that convexity is required in the definition of $\mathcal{A}$ given above (3.2.11). Hence, by regularizing functions in $\mathcal{A}$, we can see $\mathcal{A}_{\mathsf{H}} = \mathcal{A}$ when H is convex. It is also clear that $\mathcal{A}$ is convex. Due to (3.3.8), (3.3.10), and (3.3.12), we have $\overline{F}_N(t, \cdot) \in \mathcal{A}$ for all t and N. This completes the proof of the first part of the lemma.

Now, let H be given in (3.2.17). By computing the limit of $\varepsilon^{-1}(\mathsf{H}(q + \varepsilon a) - \mathsf{H}(q))$, we can see $a \cdot \mathcal{D}\mathsf{H}(q) = pa \cdot q^{\circ p-1}$ where $\circ$ denotes the Hadamard product. Differentiate one more time to get

$$a \cdot \mathcal{D}(a \cdot \mathcal{D}\mathsf{H})(q) = p(p-1)(a^{\circ 2}) \cdot (q^{\circ p-2}) \tag{3.4.8}$$

for all $a \in \mathbb{S}^K$ and $q \in \mathbb{S}^K_+$. If $p = 1$ or p is even, this quantity is nonnegative. Hence the convexity of H follows. □

3.5. Convergence to the weak solution

The goal of this section is to prove Theorem 3.2.1. The plan is to first prove the convergence of $\overline{F}_N$ assuming the existence of a weak solution f to (3.2.5) with $f(0, \cdot) = \psi$. Next, we prove the existence of solutions by using a similar argument. We adopt this plan because notation is much simpler in the first part, and the two parts are independent. Theorem 3.2.1 follows from Proposition 3.5.1 and Proposition 3.5.2 proved in Section 3.5.1 and Section 3.5.2, respectively.

3.5.1. Convergence when assuming existence of solutions

Let us assume f is a weak solution to (3.2.5) satisfying $f(0, \cdot) = \psi$. We want to show that $\overline{F}_N$ converges to f as $N \to \infty$. The goal can be summarized as follows.

Proposition 3.5.1. *In addition to the assumptions in Theorem 3.2.1, we assume that there is a unique weak solution f to* (3.2.5) *with $f(0, \cdot) = \psi$. Then, there is $C > 0$ such that* (3.2.12) *holds for all $M \geqslant 1$ and all $N \in \mathbb{N}$.*

Proof. Step 1. For $N \in \mathbb{N}$, we set

$$r_N = \partial_t \overline{F}_N - \mathsf{H}\big(\nabla \overline{F}_N\big). \tag{3.5.1}$$

For $\delta > 0$, define $\phi_\delta : \mathbb{R} \to [0, \infty)$ by

$$\phi_\delta(s) = (\delta + s^2)^{\frac{1}{2}}, \tag{3.5.2}$$

which serves as a smooth approximation of the absolute value. Since $\overline{F}_N$ is Lipschitz uniformly in N due to (3.3.8), we can set $M = \|f\|_{\mathrm{Lip}} \vee \sup_{N\in\mathbb{N}} \|\overline{F}_N\|_{\mathrm{Lip}}$. Then, we apply Lemma 3.4.3 to $\overline{F}_N, f, M, \phi_\delta$, and any choice of T, η to see that

$$\frac{\mathrm{d}}{\mathrm{d}t} J_\delta(t) \leqslant \int_{D_t} \phi'_\delta\big(\overline{F}_N - f\big)|r_N| \leqslant \int_{D_t} |r_N|, \quad \forall t \in [0, T], \tag{3.5.3}$$

where

$$J_\delta(t) = \int_{D_t} \phi_\delta\big(\overline{F}_N - f\big)(t, h)\mathrm{d}h, \tag{3.5.4}$$

for D_t given in (3.4.2). Also recall the definition of R in Lemma 3.4.3.

Step 2. We estimate $\int_{D_t} |r_N|$. Due to the definition of r_N in (3.5.1), Proposition 3.3.1 gives an upper bound for $|r_N|^2$. Hence, writing

$$\gamma = K(K+1)/2, \tag{3.5.5}$$

we have

$$\begin{aligned}\int_{D_t} |r_N| &\leqslant |D_t|^{\frac{1}{2}} \bigg(\int_{D_t} |r_N|^2\bigg)^{\frac{1}{2}} \\ &\leqslant CT^{\gamma/2}\bigg(N^{-\frac{1}{4}} \int_{D_t} \kappa(h)\big(\Delta \overline{F}_N + C|h^{-1}|\big)^{\frac{1}{4}}\mathrm{d}h + \int_{D_t} \mathbb{E}\big|\nabla F_N - \nabla \overline{F}_N\big|^2 \mathrm{d}h\bigg)^{\frac{1}{2}}.\end{aligned} \tag{3.5.6}$$

Here and henceforth, we absorb the constant R in the definition of D_t in (3.4.2) into the constant C. To bound the first integral in (3.5.6), recall the definition of $\kappa(h)$ in (3.3.1), use the definition of D_t and invoke Hölder's inequality to see

$$\int_{D_t} \kappa(h)\big(\Delta\overline{F}_N + C|h^{-1}|\big)^{\frac{1}{4}}\mathrm{d}h \leqslant C\eta^{-1}T|D_t|^{\frac{3}{4}}\bigg(\int_{D_t}\Delta\overline{F}_N + |h^{-1}|\bigg)^{\frac{1}{4}}.$$

In view of (3.3.8), using integration by parts, we have

$$\int_{D_t}\Delta\overline{F}_N \leqslant CT^{\gamma-1}.$$

The integral $\int_{D_t}|h^{-1}|$ is bounded by $C\eta^{-1}T^{\gamma}$. Therefore, we obtain

$$\int_{D_t} \kappa(h)\big(\Delta\overline{F}_N + C|h^{-1}|\big)^{\frac{1}{4}}\mathrm{d}h \leqslant C\eta^{-\frac{5}{4}}T^{1+\gamma}.$$

To avoid heavy notation, let us write

$$\overline{\mathcal{K}} = \frac{\mathcal{K}_{RT,n}}{T^{\beta}}, \qquad \mathcal{L} = \mathcal{L}_{\psi,RT,n}. \tag{3.5.7}$$

Here, β is given in the assumption of Theorem 3.2.1. For the last integral in (3.5.6), we will show in Step 4 that

$$\mathbb{E}\int_{D_t}\big|\nabla(F_N - \overline{F}_N)\big|^2 \leqslant CT^{\gamma+\beta}\eta^{-\frac{3}{2}}\overline{\mathcal{K}}. \tag{3.5.8}$$

These estimates imply that

$$\int_{D_t}|r_N| \leqslant CT^{\gamma+\frac{\beta\vee 1}{2}}\eta^{-\frac{3}{4}}\big(N^{-\frac{1}{8}} + \overline{\mathcal{K}}^{\frac{1}{2}}\big). \tag{3.5.9}$$

Step 3. We estimate $J_\delta(t)$, extend the integration from over D_t to $\mathbb{S}^K_{+,R(T-t)}$ (defined in

(3.2.7)), and conclude the result. Use (3.5.9) and (3.5.3) to see

$$J_\delta(t) \leqslant J_\delta(0) + CT^\alpha \eta^{-\frac{3}{4}}\big(N^{-\frac{1}{8}} + \overline{\mathcal{K}}^{\frac{1}{2}}\big), \quad t \in [0,T], \tag{3.5.10}$$

where we set

$$\alpha = \gamma + \frac{\beta \vee 1}{2} + 1. \tag{3.5.11}$$

Recall definitions (3.2.9), (3.5.2) and (3.5.4). Hence, for $t = 0$, we have

$$\lim_{\delta \to 0} J_\delta(0) = \int_{D_0} \big|\overline{F}_N(0,h) - f(0,h)\big| \mathrm{d}h \leqslant CT^\gamma \mathcal{L}.$$

Sending $\delta \to 0$ in (3.5.10) and using the above display, we derive that

$$\sup_{t\in[0,T]} \int_{D_t} \big|\overline{F}_N(t,h) - f(t,h)\big| \mathrm{d}h \leqslant CT^\alpha \Big(\mathcal{L} + \eta^{-\frac{3}{4}}\big(N^{-\frac{1}{8}} + \overline{\mathcal{K}}^{\frac{1}{2}}\big)\Big).$$

Due to (3.3.8) and the fact that $\overline{F}_N(0,0) = 0$, we have $|\overline{F}_N(t,h)| \leqslant C(t + |h|)$ uniformly in N. By $\overline{F}_N(0,0) = 0$ and the assumption on ψ in Theorem 3.2.1, we can see $\psi(0) = 0$. Since $f(0,\cdot) = \psi$ and the definition of weak solutions requires f to be Lipschitz, we have $|f(t,h)| \leqslant C(t + |h|)$. In addition, the measure of the set $\mathbb{S}^K_{+,R(T-t)} \setminus D_t$ is bounded by $CT^{\gamma-1}\eta$. Hence, we have

$$\sup_{t\in[0,T]} \int_{\mathbb{S}^K_{+,R(T-t)} \setminus D_t} \big|\overline{F}_N(t,h) - f(t,h)\big| \mathrm{d}h \leqslant \sup_{t\in[0,T]} \int_{\mathbb{S}^K_{+,R(T-t)} \setminus D_t} CT \leqslant CT^\gamma \eta,$$

Therefore, we obtain

$$\sup_{t\in[0,T]} \int_{\mathbb{S}^K_{+,R(T-t)}} \big|\overline{F}_N(t,h) - f(t,h)\big| \mathrm{d}h \leqslant CT^\alpha \Big(\eta + \mathcal{L} + \eta^{-\frac{3}{4}}\big(N^{-\frac{1}{8}} + \overline{\mathcal{K}}^{\frac{1}{2}}\big)\Big).$$

Let us now specify T and η. We set T proportional to M to ensure $[0,M] \times \mathbb{S}^K_{+,M} \subseteq \{(t,h) : t \in [0,T],\ h \in \mathbb{S}^K_{+,R(T-t)}\}$. Inserting this T and $\eta = (N^{-\frac{1}{8}} + \overline{\mathcal{K}}^{\frac{1}{2}})^{\frac{4}{7}}$ into the above display to see

$$\sup_{t\in[0,M]}\int_{\mathbb{S}^K_{+,M}}\big|\overline{F}_N(t,h)-f(t,h)\big|\mathrm{d}h\leqslant CM^\alpha\big(\mathcal{L}+N^{-\frac{1}{14}}+\overline{\mathcal{K}}^{\frac{2}{7}}\big). \tag{3.5.12}$$

Recall the notation (3.5.5), (3.5.7) and (3.5.11). This gives the desired result (3.2.12).

Step 4. To complete the proof, it remains to verify (3.5.8). Integrating by parts, we have

$$\begin{aligned}\int_{D_t}\big|\nabla(F_N-\overline{F}_N)\big|^2&=\int_{\partial D_t}(F_N-\overline{F}_N)\nabla(F_N-\overline{F}_N)\cdot\mathbf{n}-\int_{D_t}(F_N-\overline{F}_N)\Delta(F_N-\overline{F}_N)\\&\leqslant\|F_N-\overline{F}_N\|_{L^\infty([0,RT]\times\mathbb{S}^K_{+,RT})}\bigg(\int_{\partial D_t}\big|\nabla(F_N-\overline{F}_N)\big|+\int_{D_t}\big|\Delta(F_N-\overline{F}_N)\big|\bigg),\end{aligned} \tag{3.5.13}$$

Let us estimate the last integral. The lower bound (3.3.12) shows $\Delta\overline{F}_N\geqslant 0$, and the lower bound (3.3.13) implies that

$$\Delta F_N+CN^{-\frac{1}{2}}|Z||h^{-1}|^{\frac{3}{2}}\geqslant 0.$$

These yield

$$\begin{aligned}\int_{D_t}\big|\Delta(F_N-\overline{F}_N)\big|&\leqslant\int_{D_t}\big|\Delta F_N\big|+\big|\Delta\overline{F}_N\big|\\&\leqslant CT^\gamma N^{-\frac{1}{2}}\eta^{-\frac{3}{2}}|Z|+\int_{D_t}\big(\Delta F_N+\Delta\overline{F}_N\big).\end{aligned}$$

Applying integration by parts to the last integral and using (3.3.8) and (3.3.9), we can see that

$$\int_{D_t}\big(\Delta F_N+\Delta\overline{F}_N\big)\leqslant\int_{\partial D_t}|\nabla F_N|+|\nabla\overline{F}_N|\leqslant CT^{\gamma-1}\big(1+N^{-\frac{1}{2}}\eta^{-\frac{1}{2}}|Z|\big).$$

This display also serves as a bound for the first integral in (3.5.13). Insert the above two displays into (3.5.13) to get

$$\int_{D_t} \left|\nabla(F_N - \overline{F}_N)\right|^2 \leqslant CT^\gamma \|F_N - \overline{F}_N\|_{L^\infty([0,RT]\times\mathbb{S}^K_{+,RT})}\eta^{-\frac{3}{2}}\Big(1 + N^{-\frac{1}{2}}|Z|\Big).$$

Recall (3.2.8) and (3.5.7). Take expectations on both sides of this inequality and invoke the Cauchy–Schwarz inequality to conclude (3.5.8).

□

3.5.2. Existence of weak solutions

To complete the proof of Theorem 3.2.1, we need the following existence result.

Proposition 3.5.2. *Under the assumptions in Theorem 3.2.1, there is a unique weak solution f to* (3.2.5) *with* $f(0,\cdot) = \psi$.

Proof. The uniqueness part follows from Proposition 3.4.2. Hence, we only need to prove the existence. We first show that $(\overline{F}_N)_{N\in\mathbb{N}}$ is a Cauchy sequence in the local uniform topology and then verify that the limit is a weak solution.

Step 1. We show that the sequence $(\overline{F}_N)_{N\in\mathbb{N}}$ is Cauchy. We proceed similarly as in the previous subsection. Recall the definition of r_N in (3.5.1) and ϕ_δ in (3.5.2). Let $N, N' \in \mathbb{N}$. Now, setting $M = \sup_{N\in\mathbb{N}} \|\overline{F}_N\|_{\mathrm{Lip}}$ and applying Lemma 3.4.3 to $\overline{F}_N$ and $\overline{F}_{N'}$, we obtain

$$\frac{\mathrm{d}}{\mathrm{d}t} J_\delta(t) \leqslant \int_{D_t} \phi'_\delta\big(\overline{F}_N - \overline{F}_{N'}\big)|r_N - r_{N'}| \leqslant \int_{D_t} |r_N| + |r_{N'}|$$

where

$$J_\delta(t) = \int_{D_t} \phi_\delta\big(\overline{F}_N - \overline{F}_{N'}\big)(t,h)\mathrm{d}h.$$

The rest follows exactly the same procedure after (3.5.4) in the previous section. The only difference is that we have more terms due to the presence of $\overline{F}_{N'}$, but they are treated in the same way as for $\overline{F}_N$. Similar to (3.5.12), one can see that eventually we obtain

$$\sup_{t\in[0,M]}\int_{S^K_{+,M}}\left|\overline{F}_N(t,h)-\overline{F}_{N'}(t,h)\right|\mathrm{d}h \leqslant CM^\alpha\Big(\mathcal{L}_{\psi,CM,N}+N^{-\frac{1}{14}}+(\mathcal{K}_{CM,N}/M^\beta)^{\frac{2}{7}}$$
$$+\mathcal{L}_{\psi,CM,N'}+N'^{-\frac{1}{14}}+(\mathcal{K}_{CM,N'}/M^\beta)^{\frac{2}{7}}\Big).$$

Hence, by the assumption of Theorem 3.2.1 on the decay of $\mathcal{K}_{M,N}$ and $\mathcal{L}_{\psi,M,N}$, we know that $(\overline{F}_N)_{N\in\mathbb{N}}$ is Cauchy in local $L^\infty_t L^1_h$. Due to the argument in Remark 3.2.4, we can upgrade this to $(\overline{F}_N)_{N\in\mathbb{N}}$ being Cauchy in local $L^\infty_t L^\infty_h$. Let us denote the limit by f.

Step 2. We verify that f is a weak condition by checking that each property listed in Definition 3.4.1 is satisfied by f and that $f(0,\cdot)=\psi$.

Firstly, we verify that f is Lipschitz and satisfies the initial condition. Since $\overline{F}_N$ is Lipschitz uniformly in N due to (3.3.8), we can conclude that f is Lipschitz. Due to the assumption $\lim_{N\to\infty} L_{\psi,M,N}=0$, we have $f(0,\cdot)=\psi$.

Next, we show that $f(t,\cdot)\in\mathcal{A}_{\mathsf{H}}$ for every $t\geqslant 0$. By (3.3.11) and (3.3.12), we have that both $\overline{F}_N$ and f are convex in the temporal variable and convex in the spacial variable. It is well known that convexity implies convergence of derivatives at each point of differentiability. The Lipschitzness of f and Rademacher's theorem imply that f is differentiable almost everywhere (a.e.). Hence, we can deduce that $(\partial_t,\nabla)\overline{F}_N$ converges to $(\partial_t,\nabla)f$ pointwise a.e. Since $\overline{F}_N(t,\cdot)\in\mathcal{A}_{\mathsf{H}}$ for every t and N, the claim can be verified by passing to the limit.

Lastly, we show that f satisfies (3.2.5) a.e. Since $\overline{F}_N$ is Lipschitz uniformly in N due to (3.3.8) and H is continuous, the bounded convergence theorem implies that, for any compact $B\subseteq\mathbb{S}^K_{++}$ and t a.e.,

$$\int_B\Big|\partial_t f-\mathsf{H}(\nabla f)\Big|(t,h)\mathrm{d}h=\lim_{n\to\infty}\int_B\Big|\partial_t\overline{F}_N-\mathsf{H}(\nabla\overline{F}_N)\Big|(t,h)\mathrm{d}h.$$

We want to show that the right hand side is zero. Recall the definition of D_t in (3.4.2). By choosing T and δ in D_t suitably, we can ensure $B \subseteq D_t$. Then, by (3.5.1), (3.5.9) and the assumption $\lim_{N\to\infty} \mathcal{K}_{M,N} = 0$ in the statement of Theorem 3.2.1, we conclude that the right hand side of the above display is zero. Since B and t are arbitrary, we conclude that $\partial_t f - \mathsf{H}(\nabla f) = 0$ a.e.

□

3.6. Viscosity solutions of Hamilton–Jacobi equations

In this section, we give the precise definition of viscosity solutions. After that, we prove the comparison principle which ensures the uniqueness of solutions. In addition, we verify that the Hopf formula is a solution. Classical references include [60, 42]. See also [15, 84]. Here, we follow the approach in [96].

A function $f : \mathbb{R}_+ \times \mathbb{S}^K_+ \to \mathbb{R}$ is said to be nondecreasing if $f(t,x) - f(t',x') \geqslant 0$ whenever $t \geqslant t'$ and $x - x' \in \mathbb{S}^K_+$. A function $\psi : \mathbb{S}^K_+ \to \mathbb{R}$ is said to be nondecreasing if $\psi(x) - \psi(x') \geqslant 0$ whenever $x - x' \in \mathbb{S}^K_+$.

Definition 3.6.1.

1. A nondecreasing Lipschitz function $f : \mathbb{R}_+ \times \mathbb{S}^K_+ \to \mathbb{R}$ is a viscosity subsolution to (3.2.5) if for every $(t,x) \in (0,\infty) \times \mathbb{S}^K_+$ and every smooth $\phi : (0,\infty) \times \mathbb{S}^K_+ \to \mathbb{R}$ such that $f - \phi$ has a local maximum at (t,x), we have

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

2. A nondecreasing Lipschitz function $f : \mathbb{R}_+ \times \mathbb{S}^K_+ \to \mathbb{R}$ is a viscosity supersolution to (3.2.5) if for every $(t,x) \in (0,\infty) \times \mathbb{S}^K_+$ and every smooth $\phi : (0,\infty) \times \mathbb{S}^K_+ \to \mathbb{R}$ such

that $f - \phi$ has a local minimum at (t, x), we have

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

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

3. A nondecreasing Lipschitz function $f : \mathbb{R}_+ \times \mathbb{S}^K_+ \to \mathbb{R}$ is a viscosity solution to (3.2.5) if f is both a viscosity subsolution and supersolution.

Remark 3.6.2. The restriction $|q| \leqslant \|f\|_{\mathrm{Lip}}$ under the infimum in Definition 3.6.1 (2) can be replaced by $|q| \leqslant \|f\|_{\mathrm{Lip}} + c$ for any $c \geqslant 0$. Indeed, since f is assumed to be Lipschitz, we can always restrict H to the set $\{q \in \mathbb{S}^K_+ : \ |q| \leqslant c'\}$ without altering the equation (3.2.5) as long as $c' \geqslant \|f\|_{\mathrm{Lip}}$. Aside from this heuristic, one can straightforwardly check that the choice of c does not affect the results in this and the next sections.

Remark 3.6.3. The only properties of H used in this section are the positivity $\mathsf{H} \geqslant 0$, local Lipschitzness as in (3.6.6) and nondecreasingness given by Lemma 3.4.4. The following two propositions are still valid for general H satisfying these three properties.

Remark 3.6.4. It is easy to see that, in Definition 3.6.1, replacing the phrases "local maximum" and "local minimum" by "strict local maximum" and "strict local minimum", respectively, yields an equivalent definition.

Proposition 3.6.5 (Comparison principle)**.** *If u is a subsolution and v is a supersolution to* (3.2.5)*, then*

$$\sup_{\mathbb{R}_+ \times \mathbb{S}^K_+} (u - v) = \sup_{\{0\} \times \mathbb{S}^K_+} (u - v).$$

Proposition 3.6.6 (Hopf formula)**.** *Suppose $\psi : \mathbb{S}^K_+ \to \mathbb{R}$ is convex, Lipschitz and nondecreasing. Let f be given in* (3.2.14)*. Then f is a viscosity solution to* (3.2.5) *with initial* *condition* $f(0,\cdot) = \psi$.

3.6.1. Proof of Proposition 3.6.5

Let us argue by contradiction and assume

$$\sup_{\mathbb{R}_+ \times \mathbb{S}^K_+} (u - v) > \sup_{\{0\} \times \mathbb{S}^K_+} (u - v). \tag{3.6.1}$$

We start by modifying u. For $\varepsilon \in (0,1)$ to be specified later, we set

$$u_\varepsilon(t,x) = u(t,x) + \varepsilon \mathrm{tr}(x) - C\varepsilon t, \quad \forall (t,x) \in \mathbb{R}_+ \times \mathbb{S}^K_+$$

where tr stands for the trace. Let I be the $K \times K$ identity matrix. By choosing C large and then ε small, we can ensure that, if $u_\varepsilon - \phi$ attains a local maximum at (t,x), we have

$$\begin{cases} (\partial_t \phi - \mathsf{H}(\nabla \phi))(t,x) \leqslant -2\varepsilon, & \text{if } x \in \mathbb{S}^K_{++} \\ (\nabla \phi - \varepsilon I)(t,x) \in \mathbb{S}^K_+, & \text{if } x \in \mathbb{S}^K_+ \setminus \mathbb{S}^K_{++}. \end{cases} \tag{3.6.2}$$

Since $u(t,\cdot)$ in nondecreasing for each t, we also have

$$u_\varepsilon(t, x+y) - u_\varepsilon(t,x) \geqslant \varepsilon \mathrm{tr}(y), \quad y \in \mathbb{S}^K_+. \tag{3.6.3}$$

With ε sufficiently small chosen, (3.6.1) still holds with u replaced by u_ε. Next, we replace u_ε by $u_\varepsilon - \frac{\delta}{T-t}$, where δ is chosen small enough and $T > 1$ is chosen large enough by (3.6.1) to ensure that

$$\sup_{[0,T) \times \mathbb{S}^K_+} (u_\varepsilon - v) > \sup_{\{0\} \times \mathbb{S}^K_+} (u_\varepsilon - v). \tag{3.6.4}$$

Also, note that (3.6.2) still holds. In addition, we have, for every $M > 0$,

$$\lim_{\eta \to 0} \sup_{[T-\eta, T) \times \mathbb{S}^K_{+,M}} u_\varepsilon = -\infty. \tag{3.6.5}$$

Next, we introduce some parameters and auxiliary functions. By the formula for H in (3.2.6), there is a constant C_{H} such that

$$\big|\mathsf{H}(a)-\mathsf{H}(b)\big| \leqslant C_{\mathsf{H}}|a-b|(|a|+|b|)^{p-1}, \quad \forall a,b\in\mathbb{S}^K_+. \tag{3.6.6}$$

Let

$$L = 1+\|u\|_{\mathrm{Lip}}+\|v\|_{\mathrm{Lip}}, \qquad K = C_{\mathsf{H}}(4L)^{p-1}. \tag{3.6.7}$$

Due to the definition of u_ε, the following holds for all $(t,x)\in[0,T)\times\mathbb{S}^K_+$,

$$u_\varepsilon(t,x)\leqslant C+L|x|, \qquad \|\nabla u_\varepsilon(t,x)\|\leqslant L. \tag{3.6.8}$$

By (3.6.4), there is $(\bar{t},\overline{x})$ such that

$$\big(u_\varepsilon - v\big)(\bar{t},\overline{x}) > \sup_{\{0\}\times\mathbb{S}^K_+}(u_\varepsilon - v). \tag{3.6.9}$$

Let us set

$$R = \big(|\overline{x}|^2+1\big)^{\frac{1}{2}} + K\bar{t}.$$

Take $\chi:\mathbb{R}\to\mathbb{R}_+$ to be a smooth function satisfying

$$(r-1)_+\leqslant\chi(r)\leqslant r_+, \quad |\chi'(r)|\leqslant 1, \quad \forall r\in\mathbb{R}, \tag{3.6.10}$$

where the positive sign in the subscript indicates taking the positive part. The function χ can be viewed as a smoothed version of $r\mapsto r_+$. Define $\eta:[0,T)\times\mathbb{S}^K_+\to\mathbb{R}$ by

$$\eta(t,x) = 2L\chi\Big((|x|^2+1)^{\frac{1}{2}}+Kt-R\Big). \tag{3.6.11}$$

We claim

$$\sup_{[0,T)\times\mathbb{S}_+^K}(u_\varepsilon - v - \eta) = \sup_{\{0\}\times\mathbb{S}_+^K}(u_\varepsilon - v - \eta). \tag{3.6.12}$$

On the other hand, due to (3.6.9) and the definitions of R and η, we have

$$\sup_{[0,T)\times\mathbb{S}_+^K}(u_\varepsilon - v - \eta) \geqslant \big(u_\varepsilon - v\big)(\bar{t},\bar{x}) > \sup_{\{0\}\times\mathbb{S}_+^K}(u_\varepsilon - v) \geqslant \sup_{\{0\}\times\mathbb{S}_+^K}(u_\varepsilon - v - \eta),$$

which contradicts (3.6.12). Hence, the proof is complete once the claim (3.6.12) is verified.

Proof of (3.6.12)

Again we argue by contradiction and assume

$$\sup_{[0,T)\times\mathbb{S}_+^K}(u_\varepsilon - v - \eta) > \sup_{\{0\}\times\mathbb{S}_+^K}(u_\varepsilon - v - \eta). \tag{3.6.13}$$

We are going to employ the classical trick of "doubling the variables". For $\alpha \in (0,1)$, we introduce

$$\Psi_\alpha(t,x,t,x') = u_\varepsilon(t,x) - v(t',x') - \phi_\alpha(t,x,t',x'), \quad \forall t \in [0,T),\ t' > 0,\ x,x' \in \mathbb{S}_+^K.$$

where

$$\phi_\alpha(t,x,t',x') = \frac{1}{2\alpha}\big(|t-t'|^2 + |x-x'|^2\big) + \eta(t,x).$$

Step 1. We show that there exists a maximizer $(t_\alpha, x_\alpha, t'_\alpha, x'_\alpha)$ of Ψ_α, and they converge as $\alpha \to 0$. To start, we seek an upper bound for Ψ_α. The nondecreasingness of v gives $-v(t,x) \leqslant -v(0,0)$. The definition of η in (3.6.11) shows $\eta(t,x) \geqslant 2L(|x| + Kt - R - 1)$. Using these and the first inequality in (3.6.8), we have

$$\Psi_\alpha(t,x,t',x') \leqslant C - L|x| - \frac{1}{2\alpha}\big(|t-t'|^2 + |x-x'|^2\big) - 2KLt.$$

Here and henceforth, we absorb L, K and R into C. Now, one can see the existence of a maximizer $(t_\alpha, x_\alpha, t'_\alpha, x'_\alpha)$. Then, we have

$$\Psi_\alpha(t_\alpha, x_\alpha, t'_\alpha, x'_\alpha) \geqslant \Psi_\alpha(0,0,0,0) = u_\varepsilon(0,0) - v(0,0).$$

Combine the above two displays to see that, for all $\alpha < 1$, these points $(t_\alpha, x_\alpha, t'_\alpha, x'_\alpha)$ lie in a bounded set and

$$|t_\alpha - t'_\alpha|^2 + |x_\alpha - x'_\alpha|^2 \leqslant C\alpha.$$

By passing to a subsequence, we can assume there is t_0 and x_0 such that $t_\alpha, t'_\alpha \to t_0$ and $x_\alpha, x'_\alpha \to x_0$ as $\alpha \to 0$.

In view of (3.6.5), we must have $t_0 < T$. The maximality of $(t_\alpha, x_\alpha, t'_\alpha, x'_\alpha)$ yields

$$\begin{aligned}\big(u_\varepsilon - v - \eta\big)(t_0, x_0) &\leqslant \sup_{[0,T)\times\mathbb{S}^K_+} (u_\varepsilon - v - \eta)\\ &\leqslant \Psi_\alpha(t_\alpha, x_\alpha, t'_\alpha, x'_\alpha) \leqslant u_\varepsilon(t_\alpha, x_\alpha) - v(t'_\alpha, x'_\alpha) - \eta(t_\alpha, x_\alpha).\end{aligned}$$

Take $\alpha \to 0$ and use the continuity of u_ε, v and η to see

$$\big(u_\varepsilon - v - \eta\big)(t_0, x_0) = \sup_{[0,T)\times\mathbb{S}^K_+} (u_\varepsilon - v - \eta).$$

By (3.6.13), we must have $t_0 > 0$. Henceforth, we fix a sufficiently small α so that $t_\alpha, t'_\alpha > 0$.

Step 2. For this fixed α, note that

$$(t,x) \mapsto u_\varepsilon(t,x) - v(t'_\alpha, x'_\alpha) - \phi_\alpha(t, x, t'_\alpha, x'_\alpha) \tag{3.6.14}$$

has a local maximum at (t_α, x_α). We argue that

$$x_\alpha \in \mathbb{S}^K_{++}. \tag{3.6.15}$$

Otherwise, there is $y \in \mathbb{S}^K_+$ with $|y| = 1$ such that

$$y \cdot x_\alpha = 0. \tag{3.6.16}$$

Under this assumption, we want to derive a contradiction to the fact that the maximum is achieved (t_α, x_α). For $\delta > 0$, using (3.6.3), we can see

$$\begin{aligned} &u_\varepsilon(t_\alpha, x_\alpha + \delta y) - \phi_\alpha(t_\alpha, x_\alpha + \delta y, t'_\alpha, x'_\alpha) - \Big(u_\varepsilon(t_\alpha, x_\alpha) - \phi_\alpha(t_\alpha, x_\alpha, t'_\alpha, x'_\alpha)\Big) \\ &\qquad \geqslant \varepsilon\delta \mathrm{tr}(y) - \frac{1}{2\alpha}\Big(2\delta y \cdot (x_\alpha - x'_\alpha) + \delta^2\Big) - \Big(\eta(t_\alpha, x_\alpha + \delta y) - \eta(t_\alpha, x_\alpha)\Big). \end{aligned} \tag{3.6.17}$$

The definition of η in (3.6.11) allows us to compute

$$\nabla\eta(t,x) = \frac{x}{(|x|^2+1)^{\frac{1}{2}}} 2L\chi'\Big((|x|^2+1)^{\frac{1}{2}} + Lt - R\Big). \tag{3.6.18}$$

By (3.6.16), we have $y \cdot \nabla\eta(t_\alpha, x_\alpha) = 0$. This along with Taylor's theorem implies

$$\eta(t_\alpha, x_\alpha + \delta y) - \eta(t_\alpha, x_\alpha) = \mathcal{O}(\delta^2).$$

Apply this, (3.6.16) and $y \cdot x'_\alpha \geqslant 0$ to see that (3.6.17) is bounded below by

$$\varepsilon\delta\mathrm{tr}(y) - \mathcal{O}(\delta^2).$$

Since $\varepsilon > 0$ and $\mathrm{tr}(y) > 0$, this is strictly positive for δ small. This contradicts the fact that (3.6.14) achieves a local maximum at (t_α, x_α). By contradiction, we must have (3.6.15).

Using this, (3.6.2), and the maximality of (3.6.14) at (t_α, x_α), we obtain

$$\frac{1}{\alpha}(t_\alpha - t'_\alpha) + \partial_t \eta(t_\alpha, x_\alpha) - \mathsf{H}\left(\frac{1}{\alpha}(x_\alpha - x'_\alpha) + \nabla\eta(t_\alpha, x_\alpha)\right) \leqslant -2\varepsilon. \tag{3.6.19}$$

Step 3. Still for this fixed α, the function

$$(t', x') \mapsto v(t', x') - u_\varepsilon(t_\alpha, x_\alpha) + \phi_\alpha(t_\alpha, x_\alpha, t', x')$$

attains a local minimum at (t'_α, x'_α). Note that $-\nabla_{x'}\phi_\alpha(t_\alpha, x_\alpha, t', x') = \frac{1}{\alpha}(x_\alpha - x')$. We claim that there is $a \in \mathbb{S}^K_+$ such that

$$a - \frac{1}{\alpha}(x_\alpha - x'_\alpha) \in \mathbb{S}^K_+, \tag{3.6.20}$$

$$|a| \leqslant \|v\|_{\mathrm{Lip}}, \tag{3.6.21}$$

$$\frac{1}{\alpha}(t_\alpha - t'_\alpha) - \mathsf{H}(a) \geqslant -\varepsilon. \tag{3.6.22}$$

If $x'_\alpha \in \mathbb{S}^K_{++}$, then by setting $a = \frac{1}{\alpha}(x_\alpha - x'_\alpha)$, we clearly have (3.6.20). In this case, the local minimum is achieved at an interior point x'_α. Since v is nondecreasing, we can see $\frac{1}{\alpha}(x_\alpha - x'_\alpha) \in \mathbb{S}^K_+$ and thus $a \in \mathbb{S}^K_+$. Then (3.6.22) follows from the definition of supersolutions. If $v(t'_\alpha, \cdot)$ is differentiable at x'_α, then the minimality at x'_α implies $\frac{1}{\alpha}(x_\alpha - x'_\alpha) = \nabla v(t'_\alpha, x'_\alpha)$ and hence (3.6.21) holds. If x'_α is not a point of differentiability, then (3.6.21) still holds by a regularizing argument.

If $x'_\alpha \notin \mathbb{S}^K_{++}$, namely $x'_\alpha \in \mathbb{S}^K_+ \setminus \mathbb{S}^K_{++}$, then the existence of a and (3.6.20)–(3.6.22) directly follow from the boundary condition in the definition of supersolutions.

Step 4. We compare (3.6.19) with (3.6.22) to derive a contradiction. To start, we derive some estimates. For simplicity, we write

$$b = \nabla\eta(t_\alpha, x_\alpha).$$

Recall the definition of the constant L in (3.6.7). Due to (3.6.18) and the second inequality in (3.6.10), we get $|b| \leqslant 2L$. By (3.6.21), we have $|a| \leqslant L$. These along with (3.6.6) yield

$$\left|\mathsf{H}(a+b) - \mathsf{H}(a)\right| \leqslant C_{\mathsf{H}}|b|(4L)^{p-1}.$$

Using the definition of η in (3.6.11), we can see

$$\partial_t\eta(t_\alpha, x_\alpha) \geqslant K|\nabla\eta(t_\alpha, x_\alpha)| = K|b|.$$

The above two displays together with the definition of K in (3.6.7) imply

$$\partial_t\eta(t_\alpha, x_\alpha) - \mathsf{H}(a+b) + \mathsf{H}(a) \geqslant 0. \tag{3.6.23}$$

On the other hand, from (3.6.19) and (3.6.20), using the monotonicity of H in Lemma 3.4.4, we have

$$\frac{1}{\alpha}(t_\alpha - t'_\alpha) + \partial_t\eta(t_\alpha, x_\alpha) - \mathsf{H}(a+b) \leqslant -2\varepsilon.$$

Subtract (3.6.22) from the above display to obtain

$$\partial_t\eta(t_\alpha, x_\alpha) - \mathsf{H}(a+b) + \mathsf{H}(a) \leqslant -\varepsilon.$$

This contradicts (3.6.23) and thus the proof of (3.6.12) is complete.

3.6.2. Proof of Proposition 3.6.6

Let us rewrite the Hopf formula (3.2.14) as

$$f(t,x) = \sup_{z\in\mathbb{S}^K_+}\{z\cdot x - \psi^*(z) + t\mathsf{H}(z)\} \tag{3.6.24}$$

$$= (\psi^* - t\mathsf{H})^*(x). \tag{3.6.25}$$

Here the superscript $*$ denotes the Fenchel transformation over $\mathbb{S}^K_+$, namely,

$$u^*(x) = \sup_{y\in\mathbb{S}^K_+}\{y\cdot x - u(y)\}, \quad \forall x\in\mathbb{S}^K_+. \tag{3.6.26}$$

We check the following in order: nondecreasingness, initial condition, semigroup property (or dynamic programming principle), Lipschitzness, f being a subsolution, and f being a supersolution.

Nondecreasingness

Since the supremum in (3.6.24) is taken over $\mathbb{S}^K_+$, it is clear from Lemma 3.3.3 that $f(t,\cdot)$ is nondecreasing. By the formula of H in (3.2.6) and the Schur product theorem, we have $\mathsf{H}(z)\geqslant 0$ for all $z\in\mathbb{S}^K_+$. Hence, from the formula (3.2.14), we can see f is also nondecreasing in t.

Verification of the initial condition

The desired identity

$$\psi(x) = \sup_{z\in\mathbb{S}^K_+}\inf_{y\in\mathbb{S}^K_+}\left\{z\cdot(x-y)+\psi(x)\right\} = \psi^{**}(x), \quad \forall x\in\mathbb{S}^K_+.$$

follows from a version of Fenchel–Moreau identity stated in Proposition 3.9.1

Semigroup property

Let f be given in (3.6.24). We want to show, for all $s \geqslant 0$,

$$f(t+s,x) = \sup_{z\in\mathbb{S}^K_+}\inf_{y\in\mathbb{S}^K_+}\left\{z\cdot(x-y)+f(t,y)+s\mathsf{H}(z)\right\},$$

or, in a more compact form,

$$f(t+s,\cdot) = \big(f^*(t,\cdot)-s\mathsf{H}\big)^*. \tag{3.6.27}$$

In view of the Hopf formula (3.6.25), this is equivalent to

$$\big(\psi^*-(t+s)\mathsf{H}\big)^* = \big((\psi^*-t\mathsf{H})^{**}-s\mathsf{H}\big)^*. \tag{3.6.28}$$

From the definition of the Fenchel transform (3.6.26), it can be seen that, for any u,

$$u^{**} \leqslant u. \tag{3.6.29}$$

Since the Fenchel transform is order-reversing, (3.6.29) implies that

$$\big((\psi^*-t\mathsf{H})^{**}-s\mathsf{H}\big)^* \geqslant \big(\psi^*-(t+s)\mathsf{H}\big)^*. \tag{3.6.30}$$

To see the other direction, we use (3.6.29) to get

$$\frac{s}{t+s}\psi^* + \frac{t}{t+s}\big(\psi^*-(t+s)\mathsf{H}\big)^{**} \leqslant \psi^*-t\mathsf{H}.$$

For any u, it can be readily checked that u^* is convex and lower semi-continuous. Using the argument in Section 3.6.2, we can deduce that u^* is non-decreasing. Hence the left hand side of the above display satisfies the condition in Proposition 3.9.1. Therefore, taking the

Fenchel transform twice in the above display and applying Proposition 3.9.1, we have

$$\frac{s}{t+s}\psi^* + \frac{t}{t+s}\big(\psi^* - (t+s)\mathsf{H}\big)^{**} \leqslant (\psi^* - t\mathsf{H})^{**}.$$

Reorder terms and then use (3.6.29) to see

$$\big(\psi^* - (t+s)\mathsf{H}\big)^{**} - (\psi^* - t\mathsf{H})^{**} \leqslant \frac{s}{t}\Big((\psi^* - t\mathsf{H})^{**} - \psi^*\Big) \leqslant -s\mathsf{H}.$$

This immediately gives

$$\big(\psi^* - (t+s)\mathsf{H}\big)^{**} \leqslant (\psi^* - t\mathsf{H})^{**} - s\mathsf{H}.$$

Taking the Fenchel transform on both sides and invoking Proposition 3.9.1, we have

$$\big(\psi^* - (t+s)\mathsf{H}\big)^{*} \geqslant \big((\psi^* - t\mathsf{H})^{**} - s\mathsf{H}\big)^{*}.$$

Here, we also used the order-reversing property of the Fenchel transform. This together with (3.6.30) verifies (3.6.28).

Lipschitzness

Since ψ is Lipschitz, we have $\psi^*(z) = \infty$ outside the compact set $\{z \in \mathbb{S}^K_+ : |z| \leqslant \|\psi\|_{\mathrm{Lip}}\}$. This together with (3.6.24) implies that for each $x \in \mathbb{S}^K_+$, there is $z \in \mathbb{S}^K_+$ with $|z| \leqslant \|\psi\|_{\mathrm{Lip}}$ such that

$$f(t,x) = z \cdot x - \psi^*(z) + t\mathsf{H}(z).$$

This yields that, for any $x' \in \mathbb{S}^K_+$,

$$f(t,x) - f(t,x') \leqslant z \cdot (x - x') \leqslant \|\psi\|_{\mathrm{Lip}}|x - x'|.$$

By symmetry, we conclude that f is Lipschitz in x, and the Lipschitz coefficient is uniform in t.

To show the Lipschitzness in t, we fix any $x \in \mathbb{S}^K_+$. Then, we have, for some $z \in \mathbb{S}^K_+$ with $|z| \leqslant \|\psi\|_{\mathrm{Lip}}$,

$$\begin{aligned} f(t,x) = z \cdot x - \psi^*(z) + t\mathsf{H}(z) &\leqslant f(t',x) + (t-t')\mathsf{H}(z) \\ &\leqslant f(t',x) + |t'-t|\left(\sup_{|z|\leqslant\|\psi\|_{\mathrm{Lip}}} |\mathsf{H}(z)|\right). \end{aligned}$$

Again by symmetry, the Lipschitzness in t is obtained, and its coefficient is independent of x.

The Hopf formula is a subsolution

Let $\phi : (0,\infty) \times \mathbb{S}^K_+ \to \mathbb{R}$ be smooth. Suppose $f - \phi$ achieves a local maximum at $(t,x) \in (0,\infty) \times \mathbb{S}^K_+$. Since ψ is Lipschitz, we can see ψ^* is infinite outside a compact set. Hence, by (3.6.24), there is $\overline{z} \in \mathbb{S}^K_+$ such that

$$f(t,x) = \overline{z} \cdot x - \psi^*(\overline{z}) + t\mathsf{H}(\overline{z}).$$

For the case $x \in \mathbb{S}^K_{++}$, by (3.6.24), we have, for $s \in [0,t]$ and $h \in \mathbb{S}^K$ sufficiently small,

$$f(t,x) \leqslant f(t-s, x+h) - \overline{z} \cdot h + s\mathsf{H}(\overline{z}).$$

By the assumption on ϕ, we have

$$f(t-s,x+h) - \phi(t-s,x+h) \leqslant f(t,x) - \phi(t,x).$$

for small $s \in [0,t]$ and small $h \in \mathbb{S}^K$. Combine the above two inequalities to get

$$\phi(t,x) - \phi(t-s,x+h) \leqslant -\overline{z} \cdot h + s\mathsf{H}(\overline{z}). \tag{3.6.31}$$

Set $s = 0$ and vary h to see

$$\overline{z} = \nabla\phi(t, x). \tag{3.6.32}$$

Then, we set $h = 0$ in (3.6.31), take $s \to 0$ and insert (3.6.32) to obtain

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

If $x \in \mathbb{S}^K_+ \setminus \mathbb{S}^K_{++}$, then (3.6.31) still holds for $h \in \mathbb{S}^K_+$. Set $s = 0$ and vary h, we can see $a \cdot \nabla\phi(t, x) \geqslant a \cdot \overline{z}$ for all $a \in \mathbb{S}^K_+$. Since $\overline{z} \in \mathbb{S}^K_+$, Lemma 3.3.3 implies that $\nabla\phi(t, x) \in \mathbb{S}^K_+$.

The Hopf formula is a supersolution

The idea of proof in this part can be seen in [84, Proof of Proposition 1]. Let $(t, x) \in (0, \infty) \times \mathbb{S}^K_+$ be a local minimum point for $f - \phi$. Due to (3.6.24), f is convex in both variables. Since $\mathbb{S}^K_+$ is also convex, we have, for all $(t', x') \in (0, \infty) \times \mathbb{S}^K_+$ and all $\lambda \in [0, 1]$,

$$f(t', x') - f(t, x) \geqslant \frac{1}{\lambda}\Big(f\big(t + \lambda(t' - t), x + \lambda(x' - x)\big) - f(t, x)\Big).$$

For any fixed (t', x') and sufficiently small λ, the assumption that $f - \phi$ has a local minimum at (t, x) gives

$$f\big(t + \lambda(t' - t), x + \lambda(x' - x)\big) - f(t, x) \geqslant \phi\big(t + \lambda(t' - t), x + \lambda(x' - x)\big) - \phi(t, x).$$

Using the above two displays and setting $\lambda \to 0$, we obtain

$$f(t', x') - f(t, x) \geqslant r(t' - t) + \big(\nabla\phi(t, x)\big) \cdot (x' - x) \tag{3.6.33}$$

where

$$r = \partial_t\phi(t, x). \tag{3.6.34}$$

Before proceeding, we make a digression to convex analysis. Most of the definitions and results we need are given in Appendix 3.9. For each fixed $t \geqslant 0$, it can be seen from (3.6.24) that $f(t,\cdot)$ is convex. Setting $t' = t$ in (3.6.33), we have $\nabla\phi(t,x) \in \partial f(t,x)$ which stands for the subdifferential of $f(t,\cdot)$ at x. Its definition is given in (7.3.2). Invoking Lemma 3.9.6, we can express

$$\nabla\phi(t,x) = a + b \tag{3.6.35}$$

where $b \in \mathbf{n}(x)$, the outer normal cone at x, defined in (3.9.2); and a belongs to the closed convex hull of limit points of the form $\lim_{n\to\infty} \nabla f(t,x_n)$ where $\lim_{n\to\infty} x_n = x$ and $f(t,\cdot)$ is differentiable at each x_n. Since f is nondecreasing and Lipschitz, we have

$$a \in \mathbb{S}^K_+, \qquad |a| \leqslant \|f\|_{\mathrm{Lip}}. \tag{3.6.36}$$

By the definition of $\mathbf{n}(x)$ and Lemma 3.3.3, it can seen that $-b \in \mathbb{S}^K_+$. This along with (3.6.35) implies

$$a \in \nabla\phi(t,x) + \mathbb{S}^K_+. \tag{3.6.37}$$

By Lemma 3.9.6, the definition of a and an easy observation that $0 \in \mathbf{n}(x)$, we can deduce that $a \in \partial f(t,x)$, which due to the definition of subdifferential in (7.3.2) further implies

$$f(t,x') - f(t,x) \geqslant a\cdot(x'-x), \quad \forall x' \in \mathbb{S}^K_+.$$

Set $x' = x$ in (3.6.33) and use the above display to get

$$f(t',x') - f(t,x) \geqslant r(t'-t) + a\cdot(x'-x), \quad \forall t' \geqslant 0,\, x' \in \mathbb{S}^K_+. \tag{3.6.38}$$

Now, we return to the proof. For each $s \geqslant 0$, we define

$$\eta_s(x') = f(t,x) - rs + a \cdot (x' - x), \quad \forall x' \in \mathbb{S}^K_+.$$

Setting $t' = t - s$ in (3.6.38), for $s \in [0,t]$, we have

$$f(t-s, x') \geqslant \eta_s(x'), \quad \forall x' \in \mathbb{S}^K_+.$$

Applying the order-reversing property of the Fenchel transform twice, we obtain from the above display that

$$\left(f^*(t-s,\cdot) - s\mathsf{H}\right)^* \geqslant \left(\eta_s^* - s\mathsf{H}\right)^*.$$

Due to the semigroup property (3.6.27), this yields

$$f(t,\cdot) \geqslant \left(\eta_s^* - s\mathsf{H}\right)^*, \quad \forall s \in [0,t].$$

By (3.6.36) and the definition of the Fenchel transform in (3.6.26), the above yields

$$f(t,x) \geqslant a \cdot x - \eta_s^*(a) + s\mathsf{H}(a).$$

On the other hand, using the definition of η_s, we can compute

$$\eta_s^*(a) = -f(t,x) + rs + a \cdot x.$$

Combine the above two displays with (3.6.34) and that these hold for all $s \in [0,t]$ to see

$$\left(\partial_t \phi - \mathsf{H}(a)\right)(t,x) \geqslant 0.$$

If $x \in \mathbb{S}^K_{++}$, then (3.6.36), (3.6.37) and the nondecreasingness of H in Lemma 3.4.4 imply

$\mathsf{H}(a) \geqslant \mathsf{H}(\nabla\phi(t,x))$. If $x \in \mathbb{S}^K_+ \setminus \mathbb{S}^K_{++}$, then those same ingredients yield $\mathsf{H}(a) \geqslant \inf \mathsf{H}(q)$ where the infimum is described in (2) in Definition 3.6.1. These along with the above display verify that f is a supersolution.

3.7. Convergence to the viscosity solution

The goal of this section is to prove Theorem 3.2.2. We first state the main result of this section and deduce Theorem 3.2.2 from it.

Proposition 3.7.1. *Under the assumptions in Theorem 3.2.2, suppose that a subsequence of $(\overline{F}_N)_{N\in\mathbb{N}}$ converges locally uniformly to some function $f : \mathbb{R}_+ \times \mathbb{S}^K_+ \to \mathbb{R}$. Then, f is a viscosity subsolution to* (3.2.5) *with $f(0,\cdot) = \psi$. If H is convex, then f is also a supersolution and thus the unique viscosity solution to* (3.2.5)*.*

Remark 3.7.2. In fact, any subsequential limit f of $(\overline{F}_N)_{N\in\mathbb{N}}$ satisfies the following: if $f - \phi$ achieves a strict local maximum at $(t,x) \in (0,\infty) \times \mathbb{S}^K_{++}$ for a smooth function ϕ, then it holds that

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

which is stronger than Definition 3.6.1 (1).

Proof of Theorem 3.2.2. By (3.3.12), (3.3.8), (3.3.10) and the assumption that $\overline{F}_N(0,\cdot)$ converges to ψ pointwise, we have that ψ is convex, Lipschitz and nondecreasing. Hence, Proposition 3.6.6 implies that there is a Lipschitz viscosity solution f to the Hamilton–Jacobi equation (3.2.5) with $f(0,\cdot) = \psi$. Proposition 3.6.5 ensures the uniqueness.

Since $\overline{F}_N(0,0) = 0$ for all N and $(\overline{F}_N)_{N\geqslant 1}$ is Lipschitz uniformly in N due to (3.3.8), the Arzelà–Ascoli theorem guarantees that any subsequence of $(\overline{F}_N)_{N\geqslant 1}$ has a further subsequence that converges in the local uniform topology to some function g. In addition, we can see that g is Lipschitz. The assumption on ψ in Theorem 3.2.2 ensures that $g(0,\cdot) = \psi$. Proposition 3.7.1 implies that g is a viscosity subsolution to (3.2.5). The upper bound in Theorem 3.2.2 then follows from Proposition 3.6.5. When H is convex, using similar arguments, we can obtain an lower bound. □

We prove the subsolution part of Proposition 3.7.1 and Remark 3.7.2 in Section 3.7.1 and the supersolution part of Proposition 3.7.1 in Section 3.7.2.

3.7.1. The limit is a subsolution

To lighten notation, we assume $\overline{F}_N$ converges to f locally uniformly. We want to show f is subsolution to (3.2.5). We recall Remark 3.6.4 and assumes that $f - \phi$ achieves a strict local maximum at $(t, h) \in (0, \infty) \times \mathbb{S}^K_+$ for some smooth function ϕ.

First, we consider the case where $h \in \mathbb{S}^K_+ \setminus \mathbb{S}^K_{++}$. Then, there is a sequence $\big((t_N, h_N)\big)_{N\in\mathbb{N}}$ in $(0, \infty) \times \mathbb{S}^K_+$ such that (t_N, h_N) converges to (t, h) and $\overline{F}_N - \phi$ has a local maximum at (t_N, h_N). Note that $a + h_N \in \mathbb{S}^K_+$ for all $a \in \mathbb{S}^K_+$. So, we can differentiate $\overline{F}_N - \phi$ along any direction $a \in \mathbb{S}^K_+$ to see

$$a \cdot \nabla\big(\overline{F}_N - \phi\big)(t_N, h_N) \leqslant 0, \quad \forall a \in \mathbb{S}^K_+.$$

In view of (3.3.10), this implies

$$a \cdot \nabla\phi(t_N, h_N) \geqslant 0, \quad \forall a \in \mathbb{S}^K_+.$$

Setting $N \to \infty$, by Lemma 3.3.3, we have $\nabla\phi(t, h) \in \mathbb{S}^K_+$, verifying the boundary condition for subsolutions.

Now, we study the case where $h \in \mathbb{S}^K_{++}$. In the following, the constant C is allowed to depend on t, h, f, ϕ. We set

$$M = (t \vee |h|) + 1, \tag{3.7.1}$$

$$\gamma = K(K+1)/2, \tag{3.7.2}$$

$$\delta_N = \|\overline{F}_N - f\|_{L^\infty([0,M]\times \mathbb{S}^K_{+,M})}^{\frac{1}{4}} + \mathcal{K}_{M,N}^{\frac{1}{2}}, \tag{3.7.3}$$

where $\mathcal{K}_{M,N}$ is defined in (3.2.8) and $\mathbb{S}^K_{+,M}$ is given in (3.2.7). By the convergence of $\overline{F}_N$ to f and the assumption (3.2.13), we have $\lim_{N\to\infty} \delta_N = 0$. Let us introduce

$$\widetilde{\phi}(t', h') = \phi(t', h') + |t' - t|^2 + |h' - h|^2. \tag{3.7.4}$$

It is immediate that $f - \widetilde{\phi}$ has a local maximum at (t, h). Due to (3.7.3), for all $(t', h') \in [0, M] \times \mathbb{S}^K_{+,M}$, we have

$$\begin{aligned} \big(\overline{F}_N - \widetilde{\phi}\big)(t', h') &\leqslant \big(f - \phi\big)(t', h') - |t' - t|^2 - |h' - h|^2 + \delta_N^4, \\ \big(\overline{F}_N - \widetilde{\phi}\big)(t, h) &\geqslant \big(f - \phi\big)(t, h) - \delta_N^4. \end{aligned}$$

Since $\overline{F}_N$ converges locally uniformly to f, for N large, there is a sequence of (t_N, h_N) in $(0, \infty) \times \mathbb{S}^K_{++}$, at which $\overline{F}_N - \widetilde{\phi}$ attains a local maximum, and which converges to (t, h). From the above display and the fact that $f - \phi$ attains a local maximum at (t, h), we can deduce that

$$|t_N - t|^2 + |h_N - h|^2 \leqslant 2\delta_N^4. \tag{3.7.5}$$

By the definition of (t_N, h_N), we also have

$$\partial_t\big(\overline{F}_N - \widetilde{\phi}\big)(t_N, h_N) = 0, \quad \nabla\big(\overline{F}_N - \widetilde{\phi}\big)(t_N, h_N) = 0. \tag{3.7.6}$$

We want to apply Proposition 3.3.1. However the concentration estimate we have is for $F_N - \overline{F}_N$ not for $\nabla(F_N - \overline{F}_N)$. Therefore, we need to do a local average by introducing

$$D_N = \mathbb{S}^K_{+,\delta_N}, \tag{3.7.7}$$

$$G_N(t', h') = |D_N|^{-1} \int_{h'+D_N} \overline{F}_N(t', h'')\mathrm{d}h''. \tag{3.7.8}$$

It is clear that G_N converges locally uniformly to f. Hence, there is $(t'_N, h'_N) \in (0,\infty)\times\mathbb{S}^K_{++}$ converging to (t,h) such that $G_N - \widetilde{\phi}$ has a local maximum at (t'_N, h'_N). Consequently, we have

$$\partial_t\big(G_N - \widetilde{\phi}\big)(t'_N, h'_N) = 0, \quad \nabla\big(G_N - \widetilde{\phi}\big)(t'_N, h'_N) = 0, \tag{3.7.9}$$

$$a \cdot \nabla\Big(a \cdot \nabla\big(G_N - \widetilde{\phi}\big)\Big)(t'_N, h'_N) \leqslant 0, \quad \forall a \in \mathbb{S}^K. \tag{3.7.10}$$

Repeating the argument in the derivation of (3.7.5) yields

$$|t'_N - t|^2 + |h'_N - h|^2 \leqslant 2\delta_N^4. \tag{3.7.11}$$

We need the following estimates:

$$\int_{h'_N+D_N} \mathbb{E}\Big|\nabla F_N - \nabla\overline{F}_N\Big|^2(t'_N, h')\mathrm{d}h' \leqslant C\delta_N^{\gamma+1}, \tag{3.7.12}$$

$$\int_{h'_N+D_N} \Big|\nabla\overline{F}_N(t'_N, h') - \nabla G_N(t'_N, h'_N)\Big|^2\mathrm{d}h' \leqslant C\delta_N^{\gamma+1}. \tag{3.7.13}$$

From the definition of H in (3.2.6), we can see that $|\mathsf{H}(a) - \mathsf{H}(b)| \leqslant C|a-b|(|a|\vee|b|)^{p-1}$ for all $a, b \in \mathbb{S}^K_+$. By this, Jensen's inequality and (3.7.13), we have

$$\begin{aligned} &\left|\ |D_N|^{-1}\int_{h'_N+D_N} \mathsf{H}\big(\nabla\overline{F}_N(t'_N, h')\big)\mathrm{d}h' - \mathsf{H}\big(\nabla G_N(t'_N, h'_N)\big)\right| \\ &\quad\leqslant C\bigg(|D_N|^{-1}\int_{h'_N+D_N}\Big|\nabla\overline{F}_N(t'_N, h') - \nabla G_N(t'_N, h'_N)\Big|^2\mathrm{d}h'\bigg)^{\frac{1}{2}} \leqslant C\delta_N^{\frac{1}{2}}. \end{aligned} \tag{3.7.14}$$

Here, we used the following fact due to (3.7.2) and (3.7.7)

$$|D_N| = C\delta_N^{\gamma}.$$

Recall the definition of κ in (3.3.1). Due to $h \in \mathbb{S}^K_{++}$, (3.7.7) and (3.7.11), we know that $\kappa(h') \leqslant C$ for all $h' \in h'_N + D_N$ and N large. Take average of $(\partial_t \overline{F}_N - \mathsf{H}(\nabla \overline{F}_N))(t'_N, h')$ over $h'_N + D_N$, and use Proposition 3.3.1 and (3.7.14) to see

$$\begin{aligned}
&\Big|\partial_t G_N - \mathsf{H}\big(\nabla G_N\big)\Big|(t'_N, h'_N) \leqslant C\delta_N^{\frac{1}{2}} \\
&+ C\left(N^{-\frac{1}{4}} \fint_{h'_N + D_N} \big(\Delta \overline{F}_N(t'_N, h') + 1\big)^{\frac{1}{4}} \mathrm{d}h' + \fint_{h'_N + D_N} \mathbb{E}\Big|\nabla F_N - \nabla \overline{F}_N\Big|^2 (t'_N, h')\mathrm{d}h'\right)^{\frac{1}{2}}
\end{aligned}$$

where $\fint_{h'_N + D_N} = |D_N|^{-1}\int_{h'_N + D_N}$. By Jensen's inequality, (3.7.8) and (3.7.10), we have

$$\fint_{h'_N + D_N} \big(\Delta \overline{F}_N(t'_N, \cdot) + 1\big)^{\frac{1}{4}} \leqslant \Big(\Delta G_N(t'_N, h'_N) + 1\Big)^{\frac{1}{4}} \leqslant C.$$

The above two displays along with (3.7.12) give

$$\Big|\partial_t G_N - \mathsf{H}\big(\nabla G_N\big)\Big|(t'_N, h'_N) \leqslant C\Big(\delta_N^{\frac{1}{2}} + N^{-\frac{1}{8}}\Big).$$

Using (3.7.11) and (3.7.9), and sending N to ∞, we obtain

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

Due to (3.7.4), the derivatives of $\widetilde{\phi}$ coincide with those of ϕ at (t, h). This finishes the core of the verification of that f is a subsolution and the claim in Remark 3.7.2.

To complete the proof, we derive (3.7.12) and (3.7.13).

Proof of (3.7.12). For any smooth $g : \mathbb{S}^K_+ \to \mathbb{R}$ and any $D \subseteq \mathbb{S}^K_+$ with Lipschitz boundary, integration by parts gives

$$\int_D |\nabla g|^2 = \int_{\partial D} g\nabla g \cdot \mathbf{n} - \int_D g\Delta g, \tag{3.7.15}$$

where $\mathbf{n}$ is the outer normal on ∂D. To lighten our notation, the time variable is always evaluated at t'_N in this proof. Apply (3.7.15) to get

$$\begin{aligned}\int_{h'_N+D_N} \left|\nabla F_N - \nabla \overline{F}_N\right|^2 &\leqslant \|F_N - \overline{F}_N\|_{L^\infty(h'_N+D_N)} \\ &\quad \times \left(\int_{\partial(h'_N+D_N)} \left|\nabla F_N - \nabla \overline{F}_N\right| + \int_{h'_N+D_N} \left|\Delta F_N - \Delta \overline{F}_N\right| \right).\end{aligned} \tag{3.7.16}$$

By $\lim_{N\to\infty} h'_N = h$ (due to (3.7.11)), $h \in \mathbb{S}^K_{++}$ and (3.7.7), we have $|h'^{-1}| \leqslant C$ for all $h' \in h'_N + D_N$ for large N, Using this, (3.3.12) and (3.3.13), we get, for all $h' \in h'_N + D_N$,

$$\left|\Delta F_N - \Delta \overline{F}_N\right| \leqslant \Delta F_N + \Delta \overline{F}_N + CN^{-\frac{1}{2}}|Z|.$$

Applying this and integration by parts to obtain

$$\int_{h'_N+D_N} \left|\Delta F_N - \Delta \overline{F}_N\right| \leqslant C\delta_N^\gamma N^{-\frac{1}{2}}|Z| + \int_{\partial(h'_N+D_N)} |\nabla F_N| + |\nabla \overline{F}_N|.$$

Then, using this display, (3.3.8) and (3.3.9), we can bound the two integrals in (3.7.16) by $C\delta_N^{\gamma-1}(1 + N^{-\frac{1}{2}}|Z|)$. As a result, by taking expectations and invoking the Cauchy–Schwarz inequality in (3.7.16), we obtain

$$\mathbb{E}\int_{h'_N+D_N} \left|\nabla F_N - \nabla \overline{F}_N\right|^2 \leqslant C\delta_N^{\gamma-1}\left(\mathbb{E}\|F_N - \overline{F}_N\|^2_{L^\infty(h'_N+D_N)}\right)^{\frac{1}{2}}. \tag{3.7.17}$$

Recall that the time variable is evaluated at t'_N. By (3.7.1), (3.7.11) and (3.7.7), we have $\{t'_N\} \times (h'_N + D_N) \subseteq [0, M] \times \mathbb{S}^K_{+,M}$ for large N. Hence, the desired result (3.7.12) follows from (3.7.3) and the definition (3.2.8).

□

Proof of (3.7.13). To prepare, we start by showing that, for h' satisfying $|h'-h_N| \leqslant C^{-1}$,

$$\left|\overline{F}_N(t_N,h') - \overline{F}_N(t_N,h_N) - (h'-h_N)\cdot\nabla\overline{F}_N(t_N,h_N)\right| \leqslant C|h'-h_N|^2. \tag{3.7.18}$$

By Taylor expansion, we have

$$\begin{aligned}\overline{F}_N(t_N,h') - \overline{F}_N(t_N,h_N) &= (h'-h_N)\cdot\nabla\overline{F}_N(t_N,h_N)\\ &+ \int_0^1 (1-r)\mathcal{D}^2_{h'-h_N}\overline{F}_N(t_N,h_N+(h'-h_N)r)\mathrm{d}r\end{aligned} \tag{3.7.19}$$

where we write

$$\mathcal{D}^2_a\overline{F}_N = a\cdot\nabla\big(a\cdot\nabla\overline{F}_N\big), \quad \forall a\in\mathbb{S}^K.$$

A similar equation also holds with $\overline{F}_N$ replaced by $\widetilde{\phi}$. Take the difference of these two equations and use (3.7.6) and the fact that $\overline{F}_N - \widetilde{\phi}$ has a local maximum at (t_N,h_N) to see

$$\begin{aligned}&\int_0^1 (1-r)\mathcal{D}^2_{h'-h_N}\overline{F}_N(t_N,h_N+(h'-h_N)r)\mathrm{d}r\\ \leqslant &\int_0^1 (1-r)\mathcal{D}^2_{h'-h_N}\widetilde{\phi}(t_N,h_N+(h'-h_N)r)\mathrm{d}r.\end{aligned}$$

Since $\widetilde{\phi}$ has locally bounded derivatives, by the above display and (3.3.12), there is C such that the following holds for all h' with $|h'-h_N| \leqslant C^{-1}$

$$\left|\int_0^1 (1-r)\mathcal{D}^2_{h'-h_N}\overline{F}_N(t_N,h_N+(h'-h_N)r)\mathrm{d}r\right| \leqslant C|h'-h_N|^2.$$

Inserting this into (3.7.19) gives (3.7.18).

Now, we are ready to prove (3.7.13). Let us set

$$g_N(h') = \overline{F}_N(t'_N,h') - \overline{F}_N(t'_N,h'_N) - (h'-h'_N)\cdot\nabla G_N(t'_N,h'_N).$$

Note that, to probe (3.7.13), it is sufficient to estimate $\int_{h'_N+D_N} |\nabla g_N|^2$. Using (3.3.8) and (3.7.8), we can see

$$|\nabla G_N(t', h')| \leqslant C, \quad \forall t',\ h'. \tag{3.7.20}$$

$$|\nabla g_N(h')| \leqslant C, \quad \forall h' \in h'_N + D_N. \tag{3.7.21}$$

Apply (3.7.15) to g_N to obtain

$$\int_{h'_N+D_N} |\nabla g_N|^2 \leqslant \|g_N\|_{L^\infty(h'_N+D_N)} \Big(\int_{\partial(h'_N+D_N)} |\nabla g_N| + \int_{h'_N+D_N} |\Delta g_N| \Big).$$

By (3.7.21), the first integral on the left is bounded by $C\delta_N^{\gamma-1}$. Since $\Delta g_N = \Delta \overline{F}_N(t'_N, \cdot)$, by (3.3.12), we can see $|\Delta g_N| = \Delta g_N$. Integrating by parts and applying (3.7.21) again, we deduce that the last integral in the above display is also bounded by $C\delta_N^{\gamma-1}$. Hence, we arrive at

$$\int_{h'_N+D_N} |\nabla g_N|^2 \leqslant C\delta_N^{\gamma-1} \|g_N\|_{L^\infty(h'_N+D_N)}. \tag{3.7.22}$$

It remains to estimate $\|g_N\|_{L^\infty(h'_N+D_N)}$. We want to compare g_N with

$$\overline{F}_N(t_N, h') - \overline{F}_N(t_N, h_N) - (h' - h_N) \cdot \nabla G_N(t_N, h_N).$$

To start, using (3.7.8), we can compute, for all a, t', h',

$$\begin{aligned} a \cdot \nabla G_N(t', h') &= |D_N|^{-1} \int_{D_N} a \cdot \nabla \overline{F}_N(t', h' + h'')\mathrm{d}h'' \\ &= |D_N|^{-1} \int_{\partial D_N} \overline{F}_N(t', h' + h'') a \cdot \mathbf{n}\mathcal{S}(\mathrm{d}h'') \end{aligned} \tag{3.7.23}$$

where in the last equality we used integration by parts and $\mathcal{S}$ denotes the surface measure on ∂D_N. Now, we estimate

$$\Big|(h' - h_N) \cdot \nabla G_N(t_N, h_N) - (h' - h'_N) \cdot \nabla G_N(t'_N, h'_N)\Big| \tag{3.7.24}$$
$$\leqslant |h_N - h'_N|\big|\nabla G_N(t_N, h_N)\big| + \Big|(h' - h'_N) \cdot \Big(\nabla G_N(t_N, h_N) - \nabla G_N(t'_N, h'_N)\Big)\Big|.$$

The first term after the inequality sign is bounded by $|h_N - h'_N|$ due to (3.7.20). Using (3.3.8) and (3.7.23), we can bound the second term by

$$|D_N|^{-1} \int_{\partial D_N} \Big(|t_N - t'_N| + |h_N - h'_N|\Big)|h' - h'_N| \leqslant C|t_N - t'_N| + C|h_N - h'_N|,$$

for all $h' \in h'_N + D_N$. Hence, we conclude that (3.7.24) is bounded by the right hand of the above display with a larger constant. This along with (3.3.8) implies that

$$\|g_N\|_{L^\infty(h'_N+D_N)} \leqslant C|t_N - t'_N| + C|h_N - h'_N|$$
$$+ \sup_{h' \in h'_N + D_N} \Big|\overline{F}_N(t_N, h') - \overline{F}_N(t_N, h_N) - (h' - h_N) \cdot \nabla G_N(t_N, h_N)\Big|.$$

By (3.7.18) and the definition of D_N in (3.7.7), the supremum above can be bounded by

$$C\big(\delta_N + |h_N - h'_N|\big)^2$$
$$+ \sup_{h' \in h'_N + D_N} \big|(h' - h_N) \cdot \nabla\overline{F}_N(t_N, h_N) - (h' - h_N) \cdot \nabla G_N(t_N, h_N)\big|.$$

We claim that

$$\sup_{h' \in h'_N + D_N} \big|(h' - h_N) \cdot \nabla\overline{F}_N(t_N, h_N) - (h' - h_N) \cdot \nabla G_N(t_N, h_N)\big| \leqslant C\delta_N^2. \tag{3.7.25}$$

This along with (3.7.5) and (3.7.11) implies that $\|g_N\|_{L^\infty(h'_N+D_N)} \leqslant C\delta_N^2$. Plug this into (3.7.22), and we obtain (3.7.13).

To complete the proof, we verify the claim (3.7.25). Using integration by parts, we can see

$$(h' - h_N) \cdot \nabla \overline{F}_N(t_N, h_N) = |D_N|^{-1} \int_{\partial D_N} \Big(h'' \cdot \nabla \overline{F}_N(t_N, h_N) \Big) (h' - h_N) \cdot \mathbf{n} \mathcal{S}(\mathrm{d}h'').$$

Using the formula (3.7.23) and $\int_{\partial D_N} c \cdot \mathbf{n} = 0$ for any constant vector c, we can also get

$$\begin{aligned} &(h' - h_N) \cdot \nabla G_N(t_N, h_N) = \\ &\qquad |D_N|^{-1} \int_{\partial D_N} \Big(\overline{F}_N(t_N, h_N + h'') - \overline{F}_N(t_N, h_N) \Big) (h' - h_N) \cdot \mathbf{n} \mathcal{S}(\mathrm{d}h''). \end{aligned}$$

Taking the difference of the above two equations and using (3.7.18), we can see the left hand side of (3.7.25) is bounded by

$$C \sup_{h' \in h'_N + D_N} \delta_N |h' - h_N| \leqslant C \delta_N \big(\delta_N + |h_N - h'_N| \big).$$

Now, (3.7.25) follows from (3.7.5) and (3.7.11).

□

3.7.2. The limit is a supersolution when H is convex

Under the additional assumption that H is convex, we show that any subsequential limit of $\overline{F}_N$ is a supersolution. For simplicity of notation, we again assume the entire sequence $(\overline{F}_N)_{N \in \mathbb{N}}$ converges locally uniformly to f. We recall Remark 3.6.4 and assume that $f - \phi$ achieves a strict local minimum at $(t, h) \in (0, \infty) \times \mathbb{S}^K_+$. Recall M from (3.7.1). Let us redefine

$$\delta_N = \max\{N^{-\frac{1}{6}}, \mathcal{K}^{\frac{2}{5}}_{M,N}\}, \tag{3.7.26}$$

$$D_N = \delta_N I + \mathbb{S}^K_{+,\delta_N},$$

$$G_N(t', h') = |D_N|^{-1} \int_{h' + D_N} \overline{F}_N(t', h'') \mathrm{d}h'', \quad \forall (t', h') \in \mathbb{R}_+ \times \mathbb{S}^K_+. \tag{3.7.27}$$

Note that in the definition of G_N, the integration is over a region away from h' to avoid the singularity present in the right hand side of the estimate in Proposition 3.3.1. It is clear that G_N converges locally uniformly to f. Then, there is a sequence $(t_N, h_N) \in (0,\infty)\times\mathbb{S}^K_+$ such that $\lim_{N\to\infty}(t_N, h_N) = (t, h)$ and $G_N - \phi$ has a local minimum at (t_N, h_N). Since H is convex, we integrate both sides of the inequality in Proposition 3.3.1 and use Jensen's inequality to see

$$\begin{aligned} \Big(\partial_t G_N - \mathsf{H}\big(\nabla G_N\big)\Big)(t_N, h_N) &\geqslant \fint_{h_N+D_N}\Big(\partial_t \overline{F}_N - \mathsf{H}(\nabla \overline{F}_N)\Big)(t_N, h')\mathrm{d}h' \\ \geqslant -C\Bigg(\fint_{h_N+D_N}\frac{|h'^{-1}|}{N^{\frac{1}{4}}}\big(\Delta\overline{F}_N + |h'^{-1}|\big)^{\frac{1}{4}}\mathrm{d}h' &+ \fint_{h_N+D_N}\mathbb{E}\Big|\nabla F_N - \nabla\overline{F}_N\Big|^2\Bigg)^{\frac{1}{2}} \end{aligned} \tag{3.7.28}$$

where $\fint_{h_N+D_N} = |D_N|^{-1}\int_{h_N+D_N}$ and the time variable is evaluated at t_N in (3.7.28).

Let us estimate the integrals in (3.7.28). The definition of D_N implies that

$$|h'^{-1}| \leqslant C\delta_N^{-1}, \quad \forall h' \in h_N + D_N. \tag{3.7.29}$$

Integrate by parts and use (3.3.8) to see

$$\Delta G_N(t_N, h_N) = \fint_{h_N+D_N}\Delta\overline{F}_N(t_N, h')\mathrm{d}h' \leqslant |D_N|^{-1}\int_{\partial(h_N+D_N)}\big|\nabla\overline{F}_N(t_N,\cdot)\big| \leqslant C\delta_N^{-1}.$$

The above two displays together with Jensen's inequality and (3.7.26) implies that

$$\begin{aligned} &\fint_{h_N+D_N} N^{-\frac{1}{4}}|h'^{-1}|\big(\Delta\overline{F}_N(t_N, h') + |h'^{-1}|\big)^{\frac{1}{4}}\mathrm{d}h' \\ &\qquad\leqslant CN^{-\frac{1}{4}}\delta_N^{-1}\Big(\Delta G_N(t_N, h_N) + \delta_N^{-1}\Big)^{\frac{1}{4}} \leqslant C\delta_N^{\frac{1}{4}}. \end{aligned} \tag{3.7.30}$$

To estimate the last integral in (3.7.28), we use the same argument in the proof of (3.7.12). The only difference is that since now it is possible that $h \in \mathbb{S}^K_+\setminus\mathbb{S}^K_{++}$, the singularity in the estimate (3.3.9) takes effect. Due to (3.7.29), compared with (3.7.17), there is an additional

$\delta_N^{-\frac{1}{2}}$. For N large, we have

$$\mathbb{E}\int_{h_N+D_N}\left|\nabla F_N-\nabla\overline{F}_N\right|^2 \leqslant C\delta_N^{\gamma-\frac{3}{2}}\left(\mathbb{E}\|F_N-\overline{F}_N\|^2_{L^\infty(h_N+D_N)}\right)^{\frac{1}{2}}$$

$$\leqslant C\delta_N^{\gamma-\frac{3}{2}}\mathcal{K}_{M,N}\leqslant C\delta_N^{\gamma+1}, \tag{3.7.31}$$

where we used (3.2.8) and (3.7.1) in the penultimate inequality, and (3.7.26) in the last inequality. Inserting (3.7.30) and (3.7.31) into (3.7.28), we obtain

$$\Big(\partial_t G_N-\mathsf{H}\big(\nabla G_N\big)\Big)(t_N,h_N)\geqslant -C\delta_N^{\frac{1}{8}}. \tag{3.7.32}$$

First suppose that there are infinitely many (t_N,h_N) with $h_N\in\mathbb{S}^K_{++}$. Since first derivatives of G_N coincides with ϕ at those (t_N,h_N), by taking $N\to\infty$ and using the smoothness of ϕ, we obtain from (3.7.32) that

$$\Big(\partial_t\phi-\mathsf{H}\big(\nabla\phi\big)\Big)(t,h)\geqslant 0. \tag{3.7.33}$$

If there are infinitely many (t_N,h_N) with $h_N\in\mathbb{S}^K_+\setminus\mathbb{S}^K_{++}$, then we must have $h\in\mathbb{S}^K_+\setminus\mathbb{S}^K_{++}$. Due to $t\in(0,\infty)$ and $\lim_{N\to\infty}t_N=t$, for large N, we have $t_N\in(0,\infty)$. Since $G_N-\phi$ has a local minimum at (t_N,h_N), we have

$$\big(\partial_t G_N-\partial_t\phi\big)(t_N,h_N)=0, \tag{3.7.34}$$

$$\big(\nabla G_N-\nabla\phi\big)(t_N,h_N)\in\mathbb{S}^K_+. \tag{3.7.35}$$

We also used Lemma 3.3.3 in deriving (3.7.35). By the definition of G_N in (3.7.27), the nondecreasingness of $\overline{F}_N$ in (3.3.10), and the uniform Lipschitzness of $\overline{F}_N$ in (3.3.8), we have, for all $N\in\mathbb{N}$,

$$\nabla G_N\in\mathbb{S}^K_+,\qquad |\nabla G_N|\leqslant\|\overline{F}_N\|_{\mathrm{Lip}}\leqslant C, \tag{3.7.36}$$

where the last constant C is absolute. In addition, due to (3.3.12), G_N is convex in the second variable, which yields

$$y \cdot \nabla G_N(t_N, h_N) \leqslant G_N(t_N, h_N + y) - G_N(t_N, h_N), \quad \forall y \in \mathbb{S}^K_+.$$

Let a be any subsequential limit of $\big(\nabla G_N(t_N, h_N)\big)_{N\in\mathbb{N}}$. Replace y by $y_N = \nabla G_N(t_N, h_N)$ in the above display and use $\lim_{N\to\infty}(t_N, h_N) = (t, h)$ and the local uniform convergence of G_N towards f to see

$$|a|^2 \leqslant f(t, h + a) - f(t, h).$$

The Lipschitzness of f implies

$$|a| \leqslant \|f\|_{\mathrm{Lip}}. \tag{3.7.37}$$

We extract a subsequence from $\big(\nabla G_N(t_N, h_N)\big)_{N\in\mathbb{N}}$, along which

$$\liminf_{N\to\infty} \mathsf{H}\big(\nabla G_N(t_N, h_N)\big)$$

is achieved. Denote by a the further subsequential limit of this minimizing sequence. By this and the continuity of H, we obtain

$$\liminf_{N\to\infty} \mathsf{H}\big(\nabla G_N(t_N, h_N)\big) = \mathsf{H}(a). \tag{3.7.38}$$

Due to (3.7.35), (3.7.36) and $\lim_{N\to\infty}(t_N, h_N) = (t, h)$, we also have

$$a - \nabla\phi(t, h) \in \mathbb{S}^K_+, \qquad a \in \mathbb{S}^K_+. \tag{3.7.39}$$

Recall the quantity $\inf \mathsf{H}(q)$ for the boundary condition in (2) of Definition 3.6.1. By (3.7.37) and (3.7.39), we have

$$\mathsf{H}(a) \geqslant \inf \mathsf{H}(q).$$

Use this, (3.7.34), (3.7.38) and (3.7.32) to get

$$\begin{aligned}\Big(\partial_t \phi - \inf \mathsf{H}(q)\Big)(t,h) &\geqslant \lim_{N\to\infty} \partial_t G_N(t_N,h_N) - \mathsf{H}(a) \\ &= \lim_{N\to\infty} \partial_t G_N(t_N,h_N) - \liminf_{N\to\infty} \mathsf{H}\big(\nabla G_N(t_N,h_N)\big) \\ &\geqslant \limsup_{N\to\infty} \Big(\partial_t G_N - \mathsf{H}\big(\nabla G_N\big)\Big)(t_N,h_N) \geqslant 0.\end{aligned}$$

This along with (3.7.33) completes our verification that f is a supersolution.

3.8. Nonsymmetric matrix inference

The goal of this appendix is to demonstrate a case where H is not convex, yet the assumptions on $\mathcal{A}_{\mathsf{H}}$ in Theorem 3.2.1 are satisfied. Let X_1 and X_2 be two random vectors in $\mathbb{R}^N$. The task is to infer the nonsymmetric matrix $X_1 X_2^\intercal$ from the noisy observation

$$Y = \sqrt{\frac{2t}{N}} X_1 X_2^\intercal + W \in \mathbb{R}^{N\times N}. \tag{3.8.1}$$

Let $X = \mathsf{diag}(X_1, X_2) \in \mathbb{R}^{2N\times 2}$. We can compute

$$X \otimes X = \mathsf{diag}\big(X_1 \otimes X_1,\ X_1 \otimes X_2,\ X_2 \otimes X_1,\ X_2 \otimes X_2\big) \in \mathbb{R}^{4N^2\times 4}.$$

Let $A = (0,1,0,0) \in \mathbb{R}^4$. Then note that the non-zero entries of $(X\otimes X)A$ are those from $X_1 \otimes X_2$, which are exactly the entries of $X_1 X_2^\intercal$. As observed in [103], the model (3.8.1) is equivalent to the model

$$Y = \sqrt{\frac{2t}{N}} X^{\otimes 2} A + W \in \mathbb{R}^{4N^2\times 1},$$

which is a special case of (3.1.1).

By the formula of H in (3.2.6), we can compute $\mathsf{H}(q) = q_{11}q_{22}$ and thus $\mathcal{D}\mathsf{H}(q) = \mathsf{diag}(q_{22}, q_{11})$ for all $q \in \mathbb{S}^2_+$. Recall the set $\mathcal{A}$ defined above (3.2.11). Then for smooth $\phi \in \mathcal{A}$, using the basis (3.4.1), we can obtain

$$\nabla \cdot \mathcal{D}\mathsf{H}(\nabla\phi) = 2e^{11} \cdot \nabla(e^{22} \cdot \nabla\phi).$$

Hence, formally, $\mathcal{A}_\mathsf{H}$ consists of those $\phi \in \mathcal{A}$ whose second order derivative as on the left of the above is nonnegative. By standard arguments involving test functions, we can see $\mathcal{A}_\mathsf{H}$ is indeed convex. Then, we show $\overline{F}_N(t, \cdot) \in \mathcal{A}_\mathsf{H}$ for all t and all N. In the proof of (3.3.12), we used [94, (3.27)] to compute $a \cdot \nabla(a \cdot \nabla\overline{F}_N)$. A slight modification of [94, (3.27)] gives

$$\begin{aligned}
&Na \cdot \nabla(b \cdot \nabla\overline{F}_N) \\
&= \mathbb{E}\left\langle \big(a \cdot x^\intercal x'\big)\big(b \cdot x^\intercal x'\big)\right\rangle - 2\mathbb{E}\left\langle \big(a \cdot x^\intercal x'\big)\big(b \cdot x^\intercal x''\big)\right\rangle + \mathbb{E}\left\langle a \cdot x^\intercal x'\right\rangle \left\langle b \cdot x^\intercal x'\right\rangle,
\end{aligned}$$

for $a, b \in \mathbb{S}^2$. By the definition of X in this model, under the Gibbs measure $\langle\cdot\rangle$, we can write $x = \mathsf{diag}(x_1, x_2)$ with $x_1, x_2 \in \mathbb{R}^N$. Replace a and b by e^{11} and e^{22} respectively in the above display to see $Ne^{11} \cdot \nabla(e^{22} \cdot \overline{F}_N)$ is given by

$$\begin{aligned}
&\mathbb{E}\left\langle \big(x_1 \cdot x_1'\big)\big(x_2 \cdot x_2'\big)\right\rangle - 2\mathbb{E}\left\langle \big(x_1 \cdot x_1'\big)\big(x_2 \cdot x_2''\big)\right\rangle + \mathbb{E}\left\langle x_1 \cdot x_1'\right\rangle \left\langle x_2 \cdot x_2'\right\rangle \\
&= \mathbb{E}\sum_{m,n=1}^N \Big(\langle x_{1,m}x_{2,n}\rangle^2 - 2\langle x_{1,m}x_{2,n}\rangle \langle x_{1,m}\rangle \langle x_{2,n}\rangle + \langle x_{1,m}\rangle \langle x_{2,n}\rangle^2 \Big) \geqslant 0.
\end{aligned}$$

This shows that the assumptions on $\mathcal{A}_\mathsf{H}$ in Theorem 3.2.1 are satisfied despite the fact that H is not convex in this case.

3.9. Fenchel–Moreau identity

The goal is to prove the following version of the Fenchel–Moreau identity on $\mathbb{S}^K_+$. More general versions on self-dual cones in possibly infinite dimensional Hilbert spaces can be seen in [38]. Here, for completeness, we prove this using arguments more specific to matrices. Recall the Fenchel transformation over $\mathbb{S}^K_+$ defined in (3.6.26), and the sense of nondecreasingness in (3.2.10).

Proposition 3.9.1 (Fenchel–Moreau identity)**.** *Let $u : \mathbb{S}^K_+ \to (-\infty, +\infty]$ be a function not identically equal to $+\infty$. Then, $u^{**} = u$ if and only if u is convex, l.s.c. (lower semi-continuous), and nondecreasing.*

It is easy to see that v^* is convex and l.s.c. for any function v. In addition by Lemma 3.3.3, we can see that v^* is also nondecreasing. Hence, to prove Proposition 3.9.1, it suffices to show the following.

Lemma 3.9.2. *If $u : \mathbb{S}^K_+ \to (-\infty, +\infty]$ is convex, l.s.c., nondecreasing and not identically $+\infty$, then $u^{**} = u$.*

The rest of this section is devoted to proving Lemma 3.9.2. Henceforth, we assume that u satisfies the condition imposed in this lemma.

3.9.1. Preliminaries

We introduce some notation and classical results. We extend u to $\mathbb{S}^K \cong \mathbb{R}^{K(K+1)/2}$ by setting the value outside $\mathbb{S}^K_+$ to be ∞. Denote by $\circledast$ the usual conjugate with the sup over $\mathbb{S}^K$. The extension of u gives $u^{\circledast} = u^*$. By the regular Fenchel-Moreau theorem, we have

$$u(x) = \sup_{y \in \mathbb{S}^K} \{y \cdot x - u^*(y)\}, \quad \forall x \in \mathbb{S}^K.$$

We want to show, whenever $x \in \mathbb{S}^K_+$, the sup above can be taken over $\mathbb{S}^K_+$.

Denote by $\Omega = \mathsf{dom}\, u = \{x \in \mathbb{S}^K : \ u(x) < +\infty\}$ the effective domain of u. For any $A \subseteq \mathbb{S}^K$, $\mathsf{int}\, A$, $\mathsf{cl}\, A$, $\mathsf{bd}\, A$ and conv stand for the interior, closure, boundary, and convex hull of A, respectively. For each $y \in \mathbb{S}^K$, we define the subdifferential of u at x by

$$\partial u(y) = \{z \in \mathbb{S}^K : \ u(y') \geqslant u(y) + z \cdot (y' - y), \ \forall y' \in \mathbb{S}^K\}. \tag{3.9.1}$$

The outer normal cone to Ω at $y \in \mathbb{S}^K$ is given by

$$\mathbf{n}(y) = \{z \in \mathbb{S}^K : \ z \cdot (y' - y) \leqslant 0, \ \forall y' \in \Omega\}. \tag{3.9.2}$$

Define

$$D = \{x \in \Omega : \ u \text{ is differentiable at } x\}.$$

For $a \in \mathbb{S}^K$ and $\nu \in \mathbb{R}$, we define the affine function $L_{a,\nu}$ by $L_{a,\nu}(x) = a \cdot x + \nu$.

We recall some useful lemmas, all of which are classical.

Lemma 3.9.3. *For a convex set A, if $y \in \mathsf{cl}\, A$ and $y' \in \mathsf{int}\, A$, then $\lambda y + (1-\lambda)y' \in \mathsf{int}\, A$ for all $\lambda \in [0, 1)$.*

Lemma 3.9.4. *Let $x \in \mathbb{S}^K_+$ and $y \in \Omega$. For every $\alpha \in (0, 1)$, set $x_\alpha = (1-\alpha)x + \alpha y$. Then $\lim_{\alpha \to 0} u(x_\alpha) = u(x)$.*

Lemma 3.9.5. *The set $\mathsf{int}\, \Omega \setminus D$ has Lebesgue measure zero.*

Lemma 3.9.6. *If $\mathsf{int}\, \Omega \neq \emptyset$, then*

$$\partial u(y) = \mathsf{cl}\left(\mathsf{conv}\, A(y)\right) + \mathbf{n}(y), \quad \forall y \in \Omega,$$

where $A(y)$ is the set of all limits of sequences $\left(\nabla u(y_n)\right)_{n=1}^{\infty}$ with $\lim_{n\to\infty} y_n = y$ and $y_n \in D$ for all n.

Lemma 3.9.7. *If $\partial u(y) \cap \mathbb{S}^K_+ \neq \emptyset$, then $u^{**}(y) = u(y)$.*

Lemma 3.9.8. *For every $x \in \mathbb{S}^K$, we have $u^{**}(x) = \sup L_{a,\nu}(x)$, where the supremum is taken over the set $\{L_{a,\nu} : \ a \in \mathbb{S}^K_+, \ \nu \in \mathbb{R}, \ L_{a,\nu} \leqslant u\}$.*

Lemma 3.9.3, 3.9.6, and 3.9.7 can be derived from [105, Theorem 6.1, 25.6, and 23.5], respectively. Lemma 3.9.4 is borrowed from [23, Proposition 9.14]. The density claim in

Lemma 3.9.5 follows from [105, Theorem 25.5]. The idea to verify the boundedness assertion can be seen in the proof of [74, Proposition 6.2.2 in Chapter D]. Lemma 3.9.8 can be verified using the definitions of u^{**} and $\sup L_{a,\nu}(\cdot)$.

In Section 3.9.2, we prove Lemma 3.9.2 under an additional assumption that $\mathsf{int}\,\Omega \neq \emptyset$. We consider the case $\mathsf{int}\,\Omega = \emptyset$ in Section 3.9.3.

3.9.2. Case 1: nonempty interior

Assuming $\mathsf{int}\,\Omega \neq \emptyset$, we want to show that the identity $u^{**} = u$ holds for all $x \in \mathbb{S}_+^K$. We proceed in steps and show this identity holds on $\mathsf{cl}\,\Omega$ and then on $\mathbb{S}_+^K$.

Analysis on $\mathsf{cl}\,\Omega$

At every $x \in D$, due to the nondecreasingness of u, we have $a \cdot \nabla u(x) \geqslant 0$ for all $a \in \mathbb{S}_+^K$. Then, Lemma 3.3.3 implies $\nabla u(x) \in \mathbb{S}_+^K$ at every $x \in D$. By Lemma 3.9.7, we conclude $u^{**}(x) = u(x)$ for all $x \in D$.

Now for each $x \in \mathsf{cl}\,\Omega$, since $\mathsf{int}\,\Omega$ is convex and nonempty, by Lemma 3.9.5 and an argument using Fubini's theorem, we can see that there is $x' \in \Omega$ such that $x_\alpha = (1-\alpha)x + \alpha x' \in D$ for every $\alpha \in (0,1)$. Since both u^{**} and u are convex and l.s.c., Lemma 3.9.4 implies that $u^{**}(x) = u(x)$ for all $x \in \mathsf{cl}\,\Omega$.

Analysis on $\mathbb{S}_+^K$

Let $x \in \mathbb{S}_+^K \setminus \mathsf{cl}\,\Omega$. Hence, we have $u(x) = \infty$. Define

$$\lambda' = \sup\{\lambda \in [0, +\infty) : \ \lambda x \in \mathsf{cl}\,\Omega\}.$$

By $x \notin \mathsf{cl}\,\Omega$, $0 \in \mathsf{cl}\,\Omega$ and the convexity of $\mathsf{cl}\,\Omega$, we must have

$$\lambda' < 1. \tag{3.9.3}$$

Set $x' = \lambda' x$. It is clear that $x' \in \mathsf{cl}\,\Omega$ and satisfies (3.9.4). The definition of λ' also ensures $x' \notin \mathsf{int}\,\Omega$. There are two cases, either $x' \in \Omega$ or not.

When $x' \notin \Omega$, by $u^{**} = u$ on $\mathsf{cl}\,\Omega$ and Lemma 3.9.8, there is a sequence of affine functions $(L_{a_n,\nu_n})_{n=1}^{\infty}$ such that $a_n \in \mathbb{S}_+^K$, $u \geqslant L_{a_n,\nu_n}$ for all n and $u(x') = \lim_{n\to\infty} L_{a_n,\nu_n}(x') = \infty$. By the definition of x' and (3.9.3), we can see

$$L_{a_n,\nu_n}(x) = L_{a_n,\nu_n}(x') + (1-\lambda')a_n \cdot x \geqslant L_{a_n,\nu_n}(x').$$

Hence, we also have $u(x) = \lim_{n\to\infty} L_{a_n,\nu_n}(x) = \infty$. This together with Lemma 3.9.8 shows $u^{**} = u$ at this x.

Now, we turn to the case where $x' \in \Omega$. We need the next lemma.

Lemma 3.9.9. *For every $x \in \mathsf{bd}\,\Omega$ satisfying*

$$\lambda x \notin \mathsf{cl} \quad \forall \lambda > 1, \tag{3.9.4}$$

we have $\big(\mathbf{n}(x) \cap \mathbb{S}_+^K\big) \setminus \{0\} \neq \emptyset$.

Since x' satisfies (3.9.4), this lemma implies that there is $z \in \mathbf{n}(x') \cap \mathbb{S}_+^K$ with $z \neq 0$. The definition (3.9.2) yields

$$z \cdot (y - x') \leqslant 0, \quad \forall y \in \Omega. \tag{3.9.5}$$

Since we clearly have $0 \in \Omega$, we have $z \cdot x' \geqslant 0$. We claim that actually

$$z \cdot x' > 0. \tag{3.9.6}$$

Otherwise, we have $z \cdot x' = 0$. Since there is $x_0 \in \mathsf{int}\,\Omega \subseteq \mathbb{S}_{++}^K$, we can see that there is $\varepsilon > 0$ sufficiently small such that $x_0 - \varepsilon z \in \mathbb{S}_+^K$. The nondecreasingness of u yields $\varepsilon z \in \Omega$. Replacing y by εz in (3.9.5) and using $z \cdot x' = 0$, we have $\varepsilon |z|^2 \leqslant 0$, contradicting $z \neq 0$. Hence, we have (3.9.6).

By $u^{**} = u$ on $\mathsf{cl}\,\Omega$ and Lemma 3.9.8, we can find an affine function $L_{a,\nu}$ with $a \in \mathbb{S}_+^K$ such that $u \geqslant L_{a,\nu}$. Now, for each $\rho \geqslant 0$, define

$$\mathcal{L}_\rho = L_{a+\rho z,\ \nu-\rho z\cdot x'}.$$

Due to (3.9.5), we can see

$$\mathcal{L}_\rho(y) = L_{a,\nu}(y) + \rho z \cdot (y - x') \leqslant L_{a,\nu}(y) \leqslant u(y), \quad \forall y \in \Omega.$$

Since $u = \infty$ outside Ω, we thus have $\mathcal{L}_\rho \leqslant u$. On the other hand, we can compute

$$\mathcal{L}_\rho(x) = L_{a,\nu}(x) + \rho z \cdot (x - x') = L_{a,\nu}(x) + \rho(\lambda'^{-1} - 1) z \cdot x'.$$

By (3.9.3) and (3.9.6), we have $\lim_{\rho\to\infty} \mathcal{L}_\rho(x) = \infty = u(x)$. By Lemma 3.9.8, this shows that $u^{**} = u$ holds at $x \in \mathbb{S}^K_+ \setminus \mathsf{cl}\,\Omega$. Together with previous results, we conclude that $u^{**} = u$ holds on $\mathbb{S}^K_+$ under the assumption $\mathsf{int}\,\Omega \neq \emptyset$.

To complete the proof of Lemma 3.9.2 under the additional assumption $\mathsf{int}\,\Omega \neq \emptyset$, it remains to prove Lemma 3.9.9.

Proof of Lemma 3.9.9. Fix $x \in \Omega \setminus \mathsf{int}\,\Omega$ satisfying (3.9.4).

Step 1. We show that for every Euclidean ball $B \subseteq \mathbb{S}^K$ centered at x, there is $\overline{x} \in \mathbb{S}^K_{++} \cap \mathsf{bd}\,\Omega$. By (3.9.4), there is some $\lambda > 1$ such that $x' = \lambda x \in B \setminus \mathsf{cl}\,\Omega$. By $\mathsf{int}\,\Omega \neq \emptyset$ and Lemma 3.9.3, there is $x'' \in B \cap \mathsf{int}\,\Omega \subseteq \mathbb{S}^K_{++}$. For $\rho \in [0,1]$, set

$$x_\rho = \rho x' + (1-\rho) x'' \in B.$$

Set $\rho_0 = \sup\{\rho \in [0,1] : \ x_\rho \in \mathsf{int}\,\Omega\}$. We can see x_{ρ_0} lies in the closure but not the interior of Ω, and thus $x_{\rho_0} \in B \cap \mathsf{bd}\,\Omega$. In addition, since $x' \notin \mathsf{cl}\,\Omega$, we must have $\rho_0 < 1$ and hence $x_{\rho_0} \in \mathbb{S}^K_{++}$ due to $x'' \in \mathbb{S}^K_{++}$. We conclude that $x_{\rho_0} \in B \cap \mathbb{S}^K_{++} \cap \mathsf{bd}\,\Omega$ is the point $\overline{x}$ we want.

Step 2. By the construction above, we can find a sequence $(x_n)_{n=1}^{\infty}$ such that $x_n \in \mathbb{S}^K_{++} \cap \mathsf{bd}\,\Omega$ and $\lim_{n\to\infty} x_n = x$. We want to show $\mathbf{n}(x_n) \subseteq \mathbb{S}^K_+$ using the following lemma.

Lemma 3.9.10. *If $y \in \mathbb{S}^K_{++} \cap \mathsf{bd}\,\Omega$, then $\mathbf{n}(y) \subseteq \mathbb{S}^K_+$.*

Proof. Since $y \in \mathsf{bd}\,\Omega$, using $\mathsf{int}\,\Omega \neq \emptyset$ and Lemma 3.9.3, we can find $y_\varepsilon \in \mathsf{int}\,\Omega$ such that $|y_\varepsilon - y| < \varepsilon$ for each $\varepsilon > 0$. By this and $y \in \mathbb{S}^K_{++}$, there are $\varepsilon_0, \delta_0 > 0$ such that

$$y_\varepsilon - \delta_0 I \in \mathbb{S}^K_+, \quad \varepsilon \in (0, \varepsilon_0).$$

This further implies that there is $\delta > 0$ such that

$$y_\varepsilon - a \in \mathbb{S}^K_+, \quad \forall \varepsilon \in (0, \varepsilon_0) \quad , \forall a \in \mathbb{S}^K_+ \text{ satisfying } |a| \leqslant \delta.$$

Since u is nondecreasing and $y_\varepsilon \in \Omega$, we have $y_\varepsilon - a \in \Omega$ for any a described above. Let $z \in \mathbf{n}(y)$. The definition (3.9.2) yields $z \cdot (y_\varepsilon - a - y) \leqslant 0$ and thus

$$z \cdot a \geqslant -|z|\varepsilon.$$

Sending $\varepsilon \to 0$ and varying a, we conclude using Lemma 3.3.3 that $z \in \mathbb{S}^K_+$.

□

This lemma immediately implies that $\mathbf{n}(x_n) \subseteq \mathbb{S}^K_+$. For each n, pick $z_n \in \mathbf{n}(x_n) \cap \mathbb{S}^K_+$ with $|z_n| = 1$. By extracting a subsequence, we may assume $\lim_{n\to\infty} z_n = z$ for some $z \in \mathbb{S}^K_+$ satisfying $|z| = 1$. Since $z_n \in \mathbf{n}(x_n)$, we have

$$z_n \cdot (y - x_n) \leqslant 0, \quad \forall y \in \mathbb{S}^K.$$

Set $n \to \infty$, recall that $\lim_{n\to\infty} x_n = x$, and we obtain $z \cdot (y - x) \leqslant 0$ for all $y \in \mathbb{S}^K$. This proves Lemma 3.9.9.

□

3.9.3. Case 2: empty interior

To complete the proof of Lemma 3.9.2, let us investigate the situation where $\operatorname{int}\Omega = \emptyset$. The case $\Omega = \{0\}$ is easy to handle. So, we assume $\operatorname{int}\Omega = \emptyset$ and $\Omega \setminus \{0\} \neq \emptyset$. Set

$$J = \max\{\mathsf{rank}(x) : \ x \in \Omega\}, \tag{3.9.7}$$

where $\mathsf{rank}(x)$ is the rank of the matrix x. By $\Omega \setminus \{0\} \neq \emptyset$, we have $J \geqslant 1$.

Step 1. We show $J < K$. Otherwise, there is $x \in \Omega$ with $\mathsf{rank}(x) = K$. Hence, we have $x \in \mathbb{S}^K_{++}$. Therefore, there is $\delta > 0$ such that $x - y \in \mathbb{S}^K_{++}$, for all $y \in \mathbb{S}^K_+$ with $|y| \leqslant \delta$. This contradicts the assumption that $\operatorname{int}\Omega = \emptyset$.

For each $n \in \mathbb{N}$, we denote the $n \times n$ zero matrix by $\mathsf{0}_n$. Fix any $x \in \Omega$ with $\mathsf{rank}(x) = J$. Without loss of generality, by an orthogonal transformation, we may assume $x = \mathsf{diag}(\lambda_1, \lambda_2, \ldots, \lambda_J, \mathsf{0}_{K-J})$, where $\lambda_j > 0$ for all $1 \leqslant j \leqslant J$.

Step 2. We show that for every $y \in \Omega$, there is $y^\circ \in \mathbb{S}^J_+$ such that

$$y = \mathsf{diag}(y^\circ, \mathsf{0}_{K-J}). \tag{3.9.8}$$

Otherwise, there is $y \in \Omega$ with $y_{ij} \neq 0$ for some $i > J$ or $j > J$. Since $y \in \mathbb{S}^K_+$ is positive semidefinite, we must have $y_{ii} > 0$ for some $i > J$. By reordering, we assume $i = J + 1$. Note that this reordering preserves x. We want to show $\mathsf{rank}(x + y) > J$. Let $\widehat{y} = (y_{ij})_{1 \leqslant i,j \leqslant J+1} \in \mathbb{S}^{J+1}_+$ be a portion of y, and $\widehat{x}$ be similarly defined. It suffices to show $\mathsf{rank}(\widehat{x} + \widehat{y}) = J + 1$. We further reduce this to verifying $\widehat{x} + \widehat{y} \in \mathbb{S}^{J+1}_{++}$ and thus showing

$$v^\mathsf{T}(\widehat{x} + \widehat{y})v > 0 \tag{3.9.9}$$

for all $v \in \mathbb{R}^{J+1} \setminus \{0\}$.

First, we consider the case where $v_j \neq 0$ for some $1 \leqslant j \leqslant J$. Since $\widehat{x} = \mathsf{diag}(\lambda_1, \dots, \lambda_J, 0)$ and each λ_j is positive, we have $v^\intercal \widehat{x} v = \sum_{j=1}^J \lambda_j v_j^2 > 0$, verifying (3.9.9). Now, suppose $v_j = 0$ for all $1 \leqslant j \leqslant J$. Due to $v \in \mathbb{R}^{J+1} \setminus \{0\}$, we must have $v_{J+1} \neq 0$. Since $y_{J+1,J+1} > 0$, we obtain $v^\intercal \widehat{y} v = y_{J+1,J+1} v_{J+1}^2 > 0$. In conclusion, (3.9.9) holds.

Therefore, $\mathsf{rank}(\widehat{x} + \widehat{y}) = J + 1$, and thus $\mathsf{rank}(x + y) > J$. By the convexity of Ω, we see that $\frac{1}{2}(x + y) \in \Omega$. But this contradicts (3.9.7). Hence, by contradiction, y is of the form (3.9.8) for all $y \in \Omega$.

Step 3. We apply the result in the previous section. Define

$$\mathcal{C} = \{\mathsf{diag}(y^\circ, \mathsf{0}_{K-J}) : \ y^\circ \in \mathbb{S}_+^J\} \subseteq \mathbb{S}_+^K.$$

By the result from Step 2, we have $\Omega \subseteq \mathcal{C}$. Identifying $\mathcal{C}$ with $\mathbb{S}_+^J$, we can view u as a map from $\mathbb{S}_+^J$ to $(-\infty, \infty]$. By (3.9.7), the interior of Ω relative to $\mathbb{S}_+^J$ is nonempty. Hence, applying the result for case with nonempty interior, comparing with $u^{**} = u$, we have

$$u(x) = \sup_{z \in \mathcal{C}} \{z \cdot x - u^*(z)\}, \quad \forall x \in \mathcal{C}. \tag{3.9.10}$$

Since $u \geqslant u^{**}$, we have $u^{**} = u$ on $\mathcal{C}$.

Step 4. To complete the proof, we show that $u^{**} = u$ holds on $\mathbb{S}_+^K \setminus \mathcal{C}$. Let us set $z = \mathsf{diag}\{\mathsf{0}_J, I_{K-J}\}$ where I_{K-J} is the $(K - J) \times (K - J)$ identity matrix. Fix any $x \in \mathbb{S}_+^K \setminus \mathcal{C}$. Due to $x \notin \mathcal{C}$, there is some $i > J$ or $j > J$ such that $x_{ij} \neq 0$. Since x is positive semidefinite, we must have $x_{ii} > 0$ for some $i > J$. Therefore, we get

$$z \cdot x > 0. \tag{3.9.11}$$

By (3.9.10), there is an affine function $L_{a,\nu}$ with $a \in \mathcal{C} \subseteq \mathbb{S}_+^K$ such that $u \geqslant L_{a,\nu}$ on $\mathcal{C}$. Now, for every $\rho \geqslant 0$, we define

$$\mathcal{L}_\rho = L_{a+\rho z,\nu}.$$

By the definition of z, we can compute

$$\mathcal{L}_\rho(y) = L_{a,\nu}(y) + z \cdot y = L_{a,\nu}(y) \leqslant u(y), \quad \forall y \in \mathcal{C}.$$

Since $u = \infty$ outside $\mathcal{C}$, we then get $\mathcal{L}_\rho \leqslant u$. On the other hand, (3.9.11) implies that

$$\mathcal{L}_\rho(x) = L_{a,\nu}(y) + \rho z \cdot y$$

converges to ∞ as $\rho \to \infty$. Then, Lemma 3.9.7 implies $u^{**} = u$ at $x \in \mathbb{S}^K_+ \setminus \mathcal{C}$.

3.10. Concentration in the special case

In this appendix, we prove a concentration result assuming X has i.i.d. and bounded entries. The following lemma works for any fixed interaction matrix $A \in \mathbb{R}^{K^p \times L}$ in (3.1.1). Recall the definition of $\mathcal{K}_{M,N}$ in (3.2.8).

Lemma 3.10.1. *Assume that X consists of i.i.d. entries and $|X_{ij}| \leqslant 1$ for all i and j. Then, there is $C > 0$ such that the following holds for all $M \geqslant 1$ and $n \in \mathbb{N}$,*

$$\mathcal{K}_{M,N} \leqslant CN^{-\frac{1}{2}}\big(M + \sqrt{\log N}\big).$$

3.10.1. Proof of Lemma 3.10.1

The plan is to first obtain an estimate of $\mathbb{E}e^{\lambda^2 N|F_N - \overline{F}_N|^2}$ for small $\lambda > 0$ pointwise at each $(t,h) \in [0,M] \times \mathbb{S}^K_{+,M}$. Then, we use an ε-net argument to bound $\mathbb{E}\sup_{(t,h)\in[0,M]\times\mathbb{S}^K_{+,M}} e^{\lambda^2 N|F_N - \overline{F}_N|}$. The desired result follows from Jensen's inequality.

Pointwise estimate

Let $(t,h) \in [0,M] \times \mathbb{S}^K_{+,M}$. Denote by $G = (W,Z)$ the Gaussian vector consisting of all Gaussian random variables in F_N. We also write $\mathbb{E}_G$, $\mathbb{E}_X$ as the expectation integrating over G, X, respectively. Let $\lambda > 0$ be chosen later. Using the Cauchy–Schwarz inequality, we have

$$\begin{aligned}\mathbb{E}e^{\lambda|F_N-\overline{F}_N|} &\leqslant \mathbb{E}\Big(e^{\lambda|F_N-\mathbb{E}_X F_N|}e^{\lambda|\mathbb{E}_X F_N-\mathbb{E}_{X,G}F_N|}\Big)\\ &= \Big(\mathbb{E}e^{2\lambda|F_N-\mathbb{E}_X F_N|}\Big)^{\frac{1}{2}}\Big(\mathbb{E}e^{2\lambda|\mathbb{E}_X F_N-\mathbb{E}_{X,G}F_N|}\Big)^{\frac{1}{2}}.\end{aligned} \tag{3.10.1}$$

To treat the last term, we will use the Gaussian concentration inequality. Let us use the multi-index notation (3.2.15). By (3.2.3) and (3.2.4), we can compute

$$\partial_{W_{\mathbf{i}}}F_N = \frac{1}{N}\sqrt{\frac{2t}{N^{p-1}}}\left\langle \widetilde{x}_{\mathbf{i}}\right\rangle, \quad \partial_{Z_{ij}}F_N = \frac{1}{N}\sum_{k=1}^K \big(\sqrt{2h}\big)_{kj}\left\langle x_{ik}\right\rangle.$$

Here $\widetilde{x}$ is defined in (3.2.2). Therefore, by (3.2.1), we have

$$\begin{aligned}|\nabla_G F_N|^2 &= \sum_{\mathbf{i}}|\partial_{W_{\mathbf{i}}}F_N|^2 + \sum_{i=1}^N\sum_{j=1}^K|\partial_{Z_{ij}}F_N|^2\\ &= \frac{2t}{N^{p+1}}\left\langle \widetilde{x}\cdot\widetilde{x}'\right\rangle + \frac{2}{N^2}h\cdot\left\langle x^\intercal x'\right\rangle \leqslant CMN^{-1}.\end{aligned}$$

Invoking [27, Theorem 5.5], we obtain

$$\mathbb{E}_G e^{\lambda|\mathbb{E}_X F_N-\mathbb{E}_{X,G}F_N|} \leqslant e^{C\lambda^2 MN^{-1}}. \tag{3.10.2}$$

Then, we treat the first two terms in (3.10.1). Let us first compute $\partial_{X_{ij}}F_N$. By (3.2.4), we can compute

$$\partial_{X_{ij}}F_N = \frac{1}{N}\left\langle \frac{2t}{N^p}\partial_{X_{ij}}\big(\widetilde{x}\cdot\widetilde{X}\big) + 2\partial_{X_{ij}}\Big(h\cdot\big(x^\intercal X\big)\Big)\right\rangle.$$

Due to the boundedness assumption $|X_{i,\cdot}| \leqslant \sqrt{K}$ (and thus $|x_{i,\cdot}| \leqslant \sqrt{K}$ under the distribution $\langle \cdot \rangle$), we can verify

$$\left|\nabla_{X_{i,\cdot}} F_N\right| \leqslant CMN^{-1}.$$

Using the boundedness again and [27, Theorem 6.2] (see the penultimate display in its proof), we obtain

$$\mathbb{E}_X e^{\lambda|F_N - \mathbb{E}_X F_N|} \leqslant Ce^{C\lambda^2 M^2 N^{-1}}. \tag{3.10.3}$$

In conclusion, (3.10.1), (3.10.2) and (3.10.3), with λ replaced by $\lambda\sqrt{N}$, yield

$$\mathbb{E}e^{\lambda\sqrt{N}|F_N - \overline{F}_N|} \leqslant Ce^{C\lambda^2 M^2}.$$

Then, [116, Proposition 2.5.2] implies that, for λ sufficiently small,

$$\mathbb{E}e^{\lambda^2 N|F_N - \overline{F}_N|^2} \leqslant Ce^{C\lambda^2 M^2}. \tag{3.10.4}$$

Application of an ε-net argument

The goal is to upgrade (3.10.4) to a bound on $\mathbb{E}\sup_{(t,h)\in[0,M]\times\mathbb{S}^K_{+,M}} e^{\lambda^2 N|F_N - \overline{F}_N|^2}$. The estimates (3.3.8) and (3.3.9) imply that, for $|t-t'| + |h-h'| \leqslant 1$,

$$|F_N(t,h) - F_N(t',h')| \leqslant C\Big(1 + N^{-\frac{1}{2}}\big(\|WA^\intercal\| + |Z|\big)\Big)\big(|t-t'|^{\frac{1}{2}} + |h-h'|^{\frac{1}{2}}\big).$$

For $\varepsilon \in (0,1]$, viewing $\mathbb{S}^K_{+,M}$ as a subset of $\mathbb{R}^{K(K+1)/2}$, we introduce the ε-net

$$A_\varepsilon = \{\varepsilon, 2\varepsilon, 3\varepsilon \dots\}^{1+K(K+1)/2} \cap \Big([0,M] \times \mathbb{S}^K_{+,M}\Big).$$

Hence, for λ small, we have

$$
\begin{aligned}
&\mathbb{E} \sup_{(t,h)\in[0,M]\times\mathbb{S}^K_{+,M}} e^{\lambda^2 N|F_N-\overline{F}_N|^2} \\
&\leqslant \mathbb{E}\exp\Big(C\lambda^2\varepsilon\big(\sqrt{N}+\|WA^\intercal\|+|Z|\big)^2\Big)\sup_{(t,h)\in A_\varepsilon} e^{\lambda^2 N|F_N-\overline{F}_N|^2} \\
&\leqslant \bigg(\mathbb{E}\exp\Big(C\lambda^2\varepsilon\big(\sqrt{N}+\|WA^\intercal\|+|Z|\big)^2\Big)\bigg)^{\frac{1}{2}}\bigg(\mathbb{E}\sup_{(t,h)\in A_\varepsilon} e^{2\lambda^2 N|F_N-\overline{F}_N|^2}\bigg)^{\frac{1}{2}}
\end{aligned}
\tag{3.10.5}
$$

where we used the Cauchy–Schwarz inequality in the second inequality. Since $|A_\varepsilon| \leqslant (M/\varepsilon)^{1+K(K+1)/2}$, using the union bound and (3.10.4), we have,

$$
\bigg(\mathbb{E}\sup_{(t,h)\in A_\varepsilon} e^{2\lambda^2 N|F_N-\overline{F}_N|^2}\bigg)^{\frac{1}{2}} \leqslant C(M/\varepsilon)^C e^{C\lambda^2 M^2}, \quad \lambda\in\mathbb{R}. \tag{3.10.6}
$$

Set $\varepsilon = C^{-1}N^{-1}$ in (3.10.5) with C therein, and use (3.10.6) to see

$$
\begin{aligned}
&\mathbb{E} \sup_{(t,h)\in[0,M]\times\mathbb{S}^K_{+,M}} e^{\lambda^2 N|F_N-\overline{F}_N|^2} \\
&\leqslant C(MN)^C e^{C\lambda^2 M^2}\bigg[\mathbb{E}\exp\Big(\lambda^2\big(1+N^{-\frac{1}{2}}(\|WA^\intercal\|+|Z|)\big)^2\Big)\bigg]^{\frac{1}{2}}.
\end{aligned}
$$

We claim that, for small $\lambda > 0$,

$$
\mathbb{E}\exp\Big(\lambda^2\big(1+N^{-\frac{1}{2}}(\|WA^\intercal\|+|Z|)\big)^2\Big) \leqslant C. \tag{3.10.7}
$$

This immediately gives

$$
\mathbb{E} \sup_{(t,h)\in[0,M]\times\mathbb{S}^K_{+,M}} e^{\lambda^2 n|F_N-\overline{F}_N|^2} \leqslant C(MN)^C e^{C\lambda^2 M^2}.
$$

Finally, using Jensen's inequality, we conclude that

$$\begin{aligned}\mathbb{E}\sup_{(t,h)\in[0,M]\times\mathbb{S}^K_{+,M}}|F_N-\overline{F}_N|^2 &\leqslant \lambda^{-2}N^{-1}\log\left(\mathbb{E}\sup_{(t,h)\in[0,M]\times\mathbb{S}^K_{+,M}}e^{\lambda^2N|F_N-\overline{F}_N|^2}\right)\\ &\leqslant CN^{-1}(M^2+\log N),\end{aligned}$$

as desired. The proof will be complete once (3.10.7) is verified.

Proof of (3.10.7)

We want to bound exponential moments of $\|WA^\intercal\|^2$ and $|Z|^2$. Using the fact that Z is standard Gaussian in $\mathbb{R}^N$, we have, for λ small,

$$\mathbb{E}e^{\lambda^2N^{-1}|Z|^2}\leqslant C. \tag{3.10.8}$$

Now, we turn to bound $\mathbb{E}e^{\lambda^2N^{-1}\|WA^\intercal\|^2}$. For each $\varepsilon>0$, there is a finite set $B\subseteq\mathbb{S}^{NK-1}$ such that for each $y\in\mathbb{S}^{NK-1}$ there is $z\in B$ satisfying $|y-z|\leqslant\varepsilon$. In addition, the size of B is bounded by a^{NK} for some constant $a>0$ depending only on ε. The construction of B is classical and can be seen, for instance, in [116, Corollary 4.2.13]. Using the property of B, we can see that for each $(y_1,\, y_2,\, \dots\, y_p)\in(\mathbb{S}^{NK-1})^p$ there is $(z_1,\, z_2,\, \dots\, z_p)\in B^p$ such that

$$\left|(WA^\intercal)\cdot(y_1\otimes y_2\otimes\cdots\otimes y_p)-(WA^\intercal)\cdot(z_1\otimes z_2\otimes\cdots\otimes z_p)\right|\leqslant p\varepsilon\|WA^\intercal\|.$$

By this and fixing $\varepsilon=\frac{1}{2p}$, from the definition (3.3.7), we obtain

$$\|WA^\intercal\|\leqslant 2\sup_{(z_1,\,z_2,\,\dots\,z_p)\in B^p}(WA^\intercal)\cdot(z_1\otimes z_2\otimes\cdots\otimes z_p).$$

Note that $(WA^\intercal)\cdot(z_1\otimes z_2\otimes\cdots\otimes z_p)$ is a centered Gaussian with variance bounded by a constant C depending only on A. Therefore, there is $\gamma>0$ such that

$$\mathbb{P}\Big\{(WA^\intercal)\cdot(z_1\otimes z_2\otimes\cdots\otimes z_p)\geqslant t\Big\}\leqslant 2e^{-\gamma t^2}.$$

Combine the above two displays and apply the union bound to see

$$\mathbb{P}\Big\{e^{\lambda^2 N^{-1}\|WA^\intercal\|^2} \geqslant t\Big\} \leqslant 2\bigg(\frac{a^{pK}}{t^c}\bigg)^N$$

for some constant $c > 0$ that absorbs λ and γ. Writing $b = a^{\frac{pK}{c}}$, we have, for N large,

$$\mathbb{E}e^{\lambda^2 N^{-1}\|WA^\intercal\|^2} = \int_0^\infty \mathbb{P}\{e^{\lambda^2 N^{-1}\|WA^\intercal\|^2} \geqslant t\}\mathrm{d}t \leqslant b + \int_b^\infty 2\bigg(\frac{b}{t}\bigg)^{cN}\mathrm{d}t = b + \frac{2b}{cN-1},$$

which is bounded uniformly for large N. This and (3.10.8) imply (3.10.7).

CHAPTER 4

Statistical inference of finite-rank tensors

This chapter is essentially borrowed from [37], joint with Jean-Christophe Mourrat and Hong-Bin Chen.

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.

4.1. Introduction

4.1.1. 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

$$Y := \sqrt{\frac{2t}{N^{\mathsf{p}-1}}} X^{\otimes \mathsf{p}} A + W \ \in \mathbb{R}^{N^{\mathsf{p}} \times L}, \tag{4.1.1}$$

where $\otimes$ denotes the tensor product of matrices, and $W \in \mathbb{R}^{N^{\mathsf{p}} \times L}$, independent of the randomness of X, 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

$$|X| \leqslant \sqrt{NK}. \tag{4.1.2}$$

For instance, (4.1.2) is satisfied if every entry of X has its absolute value bounded by 1. 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 [10].

In order to analyze this model, we start by enriching the system by adding an additional observation $\overline{Y} = X\sqrt{2h} + Z$ for $h\in\mathbb{S}_+^K$, where $\mathbb{S}_+^K$ is the set of $K\times K$ symmetric positive semi-definite matrices, and $Z\in\mathbb{R}^{N\times K}$, independent of all other sources of randomness previously introduced, consists of i.i.d. standard Gaussian entries. Then, the law of X conditioned on observing Y and $\overline{Y}$ is a Gibbs measure proportional to $e^{H_N(t,h,x)}\mathrm{d}P_N^X(x)$ with Hamiltonian

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

The corresponding free energy is

$$F_N(t,h) := \frac{1}{N}\log\int_{\mathbb{R}^{N\times K}} e^{H_N(t,h,x)}\,\mathrm{d}P_N^X(x). \tag{4.1.3}$$

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

$\mathsf{H} : \mathbb{S}^K_+ \to \mathbb{R}$ be the mapping such that, for every $q \in \mathbb{S}^K_+$,

$$\mathsf{H}(q) := (AA^\intercal) \cdot q^{\otimes \mathsf{p}}. \tag{4.1.4}$$

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 4.1.1. *In addition to* (4.1.2)*, suppose that*

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

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

$$\lim_{N\to\infty} \overline{F}_N(t,h) = \sup_{h''\in\mathbb{S}^K_+} \inf_{h'\in\mathbb{S}^K_+} \left\{h'' \cdot (h-h') + \psi(h') + t\mathsf{H}(h'')\right\}. \tag{4.1.5}$$

Remark 4.1.2. 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 4.2.1).

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^X_N 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 [39, 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 [39]). 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 4.6.3.

4.1.2. Related works

Many inference models can be viewed as special cases of (4.1.1). Indeed, one could argue that essentially any "fully-connected" inference problem will have the form of (4.1.1) 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 [10, 82, 12, 95, 94], the spiked Wishart model [90, 14, 79, 86, 36], the stochastic block model (or community detection problem) [82, 87, 104], the inference of second order matrix tensor products [103], and the inference of higher order vector tensor products [83, 12, 95]. The model closest to (4.1.1) is the inference of finite-rank even-order tensor products studied in [85]. The case of tensors of odd order was left open there, see [85, Section 7]. In Section 4.5.2, 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 [39].

Many of the results mentioned above were obtained by the powerful method of adaptive interpolation introduced in [12, 13] and refined in subsequent works. In [103], 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 ([103, Proposition 4]) has a similar counterpart that plays a crucial role in this work (Lemma 4.2.2 and Proposition 4.4.7).

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 [95, 94], and can also inform the analysis of spin glass models [92, 98, 96, 93]; related considerations also appeared in the physics literature [69, 71, 22, 21].

The setting of the present paper is identical to that of [39], 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 [39], 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 $[0,\infty)\times\mathbb{S}^K_+$, 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 4.2, 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 4.3, 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 4.4. Lastly, Section 4.5 contains the proof of Theorem 4.1.1 and an application to the model (4.5.7).

Acknowledgements

We would like to warmly thank Stefano Bianchini for providing us with the idea for the proof of Proposition 4.6.2. JCM was partially supported by the NSF grant DMS-1954357.

4.2. Properties of the free energy

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

For any measurable $g:\mathbb{R}^{N\times K}\to\mathbb{R}^m$ for some $m\in\mathbb{N}$, we denote by $\langle g(x)\rangle$ the expectation of g, coordinatewise, with respect to the Gibbs measure proportional to $e^{H_N(t,h,x)}\,\mathrm{d}P^X_N(x)$, 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 T.

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

Using the expression (4.1.3), we can compute that

$$\partial_t \overline{F}_N = \frac{1}{N^{\mathsf{p}}} \mathbb{E} \left\langle x^{\otimes \mathsf{p}} A \cdot x'^{\otimes \mathsf{p}} A \right\rangle = \frac{1}{N^{\mathsf{p}}} \mathbb{E} \left[\left\langle x^{\otimes \mathsf{p}} A \right\rangle \cdot \left\langle x^{\otimes \mathsf{p}} A \right\rangle \right], \tag{4.2.1}$$

$$\nabla \overline{F}_N = \frac{1}{N} \mathbb{E} \left\langle x^{\mathsf{T}} x' \right\rangle = \frac{1}{N} \mathbb{E} \left[\langle x \rangle^{\mathsf{T}} \langle x \rangle \right]. \tag{4.2.2}$$

This computation involves the Nishimori identity, the Gaussian integration by parts, and the independence of replicas with respect to the Gibbs measure. For details, we refer to [39, (3.5)-(3.6)]. Recalling the definition of H in (4.1.4), we obtain that $\overline{F}_N$ satisfies

$$\partial_t \overline{F}_N - \mathsf{H}\big(\nabla \overline{F}_N\big) = \frac{1}{N^{\mathsf{p}}} \bigg(\mathbb{E}\big\langle \mathsf{H}(x^{\mathsf{T}} x') \big\rangle - \mathsf{H}\big(\mathbb{E}\langle x^{\mathsf{T}} x' \rangle \big) \bigg),$$

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

$$\partial_t f - \mathsf{H}(\nabla f) = 0 \quad \text{in } [0, \infty) \times \mathbb{S}^K_+. \tag{4.2.3}$$

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 (4.2.3); and then that this solution admits the variational representation appearing on the right side of (4.1.5).

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

4.2.1. Derivatives of free energy

We record basic results on the derivatives of $\overline{F}_N$.

Lemma 4.2.1. *For each $N \in \mathbb{N}$, the function $\overline{F}_N$ is C^1 and the following holds:*

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

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

Proof. It follows from (4.2.1) and (4.2.2), along with the assumption (4.1.2). □

The first display in Lemma 4.2.1 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 $\mathbb{S}^K$ and on $\mathbb{R} \times \mathbb{S}^K$, we declare

$$h_1 \leqslant h_2 \iff h_2 - h_1 \in \mathbb{S}^K_+; \tag{4.2.4}$$

$$(t_1, h_1) \leqslant (t_2, h_2) \iff (t_2, h_2) - (t_1, h_1) \in [0,\infty) \times \mathbb{S}^K_+. \tag{4.2.5}$$

As a consequence of Lemma 4.2.1 and the mean value theorem, we have that

$$\overline{F}_N \text{ is nondecreasing}, \quad \forall N \tag{4.2.6}$$

in the sense given in (4.2.5).

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

Lemma 4.2.2. *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. For $k = 1, 2$, denoting $\langle\,\cdot\,\rangle$ evaluated at (t_k, h_k) by $\langle\,\cdot\,\rangle_k$, we have

$$\langle g(x)\rangle_k = \mathbb{E}[g(X)\,|\,Y_k] \tag{4.2.7}$$

for any measurable function g satisfying $\mathbb{E}|g(X)| < \infty$. For any matrix y, we write $\mathbf{c}(y) := y^{\intercal}y$. Note that $\mathbf{c}(X^{\otimes \mathsf{p}}A) \in \mathbb{R}^{L\times L}$ and $\mathbf{c}(X) \in \mathbb{R}^{K\times K}$. Then, we have

$$(\partial_t, \nabla)\overline{F}_N(t_k, h_k) = \mathbb{E}\left(\frac{1}{N^{\mathsf{p}}}\mathrm{tr}\,\mathbf{c}\left(\left\langle X^{\otimes \mathsf{p}}A\right\rangle_k\right),\ \frac{1}{N}\mathbf{c}\left(\langle X\rangle_k\right)\right).$$

Hence, it suffices to show that, for any measurable g satisfying $\mathbb{E}|g(X)| < \infty$,

$$\mathbb{E}\,\mathbf{c}\left(\langle g(X)\rangle_1\right) \leqslant \mathbb{E}\,\mathbf{c}\left(\langle g(X)\rangle_2\right). \tag{4.2.8}$$

Indeed, in view of the previous display, the desired result follows from taking g to be $g(q) = q^{\otimes \mathsf{p}}A$ and then the identity map.

To compare the two sides in (4.2.8), we introduce

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

where W' and Z' have i.i.d. standard Gaussian entries, independent of randomness previously introduced. We claim that

$$\mathbb{E}[g(X)\,|\,Y_2] \overset{\mathrm{d}}{=} \mathbb{E}[g(X)\,|\,Y_1, Y'], \tag{4.2.9}$$

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

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

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

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

In view of (4.2.7) and (4.2.9), this yields (4.2.8) and thus the desired result.

It remains to prove (4.2.9). 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

$$\frac{2t_2}{N^{\mathsf{p}-1}}(x^{\otimes\mathsf{p}}A)\cdot(X^{\otimes\mathsf{p}}A) + \frac{1}{\sqrt{N^{\mathsf{p}-1}}}(x^{\otimes\mathsf{p}}A)\cdot\sqrt{2t_2}W_2 - \frac{t_2}{N^{\mathsf{p}-1}}|x^{\otimes\mathsf{p}}A|^2$$
$$+2h_2\cdot(x^{\intercal}X) + (Z_2\sqrt{2h_2})\cdot x - h_2\cdot(x^{\intercal}x),$$

while the Hamiltonian for the right-hand side is

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

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 (4.2.9). $\square$

4.2.2. Convexity

In this subsection, we show the following.

Lemma 4.2.3. *For each $N \in \mathbb{N}$, the function $\overline{F}_N : [0,\infty) \times \mathbb{S}^K_+ \to \mathbb{R}$ is convex.*

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

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

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

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

Again, this computation uses the Nishimori identity and the Gaussian integration by parts. Details for deriving the third identity above can be seen in the derivation of [94, (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

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

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

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

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

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

□

4.3. Some results of convex analysis

As mentioned above, our approach to proving Theorem 4.1.1 relies on the identification of the limit of $\overline{F}_N$ as the unique viscosity solution to (4.2.3). The uniqueness criterion we will use for this purpose is inspired by that described in Appendix 4.6. Compared with the setting explored there, equation (4.2.3) poses additional difficulties that are caused by the fact that the domain $\mathbb{S}^K_+$ of the "space" variable has a boundary. This is compounded by the fact that the relevant order on $\mathbb{S}^K_+$ is not total. The main purpose of this section is to demonstrate Proposition 4.3.9, 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 4.4.

4.3.1. Preliminaries

We start by recalling basic definitions and results from convex analysis. Since we need results for both functions defined on $\mathbb{S}^K_+$ and functions on $[0, \infty) \times \mathbb{S}^K_+$, 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 $\mathbb{R} \times \mathbb{S}^K$ or $\mathbb{S}^K$, 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

$$\partial u(x) := \Big\{ y \in \mathscr{H} : \ u(x') \geqslant u(x) + y \cdot (x' - x), \ \forall x' \in \mathscr{H} \Big\}. \tag{4.3.1}$$

The effective domain of u is

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

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

$$\mathbf{n}_{\mathscr{S}}(x) := \big\{ y \in \mathscr{H} : \ y \cdot (x' - x) \leqslant 0, \ \forall x' \in \mathscr{S} \big\}. \tag{4.3.2}$$

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 [105, Theorem 25.6] for a proof.

Lemma 4.3.1. *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 $\lim_{i\to\infty} x_i = x$.

Note that when x is in the interior of $\mathsf{dom}\, u$, we have $\mathbf{n}_{\mathsf{dom}\, u}(x) = \{0\}$.

We also record two classical results which, while not relevant to the proof of Proposition 4.3.9, 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 4.3.2. *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 4.3.1. □

Lemma 4.3.3. *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 $\lambda \to 0$. $\square$

To apply these results to the study of solutions to (4.2.3), we make the following remark.

Remark 4.3.4. Any convex function $f : [0, \infty) \times \mathbb{S}^K_+ \to \mathbb{R}$ can be extended in a standard way to a convex function $\overline{f} : \mathbb{R} \times \mathbb{S}^K \to \mathbb{R} \cup \{\infty\}$ by setting $\overline{f} = f$ on $[0, \infty) \times \mathbb{S}^K_+$ and $\overline{f} = \infty$ elsewhere. Note that $\overline{f}$ is proper and its effective domain is $[0, \infty) \times \mathbb{S}^K_+$ 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 $\mathscr{H} = \mathbb{R} \times \mathbb{S}^K$ and $D = (\partial_t, \nabla)$. Similar treatments can be taken for any convex function $\psi : \mathbb{S}^K_+ \to \mathbb{R}$.

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

Lemma 4.3.5. *For every $a \in \mathbb{S}^K$, we have $a \in \mathbb{S}^K_+$ if and only if $a \cdot b \geqslant 0$ for all $b \in \mathbb{S}^K_+$.*

Lemma 4.3.6. *For every $t \geqslant 0$ and $x \in \mathbb{S}^K_+$, we have $\mathbf{n}_{\mathbb{S}^K_+}(x) \subseteq -\mathbb{S}^K_+$ and $\mathbf{n}_{[0,\infty)\times\mathbb{S}^K_+}(t, x) \subseteq -([0, \infty) \times \mathbb{S}^K_+)$.*

The first lemma is an application of the diagonalizability of real symmetric matrices (see e.g. [94, Lemma 2.2]), and the second lemma is a consequence of the first lemma and the definition of outer normal cones in (4.3.2).

4.3.2. Nondecreasing gradients

The key result of this subsection is Proposition 4.3.9. To state it, it is convenient to introduce the following definitions. Recall the partial orders defined in (4.2.4) and (4.2.5).

Definition 4.3.7 (Nondecreasingness)**.** A real-valued function g defined on $\mathbb{S}^K_+$ or $[0, \infty) \times \mathbb{S}^K_+$ is said to be nondecreasing if $g(y_1) \leqslant g(y_2)$ whenever $y_1 \leqslant y_2$.

Definition 4.3.8 (Nondecreasing gradients)**.** A Lipschitz function $f : [0, \infty) \times \mathbb{S}^K_+ \to \mathbb{R}$ is said to have nondecreasing gradients if, for every (t_1, x_1) and (t_2, x_2) that are points of differentiability of f and satisfy $(t_1, x_1) \leqslant (t_2, x_2)$, it holds that

$$(\partial_t, \nabla)f(t_1, x_1) \leqslant (\partial_t, \nabla)f(t_2, x_2). \tag{4.3.3}$$

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

Proposition 4.3.9. *Suppose that $f : [0, \infty) \times \mathbb{S}^K_+ \to \mathbb{R}$ is nondecreasing, Lipschitz, convex, and has nondecreasing gradients. Then, for every $(t, x) \in [0, \infty) \times \mathbb{S}^K_+$, there exists $(b, q) \in \partial f(t, x) \cap [0, \infty) \times \mathbb{S}^K_+$ such that $|(b, q)| \leqslant \|f\|_{\mathrm{Lip}}$ and*

$$\text{for every } (a, p) \in \partial f(t, x), \quad (a, p) \leqslant (b, q). \tag{4.3.4}$$

In addition, if f satisfies (4.2.3) *on a dense set, then (b, q) can be chosen to satisfy $b - \mathsf{H}(q) = 0$.*

Remark 4.3.10. In the statement of Proposition 4.3.9, the precise interpretation of the phrase that f satisfies (4.2.3) on a dense set is that the set

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

is dense in $[0, \infty) \times \mathbb{S}^K_+$. We point out that one could equivalently replace this condition by the condition that f satisfies (4.2.3) at every point of differentiability in $(0, \infty) \times \mathbb{S}^K_{++}$. Indeed, one direction of this equivalence is immediate, since every Lipschitz function is differentiable almost everywhere. Conversely, if $(t, x) \in (0, \infty) \times \mathbb{S}^K_{++}$ 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 (4.2.3) is satisfied at (t_n, x_n). Then every subsequential limit of $(\partial_t f, \nabla f)(t_n, x_n)$, say $(a,p) \in \mathbb{R} \times \mathbb{S}^K$, 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 4.3.9. Let $(t,x) \in [0,\infty) \times \mathbb{S}^K_+$. We start by fixing some $(s_0, y_0) \in (0,\infty) \times \mathbb{S}^K_{++}$ such that $|(s_0,y_0)| = 1$. Note that

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

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

$$\left|(t_{0,j}, x_{0,j}) - \left((t,x) + j^{-1}(s_0,y_0)\right)\right| \leqslant j^{-2}, \quad \forall j \in \mathbb{N}. \tag{4.3.5}$$

If, in addition, f satisfies (4.2.3) 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 4.3.1, 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 $(b,q) \in [0,\infty) \times \mathbb{S}^K_+$. Continuity of H implies that $b - \mathsf{H}(q) = 0$ if f satisfies (4.2.3) on a dense set. It remains to show (4.3.4).

We apply Lemma 4.3.1 to the standard extension of f (see Remark 4.3.4). Note that $\mathsf{dom}\, f = [0,\infty) \times \mathbb{S}^K_+$. Let S be the corresponding set at (t,x) in this lemma. Then, due to this and Lemma 4.3.6, 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 (4.3.4) for $(a,p) \in \mathsf{cl}(\mathsf{conv}\, S)$. In fact, since the condition on (a,p) in (4.3.4) is stable under convex combinations and passage to the limit, it suffices to show (4.3.4) 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

$$\lim_{i\to\infty} (\partial_t, \nabla) f(t_i, x_i) = (a, p). \tag{4.3.6}$$

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

$$(t_i, x_i) \leqslant (t_{0,j}, x_{0,j}), \quad \forall i \geqslant i(j). \tag{4.3.7}$$

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

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

By this and (4.3.5), we have that, for every $\mathbf{a} \in [0, \infty) \times \mathbb{S}^K_+$,

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

The right-hand side is nonnegative for sufficiently large j. Lemma 4.3.5 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 (4.3.7). Using (4.3.6), (4.3.7) and the property (4.3.3), 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. □

4.4. Viscosity solutions

In this section, we study the Hamilton-Jacobi equation (4.2.3). 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 4.4.7. 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 4.5.

Recall that the notion of nondecreasing functions was introduced in Definition 4.3.7.

Definition 4.4.1 (Viscosity solutions)**.**

1. A nondecreasing Lipschitz function $f : [0,\infty) \times \mathbb{S}^K_+ \to \mathbb{R}$ is a viscosity subsolution to (4.2.3) if for every $(t,h) \in (0,\infty) \times \mathbb{S}^K_+$ and every smooth $\phi : (0,\infty) \times \mathbb{S}^K_+ \to \mathbb{R}$ such that $f - \phi$ has a local maximum at (t,h), we have

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

2. A nondecreasing Lipschitz function $f : [0,\infty) \times \mathbb{S}^K_+ \to \mathbb{R}$ is a viscosity supersolution to (4.2.3) if for every $(t,h) \in (0,\infty) \times \mathbb{S}^K_+$ and every smooth $\phi : (0,\infty) \times \mathbb{S}^K_+ \to \mathbb{R}$ such that $f - \phi$ has a local minimum at (t,h), we have

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

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

3. A nondecreasing Lipschitz function $f : [0,\infty) \times \mathbb{S}^K_+ \to \mathbb{R}$ is a viscosity solution to (4.2.3)

if f is both a viscosity subsolution and supersolution.

Remarks 4.4.3 and 4.4.4 below aim to provide a somewhat more intuitive understanding of Definition 4.4.1. Before doing so, we record the following observation.

Lemma 4.4.2. *The function* $\mathsf{H} : \mathbb{S}^K_+ \to \mathbb{R}$ *given in* (4.1.4) *is nondecreasing.*

Proof. Let $a, b \in \mathbb{S}^K_+$ be such that $a \leqslant b$. Recalling that the tensor product of two positive semidefinite matrices is positive semidefinite, see for instance [117, Theorem 7.20], one can show by induction on p that $a^{\otimes \mathsf{p}} \leqslant b^{\otimes \mathsf{p}}$. Since $AA^\intercal \in \mathbb{S}^{K^\mathsf{p}}_+$, we can use Lemma 4.3.5 to obtain that $\mathsf{H}(a) \leqslant \mathsf{H}(b)$, as desired. □

Remark 4.4.3. Given a nondecreasing Lipschitz function f, define the extension of H by

$$\overline{\mathsf{H}}(p) := \inf \Big\{ \mathsf{H}(q) : q \geqslant p,\, q \in \mathbb{S}^K_+,\, |q| \leqslant \|f\|_{\mathrm{Lip}} \Big\}, \quad \forall p \in \mathbb{S}^K. \tag{4.4.1}$$

As usual, the infimum over an empty set is understood to be ∞. Note that $\overline{\mathsf{H}} : \mathbb{S}^K \to \mathbb{R} \cup \{\infty\}$ is lower semi-continuous and agrees with H on $\mathbb{S}^K_+$ due to Lemma 4.4.2. Then, Definition 4.4.1 (2) can be reformulated as follows: f is a viscosity supersolution if for every $(t, h) \in (0, \infty) \times \mathbb{S}^K_+$ and every smooth $\phi : (0, \infty) \times \mathbb{S}^K_+$ 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 $h \in \mathbb{S}^K_{++}$ and $h \in \mathbb{S}^K_+ \setminus \mathbb{S}^K_{++}$.

Remark 4.4.4. Further simplifications of boundary conditions can be made. After the submission of this paper, [41] considers solutions defined to satisfy the equation in the viscosity sense everywhere including the boundary without any additional boundary condition imposed. Under this definition, the comparison principle and the existence of solutions still hold. Moreover, the solution admits a representation by the Hopf–Lax formula given the convexity of the nonlinearity, or the Hopf formula given the convexity of the initial condition. All properties needed in this work are still satisfied. One can work with this definition, and the main results in this work are still valid.

Let us briefly describe the simplification. Due to Lipschitzness of $\overline{F}_N$ uniformly in N (Lemma 4.2.1), we can work with a regularized nonlinearity $\mathsf{H}^{\mathsf{reg}} : \mathbb{S}^D_+ \to \mathbb{R}$ which coincides with H on a ball intersected with $\mathbb{S}^D_+$ with sufficiently large radius. In a similar way as in [41, Lemma 4.2], $\mathsf{H}^{\mathsf{reg}}$ can be constructed to be Lipschitz and nondecreasing. Then, we extend $\mathsf{H}^{\mathsf{reg}}$ to

$$\mathsf{H}^{\mathsf{ext}}(p) := \inf\left\{\mathsf{H}^{\mathsf{reg}}(q) : q \geqslant p,\, q \in \mathbb{S}^K_+\right\}, \quad \forall p \in \mathbb{S}^K.$$

One can check, similarly as in [41, Lemma 4.4], that $\mathsf{H}^{\mathsf{ext}}$ is Lipschitz and and nondecreasing. Then, the conditions for viscosity subsolutions and supersolutions can be replaced by

$$\left(\partial_t \phi - \mathsf{H}^{\mathsf{ext}}(\nabla \phi)\right)(t,h) \leqslant 0,$$
$$\left(\partial_t \phi - \mathsf{H}^{\mathsf{ext}}(\nabla \phi)\right)(t,h) \geqslant 0,$$

respectively, without the need to distinguish between $h \in \mathbb{S}^K_+ \setminus \mathbb{S}^K_{++}$ and $h \in \mathbb{S}^K_{++}$. The key property needed for this simplification in [41] is the monotonicity of the nonlinearity.

We turn to the well-posedness of equation (4.2.3). We first state a comparison principle, which ensures in particular that there is at most one viscosity solution with a given initial condition.

Proposition 4.4.5 (Comparison principle)**.** *If u is a subsolution and v is a supersolution to* (4.2.3)*, then*

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

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

Proposition 4.4.6 (Hopf formula)**.** *Let $\psi : \mathbb{S}^K_+ \to \mathbb{R}$ be convex, Lipschitz and nondecreasing, and let f be given by*

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

Then, the function f is a viscosity solution to (4.2.3) *with initial condition $f(0,\cdot) = \psi$.*

For the proofs of these two propositions, we refer to [39, Section 6].

In the remainder 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 4.2.

4.4.1. Identification criterion

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

Proposition 4.4.7. *Let $f : [0,\infty) \times \mathbb{S}^K_+ \to \mathbb{R}$ be nondecreasing, Lipschitz, convex, and have nondecreasing gradients. Suppose that $\psi = f(0,\cdot)$ is C^1 and that f satisfies* (4.2.3) *on a dense subset. Then, f is a viscosity solution to* (4.2.3) *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 4.6. 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 4.6, both assumptions are necessary; see in particular Remark 4.6.3 there.

Compared with the Euclidean setting discussed in Appendix 4.6, the existence of the boundary of $\mathbb{S}^K_+$ complicates the arguments. Indeed, in view of Lemma 4.3.1, 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 4.6.2. It turns out that for Proposition 4.4.7, a work-around is available by exploiting the assumption that the function f has nondecreasing gradients.

As preparation for this, we use Proposition 4.3.9 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 4.6.2.

Lemma 4.4.8. *Under the assumptions in Proposition 4.4.7, for every $y \in \mathbb{S}^K_+$ and every $p \in \partial\psi(y)$, there is $(b, p') \in [0, \infty) \times \mathbb{S}^K_+$ 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 $\psi : \mathbb{S}^K_+ \to \mathbb{R}$ is C^1, by Lemma 4.3.1 and setting $p' = \nabla\psi(y)$, we have

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

This implies that

$$p = p' + n \tag{4.4.2}$$

for some

$$n \in \mathbf{n}_{\mathbb{S}^K_+}(y). \tag{4.4.3}$$

Due to Lemma 4.3.6, we have $-n \in \mathbb{S}^K_+$, that is,

$$p \leqslant p'. \tag{4.4.4}$$

The same argument also yields that,

$$\text{for every } q' \in \partial\psi(y), \quad q' \leqslant p'. \tag{4.4.5}$$

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

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

which due to the convexity of ψ implies that $(0, p') \in \partial f(0, y)$. Let

$$(b, q) \in \partial f(0, y) \tag{4.4.6}$$

be as described in Proposition 4.3.9, for f at the point $(0, y)$. Then, the following properties hold

$$(0, p') \leqslant (b, q), \tag{4.4.7}$$

$$|(b, q)| \leqslant \|f\|_{\mathrm{Lip}}, \quad b - \mathsf{H}(q) = 0. \tag{4.4.8}$$

Since $f(0, \cdot) = \psi$, we must have $q \in \partial\psi(y)$. Combining (4.4.5) and (4.4.7), we see that

$$p' = q. \tag{4.4.9}$$

We are now ready to conclude. By (4.4.3) and the definition of outer normal in (4.3.2), we can verify that

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

This along with Lemma 4.3.1, (4.4.6) and (4.4.9) implies

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

The lemma then follows from this display, (4.4.2), (4.4.4), (4.4.8) and (4.4.9). $\square$

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

Proof of Proposition 4.4.7. We check that f must be a subsolution to (4.2.3). Let $\phi \in C^\infty((0,\infty)\times\mathbb{S}^K_+)$, and $(t,x) \in (0,\infty)\times\mathbb{S}^K_+$ be such that $f-\phi$ has a local maximum at (t,x). If $x \in \mathbb{S}^K_+ \setminus \mathbb{S}^K_{++}$, since, for each $a \in \mathbb{S}^K_+$ 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 $a \in \mathbb{S}^K_+$. By Lemma 4.3.5, this implies that $\nabla\phi(t,x) \in \mathbb{S}^K_+$. If $x \in \mathbb{S}^K_{++}$, 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 $\{(\partial_t\phi, \nabla\phi)(t,x)\}$, 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 also Remark 4.3.10, 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 (4.2.3). Fix any (t,x), and any

$$(a,p) \in \partial f(t,x). \tag{4.4.10}$$

Recall Remark 4.4.3 and the extension $\overline{\mathsf{H}}$ defined there. Taking (t,x) and ϕ as in Definition 4.4.1 (2), we can use Lemma 4.3.3 to see that $(\partial_t\phi(t,x), \nabla\phi(t,x)) \in \partial f(t,x)$. Therefore, it suffices to show that

$$a - \overline{\mathsf{H}}(p) \geqslant 0. \tag{4.4.11}$$

We proceed in four steps.

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

$$\inf_{y \in \mathbb{S}^K_+} \big(f_\varepsilon(0, y) - y \cdot p \big) \tag{4.4.12}$$

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

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

Note that we are working with a slightly different perturbation of f from the one in Step 3 in the proof of Proposition 4.6.2. The purpose is to ensure that the perturbative term is differentiable everywhere so that Lemma 4.3.2 is applicable. 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

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

which implies that

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

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 $y_\varepsilon \in \mathbb{S}^K_+$.

Step 2. We show

$$\lim_{\varepsilon \to 0} \varepsilon \sqrt{1 + |y_\varepsilon|^2} = 0. \tag{4.4.13}$$

We first observe that

$$\limsup_{\varepsilon\to 0}\inf_{y\in\mathbb{S}^K_+}\left(f(0,y)+\varepsilon\sqrt{1+|y|^2}-y\cdot p\right)=\inf_{y\in\mathbb{S}^K_+}\left(f(0,y)-y\cdot p\right). \tag{4.4.14}$$

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

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

and the other direction of the inequality in (4.4.14) is obvious. Since y_ε achieves the infimum on the left-hand side of (4.4.14) and also satisfies

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

we conclude that (4.4.13) 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 (4.4.12), we have that $p\in\partial\psi_\varepsilon(y_\varepsilon)$. Lemma 4.3.2 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

$$|p-p_\varepsilon|\leqslant\varepsilon. \tag{4.4.15}$$

By Lemma 4.4.8 applied to p_ε, there exists $(b_\varepsilon, p'_\varepsilon) \in [0,\infty) \times \mathbb{S}^K_+$ such that

$$(b_\varepsilon, p_\varepsilon) \in \partial f(0, y), \tag{4.4.16}$$

$$p_\varepsilon \leqslant p'_\varepsilon, \quad p'_\varepsilon \in \mathbb{S}^K_+, \quad |p'_\varepsilon| \leqslant \|f\|_{\mathrm{Lip}} \tag{4.4.17}$$

$$b_\varepsilon - \mathsf{H}(p'_\varepsilon) = 0. \tag{4.4.18}$$

Step 4. We are now ready to prove (4.4.11). Define $h : \lambda \mapsto f\left(\lambda(t,x) + (1-\lambda)(0, y_\varepsilon)\right)$ on $[0,1]$. Clearly, h is convex. By (4.4.16), 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 (4.4.10), 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 (4.4.15) and (4.4.13) implies that, as ε tends to zero,

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

By (4.4.18), the definition of $\overline{\mathsf{H}}$ in (4.4.1), and (4.4.17), 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 (4.4.15) together with the two previous displays yields that

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

and (4.4.11) follows by letting ε tend to zero. $\square$

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 4.4.9. *Let $f : [0,\infty) \times \mathbb{S}^K_+ \to \mathbb{R}$ 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 $\phi \in C^\infty((0,\infty) \times \mathbb{S}^K_+)$ and $(t,x) \in (0,\infty) \times \mathbb{S}^K_{++}$ 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 (4.2.3).

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 $(t',x') \in (0,\infty) \times \mathbb{S}^K_+$,

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

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 4.4.7 and Remark 4.3.10, it therefore suffices to show that the set

$$\left\{(t,x) \in (0,\infty) \times \mathbb{S}^K_{++} : \exists\, \phi \in C^\infty((0,\infty) \times \mathbb{S}^K_+) \text{ s.t. } (t,x) \text{ is a local maximum of } f - \phi\right\} \tag{4.4.19}$$

is dense in $[0,\infty) \times \mathbb{S}^K_+$. (The additional restriction that the local maximum be strict is easily addressed a posteriori.) Since the closure of $(0,\infty) \times \mathbb{S}^K_{++}$ is $[0,\infty) \times \mathbb{S}^K_+$, it suffices to show that the set in (4.4.19) is dense in $(0,\infty) \times \mathbb{S}^K_{++}$. We fix any $(t,x) \in (0,\infty) \times \mathbb{S}^K_{++}$, 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 $(t,x) \in (0,\infty) \times \mathbb{S}^K_{++}$, we have that $(t_\alpha, x_\alpha) \in (0,\infty) \times \mathbb{S}^K_{++}$ for every sufficiently large α. Hence (t_α, x_α) belongs to the set in (4.4.19), and we conclude that the set in (4.4.19) is a dense subset of $[0,\infty) \times \mathbb{S}^K_{+}$. □

4.5. Convergence and application

The main goal of this section is to prove Theorem 4.1.1, using the tools developed in the previous section. For illustration, we also apply the theorem to a specific model.

4.5.1. Convergence

In view of Proposition 4.4.6, Theorem 4.1.1 follows from the next theorem.

Theorem 4.5.1. *Under the conditions of Theorem 4.1.1, the function $\overline{F}_N$ converges pointwise to the unique viscosity solution to* (4.2.3) *with initial condition ψ.*

In order to prove this result, we start by recalling from [39, Proposition 3.1] (cf. also [94, Proposition 1.2]) that the function $\overline{F}_N$ satisfies an approximate form of the equation. In (4.5.1), we implicitly understand that the relevant functions are evaluated at $(t,h) \in [0,\infty) \times \mathbb{S}^K_{+}$.

Proposition 4.5.2 (Approximate Hamilton-Jacobi equation)**.** *There exists $C < \infty$ such* *that for every $N \geqslant 1$ and uniformly over $[0,\infty) \times \mathbb{S}^K_+$,*

$$\left|\partial_t \overline{F}_N - \mathsf{H}(\nabla \overline{F}_N)\right|^2 \leqslant C\kappa(h) N^{-\frac{1}{4}} \left(\Delta \overline{F}_N + |h^{-1}|\right)^{\frac{1}{4}} + C\mathbb{E}\left[\left|\nabla F_N - \nabla \overline{F}_N\right|^2\right], \qquad (4.5.1)$$

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

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

Proof of Theorem 4.5.1. Since $\overline{F}_N$ is Lipschitz uniformly in N by Lemma 4.2.1, 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 (4.2.3) and the uniqueness is ensured by Proposition 4.4.5. For convenience, we assume that the whole sequence $(\overline{F}_N)_{N\in\mathbb{N}}$ converges to f.

Lemmas 4.2.1 and 4.2.3 (see also (4.2.6)) 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 4.2.2 yields that f has nondecreasing gradients. Let $(t,h) \in (0,\infty) \times \mathbb{S}^K_{++}$ and $\phi \in C^\infty((0,\infty) \times \mathbb{S}^K_+)$ be such that $f - \phi$ has a strict local maximum at (t,h). By Corollary 4.4.9, it suffices to show that

$$(\partial_t \phi - \mathsf{H}(\nabla \phi))(t,h) = 0. \qquad (4.5.2)$$

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

$$(\partial_t \overline{F}_N - \partial_t \phi)(t_N, h_N) = 0 \quad \text{and} \quad (\nabla \overline{F}_N - \nabla \phi)(t_N, h_N) = 0. \tag{4.5.3}$$

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 $h' \in \mathbb{S}^K_+$ with $|h'| \leqslant C^{-1}$, we have

$$0 \leqslant \overline{F}_N(t_N, h_N + h') - \overline{F}_N(t_N, h_N) - h' \cdot \nabla \overline{F}_N(t_N, h_N) \leqslant C|h'|^2. \tag{4.5.4}$$

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

$$\begin{aligned}
&\overline{F}_N(t_N, h_N + h') - \overline{F}_N(t_N, h_N) \\
&\qquad = h' \cdot \nabla \overline{F}_N(t_N, h_N) + \int_0^1 (1-s) h' \cdot \nabla \left(h' \cdot \nabla \overline{F}_N \right)(t_N, h_N + sh') \, \mathrm{d}s.
\end{aligned} \tag{4.5.5}$$

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 (4.5.3), 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

(4.5.5) once more, we obtain (4.5.4).

Step 2. Let

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

In this step, we show that

$$\mathbb{E}\left[|\nabla F_N - \nabla \overline{F}_N|^2(t_N, h_N)\right] \leqslant C \left(\mathbb{E}\left[\sup_D |F_N - \overline{F}_N|^2\right]\right)^{\frac{1}{2}}. \tag{4.5.6}$$

We recall from [39, (3.13)] that, for every $a \in \mathbb{S}^K$ and $(t', h') \in [0,\infty) \times \mathbb{S}^K_+$ 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 $Z \in \mathbb{R}^{N\times K}$ is the matrix of independent standard Gaussians appearing in the definition of $\overline{Y}$ (see the second paragraph in Section 4.1.1). Applying Taylor's formula as in Step 1, it is readily verified that 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 (4.5.4), 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 the random matrix

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

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 (4.5.6).

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 4.5.2, (4.5.4), and (4.5.6), 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 (4.5.3) and the fact that the function ϕ is smooth, this yields (4.5.2), and thus completes the proof. □

4.5.2. Application

We study the model considered in [85], which corresponds to (4.1.1) 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

$$Y_{\mathbf{i}} = \sqrt{\frac{2t}{N^{\mathsf{p}-1}}} \sum_{j=1}^{K} \prod_{n=1}^{\mathsf{p}} X_{i_n, j} + W_{\mathbf{i}}, \quad \mathbf{i} \in \{1, \cdots, N\}^{\mathsf{p}}, \tag{4.5.7}$$

where $X \in \mathbb{R}^{N\times K}$ is assumed to have i.i.d. row vectors with norms bounded by $\sqrt{K}$ almost surely. Hence, the condition in (4.1.2) is satisfied. For even p, the limit of the free energy associated with this model has been proved to satisfy a variational formula in [85]. When p is odd, the situation is more difficult; in [39], it was only proven that the limit is bounded above by a variational formula. Here, we will apply Theorem 4.1.1 to treat both even and odd values of p.

Recall the definition of H in (4.1.4). In this case, the nonlinearity H is given by

$$\mathsf{H}(q) = \sum_{k,k'=1}^{K} (q_{k,k'})^{\mathsf{p}}, \quad \forall q \in \mathbb{S}^K_+. \tag{4.5.8}$$

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 (4.1.3), we have

$$\psi(h) = \mathbb{E}\log \int_{\mathbb{R}^{1\times K}} \exp\left(2h\cdot (x^\intercal X_{1,\cdot}) + \sqrt{2h}\cdot (x^\intercal Z) - h\cdot (x^\intercal x)\right)\mathrm{d}P(x), \quad \forall h \in \mathbb{S}^K_+, \tag{4.5.9}$$

where P is the law of the first row vector $X_{1,\cdot} = (X_{1,k})_{1\leqslant k\leqslant K}$. By Lemma 4.2.1, ψ is C^1. The concentration condition $\lim_{N\to\infty}\mathbb{E}\|F_N - \overline{F}_N\|^2_{L^\infty(D)} = 0$ for each compact $D \subseteq [0,\infty)\times\mathbb{S}^K_+$ is proved in [39, Lemma C.1]. Hence, the next result follows from Theorem 4.1.1.

Corollary 4.5.3. *Under the assumption* (4.1.2)*, in the model described above with* $\mathsf{p}\in\mathbb{N}$*, it holds that, for every* $(t,h)\in[0,\infty)\times\mathbb{S}^K_+$*,*

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

for H *and* ψ *given in* (4.5.8) *and* (4.5.9)*, respectively.*

4.5.3. Simplification of the variational formula

We describe a way of simplifying the formula (4.1.5) under the additional assumption that the mapping H in (4.1.4) only depends on the diagonal entries of its argument.

We introduce the linear map $\mathsf{diag}:\mathbb{R}^K\to\mathbb{S}^K$ defined by $\mathsf{diag}x = \mathrm{diag}(x_1,\dots,x_K)$. Its adjoint $\mathsf{diag}^*:\mathbb{S}^K\to\mathbb{R}^K$ is given by $\mathsf{diag}^*h = (h_{11},\dots,h_{KK})$ for $h\in\mathbb{S}^K$. Note that $\mathsf{diag}^*\mathsf{diag}$ is the identity map on $\mathbb{R}^K$, and $\mathsf{diag}\mathsf{diag}^*h = \mathrm{diag}(h_{11},\dots,h_{KK})$ for every $h\in\mathbb{S}^K$. The additional assumption on H can be reformulated as

$$\mathsf{H}(q) = \mathsf{H}(\mathsf{diag}\mathsf{diag}^*q), \quad \forall q\in\mathbb{S}^K_+. \tag{4.5.10}$$

For $x, x' \in \mathbb{R}^K$, we write $x \cdot x' = \sum_{i=1}^K x_i x_i'$. We set $\mathbb{R}^K_+ = [0,\infty)^K$. Note that $\mathsf{diag}(\mathbb{R}^K_+)$ contains exactly the diagonal matrices in $\mathbb{S}^K_+$. For F_N given in (4.1.3), we want to show that, under the assumptions of Theorem 4.1.1 and for every $t \geqslant 0$ and $x \in \mathbb{R}^K_+$, we have

$$\lim_{N\to\infty} \overline{F}_N(t, \mathsf{diag}x) = \sup_{x'' \in \mathbb{R}^K_+} \inf_{x' \in \mathbb{R}^K_+} \left\{ x'' \cdot \left(x - x'\right) + \psi\left(\mathsf{diag}x'\right) + t\mathsf{H}\left(\mathsf{diag}x''\right) \right\}. \qquad (4.5.11)$$

In particular, setting $x = 0$, we obtain a simpler representation of the limit of the original free energy F_N°.

The proof of this statement can be achieved by working with the following Hamilton–Jacobi equation:

$$\partial_t g - \mathsf{H}(\mathsf{diag}\nabla g) = 0, \quad \text{on } [0,\infty) \times \mathbb{R}^K_+. \qquad (4.5.12)$$

The well-posedness of this equation and the representation of the solution by the Hopf formula can be established in a similar way (see [41, Section 2]). A corresponding identification criterion for solutions, as stated in Proposition 4.4.7, can also be obtained. There, the partial order defining the notion of nondecreasingness, as in (4.2.4) and (4.2.5), is now induced by the cone $\mathbb{R}^K_+$. Lastly, for any differentiable function $\phi : \mathbb{S}^K_+ \to \mathbb{R}$, we can verify that,

$$\nabla \phi^{\mathsf{diag}}(x) = \mathsf{diag}^* \nabla \phi(h)\big|_{h=\mathsf{diag}x}, \quad \forall x \in \mathbb{R}^K_+,$$

where $\phi^{\mathsf{diag}} : \mathbb{R}^K_+ \to \mathbb{R}$ is given by $\phi^{\mathsf{diag}} = \phi(\mathsf{diag}\,\cdot\,)$.

Hence, setting $\overline{F}_N^{\mathsf{diag}}(t,x) = \overline{F}_N(t, \mathsf{diag}x)$, and using Proposition 4.5.2 and (4.5.10), we can see that $\overline{F}_N^{\mathsf{diag}}$ approximately solves (4.5.12) and that a similar estimate in Proposition 4.5.2 holds for $\overline{F}_N^{\mathsf{diag}}$. Then, the same argument as in the proof of Theorem 4.5.1 yields that $\overline{F}_N^{\mathsf{diag}}$ converges to the unique viscosity solution of (4.5.12) with initial condition $\psi(\mathsf{diag}\,\cdot\,)$. Due to the convexity of $\psi(\mathsf{diag}\,\cdot\,)$, the solution admits a representation by the Hopf formula, which is exactly the right-hand side in (4.5.11).

As a concrete example inspired by [6], suppose as in the previous subsection that $X \in \mathbb{R}^{N\times K}$ has i.i.d. row vectors with norm bounded by $\sqrt{K}$, but this time we observe, for each $i, j \in \{1, \ldots, N\}$ and $k \in \{1, \ldots, K-1\}$, the quantity

$$\sqrt{\frac{2t}{N}} X_{i,k} X_{j,k+1} + W_{i,j,k},$$

where $(W_{i,j,k})_{i,j\leqslant N, k<K}$ are independent standard Gaussians, independent of X. This can be mapped into our setting by choosing $\mathsf{p} = 2$, $L = K-1$, $A \in \mathbb{R}^{K^2\times(K-1)}$ given by $A_{(k,l),r} = 1$ if $r = k = l-1$ and zero otherwise. With this choice of A, the function H takes the form

$$\mathsf{H}(q) = \sum_{k=1}^{K-1} q_{k,k}\, q_{k+1,k+1} = \mathsf{H}(\mathsf{diag}\mathsf{diag}^* q), \qquad \forall q \in \mathbb{S}^K_+.$$

We thus obtain that the limit free energy $F_N^\circ(t) = F_N(t, 0)$ is given by

$$\lim_{N\to\infty} F_N^\circ(t) = \sup_{x'\in\mathbb{R}^K_+} \inf_{x\in\mathbb{R}^K_+} \left\{ \psi^{\mathsf{diag}}(x) - x\cdot x' + t \sum_{k=1}^{K-1} x'_k x'_{k+1} \right\}. \tag{4.5.13}$$

Moreover, under the additional assumption that the coordinates of the vector $(X_{1,k})_{1\leqslant k\leqslant K}$ are independent, the initial condition ψ^{diag} can be decomposed into a sum of functions of one variable: there exist convex and nondecreasing functions $\psi_1, \ldots, \psi_K : \mathbb{R}_+ \to \mathbb{R}$ such that for every $x \in \mathbb{R}^K_+$,

$$\psi^{\mathsf{diag}}(x) = \sum_{k=1}^{K} \psi_k(x_k).$$

(Cases in which different layers have different lengths, say for instance $X_{i,k} = 0$ for every $i > \alpha_k N$ for some fixed $\alpha_k \in (0,1)$, can be covered as well, and this translates into multiplying each ψ_k by a suitable scalar.) Under these conditions, the formula (4.5.13) can be further simplified, as we now explain. For each $x \in \mathbb{R}^K$, we denote by $x_o = (x_1, x_3, \ldots, x_{2\cdot\lfloor (K-1)/2\rfloor+1})$ and $x_e = (x_2, x_4, \ldots, x_{2\cdot\lfloor K/2\rfloor})$ respectively the odd and even coordinates of the vector x, and for each $k \in \{1, \ldots, K\}$ and $y \geqslant 0$, we set

$$\psi_k^*(y) := \sup_{x \geqslant 0}(xy - \psi_k(x)).$$

By [105, Theorem 12.4], we have that $\psi_k^{**} = \psi_k$. Moreover, we can write

$$\inf_{x_e} \left\{ \sum_{k=1}^{K} \psi_k(x_k) - x \cdot x' + t \sum_{k=1}^{K-1} x_k' x_{k+1}' \right\} = \sum_{k \text{ odd}} \left(\psi_k(x_k) - x_k x_k'\right) - \sum_{k \text{ even}} \psi_k^*(x_k') + t \sum_{k=1}^{K-1} x_k' x_{k+1}',$$

and observe that the optimization problems over x_o and x_e' are separated. We can thus interchange $\sup_{x_e'}$ and $\inf_{x_o}$ to get that

$$\begin{aligned}\lim_{N \to \infty} F_N^\circ(t) &= \sup_{x_o'} \inf_{x_o} \sup_{x_e'} \left\{ \sum_{k \text{ odd}} \left(\psi_k(x_k) - x_k x_k'\right) - \sum_{k \text{ even}} \psi_k^*(x_k') + t \sum_{k=1}^{K-1} x_k' x_{k+1}' \right\} \\ &= \sup_{x_o'} \inf_{x_o} \left\{ \sum_{k \text{ odd}} \left(\psi_k(x_k) - x_k x_k'\right) + \sum_{k \text{ even}} \psi_k\left(t x_{k-1}' + t x_{k+1}'\right) \right\},\end{aligned}$$

using that $\psi_k^{**} = \psi_k$, and with the understanding that $x_{K+1} = 0$. Similar formulas were first obtained in [6].

4.6. 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 $[0, \infty) \times \mathbb{S}_+^K$ in Proposition 4.4.7.

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

$$\partial_t f - \mathsf{H}(\nabla f) = 0 \qquad \text{on} \;\; [0, \infty) \times \mathbb{R}^d. \tag{4.6.1}$$

Definition 4.6.1.

1. A continuous function $f : [0, \infty) \times \mathbb{R}^d \to \mathbb{R}$ is a viscosity subsolution to (4.2.3) 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 (4.2.3) 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 (4.2.3) if f is both a viscosity subsolution and supersolution.

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

Proposition 4.6.2. *Let $f : [0, \infty) \times \mathbb{R}^d \to \mathbb{R}$ be Lipschitz and convex. Suppose that f satisfies* (4.6.1) *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* (4.6.1) *with initial condition $f(0, \cdot)$.*

Remark 4.6.3. In Proposition 4.6.2, 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 4.6.2. Recall the definition of subdifferential in (4.3.1). We decompose the proof into three steps.

Step 1. We check that f must be a subsolution to (4.6.1). 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 along with the convexity of f 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 4.3.10, 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 (4.6.1). Let $(a,p) \in \partial f(t,x)$. In view of Lemma 4.3.3, it is sufficient to show that

$$a - \mathsf{H}(p) \geqslant 0. \tag{4.6.2}$$

Since $(a,p) \in \partial f(t,x)$ and f is convex, 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

$$\inf_{y \in \mathbb{R}^d} (f(0,y) - y \cdot p) \tag{4.6.3}$$

is achieved, and we denote by y a point realizing the infimum. By arguing as in Remark 4.3.10, 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 (4.6.3), we must have $p' = \partial_y f(0,y) = p$, and thus $(b,p) \in \partial f(0,y)$ with $b - \mathsf{H}(p) = 0$.

Due to the convexity of f, the mapping $g : \lambda \mapsto f(\lambda(t,x) + (1-\lambda)(0,y))$ is convex over the interval $[0,1]$. Since $(b,p) \in \partial f(0,y)$, the right derivative of g at 0 satisfies

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

Due to $(a,p) \in \partial f(t,x)$, 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 (4.6.2), as desired.

Step 3. To conclude, there remains to consider the case when the infimum in (4.6.3) is not achieved. For every $\varepsilon > 0$, we consider

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

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

$$|\nabla f(0, y_\varepsilon) - p| \leqslant \varepsilon. \tag{4.6.4}$$

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

$$\lim_{\varepsilon\to 0} \varepsilon|y_\varepsilon| = 0. \tag{4.6.5}$$

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 (4.6.4) and (4.6.5), 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 (4.6.4) and the continuity of H, we obtain (4.6.2). □

CHAPTER 5

Limiting free energy of multi-layer generalized linear models

This chapter is essentially borrowed from [40], joint with Hong-Bin Chen.

Abstract. We compute the high-dimensional limit of the free energy associated with a multi-layer generalized linear model. Under certain technical assumptions, we identify the limit in terms of a variational formula. The approach is to first show that the limit is a solution to a Hamilton–Jacobi equation whose initial condition is related to the limiting free energy of a model with one fewer layer. Then, we conclude by an iteration.

5.1. Introduction

5.1.1. Setting

Let us describe the model. For $n \in \mathbb{N}$, let X be an $\mathbb{R}^n$-valued random vector with distribution P_X, serving as the original signal. Fix any $L \in \mathbb{N}$ as the number of layers. For $l \in \{0, 1, 2, \ldots, L\}$, let $n_l = n_l(n) \in \mathbb{N}$ be the dimension of the signal at the l-th layer. We assume that $n_0 = n$ and

$$\lim_{n \to \infty} \frac{n_l}{n} = \alpha_l > 0, \tag{5.1.1}$$

for some $\alpha_l > 0$. In particular, we have that $\alpha_0 = 1$.

For each $l \in \{1, 2, \ldots, L\}$, let

- $\varphi_l : \mathbb{R} \times \mathbb{R}^{k_l} \to \mathbb{R}$ be a measurable function for some fixed $k_l \in \mathbb{N}$ (independent of n);
- $(A_j^{(l)})_{1 \leqslant j \leqslant n_l}$ be a finite sequence of $\mathbb{R}^{k_l}$-valued random vectors, all together with law $P_{A^{(l)}}$;
- $\Phi^{(l)}$ be an $n_l \times n_{l-1}$ random matrix with law $P_{\Phi^{(l)}}$.

For $l' \geqslant l$, we also write

$$A^{[l,l']} = \left(A^{(m)}\right)_{l \leqslant m \leqslant l'}, \qquad \Phi^{[l,l']} = \left(\Phi^{(m)}\right)_{l \leqslant m \leqslant l'}, \tag{5.1.2}$$

and denote their laws by $P_{A^{[l,l']}}$ and $P_{\Phi^{[l,l']}}$, respectively.

Starting with $X^{(0)} = X$, we iteratively define, for each $l \in \{1, \dots, L\}$,

$$X_j^{(l)} = \varphi_l \left(\frac{1}{\sqrt{n_{l-1}}} \sum_{k=1}^{n_{l-1}} \Phi_{jk}^{(l)} X_k^{(l-1)},\ A_j^{(l)} \right), \quad \forall 1 \leqslant j \leqslant n_l.$$

Viewing the action of φ_l component-wise, we also write

$$X^{(l)} = \varphi_l \left(\frac{1}{\sqrt{n_{l-1}}} \Phi^{(l)} X^{(l-1)},\ A^{(l)} \right). \tag{5.1.3}$$

For $\beta \geqslant 0$, the observable is given by

$$Y^\circ = \sqrt{\beta} X^{(L)} + Z \tag{5.1.4}$$

where Z is an n_L-dimensional standard Gaussian vector. The inference task is to recover X based on the knowledge of Y°, $(\varphi_l)_{1 \leqslant l \leqslant L}$ and $\Phi^{[1,L]}$.

Using (5.1.3) iteratively, we can find a deterministic function ζ_{L-1} such that $X^{(L-1)} = \zeta_{L-1}(X, A^{[1,L-1]}, \Phi^{[1,L-1]})$. We introduce the shorthand notation:

$$x^{(L-1)} = \zeta_{L-1}\left(x, a, \Phi^{[1,L-1]}\right), \quad \forall x \in \mathbb{R}^n,\ a = \left(a^{(1)}, \dots, a^{(L-1)}\right) \in \prod_{l=1}^{L-1} \mathbb{R}^{n_l \times k_l}. \tag{5.1.5}$$

We emphasize that $x^{(L-1)}$ is random due to the presence of $\Phi^{[1,L-1]}$ and also depends on the input x and a. By Bayes' rule, the law of $(X, A^{[1,L-1]})$ conditioned on $(Y^\circ, \Phi^{[1,L]})$ is given by

$$\frac{1}{\mathcal{Z}^\circ_{\beta,L,n}}\mathcal{P}_{\beta,L,n}\left(Y^\circ\middle|\frac{1}{\sqrt{n_{L-1}}}\Phi^{(L)}x^{(L-1)}\right)\mathrm{d}P_X(x)\mathrm{d}P_{A^{[1,L-1]}}(a)$$

where

$$\mathcal{P}_{\beta,L,n}(y|z) = \int e^{-\frac{1}{2}|y-\sqrt{\beta}\varphi_L(z,a^{(L)})|^2}\mathrm{d}P_{A^{(L)}}\left(a^{(L)}\right), \quad \forall y,z\in\mathbb{R}^{n_L}, \tag{5.1.6}$$

$$\mathcal{Z}^\circ_{\beta,L,n} = \int \mathcal{P}_{\beta,L,n}\left(Y^\circ\middle|\frac{1}{\sqrt{n_{L-1}}}\Phi^{(L)}x^{(L-1)}\right)\mathrm{d}P_X(x)\mathrm{d}P_{A^{[1,L-1]}}(a). \tag{5.1.7}$$

The normalizing factor $\mathcal{Z}^\circ_{\beta,L,n}$ is called the partition function. The central object to study is the free energy

$$F^\circ_{\beta,L,n} = \frac{1}{n}\log\mathcal{Z}^\circ_{\beta,L,n}. \tag{5.1.8}$$

To compute the limit of $\mathbb{E}F^\circ_{\beta,L,n}$ as $n\to\infty$, we make the following assumptions:

(H1) X has i.i.d. entries, and the law of X_1 is supported on $[-1,1]$, independent of n and satisfies that $X_1\neq 0$ with positive probability;

(H2) for every $l\in\{1,\dots,L\}$, φ_l is bounded, not identically zero, and continuously differentiable with bounded derivatives up to the 2^l-th order;

(H3) for every $l\in\{1,\dots,L\}$, $\Phi^{(l)}$ consists of independent standard Gaussian entries, and $(A^{(l)}_j)_{1\leqslant j\leqslant n_l}$ consists of i.i.d. $\mathbb{R}^{k_l}$-valued random vectors with a fixed law and bounded a.s.

To state the main result, we need more definitions. Throughout this work, we set

$$\mathbb{R}_+ = [0,\infty). \tag{5.1.9}$$

For every $l \in \{0, 1, \ldots, L\}$ and $n \in \mathbb{N}$, define

$$\rho_{l,n} = \frac{1}{n_l} \mathbb{E} \left| X^{(l)} \right|^2. \tag{5.1.10}$$

Due to Lemma 5.5.1 to be proved later, the following limit exists

$$\lim_{n \to \infty} \rho_{l,n} = \rho_l \tag{5.1.11}$$

for some $\rho_l > 0$ with explicit expression. Let P_{X_1} be the law of X_1 and Z_1' be a standard Gaussian random variable. Set

$$\Psi_0(r) = \mathbb{E} \log \int_{\mathbb{R}} e^{r X_1 x_1 + \sqrt{r} Z_1' x_1 - \frac{r}{2} |x_1|^2} \mathrm{d}P_{X_1}(x_1), \quad \forall r \in \mathbb{R}_+. \tag{5.1.12}$$

For every $l \in \{1, \ldots, L\}$, $\rho \geqslant 0$ and $h = (h_1, h_2) \in [0, \rho] \times \mathbb{R}_+$, define $\Psi_l(h; \rho)$ to be

$$\mathbb{E} \log \int \widetilde{\mathcal{P}}_{h_2,l} \left(\sqrt{h_2} \varphi_l \left(\sqrt{h_1} V_1 + \sqrt{\rho - h_1} W_1, A_1^{(l)} \right) + Z_1 \middle| \sqrt{h_1} V_1 + \sqrt{\rho - h_1} w \right) \mathrm{d}P_{W_1}(w), \tag{5.1.13}$$

where V_1, W_1, Z_1 are independent standard Gaussian random variables and

$$\widetilde{\mathcal{P}}_{h_2,l}(y|z) = \int_{\mathbb{R}^{k_l}} e^{-\frac{1}{2}|y - \sqrt{h_2}\varphi_l(z, a_1^{(l)})|^2} \mathrm{d}P_{A_1^{(l)}} \left(a_1^{(l)} \right), \quad \forall y, z \in \mathbb{R}. \tag{5.1.14}$$

Now, we are ready to state the main result.

Theorem 5.1.1. *Under assumptions (H1)–(H3), it holds that*

$$\lim_{n \to \infty} \mathbb{E} F^{\circ}_{\beta,L,n} = \sup_{z^{(L)}} \inf_{y^{(L)}} \sup_{z^{(L-1)}} \inf_{y^{(L-1)}} \cdots \sup_{z^{(1)}} \inf_{y^{(1)}} \phi_L \left(\beta; y^{(1)}, \cdots, y^{(L)}; z^{(1)}, \cdots, z^{(L)} \right) \tag{5.1.15}$$

where $\sup_{z^{(l)}}$ *is taken over* $z^{(l)} \in \mathbb{R}_+ \times [0, \frac{\alpha_{l-1}\rho_{l-1}}{2}]$, $\inf_{y^{(l)}}$ *is taken over* $y^{(l)} \in [0, \rho_{l-1}] \times \mathbb{R}_+$, *and*

$$\phi_L\left(\beta; y^{(1)}, \cdots, y^{(L)}; z^{(1)}, \cdots, z^{(L)}\right) \tag{5.1.16}$$
$$= \alpha_L \Psi_L\left(y_1^{(L)}, \beta; \rho_{L-1}\right) + \sum_{l=1}^{L-1} \alpha_l \Psi_l\left(y_1^{(l)}, y_2^{(l+1)}; \rho_{l-1}\right) + \Psi_0\left(y_2^{(1)}\right)$$
$$+ \sum_{l=1}^{L}\left(-y^{(l)} \cdot z^{(l)} + \frac{2}{\alpha_{l-1}} z_1^{(l)} z_2^{(l)}\right) + \sum_{l=2}^{L} \frac{\alpha_{l-1}}{2}\left(1 + \rho_{l-1} y_2^{(l)}\right).$$

We briefly comment on hypotheses (H1)–(H3).

The nonzero assumptions in (H1) and (H2) are reasonable in the setting of statistical inference where only non-constant signals are interesting. They are also purely technical in order to ensure that ρ_l in (5.1.11) is nonzero and thus some domain (defined in (5.2.1)) we work on is non-degenerate. In general, one can always consider a reduced model obtained from the original one by starting from the first layer after which all layers including itself contain nonzero signals. Alternatively, small constants can be added to fulfill the nonzero assumptions, and the effect of these constants are traceable through explicit formulae.

The assumption that X has i.i.d. entries in (H1) and the assumption on the differentiability of φ_l in (H2) are mainly used in deriving concentration results in Section 5.5. We believe that results similar to Theorem 5.1.1 are still valid under different or weaker assumptions. For instance, when X is uniformly distributed on the centered n-sphere with radius $\sqrt{n}$, concentration results needed here are expected to hold. The high order of differentiability in (H2) is needed in an iterative application of the Gaussian integration by parts due to the presence of multiple layers. We remark that in the 2-layer setting, a careful treatment only needs φ_1 and φ_2 to be twice continuously differentiable with bounded derivatives, as done in [65], while (H2) requires φ_2 to be continuously differentiable up to the fourth order. Since we are considering general cases, we resort to (H2) for convenience.

On the other hand, many results in this work do not require assumptions as strong as (H1) and (H2). Hence, whenever possible, we will instead assume the following, together with (H3):

(h1) for every $n \in \mathbb{N}$, $|X| \leqslant \sqrt{n}$ a.s.;

(h2) for every $l \in \{1, \ldots, L\}$, φ_l is bounded and twice continuously differentiable with bounded derivatives.

5.1.2. Related works

Generalized linear models are relevant in many fields including signal processing, statistical learning, and neural networks. Its multi-layer setup models a type of feed-forward neural network, which captures some of the key features of deep learning. For more details on these connections, we refer to [11, 65] and references therein. Recent progress in rigorous studies of information-theoretical aspects of these models have been made using methods originated from statistical physics. The mutual information of a model, a key quantity in these investigations, is related to the free energy via a simple additive relation. Therefore, the high-dimensional limit of the free energy is the central object in these approaches. Variational formulae for the free energy have been rigorously proven in the one-layer setting in [11] and the two-layer setting in [65].

The two works just mentioned above employed the powerful adaptive interpolation method introduced in [12, 13], which can be seen as an evolution from the classic interpolation method in statistical physics. This new method has proven to be successful and versatile in treating many different models and settings [52, 14, 86, 85, 103].

The approach adopted in this work is based on identifying an enriched version of the original free energy with a solution to a certain Hamilton–Jacobi equation determined by the model. This approach was first introduced in [95, 94] and has been applied also to the study of spin glass models [97, 98, 96, 93]. Similar considerations in physics also appeared in [69, 71, 22, 21].

In treating statistical inference problems, two notions of solutions have been considered. One is the viscosity solution used in [95, 39, 37], and the other is the weak solution in [94, 36, 39]. In this paper, we take the latter approach due to the convenience and simplicity in dealing with boundary conditions under the notion of weak solutions.

Compared with [94, 36, 39], the novelty here lies in an iterative argument to treat the multi-layer setting. Let us explain this briefly. After enriching the L-layer model and verifying some concentration results, we can show that the corresponding free energy converges to the unique solution of a certain Hamilton–Jacobi equation whose initial condition is determined by the limiting free energy associated with the $(L-1)$-layer model. Then, the desired result naturally follows from an iteration of this result applied to each layer. Apart from this, different from [94, 36, 39], the Hamilton–Jacobi equation considered here is defined over a domain where the range of spacial variables depends on time. Accordingly, treatments used previously have to be adjusted.

The rest of the paper is organized as follows. In Section 5.2, we enrich the model and derive that the enriched free energy satisfies an approximate Hamilton–Jacobi equation. We also record some basic properties of the derivatives of the free energy. In Section 5.3, we give the definition of weak solutions and prove the existence and uniqueness. In particular, the existence is furnished by a variational formula known as the Hopf formula. Using these, we prove the key convergence result of the enriched free energy in Section 5.4, which is used in an iterative argument to prove Theorem 5.1.1. Lastly, we collect auxiliary results in Section 5.5, including the convergence in (5.1.11), concentration of the norm of $X^{(L)}$, and concentration of the free energy.

Acknowledgement

We warmly thank Jean–Christophe Mourrat for many helpful comments and discussions.

5.2. Approximate Hamilton–Jacobi equations

In this section, we enrich the model and derive that the associated free energy satisfies an approximate Hamilton–Jacobi equation, which is stated in Proposition 5.2.1. We also record basic properties of derivatives of the free energy in Lemma 5.2.2.

5.2.1. Enrichment

Recall the notation $\mathbb{R}_+$ in (5.1.9) and $\rho_{l,n}$ defined in (5.1.10). For $\rho > 0$, define

$$\Omega_\rho = \{(t, h_1, h_2) \in \mathbb{R}^3_+ : h_1 \leqslant \rho(1-t),\ t \leqslant 1\} \tag{5.2.1}$$

where there is no restriction on h_2. For $(t, h) \in \Omega_{\rho_{L-1,n}}$, define

$$S = \sqrt{\frac{t}{n_{L-1}}}\Phi^{(L)}X^{(L-1)} + \sqrt{h_1}V + \sqrt{\rho_{L-1,n} - \rho_{L-1,n}t - h_1}W, \tag{5.2.2}$$

$$s = \sqrt{\frac{t}{n_{L-1}}}\Phi^{(L)}x^{(L-1)} + \sqrt{h_1}V + \sqrt{\rho_{L-1,n} - \rho_{L-1,n}t - h_1}w, \tag{5.2.3}$$

$$Y = \sqrt{\beta}\varphi_L\left(S, A^{(L)}\right) + Z, \tag{5.2.4}$$

$$Y' = \sqrt{h_2}X^{(L-1)} + Z', \tag{5.2.5}$$

where $w \in \mathbb{R}^{n_L}$, $x^{(L-1)}$ is given in (5.1.5), V, W are independent n_L-dimensional standard Gaussian vectors, Z is given in (5.1.4), and Z' is an n_{L-1}-dimensional standard Gaussian vector. Due to (5.1.5) and (5.2.3), s depends on $(x, w, a, \Phi^{[1,L]}, V)$.

Recall $\mathcal{P}_{\beta,L,n}$ given in (5.1.6). We introduce the following Hamiltonian

$$H_{\beta,L,n}(x, w, a) = \log \mathcal{P}_{\beta,L,n}\big(Y\big|s\big) + \sqrt{h_2}Y' \cdot x^{(L-1)} - \frac{h_2}{2}\left|x^{(L-1)}\right|^2, \tag{5.2.6}$$

where $x \in \mathbb{R}^n$, $w \in \mathbb{R}^{n_L}$, a and $x^{(L-1)}$ are given in (5.1.5). Define the associated partition function

$$\mathcal{Z}_{\beta,L,n} = \int e^{H_{\beta,L,n}(x,w,a)}\mathrm{d}P_X(x)\mathrm{d}P_W(w)\mathrm{d}P_{A^{[1,L-1]}}(a) \tag{5.2.7}$$

and consider the corresponding free energy

$$F_{\beta,L,n} = \frac{1}{n}\log \mathcal{Z}_{\beta,L,n} \tag{5.2.8}$$

and $\overline{F}_{\beta,L,n} = \mathbb{E}F_{\beta,L,n}$ where $\mathbb{E}$ is over $Y, Y', V, \Phi^{[1,L]}$ (recall that $x^{(L-1)}$ depends on $\Phi^{[1,L-1]}$ as in (5.1.5)). The domain of $F_{\beta,L,n}$ is $\Omega_{\rho_{L-1,n}}$ defined in (5.2.1).

We often make the dependence of $F_{\beta,L,n}$ on $(t,h) \in \Omega_{\rho_{L-1,n}}$ explicit, and write $F_{\beta,L,n}(t,h)$. Comparing with the definitions of $\mathcal{Z}^{\circ}_{\beta,L,n}$ in (5.1.7) and $F^{\circ}_{\beta,L,n}$ in (5.1.8), we can verify that $\mathcal{Z}^{\circ}_{\beta,L,n} = \mathcal{Z}_{\beta,L,n}(1,0)$ and $F^{\circ}_{\beta,L,n} = F_{\beta,L,n}(1,0)$ evaluated at $t=1, h=0$. Hence, we view $F_{\beta,L,n}$ as the free energy associated with an enriched model. Note that the following holds

$$\mathbb{E}F^{\circ}_{\beta,L,n} = \overline{F}_{\beta,L,n}(1,0). \tag{5.2.9}$$

Throughout this work, we interpret t as the "temporal variable" and $h = (h_1, h_2)$ as the "spacial variable". Moreover, we use the short hand notation $\partial_i = \partial_{h_i}$ for $i = 1,2$, and denote by $\nabla = (\partial_1, \partial_2)$ the gradient operator. Define $\mathsf{H}_L : \mathbb{R}^2 \to \mathbb{R}$ by

$$\mathsf{H}_L(p) = \frac{2}{\alpha_{L-1}} p_1 p_2. \tag{5.2.10}$$

The main goal is to prove the following proposition.

Proposition 5.2.1. *Assume (h1), (h2) and (H3) for some $L \in \mathbb{N}$. For every $\beta \geqslant 0$ and every $n \in \mathbb{N}$, the function $(t,h) \mapsto \overline{F}_{\beta,L,n}(t,h)$ is differentiable in $\Omega_{\rho_{L-1,n}} \setminus \{h_1 = \rho_{L-1,n}(1-t)\}$ and there is a constant C such that, for all $(t,h) \in \Omega_{\rho_{L-1,n}} \setminus \{h_1 = \rho_{L-1,n}(1-t)\}$,*

$$\left|\partial_t \overline{F}_{\beta,L,n} - \mathsf{H}_L\left(\nabla \overline{F}_{\beta,L,n}\right)\right| \leqslant C\left(\frac{1}{n}\partial_2^2 \overline{F}_{\beta,L,n} + \mathbb{E}\big(\partial_2 F_{\beta,L,n} - \partial_2 \overline{F}_{\beta,L,n}\big)^2\right)^{\frac{1}{2}} + a_n,$$

where

$$a_n \leqslant C\left(n\mathbb{E}\left(\frac{\left|X^{(L-1)}\right|^2}{n_{L-1}} - \rho_{L-1,n}\right)^2\right)^{\frac{1}{2}} \left(\mathbb{E}\left(F_{\beta,L,n} - \overline{F}_{\beta,L,n}\right)^2\right)^{\frac{1}{2}} + C\left|\frac{n_{L-1}}{n} - \alpha_{L-1}\right|. \tag{5.2.11}$$

This suggests that the limiting Hamilton–Jacobi equation should be

$$\partial_t f - \mathsf{H}_L(\nabla f) = 0, \tag{5.2.12}$$

which will be studied in the next section.

5.2.2. Proof of Proposition 5.2.1

Recall $\mathcal{P}_{\beta,L,n}$ defined in (5.1.6). For simplicity of notation, we write $H = H_{\beta,L,n}$, $\mathcal{Z} = \mathcal{Z}_{\beta,L,n}$, $F = F_{\beta,L,n}$, $\mathcal{P} = \mathcal{P}_{\beta,L,n}$ and $\rho = \rho_{L-1,n}$. For any measurable function $g : \mathbb{R}^n \times \mathbb{R}^{n_L} \times (\prod_{l=1}^{L-1} \mathbb{R}^{n_l \times k_l}) \to \mathbb{R}$, we define

$$\langle g(x,w,a)\rangle = \frac{1}{\mathcal{Z}} \int g(x,w,a) e^{H(x,w,a)} \mathrm{d}P_X(x)\mathrm{d}P_W(w)\mathrm{d}P_{A^{[1,L-1]}}(a).$$

In other words, $\langle\,\cdot\,\rangle$ is the Gibbs measure with Hamiltonian H and reference measure $\mathrm{d}P_X(x)\mathrm{d}P_W(w)\mathrm{d}P_{A^{[1,L-1]}}(a)$.

Preliminaries

We will repeatedly use two basic tools in our computations: the Gaussian integration by parts and the Nishimori identity. The simplest form of the Gaussian integration by parts can be stated as follows. For a standard Gaussian random variable U and a differentiable function $g : \mathbb{R} \to \mathbb{R}$ satisfying $\mathbb{E}|g'(U)| < \infty$, it holds that

$$\mathbb{E}[Ug(U)] = \mathbb{E}g'(U),$$

which can be seen easily by rewriting the expectation as an integration with respect to the Gaussian density and performing the classic integration by parts. For the purpose of this work, a straightforward extension of the above to standard Gaussian vectors is sufficient.

Using the definition of H and Bayes' rule, we can see that the law of $X, W, A^{[1,L-1]}$ conditioned on $Y, Y', V, \Phi^{[1,L]}$ is given exactly by the Gibbs measure $\langle \cdot \rangle$, namely,

$$\left\langle g\left(x, w, a, Y, Y', V, \Phi^{[1,L]}\right)\right\rangle = \mathbb{E}\left[g\left(X, W, A^{[1,L-1]}, Y, Y', V, \Phi^{[1,L]}\right)\Big| Y, Y', V, \Phi^{[1,L]}\right],$$

for suitable measurable function g. The above immediately implies the Nishimori identity that, for suitable g,

$$\mathbb{E}\left\langle g\left(x, w, a, Y, Y', V, \Phi^{[1,L]}\right)\right\rangle = \mathbb{E}g\left(X, W, A^{[1,L-1]}, Y, Y', V, \Phi^{[1,L]}\right).$$

Independent copies of (x, w, a) with respect to the Gibbs measure are called replicas and often denoted as (x', w', a'), (x'', w'', a''), etc. When multiple replicas are present, the above identity can be extended in a straightforward way allowing us to replace one set of the replicas by $(X, W, A^{[1,L-1]})$, and vice versa. For instance, we have that

$$\mathbb{E}\left\langle g\left(x, w, a, x', w', a', Y, Y', V, \Phi^{[1,L]}\right)\right\rangle = \mathbb{E}\left\langle g\left(x, w, a, X, W, A^{[1,L-1]}, Y, Y', V, \Phi^{[1,L]}\right)\right\rangle.$$

Computation of $\partial_t \overline{F}$

Recall $H(x, w, a)$ in (5.2.6) and let us also write

$$H(x, w, a; y, y') = \log \mathcal{P}(y|s) + \sqrt{h_2} y' \cdot x^{(L-1)} - \frac{h_2}{2}\left|x^{(L-1)}\right|^2.$$

Hence, we have that $H(x, w, a) = H(x, w, a; Y, Y')$, and for each fixed x, w, y, y', the only randomness of $H(x, w, a; y, y')$ comes from $\Phi^{[1,L]}$ (in s and $x^{(L-1)}$) and V (in s).

We can verify that the conditioned law of (Y, Y') given $(\Phi^{[1,L]}, V)$ is given by

$$\left(\frac{1}{(2\pi)^{\frac{n_L}{2}}} \int e^{H(x,w,a;y,y')} \mathrm{d}P_X(x) \mathrm{d}P_W(w) \mathrm{d}P_{A^{[1,L-1]}}(a)\right) \mathrm{d}y \mathrm{d}y', \tag{5.2.13}$$

where we recall that W is Gaussian. Recall the partition function (5.2.7) and we introduce

$$\mathcal{Z}(y,y') = \int e^{H(x,w,a;y,y')} \mathrm{d}P_X(x)\mathrm{d}P_W(w)\mathrm{d}P_{A^{[1,L-1]}}(a).$$

Then, note that $\mathcal{Z} = \mathcal{Z}(Y,Y')$ and the only randomness of $\mathcal{Z}(y,y')$ is from $\Phi^{[1,L]}$ and V.

We introduce the shorthand notation

$$\mathrm{d}\widetilde{P}_{y,y'} = \frac{1}{(2\pi)^{\frac{n_L}{2}}} \mathrm{d}y\,\mathrm{d}y'\,\mathrm{d}P_X(\widetilde{x})\,\mathrm{d}P_W(\widetilde{w})\,\mathrm{d}P_{A^{[1,L-1]}}(\widetilde{a}) \tag{5.2.14}$$

which is a measure that integrates y, y' and all variables with tildes $\widetilde{x}, \widetilde{w}, \widetilde{a}$. Using these and (5.2.8), we can write that

$$\overline{F} = \frac{1}{n}\mathbb{E}\Big[\int e^{H(\widetilde{x},\widetilde{w},\widetilde{a};y,y')} \log \mathcal{Z}(y,y')\mathrm{d}\widetilde{P}_{y,y'}\Big] \tag{5.2.15}$$

where the expectation $\mathbb{E}$ is taken over the remaining randomness, namely, $\Phi^{[1,L]}$ and V. To lighten the notation further, we write $H(\widetilde{-};y,y') = H(\widetilde{x},\widetilde{w},\widetilde{a};y,y')$ and $H(-;y,y') = H(x,w,a;y,y')$.

Due to the dependence of $H(\widetilde{-};y,y')$ and $\mathcal{Z}(y,y')$ on t, differentiating $\overline{F}$ as in (5.2.15) with respect to t yields that

$$\begin{aligned}
\partial_t \overline{F} &= \frac{1}{n}\mathbb{E}\Big[\int \mathrm{d}\widetilde{P}_{y,y'}\Big(\partial_t H(\widetilde{-};y,y')\Big)e^{H(\widetilde{-};y,y')} \log \mathcal{Z}(y,y')\Big] \\
&\quad + \frac{1}{n}\mathbb{E}\Big[\langle \partial_t H(-;y,y')\big|_{y=Y,\,y'=Y'}\rangle\Big] \\
&= \mathtt{I}_t + \mathtt{II}_t.
\end{aligned} \tag{5.2.16}$$

Here on the second line, the Gibbs measure is the one associated with the Hamiltonian (5.2.6) and thus only integrates over the variables x, w, a. To evaluate the above, we define

$$u_y(x) = \log \mathcal{P}(y|x) \tag{5.2.17}$$

and denote by ∇u_y and Δu_y the gradient and Laplacian of u_y with differentiation in x, respectively. Then, using (5.2.3) and (5.2.6), we can compute that

$$\begin{aligned} \partial_t H(-; y, y') &= \big(\partial_t s\big) \cdot \nabla u_y(s) \\ &= \frac{1}{2}\bigg(\frac{1}{\sqrt{tn_{L-1}}}\Phi^{(L)}x^{(L-1)} - \frac{\rho}{\sqrt{\rho(1-t)-h_1}}w\bigg) \cdot \nabla u_y(s). \end{aligned} \tag{5.2.18}$$

We write $\widetilde{s}$ and $\widetilde{x}^{(L-1)}$ to be s and $x^{(L-1)}$, respectively, with x, w, a therein replaced by $\widetilde{x}, \widetilde{w}, \widetilde{a}$. Hence, we have that $\mathtt{I}_t$ is equal to

$$\frac{1}{2n}\mathbb{E}\bigg[\int \mathrm{d}\widetilde{P}_{y,y'}\bigg(\frac{1}{\sqrt{tn_{L-1}}}\Phi^{(L)}\widetilde{x}^{(L-1)} - \frac{\rho}{\sqrt{\rho(1-t)-h_1}}\widetilde{w}\bigg) \cdot \nabla u_y(\widetilde{s})e^{H(\widetilde{-};y,y')}\log \mathcal{Z}(y,y')\bigg].$$

Recall that $\widetilde{s}$ and $H(\widetilde{-}; y, y')$ depend on $\Phi^{(L)}$ and $\widetilde{w}$, and that $\mathcal{Z}(y, y')$ depends on $\Phi^{(L)}$. Since $\widetilde{w}$ under $\mathrm{d}\widetilde{P}_{y,y'}$ and $\Phi^{(L)}$ under $\mathbb{E}$ are standard Gaussian vectors, we can obtain by performing the Gaussian integration by parts with one $\widetilde{w}$ and $\Phi^{(L)}$ that

$$\begin{aligned} \mathtt{I}_t = a_n' &+ \frac{1}{2n}\mathbb{E}\bigg[\frac{1}{\mathcal{Z}(y,y')}\int \mathrm{d}\widetilde{P}_{y,y'}\mathrm{d}P_X(x)\mathrm{d}P_W(w)\mathrm{d}P_{A^{[1,L-1]}}(a) \\ &\bigg(\frac{1}{n_{L-1}}\widetilde{x}^{(L-1)} \cdot x^{(L-1)}\bigg)\big(\nabla u_y(\widetilde{s}) \cdot \nabla u_y(s)\big)e^{H(\widetilde{-};y,y')}e^{H(-;y,y')}\bigg] \\ &= a_n' + \frac{1}{2}\mathbb{E}\bigg\langle\bigg(\frac{1}{n_{L-1}}X^{(L-1)} \cdot x^{(L-1)}\bigg)\bigg(\frac{1}{n}\nabla u_Y(S) \cdot \nabla u_Y(s)\bigg)\bigg\rangle \end{aligned} \tag{5.2.19}$$

where

$$\begin{aligned} a_n' &= \frac{1}{2n}\mathbb{E}\bigg[\int \mathrm{d}\widetilde{P}_{y,y'}\bigg(\frac{1}{n_{L-1}}\Big|\widetilde{x}^{(L-1)}\Big|^2 - \rho\bigg)\big(\Delta u_y(\widetilde{s}) + |\nabla u_y(\widetilde{s})|^2\big)e^{H(\widetilde{-};y,y')}\log \mathcal{Z}(y,y')\bigg] \\ &= \frac{1}{2n}\mathbb{E}\bigg[\bigg(\frac{1}{n_{L-1}}\Big|X^{(L-1)}\Big|^2 - \rho\bigg)\big(\Delta u_Y(S) + |\nabla u_Y(S)|^2\big)\log \mathcal{Z}(Y,Y')\bigg]. \end{aligned} \tag{5.2.20}$$

Here, in deriving (5.2.19) and (5.2.20), we used (5.2.14) and the observation that replacing $\widetilde{x}, \widetilde{w}, \widetilde{a}$ by $X, W, A^{[1,L-1]}$ in $\widetilde{x}^{(L-1)}, \widetilde{s}$ yields $X^{(L-1)}, S$. We claim that $\mathtt{II}_t = 0$ and postpone its proof. Then, combining the above gives that

$$\partial_t \overline{F} = \frac{1}{2}\mathbb{E}\Big\langle \Big(\frac{1}{n_{L-1}} X^{(L-1)} \cdot x^{(L-1)}\Big)\Big(\frac{1}{n}\nabla u_Y(S) \cdot \nabla u_Y(s)\Big)\rangle + a_n'. \tag{5.2.21}$$

Computation of $\partial_1 \overline{F}$

Similarly, by (5.2.15), we have that

$$\begin{aligned}\partial_1 \overline{F} &= \frac{1}{n}\mathbb{E}\Big[\int \mathrm{d}\widetilde{P}_{y,y'}\Big(\partial_1 H(\widetilde{-}; y, y')\Big) e^{H(\widetilde{-};y,y')} \log \mathcal{Z}(y, y')\Big] \\ &\quad + \frac{1}{n}\mathbb{E}\Big[\langle \partial_1 H(-; y, y')\big|_{y=Y,\, y'=Y'}\rangle\Big] \\ &= \mathtt{I}_{h_1} + \mathtt{II}_{h_1}.\end{aligned} \tag{5.2.22}$$

To compute $\mathtt{I}_{h_1}$, we start with

$$\begin{aligned}\partial_1 H(-; y, y') &= \big(\partial_1 s\big) \cdot \nabla u_y(s) \\ &= \frac{1}{2}\Big(\frac{1}{\sqrt{h_1}} V - \frac{1}{\sqrt{\rho(1-t) - h_1}} w\Big) \cdot \nabla u_y(s),\end{aligned} \tag{5.2.23}$$

which gives that

$$\mathtt{I}_{h_1} = \frac{1}{2n}\mathbb{E}\Big[\int \mathrm{d}\widetilde{P}_{y,y'}\Big(\frac{1}{\sqrt{h_1}} V - \frac{1}{\sqrt{\rho(1-t) - h_1}} \widetilde{w}\Big) \cdot \nabla u_y(\widetilde{s}) e^{H(\widetilde{-};y,y')} \log \mathcal{Z}(y, y')\Big].$$

Using Gaussian integration by parts on V and $\widetilde{w}$, we obtain that

$$\begin{aligned}\mathtt{I}_{h_1} = \frac{1}{2n}\mathbb{E}\Big[\int \mathrm{d}\widetilde{P}_{y,y'}\big(1-1\big)\Big(\Delta u_y(\widetilde{s}) + |\nabla u_y(\widetilde{s})|^2\Big) e^{H(\widetilde{-};y,y')} \log \mathcal{Z}(y, y')\Big] \\ + \frac{1}{2n}\mathbb{E}\Big[\frac{1}{\mathcal{Z}(y, y')}\int \mathrm{d}\widetilde{P}_{y,y'} \mathrm{d}P_X(x) \mathrm{d}P_W(w) \mathrm{d}P_{A^{[1,L-1]}}(a) \\ \nabla u_y(\widetilde{s}) \cdot \nabla u_y(s) e^{H(\widetilde{-};y,y')} e^{H(-;y,y')}\Big] \\ = \frac{1}{2n}\mathbb{E}\left\langle \nabla u_Y(S) \cdot \nabla u_Y(s)\right\rangle.\end{aligned}$$

Here, in the last equality, we used the same argument as in obtaining (5.2.19). Again, we claim that $\mathtt{II}_{h_1} = 0$ and postpone its proof. This together with the above yields that

$$\partial_1 \overline{F} = \frac{1}{2}\mathbb{E}\left\langle \frac{1}{n}\nabla u_Y(S) \cdot \nabla u_Y(s) \right\rangle. \tag{5.2.24}$$

Computation of $\partial_2 \overline{F}$

Using (5.2.6), (5.2.7) and (5.2.8), we can compute that

$$\partial_2 F = \frac{1}{n}\langle \partial_2 H(x,w,a)\rangle = \frac{1}{2n}\langle 2X^{(L-1)} \cdot x^{(L-1)} + \frac{1}{\sqrt{h_2}} Z' \cdot x^{(L-1)} - x^{(L-1)} \cdot x^{(L-1)}\rangle. \tag{5.2.25}$$

Using Gaussian integration by parts on Z' and the Nishimori identity, we get that

$$\begin{aligned}\partial_2 \overline{F} &= \frac{1}{2n}\mathbb{E}\langle 2X^{(L-1)} \cdot x^{(L-1)} + \left(x^{(L-1)} - x'^{(L-1)}\right) \cdot x^{(L-1)} - x^{(L-1)} \cdot x^{(L-1)}\rangle \\ &= \frac{1}{2n}\mathbb{E}\langle X^{(L-1)} \cdot x^{(L-1)}\rangle,\end{aligned} \tag{5.2.26}$$

where $x'^{(L-1)}$ is a replica of $x^{(L-1)}$ obtained by replacing x, a in (5.1.5) by replicas x', a'.

Deriving the equation

By (5.2.10), (5.2.24) and (5.2.26), we have

$$\begin{aligned}&\left| \mathsf{H}_L(\nabla \overline{F}) - \frac{1}{2}\mathbb{E}\left\langle \frac{1}{n_{L-1}} X^{(L-1)} \cdot x^{(L-1)} \right\rangle \mathbb{E}\left\langle \frac{1}{n}\nabla u_Y(S) \cdot \nabla u_Y(s) \right\rangle \right| \\ &\qquad = \left| 1 - \frac{\alpha_{L-1} n}{n_{L-1}} \right| \left| \mathsf{H}_L(\nabla \overline{F}) \right|.\end{aligned}$$

By (5.2.46) and (5.2.47) both proved later and assumption (5.1.1), the above is bounded by $C|\frac{n_{L-1}}{n} - \alpha_{L-1}|$. This along with (5.2.21) implies that

$$\left|\partial_t \overline{F} - \mathsf{H}_L(\nabla \overline{F})\right| \leqslant \frac{1}{2}\sqrt{b_n} + |a_n'| + C\left|\frac{n_{L-1}}{n} - \alpha_{L-1}\right|.$$

where

$$b_n = \mathbb{V}\mathrm{ar}_{\mathbb{E}\langle\,\cdot\,\rangle}\left[\frac{1}{n_{L-1}}X^{(L-1)}\cdot x^{(L-1)}\right]\mathbb{V}\mathrm{ar}_{\mathbb{E}\langle\,\cdot\,\rangle}\left[\frac{1}{n}\nabla u_Y(S)\cdot\nabla u_Y(s)\right]$$

with variances taken with respect to $\mathbb{E}\langle\,\cdot\,\rangle$. Then, the desired results follows, once we prove that

$$|a_n'| \leqslant C\left(n\mathbb{E}\left(\frac{\left|X^{(L-1)}\right|^2}{n_{L-1}}-\rho\right)^2\right)^{\frac{1}{2}}\left(\mathbb{E}\left(F-\overline{F}\right)^2\right)^{\frac{1}{2}}, \tag{5.2.27}$$

$$\mathbb{V}\mathrm{ar}_{\mathbb{E}\langle\,\cdot\,\rangle}\left[\frac{1}{n}\nabla u_Y(S)\cdot\nabla u_Y(s)\right] \leqslant C, \tag{5.2.28}$$

$$\mathbb{V}\mathrm{ar}_{\mathbb{E}\langle\,\cdot\,\rangle}\left[\frac{1}{n_{L-1}}X^{(L-1)}\cdot x^{(L-1)}\right] \leqslant C\left(\frac{1}{n}\partial_2^2\overline{F}+\mathbb{E}\big(\partial_2 F-\partial_2\overline{F}\big)^2\right). \tag{5.2.29}$$

To complete the proof, it remains to verify that $\mathtt{II}_t = \mathtt{II}_{h_1} = 0$ and prove the above assertions.

Evaluating $\mathtt{II}_t$ and $\mathtt{II}_{h_1}$

Recall the definition of $\mathtt{II}_t$ in (5.2.16). By the Nishimori identity, we have that

$$\mathtt{II}_t = \frac{1}{n}\mathbb{E}\left\langle \partial_t H(x,w,a;y,y')|_{y=Y,\,y'=Y'}\right\rangle = \frac{1}{n}\mathbb{E}\left[\partial_t H\left(X,W,A^{[1,L-1]};y,y'\right)\Big|_{y=Y,\,y'=Y'}\right].$$

Using (5.2.18) and the conditional law of (Y,Y') in (5.2.13) together with the notation $\mathrm{d}\widetilde{P}_{y,y'}$ given in (5.2.14), we obtain that

$$\begin{aligned}
\mathtt{II}_t &= \frac{1}{2n}\mathbb{E}\left[\left(\frac{1}{\sqrt{tn_{L-1}}}\Phi^{(L)}X^{(L-1)}-\frac{\rho}{\sqrt{\rho(1-t)-h_1}}W\right)\cdot\nabla u_Y(S)\right]\\
&= \frac{1}{2n}\mathbb{E}\left[\int \mathrm{d}\widetilde{P}_{y,y'}e^{H(\widetilde{\cdot};y,y')}\left(\frac{1}{\sqrt{tn_{L-1}}}\Phi^{(L)}\widetilde{x}^{(L-1)}-\frac{\rho}{\sqrt{\rho(1-t)-h_1}}\widetilde{w}\right)\cdot\nabla u_y(\widetilde{s})\right]\\
&= \frac{1}{2n}\mathbb{E}\left[\int \mathrm{d}\widetilde{P}_{y,y'}e^{H(\widetilde{\cdot};y,y')}\left(\frac{1}{n_{L-1}}\left|\widetilde{x}^{(L-1)}\right|^2-\rho\right)\left(\Delta u_y(\widetilde{s})+|\nabla u_y(\widetilde{s})|^2\right)\right]
\end{aligned}$$

where in the third equality we used the Gaussian integration by parts on $\Phi^{(L)}$ and $\widetilde{w}$ (recall that under $\mathrm{d}\widetilde{P}_{y,y'}$, $\widetilde{w}$ is a standard Gaussian vector).

Due to the definition of u_y in (5.2.17), we can compute that

$$\Delta u_y(\widetilde{s}) + |\nabla u_y(\widetilde{s})|^2 = \frac{\Delta \mathcal{P}(y|\widetilde{s})}{\mathcal{P}(y|\widetilde{s})}, \tag{5.2.30}$$

where we recall that all derivatives are carried out in the second argument. Hence, we get that

$$\mathtt{II}_t = \frac{1}{2n}\mathbb{E}\left[\left(\frac{1}{n_{L-1}}\left|X^{(L-1)}\right|^2 - \rho\right)\mathbb{E}\left[\frac{\Delta\mathcal{P}(Y|S)}{\mathcal{P}(Y|S)}\middle|X^{(L-1)}, S\right]\right]. \tag{5.2.31}$$

In view of the definition of Y in (5.2.4) and the formula for $\mathcal{P}$ in (5.1.6), we can see that, conditioned on $X^{(L-1)}, S$, the law of Y has a Lebesgue density given by $(2\pi)^{-\frac{n_L}{2}}\mathcal{P}(y|S)$, namely, for any bounded measurable function g,

$$\mathbb{E}\left[g\left(Y, X^{(L-1)}, S\right)\middle|X^{(L-1)}, S\right] = \frac{1}{(2\pi)^{\frac{n_L}{2}}}\int g\left(y, X^{(L-1)}, S\right)\mathcal{P}(y|S)\mathrm{d}y. \tag{5.2.32}$$

Let us write

$$\Delta\mathcal{P}(y|S) = \sum_{j=1}^{n_L}\partial_j^2\mathcal{P}(y|S) \tag{5.2.33}$$

where again the derivatives are in the second argument. We can compute that

$$\partial_j^2\mathcal{P}(y|S) = \int \Gamma_j\left(y_j, S_j, a_j^{(L)}\right)e^{-\frac{1}{2}|y-\sqrt{\beta}\varphi_L(S,a^{(L)})|^2}\mathrm{d}P_{A^{(L)}}\left(a^{(L)}\right) \tag{5.2.34}$$

with

$$\Gamma_j\left(y_j, S_j, a_j^{(L)}\right) = \beta\left(\left(y_j - \sqrt{\beta}\varphi_j\right)^2 - 1\right)(\varphi_j')^2 + \sqrt{\beta}\left(y_j - \sqrt{\beta}\varphi_j\right)\varphi_j'' \tag{5.2.35}$$

where we used the shorthand notation $\varphi_j = \varphi_L(S_j, a_j^{(L)})$, $\varphi_j' = \varphi_L'(S_j, a_j^{(L)})$, and $\varphi_j'' = \varphi_L''(S_j, a_j^{(L)})$. Recall that φ_L acts component-wise on $(S, a^{(L)})$, namely, $\varphi_L(S, a^{(L)}) = (\varphi_L(S_j, a_j^{(L)}))_{1\leqslant j\leqslant n_L}$. Using this and the assumption that $(A_j^{(L)})_{1\leqslant j\leqslant n_L}$ are i.i.d. as in (H3), we have that

$$\frac{\partial_j^2 \mathcal{P}(y|S)}{\mathcal{P}(y|S)} = \frac{\int \Gamma_j\left(y_j, S_j, a_j^{(L)}\right) e^{-\frac{1}{2}|y_j - \sqrt{\beta}\varphi_L(S_j, a_j^{(L)})|^2} \mathrm{d}P_{A_j^{(L)}}\left(a_j^{(L)}\right)}{\int e^{-\frac{1}{2}|y_j - \sqrt{\beta}\varphi_L(S_j, a_j^{(L)})|^2} \mathrm{d}P_{A_j^{(L)}}\left(a_j^{(L)}\right)}. \tag{5.2.36}$$

Using this, (5.2.34) and (5.2.35), we can see that

$$\frac{1}{(2\pi)^{\frac{n_L}{2}}} \int \frac{\partial_j^2 \mathcal{P}(y|S)}{\mathcal{P}(y|S)} \mathcal{P}(y|S)\mathrm{d}y = 0, \tag{5.2.37}$$

$$\frac{1}{(2\pi)^{\frac{n_L}{2}}} \int \frac{\partial_i^2 \mathcal{P}(y|S)}{\mathcal{P}(y|S)} \frac{\partial_j^2 \mathcal{P}(y|S)}{\mathcal{P}(y|S)} \mathcal{P}(y|S)\mathrm{d}y = 0, \quad i \neq j. \tag{5.2.38}$$

The second equation will be used later. Now, by (5.2.32) and (5.2.37), we have that

$$\mathbb{E}\left[\frac{\partial_j^2 \mathcal{P}(Y|S)}{\mathcal{P}(Y|S)} \middle| X^{(L-1)}, S\right] = 0, \quad \forall j \in \{1, \ldots, n_L\},$$

which together with (5.2.31) implies that $\mathtt{II}_t = 0$.

It remains to show $\mathtt{II}_{h_1} = 0$. Recall the definition of $\mathtt{II}_{h_1}$ in (5.2.22). The Nishimori identity gives that

$$\mathtt{II}_{h_1} = \frac{1}{n}\mathbb{E}\left[\partial_1 H\left(X, W, A^{[1,L-1]}; y, y'\right)\Big|_{y=Y,\, y'=Y'}\right].$$

Using (5.2.23) and a similar argument used above, we have that

$$\begin{aligned}
\mathtt{II}_{h_1} &= \frac{1}{2n}\mathbb{E}\Bigg[\int \mathrm{d}\widetilde{P}_{y,y'}\left(\frac{1}{\sqrt{h_1}}V - \frac{1}{\sqrt{\rho(1-t)-h_1}}\widetilde{w}\right) \cdot \nabla u_y(\widetilde{s}) e^{H(\widetilde{\cdot}; y, y')}\Bigg] \\
&= \frac{1}{2n}\mathbb{E}\left[\int \mathrm{d}\widetilde{P}_{y,y'}\,(1-1)\left(\Delta u_y(\widetilde{s}) + |\nabla u_y(\widetilde{s})|^2\right) e^{H(\widetilde{\cdot}; y, y')}\right] = 0
\end{aligned}$$

where the second equality follows from the Gaussian integration by parts applied to V and $\widetilde{w}$.

Proof of (5.2.27)

Using (5.2.8) and a computation similar to (5.2.30), we rewrite a_n' in (5.2.20) as

$$a_n' = \frac{1}{2}\mathbb{E}\left[\left(\frac{1}{n_{L-1}}\left|X^{(L-1)}\right|^2 - \rho\right)\frac{\Delta\mathcal{P}(Y|S)}{\mathcal{P}(Y|S)}F\right].$$

Since $\mathtt{II}_t = 0$ as shown above, using the formula (5.2.31), we then have

$$a_n' = \frac{1}{2}\mathbb{E}\left[\left(\frac{1}{n_{L-1}}\left|X^{(L-1)}\right|^2 - \rho\right)\frac{\Delta\mathcal{P}(Y|S)}{\mathcal{P}(Y|S)}\left(F - \overline{F}\right)\right].$$

By the Cauchy–Schwarz inequality,

$$|a_n'| \leqslant \frac{1}{2}\left(\mathbb{E}\left[\left(\frac{1}{n_{L-1}}\left|X^{(L-1)}\right|^2 - \rho\right)^2\left(\frac{\Delta\mathcal{P}(Y|S)}{\mathcal{P}(Y|S)}\right)^2\right]\right)^{\frac{1}{2}}\left(\mathbb{E}\left(F - \overline{F}\right)^2\right)^{\frac{1}{2}}. \qquad (5.2.39)$$

Now, to prove (5.2.27), it suffices to bound the first expectation on the right.

By (5.2.32), we have that

$$\mathbb{E}\left[\left(\frac{\Delta\mathcal{P}(Y|S)}{\mathcal{P}(Y|S)}\right)^2\middle|X^{(L-1)}, S\right] = \frac{1}{(2\pi)^{\frac{n_{L-1}}{2}}}\int\left(\frac{\Delta\mathcal{P}(y|S)}{\mathcal{P}(y|S)}\right)^2\mathcal{P}(y|S)\mathrm{d}y.$$

Recall the notation (5.2.33). Then, (5.2.38) implies that

$$\mathbb{E}\left[\left(\frac{\Delta\mathcal{P}(Y|S)}{\mathcal{P}(Y|S)}\right)^2\middle|X^{(L-1)}, S\right] = \frac{1}{(2\pi)^{\frac{n_{L-1}}{2}}}\sum_{j=1}^{n_L}\int\left(\frac{\partial_j^2\mathcal{P}(y|S)}{\mathcal{P}(y|S)}\right)^2\mathcal{P}(y|S)\mathrm{d}y.$$

Using Jensen's inequality to the integral in (5.2.36), we have that

$$\int\left(\frac{\partial_j^2\mathcal{P}(y|S)}{\mathcal{P}(y|S)}\right)^2\mathcal{P}(y|S)\mathrm{d}y \leqslant \int\left(\Gamma_j\left(y_j, S_j, a_j^{(L)}\right)\right)^2 e^{-\frac{1}{2}|y-\sqrt{\beta}\varphi_L(S,a)|^2}\mathrm{d}P_{A^{(L)}}(a)\mathrm{d}y.$$

By the boundedness assumption in (h2) and the formula for Γ_j in (5.2.35), we obtain that

$$\mathbb{E}\left[\left(\frac{\Delta\mathcal{P}(Y|S)}{\mathcal{P}(Y|S)}\right)^2 \middle| X^{(L-1)}, S\right] \leqslant Cn_L,$$

which implies that

$$\mathbb{E}\left[\left(\frac{1}{n_{L-1}}\left|X^{(L-1)}\right|^2 - \rho\right)^2\left(\frac{\Delta\mathcal{P}(Y|S)}{\mathcal{P}(Y|S)}\right)^2\right] \leqslant Cn_L\mathbb{E}\left(\frac{1}{n_{L-1}}\left|X^{(L-1)}\right|^2 - \rho\right)^2.$$

Inserting this to (5.2.39) yields (5.2.27).

Proof of (5.2.28)

Recalling the definitions of u in (5.2.17) and $\mathcal{P}$ in (5.1.6), we can see that

$$\nabla u_Y(s) = \left(\frac{\int\left(Y_j - \varphi_L\left(s_j, a_j^{(L)}\right)\right)\varphi_L'\left(s_j, a_j^{(L)}\right)e^{-\frac{1}{2}|Y-\sqrt{\beta}\varphi_L(s,a^{(L)})|^2}\mathrm{d}P_{A^{(L)}}\left(a^{(L)}\right)}{\int e^{-\frac{1}{2}|Y-\sqrt{\beta}\varphi_L(s,a^{(L)})|^2}\mathrm{d}P_{A^{(L)}}\left(a^{(L)}\right)}\right)_{1\leqslant j\leqslant n_L}$$

where φ' is the derivative with respect to its first argument. Recall the definition of Y in (5.2.4). Using the boundedness of φ_L and its derivatives ensured by (h2), we can see that

$$|\nabla u_Y(s)| \leqslant C(\sqrt{n_L} + |Z|). \tag{5.2.40}$$

This computation also gives that

$$|\nabla u_Y(S)| \leqslant C(\sqrt{n_L} + |Z|)$$

which together with (5.2.40) verifies (5.2.28).

Proof of (5.2.29)

For simplicity, we write

$$\overline{X} = X^{(L-1)}, \qquad \overline{x} = x^{(L-1)}. \tag{5.2.41}$$

Using the formula for $\partial_2 F$ in (5.2.25), we can compute that

$$n\partial_2^2 F = \left\langle (\partial_2 H(x,w,a))^2 \right\rangle - \langle \partial_2 H(x,w,a) \rangle^2 - \frac{1}{4h_2^{\frac{3}{2}}} \langle Z' \cdot \overline{x} \rangle \tag{5.2.42}$$

Inserting (5.2.25) into the second term on the right and applying the Gaussian integration by parts to the last term, we obtain that

$$n\partial_2^2 \overline{F} = \mathbb{E} \left\langle (\partial_2 H(x,w,a))^2 \right\rangle - n^2 \mathbb{E}(\partial_2 F)^2 - \frac{1}{4h_2} \mathbb{E} \left\langle |\overline{x}|^2 \right\rangle + \frac{1}{4h_2} \mathbb{E} | \langle \overline{x} \rangle |^2, \tag{5.2.43}$$

where, to get the last term, we also invoked the Nishimori identity. We claim that

$$\mathbb{E} \left\langle (\partial_2 H(x,w,a))^2 \right\rangle \geqslant \frac{1}{4} \mathbb{E} \left\langle (\overline{x} \cdot \overline{x}')^2 \right\rangle + \frac{1}{4h_2} \mathbb{E} \left\langle |\overline{x}|^2 \right\rangle, \tag{5.2.44}$$

and postpone its proof. Now, insert (5.2.44) into (5.2.43) to see that

$$n\partial_2^2 \overline{F} \geqslant \frac{1}{4} \mathbb{E} \left\langle (\overline{x} \cdot \overline{x}')^2 \right\rangle - n^2 \mathbb{E}(\partial_2 F)^2.$$

By (5.2.26), we have that

$$\mathbb{V}\mathrm{ar}_{\mathbb{E}\langle \cdot \rangle} \left[X^{(L-1)} \cdot x^{(L-1)} \right] = \mathbb{E} \left\langle (\overline{x} \cdot \overline{x}')^2 \right\rangle - \left(\mathbb{E} \left\langle \overline{x} \cdot \overline{x}' \right\rangle \right)^2 = \mathbb{E} \left\langle (\overline{x} \cdot \overline{x}')^2 \right\rangle - 4n^2 (\partial_2 \overline{F}_n)^2.$$

Then, (5.2.29) follows from the above two displays along with (5.1.1).

It remains to derive (5.2.44). Using the expression of $\partial_2 H$ in (5.2.25), we have that

$$\begin{aligned} \mathbb{E} \left\langle (\partial_2 H(x,w,a))^2 \right\rangle &= \mathbb{E} \left\langle \left(\frac{1}{2\sqrt{h_2}} Z' \cdot \overline{x} + \overline{x} \cdot \overline{X} - \frac{1}{2} |\overline{x}|^2 \right)^2 \right\rangle \\ &= \mathbb{E} \Big\langle \frac{1}{4h_2} (Z' \cdot \overline{x})^2 + (\overline{x} \cdot \overline{X})^2 + \frac{1}{4} |\overline{x}|^4 \\ &\quad + \frac{1}{\sqrt{h_2}} (Z' \cdot \overline{x})(\overline{x} \cdot \overline{X}) - \frac{1}{2\sqrt{h_2}} (Z' \cdot \overline{x}) |\overline{x}|^2 - (\overline{x} \cdot \overline{X}) |\overline{x}|^2 \Big\rangle \end{aligned} \tag{5.2.45}$$

The first term on the last line can be rewritten as

$$\mathbb{E}\left\langle \frac{1}{4h_2}(Z'\cdot\overline{x})^2\right\rangle = \sum_{i,j=1}^{n_{L-1}} \frac{1}{4h_2}\mathbb{E}\left\langle Z_i'Z_j'\overline{x}_i\overline{x}_j\right\rangle.$$

If $i\neq j$, the Gaussian integration by parts yields that

$$\frac{1}{h_2}\mathbb{E}\left\langle Z_i'Z_j'\overline{x}_i\overline{x}_j\right\rangle = \mathbb{E}\left\langle \overline{x}_i\overline{x}_j(\overline{x}_i-\overline{x}_i')(\overline{x}_j+\overline{x}_j'-2\overline{x}_j'')\right\rangle.$$

If $i=j$, we have that

$$\frac{1}{h_2}\mathbb{E}\left\langle Z_i'Z_i'\overline{x}_i\overline{x}_i\right\rangle = \mathbb{E}\left\langle \overline{x}_i\overline{x}_i(\overline{x}_i-\overline{x}_i')(\overline{x}_i+\overline{x}_i'-2\overline{x}_i'')\right\rangle + \frac{1}{h_2}\mathbb{E}\left\langle \overline{x}_i^2\right\rangle.$$

The above three displays combined give that

$$\mathbb{E}\left\langle \frac{1}{4h_2}(Z'\cdot\overline{x})^2\right\rangle = \frac{1}{4}\mathbb{E}\left\langle |\overline{x}|^4-2|\overline{x}|^2(\overline{x}\cdot\overline{x}')-(\overline{x}\cdot\overline{x}')^2+2(\overline{x}\cdot\overline{x}')(\overline{x}\cdot\overline{x}'')\right\rangle + \frac{1}{4h_2}\mathbb{E}\left\langle |\overline{x}|^2\right\rangle.$$

Other terms can be computed using the Nishimori identity and the Gaussian integration by parts. We shall omit the details but only list the results:

$$\begin{aligned}
&\mathbb{E}\left\langle (\overline{x}\cdot\overline{X})^2\right\rangle = \mathbb{E}\left\langle (\overline{x}\cdot\overline{x}')^2\right\rangle,\\
&\mathbb{E}\left\langle \frac{1}{\sqrt{h_2}}(Z'\cdot\overline{x})(\overline{x}\cdot\overline{X})\right\rangle = \mathbb{E}\left\langle |\overline{x}|^2(\overline{x}\cdot\overline{x}')-(\overline{x}\cdot\overline{x}')(\overline{x}\cdot\overline{x}'')\right\rangle,\\
&\mathbb{E}\left\langle \frac{1}{\sqrt{h_2}}(Z'\cdot\overline{x})|\overline{x}|^2\right\rangle = \mathbb{E}\left\langle |\overline{x}|^4-|\overline{x}|^2(\overline{x}\cdot\overline{x}')\right\rangle,\\
&\mathbb{E}\left\langle (\overline{x}\cdot\overline{X})|\overline{x}|^2\right\rangle = \mathbb{E}\left\langle |\overline{x}|^2(\overline{x}\cdot\overline{x}')\right\rangle.
\end{aligned}$$

Inserting these computations into (5.2.45) yields that

$$\mathbb{E}\left\langle (\partial_2 H(x,w,a))^2\right\rangle = \frac{1}{4}\mathbb{E}\left\langle (\overline{x}\cdot\overline{x}')^2\right\rangle + \frac{1}{2}\mathbb{E}\left\langle (\overline{x}\cdot\overline{x}')^2-(\overline{x}\cdot\overline{x}')(\overline{x}\cdot\overline{x}'')\right\rangle + \frac{1}{4h_2}\mathbb{E}\left\langle |\overline{x}|^2\right\rangle.$$

Apply the Cauchy–Schwarz inequality and the symmetry of replicas to see that

$$\mathbb{E}\left\langle(\overline{x}\cdot\overline{x}')(\overline{x}\cdot\overline{x}'')\right\rangle \leqslant \frac{1}{2}\mathbb{E}\left\langle(\overline{x}\cdot\overline{x}')^2\right\rangle + \frac{1}{2}\mathbb{E}\left\langle(\overline{x}\cdot\overline{x}'')^2\right\rangle = \mathbb{E}\left\langle(\overline{x}\cdot\overline{x}')^2\right\rangle.$$

These two displays imply (5.2.44).

5.2.3. Estimates of derivatives

We collect useful properties of derivatives of $\overline{F}_{\beta,L,n}$ and $F_{\beta,L,n}$ in the following lemma.

Lemma 5.2.2. *Assume (h1), (h2) and (H3) for some $L\in\mathbb{N}$. For every $\beta\geqslant 0$ and every $n\in\mathbb{N}$, there is a constant C such that the following holds for all $n\in\mathbb{N}$ and all $(t,h)\in\Omega_{\rho_{L-1,n}}\setminus\{h_1=\rho_{L-1,n}(1-t)\}$,*

$$\partial_1\overline{F}_{\beta,L,n}\in[0,C]; \tag{5.2.46}$$

$$\partial_2\overline{F}_{\beta,L,n}\in\left[0,\frac{n_{L-1}\rho_{L-1,n}}{2n}\right]\subseteq[0,C]; \tag{5.2.47}$$

$$|\partial_2 F_{\beta,L,n}|\leqslant C\left(1+n^{-\frac{1}{2}}h_2^{-\frac{1}{2}}|Z'|\right); \tag{5.2.48}$$

$$\partial_i\partial_j\overline{F}_{\beta,L,n}\geqslant 0,\quad \forall i,j=1,2; \tag{5.2.49}$$

$$\partial_2^2 F_{\beta,L,n}\geqslant -Cn^{-\frac{1}{2}}h_2^{-\frac{3}{2}}|Z'|. \tag{5.2.50}$$

Let us prove these assertions. Again, for simplicity, we write $F=F_{\beta,L,n}$ in the proofs below.

Proof of (5.2.46)

By (5.2.24) and the Nishimori identity, we can see that

$$\partial_1\overline{F}=\frac{1}{2n}\mathbb{E}\left|\langle\nabla u_Y(s)\rangle\right|^2\geqslant 0.$$

Due to (5.2.40), it is also bounded.

Proof of (5.2.47)

The first range follows from the formula for $\partial_2\overline{F}$ in (5.2.26), the definition of $\rho_{L-1,n}$ in (5.1.10), the Cauchy–Schwarz inequality and the Nishimori identity. The boundedness is clear from the observation that there is a constant C such that, a.s.,

$$\left|X^{(L-1)}\right|,\ \left|x^{(L-1)}\right| \leqslant C\sqrt{n} \tag{5.2.51}$$

which is ensured by (5.1.1), (h1) and (h2).

Proof of (5.2.48)

In view of (5.2.25), this is valid due to (5.2.51).

Proof of (5.2.49)

We first show that $\partial_1\partial_2\overline{F} \geqslant 0$. Recall the formula for $\partial_1\overline{F}$ in (5.2.24). Let use write $\overline{u} = \nabla u_Y(s)$ and $\overline{U} = \nabla u_Y(S)$. We also adopt the notation (5.2.41). Then, we compute that

$$\begin{aligned}\partial_1\partial_2\overline{F} &= (2n)^{-1}\partial_2\mathbb{E}\left\langle \overline{u}\cdot\overline{U}\right\rangle \\ &= (4n)^{-1}\mathbb{E}\langle(\overline{u}\cdot\overline{U})\big((h_2)^{-\frac{1}{2}}Z'\cdot\overline{x} + 2\overline{x}\cdot\overline{X} - \overline{x}\cdot\overline{x}\big) \\ &\qquad - (\overline{u}\cdot\overline{U})\big((h_2)^{-\frac{1}{2}}Z'\cdot\overline{x}' + 2\overline{x}'\cdot\overline{X} - \overline{x}'\cdot\overline{x}'\big)\rangle.\end{aligned}$$

Perform the Gaussian integration by parts on Z' to get that

$$\begin{aligned}\partial_1\partial_2\overline{F} = (4n)^{-1}\mathbb{E}\langle(\overline{u}\cdot\overline{U})&\big((\overline{x}-\overline{x}')\cdot\overline{x} + 2\overline{x}\cdot\overline{X} - \overline{x}\cdot\overline{x}\big) \\ &- (\overline{u}\cdot\overline{U})\big((\overline{x}+\overline{x}'-2\overline{x}'')\cdot\overline{x}' + 2\overline{x}'\cdot\overline{X} - \overline{x}'\cdot\overline{x}'\big)\rangle.\end{aligned}$$

Using the Nishimori identity to replace $\overline{U}$ and $\overline{X}$ by replicas and invoking the symmetry of replicas, we arrive at

$$\partial_1\partial_2\overline{F} = (2n)^{-1}\mathbb{E}\left\langle(\overline{u}\cdot\overline{u}')(\overline{x}\cdot\overline{x}') - 2(\overline{u}\cdot\overline{u}')(\overline{x}\cdot\overline{x}'') + (\overline{u}\cdot\overline{u}')(\overline{x}''\cdot\overline{x}''')\right\rangle$$
$$= (2n)^{-1}\mathbb{E}\left|\left\langle\overline{u}\,\overline{x}^{\intercal}\right\rangle - \left\langle\overline{u}\right\rangle\left\langle\overline{x}\right\rangle^{\intercal}\right|^2 \geqslant 0.$$

The computation for $\partial_2^2\overline{F} \geqslant 0$ is exactly the same with $\overline{U}, \overline{u}$ above replaced by $\overline{X}, \overline{x}$. The verification of $\partial_1^2\overline{F} \geqslant 0$ follows the same procedure but is computationally more involved. We refer to the proof of [11, Proposition 18 in its supplementary material] for details.

Proof of (5.2.50)

Notice that the first two terms on the right of formula (5.2.42) for $\partial_2^2 F$ form a variance. Then, the desired lower bound follows from (5.2.51).

5.3. Weak solutions

We consider the equation (6.1.2) over Ω_ρ defined in (5.2.1) for some $\rho > 0$. We give the definition of weak solutions, and prove the uniqueness and existence of weak solutions. Uniqueness is ensured by Proposition 5.3.2. Proposition 5.3.3 furnishes the existence part by providing a variational formula known as the Hopf formula. After stating these, we prove the two propositions in the ensuing subsections.

We endow measurable subsets of Euclidean spaces with the Lebesgue measure. In what follows, the phrase "almost everywhere" or "almost every" (a.e.) is understood with respect to the Lebesgue measure. We denote by $\mathsf{int}\,\Omega_\rho$ the interior of Ω_ρ. In this section, for convenience, we also denote the spacial variable by x instead of h.

Definition 5.3.1. For $L \in \mathbb{N}$ and $\rho > 0$, a function $f : \Omega_\rho \to \mathbb{R}$ is a weak solution of (6.1.2) if

1. f is Lipschitz, and $\partial_1 f \geqslant 0$, $\partial_2 f \in [0, \frac{\alpha_{L-1}\rho}{2}]$ a.e.;

2. f satisfies (6.1.2) a.e.;

3. for all $(t,x) \in \operatorname{int}\Omega_\rho$ and all sufficiently small $\lambda \geqslant 0$, it holds that

$$f(t, x+\lambda e_1+\lambda e_2)+f(t,x)-f(t,x+\lambda e_1)-f(t,x+\lambda e_2) \geqslant 0. \tag{5.3.1}$$

By Rademacher's theorem, condition (1) implies that f is differentiable a.e. Condition (2) is understood in the sense that, outside a set with zero measure, f is differentiable and its derivatives satisfy equation (6.1.2). In (3), $\{e_1, e_2\}$ is the standard basis for $\mathbb{R}^2$. Condition (3) can be interpreted as a type of partial convexity. For a smooth radial bump function $\xi : \mathbb{R}^2 \to \mathbb{R}$ supported on the unit disk satisfying $\xi \in [0,1]$ and $\int \xi = 1$, introduce, for every $\varepsilon \in (0,1)$,

$$\xi_\varepsilon(x) = \varepsilon^{-2}\xi\left(\varepsilon^{-1}x\right), \quad \forall x \in \mathbb{R}^2. \tag{5.3.2}$$

If f is a weak solution, then condition (3) along with the continuity of f implies that

$$\partial_1\partial_2(f(t,\cdot) * \xi_\varepsilon)(x) \geqslant 0, \tag{5.3.3}$$

for every (t,x) in

$$\Omega_{\rho,\varepsilon} = \left\{t \in \left[0, 1-\frac{2}{\rho}\varepsilon\right],\ x_1 \in [\varepsilon, \rho(1-t)-\varepsilon],\ x_2 \in [\varepsilon, \infty)\right\}, \tag{5.3.4}$$

where the convolution in (5.3.3) is taken in terms of the spacial variable.

The main results of this section are stated below.

Proposition 5.3.2. *Given a Lipschitz function $\psi : [0,\rho] \times \mathbb{R}_+ \to \mathbb{R}$, there is at most one weak solution f of (6.1.2) satisfying $f(0,\cdot) = \psi$.*

Proposition 5.3.3. *Let $\psi_1 : [0,\rho] \to \mathbb{R}$ and $\psi_2 : \mathbb{R}_+ \to \mathbb{R}$ be Lipschitz, nondecreasing and* *convex. In addition, suppose that*

$$\partial_2\psi_2 \in \left[0, \frac{\alpha_{L-1}\rho}{2}\right], \quad a.e. \tag{5.3.5}$$

Define $\psi : [0,\rho] \times \mathbb{R}_+ \to \mathbb{R}$ *by*

$$\psi(x) = \psi_1(x_1) + \psi_2(x_2), \quad \forall x \in [0,\rho) \times \mathbb{R}_+. \tag{5.3.6}$$

Then, the formula

$$f(t,x) = \sup_{z\in\mathbb{R}_+\times[0,\frac{\alpha_{L-1}\rho}{2}]} \inf_{y\in[0,\rho]\times\mathbb{R}_+} \{z\cdot(x-y) + \psi(y) + t\mathsf{H}_L(z)\}, \quad \forall (t,x) \in \Omega_\rho, \tag{5.3.7}$$

gives a weak solution of (6.1.2) *satisfying* $f(0,\cdot) = \psi$.

The expression in (5.3.7) is known as the Hopf formula [15, 84].

5.3.1. Proof of Proposition 5.3.2

The idea of this proof can be seen in [60, Section 3.3.3]. Let f and g be weak solutions to (6.1.2). Setting $w = f - g$, we have that

$$\partial_t w = \mathsf{H}_L(\nabla f) - \mathsf{H}_L(\nabla g) = b\cdot\nabla w$$

where the vector b is given by

$$b = \frac{2}{\alpha_{L-1}}\big(\partial_2 g,\ \partial_1 f\big). \tag{5.3.8}$$

For some smooth function $\phi : \mathbb{R}_+ \to \mathbb{R}_+$ to be chosen later, we set $v = \phi(w)$, which, by the chain rule, satisfies that

$$\partial_t v = b\cdot\nabla v. \tag{5.3.9}$$

Then, we regularize b by setting $b_\varepsilon = b * \xi_\varepsilon$ for the mollifier ξ_ε introduced in (5.3.2), where we understand that the convolution is taken with respect to the spacial variable. On $\Omega_{\rho,\varepsilon}$ given in (5.3.4), the equation (5.3.9) can be rewritten as

$$\partial_t v = \mathsf{div}(v b_\varepsilon) - v\,\mathsf{div}\, b_\varepsilon + (b - b_\varepsilon)\cdot \nabla v. \tag{5.3.10}$$

Before proceeding further, we need to estimate some terms related to this display.

Definition 5.3.1 (3) and (5.3.3) imply that, for all $(t,x) \in \Omega_{\rho,\varepsilon}$,

$$\partial_1\partial_2 f_\varepsilon(t,x),\ \partial_1\partial_2 g_\varepsilon(t,x) \geqslant 0,$$

and thus

$$\mathsf{div}\, b_\varepsilon \geqslant 0, \quad \forall (t,x) \in \Omega_{\rho,\varepsilon}. \tag{5.3.11}$$

By the definitions of f_ε and g_ε, we also have that

$$|\nabla f_\varepsilon| \leqslant \|f\|_{\mathrm{Lip}}, \quad |\nabla g_\varepsilon| \leqslant \|g\|_{\mathrm{Lip}}. \tag{5.3.12}$$

Let us fix a constant R to satisfy

$$R > \sup\left\{|\nabla \mathsf{H}_L(p)| : p \in \mathbb{R}^2_+,\ |p| \leqslant \|f\|_{\mathrm{Lip}} + \|g\|_{\mathrm{Lip}}\right\}. \tag{5.3.13}$$

Fix any $\eta > 0$ and define, for $t \in [0, 1 - \frac{2}{\rho}\eta]$,

$$\begin{aligned} D_t &= [\eta, \rho(1-t) - \eta] \times [\eta, R(1-t)], \\ \Gamma_{1,t} &= [\eta, \rho(1-t) - \eta] \times \{R(1-t)\}, \\ \Gamma_{2,t} &= \{\rho(1-t) - \eta\} \times [\eta, R(1-t)]. \end{aligned} \tag{5.3.14}$$

Now, we introduce, for $t \in [0, 1 - \frac{2}{\rho}\eta]$,

$$J(t) = \int_{D_t} v(t, x) \mathrm{d}x.$$

We emphasize that J depends on η. Choose $\varepsilon < \eta$ to ensure that $\bigcup_{t \in [0,1-\frac{2}{\rho}\eta]} (\{t\} \times D_t) \subseteq \Omega_{\rho,\varepsilon}$. Using (5.3.10) and integration by parts on the integral of $\mathsf{div}(vb_\varepsilon)$, we can compute that

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t} J(t) &= \int_{D_t} \partial_t v - R \int_{\Gamma_{1,t}} v - \rho \int_{\Gamma_{2,t}} v \\
&= \int_{\Gamma_{1,t}} (\mathbf{n} \cdot b_\varepsilon - R) v + \int_{\Gamma_{2,t}} (\mathbf{n} \cdot b_\varepsilon - \rho) v \\
&\quad + \int_{\partial D_t \setminus \Gamma_t} (\mathbf{n} \cdot b_\varepsilon) v + \int_{D_t} v(-\mathsf{div}\, b_\varepsilon) + \int_{D_t} (b - b_\varepsilon) \cdot \nabla v,
\end{aligned}$$

where $\mathbf{n}$ stands for the outer normal vector. Then, $\mathbf{n} = (0, 1)$ on $\Gamma_{1,t}$ and $\mathbf{n} = (1, 0)$ on $\Gamma_{2,t}$. We treat the integrals after the second equality individually. Due to (5.3.8), (5.3.12) and (5.3.13), the first integral is nonpositive. By Definition 5.3.1 (1) and (5.3.8), the second integral is nonpositive. Note that on $\partial D_t \setminus \Gamma_t$, we have $-\mathbf{n} \in \mathbb{R}^2_+$. By Definition 5.3.1 (1), we can infer from the definition of b_ε that $b_\varepsilon \in \mathbb{R}^2_+$ on $\partial D_t \setminus \Gamma_t$, which implies that the third integral is nonpositive. In view of (5.3.11), the fourth integral is again nonpositive. The last one is $o_\varepsilon(1)$. Therefore, sending $\varepsilon \to 0$, we conclude that, for $t \in [0, 1 - \frac{2}{\rho}\eta]$,

$$\frac{\mathrm{d}}{\mathrm{d}t} J(t) \leqslant 0. \tag{5.3.15}$$

Since $w(0, x) = f(0, x) - g(0, x) = 0$, we have $\|w(\delta, \cdot)\|_\infty \leqslant \delta(\|f\|_{\mathrm{Lip}} + \|g\|_{\mathrm{Lip}})$, for each $\delta > 0$. Let us choose $\phi = \phi_\delta$ to satisfy

$$\begin{cases}
\phi_\delta(z) = 0, & \text{if } |z| \leqslant \delta(\|f\|_{\mathrm{Lip}} + \|g\|_{\mathrm{Lip}}), \\
\phi_\delta(z) > 0, & \text{otherwise.}
\end{cases}$$

Therefore, due to $v = \phi_\delta(w)$, we have that

$$J(\delta) = \int_{D_\delta} v(\delta, x)\mathrm{d}x = \int_{D_\delta} \phi_\delta(w(\delta, x))\mathrm{d}x = 0.$$

Since $J(t)$ is nonnegative, (5.3.15) implies that $J_\delta(t) = 0$ for all $t \in [\delta, 1 - \frac{2}{\rho}\eta]$. This together with the definition of ϕ guarantees that

$$|f(t,x) - g(t,x)| \leqslant \delta(\|f\|_{\mathrm{Lip}} + \|g\|_{\mathrm{Lip}}), \quad \forall x \in D_t,\ \forall t \in \left[\delta, 1 - \frac{2}{\rho}\eta\right].$$

Recall the definition of D_t in (5.3.14) which depends on η. Taking $\delta \to 0$, $\eta \to 0$ and $R \to \infty$, we conclude that $f = g$.

5.3.2. Proof of Proposition 5.3.3

Let us extend ψ_1 to be defined on $\mathbb{R}_+$ by setting

$$\psi_1(x_1) = \infty, \quad \forall x_1 \in \mathbb{R}_+ \setminus [0, \rho]. \tag{5.3.16}$$

Then, ψ_1 is still convex and nondecreasing. For $u : \mathbb{R}^2_+ \to \mathbb{R} \cup \{\infty\}$, the Fenchel transformation is defined by

$$u^*(x) = \sup_{y \in \mathbb{R}^2_+} \{y \cdot x - u(y)\}, \quad \forall x \in \mathbb{R}^2_+. \tag{5.3.17}$$

Hence, we can rewrite the Hopf formula (5.3.7) as

$$\begin{aligned} f(t,x) &= \sup_{z \in \mathbb{R}_+ \times [0, \frac{\alpha_{L-1}\rho}{2}]} \inf_{y \in \mathbb{R}^2_+} \left\{z \cdot (x - y) + \psi(y) + t\mathsf{H}_L(z)\right\} \\ &= \sup_{z \in \mathbb{R}_+ \times [0, \frac{\alpha_{L-1}\rho}{2}]} \{z \cdot x - \psi^*(z) + t\mathsf{H}_L(z)\}. \end{aligned} \tag{5.3.18}$$

We first show that f is indeed finite on Ω_ρ. From (5.3.6), it follows that

$$\psi^*(z) = \psi_1^*(z_1) + \psi_2^*(z_2), \quad \forall z \in \mathbb{R}_+^2, \tag{5.3.19}$$

where the Fenchel transforms on the right-hand side are for functions defined on $\mathbb{R}_+$ which are defined analogously to (5.3.17). By the assumption that ψ_1 is Lipschitz and nondecreasing, there is some $R \geqslant 0$ such that

$$0 \leqslant \psi_1(r) - \psi(r') \leqslant R(r - r'), \qquad \forall r \geqslant r', \quad r, r' \in [0, \rho].$$

Due to the extension in (5.3.16), we have that

$$\psi_1^*(z_1) = \sup_{y_1 \in [0,\rho]} \{y_1 z_1 - \psi_1(y_1)\}.$$

The above two displays imply that

$$\psi_1^*(z_1) = \rho z_1 - \psi_1(\rho), \quad \forall z_1 \geqslant R. \tag{5.3.20}$$

On the other hand, due to (5.3.5),

$$\psi_2^*(z_2) = \infty, \quad \forall z_2 > \frac{\alpha_{L-1}\rho}{2}. \tag{5.3.21}$$

Using this, (5.3.19) and the expression of H_L in (5.2.10), we rewrite (5.3.18) as

$$f(t,x) = \sup_{z_2 \in [0, \frac{\alpha_{L-1}\rho}{2}]} \left\{ z_2 x_2 - \psi_2^*(z_2) + \sup_{z_1 \in \mathbb{R}_+} \left\{ z_1 x_1 - \psi_1^*(z_1) + \frac{2t}{\alpha_{L-1}} z_1 z_2 \right\} \right\}. \tag{5.3.22}$$

We show that the second sup can be restricted to $z_1 \in [0, R]$. Given $(t,x) \in \Omega_\rho$, we have that $x_1 \in [0, \rho(1-t)]$ due to the definition of Ω_ρ in (5.2.1). This implies that

$$x_1 + \frac{2t}{\alpha_{L-1}} z_2 - \rho \leqslant 0, \qquad \forall (t,x) \in \Omega_\rho, \ z_2 \in \left[0, \frac{\alpha_{L-1}\rho}{2}\right],$$

which together with (5.3.20) shows that

$$\sup_{z_1\in[R,\infty)}\left\{z_1x_1-\psi_1^*(z_1)+\frac{2t}{\alpha_{L-1}}z_1z_2\right\}=\sup_{z_1\in[R,\infty)}\left\{\left(x_1+\frac{2t}{\alpha_{L-1}}z_2-\rho\right)z_1+\psi_1(\rho)\right\}$$
$$\leqslant\left(x_1+\frac{2t}{\alpha_{L-1}}z_2-\rho\right)R+\psi_1(\rho)=Rx_1-\psi_1^*(R)+\frac{2t}{\alpha_{L-1}}Rz_2,$$

In other words, the above sup is achieved at $z_1=R$. Hence, the second sup in (5.3.22) can be taken over $z_1\in[0,R]$ and thus the sup in (5.3.18) can be restricted to z belonging to the compact set

$$K=[0,R]\times\left[0,\frac{\alpha_{L-1}\rho}{2}\right]. \tag{5.3.23}$$

Therefore, due to the easy observation that ψ^* is nondecreasing and lower semi-continuous, we can see that f is finite on Ω_ρ, and furthermore, for every $(t,x)\in\Omega_\rho$,

$$f(t,x)=z\cdot x-\psi^*(z)+t\mathsf{H}_L(z),\qquad \exists z\in K. \tag{5.3.24}$$

In the following, we verify that (5.3.18) is a weak solution by checking the initial condition, and conditions (1), (2), (3) in Definition 5.3.1.

Initial condition

Using (5.3.19) and (5.3.21), the expression in (5.3.18) at $t=0$ becomes

$$f(0,x)=\sup_{z\in\mathbb{R}_+^2}\{z\cdot x-\psi^*(z)\}=\psi^{**}(x).$$

Since it is clear from the assumption that the extended ψ is lower semi-continuous, nondecreasing and convex, the Fenchel–Moreau biconjugation identity (cf. [105, Theorem 12.4], and [38, Theorem 2.2] for more general cones) ensures that

$$\psi(x)=\psi^{**}(x),\quad \forall x\in\mathbb{R}_+^2.$$

In particular, we have $f(0,\cdot) = \psi$ on Ω_ρ.

Condition (1)

Let $(t,x) \in \Omega_\rho$ and $z \in K$ be given by (5.3.24). Using this and (5.3.18) for $(t',x') \in \Omega_\rho$, we have

$$f(t,x) - f(t',x') \leqslant z \cdot (x - x') + \mathsf{H}_L(z)(t - t'). \tag{5.3.25}$$

A similar equality holds for some $z' \in K$ when interchanging $(t,x),(t',x')$. By the compactness of K, we can see that f is Lipschitz. Due to Rademacher's theorem, f is differentiable a.e. Using (5.3.25) and the definition of K in (5.3.23), we can also see that

$$\partial_1 f \in [0,R], \qquad \partial_2 f \in [0, \frac{\alpha_{L-1}\rho}{2}], \qquad \text{a.e.,}$$

which completes the verification of Definition 5.3.1 (1).

Condition (2)

We want to verify that (5.3.18) satisfies (6.1.2) almost everywhere. Let (t,x) be a point at which f is differentiable. We can assume that $(t,x) \in \operatorname{\mathsf{int}}\Omega_\rho \subseteq (0,\infty)^3$, because otherwise (t,x) belongs to a set with Lebesgue measure zero. Let z be given by (5.3.24). By this and (5.3.18), for $s \in \mathbb{R}$ and $h \in \mathbb{R}^2$ sufficiently small,

$$f(t+s, x+h) - f(t,x) \geqslant z \cdot h + s\mathsf{H}_L(z). \tag{5.3.26}$$

Set $s = 0$ and vary h to see that

$$z = \nabla f(t,x).$$

Then, we set $h = 0$ in (5.3.26), vary s and use the above display to obtain

$$\partial_t f(t,x) = \mathsf{H}_L(\nabla f(t,x)).$$

Condition (3)

Let $(t,x) \in \operatorname{int}\Omega_\rho$ and $\lambda \in \mathbb{R}$ be sufficiently small. Due to (5.3.24), there are z, z' such that

$$\begin{aligned} f(t, x+\lambda e_1) &= z \cdot (x+\lambda e_1) - \psi^*(z) + t\mathsf{H}_L(z), \\ f(t, x+\lambda e_2) &= z' \cdot (x+\lambda e_2) - \psi^*(z') + t\mathsf{H}_L(z'). \end{aligned} \tag{5.3.27}$$

Case 1: $(z_1, z_2) \leqslant (z_1', z_2')$ or $(z_1, z_2) \geqslant (z_1', z_2')$. Let us only treat the latter case. The other case is similar. Using (5.3.18), we have

$$f(t, x+\lambda e_1+\lambda e_2) \geqslant z \cdot (x+\lambda e_1+\lambda e_2) - \psi^*(z) + t\mathsf{H}_L(z),$$
$$f(t,x) \geqslant z' \cdot x - \psi^*(z') + t\mathsf{H}_L(z').$$

This along with (5.3.27) implies that the left hand side of (5.3.1) is bounded below by

$$\lambda z \cdot e_2 - \lambda z' \cdot e_2 = \lambda(z_2 - z_2') \geqslant 0.$$

Case 2: neither $(z_1, z_2) \leqslant (z_1', z_2')$ nor $(z_1, z_2) \geqslant (z_1', z_2')$. This condition implies that

$$(z_1 - z_1')(z_2 - z_2') < 0. \tag{5.3.28}$$

Let $\widetilde{z} = (z_1, z_2')$ and $\widetilde{z}' = (z_1', z_2)$. By (5.3.18), for each $\delta > 0$, there are $y, y' \in \mathbb{R}^2_+$ such that

$$\begin{aligned} f(t, x+\lambda e_1+\lambda e_2) &\geqslant \widetilde{z} \cdot (x+\lambda e_1+\lambda e_2 - y) + \psi(y) + t\mathsf{H}_L(\widetilde{z}) - \delta, \\ f(t,x) &\geqslant \widetilde{z}' \cdot (x - y') + \psi(y') + t\mathsf{H}_L(\widetilde{z}') - \delta. \end{aligned} \tag{5.3.29}$$

We set

$$\widetilde{y} = (y_1, y_2'), \quad \widetilde{y}' = (y_1', y_2).$$

Note that

$$\widetilde{z} \cdot y + \widetilde{z}' \cdot y' - z \cdot \widetilde{y} - z' \cdot \widetilde{y}' = 0. \tag{5.3.30}$$

From (5.3.27), we also have

$$\begin{aligned} f(t, x + \lambda e_1) &\leqslant z \cdot (x + \lambda e_1 - \widetilde{y}) + \psi(\widetilde{y}) + t\mathsf{H}_L(z), \\ f(t, x + \lambda e_2) &\leqslant z' \cdot (x + \lambda e_2 - \widetilde{y}') + \psi(\widetilde{y}') + t\mathsf{H}_L(z'). \end{aligned} \tag{5.3.31}$$

To get a lower bound for the left hand side of (5.3.1), we start by observing that, due to (5.3.30),

$$\begin{aligned} &\widetilde{z} \cdot (x + \lambda e_1 + \lambda e_2 - y) + \widetilde{z}' \cdot (x - y') - z \cdot (x + \lambda e_1 - \widetilde{y}) - z' \cdot (x + \lambda e_2 - \widetilde{y}') \\ &= (\widetilde{z} + \widetilde{z}' - z - z') \cdot x - (\widetilde{z} \cdot y + \widetilde{z}' \cdot y' - z \cdot \widetilde{y} - z' \cdot \widetilde{y}') + \lambda(z_1 + z_2' - z_1 - z_2') \\ &= 0. \end{aligned}$$

This along with (5.3.29) and (5.3.31) implies that the left hand side of (5.3.1) can be bounded from below by

$$\psi(y) + \psi(y') - \psi(\widetilde{y}) - \psi(\widetilde{y}') + t\big(\mathsf{H}_L(\widetilde{z}) + \mathsf{H}_L(\widetilde{z}') - \mathsf{H}_L(z) - \mathsf{H}_L(z')\big) - 2\delta.$$

From (5.3.6), we can see

$$\psi(y) + \psi(y') - \psi(\widetilde{y}) - \psi(\widetilde{y}') = 0.$$

Lastly, due to (5.3.28) and the definition of H_L in (5.2.10), we can compute that

$$\mathsf{H}_L(\widetilde{z}) + \mathsf{H}_L(\widetilde{z}') - \mathsf{H}_L(z) - \mathsf{H}_L(z') = -\frac{2}{\alpha_{L-1}}(z_1 - z_1')(z_2 - z_2') > 0.$$

The above three displays imply that the left hand side of (5.3.1) is bounded from below by

-2δ. The desired result follows by sending $\delta \to 0$.

5.4. Convergence of the free energy

The goal of this section is to prove Theorem 5.1.1. The key tool is Proposition 5.4.1 stated below, which ensures the convergence of $\overline{F}_{\beta,L,n}$ given the convergence of $\overline{F}_{\beta,L,n}(0,\cdot)$ and some additional conditions. The object $\overline{F}_{\beta,L,n}(0,\cdot)$ is closely related to the free energy associated with the $(L-1)$-layer model. Hence, an iteration is employed in Section 5.4.1 to complete the proof of Theorem 5.1.1.

Recall the definition of $\rho_{L-1,n}$ in (5.1.10) and of domain Ω_ρ, for $\rho > 0$, in (5.2.1).

Proposition 5.4.1. *Assume (h1), (h2) and (H3) for some $L \in \mathbb{N}$. Suppose that the following holds:*

1. *the limit* (5.1.11) *for $l = L-1$ exists for some $\rho_{L-1} > 0$;*

2. *there is a continuous $\psi_{\beta,L} : [0, \rho_{L-1}] \times \mathbb{R}_+ \to \mathbb{R}$ such that*

$$\lim_{n\to\infty} \overline{F}_{\beta,L,n}(0,h) = \psi_{\beta,L}(h), \quad \forall h \in [0, \rho_{L-1}) \times \mathbb{R}_+,$$

   and there is a weak solution $f_{\beta,L}$ to (6.1.2) *on $\Omega_{\rho_{L-1}}$ satisfying $f_{\beta,L}(0,\cdot) = \psi_{\beta,L}$;*

3. *there is $C > 0$ such that*

$$\mathbb{E}\left[\left(\frac{\left|X^{(L-1)}\right|^2}{n_{L-1}} - \rho_{L-1,n}\right)^2\right] \leqslant \frac{C}{n}, \quad \forall n \in \mathbb{N};$$

4. *for every $M \geqslant 1$,*

$$\lim_{n\to\infty} \sup_{\substack{t\in[0,1],\\ h_1\in[0,\ \rho_{L-1,n}(1-t)]}} \mathbb{E}\left[\left\|F_{\beta,L,n} - \overline{F}_{\beta,L,n}\right\|^2_{L^\infty_{h_2}([0,M])}(t,h_1)\right] = 0.$$

Then, for every $\rho' \in (0, \rho_{L-1})$,

$$\lim_{n\to\infty} \overline{F}_{\beta,L,n}(t,h) = f_{\beta,L}(t,h), \quad \forall (t,h) \in \Omega_{\rho'}.$$

We restrict to the domain $\Omega_{\rho'}$ because the pointwise limit of $\overline{F}_{\beta,L,n}$ may not be well-defined on boundary points of $\Omega_{\rho_{L-1}}$ (recall that $\overline{F}_{\beta,L,n}$ is defined on $\Omega_{\rho_{L-1,n}}$). The proof of this proposition is postponed to Section 5.4.2.

5.4.1. Iteration

Let us prove Theorem 5.1.1 using Proposition 5.4.1 together with some technical results postponed to Section 5.5.

Assuming that (H1)–(H3) hold for the model with L_0 layers, then these assumptions automatically hold for all $L \in \{1, \dots, L_0\}$. Hence, for all $L \in \{1, \dots, L_0\}$, conditions (1), (3), (4) in Proposition 5.4.1 are guaranteed to hold by Lemmas 5.5.1, 5.5.2 and 5.5.4, respectively. We will apply Proposition 5.4.1 iteratively to prove Theorem 5.1.1. Recall the definitions of $F^{\circ}_{\beta,L,n}$, Ψ_0, Ψ_l, $F_{\beta,L,n}$ in (5.1.8), (5.1.12), (5.1.13), (5.2.8), respectively, and also the important relation (5.2.9), which implies that

$$\lim_{n\to\infty} \mathbb{E} F^{\circ}_{\beta,L,n} = \lim_{n\to\infty} \overline{F}_{\beta,L,n}(1,0), \tag{5.4.1}$$

whenever one of the limits exists. Also recall the definition of α_l in (5.1.1).

Before proceeding, let us record the following result. Comparing the definitions of (5.1.6) and (5.1.14), and using the fact that $A^{(L)}$ has i.i.d. components due to (H3), we can see that, for every β, L, n,

$$\mathcal{P}_{\beta,L,n}(y|z) = \prod_{j=1}^{n_L} \widetilde{\mathcal{P}}_{\beta,L}(y_j|z_j), \quad \forall y, z \in \mathbb{R}^{n_L}. \tag{5.4.2}$$

We start with $L = 1$. Using (5.2.2)–(5.2.8) with L replaced by 1, we can compute that

$$\overline{F}_{\beta,1,n}(0,h) = \frac{1}{n}\mathbb{E}\log\int \mathcal{P}_{\beta,1,n}\left(\mathcal{Y}^{(1)}\middle|\sqrt{h_1}V + \sqrt{\rho_{0,n} - h_1}w\right)\mathrm{d}P_W(w) \\ + \frac{1}{n}\mathbb{E}\log\int e^{h_2 X\cdot x + \sqrt{h_2}Z'\cdot x - \frac{h_2}{2}|x|^2}\mathrm{d}P_X(x)$$

where

$$\mathcal{Y}^{(1)} = \sqrt{\beta}\varphi_1\left(\sqrt{h_1}V + \sqrt{\rho_{0,n} - h_1}W, A^{(1)}\right) + Z,$$

V, W, Z are independent n_1-dimensional standard Gaussian vectors, and Z' is an n-dimensional Gaussian vector. By (5.4.2), the definitions of Ψ_0, Ψ_1 in (5.1.12), (5.1.13), and the fact that X, V, W have i.i.d. entries (see (H1) for the claim about X), the above can be rewritten as

$$\overline{F}_{\beta,1,n}(0,h) = \frac{n_1}{n}\Psi_1(h_1, \beta; \rho_{0,n}) + \Psi_0(h_2),$$

which, by (5.1.1) and (5.1.11), converges pointwise to

$$\psi_{\beta,1}(h) = \alpha_1\Psi_1(h_1, \beta; \rho_0) + \Psi_0(h_2).$$

The results collected in Lemma 5.2.2 allow us to verify that $\psi_{\beta,1}$ satisfies all the conditions imposed in Proposition 5.3.3. Indeed, the above display shows that the decomposition as in (5.3.6) exists, and both components are Lipschitz, nondecreasing and convex due to (5.2.46), (5.2.47) and (5.2.49). Moreover, (5.3.5) is ensured by (5.2.47), (5.1.1), (5.1.11). Hence, Proposition 5.3.3 yields the existence of a weak solution $f_{\beta,1}$ satisfying $f_{\beta,1}(0,\cdot) = \psi_{\beta,1}$ given by the formula (5.3.7) with L, ρ, ψ there replaced by $1, \rho_0, \psi_{\beta,1}$, namely,

$$f_{\beta,1}(t,h) = \sup_{z^{(1)}\in\mathbb{R}_+\times[0,\frac{\alpha_0\rho_0}{2}]}\ \inf_{y^{(1)}\in[0,\rho_0]\times\mathbb{R}_+}\left\{z^{(1)}\cdot\left(h - y^{(1)}\right) + \psi_{\beta,1}\left(y^{(1)}\right) + t\mathsf{H}_1\left(z^{(1)}\right)\right\}$$

for every $(t,h) \in \Omega_{\rho_0}$. Inserting the previous display and the formula for H_1 in (5.2.10) into the above, and evaluating at $(t,h)=(1,0)$ yield that

$$f_{\beta,1}(1,0)=\sup_{z^{(1)}}\inf_{y^{(1)}}\left\{\alpha_1\Psi_1\left(y_1^{(1)},\beta;\rho_0\right)+\Psi_0\left(y_2^{(1)}\right)-y^{(1)}\cdot z^{(1)}+\frac{2}{\alpha_0}z_1^{(1)}z_2^{(1)}\right\}$$

which exactly matches the right-hand side of (5.1.15) for $L=1$.

The above discussion also validates condition (2) in Proposition 5.4.1. Therefore, applying this proposition yields that

$$\lim_{n\to\infty}\overline{F}_{\beta,1,n}(1,0)=f_{\beta,1}(1,0).$$

Using (5.4.1), this proves (5.1.15) for $L=1$.

Now, we assume that (5.1.15) is verified for $L-1$. Using (5.2.2)–(5.2.8), we can compute that

$$\begin{aligned}\overline{F}_{\beta,L,n}(0,h)&=\frac{1}{n}\mathbb{E}\log\int\mathcal{P}_{\beta,L,n}\left(\mathcal{Y}^{(L)}\middle|\sqrt{h_1}V+\sqrt{\rho_{L-1,n}-h_1}w\right)\mathrm{d}P_W(w)\\&\quad+\frac{1}{n}\mathbb{E}\log\int e^{\sqrt{h_2}Y'\cdot x^{(L-1)}-\frac{h_2}{2}|x^{(L-1)}|^2}\mathrm{d}P_X(x)\mathrm{d}P_{A^{[1,L-1]}}(a)\\&=\mathtt{I}_1+\mathtt{I}_2\end{aligned}$$

where

$$\mathcal{Y}^{(L)}=\sqrt{\beta}\varphi_L\left(\sqrt{h_1}V+\sqrt{\rho_{L-1,n}-h_1}W,A^{(L)}\right)+Z.$$

By (5.4.2) and the definition of Ψ_L given in (5.1.13), $\mathtt{I}_1=\frac{n_L}{n}\Psi_L(h_1,\beta;\rho_{L-1,n})$. Completing the square, we can rewrite $\mathtt{I}_2$ as

$$\mathtt{I}_2=\frac{1}{n}\mathbb{E}\log\int e^{-\frac{1}{2}|Y'-\sqrt{h_2}x^{(L-1)}|^2}\mathrm{d}P_X(x)\mathrm{d}P_{A^{[1,L-1]}}(a)+\frac{1}{n}\mathbb{E}\log e^{\frac{1}{2}|Y'|^2}.\tag{5.4.3}$$

Recall the definition of $x^{(L-1)}$ in (5.1.5). We can define $x^{(L-2)}$ in the same fashion and it is related to $x^{(L-1)}$ via

$$x^{(L-1)} = \varphi_{L-1}\left(\frac{1}{\sqrt{n_{L-2}}}\Phi^{(L-1)}x^{(L-2)}, a^{(L-1)}\right).$$

Inserting this and $\mathrm{d}P_{A^{[1,L-1]}} = \mathrm{d}P_{A^{(L-1)}}\mathrm{d}P_{A^{[1,L-2]}}$ to (5.4.3), and using (5.1.6) with β, L replaced by $h_2, L-1$, we can see that the first term in (5.4.3) is given by

$$\frac{1}{n}\mathbb{E}\log\int \mathcal{P}_{h_2,L-1,n}\left(Y'\middle|\frac{1}{\sqrt{n_{L-2}}}\Phi^{(L-1)}x^{(L-2)}\right)\mathrm{d}P_X(x)\mathrm{d}P_{A^{[1,L-2]}}(a).$$

Recall the definition of Y' in (5.2.5). Comparing it with (5.1.4), we can see that Y' is exactly the observable for the $(L-1)$-layer model with $\beta = h_2$. In view of the definition of the original free energy in (5.1.8), we can see that the above display is exactly $\mathbb{E}F^{\circ}_{h_2,L-1,n}$. Now, we turn to the second term in (5.4.3). Using the definitions of Y' in (5.2.5) and $\rho_{L-1,n}$ in (5.1.10), we can compute that this term is equal to

$$\frac{1}{2n}\mathbb{E}|Y'|^2 = \frac{n_{L-1}}{2n}(1+\rho_{L-1,n}h_2).$$

We conclude that

$$\overline{F}_{\beta,L,n}(0,h) = \frac{n_L}{n}\Psi_L(h_1,\beta;\rho_{L-1,n}) + \mathbb{E}F^{\circ}_{h_2,L-1,n} + \frac{n_{L-1}}{2n}(1+\rho_{L-1,n}h_2),$$

which, by the induction assumption, converges pointwise on $[0,\rho_{L-1})\times\mathbb{R}_+$ to

$$\psi_{\beta,L}(h) = \alpha_L\Psi_L(h_1,\beta;\rho_{L-1}) + f^{\circ}_{h_2,L-1} + \frac{\alpha_{L-1}}{2}(1+\rho_{L-1}h_2)$$

where $f^{\circ}_{h_2,L-1}$ is the right-hand side of (5.1.15) with β, L replaced by $h_2, L-1$, namely,

$$f^{\circ}_{h_2,L-1} = \sup_{z^{(L-1)}}\inf_{y^{(L-1)}}\sup_{z^{(L-2)}}\inf_{y^{(L-2)}}\cdots\sup_{z^{(1)}}\inf_{y^{(1)}}\phi_{L-1}\left(h_2;y^{(1)},\cdots,y^{(L-1)};z^{(1)},\cdots,z^{(L-1)}\right)$$

with ϕ_{L-1} defined analogously as in (5.1.16).

Again, as argued in the base case, Lemma 5.2.2 enables us to verify all conditions in Proposition 5.3.3, which gives a weak solution $f_{\beta,L}$ satisfying $f_{\beta,L}(0,\cdot) = \psi_{\beta,L}$. Moreover, $f_{\beta,L}$ is given by the formula (5.3.7) with ρ, ψ there replaced by $\rho_{L-1}, \psi_{\beta,L}$, namely,

$$f_{\beta,L}(t,h) = \sup_{z^{(L)}} \inf_{y^{(L)}} \left\{ z^{(L)} \cdot \left(h - y^{(L)} \right) + \psi_{\beta,L}\left(y^{(L)} \right) + t\mathsf{H}_L\left(z^{(L)} \right) \right\}$$

for every $(t,h) \in \Omega_{\rho_{L-1}}$, where sup is taken over $z^{(L)} \in \mathbb{R}_+ \times [0, \frac{\alpha_{L-1}\rho_{L-1}}{2}]$ and inf is taken over $y^{(L)} \in [0, \rho_{L-1}] \times \mathbb{R}_+$. Inserting the previous two displays and the expression of H_L in (5.2.10) into the above, and evaluating at $(t,h) = (1,0)$, we obtain that

$$\begin{aligned} f_{\beta,L}(1,0) = \sup_{z^{(L)}} \inf_{y^{(L)}} \sup_{z^{(L-1)}} \inf_{y^{(L-1)}} \cdots \sup_{z^{(1)}} \inf_{y^{(1)}} \Bigg\{ & - y^{(L)} \cdot z^{(L)} + \alpha_L \Psi_L\left(y_1^{(L)}, \beta; \rho_{L-1} \right) \\ & + \phi_{L-1}\left(y_2^{(L)}; y^{(1)}, \cdots, y^{(L-1)}; z^{(1)}, \cdots, z^{(L-1)} \right) \\ & + \frac{\alpha_{L-1}}{2}\left(1 + \rho_{L-1} y_2^{(L)} \right) + \frac{2}{\alpha_{L-1}} z_1^{(L)} z_2^{(L)} \Bigg\}. \end{aligned}$$

We can verify that the expression inside the curly brackets is given by (5.1.16), and thus $f_{\beta,L}(1,0)$ is exactly the right-hand side of (5.1.15).

Again, the above verifies condition (2) and allows us to apply Proposition 5.4.1 to obtain that

$$\lim_{n\to\infty} \overline{F}_{\beta,L,n}(1,0) = f_{\beta,L}(1,0)$$

which along with (5.4.1) gives (5.1.15) and completes the proof of Theorem 5.1.1.

5.4.2. Proof of Proposition 5.4.1

For lighter notation, we suppress some of the subscripts and simply write

$$F_n = F_{\beta,L,n}, \quad \psi = \psi_{\beta,L}, \quad f = f_{\beta,L}, \quad \rho = \rho_{L-1}, \quad \rho_n = \rho_{L-1,n}.$$

We remark that it suffices to show

$$\lim_{n\to\infty}\sup_{t\in[0,1]}\int_{E_{n,t}(R)}\left|\overline{F}_n(t,h)-f(t,h)\right|\mathrm{d}h=0 \tag{5.4.4}$$

for every $R>0$, where

$$E_{n,t}(R)=[0,(\rho\wedge\rho_n)(1-t)]\times[0,R(1-t)]. \tag{5.4.5}$$

Indeed, for every $\rho'<\rho$, we have $\rho'<\rho\wedge\rho_n$ for sufficiently large n due to assumption (1). Then, (5.4.4) together with Fubini's theorem implies that the integral of $|\overline{F}_n-f|$ over $\Omega_{\rho'}\cap\{h_2\leqslant R(1-t)\}$ decays to 0 as $n\to\infty$, which further implies that $\overline{F}_n$ converges to f pointwise a.e. on $\Omega_{\rho'}\cap\{h_2\leqslant R(1-t)\}$. By enlarging R, we conclude that this convergence holds pointwise everywhere on $\Omega_{\rho'}$.

Let us show (5.4.4). Henceforth, we denote by C a positive constant independent of n,t,h, which may change from instance to instance. We also absorb R and ρ into C. Define $w_n=\overline{F}_n-f$ and

$$r_n=\partial_t\overline{F}_n-\mathsf{H}_L(\nabla\overline{F}_n). \tag{5.4.6}$$

Then, by the definition of H_L in (5.2.10), we have that

$$\partial_t w_n=b_n\cdot\nabla w_n+r_n \tag{5.4.7}$$

where

$$b_n=(b_{n,1},\ b_{n,2})=\frac{2}{\alpha_{L-1}}(\partial_2 f,\ \partial_1\overline{F}_n). \tag{5.4.8}$$

For $\delta\in(0,1)$, let $\phi_\delta:\mathbb{R}\to\mathbb{R}_+$ be given by

$$\phi_\delta(x)=\left(\delta+x^2\right)^{\frac{1}{2}},\quad \forall x\in\mathbb{R}, \tag{5.4.9}$$

which serves as a smooth approximation of the absolute value. Take $v_n = \phi_\delta(w_n)$ and multiply both sides of (5.4.7) by $\phi'_\delta(w_n)$ to see

$$\partial_t v_n = b_n \cdot \nabla v_n + \phi'_\delta(w_n) r_n \tag{5.4.10}$$

The Lipschitzness of f and that of $\overline{F}_n$ uniform in n due to (5.2.46) and (5.2.47) imply that, uniformly in n, δ,

$$|\nabla v_n| \leqslant C. \tag{5.4.11}$$

By $\lim_{n\to\infty} \overline{F}_n(0,0) = \psi(0) = f(0,0)$ due to assumption (2), we also get from the aforementioned Lipschitzness that

$$\sup_{\Omega_{\rho\wedge\rho_n}\cap\{h_2\leqslant R\}} |\overline{F}_n - f| \leqslant C \tag{5.4.12}$$

uniformly in n, which implies that, uniformly in n, δ,

$$\sup_{\Omega_{\rho\wedge\rho_n}\cap\{h_2\leqslant R\}} |v_n| \leqslant C. \tag{5.4.13}$$

Recall the mollifier ξ_ε given in (5.3.2) and that the mollification is well-defined on domain $\Omega_{\rho\wedge\rho_n,\varepsilon}$ described in (5.3.4). Let us regularize b_n by setting $b^\varepsilon_{n,i} = b_{n,i} * \xi_\varepsilon$, with the convolution taken in h. For $(t,h) \in \Omega_{\rho\wedge\rho_n,\varepsilon}$, we can rewrite (5.4.10) as

$$\partial_t v_n = \mathsf{div}(v_n b^\varepsilon_n) - v_n \mathsf{div} b^\varepsilon_n + (b_n - b^\varepsilon_n) \cdot \nabla v_n + \phi'_\delta(w_n) r_n. \tag{5.4.14}$$

By (5.4.8), (5.2.46), (5.2.47) and Definition 5.3.1 (1), there is $C > 0$ such that the following hold for all n, all $\varepsilon \in (0,1)$ and all $(t,h) \in \Omega_{\rho\wedge\rho_n,\varepsilon}$,

$$\|b_n - b_n^\varepsilon\|_\infty = o_\varepsilon(1); \tag{5.4.15}$$

$$\|b_n^\varepsilon\|_\infty \leqslant \|b_n\|_\infty \leqslant C;$$

$$b_{n,1}^\varepsilon \in [0,\rho], \quad b_{n,2}^\varepsilon \in [0,C]. \tag{5.4.16}$$

Using (5.2.49) and (3) in Definition 5.3.1, we also have that, for $(t,h) \in \Omega_{\rho\wedge\rho_n,\varepsilon}$,

$$\mathsf{div} b_n^\varepsilon = \frac{2}{\alpha_{L-1}} \big(\partial_1\partial_2\big(f * \xi_\varepsilon\big) + \partial_1\partial_2\big(\overline{F}_n * \xi_\varepsilon\big)\big) \geqslant 0. \tag{5.4.17}$$

Fix $R > \sup_{n,\varepsilon} \|b_n^\varepsilon\|_\infty$. In the following, we absorb R into C. Let $\eta > 0$ be specified later. Consider the following sets, indexed by $t \in [0, 1 - \frac{2}{\rho\wedge\rho_n}\eta]$,

$$D_t = [\eta, (\rho\wedge\rho_n)(1-t) - \eta] \times [\eta, R(1-t)], \tag{5.4.18}$$

$$\Gamma_{1,t} = [\eta, (\rho\wedge\rho_n)(1-t) - \eta] \times \{R(1-t)\},$$

$$\Gamma_{2,t} = \{(\rho\wedge\rho_n)(1-t) - \eta\} \times [\eta, R(1-t)],$$

where, for simplicity, we suppressed the dependence on n, η in the notation.

Let us consider the object

$$J_\delta(t) = \int_{D_t} v_n(t,h)\mathrm{d}h = \int_{D_t} \phi_\delta\big(w_n(t,h)\big)\mathrm{d}h. \tag{5.4.19}$$

Choose $\varepsilon < \eta$ to ensure that $\bigcup_{t\in[0,1-\frac{2}{\rho\wedge\rho_n}\eta]}(\{t\}\times D_t) \subseteq \Omega_{\rho\wedge\rho_n,\varepsilon}$. Differentiate $J_\delta(t)$ in t and use (5.4.14) to see

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}J_\delta(t) &= \int_{D_t}\partial_t v_n - R\int_{\Gamma_{1,t}} v_n - \rho\wedge\rho_n\int_{\Gamma_{2,t}} v_n\\ &= \int_{\Gamma_{1,t}}(\mathbf{n}\cdot b_n^\varepsilon - R)v_n + \int_{\Gamma_{2,t}}(\mathbf{n}\cdot b_n^\varepsilon - \rho\wedge\rho_n)v_n\\ &\quad+\int_{\partial D_t\setminus(\Gamma_{1,t}\cup\Gamma_{2,t})}(\mathbf{n}\cdot b_n^\varepsilon)v_n + \int_{D_t}\Big(-v_n\mathsf{div}b_n^\varepsilon + (b_n - b_n^\varepsilon)\cdot\nabla v_n + \phi_\delta'(w_n)r_n\Big).\end{aligned}$$

Here in the second identity, we used integration by parts on the integral of $\mathsf{div}(v_n b_n^\varepsilon)$. The first integral on the second line is nonpositive due to the choice of R. Then second integral on that line is bounded from above by $C|\rho_n-\rho|$ due to (5.4.13), (5.4.16) and the fact that on $\Gamma_{2,t}$ the outer normal $\mathbf{n}=(1,0)$. On the last line of the display, the first integral is nonpositive due to that $\mathbf{n}\in-\mathbb{R}^2_+$ on $\partial D_t\setminus(\Gamma_{1,t}\cup\Gamma_{2,t})$, and (5.4.16). It is clear from (5.4.9) that $\|\phi_\delta'\|_\infty\leqslant 1$. By this, (5.4.11), (5.4.15) and (5.4.17), the integrand in the last integral is bounded from above by $C(o_\varepsilon(1)+|r_n|)$. Therefore, sending $\varepsilon\to0$, we conclude that, for $t\in[0,1-\frac{2}{\rho\wedge\rho_n}\eta]$,

$$\frac{\mathrm{d}}{\mathrm{d}t}J_\delta(t)\leqslant C|\rho_n-\rho| + \int_{D_t}|r_n|. \tag{5.4.20}$$

Recall the definition of r_n in (5.4.6). Proposition 5.2.1 gives an upper bound for $|r_n|$, which along with Jensen's inequality gives that

$$\int_{D_t}|r_n|\leqslant C\bigg(\int_{D_t}\frac{1}{n}\partial_2^2\overline{F}_n + \mathbb{E}\int_{D_t}\big(\partial_2F_n-\partial_2\overline{F}_n\big)^2\bigg)^{\frac12} + a_n \tag{5.4.21}$$

for a_n bounded as in (5.2.11). In view of (5.2.47), the first integral on the right-hand side of (5.4.21) can be bounded by Cn^{-1}. For the last integral in (5.4.21), we will show that

$$\mathbb{E}\int_{D_t}\big|\partial_2(F_n-\overline{F}_n)\big|^2\leqslant\Delta_{1,n}^2\eta^{-\frac12}, \tag{5.4.22}$$

for some $\Delta_{1,n}$ converging to 0 as $n \to \infty$. These estimates imply that

$$\int_{D_t} |r_n| \leqslant C(n^{-\frac{1}{2}} + \Delta_{1,n}\eta^{-\frac{1}{4}} + a_n).$$

This along with (5.4.20) implies that

$$J_\delta(t) \leqslant J_\delta(0) + C(|\rho_n - \rho| + n^{-\frac{1}{2}} + \Delta_{1,n}\eta^{-\frac{1}{4}} + a_n), \quad t \in \left[0, 1 - \frac{2}{\rho \wedge \rho_n}\eta\right].$$

Note that $\lim_{n\to\infty} |\rho_n - \rho| = 0$ by assumption (1) and $\lim_{n\to\infty} a_n = 0$ due to (5.2.11), assumptions (3) and (4), and (5.1.1). By (5.4.9) and (5.4.19), we have that

$$\lim_{\delta\to 0} J_\delta(0) = \int_{D_0} \left|\overline{F}_n(0,h) - f(0,h)\right| \mathrm{d}h$$

which converges to 0 as $n \to \infty$ by assumption (2), (5.4.12) and the bounded convergence theorem. Hence, sending $\delta \to 0$, we derive that

$$\sup_{t\in[0,1-\frac{2}{\rho\wedge\rho_n}\eta]} \int_{D_t} \left|\overline{F}_n(t,h) - f(t,h)\right| \mathrm{d}h \leqslant C\big(\Delta_{1,n}\eta^{-\frac{1}{4}} + \Delta_{2,n}\big),$$

for some $\Delta_{2,n}$ that decays to 0 as $n \to \infty$. We want to extend the above result from integrating over D_t to $E_{n,t}(R)$ for $t \in [0,1]$. The definitions of $E_{n,t}(R)$ in (5.4.5) and D_t in (5.4.18) give that

$$\begin{aligned} |E_{n,t}(R) \setminus D_t| &\leqslant C\eta, \quad \forall t \in \left[0,\ \frac{2}{\rho\wedge\rho_n}\eta\right], \\ |E_{n,t}(R)| &\leqslant C\eta, \quad \forall t \in \left[\frac{2}{\rho\wedge\rho_n}\eta,\ 1\right]. \end{aligned}$$

These along with (5.4.12) yield that

$$\sup_{t\in[0,1-\frac{2}{\rho\wedge\rho_n}\eta]}\int_{E_{n,t}(R)\setminus D_t}\big|\overline{F}_n(t,h)-f(t,h)\big|\mathrm{d}h\leqslant C\eta,$$
$$\sup_{t\in[1-\frac{2}{\rho\wedge\rho_n}\eta,1]}\int_{E_{n,t}(R)}\big|\overline{F}_n(t,h)-f(t,h)\big|\mathrm{d}h\leqslant C\eta.$$

Therefore, we obtain that

$$\sup_{t\in[0,1]}\int_{E_{n,t}(R)}\big|\overline{F}_n(t,h)-f(t,h)\big|\mathrm{d}h\leqslant C\big(\eta+\Delta_{1,n}\eta^{-\frac{1}{4}}+\Delta_{2,n}\big).$$

Insert $\eta=\Delta_{1,n}^{\frac{4}{5}}$ into the above display to see that the right-hand side of the above is bounded by $C(\Delta_{1,n}^{\frac{4}{5}}+\Delta_{2,n})$, which gives the desired result (5.4.4).

It remains to verify (5.4.22).

Proof of (5.4.22)

By writing

$$\mathbb{E}\int_{D_t}\big|\partial_2(F_n-\overline{F}_n)\big|^2=\int_\eta^{(\rho\wedge\rho_n)(1-t)}\left(\mathbb{E}\int_\eta^{R(1-t)}\big|\partial_2(F_n-\overline{F}_n)\big|^2\mathrm{d}h_2\right)\mathrm{d}h_1,$$

it suffices to show that the term inside the parentheses is $o(1)\eta^{-\frac{1}{2}}$ uniformly in t,h_1. Now, let us fix any (t,h_1) and investigate the integration with respect to h_2. Integration by parts yields that

$$\int_\eta^{R(1-t)} \left|\partial_2(F_n - \overline{F}_n)\right|^2 = (F_n - \overline{F}_n)\partial_2(F_n - \overline{F}_n)\big|_{h_2=R(1-t)} - (F_n - \overline{F}_n)\partial_2(F_n - \overline{F}_n)\big|_{h_2=\eta}$$
$$- \int_\eta^{R(1-t)} (F_n - \overline{F}_n)\partial_2^2(F_n - \overline{F}_n)$$
$$\leqslant \|F_n - \overline{F}_n\|_{L^\infty_{h_2}([0,R])}\Bigg(\left|\partial_2(F_n - \overline{F}_n)\big|_{h_2=R(1-t)}\right| + \left|\partial_2(F_n - \overline{F}_n)\big|_{h_2=\eta}\right|$$
$$(5.4.23)$$
$$+ \int_\eta^{R(1-t)} \left|\partial_2^2(F_n - \overline{F}_n)\right|\Bigg).$$

Let us estimate the last integral. By (5.2.49) and (5.2.50),

$$\partial_2^2\overline{F}_n \geqslant 0, \qquad \partial_2^2 F_n + Cn^{-\frac{1}{2}}h_2{}^{-\frac{3}{2}}|Z'| \geqslant 0,$$

which implies that

$$\int_\eta^{R(1-t)} \left|\partial_2^2(F_n + \overline{F}_n)\right| \leqslant \int_\eta^{R(1-t)} \left|\partial_2^2 F_n\right| + \left|\partial_2^2\overline{F}_n\right|$$
$$\leqslant \int_\eta^{R(1-t)} \left(\partial_2^2 F_n + \partial_2^2\overline{F}_n\right) + \int_\eta^{R(1-t)} 2Cn^{-\frac{1}{2}}h_2{}^{-\frac{3}{2}}|Z'|.$$

Applying integration by parts to the first integral after the second inequality gives that

$$\int_\eta^{R(1-t)} \left|\partial_2^2(F_n + \overline{F}_n)\right|$$
$$\leqslant \left(\left|\partial_2 F_n\right| + \left|\partial_2\overline{F}_n\right|\right)\Big|_{h_2=R(1-t)} - \left(\left|\partial_2 F_n\right| + \left|\partial_2\overline{F}_n\right|\right)\Big|_{h_2=\eta} + Cn^{-\frac{1}{2}}\eta^{-\frac{1}{2}}|Z'|$$
$$\leqslant C(1 + n^{-\frac{1}{2}}\eta^{-\frac{1}{2}}|Z'|)$$

where the last inequality follows from the estimates of $\partial_2\overline{F}_n$ in (5.2.47) and $\partial_2 F_n$ in (5.2.48). Insert estimates (5.2.47) and (5.2.48), and the above display into (5.4.23) to get that

$$\int_\eta^{R(1-t)} \left|\partial_2(F_n - \overline{F}_n)\right|^2 \leqslant C\|F_n - \overline{F}_n\|_{L^\infty_{h_2}([0,R])}\left(1 + n^{-\frac{1}{2}}\eta^{-\frac{1}{2}}|Z'|\right).$$

Take expectations on both sides of this inequality, invoke the Cauchy–Schwarz inequality and use assumption (4) to conclude (5.4.22).

5.5. Auxiliary results

We collect proofs of Lemma 5.5.1 which verifies (5.1.11), Lemma 5.5.2 which gives the concentration of $\frac{1}{n_l}\left|X^{(l)}\right|^2$, and Lemma 5.5.4 which shows that the concentration condition (4) in Proposition 5.4.1 always holds.

5.5.1. Convergence of the averaged norm

Recall $\rho_{l,n}$ from (5.1.10).

Lemma 5.5.1. *Assume (H1)–(H3) for some $L \in \mathbb{N}$. For each $l \in \{0, 1, \ldots, L\}$,* (5.1.11) *holds for ρ_l defined iteratively by*

$$\begin{aligned} \rho_0 &= \mathbb{E}|X_1|^2 \\ \rho_l &= \mathbb{E}\left|\varphi_l\left(\sqrt{\rho_{l-1}}\Phi_{11}^{(l)}, A_1^{(l)}\right)\right|^2. \end{aligned} \tag{5.5.1}$$

In (5.5.1), $\Phi_{11}^{(l)}$ is a standard Gaussian random variable independent of $A_1^{(l)}$. Examining the proof below, we can see that the lemma is still valid (with ρ_0 defined as a limit) if we replace (H1) and (H2) by weaker assumptions that $\frac{1}{n}|X|^2$ converges in probability together with (h1), and that φ_l is Lipschitz for all l.

Proof. It suffices to show that

$$\lim_{n\to\infty} \mathbb{E}\left|\frac{\left|X^{(l)}\right|^2}{n_l} - \rho_l\right|^2 = 0. \tag{5.5.2}$$

Since $X^{(0)}$ is assumed to consist of bounded i.i.d. entries and $\rho_0 = \mathbb{E}|X_j|^2$ for all $j = 1, 2, \ldots, n$, it is immediate that (5.5.2) holds for $l = 0$. We proceed by induction. Now, we assume that (5.5.2) holds for $l - 1$. Let us denote by $\mathbb{E}^{(l)}$ the expectation with respect to

$\Phi^{(l)}$ and $A^{(l)}$. We start by writing

$$\mathbb{E}\left|\frac{\left|X^{(l)}\right|^2}{n_l}-\rho_l\right|^2 \leqslant 2\mathbb{E}\left|\frac{\left|X^{(l)}\right|^2}{n_l}-\mathbb{E}^{(l)}\frac{\left|X^{(l)}\right|^2}{n_l}\right|^2+2\mathbb{E}\left|\mathbb{E}^{(l)}\frac{\left|X^{(l)}\right|^2}{n_l}-\rho_l\right|^2. \tag{5.5.3}$$

We start by estimating the first term on the right. It is clear from (5.1.3) that, conditioned on $X^{(l-1)}$, $(|X_j^{(l)}|^2)_{j=1}^{n_l}$ is a sequence of i.i.d. random varaibles. Hence, the first term can be rewritten as

$$2\mathbb{E}\frac{1}{n_l^2}\sum_{j=1}^{n_l}\mathbb{E}^{(l)}\left|\left|X_j^{(l)}\right|^2-\mathbb{E}^{(l)}\left|X_j^{(l)}\right|^2\right|^2.$$

Since $X_j^{(l)}$ is bounded, we can see that the first term is bounded by Cn_l^{-1}. Now, we turn to the second term. Using (5.1.3), we can compute that

$$\mathbb{E}^{(l)}\frac{\left|X^{(l)}\right|^2}{n_l}=g\left(\frac{\left|X^{(l-1)}\right|^2}{n_{l-1}}\right)$$

where

$$g(\sigma)=\mathbb{E}\left|\varphi_l\left(\sqrt{\sigma}\Phi_{11}^{(l)},A_1^{(l)}\right)\right|^2.$$

Since φ_l is assumed to have bounded derivatives, we can see that g is $\frac{1}{2}$-Hölder continuous. Rewriting (5.5.1) as $\rho_l=g(\rho_{l-1})$, we can bound the second term in (5.5.3) by

$$2\mathbb{E}\left|g\left(\frac{\left|X^{(l-1)}\right|^2}{n_{l-1}}\right)-g(\rho_{l-1})\right|^2\leqslant C\mathbb{E}\left|\frac{\left|X^{(l-1)}\right|^2}{n_{l-1}}-\rho_{l-1}\right|$$

which converges to 0 due to the induction assumption (5.5.2) for $l-1$. This finishes the induction step showing that (5.5.2) holds for l and thus completes the proof. □

5.5.2. Concentration of the norm

The goal is to show the following lemma.

Lemma 5.5.2. *Assume (H1)–(H3) for some $L \in \mathbb{N}$. There is a constant $C > 0$ such that, for every $n \in \mathbb{N}$,*

$$\mathbb{V}\mathrm{ar}\left[\frac{1}{n_L}\left|X^{(L)}\right|^2\right] \leqslant \frac{C}{n}.$$

To prove this, we need a classic result on concentration.

Lemma 5.5.3. *Let $A_1, A_2, \ldots, A_n$ be independent random variables with values in some space $\mathcal{X}$. Suppose that a function $f : \mathcal{X}^n \to \mathbb{R}$ satisfies*

$$\sup_{1\leqslant i\leqslant n} \sup_{\substack{a_1,\ldots,a_n,\\ a_i'\in\mathcal{X}}} |f(a_1,\ldots,a_n) - f(a_1,\ldots,a_{i-1},a_i',a_{i+1},\ldots,a_n)| \leqslant c$$

for some $c > 0$. Then, $\mathbb{V}\mathrm{ar}[f(A)] \leqslant \frac{1}{4}nc^2$.

This is a corollary of the Efron–Stein inequality. We refer to [27, Corollary 3.2] for a proof.

Proof of Lemma 5.5.2. Setting

$$g_L(x) = \frac{1}{n_L}|x|^2, \quad \forall x \in \mathbb{R}^{n_L},$$

we have that

$$g_L\left(X^{(L)}\right) = \frac{1}{n_L}\left|X^{(L)}\right|^2. \tag{5.5.4}$$

For $l \in \{0, 1, \ldots, L-1\}$, we can iteratively define

$$g_l(x) = \mathbb{E}\left[g_{l+1}\left(\varphi_{l+1}\left(\frac{1}{\sqrt{n_l}}\Phi^{(l+1)}x, A^{(l+1)}\right)\right)\right], \quad \forall x \in \mathbb{R}^{n_l}. \tag{5.5.5}$$

Due to (5.1.3), this implies that

$$g_l\left(X^{(l)}\right) = \mathbb{E}\left[g_{l+1}\left(X^{(l+1)}\right)\middle|X^{(l)}\right].$$

For convenience, we also set

$$X^{(-1)} = 0, \qquad g_{-1}(0) = \mathbb{E}\left[g_0\left(X^{(0)}\right)\right].$$

Iterating these yields that

$$g_{-1}\left(X^{(-1)}\right) = \mathbb{E}\left[g_L\left(X^{(L)}\right)\right] = \mathbb{E}\left[\frac{1}{n_L}\left|X^{(L)}\right|^2\right],$$

which along with (5.5.4) gives that

$$\begin{aligned}\mathbb{V}\mathrm{ar}\left[\frac{1}{n_L}\left|X^{(L)}\right|^2\right] &= \mathbb{E}\left[\left(g_L\left(X^{(L)}\right)\right)^2 - \left(g_{-1}\left(X^{(-1)}\right)\right)^2\right] \\ &= \sum_{l=0}^{L}\mathbb{E}\left[\left(g_l\left(X^{(l)}\right)\right)^2 - \left(g_{l-1}\left(X^{(l-1)}\right)\right)^2\right] \\ &= \sum_{l=0}^{L}\mathbb{E}\left[\left(g_l\left(X^{(l)}\right) - \mathbb{E}\left[g_l\left(X^{(l)}\right)\middle|X^{(l-1)}\right]\right)^2\right].\end{aligned}$$

Then, the desired result follows if we can show that, for all $l \in \{0, 1, \ldots, L\}$,

$$\mathbb{E}\left[\left(g_l\left(X^{(l)}\right) - \mathbb{E}\left[g_l\left(X^{(l)}\right)\middle|X^{(l-1)}\right]\right)^2\right] \leqslant \frac{C}{n}. \tag{5.5.6}$$

For $l = L$, since $X^{(L)}$ has i.i.d. entries when conditioned on $X^{(L-1)}$ due to (5.1.3), the left-hand side of (5.5.6) is given by

$$\begin{aligned}&\mathbb{E}\left[\mathbb{E}^{(L)}\left[\left(\frac{1}{n_L}\left|X^{(L)}\right|^2 - \mathbb{E}^{(L)}\frac{1}{n_L}\left|X^{(L)}\right|^2\right)^2\right]\right] \\ &= \frac{1}{n_L}\mathbb{E}\left[\left(\left|X_1^{(L)}\right|^2 - \mathbb{E}^{(L)}\left|X_1^{(L)}\right|^2\right)^2\right] \leqslant \frac{C}{n_L}\end{aligned}$$

where $\mathbb{E}^{(L)}$ is the expectation with respect to $\Phi^{(L)}$ and $A^{(L)}$.

Now, let $l \leqslant L-1$. Due to (5.1.3), $X^{(l)}$ has i.i.d. entries when conditioned on $X^{(l-1)}$. Recall the notation (5.1.2). Due to (5.1.3), viewing $X^{(L)}$ as a deterministic function of $\Phi^{[l+1,m]}$,

$A^{[l+1,m]}$ and $X^{(l)}$, and using (5.5.5), we can check inductively that

$$g_l\left(X^{(l)}\right) = \mathbb{E}\left[\frac{1}{n_L}\left|X^{(L)}\right|^2 \middle| X^{(l)}\right].$$

Then, using (5.1.3) and the chain rule, we can compute that for $i_l \in \{1, 2, \ldots, n_l\}$,

$$\frac{\partial g_l\left(X^{(l)}\right)}{\partial X^{(l)}_{i_l}} = \\ \frac{2}{n_L}\sum_{\mathbf{i}}\mathbb{E}\left[\dot{\varphi}^{(l+1)}_{i_{l+1}}\dot{\varphi}^{(l+2)}_{i_{l+2}}\cdots\dot{\varphi}^{(L)}_{i_L}\frac{\Phi^{(l+1)}_{i_{l+1},i_l}}{\sqrt{n_l}}\frac{\Phi^{(l+2)}_{i_{l+2},i_{l+1}}}{\sqrt{n_{l+1}}}\cdots\frac{\Phi^{(L-1)}_{i_{L-1},i_{L-2}}}{\sqrt{n_{L-2}}}\frac{\Phi^{(L)}_{i_L,i_{L-1}}}{\sqrt{n_{L-1}}}\middle| X^{(l)}\right], \tag{5.5.7}$$

where the summation is over

$$\mathbf{i} = (i_{l+1}, i_{l+2}, \ldots, i_L) \in \prod_{m=l+1}^{L}\{1, \ldots, n_m\} \tag{5.5.8}$$

and

$$\dot{\varphi}^{(m)}_{i_m} = \varphi'_m\left(\frac{1}{\sqrt{n_{m-1}}}\left(\Phi^{(m)}X^{(m-1)}\right)_{i_m}, A^{(m)}_{i_m}\right), \quad \forall i_m \in \{1, \ldots, n_m\}.$$

The derivative on φ_m is with respect to its first argument.

To proceed, we want to perform the Gaussian integration by parts one every $\Phi^{(m)}_{i_m,i_{m-1}}$ in every summand on the right-hand side of (5.5.7). The heuristics is that since $\Phi^{(m)}_{i_m,i_{m-1}}$ always appears in the form of $\frac{1}{\sqrt{n_{m-1}}}\Phi^{(m)}X^{(m-1)}$, we expect to obtain an extra factor of order $n^{-\frac{1}{2}}$ after performing one instance of integration by parts. However, due to the layered structure given in (5.1.3) and the chain rule, the differentiation involved in the process of integration by parts may produce new terms, the number of which grows as n increases. To cancel this effect, we need to perform more instances of integration by parts on Gaussian variables introduced by the chain rule.

The above heuristics is made rigorous by Corollary 5.5.7 which follows from a more general result Lemma 5.5.6. Applying Corollary 5.5.7, we obtain that each summand in (5.5.7) has its absolute value bounded by $Cn^{-(L-l)}$ where C is absolute. Due to (5.5.8), the summation in (5.5.7) is over $O(n^{L-l})$ many terms. Therefore, we conclude that, for each $i_l \in \{1, 2, \ldots, n_l\}$,

$$\left| \frac{\partial g_l \left(X^{(l)} \right)}{\partial X^{(l)}_{i_l}} \right| \leqslant \frac{C}{n}.$$

Invoking Lemma 5.5.3, we obtain that there is a constant C such that, for almost every realization of $X^{(l-1)}$,

$$\mathbb{E}\left[\left(g_l \left(X^{(l)} \right) - \mathbb{E}\left[g_l \left(X^{(l)} \right) \middle| X^{(l-1)} \right] \right)^2 \middle| X^{(l-1)} \right] \leqslant \frac{C}{n},$$

which then gives (5.5.6) and completes the proof. □

5.5.3. Concentration of the free energy

Recall the definitions of $\mathcal{P}_{\beta,L,n}$, $H_{\beta,L,n}$ and $F_{\beta,L,n}$ given in (5.1.6), (5.2.6), and (5.2.8). The goal is to show the lemma below.

Lemma 5.5.4. *Assume (H1)–(H3) for some $L \in \mathbb{N}$. For every $\beta \geqslant 0$, and $M \geqslant 1$, there is a constant $C > 0$ such that*

$$\sup_{t\in[0,1],\ h_1\in[0,\rho_n(1-t)]} \mathbb{E}\left[\left\| F_{\beta,L,n} - \overline{F}_{\beta,L,n} \right\|^2_{L^\infty_{h_2}([0,M])} (t, h_1) \right] \leqslant \frac{C}{\sqrt{n}}.$$

The remaining part of this subsection is devoted to the proof of this lemma. In addition to Lemma 5.5.3, we recall one more classic result on concentration.

Lemma 5.5.5. *Let $Z = (Z_1, Z_2, \ldots, Z_n)$ be a standard Gaussian vector and $f : \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable function. Then $\mathbb{V}\mathrm{ar}[f(Z)] \leqslant \mathbb{E}|\nabla f(Z)|^2$.*

This result is often called the Gaussian Poincaré inequality, whose proof we refer to that of [27, Theorem 3.20].

Let $h_2 \in [0, M]$. In the following, $C > 0$ denotes a deterministic constant independent of n, which may differ from line to line. We also absorb M and β into C. For simplicity, we write $H = H_{\beta,L,n}$ and $F = F_{\beta,L,n}$. In addition, we set

$$\Gamma\left(s, a^{(L)}\right) = \varphi_L\left(S, A^{(L)}\right) - \varphi_L\left(s, a^{(L)}\right), \tag{5.5.9}$$

where S and s are defined in (5.2.2) and (5.2.3), respectively, and

$$a^{(L)} \in \mathbb{R}^{n_L \times k_L} \tag{5.5.10}$$

is of the same size as $A^{(L)}$. In view of (5.1.6) and (5.2.6), note that H can be rewritten as

$$H(x, w, a) = \log\left(\int e^{-\frac{1}{2}\left|\sqrt{\beta}\Gamma\left(s, a^{(L)}\right) + Z\right|^2} \mathrm{d}P_{A^{(L)}}(a^{(L)})\right) + \sqrt{h_2}Y' \cdot x^{(L-1)} - \frac{h_2}{2}\left|X^{(L-1)}\right|^2,$$

where Y' is given in (5.2.5) and $a = (a^{(1)}, \cdots, a^{(L-1)})$ appearing in $x^{(L-1)}$ is defined in (5.1.5). We introduce the Hamiltonian

$$\begin{aligned}&\widehat{H}\left(x, w, a, a^{(L)}\right)\\ &\quad = -\frac{1}{2}\left(2\sqrt{\beta}Z \cdot \Gamma\left(s, a^{(L)}\right) + \beta\left|\Gamma\left(s, a^{(L)}\right)\right|^2\right) + \sqrt{h_2}Y' \cdot x^{(L-1)} - \frac{h_2}{2}\left|x^{(L-1)}\right|^2,\end{aligned} \tag{5.5.11}$$

and the associated free energy

$$\widehat{F} = \frac{1}{n}\log\int e^{\widehat{H}(x,w,a,a^{(L)})} \mathrm{d}P_X(x)\mathrm{d}P_W(w)\mathrm{d}P_{A^{[1,L-1]}}(a)\mathrm{d}P_{A^{(L)}}\left(a^{(L)}\right).$$

Then, using these and the definition of F in terms in of H in (5.2.8), we can see that

$$F = \widehat{F} - \frac{1}{2n}|Z|^2,$$

which implies that

$$\mathbb{V}\mathrm{ar}(F) \leqslant 2\mathbb{V}\mathrm{ar}(\widehat{F}) + 2\mathbb{V}\mathrm{ar}\left(\frac{1}{2n}|Z|^2\right) \leqslant 2\mathbb{V}\mathrm{ar}(\widehat{F}) + \frac{C}{n}, \tag{5.5.12}$$

where we used the fact that Z is a standard Gaussian vector in $\mathbb{R}^{n_L}$. Therefore, it suffices to study $\mathbb{V}\mathrm{ar}(\widehat{F})$. In the sequel, we denote by $\langle\,\cdot\,\rangle$ the Gibbs measure with Hamiltonian $\widehat{H}$.

Recall the notation (5.1.2). Note that $\widehat{F}$ is a function of $Z, Z', V, W, A^{(L)}, \Phi^{[1,L]}, X^{(L-1)}$, where the dependence on Z is in (5.5.11); $\Phi^{(L)}, X^{(L-1)}, V, W$ appear in S defined in (5.2.2); $X^{(L-1)}, Z'$ appear in Y' defined in (5.2.5); $A^{(L)}$ appears in (5.5.9); $\Phi^{[1,L-1]}$ appears in $x^{(L-1)}$ defined in (5.1.5); and finally $\Phi^{(L)}, V, x^{(L-1)}$ appear in s defined in (5.2.3).

The plan is to prove concentration of $\widehat{F}$ conditioned on subsets of these random variables, and then combine them together. The order of conditioning matters and we proceed as in [65]. Lastly, to get concentration uniformly in $h_2 \in [0, M]$, we will apply an ε-net argument.

Concentration conditioned on $V, W, A^{(L)}, \Phi^{[1,L]}, X^{(L-1)}$

Denote by $\mathbb{E}_{Z,Z'}$ the expectation with respect to only Z and Z'. We want to show that

$$\mathbb{E}_{Z,Z'}\left(\widehat{F} - \mathbb{E}_{Z,Z'}\widehat{F}\right)^2 \leqslant \frac{C}{n}, \tag{5.5.13}$$

for almost every realization of other randomness.

For simplicity, we write $\Gamma = \Gamma\left(s, a^{(L)}\right)$ from now on. We fix any realization of other randomness. Note that Z appears only in (5.5.11) and Z' appears only in Y' (defined in (5.2.5)). Then, we can compute that

$$\begin{aligned}
\left|\frac{\partial \widehat{F}}{\partial Z_j}\right| &= \frac{1}{n}\left|\left\langle \sqrt{\beta}\Gamma_j\right\rangle\right| \leqslant \frac{C}{n}, \quad \forall j \in \{1, 2, \ldots, n_L\} \\
\left|\frac{\partial \widehat{F}}{\partial Z'_i}\right| &= \frac{1}{n}\left|\sqrt{h_2}\left\langle x_i^{(L-1)}\right\rangle\right| \leqslant \frac{C}{n}, \quad \forall i \in \{1, 2, \ldots, n_{L-1}\},
\end{aligned}$$

where we used the boundedness of φ_L, and the boundedness of $x^{(L-1)}$ to get the inequalities.

Hence, we have that $|\nabla_{Z,Z'}\widehat{F}| \leqslant Cn^{-\frac{1}{2}}$ and thus, by Lemma 5.5.5, obtain (5.5.13).

Concentration conditioned on $A^{(L)}, \Phi^{[1,L]}, X^{(L-1)}$

Set $\mathbf{g} = (Z, Z', V, W, \Phi^{(L)})$, and let $\mathbb{E}_{\mathbf{g}}$ be the expectation with respect to these Gaussian random variables. We want to show that, a.s.,

$$\mathbb{E}_{\mathbf{g}}\left(\mathbb{E}_{Z,Z'}\widehat{F} - \mathbb{E}_{\mathbf{g}}\widehat{F}\right)^2 \leqslant \frac{C}{n}. \tag{5.5.14}$$

Note that V appears in both S (defined in (5.2.2)) and s (defined in (5.2.3)) in Γ and W appears only in S. Hence, in view of (5.5.9), using the boundedness for the derivatives of φ_L, we can verify that

$$\left|\frac{\partial \mathbb{E}_{Z,Z'}\widehat{F}}{\partial V_j}\right| = \frac{1}{n}\left|\mathbb{E}_{Z,Z'}\left\langle\left(\sqrt{\beta}Z_j + \beta\Gamma_j\right)\frac{\partial \Gamma_j}{\partial V_j}\right\rangle\right| \leqslant \frac{C}{n}, \quad \forall j \in \{1, 2, \ldots, n_L\},$$
$$\left|\frac{\partial \mathbb{E}_{Z,Z'}\widehat{F}}{\partial W_j}\right| = \frac{1}{n}\left|\mathbb{E}_{Z,Z'}\left\langle\left(\sqrt{\beta}Z_j + \beta\Gamma_j\right)\frac{\partial \Gamma_j}{\partial W_j}\right\rangle\right| \leqslant \frac{C}{n}, \quad \forall j \in \{1, 2, \ldots, n_L\}.$$

On the other hand, $\Phi^{(L)}$ only appear in both S and s. Due to the computation that

$$\frac{\partial \Gamma_j}{\partial \Phi_{jk}^{(L)}} = \sqrt{\frac{t}{n_{L-1}}}\left(\varphi_L'\left(S, A^{(L)}\right)X_k^{(L-1)} - \varphi_L'\left(s, a^{(L)}\right)x_k^{(L-1)}\right),$$

where φ_L' is the derivative with respect to its first argument, and the boundedness of the derivatives of φ_L, we also can show that

$$\left|\frac{\partial \mathbb{E}_{Z,Z'}\widehat{F}}{\partial \Phi_{jk}^{(L)}}\right| = \frac{1}{n}\left|\mathbb{E}_{Z,Z'}\left\langle\left(\sqrt{\beta}Z_j + \beta\Gamma_j\right)\frac{\partial \Gamma_j}{\partial \Phi_{jk}^{(L)}}\right\rangle\right| \leqslant \frac{C}{n^{\frac{3}{2}}}.$$

for all $j \in \{1, \ldots, n_L\}$ and $k \in \{1, \ldots, n_{L-1}\}$. Therefore,

$$\left|\nabla_{V,W,\Phi^{(L)}}\mathbb{E}_{Z,Z'}\widehat{F}\right|^2 = \sum_{j=1}^{n_L}\left|\frac{\partial \mathbb{E}_{Z,Z'}\widehat{F}}{\partial V_j}\right|^2 + \sum_{j=1}^{n_L}\left|\frac{\partial \mathbb{E}_{Z,Z'}\widehat{F}}{\partial W_j}\right|^2 + \sum_{j=1}^{n_L}\sum_{k=1}^{n_{L-1}}\left|\frac{\partial \mathbb{E}_{Z,Z'}\widehat{F}}{\partial \Phi_{jk}^{(L)}}\right|^2 \leqslant \frac{C}{n},$$

which together with Lemma 5.5.5 implies (5.5.14).

Concentration conditioned on $\Phi^{[1,L]}, X^{(L-1)}$

Fixing any realization of other randomness, we express $\mathbb{E}_{\mathbf{g}} \widehat{F} = g(A^{(L)})$ as a function of $A^{(L)}$. Then, we fix a realization of $A^{(L)}$ and let $A'^{(L)}$ be another realization such that $A_j^{(L)} = A'^{(L)}_j$ for all j except for some $j = i$. We want to show that there is an absolute constant C such that

$$\left| g\left(A^{(L)}\right) - g\left(A'^{(L)}\right) \right| \leqslant \frac{C}{n}, \tag{5.5.15}$$

which by Lemma 5.5.3 implies that, a.s.,

$$\mathbb{E}_{\mathbf{g}, A^{(L)}} \left(\mathbb{E}_{\mathbf{g}} \widehat{F} - \mathbb{E}_{\mathbf{g}, A^{(L)}} \widehat{F} \right)^2 \leqslant \frac{C}{n}. \tag{5.5.16}$$

We denote by $\langle \, \cdot \, \rangle_{\widehat{\mathsf{H}}}$ the Gibbs measure with $A^{(L)}$ and $\langle \, \cdot \, \rangle_{\widehat{\mathsf{H}}'}$ the Gibbs measure with $A'^{(L)}$. Using the definition of g, we can verify that

$$g\left(A^{(L)}\right) - g\left(A'^{(L)}\right) = \frac{1}{n} \mathbb{E}_{\mathbf{g}} \log \left\langle e^{\widehat{H} - \widehat{H}'} \right\rangle_{\widehat{H}'}.$$

By Jensen's inequality, we have that

$$g\left(A^{(L)}\right) - g\left(A'^{(L)}\right) \geqslant \frac{1}{n} \mathbb{E}_{\mathbf{g}} \left\langle \widehat{H} - \widehat{H}' \right\rangle_{\widehat{H}'}.$$

Symmetrically,

$$g\left(A'^{(L)}\right) - g\left(A^{(L)}\right) \geqslant \frac{1}{n} \mathbb{E}_{\mathbf{g}} \left\langle \widehat{H}' - \widehat{H} \right\rangle_{\widehat{H}}.$$

Using (5.5.9), (5.5.11) and the definitions of $A^{(L)}$ and $A'^{(L)}$, we have that

$$\widehat{H} - \widehat{H}' = \frac{1}{2} \left(\Gamma_i' - \Gamma_i \right) \left(2Z_i + \Gamma_i + \Gamma_i' \right)$$

where Γ_i and Γ_i' correspond to $A^{(L)}$ and $A'^{(L)}$, respectively. Together with the boundedness of Γ, Γ', the above three displays yield (5.5.15) and thus imply the desired result (5.5.16).

Iteration

Note that in (5.5.16), we can rewrite that

$$\mathbb{E}_{\mathbf{g},A^{(L)}}\widehat{F} = \mathbb{E}\left[\widehat{F}\middle|X^{(L-1)}, \Phi^{[1,L]}\right].$$

To proceed, we claim that

$$\mathbb{E}\left(\mathbb{E}\left[\widehat{F}\middle|X^{(l)}, \Phi^{[1,l]}\right] - \mathbb{E}\left[\widehat{F}\middle|X^{(l-1)}, \Phi^{[1,l]}\right]\right)^2 \leqslant \frac{C}{n}, \qquad \forall l \in \{0, 1, \dots, L-1\}, \tag{5.5.17}$$

$$\mathbb{E}\left(\mathbb{E}\left[\widehat{F}\middle|X^{(l-1)}, \Phi^{[1,l]}\right] - \mathbb{E}\left[\widehat{F}\middle|X^{(l-1)}, \Phi^{[1,l-1]}\right]\right)^2 \leqslant \frac{C}{n}, \qquad \forall l \in \{1, \dots, L-1\}, \tag{5.5.18}$$

where $X^{(-1)}$ and $\Phi^{[1,0]}$ are understood to be constantly 0 (or any constant). Given the above, we can iterate these to see that

$$\mathbb{E}\left(\mathbb{E}\left[\widehat{F}\middle|X^{(L-1)}, \Phi^{[1,L-1]}\right] - \mathbb{E}\left[\widehat{F}\right]\right)^2 \leqslant \frac{C}{n}. \tag{5.5.19}$$

Combining (5.5.12), (5.5.13), (5.5.14), (5.5.16) and (5.5.19) yields the pointwise concentration

$$\mathbb{E}\left[\left(F - \overline{F}\right)^2(t,h)\right] \leqslant \frac{C}{n}, \quad \forall (t,h) \in \Omega_{\rho_n} \cap \{|h_2| \leqslant M\}. \tag{5.5.20}$$

Then, let us prove the assertions (5.5.17) and (5.5.18).

Proof of (5.5.17)

Due to the expression (5.1.3) and the fact that $\widehat{F}$ depends on $X^{(l-1)}$ only through $X^{(l)}$, we can see that

$$\mathbb{E}\left[\widehat{F}\middle|X^{(l)}, \Phi^{[1,l]}\right] = \mathbb{E}\left[\widehat{F}\middle|X^{(l)}, X^{(l-1)}, \Phi^{[1,l]}\right].$$

Also, note that $X^{(l)}$ consists of i.i.d. entries when conditioned on $X^{(l-1)}$. Hence, we want to apply Lemma 5.5.3. Since each entry of $X^{(l)}$ is bounded uniformly in n, to verify the condition in Lemma 5.5.3, it suffices to obtain bounds for derivatives of $\widetilde{\mathbb{E}}\widehat{F}$ with respect to $X^{(l)}$, where $\widetilde{\mathbb{E}} = \mathbb{E}\left[\,\cdot\,|X^{(l)}, X^{(l-1)}, \Phi^{[1,l]}\right]$.

We introduce the following notation:

$$\begin{cases} \varphi_*^{(L)} = \varphi_L\left(S, A^{(L)}\right), \\ \widetilde{\varphi}_*^{(L)} = \varphi_L\left(s, a^{(L)}\right), \\ \dot{\varphi}^{(m)} = \varphi_m'\left(\frac{1}{\sqrt{n_{m-1}}}\Phi^{(m)}X^{(m-1)}, A^{(m)}\right), \quad \forall m \in \{1, \dots, L\}, \\ \dot{\widetilde{\varphi}}^{(m)} = \varphi_m'\left(\frac{1}{\sqrt{n_{m-1}}}\Phi^{(m)}x^{(m-1)}, a^{(m)}\right), \quad \forall m \in \{1, \dots, L\}, \\ \dot{\varphi}_*^{(L)} = \varphi_L'\left(S, A^{(L)}\right), \\ \dot{\widetilde{\varphi}}_*^{(L)} = \varphi_L'\left(s, a^{(L)}\right), \end{cases} \tag{5.5.21}$$

where φ_m' is the derivative with respect to its first argument. For $i_l \in \{1, \dots, n_l\}$, we can compute that

$$\frac{\partial \widetilde{\mathbb{E}}\widehat{F}}{\partial X_{i_l}^{(l)}} = -\frac{1}{n}\widetilde{\mathbb{E}}\left\langle\left(\sqrt{\beta}Z + \beta\Gamma\right)\cdot\partial_{X_{i_l}^{(l)}}\Gamma\right\rangle + \frac{1}{n}\widetilde{\mathbb{E}}\left\langle h_2 x^{(L-1)}\cdot\partial_{X_{i_l}^{(l)}}X^{(L-1)}\right\rangle$$

$$= -\frac{\sqrt{t}}{n}\sum_{\mathbf{i}}\widetilde{\mathbb{E}}\left[\left\langle\sqrt{\beta}Z_{i_L} + \beta\left(\varphi_{*,i_L}^{(L)} - \widetilde{\varphi}_{*,i_L}^{(L)}\right)\right\rangle\dot{\varphi}_{i_{l+1}}^{(l+1)}\cdots\dot{\varphi}_{i_{L-1}}^{(L-1)}\dot{\varphi}_{*,i_L}^{(L)}\,\frac{\Phi_{i_{l+1},i_l}^{(l+1)}}{\sqrt{n_l}}\,\cdots\,\frac{\Phi_{i_L,i_{L-1}}^{(L)}}{\sqrt{n_{L-1}}}\right] \tag{5.5.22}$$

$$+\frac{h_2}{n}\sum_{\mathbf{i}'}\widetilde{\mathbb{E}}\left[\left\langle x_{i_{L-1}}^{(L-1)}\right\rangle\dot{\varphi}_{i_{l+1}}^{(l+1)}\cdots\dot{\varphi}_{i_{L-1}}^{(L-1)}\,\frac{\Phi_{i_{l+1},i_l}^{(l+1)}}{\sqrt{n_l}}\,\frac{\Phi_{i_{l+2},i_{l+1}}^{(l+2)}}{\sqrt{n_{l+1}}}\,\cdots\,\frac{\Phi_{i_{L-1},i_{L-2}}^{(L-1)}}{\sqrt{n_{L-2}}}\right] \tag{5.5.23}$$

where $\sum_{\mathbf{i}}$ is over (5.5.8) and $\sum_{\mathbf{i}'}$ is over

$$\mathbf{i}' = (i_{l+1}, i_{l+2}, \dots, i_{L-1}) \in \prod_{m=l+1}^{L-1}\{1, \dots, n_m\}, \tag{5.5.24}$$

respectively. The treatments for (5.5.22) and (5.5.23) are similar to that for (5.5.7), where the main tool is the Gaussian integration by parts summarized in Corollary 5.5.7. Recall that heuristics were given below (5.5.7). Now, applying Corollary 5.5.7 to each summand in (5.5.22), we obtain that, for every $\mathbf{i}$, the summand in (5.5.22) has its absolute value bounded by $Cn^{-(L-l)}$. Since $\sum_{\mathbf{i}}$ is over $O(n^{L-l})$ many terms, we conclude that the part in (5.5.22) is bounded from both sides by Cn^{-1}. Analogous arguments can be applied to (5.5.23) to derive a similar bound. Hence,

$$\left|\frac{\partial\widetilde{\mathbb{E}}\widehat{F}}{\partial X^{(l)}_{i_l}}\right| \leqslant \frac{C}{n}, \quad \forall i_l \in \{1,\ldots,n_l\}.$$

and thus Lemma 5.5.3 yields (5.5.17).

Proof of (5.5.18)

Let us redefine $\widetilde{\mathbb{E}} = \mathbb{E}\left[\,\cdot\,|X^{(l-1)}, \Phi^{[1,l]}\right]$. For $i_l \in \{1,\ldots,n_l\}$, $i_{l-1} \in \{1,\ldots,n_{l-1}\}$, we can compute

$$\begin{aligned}\frac{\partial\widetilde{\mathbb{E}}\widehat{F}}{\partial\Phi^{(l)}_{i_l,i_{l-1}}} &= -\frac{1}{n}\widetilde{\mathbb{E}}\left\langle\left(\sqrt{\beta}Z+\beta\Gamma\right)\cdot\partial_{\Phi^{(l)}_{i_l,i_{l-1}}}\Gamma\right\rangle + \frac{1}{n}\widetilde{\mathbb{E}}\left\langle h_2x^{(L-1)}\cdot\partial_{\Phi^{(l)}_{i_l,i_{l-1}}}X^{(L-1)}\right\rangle \\ &\quad + \frac{1}{n}\widetilde{\mathbb{E}}\left\langle\left(h_2X^{(L-1)}+\sqrt{h_2}Z'-h_2x^{(L-1)}\right)\cdot\partial_{\Phi^{(l)}_{i_l,i_{l-1}}}x^{(L-1)}\right\rangle \\ &= \mathtt{I}_1+\mathtt{I}_2+\mathtt{I}_3.\end{aligned}$$

Here,

$$\begin{aligned}
\mathtt{I}_1 = &-\frac{\sqrt{t}}{n}\sum_{\mathbf{i}}\widetilde{\mathbb{E}}\Bigg[\left\langle\sqrt{\beta}Z_{i_L}+\beta\left(\varphi_{*,i_L}^{(L)}-\widetilde{\varphi}_{*,i_L}^{(L)}\right)\right\rangle\frac{X_{i_{l-1}}^{(l-1)}}{\sqrt{n_{l-1}}}\dot{\varphi}_{i_l}^{(l)}\dot{\varphi}_{i_{l+1}}^{(l+1)}\cdots\dot{\varphi}_{*,i_L}^{(L)}\\
&\qquad\times\frac{\Phi_{i_{l+1},i_l}^{(l+1)}}{\sqrt{n_l}}\cdots\frac{\Phi_{i_L,i_{L-1}}^{(L)}}{\sqrt{n_{L-1}}}\Bigg]\\
&+\frac{\sqrt{t}}{n}\sum_{\mathbf{i}}\widetilde{\mathbb{E}}\Bigg[\left\langle\left(\sqrt{\beta}Z_{i_L}+\beta\left(\varphi_{*,i_L}^{(L)}-\widetilde{\varphi}_{*,i_L}^{(L)}\right)\right)\frac{x_{i_{l-1}}^{(l-1)}}{\sqrt{n_{l-1}}}\dot{\widetilde{\varphi}}_{i_l}^{(l)}\dot{\widetilde{\varphi}}_{i_{l+1}}^{(l+1)}\cdots\dot{\widetilde{\varphi}}_{*,i_L}^{(L)}\right\rangle\\
&\qquad\times\frac{\Phi_{i_{l+1},i_l}^{(l+1)}}{\sqrt{n_l}}\cdots\frac{\Phi_{i_L,i_{L-1}}^{(L)}}{\sqrt{n_{L-1}}}\Bigg]\\
\mathtt{I}_2 = &\frac{h_2}{n}\sum_{\mathbf{i}'}\widetilde{\mathbb{E}}\left[\left\langle x_{i_{L-1}}^{(L-1)}\right\rangle\frac{X_{i_l}^{(l-1)}}{\sqrt{n_{l-1}}}\dot{\varphi}_{i_l}^{(l)}\dot{\varphi}_{i_{l+1}}^{(l+1)}\cdots\dot{\varphi}_{i_{L-1}}^{(L-1)}\frac{\Phi_{i_{l+1},i_l}^{(l+1)}}{\sqrt{n_l}}\frac{\Phi_{i_{l+2},i_{l+1}}^{(l+2)}}{\sqrt{n_{l+1}}}\cdots\frac{\Phi_{i_{L-1},i_{L-2}}^{(L-1)}}{\sqrt{n_{L-2}}}\right]\\
\mathtt{I}_3 = &\frac{1}{n}\sum_{\mathbf{i}'}\widetilde{\mathbb{E}}\Bigg[\left\langle\left(h_2X_{i_{L-1}}^{(L-1)}+\sqrt{h_2}Z'_{i_{L-1}}-h_2x_{i_{L-1}}^{(L-1)}\right)\frac{x_{i_l}^{(l-1)}}{\sqrt{n_{l-1}}}\dot{\widetilde{\varphi}}_{i_l}^{(l)}\dot{\widetilde{\varphi}}_{i_{l+1}}^{(l+1)}\cdots\dot{\widetilde{\varphi}}_{i_{L-1}}^{(L-1)}\right\rangle\\
&\qquad\times\frac{\Phi_{i_{l+1},i_l}^{(l+1)}}{\sqrt{n_l}}\frac{\Phi_{i_{l+2},i_{l+1}}^{(l+2)}}{\sqrt{n_{l+1}}}\cdots\frac{\Phi_{i_{L-1},i_{L-2}}^{(L-1)}}{\sqrt{n_{L-2}}}\Bigg]
\end{aligned}$$

where $\mathbf{i}$ and $\mathbf{i}'$ are given in (5.5.8) and (5.5.24), respectively.

Similar to the the treatments for (5.5.22) and (5.5.23), applying Corollary 5.5.7, we can see that $|\mathtt{I}_1|, |\mathtt{I}_2|, |\mathtt{I}_3| \leqslant Cn^{-\frac{3}{2}}$, which implies that

$$\left|\frac{\partial\widetilde{\mathbb{E}}\widehat{F}}{\partial\Phi_{i_l,i_{l-1}}^{(l)}}\right| \leqslant \frac{C}{n^{\frac{3}{2}}}, \quad \forall i_l \in \{1,\ldots,n_l\},\ i_{l-1}\in\{1,\ldots,n_{l-1}\}.$$

Now, we can conclude that $|\nabla_{\Phi^{(l)}}\widetilde{\mathbb{E}}\widehat{F}| \leqslant Cn^{-\frac{1}{2}}$ and thus (5.5.18) by Lemma 5.5.5.

An ε-net argument

By (5.2.47) and (5.2.48), there is C such that, for all t, h_1 and all $h_2, h_2' \in \mathbb{R}_+$ satisfying $|h_2 - h_2'| \leqslant 1$,

$$\left|\left(F-\overline{F}\right)(t,h_1,h_2)-\left(F-\overline{F}\right)(t,h_1,h_2')\right| \leqslant C\left(1+n^{-\frac{1}{2}}\left|Z'\right|\right)\left|h_2-h_2'\right|^{\frac{1}{2}}.$$

Setting $E_\varepsilon = [0, M] \cap \{\varepsilon, 2\varepsilon, 3\varepsilon, \dots\}$ for $\varepsilon \in (0,1)$, we have that, for all t, h_1,

$$\begin{aligned}
&\mathbb{E}\left[\left\|F - \overline{F}\right\|^2_{L^\infty_{h_2}([0,M])}(t, h_1)\right] \\
&\leqslant \mathbb{E}\left[\sup_{h_2 \in E_\varepsilon} \left(F - \overline{F}\right)^2 (t, h_1)\right] + \mathbb{E}\left[C\left(1 + n^{-\frac{1}{2}}\left|Z'\right|\right)^2 \varepsilon\right] \\
&\leqslant \sum_{h_2 \in E_\varepsilon} \mathbb{E}\left[\left(F - \overline{F}\right)^2 (t, h_1)\right] + C\varepsilon \leqslant C\left(\varepsilon^{-1} n^{-1} + \varepsilon\right),
\end{aligned}$$

where the last inequality follows from (5.5.20). Optimizing this by taking $\varepsilon = n^{-\frac{1}{2}}$ completes the proof of Lemma 5.5.4.

5.5.4. Multiple Gaussian integration by parts

Denote by $\langle \cdot \rangle$ the Gibbs measure with Hamiltonian $\widehat{H}$ given in (5.5.11). Recall the variables $x, w, a, a^{(L)}$ in $\widehat{H}$, and also the definition of s in (5.2.3). For $\gamma \in \mathbb{N} \cup \{0\}$, we enumerate the replicas, i.e., i.i.d. copies of $x, w, a, a^{(L)}, s$ under $\langle \cdot \rangle$, as

$$x^{|\gamma)},\ w^{|\gamma)},\ a^{|\gamma)},\ a^{(L|\gamma)},\ s^{|\gamma)}.$$

Recall the definition of $x^{(L-1)}$ in (5.1.5), and we want to extend this. Using (5.1.3) iteratively, for every $l \in \{0, \dots, L\}$, we can find a deterministic function ζ_l satisfying

$$X^{(l)} = \zeta_l\left(X^{(0)}, A^{[1,l]}, \Phi^{[1,l]}\right)$$

where we understand that $A^{[1,0]} = 0$ and $\Phi^{[1,0]} = 0$. Replacing $X^{(0)}, A^{[1,l]}$ above by x and projections of $(a, a^{(L)})$, we can define $x^{(l)}$ in a way analogous to (5.1.5):

$$x^{(l)} = \zeta_l\left(x, \pi_{[1,l]}\left(a, a^{(L)}\right), \Phi^{[1,l]}\right)$$

where $\pi_{[1,l]}$ is the projection of the first $\sum_{m=1}^{l} n_m k_m$ coordinates into $\prod_{m=1}^{l} \mathbb{R}^{n_m \times k_m}$ (recall $a = (a^{(1)}, \dots, a^{(L-1)})$ given in (5.1.5) and $a^{(L)}$ in (5.5.10)). For $\gamma \in \mathbb{N} \cup \{0\}$, we denote by $x^{(l|\gamma)}$ the γ-th replica of $x^{(l)}$. We also set $\widehat{H}^{|\gamma)}$ to be $\widehat{H}$ with variables replaced by their γ-th replicas.

Recall S in (5.2.2). For $\nu \in \mathbb{N} \cup \{0\}$, $\gamma \in \mathbb{N} \cup \{0\}$, $j \in \{1, \ldots, n_m\}$, we introduce

$$\begin{aligned}
\varphi_j^{(m|\nu)} &= \frac{\partial^\nu}{\partial r^\nu} \varphi_m \left(r, A_j^{(m)}\right) \Big|_{r=\frac{1}{\sqrt{n_{m-1}}}(\Phi^{(m)} X^{(m-1)})_j}, \quad \forall m \in \{1, \ldots, L\}, \\
\varphi_{*,j}^{(L|\nu)} &= \frac{\partial^\nu}{\partial r^\nu} \varphi_L \left(r, A_j^{(L)}\right) \Big|_{r=\frac{1}{\sqrt{n_{L-1}}}(\Phi^{(L)} S)_j}, \\
\widetilde{\varphi}_j^{(m|\nu|\gamma)} &= \frac{\partial^\nu}{\partial r^\nu} \varphi_m \left(r, a_j^{(m)}\right) \Big|_{r=\frac{1}{\sqrt{n_{m-1}}}(\Phi^{(m)} x^{(m-1|\gamma)})_j}, \quad \forall m \in \{1, \ldots, L\}, \\
\widetilde{\varphi}_{*,j}^{(L|\nu|\gamma)} &= \frac{\partial^\nu}{\partial r^\nu} \varphi_L \left(r, a_j^{(L)}\right) \Big|_{r=\frac{1}{\sqrt{n_{L-1}}}(\Phi^{(L)} s^{|\gamma)})_j}.
\end{aligned}$$

In particular, $\varphi_j^{(m|0)} = X_j^{(m)}$ and $\widetilde{\varphi}_j^{(m|0|\gamma)} = x_j^{(m|\gamma)}$ and note that these two identities can be extended to $m = 0$. Recall that Z and Z' are standard Gaussian vectors given in (5.1.4) and (5.2.5), respectively. We introduce the following collections of random variables

$$\begin{aligned}
\mathcal{Z} &= \{Z_j\}_{1 \leqslant j \leqslant n_L} \cup \{Z'_j\}_{1 \leqslant j \leqslant n_{L-1}}, \\
\mathcal{M}_j^{(l|\nu|\gamma)} &= \bigcup_{\widetilde{\nu} \in \{0,\ldots,\nu\}} \bigcup_{\widetilde{\gamma} \in \{0,\ldots,\gamma\}} \left\{\varphi_j^{(l|\widetilde{\nu})},\ \widetilde{\varphi}_j^{(l|\widetilde{\nu}|\widetilde{\gamma})}\right\}, \quad \forall l \leqslant L-1, \\
\mathcal{M}_j^{(L|\nu|\gamma)} &= \bigcup_{\widetilde{\nu} \in \{0,\ldots,\nu\}} \bigcup_{\widetilde{\gamma} \in \{0,\ldots,\gamma\}} \left\{\varphi_j^{(L|\widetilde{\nu})},\ \varphi_{*,j}^{(L|\widetilde{\nu})},\ \widetilde{\varphi}_j^{(L|\widetilde{\nu}|\widetilde{\gamma})},\ \widetilde{\varphi}_{*,j}^{(L|\widetilde{\nu}|\widetilde{\gamma})}\right\}, \\
\overline{\mathcal{M}} &= \bigcup_{l \in \{1,\ldots,L\}} \bigcup_{\nu \in \mathbb{N} \cup \{0\}} \bigcup_{\gamma \in \mathbb{N} \cup \{0\}} \bigcup_{i_l \in \{1,\ldots,n_l\}} \mathcal{M}_{i_l}^{(l|\nu|\gamma)}.
\end{aligned}$$

For $\nu_1, \nu_2, \ldots, \nu_L, \gamma \in \mathbb{N} \cup \{0\}$, we set

$$\mathcal{N}^{(\nu_1,\ldots,\nu_L|\gamma)} = \mathcal{Z} \cup \left(\bigcup_{m=1}^{L} \bigcup_{j_m=1}^{n_m} \mathcal{M}_{j_m}^{(m|\nu_m|\gamma)} \right).$$

For $d, r \in \mathbb{N}$, let $\mathbf{P}_{d,r}$ be the collection of polynomials of degree up to d over $\mathbb{R}^r$ with real coefficients. For every $P \in \mathbf{P}_{d,r}$ expressed as

$$P(x) = \sum a_{p_1,p_2,\ldots,p_r} x_1^{p_1} x_2^{p_2} \cdots x_r^{p_r}$$

where the summation is over

$$\left\{(p_1, p_2, \dots, p_r) \in (\mathbb{N} \cup \{0\})^r : \sum_{i=1}^{r} p_i \leqslant d\right\},$$

we define

$$\|P\| = \sum |a_{p_1,p_2,\dots,p_r}|.$$

Slightly abusing the notation, we view any finite subcollection $\mathcal{E} \subseteq \overline{\mathcal{M}}$ as an ordered tuple of random variables. In this notation, for any $P \in \mathbf{P}_{d,|\mathcal{E}|}$ for some d, we view $P(\mathcal{E})$ as P evaluated at $\mathcal{E}$. Lastly, for $a, b \in \mathbb{R}$, we write $a \vee b = \max\{a, b\}$.

Lemma 5.5.6. *Let $l \in \{1, \cdots, L\}$, $\nu_1, \dots, \nu_L \in \mathbb{N} \cup \{0\}$, $\gamma \in \mathbb{N} \cup \{0\}$, $\beta \geqslant 0$, $M \geqslant 1$. In addition to (h1), assume that $\Phi^{(m)}$ consists of i.i.d. standard Gaussian entries and that φ_m is bounded and continuously differentiable with bounded derivatives up to ν'_m-th order for every $m \in \{1, \dots, L\}$, where*

$$\nu'_m = \nu_m + (2^{m-l+1} - 1) \vee 0, \quad \forall m \in \{1, \dots, L\}. \tag{5.5.25}$$

Then, there are constants C, γ', d' such that the following holds.

For every $k \in \{l, \dots, L\}$, every $n \in \mathbb{N}$, every $(t, h) \in \Omega_{\rho_{L-1,n}} \cap \{|h_2| \leqslant M\}$, every $i_m \in \{1, \dots, n_m\}$ with $m \in \{l-1, \dots, k\}$, every

$$\mathcal{E} = \mathcal{N}^{(\nu_1,\dots,\nu_L|\gamma)}$$

and every $P \in \mathbf{P}_{d,|\mathcal{E}|}$, there is $P' \in \mathbf{P}_{d',|\mathcal{E}'|}$ for some

$$\mathcal{E}' = \mathcal{N}^{(\nu'_1,\dots,\nu'_L|\gamma')}$$

such that

$$\mathbb{E}^{[l,k]}\left[\langle P(\mathcal{E})\rangle \prod_{m=l}^{k} \Phi^{(m)}_{i_m,i_{m-1}}\right] = n^{-\frac{1}{2}(k-l+1)}\mathbb{E}^{[l,k]}\left\langle P'(\mathcal{E}')\right\rangle \tag{5.5.26}$$

and

$$\|P'\| \leqslant C\|P\|, \tag{5.5.27}$$

where $\mathbb{E}^{[l,k]}$ *is the expectation with respect to* $\Phi^{[l,k]}$.

Recall the notation introduced in (5.5.21). We state an immediate corollary of Lemma 5.5.6.

Corollary 5.5.7. *Assume (H1)–(H3) for some* $L \in \mathbb{N}$. *For* $l \in \{1, \ldots, L\}$, $d \in \mathbb{N}$, $\beta \geqslant 0$, $M \geqslant 1$, *there is* C *such that the following holds. Suppose that* $P \in \mathbf{P}_{d,\,2L-2l+10}$ *is a monomial with coefficient* 1 *and independent of* n. *Then, for every* $k \in \{l, \ldots, L\}$, *every* $n \in \mathbb{N}$, *every* $(t,h) \in \Omega_{\rho_{L-1,n}} \cap \{|h_2| \leqslant M\}$, *every* $i_m \in \{1, \ldots, n_m\}$ *with* $m \in \{l-1, \ldots, k\}$, *it holds that*

$$\widetilde{\mathbb{E}}\left[\langle P(\mathcal{E})\rangle \prod_{m=l}^{k} \Phi^{(m)}_{i_m,i_{m-1}}\right] \leqslant Cn^{-\frac{1}{2}(k-l+1)}$$

where

$$\mathcal{E} = \Big(Z_{i_L},\ Z'_{i_{L-1}},\ \varphi^{(L)}_{*,i_L},\ \widetilde{\varphi}^{(L)}_{*,i_L},\ \dot{\varphi}^{(L)}_{*,i_L},\ \dot{\widetilde{\varphi}}^{(L)}_{*,i_L},\ X^{(L-1)}_{i_{L-1}},\ x^{(L-1)}_{i_{L-1}},$$
$$\left(\dot{\varphi}^{(m)}_{i_m}\right)_{m=l}^{L-1},\ \left(\dot{\widetilde{\varphi}}^{(m)}_{i_m}\right)_{m=l}^{L-1},\ X^{(l-1)}_{i_{l-1}},\ x^{(l-1)}_{i_{l-1}}\Big)$$

and $\widetilde{\mathbb{E}}$ *integrates over* Z_{i_L}, $Z'_{i_{L-1}}$ *and* $(\Phi^{(m)}_{i_m,i_{m-1}})^k_{m=l}$.

Proof of Corollary 5.5.7. Comparing (5.5.21) with the notation here, we can rewrite

$$\mathcal{E} = \Big(Z_{i_L},\ Z'_{i_{L-1}},\ \varphi^{(L|0)}_{*,i_L},\ \widetilde{\varphi}^{(L|0|0)}_{*,i_L},\ \varphi^{(L|1)}_{*,i_L},\ \widetilde{\varphi}^{(L|1|0)}_{*,i_L},\ \varphi^{(L-1|0)}_{i_{L-1}},\ \widetilde{\varphi}^{(L-1|0|0)}_{i_{L-1}},$$
$$\left(\varphi^{(m|1)}_{i_m}\right)_{m=l}^{L-1},\ \left(\widetilde{\varphi}^{(m|1|0)}_{i_m}\right)_{m=l}^{L-1},\ \varphi^{(l-1|0)}_{i_{l-1}},\ \widetilde{\varphi}^{(l-1|0|0)}_{i_{l-1}}\Big).$$

Hence, we have that $\mathcal{E} \subseteq \mathcal{N}^{(1,1,\ldots,1|0)}$. This corollary follows from Lemma 5.5.6 by setting $\gamma = 0$ and $\nu_m = 1$ for all m and noticing that the differentiability condition (5.5.25) is fulfilled by assumption (H2). $\square$

Proof of Lemma 5.5.6. We use induction on l and start with the base case $l = L$. The Gaussian integration by parts yields that

$$\mathbb{E}^{(L)} \left\langle P(\mathcal{E})\Phi^{(L)}_{i_L,i_{L-1}} \right\rangle = \mathbb{E}^{(L)} \left[\partial_{\Phi^{(L)}_{i_L,i_{L-1}}} \langle P(\mathcal{E}) \rangle \right] = \sum_{\phi \in \overline{\mathcal{M}}} \mathbb{E}^{(L)} \left\langle \mathscr{P}_\phi \partial_{\Phi^{(L)}_{i_L,i_{L-1}}} \phi \right\rangle \tag{5.5.28}$$

where $\mathbb{E}^{(L)} = \mathbb{E}^{[L,L]}$, and, by the chain rule, viewing $\phi \in \overline{\mathcal{M}}$ as labels for the arguments in P and $\widehat{H}^{(\widetilde{\gamma})}$, we have that

$$\mathscr{P}_\phi = \partial_\phi P(\mathcal{E}) + P(\mathcal{E}) \left(\sum_{\widetilde{\gamma}=0}^{\gamma} \partial_\phi \widehat{H}^{|\widetilde{\gamma})} - \gamma \partial_\phi \widehat{H}^{|\gamma+1)} \right), \tag{5.5.29}$$

with

$$\begin{aligned} \partial_\phi \widehat{H}^{|\widetilde{\gamma})} = &- \sum_{j=1}^{n_L} \left(\mathbf{1}_{\phi = \varphi^{(L|0)}_{*,j}} - \mathbf{1}_{\widetilde{\phi} = \varphi^{(L|0|\widetilde{\gamma})}_{*,j}} \right) \left(\sqrt{\beta} Z_j + \beta \left(\varphi^{(L|0)}_{*,j} - \widetilde{\varphi}^{(L|0|\widetilde{\gamma})}_{*,j} \right) \right) \\ &+ \sum_{j=1}^{n_{L-1}} \mathbf{1}_{\phi = \varphi^{(L-1|0)}_j} h_2 \widetilde{\varphi}^{(L-1|0|\widetilde{\gamma})}_j \\ &+ \sum_{j=1}^{n_{L-1}} \mathbf{1}_{\phi = \varphi^{(L-1|0|\widetilde{\gamma})}_j} \left(h_2 \varphi^{(L-1|0)}_j + \sqrt{h_2} Z'_j - h_2 \widetilde{\varphi}^{(L-1|0|\widetilde{\gamma})}_j \right). \end{aligned} \tag{5.5.30}$$

Let us clarify (5.5.30). Due to the definition of $\widehat{H}$ in (5.5.11), fixing Z, Z', we can view $\widehat{H}^{|\widetilde{\gamma})}$ as a function of $\varphi_L\left(S, A^{(L)}\right)$, $\varphi_L\left(s^{|\widetilde{\gamma})}, a^{(L|\widetilde{\gamma})}\right)$, $X^{(L-1)}$, $x^{(L-1|\widetilde{\gamma})}$, or equivalently, $\varphi^{(L|0)}_*$, $\widetilde{\varphi}^{(L|0|\widetilde{\gamma})}_*$, $\varphi^{(L-1|0)}$, $\widetilde{\varphi}^{(L-1|0|\widetilde{\gamma})}$. Therefore, when viewing these as labels for the variables inside $\widehat{H}^{|\widetilde{\gamma})}$, we have (5.5.30) and the left-hand side of it is nonzero only if ϕ is an entry of those vectors.

Next, let us show that

$$\mathscr{P}_\phi \partial_{\Phi^{(L)}_{i_L,i_{L-1}}} \phi \neq 0 \quad \text{only if} \quad \phi \in \mathcal{M}^{(L|\nu_L|\gamma+1)}_{i_L}. \tag{5.5.31}$$

From (5.5.29) and (5.5.30), we can see that

$$\mathscr{P}_\phi \neq 0 \quad \text{only if} \quad \phi \in \mathcal{E} \cup \left(\bigcup_{j_{L-1}=1}^{n_{L-1}} \mathcal{M}^{(L-1|0|\gamma+1)}_{j_{L-1}} \right) \cup \left(\bigcup_{j_L=1}^{n_L} \mathcal{M}^{(L|0|\gamma+1)}_{j_L} \right). \tag{5.5.32}$$

On the other hand, due to (5.1.3), note that

$$\partial_{\Phi^{(L)}_{i_L,i_{L-1}}} \phi \neq 0 \quad \text{only if} \quad \phi \in \bigcup_{\widetilde{\nu}\in\mathbb{N}} \bigcup_{\widetilde{\gamma}\in\mathbb{N}} \mathcal{M}^{(L|\widetilde{\nu}|\widetilde{\gamma})}_{i_L}. \tag{5.5.33}$$

The intersection of sets in (5.5.32) and (5.5.33) is a subset of the set in (5.5.31). Hence, (5.5.31) is valid.

Due to (5.5.33), $\partial_{\Phi^{(L)}_{i_L,i_{L-1}}} \phi$ in (5.5.31), whenever nonzero, is of one of the four forms below, for some $\widetilde{\nu} \leqslant \nu_L$ and $\widetilde{\gamma} \leqslant \gamma + 1$,

$$\begin{cases} \partial_{\Phi^{(L)}_{i_L,i_{L-1}}} \varphi^{(L|\widetilde{\nu})}_{i_L} = \varphi^{(L|\widetilde{\nu}+1)}_{i_L} \frac{1}{\sqrt{n_{L-1}}} X^{(L-1)}_{i_{L-1}} \\ \partial_{\Phi^{(L)}_{i_L,i_{L-1}}} \widetilde{\varphi}^{(L|\widetilde{\nu}|\widetilde{\gamma})}_{i_L} = \varphi^{(L|\widetilde{\nu}+1|\widetilde{\gamma})}_{i_L} \frac{1}{\sqrt{n_{L-1}}} x^{(L-1|\widetilde{\gamma})}_{i_{L-1}} \\ \partial_{\Phi^{(L)}_{i_L,i_{L-1}}} \varphi^{(L|\widetilde{\nu})}_{*,i_L} = \varphi^{(L|\widetilde{\nu}+1)}_{*,i_L} \frac{1}{\sqrt{n_{L-1}}} S_{i_{L-1}} \\ \partial_{\Phi^{(L)}_{i_L,i_{L-1}}} \widetilde{\varphi}^{(L|\widetilde{\nu}|\widetilde{\gamma})}_{*,i_L} = \widetilde{\varphi}^{(L|\widetilde{\nu}+1|\widetilde{\gamma})}_{*,i_L} \frac{1}{\sqrt{n_{L-1}}} s^{|\widetilde{\gamma})}_{i_{L-1}} \end{cases} \tag{5.5.34}$$

Using this, (5.5.29) and (5.5.30), we can see that for

$$\mathcal{E}' = \mathcal{N}^{(\nu_1,\ldots,\nu_{L-1},\nu_L+1|\gamma+1)}$$

there is a polynomial $P'_\phi \in \mathbf{P}_{d',|\mathcal{E}'|}$ for some d' such that

$$P'_\phi(\mathcal{E}') = n^{\frac{1}{2}} \mathscr{P}_\phi \partial_{\Phi^{(L)}_{i_L,i_{L-1}}} \phi. \tag{5.5.35}$$

Here, the scalar $n^{\frac{1}{2}}$ is to make $n^{\frac{1}{2}}\partial_{\Phi^{(L)}_{i_L,i_{L-1}}} \phi$ to be of order 1. By (5.5.28) and (5.5.31), setting

$$P'(\mathcal{E}') = \sum_{\phi \in \mathcal{M}^{(L|\nu_L|\gamma+1)}_{i_L}} P'_\phi(\mathcal{E}')$$

we have

$$\mathbb{E}^{(L)} \left\langle P(\mathcal{E}) \Phi^{(L)}_{i_L,i_{L-1}} \right\rangle = n^{-\frac{1}{2}} \mathbb{E}^{(L)} \left\langle P'(\mathcal{E}') \right\rangle.$$

Using (5.5.29), (5.5.30), (5.5.34) and (5.5.35), we can see that

$$\|P'\| \leqslant C \|P\|$$

for some constant C that depends only on $L, \nu_1, \ldots, \nu_L, \gamma, \beta, M$.

Now, we consider the induction step and assume that the lemma holds for $l+1 \leqslant L$. In the following, we denote by C a constant that depends only on $l, \nu_1, \ldots, \nu_L, \gamma, \beta, M$ and may vary from line to line. Setting $\mathbb{E}^{(l)} = \mathbb{E}^{[l,l]}$ and using the induction assumption for $l+1$, we get that for

$$\mathcal{F} = \mathcal{N}^{(\nu'_1,\ldots,\nu'_L|\gamma')}$$

with some $\gamma' > 0$ and

$$\nu'_m = \nu_m + (2^{m-l} - 1) \vee 0, \quad \forall m \in \{1, \ldots, L\}, \tag{5.5.36}$$

there is $Q \in \mathbf{P}_{d',|\mathcal{F}|}$ for some d' such that

$$\mathbb{E}^{[l,k]}\left[\langle P(\mathcal{E})\rangle \prod_{m=l}^{k} \Phi^{(m)}_{i_m,i_{m-1}}\right] = \mathbb{E}^{(l)}\left[\mathbb{E}^{[l+1,k]}\left[\langle P(\mathcal{E})\rangle \prod_{m=l+1}^{k} \Phi^{(m)}_{i_m,i_{m-1}}\right] \Phi^{(l)}_{i_l,i_{l-1}}\right]$$
$$= n^{-\frac{1}{2}(k-l)}\mathbb{E}^{[l,k]}\left[\langle Q(\mathcal{F})\rangle \Phi^{(l)}_{i_l,i_{l-1}}\right] \tag{5.5.37}$$

and

$$\|Q\| \leqslant C\|P\|. \tag{5.5.38}$$

Applying the Gaussian integration by parts to the last expectation in (5.5.37) yields

$$\mathbb{E}^{[l,k]}\left[\langle Q(\mathcal{F})\rangle \Phi^{(l)}_{i_l,i_{l-1}}\right] = \mathbb{E}^{[l,k]}\left[\left\langle \partial_{\Phi^{(l)}_{i_l,i_{l-1}}} Q(\mathcal{F})\right\rangle\right] = \sum_{\phi\in\overline{\mathcal{M}}} \mathbb{E}^{[l,k]}\left\langle \mathscr{Q}_\phi \partial_{\Phi^{(l)}_{i_l,i_{l-1}}} \phi\right\rangle \tag{5.5.39}$$

where

$$\mathscr{Q}_\phi = \partial_\phi Q(\mathcal{F}) + Q(\mathcal{F})\left(\sum_{\widetilde{\gamma}=0}^{\gamma'} \partial_\phi \widehat{H}^{|\widetilde{\gamma})} - \gamma' \partial_\phi \widehat{H}^{|\gamma'+1)}\right). \tag{5.5.40}$$

Next, we show that

$$\mathscr{Q}_\phi \partial_{\Phi^{(l)}_{i_l,i_{l-1}}} \phi \neq 0 \quad \text{only if} \quad \phi \in \mathcal{M}^{(l|\nu'_l|\gamma'+1)}_{i_l} \cup \left(\bigcup_{m=l+1}^{L} \bigcup_{j_m=1}^{n_m} \mathcal{M}^{(m|\nu'_m|\gamma'+1)}_{j_m}\right). \tag{5.5.41}$$

Similar to the derivation of (5.5.32), using (5.5.40) and (5.5.30), we can see that

$$\mathscr{Q}_\phi \neq 0 \quad \text{only if} \quad \phi \in \mathcal{F} \cup \left(\bigcup_{j_{L-1}=1}^{n_{L-1}} \mathcal{M}^{(L-1|0|\gamma'+1)}_{j_{L-1}}\right) \cup \left(\bigcup_{j_L=1}^{n_L} \mathcal{M}^{(L|0|\gamma'+1)}_{j_L}\right)$$

Due to (5.1.3), note that

$$\partial_{\Phi^{(l)}_{i_l,i_{l-1}}}\phi \neq 0 \quad \text{only if} \quad \phi \in \bigcup_{\widetilde{\nu}\in\mathbb{N}}\bigcup_{\widetilde{\gamma}\in\mathbb{N}}\left(\mathcal{M}_{i_l}^{(l|\widetilde{\nu}|\widetilde{\gamma})} \cup \left(\bigcup_{m=l+1}^{L}\bigcup_{j_m=1}^{n_m}\mathcal{M}_{j_m}^{(m|\widetilde{\nu}|\widetilde{\gamma})}\right)\right).$$

The intersection of the sets in the above two displays is contained in the set in (5.5.41) and thus (5.5.41) is valid.

Then, we compute the summands in (5.5.39). Due to (5.5.41), we distinguish two cases:

$$\phi \in \mathcal{M}_{i_l}^{(l|\nu'_l|\gamma'+1)} \quad \text{or} \quad \phi \in \bigcup_{m=l+1}^{L}\bigcup_{j_m=1}^{n_m}\mathcal{M}_{j_m}^{(m|\nu'_m|\gamma'+1)}. \tag{5.5.42}$$

Let us consider the first case in (5.5.42). Since $l+1 \leqslant L$, $\partial_{\Phi^{(l)}_{i_l,i_{l-1}}}\phi$ has one of the two forms below, for $\widetilde{\nu} \leqslant \nu'_L$ and $\widetilde{\gamma} \leqslant \gamma'+1$,

$$\partial_{\Phi^{(l)}_{i_l,i_{l-1}}}\varphi_{i_l}^{(l|\widetilde{\nu})} = \varphi_{i_l}^{(l|\widetilde{\nu}+1)}\frac{1}{\sqrt{n_{l-1}}}X_{i_{l-1}}^{(l-1)},$$
$$\partial_{\Phi^{(l)}_{i_l,i_{l-1}}}\widetilde{\varphi}_{i_l}^{(l|\widetilde{\nu}|\widetilde{\gamma})} = \varphi_{i_l}^{(l|\widetilde{\nu}+1|\widetilde{\gamma})}\frac{1}{\sqrt{n_{l-1}}}x_{i_{l-1}}^{(l-1|\widetilde{\gamma})}$$

From this, (5.5.40) and (5.5.30), we can see that, for every ϕ belonging to the first set in (5.5.42), there is a polynomial $Q'_\phi \in \mathbf{P}_{d_\phi,|\mathcal{F}'|}$ for some d_ϕnd

$$\mathcal{F}' = \mathcal{N}^{(\overline{\nu}_1,\ldots,\overline{\nu}_L|\gamma'+1)} \tag{5.5.43}$$

with

$$\overline{\nu}_m = \begin{cases} \nu'_m + 1 & m \geqslant l \\ \nu'_m & m \leqslant l-1 \end{cases} \tag{5.5.44}$$

such that

$$Q'_\phi(\mathcal{F}') = n^{\frac{1}{2}} \mathscr{Q}_\phi \partial_{\Phi^{(l)}_{i_l,i_{l-1}}} \phi, \quad \forall \phi \in \mathcal{M}^{(l|\nu'_l|\gamma'+1)}_{i_l},$$
$$\|Q'_\phi\| \leqslant C\|Q\|, \quad \forall \phi \in \mathcal{M}^{(l|\nu'_l|\gamma'+1)}_{i_l}. \tag{5.5.45}$$

Therefore

$$\mathbb{E}^{[l,k]} \left\langle \mathscr{Q}_\phi \partial_{\Phi^{(l)}_{i_l,i_{l-1}}} \phi \right\rangle = n^{-\frac{1}{2}} \mathbb{E} \left\langle Q'_\phi(\mathcal{F}') \right\rangle, \quad \forall \phi \in \mathcal{M}^{(l|\nu'_l|\gamma'+1)}_{i_l}. \tag{5.5.46}$$

Now, we turn to the second case in (5.5.42). Let us assume that

$$\phi \in \mathcal{M}^{(m|\nu'_m|\gamma'+1)}_{j_m}, \quad m \in \{l+1, \dots, L\},\ j_m \in \{1, \dots, n_m\}. \tag{5.5.47}$$

Then, due to (5.1.3) and the chain rule, $\partial_{\Phi^{(l)}_{i_l,i_{l-1}}} \phi$ is one of the following, for $\widetilde{\nu} \leqslant \nu'_m, \widetilde{\gamma} \leqslant \gamma' + 1$:

$$\begin{aligned}
&\partial_{\Phi^{(l)}_{i_l,i_{l-1}}} \varphi^{(m|\widetilde{\nu})}_{j_m} \\
&\qquad = \varphi^{(m|\widetilde{\nu}+1)}_{j_m} \sum_{\mathbf{j}} \left(\prod_{\widetilde{m}=l+1}^{m} \frac{1}{\sqrt{n_{\widetilde{m}-1}}} \Phi^{(\widetilde{m})}_{j_{\widetilde{m}},j_{\widetilde{m}-1}} \varphi^{(\widetilde{m}-1|1)}_{j_{\widetilde{m}-1}} \right) \Bigg|_{j_l=i_l} \frac{1}{\sqrt{n_{l-1}}} X^{(l-1)}_{i_{l-1}}, \\
&\partial_{\Phi^{(l)}_{i_l,i_{l-1}}} \widetilde{\varphi}^{(m|\widetilde{\nu}|\widetilde{\gamma})}_{j_m} \\
&\qquad = \widetilde{\varphi}^{(m|\widetilde{\nu}+1|\widetilde{\gamma})}_{j_m} \sum_{\mathbf{j}} \left(\prod_{\widetilde{m}=l+1}^{m} \frac{1}{\sqrt{n_{\widetilde{m}-1}}} \Phi^{(\widetilde{m})}_{j_{\widetilde{m}},j_{\widetilde{m}-1}} \widetilde{\varphi}^{(\widetilde{m}-1|1|\widetilde{\gamma})}_{j_{\widetilde{m}-1}} \right) \Bigg|_{j_l=i_l} \frac{1}{\sqrt{n_{l-1}}} x^{(l-1|\widetilde{\gamma})}_{i_{l-1}}.
\end{aligned}$$

where the summation is over

$$\mathbf{j} = (j_{l+1}, j_{l+2}, \dots, j_{m-2}, j_{m-1}) \in \prod_{\widetilde{m}=l+1}^{m-1} \{1, \dots, n_{\widetilde{m}}\}. \tag{5.5.48}$$

When $m = L$, there are two more possibilities $\partial_{\Phi^{(l)}_{i_l,i_{l-1}}} \varphi^{(L|\widetilde{\nu})}_{*,j_L}$ and $\partial_{\Phi^{(l)}_{i_l,i_{l-1}}} \widetilde{\varphi}^{(L|\widetilde{\nu}|\widetilde{\gamma})}_{*,j_L}$, which are similar to the above and omitted for brevity. These computations allow us to write that

$$\mathbb{E}^{[l,k]}\left\langle \mathscr{Q}_\phi \partial_{\Phi^{(l)}_{i_l,i_{l-1}}} \phi \right\rangle = n^{-\frac{1}{2}(m-l+1)} \sum_{\mathbf{j}} \mathbb{E}^{[l,k]}\left[\langle \mathscr{Q}_\phi g_{\phi,\mathbf{j}} \rangle \prod_{\widetilde{m}=l+1}^{m} \Phi^{(\widetilde{m})}_{j_{\widetilde{m}},j_{\widetilde{m}-1}} \right] \Bigg|_{j_l=i_l} \tag{5.5.49}$$

where

$$g_{\phi,\mathbf{j}} = \begin{cases} \varphi_j^{(m|\widetilde{\nu}+1)} \left(\prod_{\widetilde{m}=l+1}^{m} \sqrt{\frac{n}{n_{\widetilde{m}-1}}} \varphi^{(\widetilde{m}-1|1)}_{j_{\widetilde{m}-1}} \right) \sqrt{\frac{n}{n_{l-1}}} X^{(l-1)}_{i_{l-1}}, & \phi = \varphi^{(m|\widetilde{\nu})}_{j_m}, \\ \widetilde{\varphi}_j^{(m|\widetilde{\nu}+1|\widetilde{\gamma})} \left(\prod_{\widetilde{m}=l+1}^{m} \sqrt{\frac{n}{n_{\widetilde{m}-1}}} \widetilde{\varphi}^{(\widetilde{m}-1|1|\widetilde{\gamma})}_{j_{\widetilde{m}-1}} \right) \sqrt{\frac{n}{n_{l-1}}} x^{(l-1|\widetilde{\gamma})}_{i_{l-1}}, & \phi = \widetilde{\varphi}^{(m|\widetilde{\nu}|\widetilde{\gamma})}_{j_m}. \end{cases} \tag{5.5.50}$$

By these, (5.5.40) and (5.5.30), there is a polynomial $Q_{\phi,\mathbf{j}} \in \mathbf{P}_{d',|\mathcal{F}'|}$ for some larger d' independent of ϕ, $\mathbf{j}$ and for $\mathcal{F}'$ in (5.5.43) such that

$$Q_{\phi,\mathbf{j}}(\mathcal{F}') = \mathscr{Q}_\phi g_{\phi,\mathbf{j}} \tag{5.5.51}$$

which, due to (5.5.50), also satisfies that

$$\|Q_{\phi,\mathbf{j}}\| \leqslant C\|Q\|. \tag{5.5.52}$$

Recall that we are considering the case (5.5.47). Insert (5.5.51) into the right-hand side of (5.5.49) and applying the induction assumption for $l+1$ to every summand there yields that

$$\begin{aligned} \mathbb{E}^{[l,k]}\left\langle \mathscr{Q}_\phi \partial_{\Phi^{(l)}_{i_l,i_{l-1}}} \phi \right\rangle &= n^{-\frac{1}{2}(m-l+1)} \sum_{\mathbf{j}} \mathbb{E}^{[l,k]}\left[\langle Q_{\phi,\mathbf{j}}(\mathcal{F}') \rangle \prod_{\widetilde{m}=l+1}^{m} \Phi^{(\widetilde{m})}_{j_{\widetilde{m}},j_{\widetilde{m}-1}} \right] \Bigg|_{j_l=i_l} \\ &= n^{-(m-l+\frac{1}{2})} \sum_{\mathbf{j}} \mathbb{E}^{[l,k]} \left\langle Q'_{\phi,\mathbf{j}}(\mathcal{E}') \right\rangle, \qquad \forall \phi \in \mathcal{M}^{(m|\nu'_m|\gamma'+1)}_{j_m} \end{aligned} \tag{5.5.53}$$

for some polynomials $Q'_{\phi,\mathbf{j}} \in \mathbf{P}_{d',|\mathcal{E}'|}$ for some larger d', and

$$\mathcal{E}' = \mathcal{N}^{(\nu''_1,\dots,\nu''_L|\gamma'')} \tag{5.5.54}$$

with some larger γ'' and

$$\nu''_m = \overline{\nu}_m + (2^{m-l} - 1) \vee 0, \quad \forall m \in \{1, \ldots, L\}, \tag{5.5.55}$$

where $\overline{\nu}_m$ is given in (5.5.44). In addition, each of these polynomials satisfies that

$$\|Q'_{\phi,\mathbf{j}}\| \leqslant C \|Q_{\phi,\mathbf{j}}\|. \tag{5.5.56}$$

Since $\sum_{\mathbf{j}}$ is a summation of $O(n^{m-l-1})$ many terms due to (5.5.48), setting

$$P'_\phi(\mathcal{E}') = n^{-(m-l-1)} \sum_{\mathbf{j}} Q'_{\phi,\mathbf{j}}(\mathcal{E}'), \tag{5.5.57}$$

and using (5.5.52) and (5.5.56), we obtain that

$$\|P'_\phi\| \leqslant C\|Q\|, \quad \forall \phi \in \mathcal{M}_{j_m}^{(m|\nu'_m|\gamma'+1)}. \tag{5.5.58}$$

Inserting (5.5.57) into (5.5.53) gives that

$$\mathbb{E}^{[l,k]} \left\langle \mathscr{Q}_\phi \partial_{\Phi^{(l)}_{i_l,i_{l-1}}} \phi \right\rangle = n^{-\frac{3}{2}} \mathbb{E}^{[l,k]} \left\langle P'_\phi(\mathcal{E}') \right\rangle, \quad \forall \phi \in \mathcal{M}_{j_m}^{(m|\nu'_m|\gamma'+1)} \tag{5.5.59}$$

for $m \in \{l+1, \ldots, L\}$, $j_m \in \{1, \ldots, n_m\}$.

Now, we are ready to conclude. Due to (5.5.41), the summation in (5.5.39) can to restricted to be over the set in (5.5.41). Also note that $\mathcal{F}' \subseteq \mathcal{E}'$ due to their definitions in (5.5.43) and (5.5.54). Using these, (5.5.46) and (5.5.59), we can rewrite the left-hand side of (5.5.39) as

$$\begin{aligned} \mathbb{E}^{[l,k]} \left[\langle Q(\mathcal{F}) \rangle \, \Phi^{(l)}_{i_l,i_{l-1}} \right] &= \left(\sum_{\phi \in \mathcal{M}_{i_l}^{(l|\nu'_l|\gamma'+1)}} + \sum_{m=l+1}^{L} \sum_{j_m=1}^{n_m} \sum_{\phi \in \mathcal{M}_{j_m}^{(m|\nu'_m|\gamma'+1)}} \right) \mathbb{E}^{[l,k]} \left\langle \mathscr{Q}_\phi \partial_{\Phi^{(l)}_{i_l,i_{l-1}}} \phi \right\rangle \\ &= n^{-\frac{1}{2}} \mathbb{E}^{[l,k]} \left\langle P'(\mathcal{E}') \right\rangle \end{aligned} \tag{5.5.60}$$

where

$$P'(\mathcal{E}') = \sum_{\phi \in \mathcal{M}_{i_l}^{(l|\nu_l'|\gamma'+1)}} Q_\phi'(\mathcal{F}') + \sum_{m=l+1}^{L} \sum_{j_m=1}^{n_m} \sum_{\phi \in \mathcal{M}_{j_m}^{(m|\nu_m'|\gamma'+1)}} n^{-1} P_\phi'(\mathcal{E}'). \tag{5.5.61}$$

Inserting (5.5.60) to (5.5.37) gives the desired result (5.5.26). Then, we verify (5.5.27). Note that $\sum_{j=1}^{n_m}$ in (5.5.61) is a summation of $O(n)$ many terms. Using this, (5.5.45), and (5.5.58), we obtain that

$$\|P'\| \leqslant C\|Q\|,$$

which along with (5.5.38) implies (5.5.27). Lastly, by (5.5.36), (5.5.44) and (5.5.55), we can see that ν_m'' in the definition of $\mathcal{E}'$ in (5.5.54) satisfies

$$\nu_m'' = \nu_m + (2^{m-l+1} - 1) \vee 0, \quad \forall m \in \{1, \ldots, L\},$$

completing the proof. □

CHAPTER 6

Hamilton–Jacobi equations from mean-field spin glass models

This chapter is essentially borrowed from [41], joint with Hong-Bin Chen.

Abstract. We establish the well-posedness of Hamilton–Jacobi equations arising from mean-field spin glass models in the viscosity sense. Originally defined on the set of monotone probability measures, these equation can be interpreted, via an isometry, to be defined on an infinite-dimensional closed convex cone with empty interior in a Hilbert space. We prove the comparison principle, and the convergence of finite-dimensional approximations furnishing the existence of solutions. Under additional convexity conditions, we show that the solution can be represented by a version of the Hopf–Lax formula, or the Hopf formula on cones. As the first step, we show the well-posedness of equations on finite-dimensional cones, which is self-contained and, we believe, is of independent interest. The key observation making our program possible is that, due to the monotonicity of the nonlinearity, boundary condition is not needed.

Previously, two notions of solutions were considered, one defined directly as the Hopf–Lax formula, and another as limits of finite-dimensional approximations. They have been proven to describe the limit of free energy in a wide class of mean-field spin glass models. This work shows that these two kinds of solutions are viscosity solutions.

6.1. Introduction

Recently, J.-C. Mourrat [92, 98, 96, 93] initiated a novel Hamilton–Jacobi equation approach to studying the limit free energy of mean-field spin glass models. After interpreting the inverse temperature as the temporal variable, and enriching the model by adding a random magnetic field with a parameter viewed as the spacial variable, one can compare the enriched free energy with solutions to a certain Cauchy problem of a Hamilton–Jacobi equation.

Let us give an overview of these equations. The spacial variable, denoted by ϱ, lives in $\mathcal{P}^{\nearrow}$

the set of monotone probability measures on $\mathbb{S}^D_+$, the cone of $D \times D$ positive semi-definite matrices (see Section 6.4.1 for definitions and properties of monotone probability measures) for some fixed $D \in \mathbb{N}$. Formally, the equation is of the following form:

$$\partial_t f - \int \xi(\partial_\varrho f) \mathrm{d}\varrho = 0, \quad \text{on } \mathbb{R}_+ \times \mathcal{P}^{\nearrow}, \tag{6.1.1}$$

where ξ is a real-valued function on $\mathbb{R}^{D\times D}$, and $\mathbb{R}_+ = [0,\infty)$ throughout.

Two notions of solutions have been considered. In [92, 98] where ξ is convex, the solution is defined by a version of the Hopf–Lax formula, which has been proven there to be equivalent to the celebrated Parisi's formula first proposed in [101] and rigorously verified in [72, 111] (see also [100, 99, 112, 113]). In [96, 93], the solution, defined as limits of finite-dimensional approximations, was shown to be an upper bound for the limiting free energy in a wide class of models.

The *ad hoc* and *extrinsic* nature of these two notions motivates us to seek an *intrinsic* definition of solutions. We want to define solutions in the viscosity sense, and establish the well-posedness of the equation, by which we mean the validity of a comparison principle and the existence of solutions. Moreover, we verify that the solution is the limit of finite-dimensional approximations, and, under certain convexity conditions, the solution admits a representation by a variational formula. In particular, we want to ensure that solutions understood in the aforementioned two notions are in fact viscosity solutions. Therefore, the framework of viscosity solutions is compatible with the existing theory.

The key difficulty is to find a natural definition of solutions in the viscosity sense so that all goals announced above are achievable. The surprising observation is that it is sufficient to simply require the solution to satisfy the equation in the viscosity sense everywhere, including the boundary without prescribing any additional condition (e.g. Neumann or Dirichlet) on the boundary. Let us expand the discussion below.

We start with some basics. To make sense of the differential $\partial_\varrho f$, we restrict $\mathcal{P}^{\nearrow}$ to $\mathcal{P}^{\nearrow}_2$, the set of monotone measures with finite second moments, and equip $\mathcal{P}_2^{\nearrow}$ with the 2-Wasserstein metric. Heuristically, the derivative $\partial_\varrho f(t,\varrho)$ describes the asymptotic behavior of $f(t,\vartheta) - f(t,\varrho)$ as ϑ tends to ϱ in the transportational sense, namely, in the Wasserstein metric. Fortunately, $\mathcal{P}_2^{\nearrow}$ can be isometrically embedded onto a closed convex cone in an L^2 space. This cone has empty interior but generates the L^2 space. So, we cannot restrict to a subspace to ensure that the cone has nonempty interior. Via this isometry, $\partial_\varrho f$ can be understood in the sense of the Fréchet derivative.

Therefore, we can interpret (6.1.1) as a special case of the Hamilton–Jacobi equation

$$\partial_t f - \mathsf{H}(\nabla f) = 0, \quad \text{on } \mathbb{R}_+ \times \mathcal{C}, \tag{6.1.2}$$

where $\mathcal{C}$ is a closed convex cone in a separable Hilbert space $\mathcal{H}$, and H is a general nonlinearity. Aside from the lack of local compactness in infinite dimensions, one important issue is to figure out a suitable boundary condition. The spin glass setting does not provide a direct hint, except for invalidating the Dirichlet boundary condition. Moreover, as aforementioned, the solution to (6.1.1) is expected to satisfy the Hopf–Lax formula under some convexity condition, and to be the limit of finite-dimensional approximations. These can be hard to verify if the boundary condition is not easy to work with. The fact that the cone in the spin glass setting has empty interior adds more difficulty.

To bypass these obstacles, we exploit the assumption that H is nondecreasing along the direction given by the dual cone of $\mathcal{C}$, which holds in the spin glass setting. Under this assumption, as aforementioned, we do not need to impose any additional condition on the boundary, and only need the equation to be satisfied in the viscosity sense (see Definition 6.1.5). This greatly simplifies our analysis and allows passing to the limit in a straightforward way. It is surprising that well-posedness holds under this simple definition because usually some boundary condition is needed.

In Section 6.2, we study (6.1.2) on general finite-dimensional cones. Under the monotonicity assumption on H, we will prove the comparison principle (implying the uniqueness of solutions), the existence of solutions, and, under extra convexity conditions, the representation of the solution as either the Hopf–Lax formula, or the Hopf formula. This section is self-contained, and we believe the results there are of independent interest.

In Section 6.3, we consider (6.1.2) on the infinite-dimensional cone relevant to the spin glass models. After establishing the comparison principle, we show that the limit of solutions to finite-dimensional approximations of (6.3) is a viscosity solution of (6.3). Here, the construction of finite-dimensional approximations has the flavor of projective limits. We also verify that the Hopf–Lax formula and the Hopf formula are stable when passed to the limit. In the last subsection, we present a way to make sense of the boundary of the cone despite the fact that it has empty interior.

In Section 6.4, we start with a brief description of mean-field spin glass models. We will introduce more definitions, basic results, and constructions, leading to an interpretation of viscosity solutions of (6.1.1) in Definition 6.4.3. Then, we derive basic properties of the nonlinear term in the equation, which allows us to combine results from other sections to prove the main result, Theorem 6.4.8. Below is a formal restatement of our main result.

Theorem 6.1.1. *Under certain assumptions on ξ and on the initial condition ψ, which are admissible in mean-field spin glass models, there is a unique viscosity solution f of the Cauchy problem of* (6.1.1)*. Moreover,*

1. *f is the limit of viscosity solutions of finite-dimensional approximations of* (6.1.1)*;*
2. *f is given by the Hopf–Lax formula* (6.4.21) *if ξ is convex on $\mathbb{S}^D_+$;*
3. *f is given by the Hopf formula* (6.4.22) *if ψ is convex.*

Accompanying this, a version of the comparison principle holds. In Remark 6.4.12, we explain in more detail that solutions considered in [92, 98, 96, 93] are viscosity solutions.

Lastly, in Section 6.5, we prove that on the cones underlying the finite-dimensional equations that approximate (6.1.1), a version of the Fenchel–Moreau biconjugation identity holds, which is needed for the validity of the Hopf formula as a solution. We believe this is also a new result.

We close this section with a discussion on related works, and a description of the general setting and definitions related to (6.1.2) that are used throughout the paper.

6.1.1. Related works

First, we briefly review existing works on Hamilton–Jacobi equations in Hilbert spaces and Wasserstein spaces.

Equations on Banach spaces satisfying the Radon–Nikodym property (in particular, separable Hilbert spaces) were initially studied in [43, 44], where the differential is understood in the Fréchet sense and the definition of viscosity solutions is a straightforward extension of definitions in finite dimensions. Comparison principles and existence results were established. Our interpretation of solutions are close in spirit to them. We will use Stegall's variational principle (restated as Theorem 6.3.7) as used in [43] to compensate for the lack of local compactness, in order to prove the comparison principle (Proposition 6.3.8). Different from [44], we directly use finite-dimensional approximations to furnish the existence result. As demonstrated in [44, Section 5], there are examples where, under an ordinary setting, finite-dimensional approximations converge to a solution of a different equation. Hence, the class of equations in this work provides an interesting example where the finite-dimensional approximations work properly. Moreover, since the domain for (6.1.2) is a closed convex cone with empty interior, simple modifications of methods for existence results in [44] may not be viable. Works with modified definitions of viscosity solutions for equations in Hilbert spaces also include [45, 46, 47, 48, 114, 61, 62].

Investigations of Hamilton–Jacobi equations on the Wasserstein space of probability measures include [33, 32, 34, 66, 7, 67, 68]. There are mainly three notions of differentiability considered in these works. Let $\mathcal{P}_2(\mathbb{R}^d)$ be the 2-Wasserstein space of probability measures on $\mathbb{R}^d$ for some $d \in \mathbb{N}$. The first way to make sense of differentiability is through defining the tangent space at each $\varrho \in \mathcal{P}_2(\mathbb{R}^d)$ by analogy to differential manifolds. The tangent space at ϱ is the closure of $\{\nabla\phi : \phi \in C_c^\infty(\mathbb{R}^d)\}$ in $L^2(\mathbb{R}^d, \varrho)$. We refer to [8] for more details. The second one, more extrinsic, starts by extending any function $g : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$ through defining $G : L^2(\Omega, \mathbb{P}) \to \mathbb{R}$ by $G[X] = g(\mathsf{Law}(X))$ for every $\mathbb{R}^d$-valued random variable $X \in L^2(\Omega, \mathbb{P})$ on some nice probability space $(\Omega, \mathbb{P})$. Then, one can make sense of the differentiability of g via the Fréchet differentiability of G. One issue is that there can be two different random variables with the same law, which leads to the situation where ϱ, ϑ can be "lifted" to X, Y, respectively, while X and Y are not optimally coupled, namely, the L^2 norm of $X - Y$ not equal to the metric distance between ϱ and ϑ. Another issue is the lack of a canonical choice of Ω. For details, we refer to [32, 68]. The third notion is based on viewing $\mathcal{P}_2(\mathbb{R}^d)$ as a geodesic metric space. Denoting by $\mathbf{d}_2$ the 2-Wasserstein metric, for any $g : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$, one can define the slope of g at ϱ by $|\nabla g| = \limsup_{\vartheta \to \varrho} \frac{|g(\vartheta) - g(\varrho)|}{\mathbf{d}_2(\vartheta, \varrho)}$. Then, one can study equations involving slopes. This notion was considered in [7].

The notion of differentiability adopted in this work is close in spirit to the second one discussed above. But, ours is more intrinsic in the sense that there is an isometry (see (6.4.3)) between $\mathcal{P}_2^{\nearrow}$ and a closed convex cone in an L^2 space. As a result of the monotonicity (see (6.4.1)) of measures in $\mathcal{P}_2^{\nearrow}$, the isometry is given by the right-continuous inverse of some analogue of the probability distribution function, which has already been observed in [93, Section 2]. Hence, in our case, we can identify $[0, 1)$ equipped with the Borel sigma-algebra and the Lebesgue measure as the canonical probability space Ω, appearing in the discussion of the second notion. It is natural and convenient to use the Hilbert space structure of $L^2([0, 1))$ to define differentiability.

To the best of our knowledge, there are no prior works on well-posedness of Hamilton–Jacobi equations on a domain with boundary in infinite-dimensions, or on a Wasserstein space over a set with boundary.

Considerations of using Hamilton–Jacobi equations to study the free energy of mean-field disordered models first appeared in physics literature [69, 71, 22, 21]. The approach was mathematically initiated in [95], and used subsequently in [94, 36, 39, 37, 40] to treat statistical inference models. There, the equations also take the form (6.1.1) but are defined on finite-dimensional cones. Similar to the equation in spin glass models, the nonlinearity is monotonic along the direction of the dual cone (which is the same cone as the cones in these models are self-dual). In these works, some additional Neumann-type conditions were imposed on the boundary. We remark that these conditions can be dropped and the results in [95, 39, 37], where solutions were defined in the viscosity sense, still hold with our simplified definition of viscosity solutions (Definition 6.1.5). Facts about viscosity solutions proven and used there can be replaced by those in Section 6.2.

In [95, 39, 37], the viscosity solution can always admit an expression as the Hopf formula. To prove this, a version of the Fenchel–Moreau biconjugation identity on cones is needed, which has been proven for a large class of cones in [38]. However, the cones pertinent to spin glass models do not fall in that class. As aforementioned, we will prove the identity on these cones in Section 6.5, following similar arguments as in [38].

Using the monotonicity of the nonlinearity, [49, 109] showed that the viscosity solution to a Hamilton–Jacobi equation on an open set Ω in finite dimensions can be extended to a viscosity solution on $\Omega \cup \{z\}$ for any regular point $z \in \partial\Omega$. The result is not applicable to our case. Instead, we study the equation directly on a closed cone.

6.1.2. General setting and definitions

Let $\mathcal{H}$ be a separable Hilbert space with inner product $\langle \cdot\,,\cdot \rangle_{\mathcal{H}}$ and associated norm $|\cdot|_{\mathcal{H}}$. Let $\mathcal{C} \subseteq \mathcal{H}$ be a closed convex cone. In addition, we assume that $\mathcal{C}$ generates $\mathcal{H}$, namely,

$$\mathsf{cl}\,(\mathcal{C} - \mathcal{C}) = \mathcal{H}, \tag{6.1.3}$$

where cl is the closure operator. The dual cone of $\mathcal{C}$ is defined to be

$$\mathcal{C}^* = \{x \in \mathcal{H} : \langle x, y \rangle_{\mathcal{H}} \geqslant 0, \ \forall y \in \mathcal{C}\}. \tag{6.1.4}$$

It is clear that $\mathcal{C}^*$ is a closed and convex cone. We recall the following classical result (c.f. [23, Corollary 6.33]).

Lemma 6.1.2. *If $\mathcal{C} \subseteq \mathcal{H}$ is a closed convex cone, then $(\mathcal{C}^*)^* = \mathcal{C}$ where*

$$(\mathcal{C}^*)^* = \{x \in \mathcal{H} : \langle x, y \rangle_{\mathcal{H}} \geqslant 0, \ \forall y \in \mathcal{C}^*\}.$$

Definition 6.1.3 (Differentiability and smoothness)**.**

1. A function $\phi : (0, \infty) \times \mathcal{C} \to \mathbb{R}$ is said to be *differentiable* at $(t, x) \in (0, \infty) \times \mathcal{C}$, if there is an element in $\mathbb{R} \times \mathcal{H}$, denoted by $(\partial_t \phi(t, x), \nabla \phi(t, x))$ and called the *differential* of ϕ at (t, x), such that

$$\phi(s, y) - \phi(t, x) = \partial_t \phi(t, x)(s - t) + \langle \nabla \phi(t, x), y - x \rangle_{\mathcal{H}} + o\left(|s - t| + |y - x|_{\mathcal{H}}\right),$$

as $(s, y) \in (0, \infty) \times \mathcal{C}$ tends to (t, x) in $\mathbb{R} \times \mathcal{H}$.

2. A function $\phi : (0, \infty) \times \mathcal{C} \to \mathbb{R}$ is said to be *smooth* if

   (a) ϕ is differentiable everywhere with differentials satisfying that, for every $(t, x) \in (0, \infty) \times \mathcal{C}$,

$$\begin{aligned}\phi(s, y) - \phi(t, x) = \partial_t \phi(t, x)(s - t) + \langle \nabla \phi(t, x), y - x \rangle_{\mathcal{H}} \\ + O\left(|s - t|^2 + |y - x|^2_{\mathcal{H}}\right),\end{aligned}$$

   as $(s, y) \in (0, \infty) \times \mathcal{C}$ tends to (t, x) in $\mathbb{R} \times \mathcal{H}$;

   (b) the function $(t, x) \mapsto (\partial_t \phi(t, x), \nabla \phi(t, x))$ is continuous from $(0, \infty) \times \mathcal{C}$ to $\mathbb{R} \times \mathcal{H}$.

3. A function $g : \mathcal{C} \to \mathbb{R}$ is said to be *differentiable* at $x \in \mathcal{C}$, if there is an element in $\mathcal{H}$, denoted by $\nabla g(x)$ and called the *differential* of g at x, such that

$$g(y) - g(x) = \langle \nabla g(x), y - x \rangle_{\mathcal{H}} + o\left(|y - x|_{\mathcal{H}}\right),$$

as $y \in \mathcal{C}$ tends to x in $\mathcal{H}$.

Remark 6.1.4. Note that the differential is defined at every point of the closed cone $\mathcal{C}$, which is needed to make sense of differentials at boundary points. Also, in infinite dimensions, $\mathcal{C}$ can have empty interior. Let us show that the differential is unique whenever it exists. Hence, the above is well-defined.

To see this, it suffices to show that, for any fixed $(t, x) \in (0, \infty) \times \mathcal{C}$, if $(r, h) \in \mathbb{R} \times \mathcal{H}$ satisfies $r(s - t) + \langle h, y - x \rangle_{\mathcal{H}} = o(|s - t| + |y - x|_{\mathcal{H}})$ for all $(s, y) \in (0, \infty) \times \mathcal{C}$, then we must have $r = 0$ and $h = 0$. It is easy to see that $r = 0$. Replacing y by $x + \varepsilon z$ for $\varepsilon > 0$ and any fixed $z \in \mathcal{C}$, and sending $\varepsilon \to 0$, we can deduce that $\langle h, z \rangle_{\mathcal{H}} = 0$ for all $z \in \mathcal{C}$, which along with (6.1.3) implies that $h = 0$.

For a closed cone $\mathcal{K} \subseteq \mathcal{H}$, a function $g : \mathcal{E} \to (-\infty, \infty]$ defined on a subset $\mathcal{E} \subseteq \mathcal{H}$ is said to be $\mathcal{K}$-*nondecreasing* (over $\mathcal{E}$) if g satisfies that

$$g(x) \geqslant g(x'), \qquad \text{for all } x, x' \in \mathcal{E} \text{ satisfying } x - x' \in \mathcal{K}. \tag{6.1.5}$$

Let $\mathsf{H} : \mathcal{H} \to \mathbb{R}$ be a continuous function. The following are conditions often imposed on H:

(A1) H is locally Lipschitz;

(A2) H is $\mathcal{C}^*$-nondecreasing.

Since we will work with equations defined on different cones, in different ambient Hilbert spaces, and with different nonlinearities, for convenience, we denote (6.1.2) by $\mathrm{HJ}(\mathcal{H}, \mathcal{C}, \mathsf{H})$. The Cauchy problem of $\mathrm{HJ}(\mathcal{H}, \mathcal{C}, \mathsf{H})$ with initial condition $\psi : \mathcal{C} \to \mathbb{R}$ is denoted by

$\mathrm{HJ}(\mathcal{H}, \mathcal{C}, \mathsf{H}; \psi)$.

Definition 6.1.5 (Viscosity solutions)**.**

1. A continuous function $f : \mathbb{R}_+ \times \mathcal{C} \to \mathbb{R}$ is a *viscosity subsolution* of $\mathrm{HJ}(\mathcal{H}, \mathcal{C}, \mathsf{H})$ if for every $(t, x) \in (0, \infty) \times \mathcal{C}$ and every smooth $\phi : (0, \infty) \times \mathcal{C} \to \mathbb{R}$ such that $f - \phi$ has a local maximum at (t, x), we have

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

2. A continuous function $f : \mathbb{R}_+ \times \mathcal{C} \to \mathbb{R}$ is a *viscosity supersolution* of $\mathrm{HJ}(\mathcal{H}, \mathcal{C}, \mathsf{H})$ if for every $(t, x) \in (0, \infty) \times \mathcal{C}$ and every smooth $\phi : (0, \infty) \times \mathcal{C} \to \mathbb{R}$ such that $f - \phi$ has a local minimum at (t, x), we have

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

3. A continuous function $f : \mathbb{R}_+ \times \mathcal{C} \to \mathbb{R}$ is a *viscosity solution* of $\mathrm{HJ}(\mathcal{H}, \mathcal{C}, \mathsf{H})$ if f is both a viscosity subsolution and supersolution.

Here, a local extremum at (t, x) is understood to be an extremum over a metric ball of some positive radius centered at (t, x) intersected with $(0, \infty) \times \mathcal{C}$.

For $\psi : \mathcal{C} \to \mathbb{R}$, we say $f : \mathbb{R}_+ \times \mathcal{C} \to \mathbb{R}$ is a viscosity solution of $\mathrm{HJ}(\mathcal{H}, \mathcal{C}, \mathsf{H}; \psi)$ if f is a viscosity solution of $\mathrm{HJ}(\mathcal{H}, \mathcal{C}, \mathsf{H}; \psi)$ and satisfies $f(0, \cdot) = \psi$.

Throughout, Lipschitzness of any real-valued function on a subset of $\mathcal{H}$ or $\mathbb{R} \times \mathcal{H}$ is defined with respect to $|\cdot|_{\mathcal{H}}$ or $|\cdot|_{\mathbb{R}\times\mathcal{H}}$, respectively. A *Lipschitz viscosity solution* is a viscosity solution that is Lipschitz.

Remark 6.1.6. Again, we note that $\mathcal{C}$ is closed, and allowed to have empty interior in infinite dimensions. The definition of viscosity solutions does not prescribe any additional boundary condition. It is surprising that, under (A1)–(A2), well-posedness can be established.

To talk about variational formulae, we also need definitions of convex conjugates and monotone conjugates. For any function $g : \mathcal{H} \to (-\infty, \infty]$, we define its *convex conjugate* by

$$g^{\circledast}(y) = \sup_{x \in \mathcal{H}} \{\langle x, y \rangle_{\mathcal{H}} - g(x)\}, \quad \forall y \in \mathcal{H}. \tag{6.1.6}$$

For $\mathcal{E} \supset \mathcal{C}$ and $g : \mathcal{E} \to (-\infty, \infty]$, we define the *monotone conjugate* (over $\mathcal{C}$) of g by

$$g^{*}(y) = \sup_{x \in \mathcal{C}} \{\langle x, y \rangle_{\mathcal{H}} - g(x)\}, \quad \forall y \in \mathcal{H}. \tag{6.1.7}$$

Throughout, for every $a, b \in \mathbb{R}$, we write $a \vee b = \max\{a, b\}$, $a \wedge b = \min\{a, b\}$, and $a_+ = a \vee 0$.

Acknowledgement

We warmly thank Jean-Christophe Mourrat for many stimulating discussions and helpful comments.

6.2. Equations on finite-dimensional cones

Throughout this section, if not otherwise specified, we assume that $\mathcal{H}$ is finite-dimensional. We consider the setting given in Section 6.1.2 and study the equation $\mathrm{HJ}(\mathcal{H}, \mathcal{C}, \mathsf{H})$. We will prove the comparison principle and the existence of solutions in Section 6.2.1. Then, we will show that the solution can be represented by the Hopf-Lax formula, if H is convex, in Section 6.2.2, or by the Hopf formula, if the initial is convex, in Section 6.2.3.

6.2.1. Comparison principle and existence of solutions

We prove the comparison principle (Proposition 6.2.1) and the existence of solutions (Proposition 6.2.3).

Proposition 6.2.1 (Comparison principle). *Under assumption* (A1), *if u is a viscosity* *subsolution and v is a viscosity supersolution of* $\mathrm{HJ}(\mathcal{H}, \mathcal{C}, \mathsf{H})$ *satisfying*

$$\sup_{t \in \mathbb{R}_+} \|u(t, \cdot)\|_{\mathrm{Lip}} < \infty, \qquad \sup_{t \in \mathbb{R}_+} \|v(t, \cdot)\|_{\mathrm{Lip}} < \infty,$$

then $\sup_{\mathbb{R}_+ \times \mathcal{C}}(u - v) = \sup_{\{0\} \times \mathcal{C}}(u - v)$.

Later, we will need a stronger version of the comparison principle stated below.

Proposition 6.2.2 (Comparison principle in a stronger form)**.** *Under assumption* (A1)*, let u be a viscosity subsolution of* $\mathrm{HJ}(\mathcal{H}, \mathcal{C}, \mathsf{H})$ *and v be a viscosity supersolution of* $\mathrm{HJ}(\mathcal{H}, \mathcal{C}', \mathsf{H})$*, with either $\mathcal{C} \subseteq \mathcal{C}'$ or $\mathcal{C}' \subseteq \mathcal{C}$. Suppose that*

$$L = \sup_{t \in \mathbb{R}_+} \|u(t, \cdot)\|_{\mathrm{Lip}} \vee \|v(t, \cdot)\|_{\mathrm{Lip}}$$

is finite. Then, for every $R > 0$ and every $M > 2L$, the function

$$\mathbb{R}_+ \times (\mathcal{C} \cap \mathcal{C}') \ni (t, x) \mapsto u(t, x) - v(t, x) - M(|x|_{\mathcal{H}} + Vt - R)_+ \tag{6.2.1}$$

achieves its global supremum on $\{0\} \times (\mathcal{C} \cap \mathcal{C}')$, where

$$V = \sup\left\{ \frac{|\mathsf{H}(y) - \mathsf{H}(y')|}{|y - y'|_{\mathcal{H}}} : |y|_{\mathcal{H}},\, |y'|_{\mathcal{H}} \leqslant 2L + 3M \right\}.$$

Let us first deduce Proposition 6.2.1 from Proposition 6.2.2.

Proof of Proposition 6.2.1. Let us argue by contradiction and assume $\sup_{\mathbb{R}_+ \times \mathcal{C}}(u - v) > \sup_{\{0\} \times \mathcal{C}}(u - v)$. Let L be given in the statement of Proposition 6.2.2. Fixing some $M > 2L$ and choosing $R > 0$ sufficiently large, we can get $\sup_{\mathbb{R}_+ \times \mathcal{C}}(u - v - \chi) > \sup_{\{0\} \times \mathcal{C}}(u - v - \chi)$ where $\chi(t, x) = M(|x|_{\mathcal{H}} + Vt - R)_+$ for $(t, x) \in \mathbb{R}_+ \times \mathcal{C}$. However, this contradicts the result ensured by Proposition 6.2.2. □

The proof below is a modification of the proof of [96, Proposition 3.2].

Proof of Proposition 6.2.2. For $\delta \in (0,1)$ to be chosen, let $\theta : \mathbb{R} \to \mathbb{R}_+$ be a nondecreasing smooth function satisfying

$$|\theta'| \leqslant 1, \qquad \text{and} \qquad (r-\delta)_+ \leqslant \theta(r) \leqslant r_+, \quad \forall r \in \mathbb{R},$$

where θ' is the derivative of θ. We define

$$\Phi(t,x) = M\theta\left(\left(\delta + |x|_{\mathcal{H}}^2\right)^{\frac{1}{2}} + Vt - R\right), \quad \forall (t,x) \in \mathbb{R}_+ \times \mathcal{C}.$$

It is immediate that

$$\sup_{(t,x)\in\mathbb{R}_+\times\mathcal{C}} |\nabla\Phi(t,x)|_{\mathcal{H}} \leqslant M, \tag{6.2.2}$$

$$\partial_t \Phi \geqslant V|\nabla\Phi|_{\mathcal{H}}, \tag{6.2.3}$$

$$\Phi(t,x) \geqslant M(|x|_{\mathcal{H}} + Vt - R - 1)_+, \quad \forall (t,x) \in \mathbb{R}_+ \times \mathcal{C}. \tag{6.2.4}$$

We argue by contradiction and assume that the function in (6.2.1) does not achieve its supremum on $\{0\} \times (\mathcal{C} \cap \mathcal{C}')$. Then, we can fix $\delta \in (0,1)$ sufficiently small and $T > 0$ sufficiently large so that

$$\sup_{[0,T)\times(\mathcal{C}\cap\mathcal{C}')} (u - v - \Phi) > \sup_{\{0\}\times(\mathcal{C}\cap\mathcal{C}')} (u - v - \Phi).$$

For $\varepsilon > 0$ to be determined, we define

$$\chi(t,x) = \Phi(t,x) + \varepsilon t + \frac{\varepsilon}{T-t}, \quad \forall (t,x) \in [0,T) \times \mathcal{C}.$$

In view of the previous display, we can choose $\varepsilon > 0$ small and further enlarge T so that

$$\sup_{[0,T)\times(\mathcal{C}\cap\mathcal{C}')} (u - v - \chi) > \sup_{\{0\}\times(\mathcal{C}\cap\mathcal{C}')} (u - v - \chi). \tag{6.2.5}$$

For each $\alpha > 1$, we introduce

$$\Psi_\alpha(t,x,t',x') = u(t,x) - v(t',x') - \frac{\alpha}{2}(|t-t'|^2 + |x-x'|^2_{\mathcal{H}}) - \chi(t,x),$$
$$\forall (t,x,t',x') \in [0,T) \times \mathcal{C} \times [0,T] \times \mathcal{C}'.$$

By the definition of L and (6.2.4), setting $C_1 = \sup_{t\in[0,T]}(|u(t,0)| \vee |v(t,0)|)$, we can see that

$$\Psi_\alpha(t,x,t',x') \leqslant C_1 + L(2|x|_{\mathcal{H}} + |x-x'|_{\mathcal{H}}) - \frac{1}{2}|x-x'|^2 - M(|x|_{\mathcal{H}} - R - 1)_+.$$

Hence, due to the requirement $M > 2L$, Ψ_α is bounded from above uniformly in $\alpha > 1$ and decays as $|x|_{\mathcal{H}}, |x'|_{\mathcal{H}} \to \infty$. Since $\mathcal{H}$ is finite-dimensional, we can see that Ψ_α achieves its supremum at some $(t_\alpha, x_\alpha, t'_\alpha, x'_\alpha)$. The above display also implies that there is C such that

$$|x_\alpha|_{\mathcal{H}},\ |x'_\alpha|_{\mathcal{H}} \leqslant C, \quad \forall \alpha > 1.$$

Setting $C_0 = \Psi_\alpha(0,0,0,0)$ which is independent of α, we have

$$C_0 \leqslant \Psi(t_\alpha, x_\alpha, t'_\alpha, x_\alpha) \leqslant C_1 + 2LC - \frac{\alpha}{2}(|t_\alpha - t'_\alpha|^2 + |x_\alpha - x'_\alpha|^2_{\mathcal{H}}).$$

From this, we can see that $\alpha(|t_\alpha - t'_\alpha|^2 + |x_\alpha - x'_\alpha|^2_{\mathcal{H}})$ is bounded as $\alpha \to \infty$. Hence, passing to a subsequence if necessary, we may assume $t_\alpha, t'_\alpha \to t_0$ and $x_\alpha, x'_\alpha \to x_0$ for some $(t_0, x_0) \in [0,T] \times (\mathcal{C} \cap \mathcal{C}')$.

Then, we show $t_0 \in (0,T)$. Since

$$C_0 \leqslant \Psi(t_\alpha, x_\alpha, t'_\alpha, x_\alpha) \leqslant C_1 + 2LC - \frac{\varepsilon}{T - t_\alpha},$$

we must have that t_α is bounded away from T uniformly in α, which implies $t_0 < T$. Since

$$\begin{aligned} u(t_\alpha, x_\alpha) - v(t'_\alpha, x'_\alpha) - \chi(t_\alpha, x_\alpha) &\geqslant \Psi_\alpha(t_\alpha, x_\alpha, t'_\alpha, x'_\alpha) \\ &\geqslant \sup_{[0,T)\times(\mathcal{C}\cap\mathcal{C}')} (u - v - \chi) \geqslant (u - v - \chi)(t_0, x_0), \end{aligned}$$

sending $\alpha \to \infty$, we deduce that

$$(u - v - \chi)(t_0, x_0) = \sup_{[0,T)\times(\mathcal{C}\cap\mathcal{C}')} (u - v - \chi).$$

This along with (6.2.5) implies that $t_0 > 0$. In conclusion, we have $t_0 \in (0, T)$, and thus $t_\alpha, t'_\alpha \in (0, T)$ for sufficiently large α. Henceforth, we fix any such α.

Before proceeding, we want to obtain a bound on $|x_\alpha - x'_\alpha|_{\mathcal{H}}$. First, we consider the case $\mathcal{C} \subseteq \mathcal{C}'$. Using $\Psi_\alpha(t_\alpha, x_\alpha, t'_\alpha, x_\alpha) - \Psi_\alpha(t_\alpha, x_\alpha, t'_\alpha, x'_\alpha) \leqslant 0$, the computation that

$$\Psi_\alpha(t_\alpha, x_\alpha, t'_\alpha, x_\alpha) - \Psi_\alpha(t_\alpha, x_\alpha, t'_\alpha, x'_\alpha) = v(t'_\alpha, x'_\alpha) - v(t'_\alpha, x_\alpha) + \frac{\alpha}{2}|x_\alpha - x'_\alpha|^2_{\mathcal{H}},$$

and the definition of L, we can get $\alpha|x_\alpha - x'_\alpha|_{\mathcal{H}} \leqslant 2L$. If $\mathcal{C}' \subseteq \mathcal{C}$, we use $\Psi_\alpha(t_\alpha, x'_\alpha, t'_\alpha, x'_\alpha) - \Psi_\alpha(t_\alpha, x_\alpha, t'_\alpha, x'_\alpha) \leqslant 0$, and

$$\begin{aligned} \Psi_\alpha(t_\alpha, x'_\alpha, t'_\alpha, x'_\alpha) - \Psi_\alpha(t_\alpha, x_\alpha, t'_\alpha, x'_\alpha) = u(t_\alpha, x'_\alpha) - u(t_\alpha, x_\alpha) + \frac{\alpha}{2}|x_\alpha - x'_\alpha|^2_{\mathcal{H}} \\ -\Phi(t_\alpha, x'_\alpha) + \Phi(t_\alpha, x_\alpha). \end{aligned}$$

By the definition of L and (6.2.2), we can conclude that, in both cases,

$$\alpha|x_\alpha - x'_\alpha|_{\mathcal{H}} \leqslant 2(L + M). \tag{6.2.6}$$

With this, we return to the proof. Since the function

$$(t, x) \mapsto \Psi_\alpha(t, x, t'_\alpha, x'_\alpha)$$

achieves its maximum at $(t_\alpha, x_\alpha) \in (0,T)\times\mathcal{C}$, by the assumption that u is subsolution, we have

$$\alpha(t_\alpha - t'_\alpha) + \varepsilon + \varepsilon(T-t_\alpha)^{-2} + \partial_t\Phi(t_\alpha, x_\alpha) - \mathsf{H}\left(\alpha(x_\alpha - x'_\alpha) + \nabla\Phi(t_\alpha, x_\alpha)\right) \leqslant 0 \quad (6.2.7)$$

On the other hand, since the function

$$(t', x') \mapsto \Psi_\alpha(t_\alpha, x_\alpha, t', x')$$

achieves its minimum at $(t'_\alpha, x'_\alpha) \in (0,\infty)\times\mathcal{C}'$, by the assumption that v is subsolution, we have

$$\alpha(t_\alpha - t'_\alpha) - \mathsf{H}\left(\alpha(x_\alpha - x'_\alpha)\right) \geqslant 0. \quad (6.2.8)$$

By (6.2.2) and (6.2.6), the arguments inside H in both (6.2.7) and (6.2.8) have norms bounded by $2L+3M$. Taking the difference of (6.2.7) and (6.2.8), and using the definition of V and (6.2.3), we obtain that

$$\varepsilon \leqslant V|\nabla\Phi(t_\alpha, x_\alpha)| - \partial_t\Phi(t_\alpha, x_\alpha) \leqslant 0,$$

contradicting the fact that $\varepsilon > 0$. Therefore, the desired result must hold. □

Proposition 6.2.3 (Existence of solutions)**.** *Under assumption* (A1)–(A2)*, for every Lipschitz* $\psi : \mathcal{C} \to \mathbb{R}$*, there is a viscosity solution* f *of* $\mathrm{HJ}(\mathcal{H}, \mathcal{C}, \mathsf{H}; \psi)$*. Moreover,* f *is Lipschitz and satisfies*

$$\sup_{t\in\mathbb{R}_+} \|f(t,\cdot)\|_{\mathrm{Lip}} = \|\psi\|_{\mathrm{Lip}}, \quad (6.2.9)$$

$$\sup_{x\in\mathcal{C}} \|f(\cdot,x)\|_{\mathrm{Lip}} \leqslant \sup_{\substack{p\in\mathcal{H}\\ |p|_{\mathcal{H}}\leqslant\|\psi\|_{\mathrm{Lip}}}} |\mathsf{H}(p)|. \quad (6.2.10)$$

Proof. Except for one modification, the existence follows from the Perron's method as in

[20, Theorem 7.1] or [42, Theorem 4.1] (see also the proof of [96, Proposition 3.4]). As commented in [42, Remark 4.5], we only need to make sure that any classical subsolution is a viscosity subsolution, stated more precisely as follows.

Lemma 6.2.4. *Under assumption* (A2)*, suppose that* $f : \mathbb{R}_+ \times \mathcal{C} \to \mathbb{R}$ *is continuously differentiable and satisfies* $\partial_t f - \mathsf{H}(\nabla f) \leqslant 0$ *(respectively,* $\partial_t f - \mathsf{H}(\nabla f) \geqslant 0$*) everywhere. Then* f *is a viscosity subsolution (respectively, supersolution) of* $\mathrm{HJ}(\mathcal{H}, \mathcal{C}, \mathsf{H})$.

Proof. We assume that $\partial_t f - \mathsf{H}(\nabla f) \leqslant 0$ everywhere and that $f - \phi$ achieves a local maximum at $(t, x) \in (0, \infty) \times \mathcal{C}$ for some smooth function ϕ. If $x \in \mathcal{C} \setminus \partial\mathcal{C}$, then we clearly have $\partial_t \phi(t, x) = \partial_t f(t, x)$ and $\nabla\phi(t, x) = \nabla f(t, x)$, which along with the assumption on f implies that $\partial_t \phi - \mathsf{H}(\nabla\phi) \leqslant 0$ at (t, x).

Now, let us consider the case $x \in \partial\mathcal{C}$. By the local maximality of $f - \phi$ at (t, x), we have

$$\begin{aligned}\phi(t', x') - \phi(t, x) &\geqslant f(t', x') - f(t, x)\\ &= \partial_t f(t, x)(t' - t) + \left\langle \nabla f(t, x), x' - x \right\rangle_{\mathcal{H}} + o(|t' - t| + |x' - x|_{\mathcal{H}}),\end{aligned}$$

for (t', x') sufficiently close to (t, x). Due to $t \in (0, \infty)$, replacing x' by x and varying t', we can see that

$$\partial_t \phi(t, x) = \partial_t f(t, x). \tag{6.2.11}$$

Then, replacing t' by t and x' by $(1 - \varepsilon)x + \varepsilon y$ for $\varepsilon \in [0, 1]$ and any fixed $y \in \mathcal{C}$, we can obtain by sending $\varepsilon \to 0$ that

$$\langle \nabla\phi(t, x), y - x \rangle_{\mathcal{H}} \geqslant \langle \nabla f(t, x), y - x \rangle_{\mathcal{H}}, \quad \forall y \in \mathcal{C}$$

which implies that $\nabla\phi(t, x) - \nabla f(t, x) \in \mathcal{C}^*$ by the definition of $\mathcal{C}^*$ in (6.1.4). Since H is $\mathcal{C}^*$-nondecreasing, we obtain $\mathsf{H}(\nabla\phi(t, x)) \geqslant \mathsf{H}(\nabla f(t, x))$. Using this, (6.2.11) and the assumption on f, we conclude again $\partial_t \phi - \mathsf{H}(\nabla\phi) \leqslant 0$ at (t,x). Hence, we conclude that f is a viscosity subsolution. The verification for supersolutions is similar, with inequalities above reversed. □

It remains to prove (6.2.9) and (6.2.10). The identity (6.2.9) has been proved in [96, Proposition 3.4]. Assuming (6.2.9), let us prove (6.2.10) here. We denote the right-hand side of (6.2.10) by L. We argue by contradiction and assume that there exist $t,t' \in [0,T)$, $x \in \mathcal{C}$, and $\delta > 0$ for some sufficiently large $T > 0$ such that

$$f(t,x) - f(t',x) > (L+\delta)|t-t'|. \tag{6.2.12}$$

For $\varepsilon > 0$ to be chosen, we let $\theta : \mathbb{R} \to \mathbb{R}_+$ be a nondecreasing smooth function satisfying $(r-\varepsilon) \vee 0 \leqslant \theta(r) \leqslant r \vee 0$ for every $r \in \mathbb{R}$. We set

$$\Phi(x) = M\theta\left((\varepsilon + |x|^2_{\mathcal{H}})^{\frac{1}{2}} - R\right) \tag{6.2.13}$$

for $M, R > 0$ to be chosen. Due to (6.2.12), by choosing R sufficiently large and ε sufficiently small, we have

$$\sup_{t,t' \in [0,T),\, x \in \mathcal{C}} f(t,x) - f(t',x) - (L+\delta)|t-t'| - \Phi(x) - \frac{\varepsilon}{T-t} - \frac{\varepsilon}{T-t'} \geqslant a \tag{6.2.14}$$

for some $a > 0$.

For every $\alpha > 1$, we consider

$$\Psi_\alpha(t,x,t',x') = f(t,x) - f(t',x') - (L+\delta)|t-t'| - \alpha|x-x'|^2_{\mathcal{H}} - \Phi(x) - \frac{\varepsilon}{T-t} - \frac{\varepsilon}{T-t'}$$

for $(t,x,t',x') \in [0,T) \times \mathcal{C} \times [0,T) \times \mathcal{C}$. Using (6.2.9), we can see that $f(t,x) - f(t',x')$ is bounded from above by

$$|f(t,0)| + |f(t',0)| + \|\psi\|_{\mathrm{Lip}}(|x|_{\mathcal{H}} + |x'|_{\mathcal{H}}) \leqslant \|\psi\|_{\mathrm{Lip}}(2|x|_{\mathcal{H}} + |x-x'|_{\mathcal{H}}) + 2\sup_{t\in[0,T]}|f(t,0)|.$$

Hence, by choosing M in Ψ sufficient large, we can ensure that Ψ_α is bounded from above and decays as $x, x' \to \infty$. Moreover, Ψ_α tends to $-\infty$ as $t, t' \to T$. These imply that Ψ_α is maximized at some $(t_\alpha, x_\alpha, t'_\alpha, x'_\alpha)$. Using $\Psi_\alpha(t_\alpha, x_\alpha, t'_\alpha, x'_\alpha) \geqslant \Psi_\alpha(t_\alpha, x_\alpha, t'_\alpha, x_\alpha)$, the computation

$$\Psi_\alpha(t_\alpha, x_\alpha, t'_\alpha, x_\alpha) - \Psi_\alpha(t_\alpha, x_\alpha, t'_\alpha, x'_\alpha) = f(t'_\alpha, x'_\alpha) - f(t'_\alpha, x_\alpha) + \alpha|x_\alpha - x'_\alpha|^2_{\mathcal{H}},$$

and (6.2.9), we can get $|x_\alpha - x'_\alpha|_{\mathcal{H}} = o(\alpha^{-1})$ as $\alpha \to \infty$. Note that we have, by (6.2.14),

$$\Psi_\alpha(t_\alpha, x_\alpha, t'_\alpha, x'_\alpha) \geqslant \sup_{t,t'\in[0,T),\, x\in\mathcal{C}} \Psi_\alpha(t, x, t', x) \geqslant a,$$

and, by (6.2.9),

$$\Psi_\alpha(t_\alpha, x_\alpha, t_\alpha, x'_\alpha) \leqslant \Psi_\alpha(t_\alpha, x_\alpha, t_\alpha, x_\alpha) + \|\psi\|_{\mathrm{Lip}}|x_\alpha - x'_\alpha|_{\mathcal{H}} \leqslant \|\psi\|_{\mathrm{Lip}}\, o\left(\alpha^{-1}\right).$$

The above two displays imply that, for α sufficiently large, we have $t_\alpha \neq t'_\alpha$ and thus at least one of them is nonzero. Now, let us fix any such α.

If $t_\alpha > t'_\alpha$, since f is a viscosity solution and since $(t,x) \mapsto \Psi_\alpha(t, x, t'_\alpha, x'_\alpha)$ achieves a local maximum at (t_α, x_α), we can get

$$\partial_t \phi(t_\alpha, x_\alpha) - \mathsf{H}\left(\nabla\phi(t_\alpha, x_\alpha)\right) \leqslant 0, \tag{6.2.15}$$

where ϕ is given by rewriting $\Psi_\alpha(t, x, t'_\alpha, x'_\alpha) = f(t,x) - \phi(t,x)$. Then, we show

$$|\nabla\phi(t_\alpha, x_\alpha)| \leqslant \|\psi\|_{\mathrm{Lip}}. \tag{6.2.16}$$

Since $f-\phi$ achieves a local maximum at (t_α, x_α), by (6.2.9), we have

$$\phi(t_\alpha, x) - \phi(t_\alpha, x_\alpha) \geqslant f(t_\alpha, x) - f(t_\alpha, x_\alpha) \geqslant -\|\psi\|_{\mathrm{Lip}}|x - x_\alpha|_{\mathcal{H}}, \quad \forall x \in \mathcal{C}.$$

For any $y \in \mathcal{C}$, replacing x by $x_\alpha + \varepsilon\lambda(y - x_\alpha)$ for $\lambda \geqslant 0$ and $\varepsilon > 0$ sufficiently small in the above, and sending $\varepsilon \to 0$, we have

$$\langle \lambda(y - x_\alpha), \nabla\phi(t_\alpha, x_\alpha)\rangle_{\mathcal{H}} \geqslant -\|\psi\|_{\mathrm{Lip}}|\lambda(y - x_\alpha)|_{\mathcal{H}}, \quad \forall \lambda \geqslant 0,\ y \in \mathcal{C}.$$

On the other hand, using the definition of Φ in (6.2.13), we can see that

$$\nabla\phi(t_\alpha, x_\alpha) = 2\alpha(x_\alpha - x'_\alpha) + \beta x_\alpha$$

for some $\beta = M\theta'((\varepsilon + |x_\alpha|^2)^{\frac{1}{2}} - R)(\varepsilon + |x_\alpha|^2)^{-\frac{1}{2}} \geqslant 0$. Setting $\lambda = 2\alpha + \beta$ and $y = \frac{2\alpha}{2\alpha+\beta}x'_\alpha$, we have $\lambda(y - x_\alpha) = -\nabla\phi(t_\alpha, x_\alpha)$. Inserting this to the previous display, we get (6.2.16). Since we can compute that

$$\partial_t\phi(t_\alpha, x_\alpha) = L + \delta + \varepsilon(T - t_\alpha)^{-2} \geqslant L + \delta,$$

we can deduce $L + \delta \leqslant L$ from (6.2.15), (6.2.16) and the definition of L. Hence, we reach a contradiction.

If $t'_\alpha > t_\alpha$, since $(t', x') \mapsto \Psi_\alpha(t_\alpha, x_\alpha, t', x')$ achieves a local maximum at (t'_α, x'_α) and since f is supersolution, we have

$$\partial_t\widetilde{\phi}(t'_\alpha, x'_\alpha) - \mathsf{H}\left(\nabla\widetilde{\phi}(t'_\alpha, x'_\alpha)\right) \geqslant 0, \tag{6.2.17}$$

for $\widetilde{\phi}$ given by rewriting $\Psi_\alpha(t_\alpha, x_\alpha, t', x') = \widetilde{\phi}(t', x') - f(t', x')$. Now, since $f - \widetilde{\phi}$ achieves a local minimum at (t'_α, x'_α), we have

$$\widetilde{\phi}(t'_\alpha, x) - \widetilde{\phi}(t'_\alpha, x'_\alpha) \leqslant f(t'_\alpha, x) - f(t'_\alpha, x'_\alpha) \leqslant \|\psi\|_{\mathrm{Lip}}|x - x'_\alpha|_{\mathcal{H}}, \quad \forall x \in \mathcal{C}.$$

Similar to the previous case, we can deduce

$$\left\langle \lambda(y - x'_\alpha), \nabla\widetilde{\phi}(t'_\alpha, x'_\alpha) \right\rangle_{\mathcal{H}} \leqslant \|\psi\|_{\mathrm{Lip}} |\lambda(y - x'_\alpha)|_{\mathcal{H}}, \quad \forall \lambda \geqslant 0,\ y \in \mathcal{C}.$$

Here, we simply have $\nabla\widetilde{\phi}(t'_\alpha, x'_\alpha) = 2\alpha(x_\alpha - x'_\alpha)$. Setting $\lambda = 2\alpha$ and $y = x_\alpha$ in the above display, we obtain $|\nabla\widetilde{\phi}(t'_\alpha, x'_\alpha)|_{\mathcal{H}} \leqslant \|\psi\|_{\mathrm{Lip}}$. Since $\partial_t\widetilde{\phi}(t'_\alpha, x'_\alpha) = -(L+\delta) - \varepsilon(T - t'_\alpha)^{-2} \leqslant -(L+\delta)$, this along with (6.2.17) and the definition of L yields $-(L+\delta) \geqslant -L$, reaching a contradiction.

In conclusion, we must have (6.2.10), completing the proof of (6.2.10) and thus the proof of Proposition 6.2.3. □

6.2.2. Hopf-Lax formula

Recall the definition of $g^{\circledast}$ in (6.1.6).

Lemma 6.2.5. *Let $\mathcal{H}$ be a possibly infinite-dimensional Hilbert space, and $\mathcal{C}$ be a closed convex cone in $\mathcal{H}$. Suppose that $g : \mathcal{H} \to (-\infty, \infty]$ is $\mathcal{C}^*$-nondecreasing, lower semicontinuous, and convex, and satisfies $g(0) < \infty$. Then,*

1. *$g^{\circledast}(y) = \infty$ for all $y \notin \mathcal{C}$;*
2. *$g(x) = \sup_{y\in\mathcal{C}}\{\langle x, y\rangle_{\mathcal{H}} - g^{\circledast}(y)\}$ for every $x \in \mathcal{H}$.*

Proof. For every $y \notin \mathcal{C}$, by Lemma 6.1.2, there is $z \in \mathcal{C}^*$ such that $\langle z, y\rangle_{\mathcal{H}} < 0$. For every $\lambda > 0$, we also have $0 \in \mathcal{C} \cap (-\lambda z + \mathcal{C}^*)$. Since g is $\mathcal{C}^*$-nondecreasing, we have $g(-\lambda z) \leqslant g(0)$, which implies

$$g^{\circledast}(y) \geqslant \lambda \langle -z, y\rangle_{\mathcal{H}} - g(-\lambda z) \geqslant \lambda \langle -z, y\rangle_{\mathcal{H}} - g(0).$$

Sending $\lambda \to \infty$, we obtain $g^{\circledast}(y) = \infty$, verifying (1). The standard Fenchel-Moreau theorem (c.f. [23, Theorem 13.32]) gives that $g(x) = \sup_{y\in\mathcal{H}}\{\langle x, y\rangle_{\mathcal{H}} - g^{\circledast}(y)\}$, which along with (1) implies (2). □

We show that when H is convex, the solution can be represented by a version of the Hopf–Lax formula on cones. For the standard version, we refer to [60, 42].

Proposition 6.2.6 (Hopf–Lax formula)**.** *In addition to* (A1)–(A2)*, suppose that* H *is convex and bounded below, and that* $\psi : \mathcal{C} \to \mathbb{R}$ *is Lipschitz and* $\mathcal{C}^*$*-nondecreasing. Let* f *be given by*

$$f(t,x) = \sup_{y \in \mathcal{C}} \left\{ \psi(y) - t\mathsf{H}^{\circledast}\left(\frac{y-x}{t}\right) \right\}, \quad \forall (t,x) \in \mathbb{R}_+ \times \mathcal{C}. \tag{6.2.18}$$

Then, f *is a Lipschitz viscosity solution of* $\mathrm{HJ}(\mathcal{H}, \mathcal{C}, \mathsf{H}; \psi)$*.*

Here, to make sense of (6.2.18) at $t = 0$, we use (6.1.6) to rewrite the right-hand side of (6.2.18) as

$$f(t,x) = \sup_{y \in \mathcal{C}} \inf_{z \in \mathcal{H}} \left\{ \psi(y) - \langle z, y - x \rangle_{\mathcal{H}} + t\mathsf{H}(z) \right\}, \quad \forall (t,x) \in \mathbb{R}_+ \times \mathcal{C}.$$

Then, we can see that, when $t = 0$, the supremum in this display must be achieved at $y = x$, implying $f(0,x) = \psi(x)$ for all $x \in \mathcal{C}$.

We devote the rest of this subsection to the proof of this proposition.

Semigroup property

We show that for all $t > s \geqslant 0$,

$$f(t,x) = \sup_{y \in \mathcal{C}} \left\{ f(s,y) - (t-s)\mathsf{H}^{\circledast}\left(\frac{y-x}{t-s}\right) \right\}, \quad \forall x \in \mathcal{C}. \tag{6.2.19}$$

The convexity of $\mathsf{H}^{\circledast}$ implies that

$$\mathsf{H}^{\circledast}\left(\frac{y-x}{t}\right) \leqslant \frac{s}{t}\mathsf{H}^{\circledast}\left(\frac{y-z}{s}\right) + \frac{t-s}{t}\mathsf{H}^{\circledast}\left(\frac{z-x}{t-s}\right), \quad \forall y, z \in \mathcal{H},$$

which along with (6.2.18) yields that

$$f(t,x) \geqslant \sup_{y,z\in\mathcal{C}} \left\{ \psi(y) - s\mathsf{H}^{\circledast}\left(\frac{y-z}{s}\right) - (t-s)\mathsf{H}^{\circledast}\left(\frac{z-x}{t-s}\right)\right\}$$
$$= \sup_{z\in\mathcal{C}} \left\{ f(s,z) - (t-s)\mathsf{H}^{\circledast}\left(\frac{z-x}{t-s}\right)\right\}.$$

To show the converse inequality, we claim that for any fixed $(t,x) \in (0,\infty)\times\mathcal{C}$, there is $y \in \mathcal{C}$ satisfying.

$$f(t,x) = \psi(y) - t\mathsf{H}^{\circledast}\left(\frac{y-x}{t}\right). \tag{6.2.20}$$

Assuming this, we set $z = \frac{s}{t}x + (1-\frac{s}{t})y$ which satisfies $\frac{z-x}{t-s} = \frac{y-x}{t} = \frac{y-z}{s}$. By this, (6.2.18), and (6.2.20), we have

$$f(s,z) - (t-s)\mathsf{H}^{\circledast}\left(\frac{z-x}{t-s}\right) \geqslant \psi(y) - s\mathsf{H}^{\circledast}\left(\frac{y-z}{s}\right) - (t-s)\mathsf{H}^{\circledast}\left(\frac{z-x}{t-s}\right)$$
$$= \psi(y) - t\mathsf{H}^{\circledast}\left(\frac{y-x}{t}\right) = f(t,x),$$

which yields the desired inequality.

It remains to verify the existence of y in (6.2.20). Fix any $\lambda > 0$ and set $x = \lambda\frac{y}{|y|_{\mathcal{H}}}$ in (6.1.6) for $\mathsf{H}^{\circledast}$ to see that

$$\mathsf{H}^{\circledast}(y) \geqslant \lambda|y|_{\mathcal{H}} - \sup_{|z|_{\mathcal{H}}\leqslant\lambda} |\mathsf{H}(z)|.$$

Since H is locally Lipschitz, the supremum on the right is finite. Hence, we can deduce that

$$\liminf_{y\to\infty} \frac{\mathsf{H}^{\circledast}(y)}{|y|_{\mathcal{H}}} = \infty. \tag{6.2.21}$$

We set $L = \|\psi\|_{\mathrm{Lip}}$. Then, the above implies the existence of $R > 0$ such that $\mathsf{H}^{\circledast}(\frac{y-x}{t}) \geqslant$

$(L+1)\frac{|y-x|_\mathcal{H}}{t}$ for all y satisfying $\frac{|y-x|_\mathcal{H}}{t} > R$. These imply that

$$\psi(y) - t\mathsf{H}^\circledast\left(\frac{y-x}{t}\right) \leqslant \psi(x) + L|y-x|_\mathcal{H} - (L+1)|y-x|_\mathcal{H} = \psi(x) - |y-x|_\mathcal{H},$$

for all y satisfying $|y-x|_\mathcal{H} > tR$. Therefore, the supremum in (6.2.18) can be taken over a bounded set. Also note that the function $y \mapsto \psi(y) - t\mathsf{H}^\circledast(\frac{y-x}{t})$ is upper semi-continuous and locally bounded from above due to $\mathsf{H}^\circledast(z) \geqslant -\mathsf{H}(0)$. Since $\mathcal{H}$ is finite-dimensional, the maximizer must exist, which ensures the existence of y in (6.2.20) and thus completes the proof of (6.2.19).

Lipschitzness

We first show the following claim: for every $(t,x) \in (0,\infty)\times\mathcal{C}$, there is $y \in \mathcal{C}$ satisfying $y - x \in \mathcal{C}$ such that

$$f(t,x) - f(t,x') \leqslant \psi(y) - \psi(y-x+x'), \quad \forall x' \in \mathcal{C}. \tag{6.2.22}$$

Fix any $(t,x) \in (0,\infty)\times\mathcal{C}$. Arguing as before, we can find $y \in \mathcal{C}$ such that (6.2.20) holds. Lemma 6.2.5 (1) ensures that $y-x \in \mathcal{C}$, and thus $y - x + x' \in \mathcal{C}$ for every $x' \in \mathcal{C}$. The Hopf–Lax formula (6.2.18) gives the lower bound

$$f(t,x') \geqslant \psi(y-x+x') - t\mathsf{H}^\circledast\left(\frac{y-x}{t}\right),$$

which along with (6.2.20) yields (6.2.22).

Now, for any $(t,x,x') \in (0,\infty)\times\mathcal{C}\times\mathcal{C}$, we apply (6.2.22) to both x and x' to see that there exist $y, y' \in \mathcal{C}$ such that

$$\psi(y'-x'+x) - \psi(y') \leqslant f(t,x) - f(t,x') \leqslant \psi(y) - \psi(y-x+x'),$$

which immediately implies that

$$\sup_{t>0} \|f(t,\cdot)\|_{\mathrm{Lip}} \leqslant \|\psi\|_{\mathrm{Lip}}. \tag{6.2.23}$$

Then, we show that

$$\sup_{x\in\mathcal{C}} \|f(\cdot,x)\|_{\mathrm{Lip}} \leqslant \max\left\{\left|\mathsf{H}^{\circledast}(0)\right|, \sup_{|p|_{\mathcal{H}}\leqslant\|\psi\|_{\mathrm{Lip}}} |\mathsf{H}(p)|\right\}. \tag{6.2.24}$$

Let us fix any $x \in \mathcal{C}$ and $t > s > 0$. Then, (6.2.19) immediately yields

$$f(t,x) \geqslant f(s,x) - (t-s)\mathsf{H}^{\circledast}(0)$$

where the last term is finite by the assumption that H is bounded below. Next, using (6.2.23), we can obtain from (6.2.19) that

$$f(t,x) \leqslant f(s,x) + \sup_{y\in\mathcal{C}}\left\{\|\psi\|_{\mathrm{Lip}}|x-y|_{\mathcal{H}} - (t-s)\mathsf{H}^{\circledast}\left(\frac{y-x}{t-s}\right)\right\}.$$

Lemma 6.2.5 (1) ensures that $\frac{y-x}{t-s} \in \mathcal{C}$. Replacing $\frac{y-x}{t-s}$ by z, and using $\|\psi\|_{\mathrm{Lip}}|z|_{\mathcal{H}} = \left\langle z, \frac{\|\psi\|_{\mathrm{Lip}}z}{|z|_{\mathcal{H}}}\right\rangle_{\mathcal{H}}$, we can bound the right-hand side of the above display by

$$\begin{aligned}(t-s)\sup_{z\in\mathcal{C}}\left\{\|\psi\|_{\mathrm{Lip}}|z|_{\mathcal{H}} - \mathsf{H}^{\circledast}(z)\right\} &\leqslant (t-s)\sup_{|p|_{\mathcal{H}}\leqslant\|\psi\|_{\mathrm{Lip}}}\sup_{z\in\mathcal{C}}\left\{\langle z,p\rangle_{\mathcal{H}} - \mathsf{H}^{\circledast}(z)\right\} \\ &= (t-s)\sup_{|p|_{\mathcal{H}}\leqslant\|\psi\|_{\mathrm{Lip}}}\mathsf{H}(p),\end{aligned}$$

where the last equality follows from Lemma 6.2.5 (2). The above three displays together yield (6.2.24).

Verification of the Hopf–Lax formula as a supersolution

Suppose $f - \phi$ achieves a local minimum at $(t, x) \in (0, \infty) \times \mathcal{C}$ for some smooth function ϕ. Then,

$$f(t-s, x+sy) - \phi(t-s, x+sy) \geqslant f(t,x) - \phi(t,x)$$

for every $y \in \mathcal{C}$ and sufficiently small $s > 0$. On the other hand, (6.2.19) implies that

$$f(t,x) \geqslant f(t-s, x+sy) - s\mathsf{H}^{\circledast}(y).$$

Combining the above two displays, we obtain that

$$\phi(t,x) - \phi(t-s, x+sy) + s\mathsf{H}^{\circledast}(y) \geqslant 0.$$

Sending $s \to 0$, we have that

$$\partial_t \phi(t,x) - \langle y, \nabla\phi(t,x)\rangle_{\mathcal{H}} + \mathsf{H}^{\circledast}(y) \geqslant 0.$$

Taking infimum over $y \in \mathcal{C}$ and using Lemma 6.2.5 (2), we obtain

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

verifying that f is supersolution.

Verification of the Hopf–Lax formula as a subsolution

Suppose that $f - \phi$ achieves a local maximum at $(t, x) \in (0, \infty) \times \mathcal{C}$. We want to show that

$$\left(\partial_t \phi - \mathsf{H}(\nabla\phi)\right)(t,x) \leqslant 0. \tag{6.2.25}$$

We argue by contradiction and assume that there is $\delta > 0$ such that

$$(\partial_t \phi - \mathsf{H}(\nabla\phi))\,(t', x') \geqslant \delta > 0,$$

for (t', x') sufficiently close to (t, x). The definition of $\mathsf{H}^{\circledast}$ (in (6.1.6)) implies that

$$\partial_t \phi(t', x') - \langle q, \nabla\phi(t', x')\rangle_{\mathcal{H}} + \mathsf{H}^{\circledast}(q) \geqslant \delta \tag{6.2.26}$$

for all such (t', x') and all $q \in \mathcal{H}$.

To proceed, we show that there is $R > 0$ such that for every $s > 0$ sufficiently small there is $x_s \in \mathcal{C}$ such that

$$f(t, x) = f(t - s, x_s) - s\mathsf{H}^{\circledast}\left(\frac{x_s - x}{s}\right), \tag{6.2.27}$$

$$|x - x_s|_{\mathcal{H}} \leqslant Rs. \tag{6.2.28}$$

In view of (6.2.23) and (6.2.24), we set $L = \|f\|_{\mathrm{Lip}} < \infty$. By (6.2.21), we can choose $R > 1$ to satisfy $\mathsf{H}^{\circledast}(z) \geqslant 2L|z|_{\mathcal{H}}$ for every $z \in \mathcal{H}$ satisfying $|z|_{\mathcal{H}} > R$. Then, for every $y \in \mathcal{C}$ satisfying $\frac{|y-x|_{\mathcal{H}}}{s} > R$, we have

$$f(t - s, y) - s\mathsf{H}^{\circledast}\left(\frac{y - x}{s}\right) \leqslant f(t, x) + Ls + L|y - x|_{\mathcal{H}} - 2L|y - x|_{\mathcal{H}} < f(t, x) + Ls(1 - R).$$

Hence, the supremum in (6.2.19) can be taken over $\{y \in \mathcal{C} : |y - x|_{\mathcal{H}} \leqslant Rs\}$. Since $\mathcal{H}$ is finite-dimensional, we can thus conclude the existence of $x_s \in \mathcal{C}$ satisfying (6.2.27) and (6.2.28).

Returning to the proof, we can compute that, for sufficiently small $s > 0$,

$$\begin{aligned}\phi(t, x) - \phi(t - s, x_s) &= \int_0^1 \frac{\mathrm{d}}{\mathrm{d}r}\phi(t + (r-1)s, rx + (1-r)x_s)\mathrm{d}r \\ &= \int_0^1 (s\partial_t\phi - \langle x_s - x, \nabla\phi\rangle)(t + (r-1)s, rx + (1-r)x_s)\mathrm{d}r.\end{aligned}$$

Using (6.2.26) with q replaced by $\frac{x_s - x}{s}$, and (6.2.27), we have

$$\phi(t,x) - \phi(t-s,x_s) \geqslant s\delta - s\mathsf{H}^{\circledast}\left(\frac{x_s - x}{s}\right) \geqslant s\delta + f(t,x) - f(t-s,x_s).$$

Rearranging terms, we arrive at that, for all $s > 0$ sufficiently small,

$$f(t-s,x_s) - \phi(t-s,x_s) \geqslant s\delta + f(t,x) - \phi(t,x),$$

contradicting the local maximality of $f - \phi$ at (t,x). Hence, (6.2.25) must hold, and thus f is a subsolution.

6.2.3. Hopf formula

Recall the definition of monotone conjugate in (6.1.7). Note that, by Lemma 6.1.2, we can verify that g^* is always $\mathcal{C}^*$-nondecreasing, which is reason for the prefix "monotone". The *monotone biconjugate* of g is given by $g^{**} = (g^*)^*$, which can be expressed as

$$g^{**}(x) = \sup_{y \in \mathcal{C}} \{\langle y, x\rangle_{\mathcal{H}} - g^*(x)\}, \quad \forall x \in \mathcal{H}. \tag{6.2.29}$$

It is easy to see that

$$g^{**}(x) \leqslant g(x), \quad \forall x \in \mathcal{C}. \tag{6.2.30}$$

Definition 6.2.7. A closed convex cone $\mathcal{C}$ is said to have the *Fenchel–Moreau property* if the following holds: for every $g : \mathcal{C} \to (-\infty, \infty]$ not identically equal to ∞, we have that $g^{**} = g$ on $\mathcal{C}$ if and only if g is convex, lower semicontinuous and $\mathcal{C}^*$-nondecreasing.

In Section 6.5, we will show that the cones relevant to the spin glass models have the Fenchel–Moreau property. The goal of this subsection is to show that a version of the Hopf formula on cones is a viscosity solution. In [75], Hopf proposed this formula as a solution given that the initial condition is convex or concave, which was later confirmed rigorously in [84].

Proposition 6.2.8 (Hopf formula)**.** *In addition to* (A1)–(A2)*, suppose that* $\mathcal{C}$ *has the Fenchel–Moreau property and that* $\psi : \mathcal{C} \to \mathbb{R}$ *is convex, Lipschitz and* $\mathcal{C}^*$*-nondecreasing. Then,* $f : \mathbb{R}_+ \times \mathcal{C} \to \mathbb{R}$ *given by*

$$f(t,x) = \sup_{z\in\mathcal{C}} \inf_{y\in\mathcal{C}} \left\{\langle z, x-y\rangle_{\mathcal{H}} + \psi(y) + t\mathsf{H}(z)\right\}, \quad \forall (t,x) \in \mathbb{R}_+ \times \mathcal{C}, \tag{6.2.31}$$

is a Lipschitz viscosity solution of $\mathrm{HJ}(\mathcal{H}, \mathcal{C}, \mathsf{H}; \psi)$*.*

We will also need the following equivalent forms of the Hopf formula (6.2.31):

$$f(t,x) = \sup_{z\in\mathcal{C}} \{\langle z, x\rangle_{\mathcal{H}} - \psi^*(z) + t\mathsf{H}(z)\} \tag{6.2.32}$$

$$= (\psi^* - t\mathsf{H})^*(x). \tag{6.2.33}$$

Remark 6.2.9. For concave initial condition ψ, one would expect that the following version of Hopf formula,

$$f(t,x) = \inf_{z\in\mathcal{C}} \sup_{y\in\mathcal{C}} \left\{\langle z, x-y\rangle_{\mathcal{H}} + \psi(y) + t\mathsf{H}(z)\right\}, \quad \forall (t,x) \in \mathbb{R}_+ \times \mathcal{C}, \tag{6.2.34}$$

is a viscosity solution. However, it is seemingly not valid here. Let us briefly explain this. In the proof of Proposition 6.2.8, we will need the assumption that H is $\mathcal{C}^*$-nondecreasing in several places, for instance, in the derivation of (6.2.40). Attempts to verify that formula (6.2.34) is a solution fail at these places, where the monotonicity of H only yields inequalities in undesired directions.

We check the following in order: initial condition, semigroup property (or dynamic programming principle), Lipschitzness, that the Hopf formula gives a subsolution, and that the Hopf formula gives a supersolution.

Verification of the initial condition

Using (6.2.33), we have $f(0,\cdot) = \psi^{**}$. Then, the Fenchel–Moreau property of $\mathcal{C}$ ensures that $\psi^{**} = \psi$.

Semigroup property

For f given in (6.2.31), we want to show, for all $s \geqslant 0$,

$$f(t+s,x) = \sup_{z\in\mathcal{C}} \inf_{y\in\mathcal{C}} \left\{ \langle z, x-y\rangle_{\mathcal{H}} + f(t,y) + s\mathsf{H}(z) \right\},$$

or, in a more compact form,

$$f(t+s,\cdot) = \left(f^*(t,\cdot) - s\mathsf{H}\right)^*. \tag{6.2.35}$$

In view of the Hopf formula (6.2.33), this is equivalent to

$$\left(\psi^* - (t+s)\mathsf{H}\right)^* = \left((\psi^* - t\mathsf{H})^{**} - s\mathsf{H}\right)^*. \tag{6.2.36}$$

Since the Fenchel transform is order-reversing, (6.2.30) implies that

$$\left((\psi^* - t\mathsf{H})^{**} - s\mathsf{H}\right)^* \geqslant \left(\psi^* - (t+s)\mathsf{H}\right)^*. \tag{6.2.37}$$

To see the other direction, we use (6.2.30) to get

$$\frac{s}{t+s}\psi^* + \frac{t}{t+s}\left(\psi^* - (t+s)\mathsf{H}\right)^{**} \leqslant \psi^* - t\mathsf{H}.$$

For any g, it can be readily checked that g^* is convex, lower semicontinuous, and $\mathcal{C}^*$-nondecreasing. Taking the monotone biconjugate in the above display and applying the Fenchel–Moreau property of $\mathcal{C}$, we have

$$\frac{s}{t+s}\psi^* + \frac{t}{t+s}\left(\psi^* - (t+s)\mathsf{H}\right)^{**} \leqslant (\psi^* - t\mathsf{H})^{**}.$$

Then, we rearrange terms and use (6.2.30) to see

$$(\psi^* - (t+s)\mathsf{H})^{**} - (\psi^* - t\mathsf{H})^{**} \leqslant \frac{s}{t}\left((\psi^* - t\mathsf{H})^{**} - \psi^*\right) \leqslant -s\mathsf{H},$$

and thus

$$(\psi^* - (t+s)\mathsf{H})^{**} \leqslant (\psi^* - t\mathsf{H})^{**} - s\mathsf{H}.$$

Taking the monotone conjugate on both sides, using its order-reversing property, and invoking the Fenchel–Moreau property of $\mathcal{C}$, we get

$$(\psi^* - (t+s)\mathsf{H})^{*} \geqslant \left((\psi^* - t\mathsf{H})^{**} - s\mathsf{H}\right)^{*},$$

which together with (6.2.37) verifies (6.2.36).

Lipschitzness

Since ψ is Lipschitz, we have $\psi^*(z) = \infty$ outside the compact set $B = \{z \in \mathcal{C} : |z|_{\mathcal{H}} \leqslant \|\psi\|_{\mathrm{Lip}}\}$. This together with (6.2.32) implies that for each $x \in \mathcal{C}$, there is $z \in B$ such that

$$f(t,x) = \langle z, x\rangle_{\mathcal{H}} - \psi^*(z) + t\mathsf{H}(z). \tag{6.2.38}$$

Using this and (6.2.32), we get that

$$f(t,x) - f(t,x') \leqslant \left\langle z, x - x'\right\rangle_{\mathcal{H}} \leqslant \|\psi\|_{\mathrm{Lip}}|x-x'|_{\mathcal{H}}, \quad \forall x' \in \mathcal{C}.$$

By symmetry, we conclude that $f(t,\cdot)$ is Lipschitz, and the Lipschitz coefficient is uniform in t.

To show the Lipschitzness in t, we fix any $x \in \mathcal{C}$. Then, we have, for some $z \in B$,

$$f(t,x) = \langle z, x\rangle_{\mathcal{H}} - \psi^*(z) + t\mathsf{H}(z) \leqslant f(t',x) + (t-t')\mathsf{H}(z)$$
$$\leqslant f(t',x) + |t'-t| \left(\sup_{|z|_{\mathcal{H}} \leqslant \|\psi\|_{\mathrm{Lip}}} |\mathsf{H}(z)| \right).$$

Again by symmetry, the Lipschitzness of $f(\cdot, x)$ is obtained, and its coefficient is independent of x.

Combining these results, we conclude that f is Lipschitz.

Verification of the Hopf formula as a subsolution

Let $\phi : (0,\infty) \times \mathcal{C} \to \mathbb{R}$ be smooth. Suppose that $f - \phi$ achieves a local maximum at $(t,x) \in (0,\infty) \times \mathcal{C}$. Arguing as above, there is $z \in \mathcal{C}$ such that (6.2.38) holds. By this and (6.2.32), we have, for $s \in [0,t]$ and $h \in \mathcal{C}$,

$$f(t,x) \leqslant f(t-s, x+h) - \langle z, h\rangle_{\mathcal{H}} + s\mathsf{H}(z).$$

By the assumption on ϕ ensures that

$$f(t-s,x+h) - \phi(t-s,x+h) \leqslant f(t,x) - \phi(t,x).$$

for small $s \in [0,t]$ and small $h \in \mathcal{C}$. Then, we combine the above two inequalities to get

$$\phi(t,x) - \phi(t-s,x+h) \leqslant - \langle z, h\rangle_{\mathcal{H}} + s\mathsf{H}(z), \tag{6.2.39}$$

for sufficiently small $s \geqslant 0$ and $h \in \mathcal{C}$. We can set $s = 0$, substitute εy for h for any $y \in \mathcal{C}$ and sufficiently small $\varepsilon > 0$, and then send $\varepsilon \to 0$ to see $\langle y, \nabla\phi(t,x) - z\rangle_{\mathcal{H}} \geqslant 0$ for all $y \in \mathcal{C}$, which implies that $\nabla\phi(t,x) - z \in \mathcal{C}^*$. Since H is $\mathcal{C}^*$-nondecreasing, this implies

$$\mathsf{H}(\nabla\phi(t,x)) \geqslant \mathsf{H}(z). \tag{6.2.40}$$

Then, we set $h = 0$ in (6.2.39), take $s \to 0$ to obtain $\partial_t \phi(t,x) \leqslant \mathsf{H}(z)$, which along with the above display gives

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

Hence, we conclude that f is a viscosity subsolution.

Verification of the Hopf formula as a supersolution

The idea of proof in this part can be seen in [84, Proof of Proposition 1]. Let $(t,x) \in (0,\infty) \times \mathcal{C}$ be a local minimum point for $f - \phi$. Due to (6.2.32), f is convex in both variables. Since $\mathcal{C}$ is also convex, we have, for all $(t',x') \in (0,\infty) \times \mathcal{C}$ and all $\lambda \in (0,1]$,

$$f(t',x') - f(t,x) \geqslant \frac{1}{\lambda} \left(f\left(t + \lambda(t'-t), x + \lambda(x'-x)\right) - f(t,x) \right).$$

For any fixed (t',x') and sufficiently small $\lambda > 0$, the local minimality of $f - \phi$ at (t,x) implies that

$$f\left(t + \lambda(t'-t), x + \lambda(x'-x)\right) - f(t,x) \geqslant \phi\left(t + \lambda(t'-t), x + \lambda(x'-x)\right) - \phi(t,x).$$

Using the above two displays and setting $\lambda \to 0$, we obtain

$$f(t',x') - f(t,x) \geqslant r(t'-t) + \left\langle \nabla \phi(t,x),\, x'-x \right\rangle_{\mathcal{H}} \tag{6.2.41}$$

where, for convenience, we set

$$r = \partial_t \phi(t,x). \tag{6.2.42}$$

Before proceeding, we make a digression to convex analysis. For every convex $g : \mathcal{H} \to$

$(-\infty, \infty]$ and every $y \in \mathcal{H}$, the subdifferential of g is defined by

$$\partial g(y) = \left\{z \in \mathcal{H} : g(y') \geqslant g(y) + \left\langle z, y' - y \right\rangle_{\mathcal{H}}, \forall y' \in \mathcal{H}\right\}. \tag{6.2.43}$$

For any convex set $\mathcal{E} \subseteq \mathcal{H}$, the outer normal cone to $\mathcal{E}$ at $y \in \mathcal{H}$ is defined to be

$$\mathbf{n}_{\mathcal{E}}(y) = \left\{z \in \mathcal{H} : \left\langle z, y' - y \right\rangle_{\mathcal{H}} \leqslant 0, \forall y' \in \mathcal{E}\right\}.$$

Since $f(t, \cdot)$ is convex, setting $t' = t$ in (6.2.41), we have

$$\nabla \phi(t, x) \in \partial f(t, x). \tag{6.2.44}$$

Here, $\partial f(t, x)$ stands for the subdifferential of $f(t, \cdot)$ at x. We need the following lemma characterizing subdifferentials in finite dimensions (c.f. [105, Theorem 25.6]).

Lemma 6.2.10. *Let $g : \mathcal{H} \to (-\infty, \infty]$ be lower semicontinuous, and convex. If $\{g < \infty\}$ has nonempty interior in $\mathcal{H}$, then*

$$\partial g(x) = \mathsf{cl}\left(\mathsf{conv}\, A(x)\right) + \mathbf{n}_{\{g<\infty\}}(x), \quad \forall x \in \{g < \infty\}$$

where $A(x)$ is the set of all limits of sequences of the form $(\nabla g(x_n))_{n=1}^{\infty}$ such that $\lim_{n\to\infty} x_n = x$ and g is differentiable at every x_n.

Since $f(t, \cdot) : \mathcal{C} \to \mathbb{R}$ is convex and continuous, extending $f(t, \cdot)$ by setting $f(t, x) = \infty$ for $x \notin \mathcal{C}$, we can ensure this extension is lower semicontinuous and convex and thus Lemma 6.2.10 is applicable to the extended $f(t, \cdot)$. Moreover, by this extension, we have $\{f(t, \cdot) < \infty\} = \mathcal{C}$ and thus $\mathbf{n}_{\{f(t,\cdot)<\infty\}}(x) = \mathbf{n}_{\mathcal{C}}(x)$, for every $x \in \mathcal{C}$.

Invoking Lemma 6.2.10 to (6.2.44), we can express

$$\nabla \phi(t, x) = a + b \tag{6.2.45}$$

where $b \in \mathbf{n}_{\mathcal{C}}(x)$ and a belongs to the closed convex hull of limit points of the form $\lim_{n\to\infty} \nabla f(t, x_n)$ where $\lim_{n\to\infty} x_n = x$ and $f(t,\cdot)$ is differentiable at each x_n.

Since the supremum in (6.2.32) is taken over $\mathcal{C}$, we can see that $f(t,\cdot)$ is $\mathcal{C}^*$-nondecreasing on $\mathcal{C}$. Then, by Lemma 6.1.2, we can see that the differential of f whenever exists always belongs to $\mathcal{C}$, which implies that

$$a \in \mathcal{C}. \tag{6.2.46}$$

By the definition of $\mathbf{n}_{\mathcal{C}}(x)$ and that of $\mathcal{C}^*$ in (6.1.4), it can seen that $-b \in \mathcal{C}^*$. This along with (6.2.45) implies

$$a \in \nabla\phi(t,x) + \mathcal{C}^*. \tag{6.2.47}$$

By Lemma 6.2.10, the definition of a and an easy observation that $0 \in \mathbf{n}(x)$, we can deduce that $a \in \partial f(t,x)$, which due to the definition of subdifferential in (6.2.43) further implies

$$f(t,x') - f(t,x) \geqslant \left\langle a, x'-x\right\rangle_{\mathcal{H}}, \quad \forall x' \in \mathcal{C}.$$

Then, we set $x' = x$ in (6.2.41) and use the above display to get

$$f(t',x') - f(t,x) \geqslant r(t'-t) + \left\langle a, x'-x\right\rangle_{\mathcal{H}}, \quad \forall (t',x') \in \mathbb{R}_+ \times \mathcal{C}. \tag{6.2.48}$$

Now, we return to the proof. For each $s \geqslant 0$, we define

$$\eta_s(x') = f(t,x) - rs + \left\langle a, x'-x\right\rangle_{\mathcal{H}}, \quad \forall x' \in \mathcal{C}.$$

Setting $t' = t - s$ in (6.2.48), for $s \in [0,t]$, we have

$$f(t-s,x') \geqslant \eta_s(x'), \quad \forall x' \in \mathcal{C}.$$

Applying the order-reversing property of the monotone conjugate twice, we obtain from the above display that

$$\left(f^*(t-s,\cdot) - s\mathsf{H}\right)^* \geqslant \left(\eta_s^* - s\mathsf{H}\right)^*.$$

Due to the semigroup property (6.2.35), this yields

$$f(t,\cdot) \geqslant \left(\eta_s^* - s\mathsf{H}\right)^*, \quad \forall s \in [0,t].$$

By (6.2.46) and the definition of the monotone conjugate in (6.1.7), the above yields

$$f(t,x) \geqslant \langle a, x\rangle_{\mathcal{H}} - \eta_s^*(a) + s\mathsf{H}(a).$$

On the other hand, using the definition of η_s, we can compute

$$\eta_s^*(a) = -f(t,x) + rs + \langle a, x\rangle_{\mathcal{H}}.$$

Combining the above two displays with (6.2.42), we arrive at

$$\left(\partial_t \phi - \mathsf{H}(a)\right)(t,x) \geqslant 0.$$

Lastly, (6.2.47) and the fact that H is $\mathcal{C}^*$-nondecreasing imply $\mathsf{H}(a) \geqslant \mathsf{H}(\nabla\phi(t,x))$, which along with the above display verifies that f is a supersolution.

6.3. Equations on an infinite-dimensional cone

For a fixed positive integer D, let $\mathbb{S}^D$ be the space of $D \times D$-symmetric matrices, and $\mathbb{S}^D_+$ be the cone of $D \times D$-symmetric positive semidefinite matrices. We equip $\mathbb{S}^D$ with the inner product $a \cdot b = \mathrm{tr}(ab)$, for all $a, b \in \mathbb{S}^D$. We can view $\mathbb{S}^D_+$ as a closed convex cone in the Hilbert space $\mathbb{S}^D$. Naturally, $\mathbb{S}^D$ is endowed with the Borel sigma-algebra generated by the norm topology. For $a, b \in \mathbb{S}^D$, we write

$$a \geqslant b, \quad \text{if } a - b \in \mathbb{S}^D_+, \tag{6.3.1}$$

which defines a partial order on $\mathbb{S}^D$.

We work with the infinite-dimensional Hilbert space

$$\mathcal{H} = L^2([0,1), \mathbb{S}^D) \tag{6.3.2}$$

namely, $\mathbb{S}^D$-valued squared integrable functions on $[0,1)$ endowed with the Borel sigma-algebra $\mathcal{B}_{[0,1)}$ and the Lebesgue measure. In addition to the Hilbert space $\mathcal{H}$, we will also need

$$L^p = L^p([0,1), \mathbb{S}^D_+) \tag{6.3.3}$$

for $p \in [1, \infty]$, whose norm is denoted by $|\cdot|_{L^p}$.

We consider the following cone

$$\mathcal{C} = \left\{\mu : [0,1) \to \mathbb{S}^D_+ \;\middle|\; \mu \text{ is right-continuous with left limits, and nondecreasing}\right\}. \tag{6.3.4}$$

Here, μ is said to be nondecreasing if $\mu(t) - \mu(s) \in \mathbb{S}^D_+$ whenever $t \geqslant s$. We view $\mathcal{C} \subseteq \mathcal{H}$ by identifying every element in $\mathcal{C}$ with its equivalence class in $\mathcal{H}$. Since $\{\mathbb{1}_{[t,1)}\}_{t\in[0,1)} \subseteq \mathcal{C}$, it is immediate that $\mathcal{C}$ spans $\mathcal{H}$. More precisely, (6.1.3) holds.

In this section, we study $\mathrm{HJ}(\mathcal{H}, \mathcal{C}, \mathsf{H})$ for $\mathcal{H}$ and $\mathcal{C}$ given above. We start by introducing more notations and basic results in Section 6.3.1. The main results of this section are scattered in subsections afterwards. The comparison principle is given in Proposition 6.3.8. In Section 6.3.3, we show that any limit of finite-dimensional approximations is a viscosity solution (Proposition 6.3.9), and provide sufficient conditions for such a convergence (Proposition 6.3.10). In Section 6.3.4, we show that the Hopf–Lax formula and the Hopf formula are stable, when passed to the limit (Propositions 6.3.12 and 6.3.13). Lastly, in Section 6.3.5, we briefly discuss a way to make sense of the boundary of $\mathcal{C}$ in a weaker notion. Results there are not needed elsewhere.

Throughout, we denote elements in $\mathcal{C}$ by μ, ν, ρ; generic elements in $\mathcal{H}$ by ι, κ; and elements in finite-dimensional spaces by x, y, z.

6.3.1. Preliminaries

We will introduce definitions and notations related to partitions of $[0,1)$, by which the finite approximations of $\mathrm{HJ}(\mathcal{H}, \mathcal{C}, \mathsf{H})$ will be indexed. Projection maps and lifting maps between finite-dimensional approximations and their infinite-dimensional counterparts will be used extensively. Their basic properties are recorded in Lemma 6.3.3. We will also need the projections of $\mathcal{C}$ and their dual cones, the properties of which are collected in Lemmas 6.3.4 and 6.3.5. Lastly, in Lemma 6.3.6, we clarify the relation between the differentiability in finite-dimensional approximations and the one in infinite dimensions.

Partitions

We denote the collection of ordered tuples as partitions of $[0,1)$ by

$$\mathfrak{J} = \cup_{n\in\mathbb{N}} \left\{(t_1, t_2, \ldots, t_n) \in (0,1]^n : 0 < t_1 < t_2 < \cdots < t_{n-1} < t_n = 1\right\}.$$

For every such tuple $j \in \mathfrak{J}$, we set $t_0 = 0$, and denote by $|j|$ the cardinality of j.

A natural partial order on $\mathfrak{J}$ is given by the set inclusion. Under this partial order, a subcollection $\widetilde{\mathfrak{J}} \subseteq \mathfrak{J}$ is said to be *directed* if for every pair $j, j' \in \widetilde{\mathfrak{J}}$, there is $j'' \in \widetilde{\mathfrak{J}}$ such that $j, j' \subseteq j''$.

For each $j \in \mathfrak{J}$, we associate a sigma-algebra $\mathcal{F}_j$ on $[0,1)$ generated by $\{\mathbb{1}_{[t_k, t_{k+1})}\}_{t_k \in j}$. A subcollection $\widetilde{\mathfrak{J}} \subseteq \mathfrak{J}$ is said to be *generating* if $\widetilde{\mathfrak{J}}$ is directed, and the collection of sigma-algebras $\{\mathcal{F}_j\}_{j\in\widetilde{\mathfrak{J}}}$ generates the Borel sigma-algebra on $[0,1)$.

Let $\mathfrak{J}_{\mathsf{unif}}$ be the collection of uniform partitions. A subcollection $\widetilde{\mathfrak{J}} \subseteq \mathfrak{J}$ is said to be *good* if $\widetilde{\mathfrak{J}} \subseteq \mathfrak{J}_{\mathsf{unif}}$ and $\widetilde{\mathfrak{J}}$ is generating. Examples of good collections of partitions include $\mathfrak{J}_{\mathsf{unif}}$ itself, and the collection of dyadic partitions.

In the following, we denote by $\mathfrak{J}_{\mathsf{gen}}$ a generic generating collection of partitions, and by $\mathfrak{J}_{\mathsf{good}}$ a generic good collection.

Then, we introduce the notions of nets and convergence of a net. For any directed subcollection $\widetilde{\mathfrak{J}} \subseteq \mathfrak{J}$, a collection of elements $(x_j)_{j\in\widetilde{\mathfrak{J}}}$, indexed by $\widetilde{\mathfrak{J}}$, from some set $\mathcal{X}$ is called a *net* in $\mathcal{X}$. If $\mathcal{X}$ is a topological space, a net $(x_j)_{j\in\widetilde{\mathfrak{J}}}$ is said to converge in $\mathcal{X}$ to x if for every neighborhood $\mathcal{N}$ of x, there is $j_{\mathcal{N}} \in \mathfrak{J}$ such that $x_j \in \mathcal{N}$ for every $j \in \widetilde{\mathfrak{J}}$ satisfying $j \supset j_{\mathcal{N}}$. In this case, we write $\lim_{j\in\widetilde{\mathfrak{J}}} x_j = x$ in $\mathcal{X}$.

For each $j \in \mathfrak{J}$ and every $\iota \in L^1$, we define

$$\iota^{(j)}(t) = \sum_{k=1}^{|j|} \mathbb{1}_{[t_{k-1},t_k)}(t) \frac{1}{t_k - t_{k-1}} \int_{t_{k-1}}^{t_k} \iota(s)\mathrm{d}s, \quad \forall t \in [0,1). \tag{6.3.5}$$

It is easy to see that $\iota^{(j)}$ is characterized by the condition expectation of ι on $\mathcal{F}_j$, namely,

$$\iota^{(j)}(U) = \mathbb{E}\left[\iota(U)|\mathcal{F}_j\right]. \tag{6.3.6}$$

Here, and throughout, U is uniform random variable on $[0,1)$ defined on the probability space $([0,1), \mathcal{B}_{[0,1)}, \mathrm{Leb})$. By Jensen's inequality, we have $\iota^{(j)} \in L^p$ if $\iota \in L^p$, for any $p \in [1,\infty)$, which also holds obviously for $p = \infty$. In particular, $\iota^{(j)} \in \mathcal{H}$ if $\iota \in \mathcal{H}$. It is straightforward to see that $\iota^{(j)} \in \mathcal{C}$ if $\iota \in \mathcal{C}$.

Projections and lifts

We introduce finite-dimensional Hilbert spaces indexed by $\mathfrak{J}$. For each $j \in \mathfrak{J}$, we define

$$\mathcal{H}^j = (\mathbb{S}^D)^{|j|} \tag{6.3.7}$$

equipped with the inner product

$$\langle x, y\rangle_{\mathcal{H}^j} = \sum_{k=1}^{|j|}(t_k - t_{k-1})x_k \cdot y_k, \quad \forall x, y \in \mathcal{H}^j. \tag{6.3.8}$$

For each $j \in \mathfrak{J}$, we define the projection map $\mathrm{p}_j : \mathcal{H} \to \mathcal{H}^j$ by

$$\mathrm{p}_j \iota = \left(\frac{1}{t_k - t_{k-1}} \int_{t_{k-1}}^{t_k} \iota(s)\mathrm{d}s\right)_{k \in \{1,\dots,|j|\}}, \quad \forall \iota \in \mathcal{H}. \tag{6.3.9}$$

Correspondingly, we define the associated lift map $\mathrm{l}_j : \mathcal{H}^j \to \mathcal{H}$:

$$\mathrm{l}_j x = \sum_{k=1}^{|j|} x_k \mathbb{1}_{[t_{k-1}, t_k)}, \quad \forall x \in \mathcal{H}^j. \tag{6.3.10}$$

We define projections and lifts acting on functions.

Definition 6.3.1 (Lifts and projects of functions)**.** Let $j \in \mathfrak{J}$.

- For any $\mathcal{E} \subseteq \mathcal{H}$ and any $g : \mathcal{E} \to \mathbb{R}$, its j-projection $g^j : \mathrm{p}_j\mathcal{E} \to \mathbb{R}$ is given by $g^j = g \circ \mathrm{l}_j$.
- For any $\mathcal{T} \times \mathcal{E} \subseteq \mathbb{R}_+ \times \mathcal{H}$ and any $f : \mathcal{T} \times \mathcal{E} \to \mathbb{R}$, its j-projection $f^j : \mathcal{T} \times \mathrm{p}_j\mathcal{E} \to \mathbb{R}$ is given by $f^j(t, \cdot) = f(t, \mathrm{l}_j(\cdot))$ for each $t \in \mathcal{T}$.
- For any $\mathcal{E} \subseteq \mathcal{H}^j$ and any function $g : \mathcal{E} \to \mathbb{R}$, its lift $g^\uparrow : \mathrm{l}_j\mathcal{E} \to \mathbb{R}$, is given by $g^\uparrow = g \circ \mathrm{p}_j$.
- for any $\mathcal{T} \times \mathcal{E} \subseteq \mathbb{R}_+ \times \mathcal{H}^j$ and any $f : \mathcal{T} \times \mathcal{E} \to \mathbb{R}$, its lift $f^\uparrow : \mathcal{T} \times \mathrm{l}_j\mathcal{E} \to \mathbb{R}$, is defined by $f^\uparrow(t, \cdot) = f(t, \mathrm{p}_j(\cdot))$ for each $t \in \mathcal{T}$.

Remark 6.3.2. Let us clarify our use of indexes. Objects with superscript j, for instance, $\mathcal{H}^j$, $\mathcal{C}^j$ (introduced later in (6.3.11)), f^j, are always projections of infinite-dimensional objects either mapped directly by p_j or induced by p_j. Superscript (j) is reserved for (6.3.5). Other objects directly associated with j or whose existence depends on j are labeled with subscript j. For example, the solution to finite-dimensional approximation corresponding to the partition j will be denoted as f_j.

We record basic properties of projections and lifts in the following lemma.

Lemma 6.3.3. *For every $j \in \mathfrak{J}$, the following hold:*

1. p_j *and* l_j *are adjoint to each other:* $\langle \mathrm{p}_j \iota, x \rangle_{\mathcal{H}^j} = \langle \iota, \mathrm{l}_j x \rangle_{\mathcal{H}}$ *for every* $\iota \in \mathcal{H}$ *and* $x \in \mathcal{H}^j$*;*
2. l_j *is isometric:* $\langle \mathrm{l}_j x, \mathrm{l}_j y \rangle_{\mathcal{H}} = \langle x, y \rangle_{\mathcal{H}^j}$ *for every* $x, y \in \mathcal{H}^j$*;*
3. $\mathrm{p}_j \mathrm{l}_j$ *is the identity map on* $\mathcal{H}^j$*:* $\mathrm{p}_j \mathrm{l}_j x = x$ *for every* $x \in \mathcal{H}^j$*;*
4. $\mathrm{l}_j \mathrm{p}_j \iota = \iota^{(j)}$ *for every* $\iota \in \mathcal{H}$*;*
5. p_j *is a contraction:* $|\mathrm{p}_j \iota|_{\mathcal{H}^j} \leqslant |\iota|_{\mathcal{H}}$*, or equivalently* $|\iota^{(j)}|_{\mathcal{H}} \leqslant |\iota|_{\mathcal{H}}$*, for every* $\iota \in \mathcal{H}$*;*
6. *if* $j' \in \mathfrak{J}$ *satisfies* $j \subseteq j'$*, then* $\mathrm{p}_j \mathrm{l}_{j'} \mathrm{p}_{j'} \iota = \mathrm{p}_j \iota$ *for every* $\iota \in \mathcal{H}$*.*

In addition, the following results on convergence hold:

7. *for every* $\iota \in \mathcal{H}$*,* $\lim_{j \in \mathfrak{J}_{\mathsf{gen}}} \iota^{(j)} = \iota$ *in* $\mathcal{H}$*;*
8. *for any net* $(\iota_j)_{j \in \mathfrak{J}_{\mathsf{gen}}}$ *in* $\mathcal{H}$*, if* $\lim_{j \in \mathfrak{J}_{\mathsf{gen}}} \iota_j = \iota$ *in* $\mathcal{H}$*, then* $\lim_{j \in \mathfrak{J}_{\mathsf{gen}}} \iota_j^{(j)} = \iota$ *in* $\mathcal{H}$*.*

Proof. Part (1). We can compute:

$$\begin{aligned}
\langle \mathrm{p}_j \iota, x \rangle_{\mathcal{H}^j} &= \sum_{k=1}^{|j|} (t_k - t_{k-1}) \left(\frac{1}{t_k - t_{k-1}} \int_{t_{k-1}}^{t_k} \iota(s) \mathrm{d}s \right) \cdot x_k \\
&= \sum_{k=1}^{|j|} \int_{t_{k-1}}^{t_k} \iota(s) \cdot x_k \mathrm{d}s = \int_0^1 \iota(s) \cdot \left(\sum_{k=1}^{|j|} \mathbb{1}_{[t_{k-1}, t_k)}(s) x_k \right) \mathrm{d}s = \langle \iota, \mathrm{l}_j x \rangle_{\mathcal{H}} .
\end{aligned}$$

Part (2). We use (6.3.10) to compute explicitly to get the desired result.

Part (3). Definitions of p_j in (6.3.9) and l_j in (6.3.10) directly yield (3).

Part (4). Comparing the definitions of p_j, l_j and $\iota^{(j)}$ in (6.3.5), we can easily deduce (4).

Part (5). We use (6.3.6) and Jensen's equality to see

$$\left|\iota^{(j)}\right|_{\mathcal{H}}^2 = \mathbb{E}\left|\iota^{(j)}(U)\right|^2 = \mathbb{E}|\mathbb{E}[\iota(U)|\mathcal{F}_j]|^2 \leqslant \mathbb{E}|\iota(U)|^2 = |\iota|_{\mathcal{H}}^2.$$

The equivalent formulation follows from (2) and (4).

Part (6). We can directly use the definitions of projections and lifts. Heuristically, j is a coarser partition and j' is a refinement of j. The map $\mathrm{l}_{j'}\mathrm{p}_{j'}$ has the effect of locally averaging ι with respect to the finer partition j'. On the other hand, p_j is defined via local averaging with respect to the coarser j. The result follows from the fact that local averaging first with respect to a finer partition and then to a coarser partition is equivalent to local averaging directly with respect to the coarser one.

Part (7). We argue by contradiction. We assume that there exists $\varepsilon > 0$ such that for every $j \in \mathfrak{J}_{\mathsf{gen}}$, there is some $j' \supset j$ satisfying $|\iota^{(j')} - \iota|_{\mathcal{H}} \geqslant \varepsilon$. Let us construct a sequence recursively. We start by choosing $j_1 \in \mathfrak{J}_{\mathsf{gen}}$ to satisfy $|\iota^{(j_1)} - \iota|_{\mathcal{H}} \geqslant \varepsilon$. For $m > 1$, we choose $j_{m+1} \supset (j_m \cup j'_m)$ such that $|\iota^{(j_{m+1})} - \iota|_{\mathcal{H}} \geqslant \varepsilon$, where we let $j'_m \in \mathfrak{J}_{\mathsf{gen}}$ be any partition satisfying $\max_{1\leqslant i\leqslant |j'_m|}\{|t_i - t_{i-1}|\} < \frac{1}{m}$. Denote this sequence by $\mathfrak{J}'_{\mathsf{gen}}$, which is clearly directed and generating.

By (6.3.6), for $j_m, j_n \in \mathfrak{J}'_{\mathsf{gen}}$ such that $n \geqslant m$, we have

$$\iota^{(j_m)}(U) = \mathbb{E}\left[\iota(U)|\mathcal{F}_{j_m}\right] = \mathbb{E}\left[\mathbb{E}\left[\iota(U)|\mathcal{F}_{j_n}\right]|\mathcal{F}_{j_m}\right] = \mathbb{E}\left[\iota^{(j_n)}(U)|\mathcal{F}_{j_m}\right],$$

which implies that $(\iota^{(j_n)})_{n\in\mathbb{N}}$ is a martingale with respect to $(\mathcal{F}_{j_n})_{n\in\mathbb{N}}$. By the martingale convergence theorem, this sequence converges to ι in $\mathcal{H}$ as $n \to \infty$, which is a contradiction to our construction of the sequence.

Part (8). By the triangle inequality and (5), we have

$$\left|\iota_j^{(j)} - \iota\right|_{\mathcal{H}} \leqslant \left|\iota_j^{(j)} - \iota^{(j)}\right|_{\mathcal{H}} + \left|\iota^{(j)} - \iota\right|_{\mathcal{H}} \leqslant |\iota_j - \iota|_{\mathcal{H}} + \left|\iota^{(j)} - \iota\right|_{\mathcal{H}}.$$

Then, (8) follows from (7). □

Cones and dual cones

For each $j \in \mathfrak{J}$, we introduce

$$\mathcal{C}^j = \{x \in \mathcal{H}^j : 0 \leqslant x_1 \leqslant x_2 \leqslant \cdots \leqslant x_{|j|}\}, \tag{6.3.11}$$

where we used the notation in (6.3.1). It is clear that $\mathcal{C}^j$ and $\mathcal{H}^j$ satisfy (6.1.3).

Recall the definition of dual cones in (6.1.4).

Lemma 6.3.4 (Characterizations of dual cones)**.**

1. *For each $j \in \mathfrak{J}$, the dual cone of $\mathcal{C}^j$ in $\mathcal{H}^j$ is*

$$(\mathcal{C}^j)^* = \left\{ x \in \mathcal{H}^j : \sum_{i=k}^{|j|} (t_i - t_{i-1}) x_i \in \mathbb{S}_+^D, \quad \forall k \in \{1, 2, \ldots, |j|\} \right\}.$$

2. *The dual cone of $\mathcal{C}$ in $\mathcal{H}$ is*

$$\mathcal{C}^* = \left\{ \iota \in \mathcal{H} : \int_t^1 \iota(s) \mathrm{d}s \in \mathbb{S}_+^D, \quad \forall t \in [0, 1) \right\}.$$

Proof. Part (1). We denote the set on the right-hand side by RHS. We first show that $(\mathcal{C}^j)^* \subseteq \mathsf{RHS}$. Let $x \in (\mathcal{C}^j)^*$. For every k and every $a \in \mathbb{S}_+^D$, we can choose $y \in \mathcal{C}^j$ such that $0 = y_1 = \cdots = y_{k-1}$ and $y_k = \cdots = y_{|j|} = a$. Then, we have

$$\sum_{i=k}^{|j|} (t_i - t_{i-1}) x_i \cdot a = \sum_{i=k}^{|j|} (t_i - t_{i-1}) x_i \cdot y_i \geqslant 0,$$

which implies that $x \in \mathsf{RHS}$. In the other direction, we assume $x \in \mathsf{RHS}$. For every $y \in \mathcal{C}^j$, by setting $y_0 = 0$, since $y_k - y_{k-1} \in \mathbb{S}^D_+$ for all k, we have

$$\sum_{i=1}^{|j|}(t_i - t_{i-1})x_i \cdot y_i = \sum_{k=1}^{|j|}\left(\sum_{i=k}^{|j|}(t_i - t_{i-1})x_i \cdot (y_k - y_{k-1})\right) \geqslant 0,$$

which gives that $x \in (\mathcal{C}^j)^*$. Now we can conclude that $(\mathcal{C}^j)^* = \mathsf{RHS}$ as desired.

Part (2). We denote the set on the right-hand side by RHS. Let $\iota \in \mathcal{C}^*$. For any $a \in \mathbb{S}^D_+$ and $t \in [0,1)$, we set $\mu = a\mathbb{1}_{[t,1)}$. It is clear that $\mu \in \mathcal{C}$. Due to $\langle \iota, \mu \rangle_{\mathcal{H}} \geqslant 0$ by duality, we deduce that $\iota \in \mathsf{RHS}$.

Now, let $\iota \in \mathsf{RHS}$. We argue by contradiction and assume $\iota \notin \mathcal{C}^*$. Then, by definition, there is $\mu \in \mathcal{C}$ such that $\langle \iota, \mu \rangle_{\mathcal{H}} < 0$. By Lemma 6.3.3 (7), there is a partition j such that $\left\langle \iota^{(j)}, \mu^{(j)} \right\rangle_{\mathcal{H}} < 0$. Due to Lemma 6.3.3 (2) and (4), this can be rewritten as $\langle \mathrm{p}_j \iota, \mathrm{p}_j \mu \rangle_{\mathcal{H}^j} < 0$.

On the other hand, by the definition of p_j in (6.3.9), we can compute that, for every k,

$$\sum_{i=k}^{|j|}(t_i - t_{i-1})(\mathrm{p}_j \iota)_i = \int_{t_{k-1}}^{1} \iota(s)\mathrm{d}s \in \mathbb{S}^D_+$$

by the assumption that $\iota \in \mathsf{RHS}$. Hence, by (1), we have $\iota \in (\mathcal{C}^j)^*$. Since μ is nondecreasing as $\mu \in \mathcal{C}$, it is easy to see that $\mathrm{p}_j \mu \in \mathcal{C}^j$. The detailed computation can be seen in (6.3.13). Therefore, we must have $\langle \mathrm{p}_j \iota, \mathrm{p}_j \mu \rangle_{\mathcal{H}^j} \geqslant 0$, reaching a contradiction. □

Lemma 6.3.5. *For every $j \in \mathfrak{J}$, the following hold:*

1. $\mathrm{l}_j(\mathcal{C}^j) \subseteq \mathcal{C}$;
2. $\mathrm{l}_j((\mathcal{C}^j)^*) \subseteq \mathcal{C}^*$;
3. $\mathrm{p}_j(\mathcal{C}) = \mathcal{C}^j$;
4. $\mathrm{p}_j(\mathcal{C}^*) = (\mathcal{C}^j)^*$;

5. $\mu \in \mu^{(j)} + \mathcal{C}^*$, *for every* $\mu \in \mathcal{C}$.

Proof. We first show that

$$\mathrm{p}_j(\mathcal{C}) \subseteq \mathcal{C}^j \tag{6.3.12}$$

and then verify each claim. For every $\mu \in \mathcal{C}$, it follows from the definition that $\mathrm{p}_j\mu \in \mathcal{H}^j$. Since μ is nondecreasing, setting $(\mathrm{p}_j\mu)_0 = 0$ by our convention, we get,

$$\begin{aligned}(\mathrm{p}_j\mu)_k - (\mathrm{p}_j\mu)_{k-1} &= \frac{1}{t_k - t_{k-1}} \int_{t_{k-1}}^{t_k} \mu(s)\mathrm{d}s - \frac{1}{t_{k-1} - t_{k-2}} \int_{t_{k-2}}^{t_{k-1}} \mu(s)\mathrm{d}s \\ &\geqslant \mu(t_{k-1}) - \mu(t_{k-1}) = 0,\end{aligned} \tag{6.3.13}$$

for $k \in \{2, \cdots, |j|\}$. Clearly when $k = 1$, $(\mathrm{p}_j\mu)_1 = \frac{1}{t_1}\int_0^{t_1} \mu(s)ds \in \mathbb{S}^D_+$. Hence, we have $\mathrm{p}_j\mu \in \mathcal{C}^j$ and thus (6.3.12).

Part (1). For any $x \in \mathcal{C}^j$, recall the definition of $\mathrm{l}_j x$ in (6.3.10). Since $x_k \geqslant x_{k-1}$ for each k, it is clear that $\mathrm{l}_j x$ is nondecreasing and thus belongs to $\mathcal{C}$.

Part (2). Let $x \in (\mathcal{C}^j)^*$. For every $\mu \in \mathcal{C}$, recalling the definition of $\mathrm{p}_j\mu$ in (6.3.9), we have

$$\begin{aligned}\int_0^1 \sum_{k=1}^{|j|} \mathbb{1}_{[t_{k-1},t_k)}(s) x_i \cdot \mu(s)\mathrm{d}s &= \sum_{k=1}^{|j|} \int_{t_{k-1}}^{t_k} x_k \cdot \mu(s)\mathrm{d}s \\ &= \sum_{k=1}^{|j|} (t_k - t_{k-1}) x_k \cdot \left(\frac{1}{t_k - t_{k-1}} \int_{t_{k-1}}^{t_k} \mu(s)\mathrm{d}s \right) \\ &= \sum_{k=1}^{|j|} (t_k - t_{k-1}) x_k \cdot (\mathrm{p}_j\mu)_k \geqslant 0,\end{aligned}$$

where the last inequality holds due to $x \in (\mathcal{C}^j)^*$ and $\mathrm{p}_j\mu \in \mathcal{C}^j$ by (6.3.12). This implies that $\mathrm{l}_j x \in \mathcal{C}^*$, and thus $\mathrm{l}_j((\mathcal{C}^j)^*) \subseteq \mathcal{C}^*$.

Part (3). For every $x \in \mathcal{C}^j$, by (1), we have $\mathrm{l}_j x \in \mathcal{C}$. Lemma 6.3.3 (3) implies that $x = \mathrm{p}_j \mathrm{l}_j x$. Hence, we get $\mathcal{C}^j \subseteq \mathrm{p}_j(\mathcal{C})$. Then, (3) follows from this and (6.3.12).

Part (4). Let $\iota \in \mathcal{C}^*$. For every $x \in \mathcal{C}^j$, we have by Lemma 6.3.3 (1) that $\langle \mathrm{p}_j\iota, x\rangle_{\mathcal{H}^j} = \langle \iota, \mathrm{l}_j x\rangle_{\mathcal{H}} \geqslant 0$ due to $\mathrm{l}_j x \in \mathcal{C}$ ensured by (1). Hence, we have $\mathrm{p}_j(\mathcal{C}^*) \subseteq (\mathcal{C}^j)^*$. For the other direction, let $x \in (\mathcal{C}^j)^*$. Lemma 6.3.3 (3) gives $x = \mathrm{p}_j\mathrm{l}_j x$. Invoking (2), we can deduce that $(\mathcal{C}^j)^* \subseteq \mathrm{p}_j(\mathcal{C}^*)$, completing the proof of (4).

Part (5). We show that $\mu - \mu^{(j)} \in \mathcal{C}^*$. Let $\tau \in [0,1)$ and $a \in \mathbb{S}^D_+$. We choose $t_{k_0} \in j$ such that $\tau \in [t_{k_0-1}, t_{k_0})$. Using the definition of $\mu^{(j)}$ in (6.3.5), we can compute that

$$
\begin{aligned}
&\int_\tau^1 \left(\mu - \mu^{(j)}\right)(s)\mathrm{d}s \\
&= \left(\int_\tau^{t_{k_0}} \mu(s)\mathrm{d}s + \int_{t_{k_0}}^1 \mu(s)\mathrm{d}s\right) - \left(\frac{t_{k_0}-\tau}{t_{k_0}-t_{k_0-1}}\int_{t_{k_0-1}}^{t_{k_0}} \mu(s)\mathrm{d}s + \int_{t_{k_0}}^1 \mu(s)\mathrm{d}s\right), \\
&= \int_\tau^{t_{k_0}} \mu(s)\mathrm{d}s - \frac{t_{k_0}-\tau}{t_{k_0}-t_{k_0-1}}\int_{t_{k_0-1}}^{t_{k_0}} \mu(s)\mathrm{d}s \\
&= (t_{k_0}-\tau)\left(\frac{1}{t_{k_0}-\tau}\int_\tau^{t_{k_0}} \mu(s)\mathrm{d}s - \frac{1}{t_{k_0}-t_{k_0-1}}\int_{t_{k_0-1}}^{t_{k_0}} \mu(s)\mathrm{d}s\right) \geqslant 0,
\end{aligned}
$$

where the last inequality follows from the fact that μ is nondecreasing. By Lemma 6.3.4 (2), we conclude that $\mu - \mu^{(j)} \in \mathcal{C}^*$ as desired. □

Derivatives

Recall Definition 6.1.3 (3) for the differentiability of functions defined on $\mathcal{C}$. We denote by ∇_j the differential operator on functions defined on $\mathcal{C}^j$.

Lemma 6.3.6. *For every $j \in \mathfrak{J}$, the following hold.*

1. *If $g : \mathcal{C} \to \mathbb{R}$ is differentiable at $\mathrm{l}_j x$ for some $x \in \mathcal{C}^j$. Then, $g^j : \mathcal{C}^j \to \mathbb{R}$ is differentiable at x and its differential is given by $\nabla_j g^j(x) = \mathrm{p}_j(\nabla g(\mathrm{l}_j x))$.*

2. *If $g : \mathcal{C}^j \to \mathbb{R}$ is differentiable at x for some $x \in \mathcal{C}^j$. Then, $g^\uparrow : \mathcal{C} \to \mathbb{R}$ is differentiable at every $\mu \in \mathcal{C}$ satisfying $\mathrm{p}_j\mu = x$ and its differential is given by $\nabla g^\uparrow(\mu) = \mathrm{l}_j(\nabla_j g(x))$.*

Proof. Part (1) Recall that by definition, $g^j = g \circ l_j$. For every $y \in \mathcal{C}^j$, we can see that

$$
\begin{aligned}
g^j(y) - g^j(x) &= g \circ \mathrm{l}_j(y) - g \circ \mathrm{l}_j(x) \\
&= \langle \nabla g(\mathrm{l}_j x), \mathrm{l}_j y - \mathrm{l}_j x \rangle_{\mathcal{H}} + o\left(|\mathrm{l}_j y - \mathrm{l}_j x|_{\mathcal{H}}\right), \\
&= \langle \mathrm{p}_j(\nabla g(\mathrm{l}_j x)), y - x \rangle_{\mathcal{H}^j} + o\left(|y - x|_{\mathcal{H}^j}\right),
\end{aligned}
$$

where the last equality follows from Lemma 6.3.3 (1) and (2).

Part (2). Recall that by definition, $g^{\uparrow} = g \circ \mathrm{p}_j$. Let $\mu \in \mathcal{C}$ satisfy $\mathrm{p}_j \mu = x$. Then for any $\nu \in \mathcal{C}$, we get

$$
\begin{aligned}
g^{\uparrow}(\nu) - g^{\uparrow}(\mu) &= g \circ \mathrm{p}_j(\nu) - g \circ \mathrm{p}_j(\mu), \\
&= \langle \nabla_j g(x), \mathrm{p}_j \nu - x \rangle_{\mathcal{H}^j} + o\left(|\mathrm{p}_j \nu - x|_{\mathcal{H}^j}\right), \\
&= \langle \mathrm{l}_j(\nabla_j g(x)), \nu - \mu \rangle_{\mathcal{H}} + o\left(|\nu - \mu|_{\mathcal{H}}\right),
\end{aligned}
$$

where we used Lemma 6.3.3 (1) and (5). □

6.3.2. Comparison principle

To compensate for the lack of compactness in infinite dimensions, we need Stegall's variational principle [110, Theorem on page 174] (see also [32, Theorem 8.8]).

Theorem 6.3.7 (Stegall's variational principle)**.** *Let $\mathcal{E}$ be a convex and weakly compact set in a separable Hilbert space $\mathcal{X}$ and $g : \mathcal{E} \to \mathbb{R}$ be an upper semi-continuous function bounded from above. Then, for every $\delta > 0$, there is $\iota \in \mathcal{X}$ satisfying $|\iota|_{\mathcal{X}} \leqslant \delta$ such that $g + \langle \iota, \cdot \rangle_{\mathcal{X}}$ achieves maximum on $\mathcal{E}$.*

Originally, $\mathcal{E}$ is only required to satisfy the Radon-Nikodym property which is weaker than being convex and weakly compact (see discussion on [110, page 173]).

The goal of this subsection to prove the following.

Proposition 6.3.8 (Comparison principle)**.** *Under assumption* (A1)*, let u be a Lipschitz viscosity subsolution and v be a Lipschitz viscosity supersolution of* $\mathrm{HJ}(\mathcal{H}, \mathcal{C}, \mathsf{H})$*. If* $u(0,\cdot) \leqslant v(0,\cdot)$*, then* $u \leqslant v$*.*

Proof of Proposition 6.3.8. It suffices to show $u(t,\cdot) - v(t,\cdot) \leqslant 0$ for all $t \in [0,T)$ for any $T > 0$. Henceforth, we fix any $T > 0$. We set $L = \|u\|_{\mathrm{Lip}} \vee \|v\|_{\mathrm{Lip}}$, $M = 2L + 3$ and V to be the Lipschitz coefficient of H restricted to the centered ball with radius $2L + M + 3$. We proceed in steps.

Step 1. Let $\theta : \mathbb{R} \to \mathbb{R}_+$ be a nondecreasing smooth function satisfying

$$|\theta'| \leqslant 1 \qquad \text{and} \qquad (r-1)_+ \leqslant \theta(r) \leqslant r_+, \quad \forall r \in \mathbb{R},$$

where θ' is the derivative of θ. For $R > 1$ to be determined, we define

$$\Phi(t,\mu) = M\theta\left(\left(1 + |\mu|^2_{\mathcal{H}}\right)^{\frac{1}{2}} + Vt - R\right), \quad \forall (t,\mu) \in \mathbb{R}_+ \times \mathcal{C}.$$

It is immediate that

$$\sup_{(t,\mu)\in\mathbb{R}_+\times\mathcal{C}} |\nabla\Phi(t,\mu)|_{\mathcal{H}} \leqslant M, \tag{6.3.14}$$

$$\partial_t \Phi \geqslant V|\nabla\Phi|_{\mathcal{H}}, \tag{6.3.15}$$

$$\Phi(t,\mu) \geqslant M(|\mu|_{\mathcal{H}} - R - 1)_+, \quad \forall (t,\mu) \in \mathbb{R}_+ \times \mathcal{C}. \tag{6.3.16}$$

For $\varepsilon, \sigma \in (0,1)$ to be determined, we consider

$$\Psi(t,\mu,t',\mu') = u(t,\mu) - v(t',\mu') - \frac{1}{2\varepsilon}(|t-t'|^2 + |\mu-\mu'|^2_{\mathcal{H}}) - \Phi(t,\mu) - \sigma t - \frac{\sigma}{T-t},$$
$$\forall (t,\mu,t',\mu') \in [0,T) \times \mathcal{C} \times \mathbb{R}_+ \times \mathcal{C}.$$

Setting $C_0 = u(0,0) - v(0,0)$, and using (6.3.16) and the definition of L, we have

$$\Psi(t,\mu,t',\mu') \leqslant C_0 + L(2|t| + 2|\mu|_{\mathcal{H}} + |t-t'| + |\mu-\mu'|_{\mathcal{H}}) - \frac{1}{2\varepsilon}(|t-t'|^2 + |\mu-\mu'|^2_{\mathcal{H}}) \tag{6.3.17}$$

$$-M(|\mu|_{\mathcal{H}} - R - 1)_+ - \frac{\sigma}{T-t}.$$

Hence, by the definition of M, Ψ is bounded from above and its supremum is achieved over a bounded set. Invoking Theorem 6.3.7, for $\delta \in (0,1)$ to be chosen, there is $(\bar{s}, \bar{\iota}, \bar{s}', \bar{\iota}') \in \mathbb{R} \times \mathcal{H} \times \mathbb{R} \times \mathcal{H}$ satisfying

$$|\bar{s}|,\ |\bar{\iota}|_{\mathcal{H}},\ |\bar{s}'|,\ |\bar{\iota}'|_{\mathcal{H}} \leqslant \delta, \tag{6.3.18}$$

such that the function

$$\overline{\Psi}(t,\mu,t',\mu') = \Psi(t,\mu,t',\mu') - \bar{s}t - \langle \bar{\iota}, \mu\rangle_{\mathcal{H}} - \bar{s}'t' - \langle \bar{\iota}', \mu'\rangle_{\mathcal{H}},$$

for all $(t,\mu,t',\mu') \in [0,T) \times \mathcal{C} \times \mathbb{R}_+ \times \mathcal{C}$ achieves its maximum at $(\bar{t}, \bar{\mu}, \bar{t}', \bar{\mu}')$.

Step 2. We derive bounds on $|\bar{\mu}|_{\mathcal{H}}$, $|\bar{\mu} - \bar{\mu}'|_{\mathcal{H}}$ and $|\bar{t} - \bar{t}'|$. Using $\overline{\Psi}(0,0,0,0) \leqslant \overline{\Psi}(\bar{t}, \bar{\mu}, \bar{t}', \bar{\mu}')$, (6.3.17) and $\bar{t} \leqslant T$, we have

$$\begin{aligned}
C_0 &\leqslant \frac{\varepsilon}{T} + \Psi(\bar{t}, \bar{\mu}, \bar{t}', \bar{\mu}') + 2\delta|\bar{\mu}|_{\mathcal{H}} + 2T\delta + \delta|\bar{t} - \bar{t}'| + \delta|\bar{\mu} - \bar{\mu}'|_{\mathcal{H}} \\
&\leqslant \frac{\varepsilon}{T} + C_0 + 2LT + (2L|\bar{\mu}|_{\mathcal{H}} - M(|\bar{\mu}|_{\mathcal{H}} - R - 1)_+) + \left(L|\bar{t} - \bar{t}'| - \frac{1}{2\varepsilon}|\bar{t} - \bar{t}'|^2\right) \\
&\qquad + \left(L|\bar{\mu} - \bar{\mu}'|_{\mathcal{H}} - \frac{1}{2\varepsilon}|\bar{\mu} - \bar{\mu}'|^2_{\mathcal{H}}\right) + 2\delta|\bar{\mu}|_{\mathcal{H}} + 2T\delta + \delta|\bar{t} - \bar{t}'| + \delta|\bar{\mu} - \bar{\mu}'|_{\mathcal{H}} \\
&\leqslant (2(L+\delta)|\bar{\mu}|_{\mathcal{H}} - M(|\bar{\mu}|_{\mathcal{H}} - R - 1)_+) + \left(\frac{\varepsilon}{T} + C_0 + 2LT + \varepsilon(L+\delta)^2 + 2T\delta\right).
\end{aligned}$$

By this and the definition of M, there is $C_1 > 0$ such that, for all $\varepsilon, \delta \in (0,1)$ and all $R > 1$,

$$|\bar{\mu}|_{\mathcal{H}} \leqslant C_1 R. \tag{6.3.19}$$

Since

$$0 \geqslant \overline{\Psi}(\bar{t}, \overline{\mu}, \bar{t}', \overline{\mu}) - \overline{\Psi}(\bar{t}, \overline{\mu}, \bar{t}', \overline{\mu}') = v(\bar{t}', \overline{\mu}') - v(\bar{t}', \overline{\mu}) + \frac{1}{2\varepsilon}|\overline{\mu} - \overline{\mu}'|^2_{\mathcal{H}} + \langle \bar{\iota}', \overline{\mu}' - \overline{\mu} \rangle_{\mathcal{H}},$$

by the definition of L and (6.3.18), we can get

$$|\overline{\mu} - \overline{\mu}'|_{\mathcal{H}} \leqslant 2(L + \delta)\varepsilon. \tag{6.3.20}$$

Similarly, by

$$0 \geqslant \overline{\Psi}(\bar{t}, \overline{\mu}, \bar{t}, \overline{\mu}') - \overline{\Psi}(\bar{t}, \overline{\mu}, \bar{t}', \overline{\mu}') = v(\bar{t}', \overline{\mu}') - v(\bar{t}, \overline{\mu}') + \frac{1}{2\varepsilon}|\bar{t} - \bar{t}'|^2 + \bar{s}'(\bar{t}' - \bar{t}),$$

we have

$$|\bar{t} - \bar{t}'| \leqslant 2(L + \delta)\varepsilon. \tag{6.3.21}$$

Step 3. We show that for every $\sigma, \varepsilon \in (0, 1)$, every $R > 1$, and sufficiently small δ, we have either $\bar{t} = 0$ or $\bar{t}' = 0$ We argue by contradiction and assume that $\bar{t} > 0$ and $\bar{t}' > 0$. Since the function

$$(t, \mu) \mapsto \overline{\Psi}(t, \mu, \bar{t}', \overline{\mu}')$$

achieves its maximum at $(\bar{t}, \overline{\mu}) \in (0, T) \times \mathcal{C}$, by the assumption that u is a subsolution, we have

$$\frac{1}{\varepsilon}(\bar{t} - \bar{t}') + \partial_t \Phi(\bar{t}, \overline{\mu}) + \sigma + \sigma(T - \bar{t})^{-2} + \bar{s} - \mathsf{H}\left(\frac{1}{\varepsilon}(\overline{\mu} - \overline{\mu}') + \nabla\Phi(\bar{t}, \overline{\mu}) + \bar{\iota}\right) \leqslant 0. \tag{6.3.22}$$

Since the function

$$(t', \mu') \mapsto \overline{\Psi}(\bar{t}, \overline{\mu}, t', \mu')$$

achieves its maximum at $(\overline{t}', \overline{\mu}') \in (0, \infty) \times \mathcal{C}$, by the assumption that v is a supersolution, we have

$$\frac{1}{\varepsilon}(\overline{t} - \overline{t}') - \overline{s}' - \mathsf{H}\left(\frac{1}{\varepsilon}(\overline{\mu} - \overline{\mu}') - \overline{\iota}'\right) \geqslant 0. \tag{6.3.23}$$

By (6.3.14), (6.3.18) and (6.3.20), for $\varepsilon, \delta \in (0,1)$, we have

$$\left|\frac{1}{\varepsilon}(\overline{\mu} - \overline{\mu}') + \nabla\Phi(\overline{t}, \overline{\mu}) + \overline{\iota}\right|_{\mathcal{H}}, \ \left|\frac{1}{\varepsilon}(\overline{\mu} - \overline{\mu}') - \overline{\iota}'\right|_{\mathcal{H}} \leqslant 2L + M + 3.$$

Taking the difference of terms in (6.3.22) and (6.3.23), by the definition of L, (6.3.15) and (6.3.18), we obtain

$$\sigma \leqslant -\overline{s} - \overline{s}' + V|\nabla\Phi(\overline{t}, \overline{\mu})|_{\mathcal{H}} + V(|\overline{\iota}|_{\mathcal{H}} + |\overline{\iota}'|_{\mathcal{H}}) - \partial_t\Phi(\overline{t}, \overline{\mu}) \leqslant 2(1+V)\delta.$$

By making δ sufficiently small, we reach a contradiction, and thus we must have either $\overline{t} = 0$ or $\overline{t}' = 0$.

Step 4. We conclude our proof. Let us consider the case $\overline{t} = 0$. Fixing any $(t, \mu) \in [0, T) \times \mathcal{C}$, by $\overline{\Psi}(t, \mu, t, \mu) \leqslant \overline{\Psi}(\overline{t}, \overline{\mu}, \overline{t}', \overline{\mu}')$, we have

$$\Psi(t, \mu, t, \mu) \leqslant \Psi(\overline{t}, \overline{\mu}, \overline{t}', \overline{\mu}') + \delta(4T + 2|\mu|_{\mathcal{H}} + 2C_1 R + 2(L+\delta)\varepsilon)$$

where we used $t, \overline{t} < T$, (6.3.18), (6.3.19) and (6.3.20). Due to $u(0, \cdot) \leqslant v(0, \cdot)$ and $\overline{t} = 0$, using (6.3.20) and (6.3.21), we can see

$$\begin{aligned}\Psi(\overline{t}, \overline{\mu}, \overline{t}', \overline{\mu}') \leqslant u(0, \overline{\mu}) - v(\overline{t}', \overline{\mu}') \leqslant v(0, \overline{\mu}) - v(\overline{t}', \overline{\mu}') &\leqslant L|\overline{t} - \overline{t}'| + L|\overline{\mu} - \overline{\mu}'|_{\mathcal{H}} \\ &\leqslant 4L(L+\delta)\varepsilon.\end{aligned}$$

Combining the above two displays and recalling the definition of Ψ, we get

$$u(t,\mu) - v(t,\mu) \leqslant \Phi(t,\mu) + \sigma t + \frac{\sigma}{T-t} + 4L(L+\delta)\varepsilon + \delta(4T + 2|\mu|_{\mathcal{H}} + 2C_1 R + 2(L+\delta)\varepsilon).$$

First sending $\delta \to 0$, then $\varepsilon, \sigma \to 0$, and finally $R \to \infty$, by the above and the definition of Φ, we obtain $u(t,\mu) - v(t,\mu) \leqslant 0$ as desired. The case $\bar{t}' = 0$ is similar. □

6.3.3. Convergence of approximations

Let us denote by $\mathrm{HJ}(\mathcal{H}, \mathcal{C}, \mathsf{H})$ the Hamilton–Jacobi equation (5.2.12) on cone $\mathcal{C}$ with nonlinearity H, and by $\mathrm{HJ}(\mathcal{H}, \mathcal{C}, \mathsf{H}; \psi)$ the equation $\mathrm{HJ}(\mathcal{H}, \mathcal{C}, \mathsf{H})$ with initial condition $\psi : \mathcal{C} \to \mathbb{R}$.

The following lemma shows that the lift of a finite-dimensional viscosity solution is a viscosity solution in infinite-dimensions.

Proposition 6.3.9 (Limit of approximations is a solution). *Suppose that* H *is continuous. For each* $j \in \mathfrak{J}_{\mathsf{gen}}$*, let* f_j *be a viscosity subsolution (respectively, supersolution) of* $\mathrm{HJ}(\mathcal{H}^j, \mathcal{C}^j, \mathsf{H}^j)$. *If* $f = \lim_{j \in \mathfrak{J}_{\mathsf{gen}}} f_j^{\uparrow}$ *in the local uniform topology, then* f *is a viscosity subsolution (respectively, supersolution) of* $\mathrm{HJ}(\mathcal{H}, \mathcal{C}, \mathsf{H})$.

Proof. Suppose that $\{f_j\}_{j \in \mathfrak{J}_{\mathsf{gen}}}$ is a collection of viscosity subsolutions. Let us assume that $f - \phi$ achieves a local maximum at $(t,\mu) \in (0,\infty) \times \mathcal{C}$ for some smooth function ϕ. We define

$$\widetilde{\phi}(s,\nu) = \phi(s,\nu) + |s-t|^2 + |\nu - \mu|^2_{\mathcal{H}}, \quad \forall (s,\nu) \in \mathbb{R}_+ \times \mathcal{C}.$$

Then, there is some $R > 0$ such that

$$f(s,\nu) - \widetilde{\phi}(s,v) = f(t,\mu) - \widetilde{\phi}(t,\mu) - |(s,\nu) - (t,\mu)|^2_{\mathbb{R}\times\mathcal{H}}, \quad \forall (s,\nu) \in B \tag{6.3.24}$$

where

$$B = \{(s,\nu) \in (0,\infty) \times \mathcal{C} : |(s,\nu) - (t,\mu)|_{\mathbb{R}\times\mathcal{H}} \leqslant 2R\}.$$

Note that $f - \widetilde{\phi}$ achieves a local maximum at (t,μ) and that the derivatives of $\widetilde{\phi}$ coincide with those of ϕ at (t,μ). For lighter notation, we replace ϕ by $\widetilde{\phi}$ henceforth. It is also clear from Definition 6.1.3 (2) that ϕ is locally Lipschitz. Hence, there is $L > 0$ such that

$$|\phi(s,\nu) - \phi(s',\nu')| \leqslant L|(s,\nu) - (s',\nu')|_{\mathbb{R}\times\mathcal{H}}, \quad \forall (s,\nu),\, (s',\nu') \in B. \tag{6.3.25}$$

For each $j \in \mathfrak{J}_{\mathsf{gen}}$, we set

$$B_j = \{(s,y) \in (0,\infty) \times \mathcal{C}^j : |(s,y) - (t, \mathrm{p}_j\mu)|_{\mathbb{R}\times\mathcal{H}^j} \leqslant R\}.$$

By making $2R < |t|$ sufficiently small, we can ensure that both B and B_j are closed. Let $(t_j, x_j) \in B_j$ be the point at which $f_j - \phi^j$ achieves the maximum over B_j. Here, ϕ^j is the j-projection of ϕ given in Definition 6.3.1.

For any $\delta \in (0,1)$, we choose $j' \in \mathfrak{J}_{\mathsf{gen}}$ such that, for all $j \in \mathfrak{J}_{\mathsf{gen}}$ satisfying $j \supset j'$,

$$\sup_B \left| f_j^{\uparrow} - f \right| < \frac{\delta^2}{4}, \tag{6.3.26}$$

$$\left| \mu - \mu^{(j)} \right|_{\mathcal{H}} < R \wedge \frac{\delta^2}{4L}. \tag{6.3.27}$$

We claim that, for all $j \in \mathfrak{J}_{\mathsf{gen}}$ satisfying $j \supset j'$,

$$|(t_j, \mathrm{l}_j x_j) - (t,\mu)|_{\mathbb{R}\times\mathcal{H}} < \delta. \tag{6.3.28}$$

We argue by contradiction and suppose that there is $j \supset j'$ such that

$$|(t_j, \mathrm{l}_j x_j) - (t,\mu)|_{\mathbb{R}\times\mathcal{H}} \geqslant \delta. \tag{6.3.29}$$

Before proceeding, we note that

$$|(t_j, \mathrm{l}_j x_j) - (t,\mu)|_{\mathbb{R}\times\mathcal{H}} \leqslant \left|(t_j, \mathrm{l}_j x_j) - \left(t, \mu^{(j)}\right)\right|_{\mathbb{R}\times\mathcal{H}} + \left|\mu - \mu^{(j)}\right|_{\mathcal{H}} \leqslant 2R \tag{6.3.30}$$

where in the last inequality we used (6.3.27), and the fact that $(t_j, x_j) \in B_j$ together with Lemma 6.3.3 (2) and (4). Then, we have

$$\begin{aligned}
f_j(t_j, x_j) - \phi^j(t_j, x_j) &= f_j^{\uparrow}(t_j, \mathrm{l}_j x_j) - \phi(t_j, \mathrm{l}_j x_j) \\
&\leqslant f(t_j, \mathrm{l}_j x_j) - \phi(t_j, \mathrm{l}_j x_j) + \frac{\delta^2}{4} \\
&\leqslant f(t,\mu) - \phi(t,\mu) - \frac{3\delta^2}{4} \\
&\leqslant f_j^{\uparrow}(t,\mu) - \phi(t,\mu) - \frac{\delta^2}{2} \\
&\leqslant f_j^{\uparrow}\left(t, \mu^{(j)}\right) - \phi\left(t, \mu^{(j)}\right) - \frac{\delta^2}{4} \\
&= f_j\left(t, \mathrm{p}_j\mu\right) - \phi^j\left(t, \mathrm{p}_j\mu\right) - \frac{\delta^2}{4}
\end{aligned}$$

where the first and the last equalities follow from the definitions of lifts and projections of functions in Definition 6.3.1 together with Lemma 6.3.3 (3) and (4); the first and third inequalities follow from (6.3.26) and the fact that $(t_j, \mathrm{l}_j x_j) \in B$ due to (6.3.30); the second inequality follows from (6.3.29) and (6.3.24); the fourth inequality follows from the observation that $f_j^{\uparrow}(t,\mu) = f_j^{\uparrow}(t,\mu^{(j)})$ due to the definition of lifts of functions and Lemma 6.3.3 (4), and (6.3.25) along with (6.3.27). The relation in the above display contradicts the fact the maximality of $f_j - \phi_j$ over B_j at (t_j, x_j). Hence, by contradiction, we must have (6.3.28) and thus

$$\lim_{j\in\mathfrak{J}_{\mathsf{gen}}} (t_j, \mathrm{l}_j x_j) = (t,\mu) \quad \text{in } (0,\infty)\times\mathcal{C}. \tag{6.3.31}$$

Using (6.3.31) and Lemma 6.3.3 (3) and (5), we also have that $\lim_{j\in\mathfrak{J}_{\mathsf{gen}}} |(t_j, x_j) - (t, \mathrm{p}_j\mu)|_{\mathbb{R}\times\mathcal{H}^j} = 0$. Hence, we deduce that, for sufficiently fine $j \in \mathfrak{J}_{\mathsf{gen}}$, (t_j, x_j) lies in the interior of B_j relative to $(0,\infty)\times\mathcal{C}^j$. Since f_j is a viscosity subsolution, we get that

$$\left(\partial_t\phi^j - \mathsf{H}^j\left(\nabla_j\phi^j\right)\right)(t_j, x_j) \leqslant 0. \tag{6.3.32}$$

Using the definition of projections of functions, Lemma 6.3.6 (1), and Lemma 6.3.3 (4), we have that

$$\partial_t\phi^j(t, x_j) = \partial_t\phi(t, \mathrm{l}_j x_j), \qquad \nabla_j\phi^j(t_j, x_j) = \mathrm{p}_j\left(\nabla\phi(t_j, \mathrm{l}_j x_j)\right),$$
$$\mathsf{H}^j\left(\nabla_j\phi^j(t_j, x_j)\right) = \mathsf{H}\left(\left(\nabla\phi(t_j, \mathrm{l}_j x_j)\right)^{(j)}\right).$$

Then, using (6.3.31), the continuity of differentials (see Definition 6.1.3 (2)), and Lemma 6.3.3 (8), we can pass (6.3.32) to the limit to obtain that

$$\left(\partial_t\phi - \mathsf{H}\left(\nabla\phi\right)\right)(t, \mu) \leqslant 0.$$

Hence, we have verified that f is a viscosity subsolution. The same argument also works for viscosity supersolutions. □

Recall that $\mathfrak{J}_{\mathsf{unif}}$ is the collection of uniform partitions of $[0, 1)$, which is generating in the sense given in Section 6.3.1.

Proposition 6.3.10 (Convergence of approximations)**.** *In addition to* (A1)–(A2)*, suppose that* $\psi : \mathcal{C} \to \mathbb{R}$ *satisfies*

$$|\psi(\mu) - \psi(\nu)| \leqslant C|\mu - \nu|_{L^p}, \tag{6.3.33}$$

for some $C > 0$ *and* $p \in [1, 2)$*. For every* $j \in \mathfrak{J}_{\mathsf{good}}$*, let* f_j *be a viscosity solution of* $\mathrm{HJ}(\mathcal{H}^j, \mathcal{C}^j, \mathsf{H}^j; \psi^j)$*. Then,* $(f_j^{\uparrow})_{j\in\mathfrak{J}_{\mathsf{good}}}$ *converges in the local uniform topology to a Lipschitz* *function* $f : \mathbb{R}_+ \times \mathcal{C} \to \mathbb{R}$ *satisfying* $f(0, \cdot) = \psi$,

$$\sup_{t \in \mathbb{R}_+} \|f(t, \cdot)\|_{\mathrm{Lip}} \leqslant \|\psi\|_{\mathrm{Lip}}, \tag{6.3.34}$$

$$\sup_{\mu \in \mathcal{C}} \|f(\cdot, \mu)\|_{\mathrm{Lip}} \leqslant \sup_{\substack{\iota \in \mathcal{H} \\ |\iota| \leqslant \|\psi\|_{\mathrm{Lip}}}} |\mathsf{H}(\iota)|. \tag{6.3.35}$$

To prove this result, we follow closely the proof of [96, Proposition 3.7]. We need the following lemma.

Lemma 6.3.11. *Under assumption* (A2)*, let* $j, j' \in \mathfrak{J}$ *satisfy* $j \subseteq j'$*. If* f_j *is a viscosity subsolution (respectively, supersolution) of* $\mathrm{HJ}(\mathcal{H}^j, \mathcal{C}^j, \mathsf{H}^j)$*, then the function defined by*

$$f_{j \to j'}(t, x) = f_j(t, \mathrm{p}_j \mathrm{l}_{j'} x), \quad \forall (t, x) \in \mathbb{R}_+ \times (\mathrm{p}_j \mathrm{l}_{j'})^{-1}(\mathcal{C}^j) \tag{6.3.36}$$

is a viscosity subsolution (respectively, supersolution) of $\mathrm{HJ}(\mathcal{H}^{j'}, (\mathrm{p}_j \mathrm{l}_{j'})^{-1}(\mathcal{C}^j), \mathsf{H}^{j'})$.

Proof. Setting $\widetilde{\mathcal{C}} = (\mathrm{p}_j \mathrm{l}_{j'})^{-1}(\mathcal{C}^j)$ for convenience, we suppose that $f_{j \to j'} - \phi$ has a local maximum at $(t, x) \in (0, \infty) \times \widetilde{\mathcal{C}}$ for some smooth function ϕ. We define

$$\phi_j(s, y) = \phi(s, x + \mathrm{p}_{j'} \mathrm{l}_j y - \mathrm{p}_{j'} \mathrm{l}_j \mathrm{p}_j \mathrm{l}_{j'} x), \quad \forall (s, y) \in \mathbb{R}_+ \times \mathcal{C}^j.$$

Using Lemma 6.3.3 (6) and (3), we can show

$$\mathrm{p}_j \mathrm{l}_{j'} \left(x + \mathrm{p}_{j'} \mathrm{l}_j y - \mathrm{p}_{j'} \mathrm{l}_j \mathrm{p}_j \mathrm{l}_{j'} x \right) = \mathrm{p}_j \mathrm{l}_j y = y \in \mathcal{C}^j \tag{6.3.37}$$

for every $y \in \mathcal{C}^j$, which implies that $x + \mathrm{p}_{j'} \mathrm{l}_j y - \mathrm{p}_{j'} \mathrm{l}_j \mathrm{p}_j \mathrm{l}_{j'} x \in \widetilde{\mathcal{C}}$ for every $y \in \mathcal{C}^j$. Setting $\overline{y} = \mathrm{p}_j \mathrm{l}_{j'} x$, we want to show that $f_j - \phi_j$ achieves a local maximum at $(t, \overline{y})$. Let us fix some $r > 0$ sufficiently small such that

$$\sup_{B_{j'}} (f_{j \to j'} - \phi) = f_{j \to j'}(t, x) - \phi(t, x) \tag{6.3.38}$$

where

$$B_{j'} = \left\{(s,z) \in (0,\infty) \times \widetilde{\mathcal{C}} : |s-t| + |z-x|_{\mathcal{H}^{j'}} \leqslant r\right\}.$$

Then, we set $B_j = \{(s,y) \in (0,\infty) \times \mathcal{C}^j : |s-t| + |y-\overline{y}|_{\mathcal{H}^j} \leqslant r\}$. Using Lemma 6.3.3 (2) and (5), we have that

$$\left|\mathrm{p}_{j'}\mathrm{l}_j y - \mathrm{p}_{j'}\mathrm{l}_j \overline{y}\right|_{\mathcal{H}^{j'}} \leqslant |y-\overline{y}|_{\mathcal{H}^j}, \quad \forall y \in \mathcal{C}^j,$$

which along with (6.3.37) implies that

$$\left(s, x + \mathrm{p}_{j'}\mathrm{l}_j y - \mathrm{p}_{j'}\mathrm{l}_j \overline{y}\right) \in B_{j'}, \quad \forall (s,y) \in B_j.$$

Using (6.3.37), the definition of $f_{j\to j'}$ in (6.3.36), and the definition of ϕ_j, we also have that for all $(s,y) \in \mathbb{R}_+ \times \mathcal{C}^j$,

$$f_j(s,y) - \phi_j(s,y) = f_{j\to j'}(s, x + \mathrm{p}_{j'}\mathrm{l}_j y - \mathrm{p}_{j'}\mathrm{l}_j \overline{y}) - \phi(s, x + \mathrm{p}_{j'}\mathrm{l}_j y - \mathrm{p}_{j'}\mathrm{l}_j \overline{y}).$$

Using this, the previous display, and (6.3.38), we obtain that

$$\sup_{B_j}(f_j - \phi_j) \leqslant \sup_{B_{j'}}(f_{j\to j'} - \phi) = f_{j\to j'}(t,x) - \phi(t,x) = f_j(t,\overline{y}) - \phi_j(t,\overline{y}),$$

which implies that $f_j - \phi_j$ achieves a local maximum at $(t,\overline{y})$.

Since f_j is a viscosity subsolution, we have

$$\left(\partial_t \phi_j - \mathsf{H}^j(\nabla_j \phi_j)\right)(t,\overline{y}) \leqslant 0.$$

Using the definition of ϕ_j, we can compute that, for any $h \in \mathcal{H}^j$ sufficiently small,

$$
\begin{aligned}
\langle h, \nabla_j \phi_j(s,y)\rangle_{\mathcal{H}^j} + o\left(|h|_{\mathcal{H}^j}\right) &= \phi_j(s, y+h) - \phi_j(s,y) \\
&= \left\langle \mathrm{p}_{j'}\mathrm{l}_j h,\, \nabla_{j'}\phi(\cdots)\right\rangle_{\mathcal{H}^{j'}} + o\left(|\mathrm{p}_{j'}\mathrm{l}_j h|_{\mathcal{H}^{j'}}\right) \\
&= \left\langle h,\, \mathrm{p}_j \mathrm{l}_{j'} \nabla_{j'}\phi(\cdots)\right\rangle_{\mathcal{H}^j} + o\left(|h|_{\mathcal{H}^j}\right), \quad \forall (s,y) \in \mathcal{C}^j,
\end{aligned}
$$

where in $(\cdots)$ we omitted $(s, x + \mathrm{p}_{j'}\mathrm{l}_j y - \mathrm{p}_{j'}\mathrm{l}_j\mathrm{p}_j\mathrm{l}_{j'}x)$, and, in the last equality, we used Lemma 6.3.3 (1) and (6) to get the term in the bracket and Lemma 6.3.3 (2) and (5) for the error term. The above display implies that $\nabla_j\phi_j(t,\overline{y}) = \mathrm{p}_j\mathrm{l}_{j'}\nabla_{j'}\phi(t,x)$. It is easy to see $\partial_t\phi_j(t,\overline{y}) = \partial_t\phi(t,x)$. These along with the previous display and the definition of H^j yield

$$
\left(\partial_t\phi - \mathsf{H}(\mathrm{l}_j\mathrm{p}_j\mathrm{l}_{j'}\nabla_{j'}\phi)\right)(t,x) \leqslant 0.
$$

We claim that

$$
\mathrm{l}_{j'}\nabla_{j'}\phi(t,x) - \mathrm{l}_j\mathrm{p}_j\mathrm{l}_{j'}\nabla_{j'}\phi(t,x) \in \mathcal{C}^*. \tag{6.3.39}
$$

Since H is $\mathcal{C}^*$-nondecreasing, recalling that $\mathsf{H}^{j'} = \mathsf{H}(\mathrm{l}_{j'}(\cdot))$, we deduce from (6.3.39) and the previous display that

$$
\left(\partial_t\phi - \mathsf{H}^{j'}(\nabla_{j'}\phi)\right)(t,x) \leqslant 0,
$$

verifying that $f_{j\to j'}$ is a viscosity subsolution of $\mathrm{HJ}(\mathcal{H}^{j'}, \widetilde{\mathcal{C}}, \mathsf{H}^{j'})$.

To prove (6.3.39), by the duality of cones, it suffices to show that

$$
\left\langle \iota,\ \mathrm{l}_{j'}\nabla_{j'}\phi(t,x) - \mathrm{l}_j\mathrm{p}_j\mathrm{l}_{j'}\nabla_{j'}\phi(t,x)\right\rangle_{\mathcal{H}} \geqslant 0, \quad \forall \iota \in \mathcal{C}.
$$

By Lemma 6.3.3 (1), the above is equivalent to

$$\left\langle \mathrm{p}_{j'}\iota - \mathrm{p}_{j'}\mathrm{l}_j\mathrm{p}_j\iota,\ \nabla_{j'}\phi(t,x)\right\rangle_{\mathcal{H}^{j'}} \geqslant 0, \quad \forall \iota \in \mathcal{C}. \tag{6.3.40}$$

Fix any $\iota \in \mathcal{C}$. Lemma 6.3.3 (6) yields

$$\mathrm{p}_j\mathrm{l}_{j'}\left(\mathrm{p}_{j'}\iota - \mathrm{p}_{j'}\mathrm{l}_j\mathrm{p}_j\iota\right) = \mathrm{p}_j\iota - \mathrm{p}_j\mathrm{l}_j\mathrm{p}_j\iota = 0. \tag{6.3.41}$$

Hence, setting $z = \mathrm{p}_{j'}\iota - \mathrm{p}_{j'}\mathrm{l}_j\mathrm{p}_j\iota$, we have $z \in \widetilde{\mathcal{C}}$, and thus $\varepsilon z + x \in \widetilde{\mathcal{C}}$ for any $\varepsilon > 0$. Since $f_{j\to j'} - \phi$ has a local maximum at (t,x), we can see that, for $\varepsilon > 0$ sufficiently small,

$$\begin{aligned}\left\langle \varepsilon z, \nabla_{j'}\phi(t,x)\right\rangle_{j'} + o(\varepsilon) = \phi(t, x+\varepsilon z) - \phi(t,x) &\geqslant f_{j\to j'}(t, x+\varepsilon z) - f_{j\to j'}(t,x) \\ &= f_j(t, \mathrm{p}_j\mathrm{l}_{j'}x + \varepsilon \mathrm{p}_j\mathrm{l}_{j'}z) - f_j(t, \mathrm{p}_j\mathrm{l}_{j'}x) = 0\end{aligned}$$

where the last equality follows from (6.3.41) and the definition of z. Sending $\varepsilon \to 0$, we can verify (6.3.40) and complete the proof for subsolutions. The argument for supersolutions is the same with inequalities reversed. □

Proof of Proposition 6.3.10. Throughout this proof, we denote by C an absolute constant, which may vary from instance to instance. Let $j, j' \subseteq \mathfrak{J}_{\mathsf{good}}$ satisfy $j \subseteq j'$, and f_j, $f_{j'}$ be viscosity solutions to $\mathrm{HJ}(\mathcal{H}^j, \mathcal{C}^j, \mathsf{H}^j; \psi^j)$, $\mathrm{HJ}(\mathcal{H}^{j'}, \mathcal{C}^{j'}, \mathsf{H}^{j'}; \psi^{j'})$, respectively. We define $f_{j\to j'}$ by (6.3.36). By Lemma 6.3.11, $f_{j\to j'}$ is a viscosity solution of $\mathrm{HJ}(\mathcal{H}^{j'},\ (\mathrm{p}_j\mathrm{l}_{j'})^{-1}(\mathcal{C}^j),\ \mathsf{H}^{j'};\ \psi^j(\mathrm{p}_j\mathrm{l}_{j'}(\cdot)))$. By Lemma 6.3.5 (3) and (1), we have

$$\mathcal{C}^{j'} \subseteq (\mathrm{p}_j\mathrm{l}_{j'})^{-1}(\mathcal{C}^j). \tag{6.3.42}$$

We claim that there is $C > 0$ such that

$$|f_{j\to j'}(t,x) - f_{j'}(t,x)| \leqslant C|j|^{-\frac{2-p}{2p}}\left(t + |x|_{\mathcal{H}^{j'}}\right), \quad \forall (t,x) \in \mathbb{R}_+ \times \mathcal{C}^{j'}. \tag{6.3.43}$$

Let us use this to derive the desired results. For $\mu \in \mathcal{C}$, we set $x = \mathrm{p}_{j'}\mu$. Lemma 6.3.3 (6) implies that $\mathrm{p}_j \mathrm{l}_{j'} x = \mathrm{p}_j \mu$. Hence, by definitions, we have

$$f_j^{\uparrow}(t, \mu) = f_j(t, \mathrm{p}_j \mu) = f_{j \to j'}(t, x)$$

and $f_{j'}^{\uparrow}(t, \mu) = f_{j'}(t, x)$. Now using (6.3.43) and Lemma 6.3.3 (5), we have

$$\left| f_j^{\uparrow}(t, \mu) - f_{j'}^{\uparrow}(t, \mu) \right| \leqslant C|j|^{-\frac{2-p}{2p}} \left(t + |\mu|_{\mathcal{H}}\right).$$

We could now conclude the existence of a limit $f(t, \mu)$ by arguing that the above together with the triangle inequality yields that $(f_j^{\uparrow}(t, \mu))_{j \in \mathfrak{J}_{\mathsf{good}}}$ is a *Cauchy net* in $\mathbb{R}$ (see [89, Definition 2.1.41] and [89, Proposition 2.1.49]). Denoting the pointwise limit by f, and passing j' to limit in the above display to see that $f_j^{\uparrow}$ converges in the local uniform topology to some $f : \mathbb{R}_+ \times \mathcal{C}$. By Lemma 6.3.3 (7), it is straightforward to see $f(0, \cdot) = \psi$.

Then, we show (6.3.34) and (6.3.35). By (6.3.33) and Hölder's inequality, we have $\|\psi\|_{\mathrm{Lip}} < C$. Proposition 6.2.3 implies that, for every j,

$$\sup_{t \in \mathbb{R}_+} \|f_j(t, \cdot)\|_{\mathrm{Lip}} = \|\psi^j\|_{\mathrm{Lip}}, \qquad \sup_{x \in \mathcal{C}^j} \|f_j(\cdot, x)\|_{\mathrm{Lip}} \leqslant \sup_{\substack{p \in \mathcal{H}^j \\ |p|_{\mathcal{H}^j} \leqslant \|\psi^j\|_{\mathrm{Lip}}}} |\mathsf{H}^j(p)|. \tag{6.3.44}$$

By the definition of ψ^j and Lemma 6.3.3 (2), we can see that, for every $x, y \in \mathcal{C}^j$,

$$|\psi^j(x) - \psi^j(y)| = |\psi(\mathrm{l}_j x) - \psi(\mathrm{l}_j y)| \leqslant \|\psi\|_{\mathrm{Lip}} |\mathrm{l}_j x - \mathrm{l}_j y|_{\mathcal{H}} = \|\psi\|_{\mathrm{Lip}} |x - y|_{\mathcal{H}^j},$$

which implies that

$$\|\psi^j\|_{\mathrm{Lip}} \leqslant \|\psi\|_{\mathrm{Lip}}, \quad \forall j \in \mathfrak{J}. \tag{6.3.45}$$

Using this, the first result in (6.3.44) and Lemma 6.3.3 (5), we have, for every $t \in \mathbb{R}_+$ and every $\mu, \nu \in \mathcal{C}$,

$$|f_j^{\uparrow}(t,\mu) - f_j^{\uparrow}(t,\nu)| = |f_j(t, \mathrm{p}_j\mu) - f_j(t, \mathrm{p}_j\nu)| \leqslant \|\psi^j\|_{\mathrm{Lip}} |\mathrm{p}_j\mu - \mathrm{p}_j\nu|_{\mathcal{H}^j} \leqslant \|\psi\|_{\mathrm{Lip}} |\mu - \nu|_{\mathcal{H}},$$

yielding (6.3.34) after passing j to the limit. To see (6.3.35), for every p satisfying the condition under supremum in the second result in (6.3.44), we have, by Lemma 6.3.3 (2), that

$$|\mathrm{l}_j p|_{\mathcal{H}} = |p|_{\mathcal{H}^j} \leqslant \|\psi^j\|_{\mathrm{Lip}} \leqslant \|\psi\|_{\mathrm{Lip}}.$$

Since $\mathsf{H}^j(p) = \mathsf{H}(\mathrm{l}_j p)$ by definition, the right-hand side of the second result in (6.3.44) is thus bounded by the right-hand side of (6.3.35). Passing j to the limit, we can verify (6.3.35).

It remains to prove (6.3.43). Due to (6.3.45), Proposition 6.2.3 yields

$$\sup_{t\in\mathbb{R}_+} \|f_j(t,\cdot)\|_{\mathrm{Lip}}, \quad \sup_{t\in\mathbb{R}_+} \|f_{j'}(t,\cdot)\|_{\mathrm{Lip}} \leqslant \|\psi\|_{\mathrm{Lip}}. \tag{6.3.46}$$

The definition of $f_{j\to j'}$ in (6.3.36) implies

$$\begin{aligned}|f_{j\to j'}(t,x) - f_{j\to j'}(t,y)| &= |f_j(t, \mathrm{p}_j\mathrm{l}_{j'}x) - f_j(t, \mathrm{p}_j\mathrm{l}_{j'}y)| \\ &\leqslant \|\psi\|_{\mathrm{Lip}} \left|\mathrm{p}_j\mathrm{l}_{j'}x - \mathrm{p}_j\mathrm{l}_{j'}y\right|_{\mathcal{H}^j} \leqslant \|\psi\|_{\mathrm{Lip}} |x-y|_{\mathcal{H}^{j'}}, \quad \forall t \geqslant 0,\ \forall x,y \in (\mathrm{p}_j\mathrm{l}_{j'})^{-1}(\mathcal{C}^j),\end{aligned}$$

where we used Lemma 6.3.3 (2) and (5) to derive the last inequality. Hence, we have

$$\sup_{\mathbb{R}_+} \|f_{j\to j'}(t,\cdot)\|_{\mathrm{Lip}} \leqslant \|\psi\|_{\mathrm{Lip}}. \tag{6.3.47}$$

Using (6.3.42) and Proposition 6.2.2 with M replaced by $2\|\psi\|_{\mathrm{Lip}} + 1$ and $R > 1$ to be determined, we have that

$$\begin{aligned}\sup_{(t,x)\in\mathbb{R}_+\times\mathcal{C}^{j'}} & f_{j\to j'}(t,x) - f_{j'}(t,x) - M(|x|_{\mathcal{H}^{j'}} + Vt - R)_+ \\ &= \sup_{x\in\mathcal{C}^{j'}} f_{j\to j'}(0,x) - f_{j'}(0,x) - M(|x|_{\mathcal{H}^{j'}} - R)_+. \end{aligned} \tag{6.3.48}$$

The term inside the supremum on right-hand side of (6.3.48) can be rewritten as

$$\psi\left((\mathrm{l}_{j'}x)^{(j)}\right) - \psi(\mathrm{l}_{j'}x) - M(|x|_{\mathcal{H}^{j'}} - R)_+,$$

where we used the definition of $f_{j\to j'}$ in (6.3.36) and Lemma 6.3.3 (4). By (6.3.33) and Hölder's inequality, we have

$$\begin{aligned}\left|\psi\left((\mathrm{l}_{j'}x)^{(j)}\right) - \psi(\mathrm{l}_{j'}x)\right| &\leqslant C\left|(\mathrm{l}_{j'}x)^{(j)} - \mathrm{l}_{j'}x\right|_{L^1}^{\frac{2-p}{p}}\left|(\mathrm{l}_{j'}x)^{(j)} - \mathrm{l}_{j'}x\right|_{\mathcal{H}}^{\frac{2p-2}{p}} \\ &\leqslant C\left|(\mathrm{l}_{j'}x)^{(j)} - \mathrm{l}_{j'}x\right|_{L^1}^{\frac{2-p}{p}}|x|_{\mathcal{H}^{j'}}^{\frac{2p-2}{p}}\end{aligned} \tag{6.3.49}$$

where we also used Lemma 6.3.3 (2) and (5) in the last inequality. Setting $J = |j|$ and $J' = |j'|$, due to $j' \supset j$ and $j, j' \in \mathfrak{J}_{\mathsf{good}} \subseteq \mathfrak{J}_{\mathsf{unif}}$, we know that there is $N \in \mathbb{N}$ such that $J' = JN$. Before estimating the L^1 norm, we remark that it suffices to assume $D = 1$, namely, $\mathrm{l}_{j'}x(s) \in \mathbb{R}_+$ for each $s \in [0,1)$. Indeed, if $D > 1$, we can use reduce the problem to the real-valued case by considering

$$s \mapsto I_D \cdot \mathrm{l}_{j'}x(s)$$

where I_D is the $D\times D$ identity matrix. This reduction is valid due to $C_K^{-1}I_D\cdot a \leqslant |a| \leqslant C_D I_D \cdot a$ for every $a \in \mathbb{S}^D_+$ and some constant $C_D > 0$. With this simplification clarified, we assume $D = 1$. Writing $j' = (t_1, t_2, \ldots, t_{J'})$ with $t_k = \frac{k}{J'}$ and $j = (s_1, \ldots, s_J)$ with $s_m = \frac{m}{J}$, we can compute that

$$\left|1_{j'}x - (1_{j'}x)^{(j)}\right|_{L^1} = \sum_{m=1}^{J} \sum_{k:s_{m-1}<t_k\leqslant s_m} (t_k - t_{k-1})\cdot \tag{6.3.50}$$

$$\left|x_k - \frac{1}{s_m - s_{m-1}} \sum_{k':s_{m-1}<t_{k'}\leqslant s_m} (t_{k'} - t_{k'-1})x_{k'}\right|$$

$$= \sum_{m=1}^{J} \sum_{k=N(m-1)+1}^{Nm} \frac{1}{JN} \left|x_k - \frac{1}{N} \sum_{k'=N(m-1)+1}^{Nm} x_{k'}\right|$$

$$\leqslant \frac{1}{JN^2} \sum_{m=1}^{J} \sum_{k=N(m-1)+1}^{Nm} \sum_{k'=N(m-1)+1}^{Nm} |x_k - x_{k'}|$$

$$= \frac{2}{JN^2} \sum_{m=1}^{J} \sum_{k,k':N(m-1)<k'<k\leqslant Nm} |x_k - x_{k'}|. \tag{6.3.51}$$

Let $B > 0$ be chosen later. Since $x_k \geqslant x_{k'} \geqslant 0$ for $k > k'$ due to $x \in \mathcal{C}^{j'}$, we have

$$\frac{2}{JN^2} \sum_{m=1}^{J} \sum_{k,k':N(m-1)<k'<k\leqslant Nm} |x_k - x_{k'}|\mathbb{1}_{|x_k|\geqslant B} \leqslant \frac{2}{JN^2} \sum_{m=1}^{J} \sum_{k,k':N(m-1)<k'<k\leqslant Nm} |x_k|\mathbb{1}_{|x_k|\geqslant B}$$

$$\leqslant \frac{2}{JN} \sum_{m=1}^{J} \sum_{k=N(m-1)+1}^{Nm} \frac{|x_k|^2}{B} = \frac{2}{B} \sum_{k=1}^{J'} \frac{1}{J'} |x_k|^2 = \frac{2}{B} |x|^2_{\mathcal{H}^{j'}}. \tag{6.3.52}$$

One the other hand, switching summations, we have

$$\frac{2}{JN^2} \sum_{m=1}^{J} \sum_{k,k':N(m-1)<k'<k\leqslant Nm} |x_k - x_{k'}|\mathbb{1}_{|x_k|\leqslant B}$$

$$= \frac{2}{JN^2} \sum_{r,r':0<r'<r\leqslant N} \sum_{m=1}^{J} |x_{N(m-1)+r} - x_{N(m-1)+r'}|\mathbb{1}_{|x_{N(m-1)+r}|\leqslant B}$$

Again using $x_k \geqslant x_{k'} \geqslant 0$ for $k > k'$ and setting $m^* = \max\{m \in \{1, \ldots, J\} : x_{N(m-1)+r} \leqslant$

$B\}$, we can see that

$$\sum_{m=1}^{J} |x_{N(m-1)+r} - x_{N(m-1)+r'}| \mathbb{1}_{|x_{N(m-1)+r}| \leqslant B} = \sum_{m=1}^{m^*} (x_{N(m-1)+r} - x_{N(m-1)+r'}) \mathbb{1}_{x_{N(m-1)+r} \leqslant B}$$
$$\leqslant x_{N(m^*-1)+r} \mathbb{1}_{x_{N(m*-1)+r} \leqslant B} \leqslant B.$$

Here in the penultimate inequality, we also used the fact that $-x_{N(m-1)+r'} + x_{N(m-2)+r} \leqslant 0$ because $N(m-1)+r' > N(m-2)+r$ due to $|r-r'| < N$. Therefore,

$$\frac{2}{JN^2} \sum_{m=1}^{J} \sum_{k,k':N(m-1)<k'<k\leqslant Nm} |x_k - x_{k'}| \mathbb{1}_{|x_k| \leqslant B} \leqslant \frac{B}{J}.$$

Inserting into (6.3.51) the above estimate combined with (6.3.52), and choosing $B = \sqrt{J}|x|_{\mathcal{H}^{j'}}$, we conclude that

$$\left| (l_{j'} x)^{(j)} - l_{j'} x \right|_{L^1} \leqslant 3J^{-\frac{1}{2}} |x|_{\mathcal{H}^{j'}}.$$

Plugging this into (6.3.49) yields

$$f_{j \to j'}(0,x) - f_{j'}(0,x) - M(|x|_{\mathcal{H}^{j'}} - R)_+ \leqslant CJ^{-\frac{2-p}{2p}} |x|_{\mathcal{H}^{j'}}, \quad \forall x \in \mathcal{C}^{j'}.$$

Due to $f_{j\to j'}(0,0) = f_{j'}(0,0) = \psi(0)$, (6.3.46), and (6.3.47), the choice of $M = \|\psi\|_{\mathrm{Lip}} + 1$ ensures that

$$f_{j \to j'}(0,x) - f_{j'}(0,x) - M(|x|_{\mathcal{H}^{j'}} - R)_+ \leqslant 2\|\psi\|_{\mathrm{Lip}} |x|_{\mathcal{H}^{j'}} - M|x|_{\mathcal{H}^{j'}} + MR$$
$$= MR - |x|_{\mathcal{H}^{j'}}, \quad \forall x \in \mathcal{C}^{j'}.$$

These two estimates implies that the left-hand side of them is bounded by $CJ^{-\frac{2-p}{2p}} MR$. Absorbing M into C and using (6.3.48), we arrive at

$$\sup_{(t,x)\in \mathbb{R}_+ \times \mathcal{C}^{j'}} f_{j\to j'}(t,x) - f_{j'}(t,x) - M(|x|_{\mathcal{H}^{j'}} + Vt - R)_+ \leqslant CJ^{-\frac{2-p}{2p}} R.$$

Replacing R by $|x|_{\mathcal{H}^{j'}} + Vt$ for each $(t,x) \in \mathbb{R}_+ \times \mathcal{C}^{j'}$, we obtain one bound for (6.3.43).

For the opposite bound, we again use (6.3.42) and Proposition 6.2.2 to get a result as in (6.3.48) with $f_{j\to j'}$ and $f_{j'}$ swapped. Then, the same arguments as above give the other bound to complete the proof of (6.3.43). □

6.3.4. Variational formulae

Recall the definition of convex conjugates in (6.1.6). For $j \in \mathfrak{J}$, $(\mathsf{H}^j)^\circledast$ is defined with respect to $\mathcal{H}^j$ as the Hilbert space; $\mathsf{H}^\circledast$ is defined with respect to $\mathcal{H}$.

The proposition below shows that the limit of finite-dimensional Hopf-Lax formulae is the infinite-dimensional Hopf-Lax formula. Then we will prove the counterpart for the Hopf formula.

Proposition 6.3.12 (Hopf-Lax formula in the limit)**.** *In addition to* (A2)*, suppose*

- $\psi : \mathcal{C} \to \mathbb{R}$ *is* $\mathcal{C}^*$*-nondecreasing and continuous;*
- $\mathsf{H} : \mathcal{H} \to \mathbb{R}$ *is lower semicontinuous, convex, and satisfies* $\mathsf{H}(\iota^{(j)}) \leqslant \mathsf{H}(\iota)$ *for every* $\iota \in \mathsf{H}$ *and every* $j \in \mathfrak{J}_{\mathsf{gen}}$*;*
- *for each* $j \in \mathfrak{J}_{\mathsf{gen}}$*,* $f_j : \mathbb{R}_+ \times \mathcal{C}^j \to (-\infty, \infty]$ *is given by*

$$f_j(t,x) = \sup_{y\in\mathcal{C}^j} \left\{ \psi^j(y) - t(\mathsf{H}^j)^\circledast\left(\frac{y-x}{t}\right) \right\}, \quad \forall (t,x) \in \mathbb{R}_+ \times \mathcal{C}^j. \tag{6.3.53}$$

If $\lim_{j\in\mathfrak{J}_{\mathsf{gen}}} f_j^{\uparrow}(t,\mu)$ *exists in* $\mathbb{R}$ *at some* $(t,\mu) \in \mathbb{R}_+ \times \mathcal{C}$*, then the limit is given by*

$$f(t,\mu) = \sup_{\nu\in\mathcal{C}} \left\{ \psi(\nu) - t\mathsf{H}^\circledast\left(\frac{\nu-\mu}{t}\right) \right\}.$$

Proof. The assumption on H allows us apply Lemma 6.2.5 (1) to H^j to see that

$(\mathsf{H}^j)^\circledast(\frac{y-\mathrm{p}_j\mu}{t}) = \infty$ if $y - \mathrm{p}_j\mu \notin \mathcal{C}^j$. Therefore, we have

$$\begin{aligned} f_j^\uparrow(t,\mu) &= \sup_{y\in \mathrm{p}_j\mu+\mathcal{C}^j}\left\{\psi^j(y) - t(\mathsf{H}^j)^\circledast\left(\frac{y-\mathrm{p}_j\mu}{t}\right)\right\} \\ &= \sup_{y\in \mathrm{p}_j\mu+\mathcal{C}^j}\inf_{z\in\mathcal{H}^j}\left\{\psi^j(y) + \langle z, \mathrm{p}_j\mu - y\rangle_{\mathcal{H}^j} + t\mathsf{H}^j(z)\right\} \\ &= \sup_{y\in \mathrm{p}_j\mu+\mathcal{C}^j}\inf_{\iota\in\mathcal{H}}\left\{\psi(\mathrm{l}_j y) + \langle \mathrm{p}_j\iota, \mathrm{p}_j\mu - y\rangle_{\mathcal{H}^j} + t\mathsf{H}^j(\mathrm{p}_j\iota)\right\} \qquad (6.3.54) \\ &= \sup_{y\in \mathrm{p}_j\mu+\mathcal{C}^j}\inf_{\iota\in\mathcal{H}}\left\{\psi(\mathrm{l}_j y) + \left\langle \iota, \mu^{(j)} - \mathrm{l}_j y\right\rangle_{\mathcal{H}^j} + t\mathsf{H}\left(\iota^{(j)}\right)\right\}, \end{aligned}$$

where in the penultimate equality, we used the easy fact that $\mathrm{p}_j\mathcal{H} = \mathcal{H}^j$; in the last equality, we used and Lemma 6.3.3 (2) and (4). Using the assumption that $\mathsf{H}(\iota^{(j)}) \leqslant \mathsf{H}(\iota)$ for all $\iota \in \mathcal{H}$, we have

$$f_j^\uparrow(t,\mu) \leqslant \sup_{y\in \mathrm{p}_j\mu+\mathcal{C}^j}\left\{\psi(\mathrm{l}_j y) - t\mathsf{H}^\circledast\left(\frac{\mathrm{l}_j y - \mu^{(j)}}{t}\right)\right\}$$

For $y \in \mathrm{p}_j\mu + \mathcal{C}^j$, we have $\mathrm{l}_j y - \mu^{(j)} \in \mathcal{C}$ by Lemma 6.3.5 (1) and Lemma 6.3.3 (4). Meanwhile, Lemma 6.3.5 (5) yields $\mu - \mu^{(j)} \in \mathcal{C}^*$. Since ψ is $\mathcal{C}^*$-nondecreasing, we obtain

$$\psi(\mathrm{l}_j y) \leqslant \psi\left(\mathrm{l}_j y - \mu^{(j)} + \mu\right), \quad \forall y \in \mathrm{p}_j\mu + \mathcal{C}^j.$$

Using this, the previous display, we have

$$f_j^\uparrow(t,\mu) \leqslant \sup_{y\in \mathrm{p}_j\mu+\mathcal{C}^j}\left\{\psi(\mathrm{l}_j y - \mu^{(j)} + \mu) - t\mathsf{H}^\circledast\left(\frac{(\mathrm{l}_j - \mu^{(j)} + \mu) - \mu}{t}\right)\right\} \leqslant f(t,\mu).$$

Passing j to the limit, we conclude that $\lim_{j\in\mathfrak{J}_{\mathsf{gen}}} f_j^\uparrow(t,\mu) \leqslant f(t,\mu)$.

For the other direction, fixing any $\varepsilon > 0$, we can find ν to satisfy

$$f(t,\mu) \leqslant \varepsilon + \psi(\nu) - t\mathsf{H}^\circledast\left(\frac{\nu-\mu}{t}\right) = \varepsilon + \psi(\nu) + \inf_{\iota\in\mathcal{H}}\left\{\langle \iota, \mu-\nu\rangle + t\mathsf{H}(\iota)\right\}.$$

Since ψ is continuous, by Lemma 6.3.3 (7), we can find $j' \in \mathfrak{J}_{\mathsf{gen}}$ such that $\psi(\nu) \leqslant \psi(\nu^{(j)}) + \varepsilon$

for all $j \supset j'$. Then, we have

$$\begin{aligned}
f(t,\mu) &\leqslant 2\varepsilon + \psi\left(\nu^{(j)}\right) + \inf_{\iota\in\mathcal{H}}\left\{\langle \iota, \mu-\nu\rangle_{\mathcal{H}} + t\mathsf{H}(\iota)\right\}\\
&\leqslant 2\varepsilon + \psi\left(\nu^{(j)}\right) + \inf_{\iota\in\mathcal{H}}\left\{\left\langle \iota^{(j)}, \mu-\nu\right\rangle_{\mathcal{H}} + t\mathsf{H}\left(\iota^{(j)}\right)\right\}\\
&= 2\varepsilon + \psi(\mathrm{l}_j(\mathrm{p}_j\nu)) + \inf_{\iota\in\mathcal{H}}\left\{\langle \mathrm{p}_j\iota, \mathrm{p}_j\mu - \mathrm{p}_j\nu\rangle_{\mathcal{H}^j} + t\mathsf{H}^j(\mathrm{p}_j\iota)\right\}\\
&\leqslant 2\varepsilon + f_j^{\uparrow}(t,\mu), \qquad \forall j\supset j'
\end{aligned}$$

where the second inequality follows from $\{\iota^{(j)} : \iota\in\mathcal{H}\}\subseteq\mathcal{H}$; on the third line we used Lemma 6.3.3 (2) and (4); the last line follows from (6.3.54). Passing j to the limit and then sending $\varepsilon\to 0$, we obtain the converse bound, which completes the proof. □

Proposition 6.3.13 (Hopf formula in the limit)**.** *Suppose*

- $\psi:\mathcal{C}\to\mathbb{R}$ *is* $\mathcal{C}^*$*-nondecreasing;*
- $\mathsf{H}:\mathcal{H}\to\mathbb{R}$ *is continuous;*
- *for each* $j\in\mathfrak{J}_{\mathsf{gen}}$*,* $f_j:\mathbb{R}_+\times\mathcal{C}^j\to(-\infty,\infty]$ *is given by*

$$f_j(t,x) = \sup_{z\in\mathcal{C}^j}\inf_{y\in\mathcal{C}^j}\left\{\langle z, x-y\rangle_{\mathcal{H}^j} + \psi^j(y) + t\mathsf{H}^j(z)\right\}, \quad \forall (t,x)\in\mathbb{R}_+\times\mathcal{C}^j. \quad (6.3.55)$$

If $\lim_{j\in\mathfrak{J}_{\mathsf{gen}}} f_j^{\uparrow}(t,\mu)$ *exists in* $\mathbb{R}$ *at some* $(t,\mu)\in\mathbb{R}_+\times\mathcal{C}$*, then the limit is given by*

$$f(t,\mu) = \sup_{\nu\in\mathcal{C}}\inf_{\rho\in\mathcal{C}}\left\{\langle \nu, \mu-\rho\rangle + \psi(\rho) + t\mathsf{H}(\nu)\right\}.$$

Proof. We start by finding a formula for $f_j^{\uparrow}(t,\mu)$. Using (6.3.55), the definitions of lifts and projections of functions, and Lemma 6.3.5 (3), we can get

$$f_j^{\uparrow}(t,\mu) = \sup_{\nu\in\mathcal{C}}\inf_{\rho\in\mathcal{C}}\left\{\langle \mathrm{p}_j\nu, \mathrm{p}_j\mu - \mathrm{p}_j\rho\rangle_{\mathcal{H}^j} + \psi(\mathrm{l}_j\mathrm{p}_j\rho) + t\mathsf{H}(\mathrm{l}_j\mathrm{p}_j\nu)\right\}.$$

Then, by Lemma 6.3.3 (2) and (4), the above becomes

$$f_j^{\uparrow}(t,\mu) = \sup_{\nu\in\mathcal{C}} \inf_{\rho\in\mathcal{C}} \left\{ \left\langle \nu^{(j)}, \mu^{(j)} - \rho^{(j)} \right\rangle_{\mathcal{H}} + \psi\left(\rho^{(j)}\right) + t\mathsf{H}\left(\nu^{(j)}\right) \right\}.$$

Fix any (t,μ). For $\varepsilon > 0$, we choose ν such that

$$\begin{aligned} f(t,\mu) &\leqslant \varepsilon + \inf_{\rho\in\mathcal{C}} \left\{ \langle \nu, \mu-\rho\rangle_{\mathcal{H}} + \psi(\rho) + t\mathsf{H}(\nu) \right\} \\ &\leqslant \varepsilon + \inf_{\rho\in\mathcal{C}} \left\{ \left\langle \nu, \mu - \rho^{(j)} \right\rangle_{\mathcal{H}} + \psi(\rho^{(j)}) + t\mathsf{H}(\nu) \right\} \end{aligned}$$

for all $j \in \mathfrak{J}$, where the last inequality follows from the fact that $\{\rho^{(j)} : \rho \in \mathcal{C}\} \subseteq \mathcal{C}$. Allowed by the continuity of H and Lemma 6.3.3 (7), we can find $j' \in \mathfrak{J}$ such that for all $j \supset j'$,

$$\langle \nu, \mu\rangle_{\mathcal{H}} + t\mathsf{H}(\nu) \leqslant \varepsilon + \left\langle \nu^{(j)}, \mu \right\rangle_{\mathcal{H}} + t\mathsf{H}\left(\nu^{(j)}\right).$$

By Lemma 6.3.3 (1), (2) and (4), we can see that $\left\langle \iota, \kappa^{(j)}\right\rangle_{\mathcal{H}} = \left\langle \iota^{(j)}, \kappa^{(j)}\right\rangle_{\mathcal{H}}$ for all $\iota, \kappa \in \mathcal{H}$, which along with the above two displays implies that

$$f(t,\mu) \leqslant 2\varepsilon + \inf_{\rho\in\mathcal{C}^j} \left\{ \left\langle \nu^{(j)}, \mu^{(j)} - \rho^{(j)} \right\rangle_{\mathcal{H}} + \psi\left(\rho^{(j)}\right) + t\mathsf{H}\left(\nu^{(j)}\right) \right\} \leqslant 2\varepsilon + f_j^{\uparrow}(t,\mu), \quad \forall j \supset j'.$$

Passing j to the limit and sending $\varepsilon \to 0$, we obtain $f(t,\mu) \leqslant \lim_{j\in\mathfrak{J}_{\mathsf{gen}}} f_j^{\uparrow}(t,\mu)$.

To see the converse inequality, fixing any $\varepsilon > 0$, we choose ν_j, for each $j \in \mathfrak{J}$, to satisfy

$$f_j^{\uparrow}(t,\mu) \leqslant \varepsilon + \inf_{\rho\in\mathcal{C}} \left\{ \left\langle \nu_j^{(j)}, \mu^{(j)} - \rho^{(j)} \right\rangle_{\mathcal{H}} + \psi\left(\rho^{(j)}\right) + t\mathsf{H}\left(\nu_j^{(j)}\right) \right\}, \quad \forall j \in \mathfrak{J}.$$

On the other hand, it is clear from the definition of $f(t,\mu)$ that

$$f(t,\mu) \geqslant \inf_{\rho\in\mathcal{C}} \left\{ \left\langle \nu_j^{(j)}, \mu - \rho \right\rangle_{\mathcal{H}} + \psi(\rho) + t\mathsf{H}\left(\nu_j^{(j)}\right) \right\}, \quad \forall j \in \mathfrak{J}.$$

By Lemma 6.3.5 (5), we have $\rho - \rho^{(j)} \in \mathcal{C}^*$. Since ψ is $\mathcal{C}^*$-nondecreasing, we obtain $\psi(\rho) \geqslant$

$\psi(\rho^{(j)})$. This along with the fact that $\left\langle \nu_j^{(j)}, \mu - \rho \right\rangle_{\mathcal{H}} = \left\langle \nu_j^{(j)}, \mu^{(j)} - \rho^{(j)} \right\rangle_{\mathcal{H}}$ yields

$$f(t,\mu) \geqslant \inf_{\rho\in\mathcal{C}} \left\{ \left\langle \nu_j^{(j)}, \mu^{(j)} - \rho^{(j)} \right\rangle_{\mathcal{H}} + \psi\left(\rho^{(j)}\right) + t\mathsf{H}\left(\nu_j^{(j)}\right) \right\} \geqslant f_j^{\uparrow}(t,\mu) - \varepsilon, \quad \forall j \in \mathfrak{J}.$$

Passing j to the limit along $\mathfrak{J}_{\mathsf{gen}}$ and sending $\varepsilon \to 0$, we get $f(t,\mu) \geqslant \lim_{j\in\mathfrak{J}_{\mathsf{gen}}} f_j^{\uparrow}(t,\mu)$, completing the proof. □

6.3.5. Weak boundary

It can be checked that $\mathcal{C}$ has empty interior in $\mathcal{H}$. Therefore, the boundary of $\mathcal{C}$ is equal to $\mathcal{C}$. On the other hand, for each $j \in \mathfrak{J}$, the interior of $\mathcal{C}^j$ is not empty. We denote its boundary by $\partial\mathcal{C}^j$.

Lemma 6.3.14 (Characterizations of $\partial\mathcal{C}^j$)**.** *Let $j \in \mathfrak{J}$ and $x \in \mathcal{C}^j$. Then, the following are equivalent:*

1. *$x \in \partial\mathcal{C}^j$;*
2. *there is $y \in (\mathcal{C}^j)^* \setminus \{0\}$ such that $\langle x, y\rangle_{\mathcal{H}^j} = 0$;*
3. *there is $k \in \{1, 2, \ldots, |j|\}$ such that $x_k = x_{k-1}$.*

For (3), recall our convention that $x_0 = 0$.

Proof. First, we show that (3) implies (2). Let I_D be the $D\times D$ identity and matrix. If $k > 1$, we set $y_k = \frac{1}{t_k - t_{k-1}} I_D$, $y_{k-1} = -\frac{1}{t_{k-1}-t_{k-2}} I_D$ and $y_i = 0$ for all $i \in \{1, 2, \ldots, |j|\} \setminus \{k-1, k\}$. If $k = 1$, we set $y_1 = I_D$ and $y_i = 0$ otherwise. By Lemma 6.3.4 (1), we have $y \in (\mathcal{C}^j)^*$. It is also clear that $y \neq 0$ and $\langle x, y\rangle_{\mathcal{H}^j} = 0$, verifying (2).

Next, we show that (2) implies (1). Assuming (2), we suppose that x is in the interior. Then, there is $\varepsilon > 0$ sufficiently small such that $x - \varepsilon y \in \mathcal{C}^j$, which implies that $\langle x - \varepsilon y, y\rangle_{\mathcal{H}^j} \geqslant 0$. However, by assumption (2), we must have $-\varepsilon|y|^2_{\mathcal{H}^j} \geqslant 0$ and thus $y = 0$, reaching a contradiction.

Finally, we show that (1) implies (3). Assuming (1), we suppose that (3) is not true. Since the coordinates of x are increasing, we can find $\delta > 0$ such that $x_k \geqslant \delta I_D + x_{k-1}$ for all k. By the finite-dimensionality, there is a constant $C > 0$ such that

$$y_k - C\varepsilon I_D \leqslant x_k \leqslant y_k + C\varepsilon I_D$$

for every $y \in \mathcal{H}^j$ satisfying $|y - x|_{\mathcal{H}^j} \leqslant \varepsilon$, for every $\varepsilon > 0$ and every $k \in \{1, , 2, \ldots, |j|\}$. Choosing ε sufficiently small, we can see that, for such y, we have $y_k \geqslant y_{k-1}$ for all k, namely $y \in \mathcal{C}^j$, which contradicts (1). □

The equivalence between (1) and (2) actually holds for more general cones in finite dimensions. It is thus natural to define a weak notion of boundary for $\mathcal{C}$.

Definition 6.3.15. The weak boundary of $\mathcal{C}$ denoted by $\partial_{\mathsf{w}}\mathcal{C}$ is defined by

$$\partial_{\mathsf{w}}\mathcal{C} = \{\mu \in \mathcal{C} : \exists \iota \in \mathcal{C}^* \setminus \{0\},\ \langle \mu, \iota \rangle_{\mathcal{H}} = 0\}.$$

When $D = 1$, for every $\mu \in \mathcal{C}$, since μ is nondecreasing, we have that μ is differentiable a.e. and we denote its derivative by $\dot{\mu}$. If $D > 1$, we can choose a basis for $\mathbb{S}^D$ consisting of elements in $\mathbb{S}^D_+$. For each a from the basis, the derivative of $s \mapsto a \cdot \mu(s)$ exists a.e. We can use these to define $\dot{\mu}$. We define the essential support of an $\mathbb{S}^D$-valued function on $[0, 1)$ as the smallest closed set relative to $[0, 1)$, outside which the function is zero a.e.

Lemma 6.3.16 (Characterization of $\partial_{\mathsf{w}}\mathcal{C}$)**.** *For $\mu \in \mathcal{C}$, it holds that $\mu \in \partial_{\mathsf{w}}\mathcal{C}$ if and only if the essential support of $\dot{\mu}$ is not $[0, 1)$.*

Proof. Let $\mu \in \mathcal{C}$. By adding a constant, we may assume $\mu(0) = 0$. For any fixed $\iota \in \mathcal{C}^*$, we set $\kappa : [0, 1) \to \mathbb{R}$ by $\kappa(t) = \int_t^1 \iota(s)\mathrm{d}s$. Then, κ is continuous, nonnegative (by Lemma 6.3.4 (2)), and differentiable with its derivative is given by $-\iota$. Since $\mu(0) = 0$ and

$\lim_{t\to 1}\kappa(t)=0$, by integration by parts, we have that

$$\langle \mu, \iota\rangle_{\mathcal{H}} = \int_0^1 \kappa(s)\dot{\mu}(s)ds.$$

First, suppose that the essential support of $\dot{\mu}$ is $[0,1)$. Note that for any nonzero ι, we must have κ is not identically zero. Then the integral above is positive, and thus $\mu \notin \partial_{\mathsf{w}}\mathcal{C}$. For the other direction, suppose that the essential support of $\dot{\mu}$ is a strict subset of $[0,1)$. This implies the existence of a nonempty open set $O \subseteq [0,1)$ on which $\dot{\mu}$ vanishes. We then choose a nonnegative and smooth κ such that $\kappa > 0$ only on a subset of O. Setting $\iota = -\dot{\kappa}$, we clearly have $\iota \in \mathcal{C}^* \setminus \{0\}$. In this case, the integral in the above display is zero, implying $\mu \in \partial_{\mathsf{w}}\mathcal{C}$. □

It is thus attempting to use $\partial_{\mathsf{w}}\mathcal{C}$ as a more suitable notion of boundary in this case, and to impose some boundary condition on $\partial_{\mathsf{w}}\mathcal{C}$. However, the argument could still be more involved than the one we presented in this section, due to the following result as an immediate consequence of Lemma 6.3.16.

Lemma 6.3.17. *For $j \in \mathfrak{J}$, then $\mathrm{l}_j x \in \partial_{\mathsf{w}}\mathcal{C}$ for every $x \in \mathcal{C}^j$; and $\mu^{(j)} \in \partial_{\mathsf{w}}\mathcal{C}$ for every $\mu \in \mathcal{C}$.*

In other words, any point from $\mathcal{C}^j$ is lifted to the boundary of $\mathcal{C}$, no matter it is in the interior of $\mathcal{C}^j$ or not. The following lemma could potentially be a remedy.

Lemma 6.3.18. *If $x \in \mathcal{C}^j \setminus \partial\mathcal{C}^j$, then there is $\mu \in \mathcal{C} \setminus \partial_{\mathsf{w}}\mathcal{C}$ such that $\mathrm{p}_j\mu = x$.*

Proof. By the equivalence between (1) and (3) in Lemma 6.3.14, we can find $\delta > 0$ such that $x_k - x_{k-1} \geqslant \delta I_D$ for all k, where I_D is the $D \times D$ identity matrix. Then, we define $\mu : [0,1) \to \mathbb{S}^D$ by

$$\mu(s) = \varepsilon I_D\left(s - \frac{t_k + t_{k-1}}{2}\right) + x_k, \quad \text{if } s \in [t_{k-1}, t_k),$$

for $\varepsilon > 0$. It is straightforward to check that $\mathrm{p}_j\mu = x$. By choosing $\varepsilon > 0$ sufficiently small, we can ensure that μ is strictly increasing on $[0, 1)$. Hence, Lemma 6.3.16 implies that $\mu \in \mathcal{C} \setminus \partial_{\mathsf{w}}\mathcal{C}$. □

6.4. Application to the spin glass setting

6.4.1. Setting and definitions

We first introduce the definition of monotone probability measures. After explaining the setting of mean-field spin glass models in detail, we give the definition of viscosity solutions of (6.1.1).

Monotone probability measures

We have already mentioned the set $\mathcal{P}^{\nearrow}$ of monotone probability measures on $\mathbb{S}^D_+$ in Section 6.1. Its definition is as follows. A probability measure μ on $\mathbb{S}^D_+$ is said to be *monotone*, if

$$\mathbb{P}\left\{a \cdot X < a \cdot X' \text{ and } b \cdot X > b \cdot X'\right\} = 0, \quad \forall a, b \in \mathbb{S}^D_+, \tag{6.4.1}$$

where X and X' are two independent $\mathbb{S}^D_+$-valued random variable with the same law μ.

For $p \in [1, \infty)$, denote by $\mathcal{P}^{\nearrow}_p$ the restriction of $\mathcal{P}^{\nearrow}$ to those probability measures with finite p-th moments. We equip $\mathcal{P}^{\nearrow}_p$ with the p-Wasserstein metric $\mathbf{d}_p$. Let us recall the definition of Wasserstein metrics. For probability measures ϱ, ϑ on some measure space $\mathcal{X}$ with finite p-th moments, the p-Wasserstein distance between ϱ, ϑ is given by

$$\mathbf{d}_p(\varrho, \vartheta) = \inf_{\pi \in \Pi(\varrho, \vartheta)} \left(\int |x - y|^p \pi(\mathrm{d}x, \mathrm{d}y)\right)^{\frac{1}{p}}, \tag{6.4.2}$$

where $\Pi(\varrho, \vartheta)$ is collection of all couplings of ϱ, ϑ, and a probability measure π on $\mathcal{X} \times \mathcal{X}$ is said to be a coupling of ϱ, ϑ if the first marginal of π is ϱ and its second marginal is ϑ.

We want to embed $\mathcal{P}^{\nearrow}_2$ isometrically onto the cone $\mathcal{C}$ given in (6.3.4) with the ambient Hilbert space $\mathcal{H}$ in (6.3.2). Throughout this section, let U be the random variable distributed uniformly over $[0,1)$. Guaranteed by [93, Propositions 2.4 and 2.5], we have that

$$\Xi : \begin{cases} \mathcal{C} \to \mathcal{P}_2^{\nearrow} \\ \mu \mapsto \mathsf{Law}(\mu(U)) \end{cases} \tag{6.4.3}$$

is an isometric bijection between $\mathcal{C}$ and $\mathcal{P}_2^{\nearrow}$. In fact, [93, Propositions 2.4 and 2.5] ensures a stronger result:

$$\mathbf{d}_p(\varrho, \vartheta) = |\Xi^{-1}(\varrho) - \Xi^{-1}(\vartheta)|_{\mathcal{H}} = \mathbb{E}|\Xi^{-1}(\varrho)(U) - \Xi^{-1}(\vartheta)(U)|^2, \quad \forall \varrho, \vartheta \in \mathcal{P}_2^{\nearrow}, \ \forall p \in [1,2]. \tag{6.4.4}$$

For $g : \mathcal{P}_2^{\nearrow} \to \mathbb{R}$ and $f : \mathbb{R}_+ \times \mathcal{P}_2^{\nearrow} \to \mathbb{R}$, the actions of Ξ on them are given by

$$\Xi g : \begin{cases} \mathcal{C} \to \mathbb{R} \\ \mu \mapsto g(\Xi(\mu)) \end{cases}, \qquad \Xi f : \begin{cases} \mathbb{R}_+ \times \mathcal{C} \to \mathbb{R} \\ (t, \mu) \mapsto f(t, \Xi(\mu)) \end{cases}. \tag{6.4.5}$$

Mean-field spin glass models

We following the setting in [93]. We consider a wide class of mean-field fully connected vector spin models. Recall that D is any positive integer. Let $\xi : \mathbb{R}^{D\times D} \to \mathbb{R}$ be locally Lipschitz. For each $N \in \mathbb{N}$, let $\mathcal{H}_N$ be a finite-dimensional Hilbert space, and $(H_N(\sigma))_{\sigma \in \mathcal{H}_N^D}$ be a centered Gaussian random field with covariance structure given by

$$\mathbb{E}[H_N(\sigma) H_N(\tau)] = N\xi\left(\frac{\sigma\tau^\intercal}{N}\right), \quad \forall \sigma, \tau \in \mathcal{H}_N^D, \tag{6.4.6}$$

where the $D \times D$ real-valued matrix $\sigma\tau^\intercal$ is given by

$$\sigma\tau^\intercal = \left(\langle \sigma_d, \tau_{d'} \rangle_{\mathcal{H}_N}\right)_{1\leqslant d, d' \leqslant D}.$$

Here, $H_N(\sigma)$ is the Hamiltonian of the configuration σ. For each $N \in \mathbb{N}$, let P_N be a probability measure on $\mathcal{H}_N^D$ and we assume that P_N is supported on the centered ball in $\mathcal{H}_N^D$ with radius $\sqrt{N}$, where the inner product on $\mathcal{H}_N^D$ is the standard one induced by Cartesian products, namely,

$$\langle \sigma, \tau \rangle_{\mathcal{H}_N^D} = \sum_{d=1}^{D} \langle \sigma_d, \tau_d \rangle_{\mathcal{H}_N}, \quad \forall \sigma, \tau \in \mathcal{H}_N^D.$$

Each σ is viewed as a configuration of spins in a system, and $H_N(\sigma)$ is the random Hamiltonian at σ. Spin configurations are distributed according to P_N.

As an example, the Sherrington–Kirkpatrick Model corresponds to $D = 1$, $\mathcal{H}_N = \mathbb{R}^N$, $\xi(r) = r^2$ and P_N is a uniform measure on $\{-1, +1\}^N$. For each $N \in \mathbb{N}$, under P_N, we can view σ as sampled uniformly from configurations of Ising spins, namely, combinations of N spins each at state -1 or $+1$. The Hamiltonian can be expressed as

$$H_N^{\mathrm{SK}}(\sigma) = \frac{1}{\sqrt{N}} \sum_{i,j=1}^{N} g_{ij} \sigma_i \sigma_j, \quad \forall \sigma \in \{-1, +1\}^N,$$

where $\{g_{ij}\}_{1 \leqslant i,j \leqslant N}$ is a collection of independent standard Gaussian random variables.

Back to the general setting, we are interested in the asymptotics, as $N \to \infty$, of the free energy

$$\frac{1}{N} \mathbb{E} \log \int \exp\left(H_N(\sigma)\right) \mathrm{d}P_N(\sigma).$$

For $t \geqslant 0$, we also set

$$F_N(t) = -\frac{1}{N} \log \int \exp\left(\sqrt{2t} H_N(\sigma) - Nt\xi\left(\frac{\sigma\sigma^\intercal}{N}\right)\right) \mathrm{d}P_N(\sigma).$$

Comparing with the previous display, the additional normalizing term $Nt\xi(\sigma\sigma^\intercal/N)$ is to ensure the exponential term has expectation equal to one. This additional term can be removed as explained in [98].

In a more involved way (see [93, Section 3]), we can enrich the spin glass model by introducing an additional magnetic field parametrized by $\varrho \in \mathcal{P}^{\nearrow}$ with finite support. Similarly, there is a quantity $F_N(t, \varrho)$ associated with enriched model satisfying $F_N(t, \delta_0) = F_N(t)$ for all $t \in \mathbb{R}$, where δ_0 is the Dirac measure at the zero matrix.

Interpretation of (6.1.1)

By the isometry between $\mathcal{P}_2^{\nearrow}$ and $\mathcal{C}$, we can make sense of ∂_μ as the Fréchet derivative in $\mathcal{H}$ discussed in Section 6.1.2. In particular, whenever exists, $\nabla \Xi f(t, \mu)$ is an L^2 function over $[0, 1)$. From another angle, note that, for every measurable $g : \mathbb{S}^D_+ \to \mathbb{R}^{D\times D}$, we have

$$\int \xi(g) \mathrm{d}\varrho = \mathbb{E}\left[\xi\left(g \circ \left(\Xi^{-1}(\varrho)\right)(U)\right)\right] = \int_0^1 \xi\left(g \circ \left(\Xi^{-1}(\varrho)\right)(s)\right) \mathrm{d}s.$$

Hence, the equation (6.1.1) can be viewed as

$$\partial_t f - \int_0^1 \xi\left(\nabla f\right) \mathrm{d}s = 0, \quad \text{on } \mathbb{R}_+ \times \mathcal{C}, \tag{6.4.7}$$

where ds denotes the Lebesgue measure.

For technical reasons, we want to consider a regularized version of ξ. Let us describe the regularization. Recall the definition of being nondecreasing along a cone in (6.1.5). A function $g : \mathbb{S}^D_+ \to \mathbb{R}$ is said to be *proper* if g is nondecreasing over $\mathbb{S}^D_+$, and for every $b \in \mathbb{S}^D_+$, the function $a \mapsto g(a + b) - g(a)$ is nondecreasing over $\mathbb{S}^D_+$.

Definition 6.4.1. A function $\overline{\xi} : \mathbb{S}^D_+ \to \mathbb{R}$ is said to be a *regularization* of $\xi : \mathbb{R}^{D\times D} \to \mathbb{R}$ in (6.4.6) if

1. $\overline{\xi}$ coincides with ξ on a subset of $\mathbb{S}^D_+$ consisting matrices with every entry in $[-1, 1]$;

2. $\overline{\xi}$ is Lipschitz and proper;

3. $\overline{\xi}$ is bounded below;

4. $\overline{\xi}$ is convex, if, in addition, ξ is convex.

Let us justify the condition (1). For simplicity, we write $\overline{F}_N(t,\varrho) = \mathbb{E}F_N(t,\varrho)$ for every N, t, ϱ. It is expected that $(t,\varrho) \mapsto \overline{F}_N(t,\varrho)$ converges as $N \to \infty$ to a solution f of (6.4.7). Due to our assumption on the support of P_N, it has been shown in [93, Proposition 3.1] that

$$\left|\overline{F}_N(t,\varrho) - \overline{F}_N(t,\vartheta)\right| \leqslant \mathbf{d}_1(\varrho,\vartheta) = \left|\Xi^{-1}(\varrho) - \Xi^{-1}(\vartheta)\right|_{L^1}, \quad \forall t \geqslant 0,\ \forall \varrho,\vartheta \in \mathcal{P}_2^{\nearrow},\ \forall N \in \mathbb{N}. \tag{6.4.8}$$

The above bound is first established for ϱ,ϑ with finite supports. Then $\overline{F}_N(t,\cdot)$ can be extended by density, and the above bound can be extended accordingly. The above bound implies that $|\nabla(\Xi\overline{F}_N)(t,\mu)|_{L^\infty} \leqslant 1$ for every N, t, μ. Passing to the limit, then same bound is expected to hold for Ξf, which means that only values of ξ on matrices with entries in $[-1,1]$ matter. In addition, by [93, Proposition 3.8], for every N and t,

$$\Xi\overline{F}_N(t,\cdot) \text{ is } \mathcal{C}^*\text{-nondecreasing} \tag{6.4.9}$$

which by the duality of cones implies that $\nabla(\Xi\overline{F}_N)(t,\mu) \in \mathcal{C}$ for every $\mu \in \mathcal{C}$. In particular, $\nabla(\Xi\overline{F}_N)(t,\mu(s)) \in \mathbb{S}^D_+$ for a.e. $s \in [0,1)$. Passing to the limit, we expect $\nabla(\Xi f)(t,\mu(s)) \in \mathbb{S}^D_+$ for a.e. $s \in [0,1)$ and every $\mu \in \mathcal{C}$. Hence, in view of (6.4.7), ξ can be further restricted to $\mathbb{S}^D_+$. Therefore, condition (1) can be justified.

Since ξ is the covariance function for Gaussian fields, there are many structures to exploit. Under the assumption that ξ admits a convergent power series expansion, [93, Propositions 6.4 and 6.6] yield that ξ is when restricted to $\mathbb{S}^D_+$. The following lemma guarantees the existence of $\overline{\xi}$.

Lemma 6.4.2. *If $\xi : \mathbb{R}^{D\times D} \to \mathbb{R}$ restricted to $\mathbb{S}^D_+$ is locally Lipschitz and proper, then ξ admits a regularization.*

Proof. We follow the construction in [93, Proposition 6.8]. There, the definition of regularizations only requires (1) and (2). Here, we will verify that $\bar{\xi}$ constructed is convex if ξ is so. For $r>0$, we set $B_{\mathrm{tr}}(r)=\{a\in\mathbb{S}^D_+ : \mathrm{tr}(a)\leqslant r\}$. Then, all $a\in\mathbb{S}^K_+$ with entries in $[-1,1]$ belong to $B_{\mathrm{tr}}(D)$. For every $a\in\mathbb{S}^D_+$, we denote by $|a|_\infty$ the largest eigenvalue of a. Setting $L=\||\nabla\xi|_\infty\|_{L^\infty(B_{\mathrm{tr}}(2D))}$, we define, for every $a\in\mathbb{S}^D_+$,

$$\bar{\xi}(a)=\begin{cases}\xi(a)\vee(\xi(0)+2L(\mathrm{tr}(a)-D)), & \text{if } a\in B_{\mathrm{tr}}(2D),\\ \xi(0)+2L(\mathrm{tr}(a)-D), & \text{if } a\notin B_{\mathrm{tr}}(2D).\end{cases}$$

Since ξ is proper, we have $\xi(a)\geqslant\xi(0)\geqslant\xi(0)+2L(\mathrm{tr}(a)-D)$ for all $a\in B_{\mathrm{tr}}(D)$. Hence, $\bar{\xi}$ coincides with ξ on matrices with entries in $[-1,1]$, verifying (1). Note that $\bar{\xi}$ is continuous on $\{a\in\mathbb{S}^D_+ : \mathrm{tr}(a)=2D\}$. Then, it is easy to check that $\bar{\xi}$ is Lipschitz. Due to the choice of L, we can also see that the gradient of $\bar{\xi}$ is nondecreasing and thus $\bar{\xi}$ is proper, verifying (2). It is clear from the construction that (3) holds.

Now, assuming that ξ is convex, we show that $\bar{\xi}$ is also convex. If $a,b\in B_{\mathrm{tr}}(2D)$ or $a,b\notin B_{\mathrm{tr}}(2D)$, it is easy to check that

$$\bar{\xi}(\lambda a+(1-\lambda)b)\leqslant\lambda\bar{\xi}(a)+(1-\lambda)\bar{\xi}(b),\quad\forall\lambda\in[0,1]. \tag{6.4.10}$$

Then, we consider $a\in B_{\mathrm{tr}}(2D)$ and $b\notin B_{\mathrm{tr}}(2D)$. If λ satisfies $\lambda a+(1-\lambda)b\notin B_{\mathrm{tr}}(2D)$, then (6.4.10) holds. Now, let λ be such that $\lambda a+(1-\lambda)b\in B_{\mathrm{tr}}(2D)$. There is $\gamma\in[0,\lambda]$ such that $c=\gamma a+(1-\gamma)b$ satisfies $\mathrm{tr}(c)=2D$. Then, for $\alpha=\frac{\lambda-\gamma}{1-\gamma}$, we have $\alpha a+(1-\alpha)c=\lambda a+(1-\lambda)b$. Since $\bar{\xi}$ is convex on $B_{\mathrm{tr}}(2D)$, the left-hand side of (6.4.10) is bounded from above by $\alpha\bar{\xi}(a)+(1-\alpha)\bar{\xi}(c)$. Since $\bar{\xi}(c)=\xi(0)+2L(\mathrm{tr}(c)-D)$ is due to $\mathrm{tr}(c)=2D$, by the definition of $\bar{\xi}$, we have $\bar{\xi}(c)\leqslant\gamma\bar{\xi}(a)+(1-\gamma)\bar{\xi}(b)$. Combining these and using the choice of α, we recover (6.4.10), verifying (4). □

Henceforth, we fix a regularization $\bar{\xi}$. We set

$$\mathcal{H}_+=\{\iota\in\mathcal{H} : \iota(s)\in\mathbb{S}^D_+,\ \text{a.e.}\ s\in[0,1]\}$$

and define $\mathsf{H}_{\overline{\xi}} : \mathcal{H}_+ \to \mathbb{R}$ by

$$\mathsf{H}_{\overline{\xi}}(\iota) = \int_0^1 \overline{\xi}(\iota(s))\mathrm{d}s, \quad \forall \iota \in \mathcal{H}_+.$$

The equation (6.4.7) with regularized nonlinearity can be expressed as

$$\partial_t f - \mathsf{H}_{\overline{\xi}}(\nabla f) = 0, \quad \text{on } \mathbb{R}_+ \times \mathcal{C}.$$

Note that $\mathsf{H}_{\overline{\xi}}$ is not defined on the entirety of $\mathcal{H}$, which is needed to apply results from previous sections. Hence, we introduce the following extension,

$$\mathsf{H}(\iota) = \inf\left\{\mathsf{H}_{\overline{\xi}}(\mu) : \mu \in \mathcal{C} \cap (\iota + \mathcal{C}^*)\right\}, \quad \forall \iota \in \mathcal{H}. \tag{6.4.11}$$

We will study properties of H in the next subsection. Now, we can conclude the subsection with the definition of solutions to (6.1.1).

Definition 6.4.3. Under the assumption that $\xi : \mathbb{R}^{D\times D} \to \mathbb{R}$ admits a regularization $\overline{\xi} : \mathbb{S}^D_+ \to \mathbb{R}$, a function $f : \mathbb{R}_+ \times \mathcal{P}_2^{\nearrow} \to \mathbb{R}$ is said to be a *viscosity subsolution* (respectively, *supersolution*) of (6.1.1) (with regularization $\overline{\xi}$), if $\Xi f : \mathbb{R}_+ \times \mathcal{C} \to \mathbb{R}$ is a viscosity subsolution (respectively, supersolution) of $\mathrm{HJ}(\mathcal{H}, \mathcal{C}, \mathsf{H})$ for $\mathcal{H}$, $\mathcal{C}$, H given in (6.3.2), (6.3.4), (6.4.11), respectively. The function f is said to be a *viscosity solution* of (6.1.1) if f is both a subsolution and a supersolution.

6.4.2. Properties of the nonlinearity

In this section, we verify a few useful properties of H. Most of them are recorded in Lemma 6.4.4 below. We will also show an alternative expression of the Hopf–Lax formula in terms of $\overline{\xi}$ in Proposition 6.4.7.

Basic properties

Lemma 6.4.4. *Let* H *be given in* (6.4.11)*. Then, the following hold:*

1. $\mathsf{H}(\mu) = \mathsf{H}_{\overline{\xi}}(\mu)$ *for every* $\mu \in \mathcal{C}$*;*

2. H *is* $\mathcal{C}^*$*-nondecreasing;*

3. H *is Lipschitz;*

4. H *is bounded below;*

5. *if* ξ *is convex, then* H *is convex and satisfies* $\mathsf{H}(\iota^{(j)}) \leqslant \mathsf{H}(\iota)$ *for every* $j \in \mathfrak{J}$ *and every* $\iota \in \mathcal{H}$*.*

Proof. Part (1). We first show that $\mathsf{H}_{\overline{\xi}}$ is $\mathcal{C}^*$-nondecreasing on $\mathcal{C}$. We argue that it suffices to show $\mathsf{H}^j_{\overline{\xi}}$ is $(\mathcal{C}^j)^*$-nondecreasing on $\mathcal{C}^j$, where $\mathsf{H}^j_{\overline{\xi}}$ is the j-projection of $\mathsf{H}_{\overline{\xi}}$. Indeed, for $\mu, \nu \in \mathcal{C}$ satisfying $\mu - \nu \in \mathcal{C}^*$, we have $\mathsf{H}_{\overline{\xi}}(\mu^{(j)}) - \mathsf{H}_{\overline{\xi}}(\nu^{(j)}) = \mathsf{H}^j_{\overline{\xi}}(\mathrm{p}_j\mu) - \mathsf{H}^j_{\overline{\xi}}(\mathrm{p}_j\nu)$. Due to Lemma 6.3.5 (4), we have $\mathrm{p}_j\mu - \mathrm{p}_j\nu \in (\mathcal{C}^j)^*$. Hence, $\mathsf{H}_{\overline{\xi}}(\mu^{(j)}) - \mathsf{H}_{\overline{\xi}}(\nu^{(j)}) \geqslant 0$ for every $j \in \mathfrak{J}$. Passing to the limit along some $\mathfrak{J}_{\mathsf{gen}}$, and using Lemma 6.3.3 (7) and the continuity of $\mathsf{H}_{\overline{\xi}}$ to conclude that $\mathsf{H}_{\overline{\xi}}(\mu) - \mathsf{H}_{\overline{\xi}}(\nu) \geqslant 0$.

With this explained, we compute the gradient of $\mathsf{H}^j_{\overline{\xi}}$. Since for every $x \in \mathcal{C}^j$, $\mathsf{H}^j_{\overline{\xi}}(x) = \sum_{k=1}^{|j|}(t_k - t_{k-1})\overline{\xi}(x_k)$ and $\overline{\xi}$ is locally Lipschitz, we have (recall that the inner product in $\mathcal{H}^j$ is given in (6.3.8))

$$\nabla_j \mathsf{H}^j_{\overline{\xi}}(x) = (\nabla\overline{\xi}(x_k))_{k=1,2,\dots,|j|}, \quad \forall x \in \mathcal{C}^j,$$

which holds almost a.e. on $\mathcal{C}^j$ endowed with the Lebesgue measure. Here ∇_j denotes the gradient of functions defined on subsets of $\mathcal{H}^j$ and ∇ on $\mathbb{S}^D_+$. Since $\overline{\xi}$ is proper, the above display implies that $\nabla_j \mathsf{H}^j_{\overline{\xi}}(x) \in \mathcal{C}^j$ a.e. We can find a full measure set in $\mathcal{C}^j \times \mathcal{C}^j$ such that $\mathsf{H}^j_{\overline{\xi}}$ is differentiable a.e. on the line segment between any two points from this set. For any x, y from this set satisfying $x - y \in (\mathcal{C}^j)^*$, we have

$$\mathsf{H}^j_{\overline{\xi}}(x) - \mathsf{H}^j_{\overline{\xi}}(y) = \int_0^1 \left\langle \nabla_j \mathsf{H}^j_{\overline{\xi}}(sx + (1-s)y),\, x - y \right\rangle_{\mathcal{H}^j} \mathrm{d}s \geqslant 0,$$

where we used the definition of dual cones to deduce that the integrand is nonnegative a.e. Using the density of such pairs, we can conclude that $\mathsf{H}^j_{\bar{\xi}}$ is $(\mathcal{C}^j)^*$-nondecreasing.

Having shown that $\mathsf{H}_{\bar{\xi}}$ is $\mathcal{C}^*$-nondecreasing, we return to the proof. Let $\mu \in \mathcal{C}$. By the definition of H, we clearly have $\mathsf{H}(\mu) \leqslant \mathsf{H}_{\bar{\xi}}(\mu)$. On the other hand, for every $\nu \in \mathcal{C} \cap (\mu + \mathcal{C}^*)$, the monotonicity of $\mathsf{H}_{\bar{\xi}}$ implies that $\mathsf{H}_{\bar{\xi}}(\mu) \leqslant \mathsf{H}_{\bar{\xi}}(\nu)$. Taking infimum in ν, we obtain $\mathsf{H}_{\bar{\xi}}(\mu) \leqslant \mathsf{H}(\mu)$ verifying (1).

Part (2). Let $\iota, \kappa \in \mathcal{H}$ satisfy $\iota - \kappa \in \mathcal{C}^*$. For every $\mu \in \mathcal{C} \cap (\iota + \mathcal{C}^*)$, it is immediate that $\mu \in \mathcal{C} \cap (\kappa + \mathcal{C}^*)$, implying $\mathsf{H}_{\bar{\xi}}(\mu) \geqslant \mathsf{H}(\kappa)$. Taking infimum over $\mu \in \mathcal{C} \cap (\kappa + \mathcal{C}^*)$, we obtain (2).

Part (3). Fix any $\iota, \iota' \in \mathcal{H}$. Let ν be the projection of $\iota - \iota'$ to $\mathcal{C}$. Since $\mathcal{C}$ is closed and convex, we have

$$\left\langle \iota - \iota' - \nu,\ \rho - \nu \right\rangle_{\mathcal{H}} \leqslant 0, \quad \forall \rho \in \mathcal{C}. \tag{6.4.12}$$

Since $s\nu \in \mathcal{C}$ for all $s \geqslant 0$, (6.4.12) yields

$$\left\langle \iota - \iota' - \nu, \nu \right\rangle_{\mathcal{H}} = 0. \tag{6.4.13}$$

Inserting this back to (6.4.12), we have $\langle \iota - \iota' - \nu, \rho \rangle_{\mathcal{H}} \leqslant 0$ for all $\rho \in \mathcal{C}$, which implies

$$\iota' - \iota + \nu \in \mathcal{C}^*,$$

For all $\mu \in \mathcal{C} \cap (\iota' + \mathcal{C}^*)$, the above display implies that $\mu + \nu \in \mathcal{C} \cap (\iota + \mathcal{C}^*)$. Since H is $\mathcal{C}^*$-nondecreasing by (2), we have

$$\mathsf{H}_{\bar{\xi}}(\mu + \nu) \geqslant \mathsf{H}(\iota), \quad \forall \mu \in \mathcal{C} \cap (\iota' + \mathcal{C}^*).$$

By (1), we get

$$|\mathsf{H}_{\overline{\xi}}(\mu+\nu) - \mathsf{H}_{\overline{\xi}}(\mu)| \leqslant \mathbb{E}|\overline{\xi}(\mu(U)+\nu(U)) - \overline{\xi}(\mu(U))| \leqslant \|\overline{\xi}\|_{\mathrm{Lip}}|\nu|_{\mathcal{H}}.$$

The above two displays imply

$$\mathsf{H}(\iota) - \mathsf{H}_{\overline{\xi}}(\mu) \leqslant \|\overline{\xi}\|_{\mathrm{Lip}}|\nu|_{\mathcal{H}}, \quad \forall \mu \in \mathcal{C} \cap (\iota' + \mathcal{C}^*).$$

Due to (6.4.13), we can see that

$$|\iota - \iota'|^2_{\mathcal{H}} = |\iota - \iota' - \nu|^2_{\mathcal{H}} + |\nu|^2_{\mathcal{H}} \geqslant |\nu|^2_{\mathcal{H}}.$$

Using this and taking supremum over $\mu \in \mathcal{C} \cap (\iota' + \mathcal{C}^*)$, we obtain

$$\mathsf{H}(\iota) - \mathsf{H}(\iota') \leqslant \|\overline{\xi}\|_{\mathrm{Lip}}|\iota - \iota'|_{\mathcal{H}}.$$

By symmetry, we conclude that H is Lipschitz.

Part (4). This is clear from Definition 6.4.1 (3) and (6.4.11).

Part (5). By Definition 6.4.1 (6.4.10), we have that $\overline{\xi}$ is convex. From (1), we can see that H is convex on $\mathcal{C}$. For every $\iota, \kappa \in \mathcal{H}$ and every $s \in [0,1]$, we have $s\mu + (1-s)\nu \in \mathcal{C} \cap (s\iota + (1-s)\kappa + \mathcal{C}^*)$ if $\mu \in \mathcal{C} \cap (\iota + \mathcal{C}^*)$ and $\nu \in \mathcal{C} \cap (\kappa + \mathcal{C}^*)$. In view of this, the convexity of H on $\mathcal{H}$ follows from its convexity on $\mathcal{C}$ and (6.4.11). To see the second claim, using Jensen's inequality, we have that, for every $\mu \in \mathcal{C}$ and every $j \in \mathfrak{J}$,

$$\mathsf{H}\left(\mu^{(j)}\right) = \sum_{k=0}^{|j|}(t_{k+1} - t_k)\overline{\xi}\left(\frac{1}{t_k - t_{k-1}}\int_{t_{k-1}}^{t_k} \mu(s)\mathrm{d}s\right) \leqslant \int_0^1 \overline{\xi}(\mu(s))\mathrm{d}s = \mathsf{H}(\mu).$$

Fix any $\iota \in \mathcal{H}$. By Lemma 6.3.3 (4) and Lemma 6.3.5, we have $\mu^{(j)} \in \mathcal{C} \cap (\iota^{(j)} + \mathcal{C}^*)$ for every $j \in \mathfrak{J}$, and every $\mu \in \mathcal{C} \cap (\iota + \mathcal{C}^*)$. Therefore, the above display along with the definition of

H implies that

$$\mathsf{H}\left(\iota^{(j)}\right) \leqslant \mathsf{H}(\mu), \quad \forall \mu \in \mathcal{C} \cap (\iota + \mathcal{C}^*).$$

Taking infimum over $\mu \in \mathcal{C} \cap (\iota + \mathcal{C}^*)$, we conclude that $\mathsf{H}(\iota^{(j)}) \leqslant \mathsf{H}(\iota)$. □

Alternative expression of the Hopf–Lax formula

The goal is to prove Proposition 6.4.7. We need several lemmas in preparation for the proof.

We introduce the notation for the *nondecreasing rearrangement*. For every $j \in \mathfrak{J}_{\mathsf{unif}}$ and $x \in \mathcal{H}^j$, we set $x_\sharp = (x_{\sigma(k)})_{k=1,2,\ldots,|j|}$ where σ is a permutation of $(1, 2, \ldots, |j|)$ satisfying $x_{\sigma(k)} - x_{\sigma(k-1)} \in \mathbb{S}^D_+$ for every $k \geqslant 2$. Using this notation, for every $j \in \mathfrak{J}_{\mathsf{unif}}$ and every $\iota \in \mathcal{H}$, we set $\iota^{(j)}_\sharp = \mathrm{l}_j((\mathrm{p}_j\iota)_\sharp)$.

Lemma 6.4.5. *For every $j \in \mathfrak{J}_{\mathsf{unif}}$,*

1. $\iota^{(j)}_\sharp \in \iota^{(j)} + \mathcal{C}^*$ *for every* $\iota \in \mathcal{H}$*;*
2. $\iota^{(j)}_\sharp \in \mathcal{C}$ *for every* $\iota \in \mathcal{H}_+$*;*
3. $\mathbb{E}h(\iota^{(j)}(U)) = \mathbb{E}h(\iota^{(j)}_\sharp(U))$ *for every real-valued function h and every $\iota \in \mathcal{H}$.*

Proof. Part (1). For every $x \in \mathcal{H}^j$, by the rearrangement inequality, we have

$$\langle x_\sharp, y\rangle_{\mathcal{H}^j} = \frac{1}{|j|}\sum_{k=1}^{|j|} x_{\sigma(k)} y_k \geqslant \frac{1}{|j|}\sum_{k=1}^{|j|} x_k y_k = \langle x, y\rangle_{\mathcal{H}^j}, \quad \forall y \in \mathcal{C}^j,$$

where σ is the permutation in the definition of $x_\sharp$. This implies that $x_\sharp - x \in (\mathcal{C}^j)^*$. By the definition of $\iota^{(j)}_\sharp$ and Lemma 6.3.3 (3), we have $\mathrm{p}_j(\iota^{(j)}_\sharp) = (\mathrm{p}_j\iota)_\sharp$. Hence, we get $\mathrm{p}_j(\iota^{(j)}_\sharp) - \mathrm{p}_j\iota \in (\mathcal{C}^j)^*$, which along with Lemma 6.3.5 (2) implies that $(\iota^{(j)}_\sharp)^{(j)} - \iota^{(j)} \in \mathcal{C}^*$. By Lemma 6.3.3 (3) and (4), we have

$$(\iota^{(j)}_\sharp)^{(j)} = \mathrm{l}_j\mathrm{p}_j\mathrm{l}_j((\mathrm{p}_j\iota)_\sharp) = \mathrm{l}_j((\mathrm{p}_j\iota)_\sharp) = \iota^{(j)}_\sharp.$$

Then, (1) follows.

Part (2). Let $\iota \in \mathcal{H}_+$. It is clear from the definition that $(\mathrm{p}_j\iota)_\sharp \in \mathcal{C}^j$. Then, by Lemma 6.3.5 (1), we get $\iota_\sharp^{(j)} \in \mathcal{C}$.

Part (3). We can compute that

$$\mathbb{E}h\left(\iota^{(j)}(U)\right) = \frac{1}{|j|}\sum_{k=1}^{|j|} h\left(|j|\int_{\frac{k-1}{|j|}}^{\frac{k}{|j|}} \iota(s)\mathrm{d}s\right) = \frac{1}{|j|}\sum_{k=1}^{|j|} h\left(|j|\int_{\frac{\sigma(k)-1}{|j|}}^{\frac{\sigma(k)}{|j|}} \iota(s)\mathrm{d}s\right) = \mathbb{E}h\left(\iota_\sharp^{(j)}(U)\right)$$

where the permutation σ is the one corresponding to the nondecreasing rearrangement of $\mathrm{p}_j\iota$. This completes the proof. □

Recall the definitions of conjugates in (6.1.6) and (6.1.7). The monotone conjugate $\overline{\xi}^*$ is defined with respect to the cone $\mathbb{S}_+^D$ in the space $\mathbb{S}^D$, namely,

$$\overline{\xi}^*(a) = \sup_{a' \in \mathbb{S}_+^D} \left\{a' \cdot a - \overline{\xi}(a')\right\}, \quad \forall a \in \mathbb{S}^D. \tag{6.4.14}$$

Next, we show the following.

Lemma 6.4.6. *For every $\mu \in \mathcal{C}$, it holds that $\mathsf{H}^{\circledast}(\mu) = \mathbb{E}\overline{\xi}^*(\mu(U))$.*

Proof. We proceed in three steps.

Step 1. Setting

$$\mathsf{H}^*_{\overline{\xi}}(\iota) = \sup_{\nu\in\mathcal{C}}\{\langle \nu, \iota\rangle_{\mathcal{H}} - \mathsf{H}_{\overline{\xi}}(\nu)\}, \quad \forall \iota \in \mathcal{H},$$

we first show that

$$\mathsf{H}^{\circledast}(\mu) = \mathsf{H}^*_{\overline{\xi}}(\mu), \quad \forall \mu \in \mathcal{C}. \tag{6.4.15}$$

Since the supremum in $\mathsf{H}^{\circledast}$ is taken over $\mathcal{H}$ and since H coincides with $\mathsf{H}_{\overline{\xi}}$ on $\mathcal{C}$ (by

Lemma 6.4.4 (1)), it is immediate that $\mathsf{H}^{\circledast}(\mu) \geqslant \mathsf{H}^*_{\bar{\xi}}(\mu)$. Fix any $\varepsilon > 0$. For every $\iota \in \mathcal{H}$, by the definition of H in (6.4.11), there is $\nu \in \mathcal{C} \cap (\iota + \mathcal{C}^*)$ such that $\mathsf{H}(\iota) \geqslant \mathsf{H}_{\bar{\xi}}(\nu) - \varepsilon$. Due to $\nu - \iota \in \mathcal{C}^*$ and $\mu \in \mathcal{C}$, we have $\langle \iota, \mu\rangle_{\mathcal{H}} \leqslant \langle \nu, \mu\rangle_{\mathcal{H}}$. Hence, we obtain

$$\langle \iota, \mu\rangle_{\mathcal{H}} - \mathsf{H}(\iota) \leqslant \langle \nu, \mu\rangle_{\mathcal{H}} - \mathsf{H}_{\bar{\xi}}(\nu) + \varepsilon.$$

Since the right-hand side of the above is bounded by $\mathsf{H}^*_{\bar{\xi}}(\mu) + \varepsilon$, taking supremum over $\iota \in \mathcal{H}$ yields $\mathsf{H}^{\circledast}(\mu) \leqslant \mathsf{H}^*_{\bar{\xi}}(\mu) + \varepsilon$. Sending $\varepsilon \to 0$, we obtain (6.4.15).

Step 2. We show

$$\mathsf{H}^*_{\bar{\xi}}(\mu) = \mathsf{H}^{\oplus}_{\bar{\xi}}(\mu), \quad \forall \mu \in \mathcal{C}, \tag{6.4.16}$$

where $\mathsf{H}^{\oplus}_{\bar{\xi}}$ is defined by

$$\mathsf{H}^{\oplus}_{\bar{\xi}}(\iota) = \sup_{\kappa \in \mathcal{H}_+} \{\langle \kappa, \iota\rangle_{\mathcal{H}} - \mathsf{H}_{\bar{\xi}}(\kappa)\}, \quad \forall \iota \in \mathcal{H}.$$

Due to $\mathcal{C} \subseteq \mathcal{H}_+$, it is easy to see that $\mathsf{H}^*_{\bar{\xi}}(\mu) \leqslant \mathsf{H}^{\oplus}_{\bar{\xi}}(\mu)$ for $\mu \in \mathcal{C}$. For any fixed $\varepsilon > 0$, by the definition of $\mathsf{H}^{\oplus}_{\bar{\xi}}$, there is $\kappa \in \mathcal{H}_+$ such that

$$\mathsf{H}^{\oplus}_{\bar{\xi}}(\mu) \leqslant \langle \kappa, \mu\rangle_{\mathcal{H}} - \mathsf{H}_{\bar{\xi}}(\kappa) + \varepsilon.$$

Using Lemma 6.3.3 (7) and the continuity of $\mathsf{H}_{\bar{\xi}}$, we can find $j \in \mathfrak{J}_{\mathsf{unif}}$ such that

$$\mathsf{H}^{\oplus}_{\bar{\xi}}(\mu) \leqslant \left\langle \kappa^{(j)}, \mu \right\rangle_{\mathcal{H}} - \mathsf{H}_{\bar{\xi}}\left(\kappa^{(j)}\right) + 2\varepsilon.$$

Lemma 6.4.5 implies that $\left\langle \kappa^{(j)}, \mu\right\rangle_{\mathcal{H}} \leqslant \langle \kappa^{(j)}_\sharp, \mu\rangle_{\mathcal{H}}$ (due to $\mu \in \mathcal{C}$), $\mathsf{H}_{\bar{\xi}}(\kappa^{(j)}) = \mathsf{H}_{\bar{\xi}}(\kappa^{(j)}_\sharp)$, and $\kappa^j_\sharp \in \mathcal{C}$. These together with the above display yields $\mathsf{H}^{\oplus}_{\bar{\xi}}(\mu) \leqslant \mathsf{H}^*_{\bar{\xi}}(\mu) + 2\varepsilon$. Since ε is arbitrary, we obtain (6.4.16).

Step 3. We show

$$\mathsf{H}^{\oplus}_{\overline{\xi}}(\iota) = \mathbb{E}\overline{\xi}^*(\iota(U)), \quad \forall \iota \in \mathcal{H}_+. \tag{6.4.17}$$

For every $\iota \in \mathcal{H}$, $\kappa \in \mathcal{H}_+$, by the definition of $\overline{\xi}^*$, we have

$$\overline{\xi}^*(\iota(s)) \geqslant \iota(s) \cdot \kappa(s) - \overline{\xi}(\kappa(s)), \quad \forall s \in [0,1).$$

Integrating in s, we get

$$\mathbb{E}\overline{\xi}^*(\iota(U)) \geqslant \langle \iota, \kappa \rangle_{\mathcal{H}} - \mathsf{H}_{\overline{\xi}}(\kappa).$$

Taking supremum over $\kappa \in \mathcal{H}_+$, we obtain $\mathbb{E}\overline{\xi}^*(\iota(U)) \geqslant \mathsf{H}^{\oplus}_{\overline{\xi}}(\iota)$ for every $\iota \in \mathcal{H}$.

For the converse inequality, note that $\overline{\xi}^*$ is lower-semicontinuous and $\overline{\xi}^*(\iota) \geqslant \overline{\xi}(0)$ by the definition of $\overline{\xi}^*$. Using Lemma 6.3.3 (7), we can extract from $\mathfrak{J}_{\mathsf{unif}}$ a sequence $(j_n)_{n=1}^{\infty}$ satisfying $\lim_{n\to\infty} \iota^{(j_n)} = \iota$ a.e. on $[0,1)$. Since $\overline{\xi}^*(\iota) \geqslant \overline{\xi}(0)$, invoking Fatou's lemma, we get

$$\mathbb{E}\overline{\xi}^*(\iota(U)) \leqslant \mathbb{E} \liminf_{n\to\infty} \overline{\xi}^* \left(\iota^{(j_n)}(U)\right) \leqslant \liminf_{n\to\infty} \mathbb{E}\overline{\xi}^* \left(\iota^{(j_n)}(U)\right).$$

Recall the definitions of $\iota^{(j)}$ in (6.3.5), and $\mathrm{p}_j\iota$ in (6.3.9). We can compute that

$$\begin{aligned}\mathbb{E}\overline{\xi}^* \left(\iota^{(j)}(U)\right) &= \sum_{k=1}^{|j|}(t_k - t_{k-1})\overline{\xi}^*((\mathrm{p}_j\iota)_k) = \sum_{k=1}^{|j|}(t_k - t_{k-1}) \sup_{x_k \in \mathbb{S}^D_+} \left\{x_k \cdot (\mathrm{p}_j\iota)_k - \overline{\xi}(x_k)\right\} \\ &= \sup_{x \in \mathcal{H}^j_+} \left\{\langle x, \mathrm{p}_j\iota\rangle_{\mathcal{H}^j} - \mathbb{E}\overline{\xi}(\mathrm{l}_j x(U))\right\} = \sup_{x \in \mathcal{H}^j_+} \left\{\langle \mathrm{l}_j x, \iota\rangle_{\mathcal{H}} - \mathsf{H}_{\overline{\xi}}(\mathrm{l}_j x)\right\} \leqslant \mathsf{H}^{\oplus}_{\overline{\xi}}(\iota),\end{aligned}$$

where $\mathcal{H}^j_+$ stands for $\mathrm{p}_j(\mathcal{H}_+) = \{x \in \mathcal{H}^j : x_k \in \mathbb{S}^D_+, \forall k\}$, and, in the last equality, we used Lemma 6.3.3 (2). The above two displays together yield $\mathbb{E}\overline{\xi}^*(\iota(U)) \leqslant \mathsf{H}^{\circledast}(\iota)$ for every $\iota \in \mathcal{H}$, verifying (6.4.17). The desired result follows from (6.4.15), (6.4.16) and (6.4.17). $\square$

Now, we are ready to prove the following.

Proposition 6.4.7 (Hopf-Lax formula in the spin glass setting)**.** *If ψ is $\mathcal{C}^*$-nondecreasing and continuous, then*

$$\sup_{\nu\in\mathcal{C}}\left\{\psi(\nu)-t\mathsf{H}^{\circledast}\left(\frac{\nu-\mu}{t}\right)\right\}=\sup_{\nu\in\mu+\mathcal{C}}\left\{\psi(\nu)-t\mathbb{E}\overline{\xi}^*\left(\frac{(\nu-\mu)(U)}{t}\right)\right\},\quad\forall\mu\in\mathcal{C}.\tag{6.4.18}$$

If $D=1$, then the right-hand side of the above is equal to

$$\sup_{\nu\in\mathcal{C}}\left\{\psi(\nu)-t\mathbb{E}\overline{\xi}^*\left(\frac{(\nu-\mu)(U)}{t}\right)\right\},\quad\forall\mu\in\mathcal{C}.\tag{6.4.19}$$

Proof. Let us denote the right-hand side in (6.4.18) by RHS. For simplicity, we omit U and write $\mathbb{E}\overline{\xi}^*(\iota)=\mathbb{E}\overline{\xi}^*(\iota(U))$ for all $\iota\in\mathcal{H}$. By Lemma 6.2.5 (1) which still holds in infinite dimensions, we have $\mathsf{H}^{\circledast}(\iota)=\infty$ if $\iota\notin\mathcal{C}$. Using this and Lemma 6.4.6, we can get the the left-hand side of (6.4.18) is equal to

$$\sup_{\nu\in\mu+\mathcal{C}}\left\{\psi(\nu)-t\mathsf{H}^{\circledast}\left(\frac{\nu-\mu}{t}\right)\right\}=\mathsf{RHS},\quad\forall\mu\in\mathcal{C},$$

verifying (6.4.18).

Now, we assume $D=1$. Denoting the term in (6.4.19) by I, to show (6.4.19), we only need to show $\mathsf{I}\leqslant\mathsf{RHS}$. Now, note that $\overline{\xi}^*(r)=\sup_{s\geqslant0}\{sr-\overline{\xi}(s)\}$. Since $\overline{\xi}$ is $\mathbb{S}^D_+$-nondecreasing (see Definition 6.4.1 (2)), we have $\overline{\xi}^*(r)=\overline{\xi}^*(0)$ for all $r<0$. For every $\kappa\in\mathcal{H}$, we define κ_+ by $\kappa_+(s)=(\kappa(s))\vee0$ for all $s\in[0,1)$. Then, we have

$$\mathsf{I}=\sup_{\nu\in\mathcal{C}}\left\{\psi(\nu)-t\mathbb{E}\overline{\xi}^*\left(\frac{(\nu-\mu)_+}{t}\right)\right\}.$$

For every $\varepsilon>0$, we can find $\nu\in\mathcal{C}$ such that

$$\mathsf{I}\leqslant\psi(\nu)-t\mathbb{E}\overline{\xi}^*(\iota)+\varepsilon,$$

where we set $\iota=\frac{1}{t}(\nu-\mu)_+\in\mathcal{H}_+$. We choose a sufficiently fine $j\in\mathfrak{J}_{\mathsf{unif}}$ satisfying $\psi(\nu)\leqslant\psi(\nu^{(j)})+\varepsilon$. Since $\overline{\xi}^*$ is convex, we have $\mathbb{E}\overline{\xi}^*(\iota^{(j)})\leqslant\mathbb{E}\overline{\xi}^*(\iota)$ by Jensen's inequality.

Hence, the above becomes

$$\mathsf{I} \leqslant \psi\left(\nu^{(j)}\right) - t\mathbb{E}\overline{\xi}^*\left(\iota^{(j)}\right) + 2\varepsilon.$$

Setting $\rho = \mu + t\iota_\sharp^{(j)}$, we have

$$\rho - \nu^{(j)} = (\mu - \mu^{(j)}) + t(\iota_\sharp^{(j)} - \iota^{(j)}) + \left(t\iota^{(j)} - (\nu - \mu)^{(j)}\right) \in \mathcal{C}^*$$

where the terms inside the first two pairs of parentheses on the right belong to $\mathcal{C}^*$ due to Lemma 6.3.5 (5) and Lemma 6.4.5 (1), respectively; and it is easy to see that the term in the last pair of parentheses belongs to $\mathcal{H}_+ \subseteq \mathcal{C}^*$. Due to $\iota \in \mathcal{H}_+$, Lemma 6.4.5 (2) implies $\rho \in \mathcal{C}$ and $\rho - \mu \in \mathcal{C}$. Since ψ is $\mathcal{C}^*$-nondecreasing, we get $\psi(\nu^{(j)}) \leqslant \psi(\rho)$. Lemma 6.4.5 (3) also gives $\mathbb{E}\overline{\xi}^*(\frac{\rho-\mu}{t}) = \mathbb{E}\overline{\xi}^*(\iota_\sharp^{(j)}) = \mathbb{E}\overline{\xi}^*(\iota^{(j)})$. These along with Lemma 6.4.6 yield

$$\mathsf{I} \leqslant \psi(\rho) - t\mathbb{E}\overline{\xi}^*\left(\frac{\rho-\mu}{t}\right) + 2\varepsilon = \psi(\rho) - t\mathsf{H}^{\circledast}\left(\frac{\rho-\mu}{t}\right) + 2\varepsilon \leqslant \mathsf{RHS} + 2\varepsilon.$$

Sending $\varepsilon \to 0$, we obtain the desired result. □

6.4.3. Proof of the main result

We state the rigorous version of Theorem 6.1.1. Recall the isometry Ξ given in (6.4.3); the action of Ξ on functions in (6.4.5); the 1-Wasserstein metric $\mathbf{d}_1$ in (6.4.2); the definition of solutions in Definition 6.4.3; Hilbert spaces $\mathcal{H}$ and $\mathcal{H}^j$, $j \in \mathfrak{J}$, in (6.3.2) and (6.3.7), respectively; cones $\mathcal{C}$ and $\mathcal{C}^j$, $j \in \mathfrak{J}$, in (6.3.4) and (6.3.11), respectively; the definition of $\mathcal{C}^*$-nondecreasingness in (6.1.5); $\overline{\xi}^*$ in (6.4.14); the definition of good collections of partitions at the beginning of Section 6.3.1; lifts and projections of functions in Definition 6.3.1.

Theorem 6.4.8. *Suppose*

- $\xi : \mathbb{R}^{D\times D} \to \mathbb{R}$ *is locally Lipschitz and its restriction to* $\mathbb{S}^D_+$ *is proper;*

- $\psi : \mathcal{P}_2^{\nearrow} \to \mathbb{R}$ *satisfies that* $\Xi\psi : \mathcal{C} \to \mathbb{R}$ *is* $\mathcal{C}^*$*-nondecreasing, and*

$$|\psi(\varrho) - \psi(\vartheta)| \leqslant \mathbf{d}_1(\varrho, \vartheta), \quad \forall \varrho, \vartheta \in \mathcal{P}_2^{\nearrow}. \tag{6.4.20}$$

Then, regularizations of ξ *exist, and, for any regularization* $\overline{\xi}$*, there is a unique Lipschitz viscosity solution* f *to* (6.1.1) *(with regularization* $\overline{\xi}$*) satisfying* $f(0, \cdot) = \psi$*. Moreover,*

1. $\Xi f = \lim_{j \in \mathfrak{J}_{\mathsf{good}}} f_j^{\uparrow}$ *in the local uniform topology, for any good collection of partitions* $\mathfrak{J}_{\mathsf{good}}$*, where* f_j *is the unique Lipschitz viscosity solution to* $\mathrm{HJ}(\mathcal{H}^j, \mathcal{C}^j, \mathsf{H}^j; (\Xi\psi)^j)$ *for every* $j \in \mathfrak{J}_{\mathsf{good}}$ *and for* H *given in* (6.4.11)*;*

2. *if* ξ *is convex, then* f *is given by the Hopf–Lax formula*

$$\Xi f(t, \mu) = \sup_{\nu \in \mu + \mathcal{C}} \left\{ \Xi\psi(\nu) - t\mathbb{E}\overline{\xi}^* \left(\frac{(\nu - \mu)(U)}{t} \right) \right\}, \quad \forall (t, \mu) \in \mathbb{R}_+ \times \mathcal{C}; \tag{6.4.21}$$

3. *if* $\Xi\psi$ *is convex, then* f *is given by the Hopf formula*

$$\Xi f(t, \mu) = \sup_{\nu \in \mathcal{C}} \inf_{\rho \in \mathcal{C}} \left\{ \langle \nu, \mu - \rho \rangle_{\mathcal{H}} + \Xi\psi(\rho) + t\mathbb{E}\overline{\xi}(\nu(U)) \right\}, \quad \forall (t, \mu) \in \mathbb{R}_+ \times \mathcal{C}. \tag{6.4.22}$$

Remark 6.4.9. In view of Definition 6.4.3 and Lemma 6.4.4, Proposition 6.3.8 supplies a comparison principle for (6.1.1).

Remark 6.4.10. We briefly comment on the assumptions on ξ and ψ. As mentioned previously, in most of interesting models, ξ given in (6.4.6) admits a convergent power series and is proper on $\mathbb{S}^D_+$ (see [93, Propositions 6.4 and 6.6]). In practice, ψ will be the limit of $\overline{F}_N(0, \cdot)$ as $N \to \infty$. Due to (6.4.8) and (6.4.9), the assumptions on ψ are natural. In general, the initial condition ψ as the limit of $\overline{F}_N(0, \cdot)$ is neither concave or convex, which renders the Hopf formula less useful. A discussion on existence of variational formulae for free energy limits is in [96, Section 6].

Remark 6.4.11. If $D = 1$, using Proposition 6.4.7, we can slightly simplify (6.4.21) into

$$\Xi f(t,\mu) = \sup_{\nu \in \mathcal{C}} \left\{ \Xi\psi(\nu) - t\mathbb{E}\overline{\xi}^* \left(\frac{(\nu-\mu)(U)}{t} \right) \right\}, \quad \forall (t,\mu) \in \mathbb{R}_+ \times \mathcal{C}. \tag{6.4.23}$$

The supremum in (6.4.21) is taken over $\mu + \mathcal{C}$, and now the supremum is simply over $\mathcal{C}$. We do not know if this simplification holds for $D > 1$.

Remark 6.4.12. Let us discuss how solutions considered in [92, 98, 96, 93] can be recast as viscosity solutions. In [92, 98] where $D = 1$, the solution to (6.1.1) was defined as (6.4.23) with $\overline{\xi}$ replaced by the original ξ. Since it is only the values of ξ over matrices in $\mathbb{S}^D_+$ with entries in $[-1, 1]$ that matters (see the discussion below Definition 6.4.1), one can work directly with the regularization $\overline{\xi}$. Then, due to Theorem 6.4.8 (2), the Hopf–Lax solution in [92, 98] is the unique viscosity solution. In [96, 93], the solution was defined as the limit of solutions of $\mathrm{HJ}(\mathcal{H}^j, \mathcal{C}^j, \mathsf{H}^j; (\Xi\psi)^j)$ indexed by $j \in \mathfrak{J}_{\mathsf{unif}}$. Although solutions of finite-dimensional equations in [96, 93] were required to satisfy the Neumann boundary condition, the theory developed there is compatible with the definition of solutions in this work. All results related to viscosity solutions in finite dimensions there can be replaced by their counterparts in this work. Moreover, some arguments can be simplified due to the simplification of the boundary condition. Therefore, with this modification, in view of Theorem 6.4.8 (1), the solution considered in [96, 93] is the unique viscosity solution.

Proof of Theorem 6.4.8. Lemma 6.4.2 guarantees the existence of regularizations. We fix any regularization $\overline{\xi}$. The properties of H are listed in Lemma 6.4.4. In particular, H satisfies (A1)–(A2). Using (6.4.4), we can rewrite (6.4.20) as

$$|\Xi\psi(\mu) - \Xi\psi(\nu)| \leqslant C|\mu - \nu|_{L^1}, \quad \forall \mu, \nu \in \mathcal{C},$$

where L^1 is given in (6.3.3). The existence of a viscosity solution f and (1) follow from Propositions 6.3.9 and Proposition 6.3.10. The latter proposition also ensures that Ξf is Lipschitz. By Proposition 6.3.8, Ξf (and thus f) is the unique viscosity solution that is

Lipschitz.

Given that ξ is convex, it is easy to see from Lemma 6.4.4 (4) and (5) that H^j is convex and bounded below for every $j \in \mathfrak{J}$. In addition, it is also clear that $(\Xi\psi)^j : \mathcal{C}^j \to \mathbb{R}$ is Lipschitz and $(\mathcal{C}^j)^*$-nondecreasing. Therefore, Proposition 6.2.6 ensures that the viscosity solution f_j of $\mathrm{HJ}(\mathcal{H}^j, \mathcal{C}^j, \mathsf{H}^j; (\Xi\psi)^j)$ admits a representation given in (6.3.53) with ψ there replaced by $\Xi\psi$. Hence, Proposition 6.3.12 along with Lemma 6.4.4 (5) and Proposition 6.4.7 yields (2).

Proposition 6.5.1 stated and proved later ensures that $\mathcal{C}^j$ has the Fenchel–Moreau property as defined in Definition 6.2.7, for each $j \in \mathfrak{J}$. Under the assumption that $\Xi\psi$ is convex, it is straightforward to see that $(\Xi\psi)^j : \mathcal{C}^j \to \mathbb{R}$ is also convex. Invoking Proposition 6.2.8, we have that the viscosity solution f_j of $\mathrm{HJ}(\mathcal{H}^j, \mathcal{C}^j, \mathsf{H}^j; (\Xi\psi)^j)$ is given by (6.3.55) with ψ there replaced by $\Xi\psi$. Then, (3) follows from Proposition 6.3.13 along with Lemma 6.4.4 (1). □

6.5. Fenchel–Moreau identity on cones

Recall Definition 6.2.7 of the Fenchel–Moreau property. To apply Proposition 6.2.8 to equations on $\mathbb{R}_+ \times \mathcal{C}^j$, $j \in \mathfrak{J}$, we need to show that $\mathcal{C}^j$ given in (6.3.11) has the Fenchel–Moreau property. Adapting the definition of monotone conjugate in (6.1.7) to $\mathcal{C}^j$ with ambient Hilbert space $\mathcal{H}^j$ given in (6.3.7), in this section, for any $g : \mathcal{C}^j \to (-\infty, \infty]$, we set

$$g^*(y) = \sup_{x \in \mathcal{C}^j} \{\langle x, y\rangle_{\mathcal{H}^j} - g(x)\}, \quad \forall y \in \mathcal{H}^j, \tag{6.5.1}$$

and $g^{**} = (g^*)^*$, where g^* is understood to be its restriction to $\mathcal{C}^j$.

Proposition 6.5.1. *For every $j \in \mathfrak{J}$, the closed convex cone $\mathcal{C}^j$ possesses the Fenchel–Moreau property: for $g : \mathcal{C}^j \to (-\infty, \infty]$ not identically equal to ∞, we have $g^{**} = g$ if and only if g is convex, lower semicontinuous and $(\mathcal{C}^j)^*$-nondecreasing.*

The proof largely follows the steps in [38]. We first recall basic results of convex analysis. Then, we show Lemma 6.5.7 which treats the case where the effective domain of g has nonempty interior. Finally, in the last subsection, we extend the result to the general case.

6.5.1. Basic results of convex analysis

For $a \in \mathcal{H}^j$ and $\nu \in \mathbb{R}$, we define the affine function $L_{a,\nu}$ with slope a and translation ν by

$$L_{a,\nu}(x) = \langle a, x \rangle_{\mathcal{H}^j} + \nu, \quad \forall x \in \mathcal{H}^j.$$

For a function $g : \mathcal{E} \to (-\infty, \infty]$ defined on a subset $\mathcal{E} \subseteq \mathcal{H}^j$, we can extend it in the standard way to $g : \mathcal{H}^j \to (-\infty, \infty]$ by setting $g(x) = \infty$ for $x \notin \mathcal{E}$. For $g : \mathcal{H}^j \to (-\infty, \infty]$, we define its *effect domain* by

$$\mathsf{dom}\, g = \left\{ x \in \mathcal{H}^j : \ g(x) < \infty \right\}.$$

Henceforth, we shall not distinguish functions defined on $\mathcal{C}^j$ from their standard extensions to $\mathcal{H}^j$. We denote by $\Gamma_0(\mathcal{E})$ the collection of convex and lower semicontinuous functions from $\mathcal{E} \subseteq \mathcal{H}^j$ to $(-\infty, \infty]$ with nonempty effect domain.

For $g : \mathcal{H}^j \to (-\infty, \infty]$ and each $x \in \mathcal{H}^j$, recall that the subdifferential of g at x is given by

$$\partial g(x) = \left\{ z \in \mathcal{H}^j : g(y) \geqslant g(x) + \langle z, y - x \rangle_{\mathcal{H}^j}, \forall y \in \mathcal{H}^j \right\}.$$

The effective domain of ∂g is defined to be

$$\mathsf{dom}\, \partial g = \left\{ x \in \mathcal{H}^j : \ \partial g(x) \neq \emptyset \right\}.$$

We now list some lemmas needed in our proofs.

Lemma 6.5.2. *For a convex set $\mathcal{E} \subseteq \mathcal{H}^j$, if $y \in \mathsf{cl}\, \mathcal{E}$ and $y' \in \mathsf{int}\, \mathcal{E}$, then $\lambda y + (1-\lambda) y' \in \mathsf{int}\, \mathcal{E}$ for all $\lambda \in [0, 1)$.*

Lemma 6.5.3. *For $g \in \Gamma_0(\mathcal{H}^j)$, it holds that $\mathsf{int}\,\mathsf{dom}\, g \subseteq \mathsf{dom}\, \partial g \subseteq \mathsf{dom}\, g$.*

Lemma 6.5.4. *Let $g \in \Gamma_0(\mathcal{H}^j)$, $x \in \mathcal{H}^j$ and $y \in \mathsf{dom}\, g$. For every $\alpha \in (0, 1)$, set $x_\alpha =$*

$(1-\alpha)x+\alpha y$. *Then* $\lim_{\alpha\to 0} g(x_\alpha) = g(x)$.

Lemma 6.5.5. *Let* $g \in \Gamma_0(\mathcal{C}^j)$, $x \in \mathcal{C}^j$ *and* $y \in \mathcal{C}^j$. *If* $y \in \partial g(x)$, *then* $g^*(y) = \langle x, y\rangle_{\mathcal{H}^j} - g(x)$.

Lemma 6.5.6. *For* $g \in \Gamma_0(\mathcal{C}^j)$ *and* $x \in \mathcal{C}^j$, *we have*

$$g^{**}(x) = \sup L_{a,\nu}(x)$$

where the supremum is taken over

$$\{(a,\nu) \in \mathcal{C}^j \times \mathbb{R} : L_{a,\nu} \leqslant g \text{ on } \mathcal{C}^j\}. \tag{6.5.2}$$

For, Lemmas 6.5.2, 6.5.3, and 6.5.4, we refer to [23, Propositions 3.35, 16.21, and 9.14]. Here, let us prove Lemma 6.5.5 and Lemma 6.5.6.

Proof of Lemma 6.5.5. By the standard extension, we have $g \in \Gamma_0(\mathcal{H}^j)$. Since $y \in \partial g(x)$, it is classically known (c.f. [23, Theorem 16.23]) that

$$\sup_{z\in\mathcal{H}^j} \{\langle z, y\rangle_{\mathcal{H}^j} - g(z)\} = \langle x, y\rangle_{\mathcal{H}^j} - g(x).$$

By assumption, we know $x \in \mathsf{dom}\,\partial g$. Hence, Lemma 6.5.3 implies $x \in \mathsf{dom}\, g$ and thus both sides above are finite. On the other hand, by the extension, we have $g(z) = \infty$ if $z \notin \mathcal{C}^j$, which yields

$$\sup_{z\in\mathcal{H}^j} \{\langle z, y\rangle_{\mathcal{H}^j} - g(z)\} = \sup_{z\in\mathcal{C}^j} \{\langle z, y\rangle_{\mathcal{H}^j} - g(z)\} = g^*(y).$$

The desired result follows from the above two displays. □

Proof of Lemma 6.5.6. For each $y \in \mathcal{C}^j$,

$$L_{y,\,-g^*(y)}(x) = \langle y, x\rangle_{\mathcal{H}^j} - g^*(y), \quad \forall x \in \mathcal{C}.$$

is an affine function with slope $y \in \mathcal{C}^j$. By (6.1.7), we can see that $L_{y,\,-g^*(y)} \leqslant g$ on $\mathcal{C}^j$. In view of the definition of g^{**} in (6.2.29), we have $g^{**}(x) \leqslant \sup L_{a,\nu}(x)$ for all $x \in \mathcal{C}^j$ where the sup is taken over the collection in (6.5.2).

For the other direction, if (α, ν) belongs to the set in (6.5.2), we have

$$\langle a, x\rangle_{\mathcal{H}^j} + \nu \leqslant g(x), \quad \forall x \in \mathcal{C}^j.$$

Rearranging and taking supremum in $x \in \mathcal{C}^j$, we get $g^*(a) \leqslant -\nu$. This yields

$$L_{a,\nu}(x) \leqslant \langle a, x\rangle_{\mathcal{H}^j} - g^*(a) \leqslant g^{**}(x),$$

which implies $\sup L_{a,\nu}(x) \leqslant g^{**}(x)$. □

The proof of Proposition 6.5.1 consists of two parts. The first part, summarized in the lemma below, concerns the case where $\mathsf{dom}\, g$ has non-empty interior.

Lemma 6.5.7. *If* $\mathsf{int}\,\mathsf{dom}\, g \neq \emptyset$*, then* $g^{**} = g$ *if and only if* g *is convex, lower semicontinuous and* $(\mathcal{C}^j)^*$*-nondecreasing.*

The next subsection is devoted to its proof. The second part deals with the case where $\mathsf{dom}\, g$ has empty interior. For this, we need more careful analysis of the structure of the boundary of $\mathcal{C}^j$. This is done in the second subsection.

6.5.2. Proof of Lemma 6.5.7

Let satisfy $\mathsf{int}\,\mathsf{dom}\, g \neq \emptyset$. It is clear from (6.2.29) that g^{**} is convex, lower semicontinuous, and $(\mathcal{C}^j)^*$-nondecreasing.

Henceforth, assuming that g is convex, lower semicontinuous, and $(\mathcal{C}^j)^*$-nondecreasing, we want to prove the converse. For convenience, we write $\Omega = \mathsf{dom}\, g$. The plan is to prove the identity $g = g^{**}$ first on $\mathsf{int}\,\Omega$, then on $\mathsf{cl}\,\Omega$, and finally on the entire $\mathcal{C}$.

Analysis on $\mathsf{int}\,\Omega$

Let $x \in \mathsf{int}\,\Omega$. By Lemma 6.5.3, we know $\partial g(x)$ is not empty. For each $v \in (\mathcal{C}^j)^*$, there is $\varepsilon > 0$ small so that $x - \varepsilon v \in \Omega$. For each $y \in \partial g(x)$, by the definition of subdifferentials and the monotonicity of g, we have

$$\langle v, y \rangle_{\mathcal{H}^j} \geqslant \frac{1}{\varepsilon}\left(g(x) - g(x - \varepsilon v)\right) \geqslant 0,$$

which implies $\emptyset \neq \partial g(x) \subseteq \mathcal{C}^j$. Invoking Lemma 6.5.5, we can deduce

$$g(x) \leqslant \sup_{y \in \mathcal{C}^j} \{\langle y, x \rangle_{\mathcal{H}^j} - g^*(y)\} = g^{**}(x).$$

On the other hand, from (6.2.29), it is easy to see that

$$g(x) \geqslant g^{**}(x), \quad \forall x \in \mathcal{C}^j. \tag{6.5.3}$$

Hence, we obtain

$$g(x) = g^{**}(x), \quad \forall x \in \mathsf{int}\,\Omega.$$

Analysis on $\mathsf{cl}\,\Omega$

Let $x \in \mathsf{cl}\,\Omega$ and choose $y \in \mathsf{int}\,\Omega$. Setting $x_\alpha = (1-\alpha)x + \alpha y$, by Lemma 6.5.2, we have $x_\alpha \in \mathsf{int}\,\Omega$ for every $\alpha \in (0, 1]$. By the result on $\mathsf{int}\,\Omega$, we have

$$g(x_\alpha) = g^{**}(x_\alpha).$$

Then, x_α belongs to $\operatorname{dom} g$ and $\operatorname{dom} g^{**}$. Applying Lemma 6.5.4 and sending $\alpha \to 0$, we get

$$g(x) = g^{**}(x), \quad \forall x \in \operatorname{cl}\Omega. \tag{6.5.4}$$

Analysis on $\mathcal{C}^j$

Due to (6.5.4), we only need to consider points outside $\operatorname{cl}\Omega$. Fixing any $x \in \mathcal{C}^j \setminus \operatorname{cl}\Omega$, we have $g(x) = \infty$. Since f is not identically equal to ∞ and $(\mathcal{C}^j)^*$-nondecreasing, we must have $0 \in \Omega$. By this, $x \notin \operatorname{cl}\Omega$ and the convexity of $\operatorname{cl}\Omega$, we must have

$$\overline{\lambda} = \sup\{\lambda \in \mathbb{R}_+ : \lambda x \in \operatorname{cl}\Omega\} < 1. \tag{6.5.5}$$

We set

$$\overline{x} = \overline{\lambda} x. \tag{6.5.6}$$

Then, we have that $\overline{x} \in \operatorname{bd}\Omega$ and $\lambda\overline{x} \notin \operatorname{cl}\Omega$ for all $\lambda > 1$.

We need to discuss two cases: $\overline{x} \in \Omega$ or not.

In the second case where $\overline{x} \notin \Omega$, we have $g(\overline{x}) = \infty$. Due to $\overline{x} \in \operatorname{cl}\Omega$ and (6.5.4), we have $g^{**}(\overline{x}) = \infty$. On the other hand, by (6.5.4) and the fact that $0 \in \Omega$, we have $g^{**}(0) = g(0)$ and thus $0 \in \operatorname{dom} g^{**}$. The convexity of g^{**} implies that

$$\infty = g^{**}(\overline{x}) \leqslant \overline{\lambda} g^{**}(x) + (1 - \overline{\lambda}) g^{**}(0).$$

Hence, we must have $g^{**}(x) = \infty$ and thus $g(x) = g^{**}(x)$ for such x.

We now consider the case where $\overline{x} \in \Omega$. For every $y \in \mathcal{H}^j$, the outer normal cone to Ω at y is defined by

$$\mathbf{n}_\Omega(y) = \{z \in \mathcal{H}^j : \langle z, y' - y\rangle_{\mathcal{H}^j} \leqslant 0, \forall y' \in \Omega\}. \tag{6.5.7}$$

We need the following result.

Lemma 6.5.8. *Assume* $\operatorname{int}\Omega \neq \emptyset$. *For every* $y \in \Omega \setminus \operatorname{int}\Omega$ *satisfying* $\lambda y \notin \operatorname{cl}\Omega$ *for all* $\lambda > 1$, *there is* $z \in \mathbf{n}_\Omega(y) \cap \mathcal{C}^j$ *such that* $\langle z, y\rangle_{\mathcal{H}^j} > 0$.

By Lemma 6.5.8 applied to $\overline{x} \in \Omega$, there is $z \in \mathcal{C}^j$ such that

$$\langle z, w - \overline{x}\rangle_{\mathcal{H}^j} \leqslant 0, \quad \forall w \in \Omega, \tag{6.5.8}$$

$$\langle z, \overline{x}\rangle_{\mathcal{H}^j} > 0. \tag{6.5.9}$$

The monotonicity of g ensures that $g(x) \geqslant g(0)$ for all $x \in \mathcal{C}^j$. For each $\rho \geqslant 0$, define

$$\mathcal{L}_\rho = L_{\rho z,\ g(0) - \rho\langle z, \overline{x}\rangle_{\mathcal{H}^j}}.$$

Due to (6.5.8), we can see that

$$\mathcal{L}_\rho(w) = \rho\,\langle z, w - \overline{x}\rangle_{\mathcal{H}^j} + g(0) \leqslant g(w), \quad \forall w \in \Omega.$$

Since we know $f\big|_{\mathcal{C}^j \setminus \Omega} = \infty$, the inequality above gives

$$\mathcal{L}_\rho \leqslant g, \quad \forall \rho \geqslant 0. \tag{6.5.10}$$

Evaluating $\mathcal{L}_\rho$ at x and using (6.5.6), we have

$$\mathcal{L}_\rho(x) = \rho\,\langle z, x - \overline{x}\rangle_{\mathcal{H}^j} + g(0) = \rho\left(\overline{\lambda}^{-1} - 1\right)\langle z, \overline{x}\rangle_{\mathcal{H}^j} + g(0).$$

By (6.5.5) and (6.5.9), we obtain

$$\lim_{\rho\to\infty} \mathcal{L}_\rho(x) = \infty.$$

This along with (6.5.10), Lemma 6.5.6 and (6.5.3) implies

$$g(x) = g^{**}(x) \quad \forall x \in \mathcal{C}^j \setminus \mathsf{cl}\,\Omega.$$

In view of this and (6.5.4), we have completed the proof of Lemma 6.5.7. It remains to prove Lemma 6.5.8.

Proof of Lemma 6.5.8. Fix y satisfying the condition. Since it is possible that $y \notin \mathsf{int}\,\mathcal{C}^j$, we want to approximate y by points in $\mathsf{bd}\,\Omega \cap \mathsf{int}\,\mathcal{C}^j$. For every open ball $B \subseteq \mathcal{H}^j$ centered at y, there is some $\lambda > 1$ such that $y' = \lambda y \in \mathcal{C}^j \cap (B \setminus \mathsf{cl}\,\Omega)$. Due to $\mathsf{int}\,\Omega \neq \emptyset$ and $y \in \Omega$, by Lemma 6.5.2, there is some $y'' \in B \cap \mathsf{int}\,\Omega \subseteq \mathsf{int}\,\mathcal{C}$. For $\rho \in [0,1]$, we set

$$y_\rho = \rho y' + (1-\rho)y'' \in B.$$

Then, we take

$$\rho_0 = \sup\{\rho \in [0,1] : y_\rho \in \mathsf{int}\,\Omega\}.$$

Since $y' \notin \mathsf{cl}\,\Omega$, we must have $\rho_0 < 1$. It can be seen that $y_{\rho_0} \in \mathsf{cl}\,\Omega \setminus \mathsf{int}\,\Omega$ and thus $y_{\rho_0} \in B \cap \mathsf{bd}\,\Omega$. Due to $y' \in \mathcal{C}^j$, $y'' \in \mathsf{int}\,\mathcal{C}^j$ and Lemma 6.5.2, we have $y_{\rho_0} \in \mathsf{int}\,\mathcal{C}^j$. In summary, we obtain $y_{\rho_0} \in B \cap \mathsf{bd}\,\Omega \cap \mathsf{int}\,\mathcal{C}^j$.

By this construction and varying the size of the open balls centered at y, we can find a sequence $(y_n)_{n=1}^{\infty}$ such that

$$y_n \in \mathsf{int}\,\mathcal{C}^j, \tag{6.5.11}$$

$$y_n \in \mathsf{bd}\,\Omega, \tag{6.5.12}$$

$$\lim_{n\to\infty} y_n = y. \tag{6.5.13}$$

Fix any n. By (6.5.11), there is $\delta > 0$ such that

$$y_n + B(0, 2\delta) \subseteq \mathcal{C}^j. \tag{6.5.14}$$

Here, for $a \in \mathcal{H}^j, r > 0$, we write $B(a,r) = \{z \in \mathcal{H}^j : \ |z - a| < r\}$. For each $\varepsilon \in (0, \delta)$, due to (6.5.12), we can also find $y_{n,\varepsilon}$ such that

$$y_{n,\varepsilon} \in \Omega, \tag{6.5.15}$$

$$|y_{n,\varepsilon} - y_n| < \varepsilon. \tag{6.5.16}$$

This and (6.5.14) imply that

$$y_{n,\varepsilon} - a \in \mathcal{C}^j, \quad \forall \varepsilon \in (0, \delta), \ a \in B(0, \delta).$$

Since g is $(\mathcal{C}^j)^*$-nondecreasing, this along with (6.5.15) implies that

$$y_{n,\varepsilon} - a \in \Omega, \quad \forall \varepsilon \in (0, \delta), \ a \in (\mathcal{C}^j)^* \cap B(0, \delta).$$

Due to (6.5.12) and $\mathsf{int}\, \Omega \neq \emptyset$, we have that $\mathbf{n}_\Omega(y_n)$ contains some nonzero vector z_n (see [23, Proposition 6.45] together with [23, Proposition 6.23 (iii)]). The definition of the outer normal cone in (6.5.7) yields

$$\langle z_n, y_{n,\varepsilon} - a - y_n \rangle_{\mathcal{H}^j} \leqslant 0,$$

which along with (6.5.16) implies

$$\langle z_n, a \rangle_{\mathcal{H}^j} \geqslant -|z_n| \varepsilon.$$

Sending $\varepsilon \to 0$ and varying $a \in (\mathcal{C}^j)^* \cap B(0, \delta)$, we conclude that

$$z_n \in \mathbf{n}_\Omega(y_n) \cap \mathcal{C}^j, \quad \forall n. \tag{6.5.17}$$

Now for each n, we rescale z_n to get $|z_n| = 1$. By passing to a subsequence, we can assume that there is $z \in \mathcal{C}^j$ such that z_n converges to z. By $z_n \in \mathbf{n}_\Omega(y_n)$, we get

$$\langle z_n, w - y_n \rangle_{\mathcal{H}^j} \leqslant 0, \quad \forall w \in \Omega.$$

The convergence of $(z_n)_{n=1}^\infty$ along with (6.5.13) implies

$$\lim_{n\to\infty} \langle z_n, w - y_n \rangle_{\mathcal{H}^j} = \langle z, w - y \rangle_{\mathcal{H}^j}, \quad \forall w \in \Omega.$$

The above two displays yield $z \in \mathbf{n}_\Omega(y) \cap \mathcal{C}^j$.

Then, we show $\langle z, y \rangle_{\mathcal{H}^j} > 0$. Fix some $x_0 \in \operatorname{int}\Omega$ and some $\varepsilon > 0$ such that $B(x_0, 2\varepsilon) \subseteq \Omega$. Let y_n and z_n be given as in the above. Due to $|z_n| = 1$, we have

$$x_0 - \varepsilon z_n \in \Omega \subseteq \mathcal{C}^j.$$

Since it is easy to see that $\mathcal{C}^j \subseteq (\mathcal{C}^j)^*$, by (6.5.17), we have $z_n \in (\mathcal{C}^j)^*$, which along with the above display implies that

$$\langle x_0 - \varepsilon z_n, z_n \rangle_{\mathcal{H}^j} \geqslant 0$$

and thus $\langle x_0, z_n \rangle_{\mathcal{H}^j} \geqslant \varepsilon$. Using $z_n \in \mathbf{n}_\Omega(y_n)$, we obtain

$$\langle y_n, z_n \rangle_{\mathcal{H}^j} \geqslant \langle x_0, z_n \rangle_{\mathcal{H}^j} \geqslant \varepsilon.$$

Passing to the limit, we conclude that $\langle z, y \rangle_{\mathcal{H}^j} > 0$, completing the proof. □

6.5.3. Proof of Proposition 6.5.1

Similar to the arguments in the beginning of the proof of Lemma 6.5.7, we only need to show the direction that $g^{**} = g$ if g is convex, lower semicontinuous and $(\mathcal{C}^j)^*$-nondecreasing. By Lemma 6.5.7, we only need to consider the case where Ω has empty interior. Recall that we have set $\Omega = \mathsf{dom}\, g$. Throughout this subsection, we assume that Ω has empty interior. We proceed in steps.

Step 1. Setting

$$N = \max\left\{\mathsf{rank}\left(x_{|j|}\right) : x \in \Omega\right\}, \tag{6.5.18}$$

we want to show $N < D$. We need the following lemma.

Lemma 6.5.9. *If there is $x \in \mathcal{C}^j$ such that $x_{|j|}$ is of full rank, then* $\mathsf{int}\,(\mathcal{C}^j \cap (x - (\mathcal{C}^j)^*)) \neq \emptyset$.

Proof. Recall the partial order induced by $\mathbb{S}^D_+$ in (6.3.1). Let x satisfy the assumption. Then, there is some constant $a > 0$ such that $x_{|j|} \geqslant aI_D$ where I_D is the $D \times D$ identity matrix. Let us define $y_k = k\delta I_D$, $k = 1, 2, \ldots, |j|$, for some $\delta > 0$ to be chosen later. Then, it is clear that $y \in \mathcal{C}^j$. We consider $B = \{z \in \mathcal{H}^j : |z_k - y_k| \leqslant r, \forall k\}$ for some $r > 0$ to be chosen later. Then, due to finite dimensionality, there is some $c > 0$ such that, for every $z \in B$,

$$-crI_D \leqslant z_k - y_k \leqslant crI_D, \quad \forall k = 1, 2, \ldots, |j|.$$

Using this, we can show that, for every $z \in B$,

$$z_k - z_{k-1} \geqslant y_k - y_{k-1} - 2crI_D = (\delta - 2cr)I_D, \quad \forall k = 1, 2, \ldots, |j|,$$

where we set $z_0 = y_0 = 0$. By choosing r sufficiently small, the above is in $\mathbb{S}^D_+$, and we have

$B \subseteq \mathcal{C}^j$. On the other hand, we also have, for $i = 1, 2, \ldots, |j|$,

$$\sum_{k=i}^{|j|}(x_k - z_k) = \sum_{k=i}^{|j|}(x_k - y_k + y_k - z_k) \geqslant \left(x_{|j|} - \sum_{k=1}^{|j|} y_k\right) + \sum_{k=i}^{|j|}(y_k - z_k)$$
$$\geqslant \left(aI_D - \sum_{k=1}^{|j|} y_k\right) - |j|crI_D = \left(a - \frac{1}{2}(1+|j|)|j|\delta - |j|cr\right) I_D,$$

which is in $\mathbb{S}_+^D$ if δ and r are chosen sufficiently small. Hence, we have $x - z \in (\mathcal{C}^j)^*$ for all $z \in B$, which is equivalent to $B \subseteq x - (\mathcal{C}^j)^*$. Since B has nonempty interior, the proof is complete. □

Since g is $(\mathcal{C}^j)^*$-nondecreasing, we have

$$\mathcal{C}^j \cap (x - (\mathcal{C}^j)^*) \subseteq \Omega, \quad \forall x \in \Omega.$$

Hence, Lemma 6.5.9 implies that if there is $x \in \Omega$ with $\mathsf{rank}(x_{|j|}) = D$, then $\mathsf{int}\,\Omega \neq \emptyset$. Therefore, under our assumption $\mathsf{int}\,\Omega = \emptyset$, we must have that $x_{|j|}$ is of rank less than D for every $x \in \Omega$. So, for N defined in (6.5.18), we must have $N < D$.

Step 2. We fix $\widetilde{x} \in \Omega$ with rank N. By changing basis, we may assume $\widetilde{x} = \mathsf{diag}(a, 0_{D-N})$ where a is a $N \times N$ diagonal matrix with positive entries and 0_{D-N} is $(D-N) \times (D-N)$ zero matrix. We set

$$\widetilde{\mathbb{S}}_+^N = \{\mathsf{diag}(a, 0_{D-N}) : a \in \mathbb{S}_+^N\} \subseteq \mathbb{S}_+^D,$$
$$\widetilde{\mathcal{C}}^j = \{x \in \mathcal{C}^j : x_k \in \widetilde{\mathbb{S}}_+^N, \forall k\}.$$

We want to show that

$$\Omega \subseteq \widetilde{\mathcal{C}}^j. \tag{6.5.19}$$

We argue by contradiction and assume that there is $y \in \Omega$ such that $y_{k'} \notin \widetilde{\mathbb{S}}_+^N$ for some

k'. Let us define $\widetilde{y}$ by $\widetilde{y}_k = y_k$ for all $k \leqslant k'$ and $\widetilde{y}_k = y_{k'}$ for all $k > k'$. We clearly have $\widetilde{y} \in \mathcal{C}^j \cap (y - (\mathcal{C}^j)^*)$ which implies that $\widetilde{y} \in \Omega$ (because g is $(\mathcal{C}^j)^*$-nondecreasing). By convexity of Ω, we must have $z = \frac{1}{2}\widetilde{x} + \frac{1}{2}\widetilde{y} \in \Omega$.

We argue that $z_{|j|}$ has rank at least $N+1$. Since $y_{k'}$ is positive semi-definite, we must have

$$(y_{k'})_{ii} > 0 \tag{6.5.20}$$

for some $i > N$. By reordering coordinates, we may assume $i = N+1$ in (6.5.20) and thus $(y_{k'})_{N+1,N+1} > 0$. Setting $\widehat{z}_{|j|} = ((z_{|j|})_{m,n})_{1 \leqslant m,n \leqslant |j|}$, it suffices to verify $v^\intercal z_{|j|} v > 0$ for all $v \in \mathbb{R}^{N+1} \setminus \{0\}$. We define $\widehat{x}_{|j|}$ and $\widehat{y}_{|j|}$ analogously. If $v_n \neq 0$ for all $n = 1, 2, \ldots, N$, we have

$$v^\intercal \widehat{z}_{|j|} v \geqslant \frac{1}{2} v^\intercal \widehat{x}_{|j|} v > 0$$

due to the fact that $\widehat{x}_{|j|} = a$ is a diagonal matrix with positive entries. If $v_n = 0$ for all $n = 1, 2, \ldots, N$, then we must have $v_{N+1} \neq 0$ and thus

$$v^\intercal \widehat{z}_{|j|} v \geqslant \frac{1}{2} v^\intercal \widehat{y}_{|j|} v = \frac{1}{2} v_{N+1}^2 (y_{|j|})_{N+1,N+1} > 0.$$

We conclude that $z_{|j|}$ has rank at least $N+1$ contradicting the definition of N. Therefore, we must have (6.5.19).

Step 3. We conclude by applying Lemma 6.5.7 to g restricted to $\widetilde{\mathcal{C}}^j$, and treating g on $\mathcal{C}^j \setminus \widetilde{\mathcal{C}}^j$ using Lemma 6.5.6.

In view of (6.5.18) and (6.5.19), applying Lemma 6.5.9 to $\widetilde{\mathcal{C}}^j$, we have that Ω has nonempty interior relative to $\widetilde{\mathcal{C}}^j$. Let $\widetilde{g}$ be the restriction of g to $\widetilde{\mathcal{C}}^j$. Define

$$\widetilde{g}^*(y) = \sup_{x \in \widetilde{\mathcal{C}}^j} \{\langle x, y \rangle_{\mathcal{H}^j} - \widetilde{g}(x)\}, \quad \forall y \in \widetilde{\mathcal{H}}^j$$

where $\widetilde{\mathcal{H}}^j = \{x \in \mathcal{H}^j : x_k \in \widetilde{\mathbb{S}}^N,\ \forall k\}$ with $\widetilde{\mathbb{S}}^N = \{\mathsf{diag}(a, 0_{K-N}) : a \in \mathbb{S}^N\}$. Since $g(x) = \infty$ for $x \notin \widetilde{\mathcal{C}}^j$ and $g = \widetilde{g}$ on $\widetilde{\mathcal{C}}^j$, we can see from the definition of g^* in (6.5.1) that

$$g^*(y) = \sup_{x \in \widetilde{\mathcal{C}}^j} \{\langle x, y\rangle_{\mathcal{H}^j} - g(x)\} = \widetilde{g^*}(y), \quad \forall y \in \widetilde{\mathcal{H}}^j,$$

which implies $g^{**}(x) \geqslant \widetilde{g^{**}}(x)$ for all $x \in \widetilde{\mathcal{C}}^j$. Since Lemma 6.5.7 implies that $\widetilde{g}(x) = \widetilde{g^{**}}(x)$ for $x \in \widetilde{\mathcal{C}}^j$, we can thus conclude that $g^{**}(x) \geqslant \widetilde{g}(x) = g(x)$ for all $x \in \widetilde{\mathcal{C}}^j$. This along with (6.5.3) yields

$$g^{**}(x) = g(x), \quad \forall x \in \widetilde{\mathcal{C}}^j. \tag{6.5.21}$$

For $x \in \mathcal{C}^j \setminus \widetilde{\mathcal{C}}^j$, arguing as above (the paragraph studying the rank of $z_{|j|}$), we can see that there is some k and some $i > N$ such that $(x_k)_{ii} > 0$. Now, setting $y_k = \mathsf{diag}(0_N, I_{K-N})$ for every k, we have $y \in \mathcal{C}^j$, $\langle y, x\rangle_{\mathcal{H}^j} > 0$ and $\langle y, z\rangle_{\mathcal{H}^j} = 0$ for all $z \in \widetilde{\mathcal{C}}^j$. We define $\mathcal{L}_\rho = \rho \langle y, \cdot\rangle_{\mathcal{H}^j} + g(0)$ for each $\rho > 0$. Since $g(z) \geqslant g(0)$ for all $z \in \mathcal{C}^j$ due to the monotonicity of g, and since $\mathcal{L}_\rho(z) = g(0)$ for all $z \in \widetilde{\mathcal{C}}^j$, we have $g(z) \geqslant \mathcal{L}_\rho(z)$ for all $z \in \widetilde{\mathcal{C}}^j$. Due to $g = \infty$ on $\mathcal{C}^j \setminus \widetilde{\mathcal{C}}^j$, we thus get

$$g(z) \geqslant \mathcal{L}_\rho(z), \quad \forall z \in \mathcal{C}^j.$$

Due to $\langle y, x\rangle_{\mathcal{H}^j} > 0$, we also have $\lim_{\rho\to\infty} \mathcal{L}_\rho(x) = \infty = g(x)$. In view of Lemma 6.5.6, this along with the above display implies that $g^{**}(x) = g(x)$ for all $x \in \mathcal{C}^j \setminus \widetilde{\mathcal{C}}^j$, which together with (6.5.21) completes the proof of Proposition 6.5.1.

CHAPTER 7

Fenchel–Moreau identities on cones

This chapter is essentially borrowed from [38], joint with Hong-Bin Chen.

Abstract. A pointed convex cone naturally induces a partial order, and further a notion of nondecreasingness for functions. We consider extended real-valued functions defined on the cone. Monotone conjugates for these functions can be defined in an analogous way to the standard convex conjugate. The only difference is that the supremum is taken over the cone instead of the entire space. We give sufficient conditions for the cone under which the corresponding Fenchel–Moreau biconjugation identity holds for proper, convex, lower semicontinuous and nondecreasing functions defined on the cone. In addition, we show that these conditions are satisfied by a class of cones known as perfect cones.

7.1. Introduction

The classical Fenchel–Moreau identity can be stated as $f = f^{**}$ for convex $f : \mathsf{H} \to (-\infty, \infty]$ satisfying a few additional regularity conditions. Here H is a Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and the convex conjugate is given by

$$f^*(x) = \sup_{y \in \mathsf{H}} \{ \langle y, x \rangle - f(y) \}, \quad \forall x \in \mathsf{H}.$$

Note that the supremum is taken over the entire space H.

On the other hand, it is well-known (c.f. [105, Theorem 12.4]) that if $f : [0, \infty)^d \to (-\infty, \infty]$ is convex with extra usual assumptions and, in addition, is nondecreasing in the sense that

$$f(x) \geqslant f(y), \quad \text{if } x - y \in [0, \infty)^d,$$

then we also have $f = f^{**}$. Here $*$ stands for the monotone conjugate defined by

$$f^*(x) = \sup_{y \in [0,\infty)^d} \{\langle y, x\rangle - f(y)\}, \quad \forall x \in [0,\infty)^d.$$

The inner product appearing above is the standard one in $\mathbb{R}^d$. The nonnegative orthant $[0,\infty)^d$ is a cone in $\mathbb{R}^d$ and the nondecreasingness can be formulated with respect to the partial order induced by this cone. Compared with the convex conjugate, the supremum above is taken over the cone.

Recently, in [36], to study a certain Hamilton–Jacobi equation with spatial variables in the set of $n \times n$ (symmetric) positive semidefinite (p.s.d.) matrices denoted by $\mathbb{S}^n_+$, a version of the Fenchel–Moreau identity on $\mathbb{S}^n_+$ is needed to verify that the unique solution admits a variational formula. The derivation of such formulae for Hamilton–Jacobi equations on entire Euclidean spaces are known and can be seen, for instance, in [15, 84]. On $\mathbb{S}^n_+$, [36, Proposition B.1] proves that $f = f^{**}$ holds if $f : \mathbb{S}^n_+ \to (-\infty, \infty]$ is convex with some usual regularity assumptions and is nondecreasing in the sense that

$$f(x) \geqslant f(y), \qquad \text{if } x - y \in \mathbb{S}^n_+.$$

Accordingly, here $*$ stands for the monotone conjugate with respect to $\mathbb{S}^n_+$ given by

$$f^*(x) = \sup_{y \in \mathbb{S}^n_+} \{\langle y, x\rangle - f(y)\}, \quad \forall x \in \mathbb{S}^n_+.$$

The inner product is the Frobenius inner product for matrices. Again, in this case, $\mathbb{S}^n_+$ can be viewed as a cone in $\mathbb{S}^n$, the space of $n \times n$ real symmetric matrices.

In view of these two examples, it is natural to pursue a generalization to an arbitrary (convex) cone $\mathcal{C}$ in a Hilbert space H. More precisely, we want to show $f = f^{**}$ for proper, lower semicontinuous and convex $f : \mathcal{C} \to (-\infty, \infty]$ which is also nondecreasing in the sense that

$$f(x) \geqslant f(y), \qquad \text{if } x - y \in \mathcal{C},$$

where

$$f^*(y) = \sup_{z \in \mathcal{C}} \{\langle z, y\rangle - f(z)\}, \quad \forall y \in \mathcal{C}^\vee, \tag{7.1.1}$$

$$f^{**}(x) = \sup_{y \in \mathcal{C}^\vee} \{\langle y, x\rangle - f^*(y)\}, \quad \forall x \in \mathcal{C}, \tag{7.1.2}$$

where $\mathcal{C}^\vee$ is the dual cone of $\mathcal{C}$.

In Theorem 7.2.2, we give sufficient conditions on $\mathcal{C}$ for $f = f^{**}$ to hold for all f satisfying the aforementioned properties. In particular, these conditions hold for a class of cones called *perfect cones* first introduced in [17] in the setting of Euclidean spaces. In short, a perfect cone is a self-dual cone satisfying that every face $\mathcal{F}$ of $\mathcal{C}$ is self-dual in the linear space spanned by $\mathcal{F}$.

The nonnegative orthant $[0, \infty)^d$ and the set of p.s.d. matrices $\mathbb{S}^n_+$ are both perfect cones. The former is easy to see using Definition 7.2.1 and the latter will be proved in Lemma 7.5.1. An example of an infinite-dimensional perfect cone is given in Lemma 7.5.3. Classical references for properties of cones and self-dual cones in Euclidean spaces or Hilbert spaces include [16, 18, 19, 24, 102]. The generality pursued in this work is also motivated by the study of Hamilton–Jacobi equations arising in mean-field disordered systems [95, 94, 92, 96, 93, 36], where the solution is defined on a set that can be identified with a cone in possibly infinite dimensions, and expected to be nondecreasing with respect to the cone.

Let us briefly comment on the connection to the theory of abstract convexity and related works. Let $\mathscr{A}$ be the collection of affine functions with slopes in $\mathcal{C}^\vee$. In view of (7.1.1) and (7.1.2), we can declare a function f on $\mathcal{C}$ to be $\mathscr{A}$-convex if f is equal to the upper envelope of all functions in $\mathscr{A}$ lying below f (see (7.3.4) and the right-hand side of (7.3.3)).

Then, by the Fenchel–Moreau theorem for abstract convexity (c.f. [73, Theorem 7.1]), the desired Fenchel–Moreau identity here is equivalent to the statement that the $\mathscr{A}$-convexity coincides with the usual notion of convexity for nondecreasing functions defined on $\mathcal{C}$. We refer to [91, 108, 73] for more details on abstract convexity. Studies of increasing functions on cones from the perspective of abstract convexity include [57, 58, 59]. The rest of the paper is organized as follows. We introduce definitions and state the main results in Section 7.2. These results will be proved in Section 7.3 and Section 7.4. Lastly, examples of perfect cones in finite dimensions and infinite dimensions are given in Section 7.5.

Acknowledgements

We thank Jean-Christophe Mourrat and Tim Hoheisel for helpful discussions. We are grateful to the anonymous referees for comments that helped to improve the paper significantly.

7.2. Definitions and main results

Let H be a real Hilbert space equipped with inner product $\langle \cdot, \cdot \rangle$ and the associated norm $|\cdot|$. We refer to an element in H sometimes as a vector, though H can be possibly infinite-dimensional. We denote the interior, the closure and the boundary relative to H by int, cl, and bd, respectively.

7.2.1. Definitions related to cones

Let $\mathcal{C}$ be a cone in H. In this work, for simplicity, we require all cones to be convex and contain the origin. Hence, $\mathcal{C}$ is a cone if and only if it satisfies

$$\alpha x + \beta y \in \mathcal{C}, \quad \forall x, y \in \mathcal{C},\ \forall \alpha, \beta \geqslant 0.$$

Naturally, $\mathcal{C}$ induces a preorder $\preceq$ on H given by

$$x \preceq y \quad \text{if and only if} \quad y - x \in \mathcal{C}.$$

We also write $x \succeq y$ if $y \preceq x$. When $\mathcal{C}$ is *pointed*, namely $\mathcal{C} \cap (-\mathcal{C}) = \{0\}$, this preorder becomes a partial order. We denote by span and $\overline{\mathsf{span}}$ the operations of taking the linear span and the closed linear span, respectively. The dual of $\mathcal{C}$ with respect to $\overline{\mathsf{span}}\,\mathcal{C}$ is given by

$$\mathcal{C}^\vee = \{x \in \overline{\mathsf{span}}\,\mathcal{C} : \ \langle x, y\rangle \geqslant 0, \, \forall y \in \mathcal{C}\}. \tag{7.2.1}$$

The cone $\mathcal{C}$ is said to be self-dual (with respect to $\overline{\mathsf{span}}\,\mathcal{C}$) provided $\mathcal{C} = \mathcal{C}^\vee$. It is clear that a self-dual cone is closed and pointed.

A subset $\mathcal{F}$ of a closed and pointed cone $\mathcal{C}$ is a *face* of $\mathcal{C}$ if $\mathcal{F}$ is a cone and satisfies that

$$\text{if } \ 0 \preceq x \preceq y \ \text{ and } \ y \in \mathcal{F}, \qquad \text{then } x \in \mathcal{F}. \tag{7.2.2}$$

Denote by $\mathcal{F}^\vee$ the dual cone of $\mathcal{F}$ in the space $\overline{\mathsf{span}}\,\mathcal{F}$. The following definition is a generalization of [17, Definition 4] from Euclidean spaces to Hilbert spaces.

Definition 7.2.1. A cone $\mathcal{C}$ is said to be *perfect* if it is self-dual and every face $\mathcal{F}$ of $\mathcal{C}$ satisfies

1. $\mathcal{F}^\vee = \mathcal{F}$;
2. $\mathcal{F}$ has nonempty interior with respect to $\overline{\mathsf{span}}\,\mathcal{F}$.

In other words, $\mathcal{F}^\vee = \mathcal{F}$ means $\mathcal{F}$ is self-dual in its own closed span. Since $\mathcal{C}$ is itself a face, a perfect cone satisfying $\overline{\mathsf{span}}\,\mathcal{C} = \mathsf{H}$ must have nonempty interior. In finite-dimensions, a self-dual cone always has nonempty interior in its own span (c.f. [23, Exercise 6.15]). Hence, if H is finite-dimensional, then (2) automatically follows from (1). Compared with [17, Definition 3] where only (1) is imposed, condition (2) is added to ensure this non-degeneracy in infinite dimensions. In Section 7.5, we give two examples of perfect cones, a finite-dimensional one and an infinite-dimensional one.

7.2.2. Definitions related to functions

The domain of a function $f : \mathcal{C} \to (-\infty, \infty]$ is defined as

$$\operatorname{dom} f = \{x \in \mathcal{C} : \; f(x) < \infty\}. \tag{7.2.3}$$

A function $f : \mathcal{C} \to (-\infty, \infty]$ is said to be $\mathcal{C}$-*nondecreasing* provided

$$f(x) \geqslant f(y), \quad \forall x \succeq y \succeq 0.$$

For any $f : \mathcal{C} \to (-\infty, \infty]$, we define the *monotone conjugate* of f by (7.1.1) and the *monotone biconjugate* of f by (7.1.2). Lastly, f is said to be *proper* if f is not identically equal to ∞. We denote by $\Gamma_{\nearrow}(\mathcal{C})$ the collection of functions on $\mathcal{C}$ with values in $(-\infty, \infty]$ that are proper, convex, lower semicontinuous (l.s.c.), and $\mathcal{C}$-nondecreasing.

7.2.3. Main results

For any closed subspace $\mathsf{H}' \subseteq \mathsf{H}$, we denote by $\mathbb{P}_{\mathsf{H}'}$ the orthogonal projection onto H'.

Theorem 7.2.2. *Assume that*

(H1) $\mathcal{C} \subseteq \mathsf{H}$ *is a closed and pointed cone satisfying* $\overline{\operatorname{span}}\, \mathcal{C} = \mathsf{H}$*;*

(H2) every face $\mathcal{F}$ *of* $\mathcal{C}$ *is closed and has nonempty interior with respect to* $\overline{\operatorname{span}}\, \mathcal{F}$*;*

(H3) for every face $\mathcal{F}$ *of* $\mathcal{C}$*, the dual cone* $\mathcal{F}^{\vee}$ *of* $\mathcal{F}$ *in the space* $\overline{\operatorname{span}}\, \mathcal{F}$ *is contained in* $\mathbb{P}_{\overline{\operatorname{span}}\, \mathcal{F}}\,(\mathcal{C}^{\vee})$.

Let $f : \mathcal{C} \to (-\infty, \infty]$ *be proper. Then,* $f = f^{**}$ *if and only if* $f \in \Gamma_{\nearrow}(\mathcal{C})$.

If $f = f^{**}$, then it is easy to see $f \in \Gamma_{\nearrow}(\mathcal{C})$ necessarily. The nontrivial part is the sufficient condition for $f = f^{**}$. As a special case, the following holds.

Corollary 7.2.3. *Suppose that* $\mathcal{C}$ *is a perfect cone satisfying* $\overline{\operatorname{span}}\, \mathcal{C} = \mathsf{H}$*. Let* $f : \mathcal{C} \to (-\infty, \infty]$ *be proper. Then,* $f = f^{**}$ *if and only if* $f \in \Gamma_{\nearrow}(\mathcal{C})$.

Let us briefly comment on hypotheses (H1)–(H3). The first two hypotheses are natural. Note that since $\mathcal{C}$ is itself a face, (H2) ensures that $\mathcal{C}$ has nonempty interior. In finite dimensions, given that $\mathcal{C}$ is closed, every face $\mathcal{F}$ is automatically closed (see [105, Corollary 18.1.1]), and the second half of (H2) also holds. Hence, (H1) implies (H2) in finite dimensions. Lastly, the proposition below shows that (H3) is nearly sharp when H is finite-dimensional.

Proposition 7.2.4. *Assume (H1) and that* H *is finite-dimensional. If* $f = f^{**}$ *for all* $f \in \Gamma_{\nearrow}(\mathcal{C})$*, then every face* $\mathcal{F}$ *of* $\mathcal{C}$ *satisfies* $\mathcal{F}^{\vee} \subseteq \mathsf{cl}\,(\mathbb{P}_{\mathsf{span}\,\mathcal{F}}(\mathcal{C}^{\vee}))$.

We believe that our results can be extended to more general scenarios. Here, we stick to the current setting for simplicity of presentation.

7.3. Preliminaries

In the first part of this section, we state some basic results that are needed throughout this work. In the second part, we prove Proposition 7.2.4. In the last part, we prove the following result.

Proposition 7.3.1. *Suppose that* $\mathcal{C}$ *is closed and pointed. Let* $f : \mathcal{C} \to (-\infty, \infty]$ *satisfy* $\mathsf{int}\,\mathsf{dom}\, f \neq \emptyset$*. Then* $f = f^{**}$ *if and only if* $f \in \Gamma_{\nearrow}(\mathcal{C})$*.*

7.3.1. Basic results of convex analysis

For $a \in \mathsf{H}$ and $\nu \in \mathbb{R}$, we define the affine function $L_{a,\nu}$ with slope a and translation ν by

$$L_{a,\nu}(x) = \langle a, x \rangle + \nu, \quad \forall x \in \mathsf{H}. \tag{7.3.1}$$

For a function $f : \mathbb{E} \to (-\infty, \infty]$ defined on a subset $\mathbb{E} \subseteq \mathsf{H}$, we can extend it in the standard way to $f : \mathsf{H} \to (-\infty, \infty]$ by setting $f(x) = \infty$ for $x \notin \mathbb{E}$. For $f : \mathsf{H} \to (-\infty, \infty]$, we define its domain by

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

Note that by the standard extension, the above definition is equivalent to (7.2.3) where only functions defined on $\mathcal{C}$ are considered. Henceforth, we shall not distinguish functions defined on $\mathcal{C}$ from their standard extensions to H. Denote by $\Gamma_0(\mathbb{E})$ the collection of proper, convex and l.s.c. functions from $\mathbb{E} \subseteq \mathsf{H}$ to $(-\infty, \infty]$. In particular, when $\mathcal{C}$ is closed, the collection $\Gamma_{\nearrow}(\mathcal{C}) \subseteq \Gamma_0(\mathcal{C})$ can be viewed as a subcollection of $\Gamma_0(\mathsf{H})$.

For $f : \mathsf{H} \to (-\infty, \infty]$ and each $x \in \mathsf{H}$, we define the subdifferential of f at x by

$$\partial f(x) = \big\{u \in \mathsf{H} : f(y) \geqslant f(x) + \langle y - x, u \rangle \,,\, \forall y \in \mathsf{H}\big\}. \tag{7.3.2}$$

The effective domain of ∂f is defined to be

$$\mathsf{dom}\, \partial f = \big\{x \in \mathsf{H} :\ \partial f(x) \neq \emptyset\big\}.$$

We now list some lemmas needed in our proofs.

Lemma 7.3.2. *For a convex set $A \subseteq \mathsf{H}$, if $y \in \mathsf{cl}\, A$ and $y' \in \mathsf{int}\, A$, then $\lambda y + (1-\lambda)y' \in \mathsf{int}\, A$ for all $\lambda \in [0, 1)$.*

Lemma 7.3.3. *For $f \in \Gamma_0(\mathsf{H})$, it holds that $\mathsf{int}\,\mathsf{dom}\, f \subseteq \mathsf{dom}\, \partial f \subseteq \mathsf{dom}\, f$.*

Lemma 7.3.4. *Let $f \in \Gamma_0(\mathsf{H})$, $x \in \mathsf{H}$ and $y \in \mathsf{dom}\, f$. For every $\alpha \in (0, 1)$, set $x_\alpha = (1-\alpha)x + \alpha y$. Then $\lim_{\alpha \to 0} f(x_\alpha) = f(x)$.*

Lemma 7.3.5. *Suppose that $\mathcal{C}$ is closed. Let $f \in \Gamma_0(\mathcal{C})$, $x \in \mathcal{C}$ and $u \in \mathcal{C}^\vee$. If $u \in \partial f(x)$, then $f^*(u) = \langle x, u \rangle - f(x)$.*

Lemma 7.3.6. *For $f \in \Gamma_0(\mathcal{C})$ and $x \in \mathcal{C}$, we have*

$$f^{**}(x) = \sup L_{a,\nu}(x) \tag{7.3.3}$$

where the supremum is taken over

$$\{(a,\nu)\in\mathcal{C}^\vee\times\mathbb{R} : L_{a,\nu}\leqslant f \ \text{ on } \mathcal{C}\}. \tag{7.3.4}$$

Lemmas 7.3.2, 7.3.3, and 7.3.4 can be found in [23] as Propositions 3.35, 16.21, and 9.14, respectively. For completeness, let us quickly prove Lemma 7.3.5 and Lemma 7.3.6.

Proof of Lemma 7.3.5. By the standard extension, we have $f\in\Gamma_0(\mathsf{H})$. Invoking [23, Theorem 16.23], it is classically known that

$$\sup_{z\in\mathsf{H}}\big\{\,\langle z,u\rangle - f(z)\big\} = \langle x,u\rangle - f(x).$$

By assumption, we know $x\in\operatorname{\mathsf{dom}}\partial f$. Hence, Lemma 7.3.3 implies $x\in\operatorname{\mathsf{dom}} f$ and thus the right hand side of the above equation is finite. Then, the supremum on the left must also be finite. On the other hand, by the extension, we have $f(z)=\infty$ if $z\notin\mathcal{C}$, which yields

$$\sup_{z\in\mathsf{H}}\big\{\,\langle z,u\rangle - f(z)\big\} = \sup_{z\in\mathcal{C}}\big\{\,\langle z,u\rangle - f(z)\big\} = f^*(u).$$

□

Proof of Lemma 7.3.6. For each $y\in\mathcal{C}^\vee$,

$$L_{y,\,-f^*(y)}(x) = \langle y,x\rangle - f^*(y),\quad \forall x\in\mathcal{C}.$$

is an affine function with slope $y\in\mathcal{C}^\vee$. By (7.1.1), we can see that $L_{y,\,-f^*(y)}\leqslant f$ on $\mathcal{C}$. In view of the definition of f^{**} in (7.1.2), we have $f^{**}(x)\leqslant \sup L_{a,\nu}(x)$ for all $x\in\mathcal{C}$ where the sup is taken over the collection in (7.3.4).

For the other direction, for $L_{a,\nu}$ satisfying $a \in \mathcal{C}^\vee$ and $L_{a,\nu} \leqslant f$, we have

$$\langle a, x \rangle + \nu \leqslant f(x), \quad \forall x \in \mathcal{C}.$$

Rearranging and taking supremum in $x \in \mathcal{C}$, we get $f^*(a) \leqslant -\nu$. This yields

$$L_{a,\nu}(x) \leqslant \langle a, x \rangle - f^*(a) \leqslant f^{**}(x),$$

which implies $\sup L_{a,\nu}(x) \leqslant f^{**}(x)$. □

7.3.2. Proof of Proposition 7.2.4

We first prove the following lemma.

Lemma 7.3.7. *For $w \in \mathcal{F}^\vee$, the function $f : \mathcal{C} \to \mathbb{R} \cup \{\infty\}$ given by*

$$f(x) = \begin{cases} \langle w, x \rangle & \text{if } x \in \mathcal{F}, \\ \infty & \text{if } x \notin \mathcal{F}. \end{cases}$$

belongs to $\Gamma_{\nearrow}(\mathcal{C})$. Moreover, if $f = f^{**}$, then

$$\langle w, x \rangle = \sup \langle v, x \rangle, \quad \forall x \in \mathcal{F} \tag{7.3.5}$$

where the supremum is taken over

$$\{v \in \mathsf{cl}\,(\mathbb{P}_{\overline{\mathsf{span}}\,\mathcal{F}}(\mathcal{C}^\vee)) : w - v \in \mathcal{F}^\vee\}. \tag{7.3.6}$$

Proof. It is clear that f is convex, proper, and l.s.c. To show f is $\mathcal{C}$-nondecreasing, let $0 \preceq x \preceq y$. Note that this implies $0 \preceq y - x \preceq y$. If $y \in \mathcal{F}$, by (7.2.2) in the definition of faces, we have $x \in \mathcal{F}$ and $y - x \in \mathcal{F}$. This along with $w \in \mathcal{F}^\vee$ yields $f(x) \leqslant f(y)$. If $y \notin \mathcal{F}$, then $f(x) \leqslant \infty = f(y)$. This verifies $f \in \Gamma_{\nearrow}(\mathcal{C})$.

Now, we want to show (7.3.5). By Lemma 7.3.6,

$$f(x) = \sup\{\langle a, x\rangle + \nu\}, \quad \forall x \in \mathcal{F} \tag{7.3.7}$$

where the supremum is taken over all $(a, \nu) \in \mathcal{C}^\vee \times \mathbb{R}$ satisfying

$$\langle w, y\rangle \geqslant \langle a, y\rangle + \nu, \quad \forall y \in \mathcal{F},$$

which is equivalent to

$$\langle w - a, \lambda y\rangle \geqslant \nu, \quad \forall \lambda \geqslant 0,\ y \in \mathcal{F}.$$

Setting $\lambda = 0$ and $\lambda \to \infty$ yield, respectively,

$$\nu \leqslant 0, \qquad \text{and} \qquad \langle w - a, y\rangle \geqslant 0, \quad \forall y \in \mathcal{F}.$$

For every such (a, ν), setting $v = \mathbb{P}_{\overline{\mathsf{span}}\,\mathcal{F}}(a)$ (which gives $\langle v, y\rangle = \langle a, y\rangle$ for all $y \in \mathcal{F}$), we thus obtain

$$\langle w, y\rangle \geqslant \langle v, y\rangle \geqslant \langle a, y\rangle + \nu, \quad \forall y \in \mathcal{F}.$$

In particular, this implies that v belongs to the set in (7.3.6). Hence, in view of (7.3.7), we conclude

$$f(x) \leqslant \sup \langle v, x\rangle, \quad \forall x \in \mathcal{F}$$

where the supremum is over the set in (7.3.6). On the other hand, for every v in the set in (7.3.6), we have $f(x) = \langle w, x\rangle \geqslant \langle v, x\rangle$ for all $x \in \mathcal{F}$. This completes the proof of (7.3.5). □

Now, we are ready to prove Proposition 7.2.4. Since H is finite-dimensional, we have $\mathsf{span}\,\mathcal{F} = \overline{\mathsf{span}}\,\mathcal{F}$. We argue by contradiction and assume that there is $w \in \mathcal{F}^\vee \setminus \mathsf{cl}\left(\mathbb{P}_{\mathsf{span}\,\mathcal{F}}(\mathcal{C}^\vee)\right)$. Then, by separation theorems, there are $\varepsilon > 0$ and $z \in \operatorname{span}\mathcal{F}$ such that

$$\langle w, z\rangle \geqslant \langle u, z\rangle + \varepsilon, \quad \forall u \in \mathbb{P}_{\operatorname{span}\mathcal{F}}(\mathcal{C}^\vee). \tag{7.3.8}$$

By [23, Proposition 6.4 (i)] and the fact that $\mathcal{F}$ is a cone, we have $\operatorname{span}\mathcal{F} = \mathcal{F} - \mathcal{F}$. Hence, there are $x, y \in \mathcal{F}$ such that $z = x - y$. By Lemma 7.3.7, we can find v from the set in (7.3.6) such that

$$\langle w, x\rangle < \langle v, x\rangle + \varepsilon.$$

On the other hand, since $\langle w, y\rangle \geqslant \langle v, y\rangle$ due to (7.3.6), we obtain from (7.3.8) that

$$\langle w, x\rangle \geqslant \langle v, x\rangle + \varepsilon,$$

contradicting the previous display.

7.3.3. Proof of Proposition 7.3.1

Let $f : \mathcal{C} \to (-\infty, \infty]$ be proper and satisfy $\operatorname{int}\operatorname{dom} f \neq \emptyset$. It is clear from (7.1.2) that f^{**} is convex, l.s.c., and $\mathcal{C}$-nondecreasing. Therefore, assuming $f = f^{**}$ and that f is proper, we have $f \in \Gamma_{\nearrow}(\mathcal{C})$.

From now on, we assume $f \in \Gamma_{\nearrow}(\mathcal{C})$ and prove the converse. For convenience, we write $\Omega = \operatorname{dom} f$. The plan is to prove the identity $f = f^{**}$ first on $\operatorname{int}\Omega$, then on $\operatorname{cl}\Omega$, and finally on the entire $\mathcal{C}$.

Analysis on $\operatorname{int}\Omega$

Let $x \in \operatorname{int}\Omega$. By Lemma 7.3.3, we know $\partial f(x)$ is not empty. For each $v \in \mathcal{C}$, there is $\varepsilon > 0$ small so that $x - \varepsilon v \in \Omega$. For each $u \in \partial f(x)$, by the definition of subdifferentials and nondecreasingness, we have

$$\langle v, u\rangle \geqslant \frac{1}{\varepsilon}\big(f(x) - f(x - \varepsilon v)\big) \geqslant 0,$$

which implies

$$\emptyset \neq \partial f(x) \subseteq \mathcal{C}^{\vee}, \quad \forall x \in \operatorname{int} \Omega. \tag{7.3.9}$$

Invoking Lemma 7.3.5, from (7.3.9) we can deduce

$$f(x) \leqslant \sup_{y \in \mathcal{C}^{\vee}} \{\langle y, x \rangle - f^*(y)\} = f^{**}(x).$$

On the other hand, from (7.1.2), it is easy to see that

$$f(x) \geqslant f^{**}(x), \quad \forall x \in \mathcal{C}. \tag{7.3.10}$$

Hence, we obtain

$$f(x) = f^{**}(x), \quad \forall x \in \operatorname{int} \Omega.$$

Analysis on $\operatorname{cl} \Omega$

Let $x \in \operatorname{cl} \Omega$ and choose $y \in \operatorname{int} \Omega$. Then, $x_\alpha = (1-\alpha)x + \alpha y$ belongs to $\operatorname{int} \Omega$ for every $\alpha \in (0, 1]$ due to Lemma 7.3.2. By the result on $\operatorname{int} \Omega$, we have

$$f(x_\alpha) = f^{**}(x_\alpha).$$

Then, x_α belongs to $\operatorname{dom} f$ and $\operatorname{dom} f^{**}$. Applying Lemma 7.3.4 and sending $\alpha \to 0$, we conclude that

$$f(x) = f^{**}(x), \quad \forall x \in \operatorname{cl} \Omega. \tag{7.3.11}$$

Analysis on $\mathcal{C}$

Due to (7.3.11), we only need to consider vectors outside $\mathsf{cl}\,\Omega$. Let $x \in \mathcal{C} \setminus \mathsf{cl}\,\Omega$, and we have $f(x) = \infty$. Since f is proper and $\mathcal{C}$-nondecreasing, we must have $0 \in \Omega$. By this, $x \notin \mathsf{cl}\,\Omega$ and the convexity of $\mathsf{cl}\,\Omega$, we must have

$$\lambda' = \sup\{\lambda \in [0, \infty) : \lambda x \in \mathsf{cl}\,\Omega\} < 1. \tag{7.3.12}$$

We set

$$x' = \lambda' x. \tag{7.3.13}$$

Then, we have that $x' \in \mathsf{bd}\,\Omega$ and $\lambda x' \notin \mathsf{cl}\,\Omega$ for all $\lambda > 1$.

We need to discuss two cases: $x' \in \Omega$ or not.

In the second case where $x' \notin \Omega$, we have $f(x') = \infty$. Due to $x' \in \mathsf{cl}\,\Omega$ and (7.3.11), we have $f^{**}(x') = \infty$. On the other hand, by (7.3.11) and the fact that $0 \in \Omega$, we have $f^{**}(0) = f(0)$ and thus $0 \in \mathsf{dom}\, f^{**}$. The convexity of f^{**} implies that

$$\infty = f^{**}(x') \leqslant \lambda' f^{**}(x) + (1 - \lambda') f^{**}(0).$$

Hence, we must have $f^{**}(x) = \infty$ and thus $f(x) = f^{**}(x)$ for such x.

We now consider the case where $x' \in \Omega$. For every $y \in \mathsf{H}$, the outer normal cone to Ω at y is defined by

$$\mathbf{n}(y) = \{z \in \mathsf{H} : \langle z, y' - y\rangle \leqslant 0, \forall y' \in \Omega\}. \tag{7.3.14}$$

We need the following result.

Lemma 7.3.8. *Assume* $\mathsf{int}\,\Omega \neq \emptyset$. *For every* $y \in \Omega \setminus \mathsf{int}\,\Omega$ *satisfying* $\lambda y \notin \mathsf{cl}\,\Omega$ *for all* $\lambda > 1$, *there is* $z \in \mathbf{n}(y) \cap \mathcal{C}^\vee$ *such that* $\langle z, y\rangle > 0$.

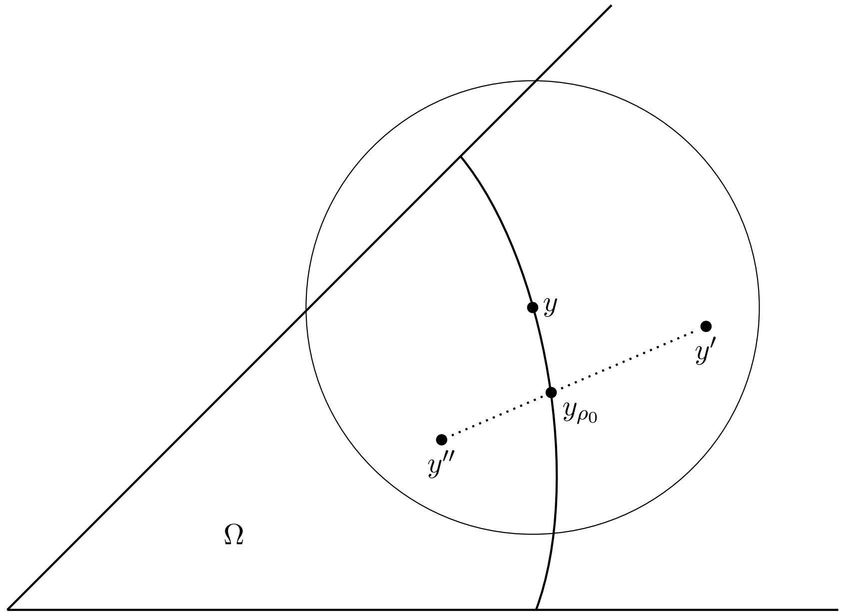

Figure 7.1: Construction of y_{ρ_0}.

Proof. Fix y satisfying the condition. It can happen that $y \notin \mathsf{int}\,\mathcal{C}$, and we want to approximate y by a point in $\mathsf{bd}\,\Omega \cap \mathsf{int}\,\mathcal{C}$. The following construction is illustrated in Figure 7.1. For every open ball $B \subseteq \mathsf{H}$ centered at y, there is some $\lambda > 1$ such that $y' = \lambda y \in \mathcal{C} \cap (B \setminus \mathsf{cl}\,\Omega)$. Due to $\mathsf{int}\,\Omega \neq \emptyset$ and $y \in \Omega$, by Lemma 7.3.2, there is some $y'' \in B \cap \mathsf{int}\,\Omega \subseteq \mathsf{int}\,\mathcal{C}$. For $\rho \in [0,1]$, we set

$$y_\rho = \rho y' + (1-\rho)y'' \in B.$$

Then, we take

$$\rho_0 = \sup\{\rho \in [0,1] : y_\rho \in \mathsf{int}\,\Omega\}.$$

Since $y' \notin \mathsf{cl}\,\Omega$, we must have $\rho_0 < 1$. It can be seen that $y_{\rho_0} \in \mathsf{cl}\,\Omega \setminus \mathsf{int}\,\Omega$ and thus $y_{\rho_0} \in B \cap \mathsf{bd}\,\Omega$. Due to $y' \in \mathcal{C}$, $y'' \in \mathsf{int}\,\mathcal{C}$ and Lemma 7.3.2, we have $y_{\rho_0} \in \mathsf{int}\,\mathcal{C}$. In summary, we obtain $y_{\rho_0} \in B \cap \mathsf{bd}\,\Omega \cap \mathsf{int}\,\mathcal{C}$.

By this construction and varying the size of the open balls centered at y, we can find a sequence $\{y_n\}_{n=1}^{\infty}$ such that

$$y_n \in \operatorname{int} \mathcal{C}, \tag{7.3.15}$$

$$y_n \in \operatorname{bd} \Omega, \tag{7.3.16}$$

$$\lim_{n\to\infty} y_n = y. \tag{7.3.17}$$

Fix any n. By (7.3.15), there is $\delta > 0$ such that

$$y_n + B(0, 2\delta) \subseteq \mathcal{C}. \tag{7.3.18}$$

Here, for $a \in \mathsf{H}, r > 0$, we write $B(a, r) = \{z \in \mathsf{H} : \ |z - a| < r\}$. For each $\varepsilon \in (0, \delta)$, due to (7.3.16), we can also find $y_{n,\varepsilon}$ such that

$$y_{n,\varepsilon} \in \Omega, \tag{7.3.19}$$

$$|y_{n,\varepsilon} - y_n| < \varepsilon. \tag{7.3.20}$$

This and (7.3.18) imply that

$$y_{n,\varepsilon} - a \in \mathcal{C}, \quad \forall \varepsilon \in (0, \delta),\ a \in B(0, \delta). \tag{7.3.21}$$

By $\mathcal{C}$-nondecreasingness, (7.3.19) and (7.3.21), we can see

$$y_{n,\varepsilon} - a \in \Omega, \quad \forall \varepsilon \in (0, \delta),\ a \in \mathcal{C} \cap B(0, \delta).$$

Due to (7.3.16) and $\operatorname{int} \Omega \neq \emptyset$, we have that $\mathbf{n}(y_n)$ contains some nonzero vector z_n (see [23, Proposition 6.45] together with [23, Proposition 6.23 (iii)]). The definition of the outer normal cone in (7.3.14) yields

$$\langle z_n, y_{n,\varepsilon} - a - y_n \rangle \leqslant 0,$$

which along with (7.3.20) implies

$$\langle z_n, a \rangle \geqslant -|z_n|\varepsilon.$$

Sending $\varepsilon \to 0$ and varying $a \in \mathcal{C} \cap B(0, \delta)$, we conclude that

$$z_n \in \mathbf{n}(y_n) \cap \mathcal{C}^\vee, \quad \forall n.$$

Now for each n, we rescale z_n to get $|z_n| = 1$. Since $\mathcal{C}^\vee \cap \mathsf{cl}\, B(0,1)$ is convex, closed, and bounded, invoking the Banach–Alaoglu–Bourbaki theorem and the Eberlein–Šmulian theorem, by passing to a subsequence, we can assume that there is $z \in \mathcal{C}^\vee$ such that z_n converges weakly to z. By $z_n \in \mathbf{n}(y_n)$, we get

$$\langle z_n, w - y_n \rangle \leqslant 0, \quad \forall w \in \Omega. \tag{7.3.22}$$

The weak convergence of $\{z_n\}_{n=1}^\infty$ along with the strong convergence in (7.3.17) implies

$$\lim_{n\to\infty} \langle z_n, w - y_n \rangle = \langle z, w - y \rangle, \quad \forall w \in \Omega.$$

The above two displays yield $z \in \mathbf{n}(y) \cap \mathcal{C}^\vee$.

Then, we show $\langle z, y \rangle > 0$. Fix some $x_0 \in \mathsf{int}\, \Omega$ and some $\varepsilon > 0$ such that $B(x_0, 2\varepsilon) \subseteq \Omega$. Let y_n and z_n be given as in the above. Due to $|z_n| = 1$, we have

$$x_0 - \varepsilon z_n \in \Omega \subseteq \mathcal{C},$$

which along with the fact that $z_n \in \mathcal{C}^\vee$ implies that

$$\langle x_0 - \varepsilon z_n, z_n \rangle \geqslant 0$$

and thus $\langle x_0, z_n \rangle \geqslant \varepsilon$. Using $z_n \in \mathbf{n}(y_n)$, we obtain

$$\langle y_n, z_n \rangle \geqslant \langle x_0, z_n \rangle \geqslant \varepsilon.$$

Passing to the limit, we conclude that $\langle z, y \rangle > 0$ completing the proof. $\square$

We now go back to our main proof and apply Lemma 7.3.8 to $x' \in \Omega$. Hence, there is $z \in \mathcal{C}^\vee$ such that

$$\left\langle z, w - x' \right\rangle \leqslant 0, \quad \forall w \in \Omega, \tag{7.3.23}$$

$$\left\langle z, x' \right\rangle > 0. \tag{7.3.24}$$

By (7.3.11) and Lemma 7.3.6 (or the simple fact that $f \geqslant f(0)$), there is an affine function $L_{a,\nu}$ with $a \in \mathcal{C}^\vee$ and $\nu \in \mathbb{R}$ such that $f \geqslant L_{a,\nu}$. For each $\rho \geqslant 0$, define

$$\mathcal{L}_\rho = L_{a+\rho z,\ \nu - \rho\langle z, x' \rangle}.$$

Due to (7.3.23), we can see that

$$\mathcal{L}_\rho(w) = L_{a,\nu}(w) + \rho \left\langle z, w - x' \right\rangle \leqslant L_{a,\nu}(w) \leqslant f(w), \quad \forall w \in \Omega.$$

Since we know $f\big|_{\mathcal{C}\setminus\Omega} = \infty$, the inequality above gives us

$$\mathcal{L}_\rho \leqslant f, \quad \forall \rho \geqslant 0. \tag{7.3.25}$$

Evaluating $\mathcal{L}_\rho$ at x and using (7.3.13), we have

$$\mathcal{L}_\rho(x) = L_{a,\nu}(x) + \rho \left\langle z, x - x' \right\rangle = L_{a,\nu}(x) + \rho(\lambda'^{-1} - 1) \left\langle z, x' \right\rangle .$$

By (7.3.12) and (7.3.24), we obtain

$$\lim_{\rho \to \infty} \mathcal{L}_\rho(x) = \infty.$$

This along with (7.3.25), Lemma 7.3.6 and (7.3.10) implies

$$f(x) = f^{**}(x) \quad \forall x \in \mathcal{C} \setminus \mathsf{cl}\, \Omega.$$

In view of this and (7.3.11), we have completed the proof of Proposition 7.3.1.

7.4. Proof of Theorem 7.2.2

We devote this section to the proof of Theorem 7.2.2. As commented in the beginning of the proof of Proposition 7.3.1 in Section 7.3.3, assuming $f = f^{**}$, we have $f \in \Gamma_{\nearrow}(\mathcal{C})$.

Now, assuming (H1)–(H3) and $f \in \Gamma_{\nearrow}(\mathcal{C})$, we want to prove $f = f^{**}$. Again, we write $\Omega = \mathsf{dom}\, f$ which is a nonempty subset of $\mathcal{C}$. Let us introduce

$$\mathcal{F}_\Omega = \left\{ \lambda y : \ \lambda \geqslant 0,\, y \in \Omega \right\}.$$

We will first show that $f = f^{**}$ holds on $\mathcal{F}_\Omega$ and then on $\mathcal{C}$.

7.4.1. Identity on $\mathcal{F}_\Omega$

We prove $f = f^{**}$ on $\mathcal{F}_\Omega$. The idea is to show Ω has nonempty interior relative to $\mathcal{F}_\Omega$ and apply Proposition 7.3.1 to f restricted to $\mathcal{F}_\Omega$. Some properties of $\mathcal{F}_\Omega$ are needed and they are stated and proved in the two lemmas below.

Lemma 7.4.1. *The set $\mathcal{F}_\Omega$ is a face of $\mathcal{C}$.*

Proof. Recall the definition of a face above Definition 7.2.1. Since in this work, we require cones to be convex, to show $\mathcal{F}_\Omega$ is a face, we start by checking it is convex. Note that for any $x_1, x_2 \in \mathcal{F}_\Omega$, there are $\lambda_1, \lambda_2 \geqslant 0$ and $y_1, y_2 \in \Omega$ such that $x_i = \lambda_i y_i$ for $i = 1, 2$. We can choose $\mu > 0$ large enough so that $\frac{\lambda_i}{\mu} y_i \preceq y_i$ for both i. Hence, by the $\mathcal{C}$-nondecreasingness of f, we have $\frac{\lambda_i}{\mu} y_i \in \Omega$ for both i. Then, for each $\alpha \in [0, 1]$, it holds that

$$\alpha x_1 + (1 - \alpha) x_2 = \mu \left(\alpha \frac{\lambda_1}{\mu} y_1 + (1 - \alpha) \frac{\lambda_2}{\mu} y_2 \right).$$

By the convexity of Ω, we have $\alpha \frac{\lambda_1}{\mu} y_1 + (1 - \alpha) \frac{\lambda_2}{\mu} y_2 \in \Omega$. Hence, we conclude that $\alpha x_1 + (1 - \alpha) x_2 \in \mathcal{F}_\Omega$, which implies that $\mathcal{F}_\Omega$ is convex. Then, it is easy to see $\mathcal{F}_\Omega$ is a cone.

Now let $0 \preceq x \preceq y$ and $y \in \mathcal{F}_\Omega$. By definition, there is $\mu > 0$ such that $\mu y \in \Omega$. We can deduce that $0 \preceq \mu x \preceq \mu y$. Again, the $\mathcal{C}$-nondecreasingness implies $\mu x \in \Omega$ and thus $x \in \mathcal{F}_\Omega$. □

Lemma 7.4.2. *Assume (H1) and (H2). The subset Ω has nonempty interior with respect to the space $\overline{\mathsf{span}}\, \mathcal{F}_\Omega$.*

Proof. For positive integers $m, n \in \mathbb{N}_+$, we set $E_{m,n} = \{mx \in \mathsf{H} : f(x) \leqslant n\}$ which is the level set $\{f \leqslant n\}$ scaled by m. We want to show

$$\mathcal{F}_\Omega = \bigcup_{m,n \in \mathbb{N}_+} E_{m,n}. \tag{7.4.1}$$

For each $x \in \mathcal{F}_\Omega$, there is $\mu > 0$ such that $y = \mu x \in \Omega$. Then, there is $n \in \mathbb{N}_+$ such that $y \in \{f \leqslant n\}$. Choose $m \in \mathbb{N}$ to satisfy $m\mu \geqslant 1$. Since f is $\mathcal{C}$-nondecreasing and $0 \preceq \frac{1}{m\mu} y \preceq y$, it yields that $\frac{1}{m\mu} y \in \{f \leqslant n\}$, which implies that $x \in E_{m,n}$. The other direction is easy by the definition of $\mathcal{F}_\Omega$. Therefore, we have verified (7.4.1).

Since f is l.s.c., we know that every $E_{m,n}$ is closed. As a closed subspace of H, the space $\overline{\mathsf{span}}\, \mathcal{F}_\Omega$ is complete. On the other hand, by (H1) and (H2), the face $\mathcal{F}_\Omega$ also has nonempty interior in $\overline{\mathsf{span}}\,\mathcal{F}_\Omega$. Hence, invoking the Baire category theorem (see [106, Section 10.2]) and taking (7.4.1) into account, we can deduce that there is a pair m, n such that $E_{m,n}$ has nonempty interior in $\overline{\mathsf{span}}\,\mathcal{F}_\Omega$. This implies that the interior of $\{f \leqslant n\} \subseteq \Omega$ relative to $\overline{\mathsf{span}}\,\mathcal{F}_\Omega$ is nonempty. Hence, we conclude that Ω has nonempty interior. □

Let us set $\mathcal{C}' = \mathcal{F}_\Omega$, $\mathsf{H}' = \overline{\mathsf{span}}\,\mathcal{F}_\Omega$ and f' be the restriction of f to $\mathcal{C}'$. Since $\Omega \subseteq \mathcal{F}_\Omega$, it is immediate that $\mathsf{dom}\, f' = \Omega \subseteq \mathcal{C}'$. Also, due to $f \in \Gamma_{\nearrow}(\mathcal{C})$, we have $f' \in \Gamma_{\nearrow}(\mathcal{C}')$. By (H1), (H2) and Lemma 7.4.1, $\mathcal{C}'$ is closed and pointed in H'. Lemma 7.4.2 guarantees that $\mathsf{dom}\, f'$ has nonempty interior in H'. Therefore, invoking Proposition 7.3.1, we obtain

$$f'(x) = f'^{*'*'}(x), \quad \forall x \in \mathcal{C}'. \tag{7.4.2}$$

Here,

$$f'^{*'}(y) = \sup_{z \in \mathcal{C}'} \{\langle z, y\rangle - f'(z)\}, \quad \forall y \in \mathcal{C}'^{\vee},$$
$$f'^{*'*'}(x) = \sup_{y \in \mathcal{C}'^{\vee}} \{\langle y, x\rangle - f'^{*'}(y)\}, \quad \forall x \in \mathcal{C}',$$

where $\mathcal{C}'^{\vee}$ is the dual cone of $\mathcal{C}'$ in H'. Due to $\Omega \subseteq \mathcal{C}'$,

$$f(z) = \infty, \quad \forall z \notin \mathcal{C}'. \tag{7.4.3}$$

By this and (7.1.1), we have

$$f^*(y) = f'^{*'}(y), \quad \forall y \in \mathcal{C}'^{\vee}.$$

Using (H3), the definition of f' and (7.4.3), we can see that

$$\{L_{a,\nu} : a \in \mathcal{C}'^{\vee},\ \nu \in \mathbb{R} \text{ such that } L_{a,\nu} \leqslant f' \text{ on } \mathcal{C}'\}$$
$$\subseteq \{L_{a,\nu} : a \in \mathcal{C}^{\vee},\ \nu \in \mathbb{R} \text{ such that } L_{a,\nu} \leqslant f \text{ on } \mathcal{C}\},$$

which together with Lemma 7.3.6 implies that

$$f^{**}(x) \geqslant f'^{*'*'}(x), \quad \forall x \in \mathcal{C}'.$$

This along with (7.4.2) and $f = f'$ on $\mathcal{C}'$ yields $f^{**} \geqslant f$ on $\mathcal{C}'$. Lastly, from (7.3.10), we conclude that

$$f(x) = f^{**}(x), \quad \forall x \in \mathcal{F}_\Omega. \tag{7.4.4}$$

7.4.2. Identity on $\mathcal{C}$

Due to (7.4.4), we only need to show $f(x) = f^{**}(x)$ for $x \in \mathcal{C} \setminus \mathcal{F}_\Omega$. To start, we record useful properties of faces in the ensuing two lemmas. Note that from the discussion below Definition 7.2.1 we have $\operatorname{int} \mathcal{C} \neq \emptyset$ if $\mathcal{C}$ is perfect.

Lemma 7.4.3. *Let $\mathcal{F}$ be a face of a cone $\mathcal{C} \subseteq \mathsf{H}$. If $\mathcal{F} \neq \mathcal{C}$, then $\mathcal{F} \cap \operatorname{int} \mathcal{C} = \emptyset$ and thus $\mathcal{F} \subseteq \operatorname{bd} \mathcal{C}$.*

Proof. Let us argue by contradiction and suppose that there is $x \in \mathcal{F} \cap \operatorname{int} \mathcal{C}$. Then for every $y \in \mathcal{C}$, we can find $\varepsilon > 0$ small so that $x - \varepsilon y \in \mathcal{C}$ and thus $0 \preceq \varepsilon y \preceq x$. Then, the definition of faces implies that $\varepsilon y \in \mathcal{F}$. Since $\mathcal{F}$ is a cone and $\varepsilon > 0$, we obtain $y \in \mathcal{F}$ which implies $\mathcal{C} \subseteq \mathcal{F}$ and thus $\mathcal{C} = \mathcal{F}$, contradicting the assumption that $\mathcal{F} \neq \mathcal{C}$. Therefore, the desired result holds. □

Lemma 7.4.4. *Assume (H1)–(H3). Let $\mathcal{F}$ be a face of $\mathcal{C}$. For every $x \in \mathcal{C} \setminus \mathcal{F}$, there is $v \in \mathcal{C}^\vee$ such that $\langle v, x \rangle > 0$ and*

$$\langle v, y \rangle = 0, \quad \forall y \in \mathcal{F}.$$

Proof. We take $\mathcal{F}'$ to be the intersection of all faces of $\mathcal{C}$ containing both $\mathcal{F}$ and x. It can be checked that $\mathcal{F}'$ is again a face of $\mathcal{C}$. Hence, $\mathcal{F}'$ is the minimal face containing both $\mathcal{F}$

and x. Let us write

$$\mathsf{H}' = \overline{\mathsf{span}}\,\mathcal{F}' \tag{7.4.5}$$

and denote by $\mathring{\mathcal{F}}'$ the interior of $\mathcal{F}'$ with respect to H'. By (H2), we have $\mathring{\mathcal{F}}' \neq \emptyset$ and that $\mathcal{F}'$ is closed. Since $\mathcal{F}$ is clearly a face of $\mathcal{F}'$, Lemma 7.4.3 applied to $\mathcal{F} \subseteq \mathcal{F}'$ yields $\mathcal{F} \cap \mathring{\mathcal{F}}' = \emptyset$.

By the Hahn–Banach separation theorem (c.f. [28, Theorem 1.6]), there are $\alpha \in \mathbb{R}$ and a nonzero vector $w \in \mathsf{H}'$ such that

$$\langle w, y \rangle \leqslant \alpha, \quad \forall y \in \mathcal{F}, \tag{7.4.6}$$

$$\langle w, z \rangle \geqslant \alpha, \quad \forall z \in \mathring{\mathcal{F}}'. \tag{7.4.7}$$

Since $\mathcal{F}'$ is closed and convex, and $\mathring{\mathcal{F}}' \neq \emptyset$, by [23, Proposition 3.36 (iii)], we have that the closure of $\mathring{\mathcal{F}}'$ is $\mathcal{F}'$. Hence, (7.4.7) becomes

$$\langle w, z \rangle \geqslant \alpha, \quad \forall z \in \mathcal{F}'. \tag{7.4.8}$$

By (7.2.2), we have $0 \in \mathcal{F}$. Due to this and $\mathcal{F} \subseteq \mathcal{F}'$, using (7.4.6) and (7.4.8), we must have $\alpha = 0$ and

$$\langle w, y \rangle = 0, \quad \forall y \in \mathcal{F}. \tag{7.4.9}$$

Then, (7.4.8) is turned into $\langle w, z \rangle \geqslant 0$ for all $z \in \mathcal{F}'$ which implies that

$$w \in \mathcal{F}'^{\vee} \tag{7.4.10}$$

where $\mathcal{F}'^{\vee}$ is the dual cone of $\mathcal{F}'$ in H'. Due to (H3), there is $v \in \mathcal{C}^{\vee}$ such that

$$\langle v, z \rangle = \langle w, z \rangle, \quad \forall z \in \mathsf{H}'. \tag{7.4.11}$$

Now, we consider the null space of the linear map $y \mapsto \langle v, y\rangle$ given by

$$\mathbb{E} = \{y \in \mathsf{H} : \langle v, y\rangle = 0\}. \tag{7.4.12}$$

We want to show $\mathbb{E} \cap \mathcal{F}'$ is a face of $\mathcal{C}$. It is clear that $\mathbb{E} \cap \mathcal{F}'$ is a cone. For $y \in \mathbb{E} \cap \mathcal{F}'$ and $z \in \mathcal{C}$ satisfying $0 \preceq z \preceq y$, by $v \in \mathcal{C}^\vee$, we obtain

$$\langle v, y - z\rangle \geqslant 0,$$
$$\langle v, z\rangle \geqslant 0.$$

Due to $y \in \mathbb{E}$, the above two displays yield $\langle v, z\rangle = 0$ which implies that $z \in \mathbb{E}$. Since $\mathcal{F}'$ is a face, by $0 \preceq z \preceq y$ and $y \in \mathcal{F}'$, we also have $z \in \mathcal{F}'$. Hence, we have $z \in \mathbb{E} \cap \mathcal{F}'$ and thus verified that $\mathbb{E} \cap \mathcal{F}'$ is a face of $\mathcal{C}$.

We claim that

$$\mathbb{E} \cap \mathcal{F}' \neq \mathcal{F}'. \tag{7.4.13}$$

Otherwise, we have $\mathcal{F}' \subseteq \mathbb{E}$, which due to (7.4.5) implies that $\mathsf{H}' \subseteq \mathbb{E}$. However, this along with (7.4.11) means that $\langle w, w\rangle = 0$ contradicting the fact that $w \neq 0$. Hence, (7.4.13) is valid.

To conclude, we argue that

$$x \notin \mathbb{E}. \tag{7.4.14}$$

Otherwise, since $\mathcal{F}'$ contains x by the definition of $\mathcal{F}'$, we have $x \in \mathbb{E} \cap \mathcal{F}'$. From (7.4.9), (7.4.11) and (7.4.12), we can deduce that $\mathcal{F} \subseteq \mathbb{E}$ and thus $\mathcal{F} \subseteq \mathbb{E} \cap \mathcal{F}'$. Therefore, $\mathbb{E} \cap \mathcal{F}'$ is a face containing both x and $\mathcal{F}$. However, this together with (7.4.13) contradicts the fact that $\mathcal{F}'$ is chosen to be the minimal face containing x and $\mathcal{F}$.

Therefore, by contradiction, we conclude that (7.4.14) must hold. Then, by $x \in \mathcal{F}'$ and (7.4.10), we must have $\langle v, x \rangle > 0$. In view of this, (7.4.9) and (7.4.11), the vector v satisfies all the desired properties. □

With these results, we resume the proof of $f = f^{**}$ on $\mathcal{C} \setminus \mathcal{F}_\Omega$. Fix any $x \in \mathcal{C} \setminus \mathcal{F}_\Omega$. For each $\rho > 0$, we set

$$\mathcal{L}_\rho = L_{\rho v,\ f(0)},$$

with $v \in \mathcal{C}^\vee$ given in Lemma 7.4.4 corresponding to this x and $\mathcal{F} = \mathcal{F}_\Omega$. This lemma implies that v is perpendicular to $\mathcal{F}_\Omega$ and thus

$$\mathcal{L}_\rho(y) = \rho \langle v, y \rangle + f(0) = f(0), \quad \forall y \in \mathcal{F}_\Omega.$$

Then, the $\mathcal{C}$-nondecreasingness of f implies that

$$f(y) \geqslant \mathcal{L}_\rho(y), \quad \forall y \in \mathcal{F}_\Omega.$$

Since we know $f = \infty$ on $\mathcal{C} \setminus \mathcal{F}_\Omega$, we obtain

$$f \geqslant \mathcal{L}_\rho, \quad \forall \rho > 0.$$

On the other hand, due to $\langle v, x \rangle > 0$ in Lemma 7.4.4, we have

$$\lim_{\rho \to \infty} \mathcal{L}_\rho(x) = \infty = f(x).$$

Hence, by the above two displays, (7.3.10) and Lemma 7.3.6, we conclude that $f(x) = f^{**}(x)$ for $x \in \mathcal{C} \setminus \mathcal{F}_\Omega$. This together with (7.4.4) completes the proof of Theorem 7.2.2.

7.4.3. Proof of Corollary 7.2.3

Recall the notion of perfect cones in Definition 7.2.1. We verify that any perfect cone $\mathcal{C}$ with $\overline{\mathsf{span}}\,\mathcal{C} = \mathsf{H}$ satisfies (H1)–(H3). Since the self-duality of $\mathcal{C}$ implies that $\mathcal{C}$ is both closed and pointed, property (H1) holds for $\mathcal{C}$. For any face $\mathcal{F}$ of $\mathcal{C}$, by Definition 7.2.1 (1), $\mathcal{F}$ is self-dual in $\overline{\mathsf{span}}\,\mathcal{F}$ and thus closed. Hence, (H2) follows from this and Definition 7.2.1 (2). Lastly, due to $\mathcal{F} \subseteq \mathcal{C}$, $\mathcal{F}^\vee = \mathcal{F}$ and $\mathcal{C}^\vee = \mathcal{C}$, it is immediate that $\mathcal{F}^\vee \subseteq \mathbb{P}_{\overline{\mathsf{span}}\,F}(\mathcal{C}^\vee)$ and thus (H3) holds. Therefore, Theorem 7.2.2 yields Corollary 7.2.3.

7.5. Examples of perfect cones

We show that the set of positive semidefinite matrices is a perfect cone, and that an infinite-dimensional circular cone is perfect.

7.5.1. Positive semidefinite matrices

Let $n \in \mathbb{N} \setminus \{0\}$ and denote by $\mathbb{S}^n$ the set of all $n \times n$ symmetric matrices, by $\mathbb{S}^n_+$ the set of all $n \times n$ positive semidefinite matrices, and by $\mathbb{S}^n_{++}$ the set of all $n \times n$ positive definite matrices. On $\mathbb{S}^n$, we define the inner product by

$$\langle x, y\rangle = \mathrm{tr}(xy), \quad \forall x, y \in \mathbb{S}^n,$$

where tr is the trace of a matrix and $x^\intercal$ is the transpose of x. Hence, $\mathbb{S}^n$ is a Hilbert space with dimension $n(n+1)/2$. The goal is the following.

Lemma 7.5.1. *For each positive integer n, the set $\mathbb{S}^n_+$ is a perfect cone in $\mathbb{S}^n$.*

To start, it is well-know that $\mathbb{S}^n_+$ is self-dual, which is attributed often to Fejér (see, e.g. [76, Theorem 7.5.4]). For completeness of presentation, we prove it below.

Lemma 7.5.2. *Let $x \in \mathbb{S}^n$. Then, $x \in \mathbb{S}^n_+$ if and only if $\langle x, y\rangle \geqslant 0$ for every $y \in \mathbb{S}^n_+$.*

Proof. If $x \in \mathbb{S}^n_+$, then for any $y \in \mathbb{S}^n_+$ we have $\langle x, y\rangle = \mathrm{tr}(\sqrt{x}\sqrt{y}\sqrt{y}\sqrt{x}) \geqslant 0$. For the other direction, by choosing an orthonormal basis, we may assume that x is diagonal. Testing by

$y \in \mathbb{S}^n_+$, we can show that all diagonal entries in x are nonnegative and thus $x \in \mathbb{S}^n_+$. $\square$

Proof of Lemma 7.5.1. Given the above lemma, we only need to verify the conditions on the faces of $\mathbb{S}^n_+$ stated in Definition 7.2.1. Let $\mathcal{F}$ be a face of $\mathbb{S}^n_+$.

The cases $\mathcal{F} = \{0\}$ and $\mathcal{F} = \mathbb{S}^n_+$ are trivial, so we assume $\{0\} \subsetneq \mathcal{F} \subsetneq \mathbb{S}^n_+$. Lemma 7.4.3 implies $\mathcal{F} \subseteq \mathsf{bd}\, \mathbb{S}^n_+ = \mathbb{S}^n_+ \setminus \mathbb{S}^n_{++}$. Set

$$m = \max\left\{\mathsf{rank}(z) : \ z \in \mathcal{F}\right\}, \tag{7.5.1}$$

where $\mathsf{rank}(z)$ is the rank of the matrix z. By our assumption on $\mathcal{F}$, we must have $1 \leqslant m < n$. For each $k \in \mathbb{N} \setminus \{0\}$, we denote by $\mathbf{0}_k$ the $k \times k$ zero matrix. Due to (7.5.1), there is $x \in \mathcal{F}$ with $\mathsf{rank}(x) = m$. By fixing a suitable orthonormal basis, we may assume

$$x = \mathsf{diag}(\lambda_1, \lambda_2, \cdots, \lambda_m, \mathbf{0}_{n-m}), \tag{7.5.2}$$

where $\lambda_j > 0$ for all $1 \leqslant j \leqslant m$.

Let us consider the following set

$$\mathbb{E} = \left\{\mathsf{diag}(y^\circ, \mathbf{0}_{n-m}) : \quad y^\circ \in \mathbb{S}^m_+\right\} \subseteq \mathbb{S}^n_+. \tag{7.5.3}$$

We now show $\mathbb{E} = \mathcal{F}$. First, we want to prove $\mathcal{F} \subseteq \mathbb{E}$. In other words, we claim that for every $y \in \mathcal{F}$, there is $y^\circ \in \mathbb{S}^m_+$ such that

$$y = \mathsf{diag}(y^\circ, \mathbf{0}_{n-m}). \tag{7.5.4}$$

Let us argue by contradiction. Suppose that (7.5.4) does not hold for all $y \in \mathcal{F}$, then we can find $y \in \mathcal{F}$ with $y_{jk} \neq 0$ for some $j > m$ or $k > m$. Assuming the former without loss of generality, we compute $v^\intercal y v$ for $v = te^j + e^k$ and vary $t \in \mathbb{R}$ where e^j and e^k belong to the standard basis for $\mathbb{R}^n$. Then, due to $y \in \mathbb{S}^n_+$, we must have $y_{jj} > 0$. By reordering the basis, we may assume $j = m+1$, and thus

$$y_{m+1,\,m+1} > 0. \tag{7.5.5}$$

Let $\widehat{y} = (y_{ij})_{1\leqslant i,\,j\leqslant m+1} \in \mathbb{S}^{m+1}_{+}$, and we define $\widehat{x}$ similarly. Then, we want to show $\mathsf{rank}(\widehat{x}+\widehat{y}) = m+1$. Let $v \in \mathbb{R}^{m+1}\setminus\{0\}$. If $v_j \neq 0$ for some $1 \leqslant j \leqslant m$, then we have

$$v^{\mathsf{T}}(\widehat{x}+\widehat{y})v \geqslant v^{\mathsf{T}}\widehat{x}v > 0.$$

The last inequality follows from (7.5.2). If $v_j = 0$ for all $1 \leqslant j \leqslant m$, then due to $v \neq 0$, we must have $v_{m+1} \neq 0$, and by (7.5.5), we get

$$v^{\mathsf{T}}(\widehat{x}+\widehat{y})v \geqslant v^{\mathsf{T}}\widehat{y}v = y_{m+1,m+1}v_{m+1}^2 > 0.$$

In conclusion, we obtain $v^{\mathsf{T}}(\widehat{x}+\widehat{y})v > 0$, which implies that $\widehat{x}+\widehat{y} \in \mathbb{S}^{m+1}_{++}$ and thus $\mathsf{rank}(x+y) \geqslant \mathsf{rank}(\widehat{x}+\widehat{y}) = m+1$. Since $\mathcal{F}$ is a cone, we have $x+y \in \mathcal{F}$. But this contradicts the maximality of m as in (7.5.1). Hence, every $y \in \mathcal{F}$ satisfies (7.5.4), and thus we verified $\mathcal{F} \subseteq \mathbb{E}$.

Now, we turn to the proof of $\mathbb{E} \subseteq \mathcal{F}$. For every y of the form (7.5.4), due to (7.5.2), there exists a small $\varepsilon > 0$ such that $x \succeq \varepsilon y \succeq 0$ where the partial order $\succeq$ is induced by the cone $\mathbb{S}^n_+$. Indeed, such ε exists because, viewing x, y as matrices in $\mathbb{S}^m_+$, we can choose ε sufficiently small so that the absolute values of eigenvalues of y is bounded by $\min_{1\leqslant j\leqslant m}\lambda_j$. Recall the definition of faces above Definition 7.2.1. Since $\mathcal{F}$ is a face, we must have $\varepsilon y \in \mathcal{F}$ and thus $y \in \mathcal{F}$. Hence, we conclude $\mathbb{E} \subseteq \mathcal{F}$.

Now, we have $\mathcal{F} = \mathbb{E}$. In view of (7.5.3), we can identify $\mathcal{F}$ with $\mathbb{S}^m_+$ and $\overline{\mathsf{span}}\,\mathcal{F}$ with $\mathbb{S}^m$. We know that $\mathbb{S}^m_+$ is self-dual in $\mathbb{S}^m$ by Lemma 7.5.2, whose interior is given by $\mathbb{S}^m_{++}$ and thus not empty. Therefore, all conditions on $\mathcal{F}$ in Definition 7.2.1 are verified. $\square$

7.5.2. An infinite-dimensional circular cone

We consider a generalization of the finite dimensional circular cone $\{x \in \mathbb{R}^{d+1} : (x_1^2 + \cdots + x_d^2)^{\frac{1}{2}} \leqslant x_0\}$. Let $\mathsf{H} = l^2(\mathbb{N})$ where the elements in $l^2(\mathbb{N})$ are precisely $x = (x_0, x_1, x_2, \dots)$ with $\sum_{i=0}^{\infty} x_i^2 < \infty$. The inner product on H is given by

$$\langle x, y\rangle = \sum_{i=0}^{\infty} x_i y_i, \quad \forall x, y \in \mathsf{H}.$$

We denote by $|\cdot|$ the associated norm. For each $x \in \mathsf{H}$, we write $x_{\geqslant 1} = (0, x_1, x_2, \dots) \in \mathsf{H}$. We consider the following cone

$$\mathcal{C} = \{x \in \mathsf{H} : \ |x_{\geqslant 1}| \leqslant x_0\}. \tag{7.5.6}$$

The desired result is stated below.

Lemma 7.5.3. *The cone $\mathcal{C}$ defined in* (7.5.6) *is perfect in* H*.*

To prove this lemma, we start with the following result.

Lemma 7.5.4. *The interior of $\mathcal{C}$ is nonempty, and given by*

$$\operatorname{int}\mathcal{C} = \{x \in \mathsf{H} : \ x_0 > 0, \, |x_{\geqslant 1}| < x_0\}. \tag{7.5.7}$$

Proof. Let y belong to the set on the right hand side of (7.5.7). Choose $\varepsilon > 0$ such that $y_0 - |y_{\geqslant 1}| > 2\varepsilon$. Then, we want to show that, for all $x \in \mathsf{H}$ satisfying $|x - y| < \varepsilon$, we have $x \in \mathcal{C}$. We can see that

$$(x_0 - y_0)^2 + |x_{\geqslant 1} - y_{\geqslant 1}|^2 = |x - y|^2 < \varepsilon^2.$$

This yields $|x_0 - y_0| < \varepsilon$ and $|x_{\geqslant 1} - y_{\geqslant 1}| < \varepsilon$. Now, using these, the property of ε and the triangle inequality, we get

$$|x_{\geqslant 1}| \leqslant |y_{\geqslant 1}| + \varepsilon < (y_0 - 2\varepsilon) + \varepsilon = y_0 - \varepsilon \leqslant x_0.$$

Hence, we have $x \in \mathcal{C}$ and can deduce that the right side of (7.5.7) is contained in $\mathsf{int}\,\mathcal{C}$. For the other direction, let $y \in \mathcal{C}$ with $|y_{\geqslant 1}| = y_0$. It is easy to see that every neighborhood of y contains a point not in $\mathcal{C}$. Therefore, we conclude that (7.5.7) holds. □

In order to prove the perfectness of $\mathcal{C}$, we need information about its faces. The next lemma classifies all faces of $\mathcal{C}$. The definition of faces are given above Definition 7.2.1.

Lemma 7.5.5. *Under the above setting, if $\mathcal{F}$ is a face of $\mathcal{C}$, then either $\mathcal{F} = \mathcal{C}$ or there is $x \in \mathsf{bd}\,\mathcal{C}$ such that $\mathcal{F} = \{\lambda x : \ \lambda \geqslant 0\}$.*

Proof. It is clear that $\mathcal{C}$ is a face of itself. Now we consider the case $\mathcal{F} \neq \mathcal{C}$. If $\mathcal{F} = \{0\}$, then there is nothing to prove. Hence, let us further assume that there is a nonzero $x \in \mathcal{F} \subseteq \mathcal{C}$. In particular, due to (7.5.6), we have $x_0 > 0$. Lemma 7.4.3 implies $\mathcal{F} \subseteq \mathsf{bd}\,\mathcal{C}$. By Lemma 7.5.4 and the definition of $\mathcal{C}$, the vector x satisfies

$$|x_{\geqslant 1}| = x_0 > 0. \tag{7.5.8}$$

By definition of faces, $\mathcal{F}$ is a cone. Due to this and $x \in \mathcal{F}$, we have

$$\mathcal{F} \supset \{\lambda x : \ \lambda \geqslant 0\}.$$

Now, we show that the above is in fact an equality. Let $y \in \mathcal{F} \setminus \{0\}$. By similar arguments as above, we have $y_0 > 0$. Rescaling if needed, we may assume $y_0 = x_0$. Recall that in this work, convexity is built into the definition of cones. Set $z = \frac{1}{2}(x + y)$. Using Jensen's inequality, by (7.5.8) and an analogous one for y, we obtain

$$x_0^2 = |z_{\geqslant 1}|^2 = \sum_{i=1}^{\infty} \left(\frac{x_i + y_i}{2}\right)^2 \leqslant \sum_{i=1}^{\infty} \frac{x_i^2 + y_i^2}{2} = x_0^2.$$

The equality holds only if $x_i = y_i$ for all i, so $y = x$ and the proof is complete. □

Proof of Lemma 7.5.3. We first show that $\mathcal{C}$ is self-dual. Recall that the dual cone is defined in (7.2.1) and denoted by $\mathcal{C}^\vee$. Let $y \in \mathcal{C}^\vee$ and we have

$$\langle x, y\rangle \geqslant 0, \quad \forall x \in \mathcal{C}. \tag{7.5.9}$$

Since $(1, 0, 0, \dots) \in \mathcal{C}$, we get $y_0 \geqslant 0$. We consider two cases depending on whether $y_0 = 0$ or not. Suppose $y_0 = 0$, for any fixed $i \geqslant 1$, we construct x' in the following way. Set $x'_0 = 1$, set $x'_i = -1$ if $y_i \geqslant 0$ and $x'_i = 1$ if $y_i < 0$, and lastly set all other entries of x' to be zero. Inserting this x' into (7.5.9) and varying i, we can see $y = 0$ and thus $y \in \mathcal{C}$. Now we consider the case where $y_0 > 0$. If $|y_{\geqslant 1}| = 0$, then this immediately implies $y \in \mathcal{C}$. If $|y_{\geqslant 1}| \neq 0$, then we set $\gamma = |y_{\geqslant 1}|^{-1} > 0$ and consider x' given by

$$x'_0 = y_0; \qquad x'_i = -\gamma y_i y_0.$$

Plugging x' into (7.5.9) and using $y_0 > 0$, we obtain $y_0 \geqslant |y_{\geqslant 1}|$ and thus $y \in \mathcal{C}$, which implies $\mathcal{C}^\vee \subseteq \mathcal{C}$. Since it is clear that $\mathcal{C} \subseteq \mathcal{C}^\vee$, we conclude that $\mathcal{C}$ is self-dual.

To show $\mathcal{C}$ is perfect, it remains to check the conditions on the faces of $\mathcal{C}$ stated in Definition 7.2.1. Recall that Lemma 7.5.4 ensures $\mathsf{int}\,\mathcal{C} \neq \emptyset$. Hence, if $\mathcal{F} = \mathcal{C}$, then $\mathcal{F}$ is self-dual and has nonempty interior with respect to $\overline{\mathsf{span}}\,\mathcal{F}$. Now if $\mathcal{F} \neq \mathcal{C}$, then Lemma 7.5.5 implies $\mathcal{F} = \{\lambda x : \lambda \geqslant 0\}$, which is one-dimensional. We can identify $\overline{\mathsf{span}}\,\mathcal{F}$ with $\mathbb{R}$ and $\mathcal{F}$ with $[0, \infty)$ in an isometric way. Now, it is easy to see that $\mathcal{F}$ is self-dual and has nonempty interior with respect to $\overline{\mathsf{span}}\,\mathcal{F}$. By Definition 7.2.1, we conclude that $\mathcal{C}$ given in (7.5.6) is perfect. □

Bibliography

[1] M. Aizenman and A. Burchard. Hölder regularity and dimension bounds for random curves. *Duke Math. J.*, 99(3):419–453, 1999.

[2] M. Aizenman, M. Harel, and R. Peled. Exponential decay of correlations in the $2d$ random field Ising model. *J. Stat. Phys.*, 2019. to appear.

[3] M. Aizenman and R. Peled. A power-law upper bound on the correlations in the $2D$ random field Ising model. *Comm. Math. Phys.*, 372(3):865–892, 2019.

[4] M. Aizenman and J. Wehr. Rounding of first-order phase transitions in systems with quenched disorder. *Phys. Rev. Lett.*, 62(21):2503–2506, 1989.

[5] M. Aizenman and J. Wehr. Rounding effects of quenched randomness on first-order phase transitions. *Comm. Math. Phys.*, 130(3):489–528, 1990.

[6] D. Alberici, F. Camilli, P. Contucci, and E. Mingione. The solution of the deep Boltzmann machine on the Nishimori line. *Comm. Math. Phys.*, 387(2):1191–1214, 2021.

[7] L. Ambrosio and J. Feng. On a class of first order Hamilton–Jacobi equations in metric spaces. *Journal of Differential Equations*, 256(7):2194–2245, 2014.

[8] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient Flows: in Metric Spaces and in the Space of Probability Measures*. Springer Science & Business Media, 2005.

[9] J. Barbier. Overlap matrix concentration in optimal bayesian inference. *arXiv preprint arXiv:1904.02808*, 2019.

[10] 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.

[11] J. Barbier, F. Krzakala, N. Macris, L. Miolane, and L. Zdeborová. Optimal errors and phase transitions in high-dimensional generalized linear models. *Proceedings of the National Academy of Sciences*, 116(12):5451–5460, 2019.

[12] 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.

[13] 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.

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

[15] M. Bardi and L. C. Evans. On Hopf's formulas for solutions of Hamilton-Jacobi equations. *Nonlinear Analysis: Theory, Methods & Applications*, 8(11):1373–1381, 1984.

[16] G. P. Barker. Faces and duality in convex cones. *Linear and Multilinear Algebra*, 6(3):161–169, 1978.

[17] G. P. Barker. Perfect cones. *Linear Algebra and its Applications*, 22:211–221, 1978.

[18] G. P. Barker. Theory of cones. *Linear Algebra and its Applications*, 39:263–291, 1981.

[19] G. P. Barker and J. Foran. Self-dual cones in Euclidean spaces. *Linear Algebra and its Applications*, 13(1-2):147–155, 1976.

[20] G. Barles. An introduction to the theory of viscosity solutions for first-order Hamilton–Jacobi equations and applications. In *Hamilton-Jacobi Equations: Approximations, Numerical Analysis and Applications*, pages 49–109. Springer, 2013.

[21] 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.

[22] 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.

[23] H. H. Bauschke, P. L. Combettes, et al. *Convex analysis and monotone operator theory in Hilbert spaces*, volume 408. Springer, 2011.

[24] J. Bellissard, B. Iochum, and R. Lima. Homogeneous and facially homogeneous self-dual cones. *Linear Algebra and its Applications*, 19(1):1–16, 1978.

[25] S. H. Benton. *The Hamilton-Jacobi equation : a global approach.* Academic Press, New York, 1977.

[26] A. Berretti. Some properties of random Ising models. *J. Statist. Phys.*, 38(3-4):483–496, 1985.

[27] S. Boucheron, G. Lugosi, and P. Massart. *Concentration inequalities: A nonasymptotic theory of independence*. Oxford university press, 2013.

[28] H. Brezis. *Functional analysis, Sobolev spaces and partial differential equations*. Springer Science & Business Media, 2010.

[29] J. Bricmont and A. Kupiainen. The hierarchical random field Ising model. *J. Statist. Phys.*, 51(5-6):1021–1032, 1988. New directions in statistical mechanics (Santa Barbara, CA, 1987).

[30] J. Bricmont and A. Kupiainen. Phase transition in the 3d random field Ising model. *Comm. Math. Phys.*, 116(4):539–572, 1988.

[31] F. Camia, J. Jiang, and C. M. Newman. A note on exponential decay in the random field Ising model. *J. Stat. Phys.*, 173(2):268–284, 2018.

[32] P. Cardaliaguet. Notes on mean field games. Technical report, Technical report, 2010.

[33] P. Cardaliaguet and M. Quincampoix. Deterministic differential games under probability knowledge of initial condition. *International Game Theory Review*, 10(01):1–16, 2008.

[34] P. Cardaliaguet and A. Souquiere. A differential game with a blind player. *SIAM Journal on Control and Optimization*, 50(4):2090–2116, 2012.

[35] S. Chatterjee. On the decay of correlations in the random field ising model. *Communications in Mathematical Physics*, 362(1):253–267, 2018.

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

[37] H.-B. Chen, J.-C. Mourrat, and J. Xia. Statistical inference of finite-rank tensors. *arXiv preprint arXiv:2104.05360*, 2021.

[38] H.-B. Chen and J. Xia. Fenchel-Moreau identities on self-dual cones. *arXiv preprint arXiv:2011.06979*, 2020.

[39] H.-B. Chen and J. Xia. Hamilton-Jacobi equations for inference of matrix tensor products. *arXiv preprint arXiv:2009.01678*, 2020.

[40] H.-B. Chen and J. Xia. Limiting free energy of multi-layer generalized linear models. *arXiv preprint arXiv:2108.12615*, 2021.

[41] H.-B. Chen and J. Xia. Hamilton-Jacobi equations from mean-field spin glasses. *arXiv preprint arXiv:2201.12732*, 2022.

[42] M. G. Crandall, H. Ishii, and P.-L. Lions. User's guide to viscosity solutions of second order partial differential equations. *Bulletin of the American mathematical society*, 27(1):1–67, 1992.

[43] M. G. Crandall and P.-L. Lions. Hamilton-Jacobi equations in infinite dimensions I. Uniqueness of viscosity solutions. *Journal of Functional Analysis*, 62(3):379–396, 1985.

[44] M. G. Crandall and P.-L. Lions. Hamilton-Jacobi equations in infinite dimensions. II. Existence of viscosity solutions. *Journal of Functional Analysis*, 65(3):368–405, 1986.

[45] M. G. Crandall and P.-L. Lions. Hamilton-Jacobi equations in infinite dimensions, III. *Journal of functional analysis*, 68(2):214–247, 1986.

[46] M. G. Crandall and P.-L. Lions. Viscosity solutions of Hamilton-Jacobi equations in infinite dimensions. IV. Hamiltonians with unbounded linear terms. *Journal of Functional Analysis*, 90(2):237–283, 1990.

[47] M. G. Crandall and P.-L. Lions. Viscosity solutions of Hamilton-Jacobi equations in infinite dimensions. V. Unbounded linear terms and B-continuous solutions. *Journal of functional analysis*, 97(2):417–465, 1991.

[48] M. G. Crandall and P.-L. Lions. Hamilton-Jacobi equations in infinite dimensions. VI. Nonlinear A and Tataru's method refined. In *Evolution Equations, Control Theory, and Biomathematics (Han sur Lesse, 1991)*, volume 155 of *Lecture Notes in Pure and Appl. Math.*, pages 51–89. Dekker, New York, 1994.

[49] M. G. Crandall and R. Newcomb. Viscosity solutions of Hamilton–Jacobi equations at the boundary. *Proceedings of the American Mathematical Society*, 94(2):283–290, 1985.

[50] M. Damron and P. Tang. Superlinearity of geodesic length in 2d critical first-passage percolation. *Preprint 2019, arXiv:1902.03302*, 2016.

[51] Derrida, B. and Shnidman, Y. Possible line of critical points for a random field ising model in dimension 2. *J. Physique Lett.*, 45(12):577–581, 1984.

[52] 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*, pages 424–432, 2016.

[53] J. Ding and M. Wirth. Correlation length of two-dimensional random field ising model via greedy lattice animal, 2021.

[54] J. Ding and J. Xia. Exponential decay of correlations in the two-dimensional random field ising model at zero temperature. *arXiv preprint arXiv:1902.03302*, 2019.

[55] J. Ding and Z. Zhuang. Long range order for random field ising and potts models, 2021.

[56] A. Douglis. Solutions in the large for multi-dimensional non linear partial differential equations of first order. *Annales de l'Institut Fourier*, 15(2):1–35, 1965.

[57] J. Dutta, J. Martinez-Legaz, and A. Rubinov. Monotonic analysis over cones: I. *Optimization*, 53(2):129–146, 2004.

[58] J. Dutta, J. Martinez-Legaz, and A. Rubinov. Monotonic analysis over cones: II. *Optimization*, 53(5-6):529–547, 2004.

[59] J. Dutta, J. E. Martínez-Legaz, and A. M. Rubinov. Monotonic analysis over cones: III. *Journal of Convex Analysis*, 15(3):561, 2008.

[60] L. C. Evans. *Partial Differential Equations*, volume 19. American Mathematical Soc., 2010.

[61] J. Feng. Large deviation for diffusions and Hamilton–Jacobi equation in Hilbert spaces. *The Annals of Probability*, 34(1):321–385, 2006.

[62] J. Feng and M. Katsoulakis. A comparison principle for Hamilton–Jacobi equations related to controlled gradient flows in infinite dimensions. *Archive for Rational Mechanics and Analysis*, 192(2):275–310, 2009.

[63] C. M. Fortuin, P. W. Kasteleyn, and J. Ginibre. Correlation inequalities on some partially ordered sets. *Comm. Math. Phys.*, 22:89–103, 1971.

[64] J. Fröhlich and J. Z. Imbrie. Improved perturbation expansion for disordered systems: beating Griffiths singularities. *Comm. Math. Phys.*, 96(2):145–180, 1984.

[65] M. Gabrié, A. Manoel, C. Luneau, J. Barbier, N. Macris, F. Krzakala, and L. Zdeborová. Entropy and mutual information in models of deep neural networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2019(12):124014, 2019.

[66] W. Gangbo, T. Nguyen, A. Tudorascu, et al. Hamilton-Jacobi equations in the Wasserstein space. *Methods and Applications of Analysis*, 15(2):155–184, 2008.

[67] W. Gangbo and A. Święch. Optimal transport and large number of particles. *Discrete & Continuous Dynamical Systems*, 34(4):1397, 2014.

[68] W. Gangbo and A. Tudorascu. On differentiability in the Wasserstein space and well-posedness for Hamilton–Jacobi equations. *Journal de Mathématiques Pures et Appliquées*, 125:119–174, 2019.

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

[70] G. Grinstein and S.-K. Ma. Roughening and lower critical dimension in the random-field ising model. *Phys. Rev. Lett.*, 49:685–688, Aug 1982.

[71] F. Guerra. Sum rules for the free energy in the mean field spin glass model. *Fields Institute Communications*, 30(11), 2001.

[72] F. Guerra. Broken replica symmetry bounds in the mean field spin glass model. *Communications in Mathematical Physics*, 233(1):1–12, 2003.

[73] N. Hadjisavvas, S. Komlósi, and S. S. Schaible. *Handbook of generalized convexity and generalized monotonicity*, volume 76. Springer Science & Business Media, 2006.

[74] J.-B. Hiriart-Urruty and C. Lemaréchal. *Fundamentals of convex analysis*. Springer Science & Business Media, 2012.

[75] E. Hopf. Generalized solutions of non-linear equations of first order. *Journal of Mathematics and Mechanics*, 14(6):951–973, 1965.

[76] R. A. Horn and C. R. Johnson. *Matrix analysis*. Cambridge university press, 2012.

[77] J. Z. Imbrie. The ground state of the three-dimensional random-field Ising model. *Comm. Math. Phys.*, 98(2):145–176, 1985.

[78] Y. Imry and S.-K. Ma. Random-field instability of the ordered state of continuous symmetry. *Phys. Rev. Lett.*, 35:1399–1401, Nov 1975.

[79] J. Kadmon and S. Ganguli. Statistical mechanics of low-rank tensor decomposition. In *Advances in Neural Information Processing Systems*, pages 8201–8212, 2018.

[80] S. N. Kruzhkov. Generalized solutions of nonlinear equations of the first order with several variables. i. *Matematicheskii Sbornik*, 112(3):394–415, 1966.

[81] S. N. Kruzhkov. Generalized solutions of nonlinear first order equations with several independent variables. ii. *Matematicheskii Sbornik*, 114(1):108–134, 1967.

[82] M. Lelarge and L. Miolane. Fundamental limits of symmetric low-rank matrix estimation. *Probability Theory and Related Fields*, 173(3-4):859–929, 2019.

[83] 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.

[84] P.-L. Lions and J.-C. Rochet. Hopf formula and multitime Hamilton-Jacobi equations. *Proceedings of the American Mathematical Society*, 96(1):79–84, 1986.

[85] 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.

[86] C. Luneau, N. Macris, and J. Barbier. High-dimensional rank-one nonsymmetric matrix decomposition: the spherical case. *arXiv preprint arXiv:2004.06975*, 2020.

[87] 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.

[88] V. Mayya and G. Reeves. Mutual information in community detection with covariate information and correlated networks. *arXiv preprint arXiv:1912.05375*, 2019.

[89] R. E. Megginson. *An Introduction to Banach Space Theory*, volume 183. Springer Science & Business Media, 2012.

[90] L. Miolane. Fundamental limits of low-rank matrix estimation: the non-symmetric case. *arXiv preprint arXiv:1702.00473*, 2017.

[91] J. J. Moreau. Inf-convolution, sous-additivité, convexité des fonctions numériques. *Journal de Mathématiques Pures et Appliquées*, 1970.

[92] J.-C. Mourrat. Parisi's formula is a Hamilton-Jacobi equation in Wasserstein space. *arXiv preprint arXiv:1906.08471*, 2019.

[93] J.-C. Mourrat. Free energy upper bound for mean-field vector spin glasses. *arXiv preprint arXiv:2010.09114*, 2020.

[94] J.-C. Mourrat. Hamilton–Jacobi equations for finite-rank matrix inference. *Annals of Applied Probability*, 30(5):2234–2260, 2020.

[95] J.-C. Mourrat. Hamilton–Jacobi equations for mean-field disordered systems. *Annales Henri Lebesgue*, 4:453–484, 2021.

[96] J.-C. Mourrat. Nonconvex interactions in mean-field spin glasses. *Probability and Mathematical Physics*, 2(2):61–119, 2021.

[97] J.-C. Mourrat. The Parisi formula is a Hamilton–Jacobi equation in Wasserstein space. *Canadian Journal of Mathematics*, page 1–23, 2021.

[98] J.-C. Mourrat and D. Panchenko. Extending the Parisi formula along a Hamilton-Jacobi equation. *Electronic Journal of Probability*, 25, 2020.

[99] D. Panchenko. *The Sherrington-Kirkpatrick model.* Springer Science & Business Media, 2013.

[100] D. Panchenko. The Parisi ultrametricity conjecture. *Annals of Mathematics*, pages 383–393, 2013.

[101] G. Parisi. A sequence of approximated solutions to the SK model for spin glasses. *Journal of Physics A: Mathematical and General*, 13(4):L115, 1980.

[102] R. C. Penney. Self-dual cones in Hilbert space. *Journal of Functional Analysis*, 21(3):305–315, 1976.

[103] G. Reeves. Information-theoretic limits for the matrix tensor product. *IEEE Journal on Selected Areas in Information Theory*, 1(3):777–798, 2020.

[104] 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.

[105] R. T. Rockafellar. *Convex Analysis*, volume 36. Princeton university press, 1970.

[106] H. Royden and P. Fitzpatrick. *Real Analysis.* Prentice Hall, 2010.

[107] S. Sheffield. Random surfaces. *Astérisque*, 304:vi+175, 2005.

[108] I. Singer. *Abstract convex analysis*, volume 25. John Wiley & Sons, 1997.

[109] P. Souganidis. A remark about viscosity solutions of Hamilton–Jacobi equations at the boundary. *Proceedings of the American Mathematical Society*, 96(2):323–329, 1986.

[110] C. Stegall. Optimization of functions on certain subsets of Banach spaces. *Mathematische Annalen*, 236(2):171–176, 1978.

[111] M. Talagrand. The Parisi formula. *Annals of mathematics*, pages 221–263, 2006.

[112] M. Talagrand. *Mean Field Models for Spin Glasses: Volume I: Basic Examples*, volume 54. Springer Science & Business Media, 2010.

[113] M. Talagrand. *Mean Field Models for Spin Glasses: Volume II: Advanced Replica-Symmetry and Low Temperature*, volume 55. Springer Science & Business Media, 2011.

[114] D. Tataru. Viscosity solutions of Hamilton-Jacobi equations with unbounded nonlinear terms. *Journal of Mathematical Analysis and Applications*, 163(2):345–392, 1992.

[115] J. van den Berg. A uniqueness condition for Gibbs measures, with application to the 2-dimensional Ising antiferromagnet. *Comm. Math. Phys.*, 152(1):161–166, 1993.

[116] R. Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press, 2018.

[117] F. Zhang. *Matrix theory.* Universitext. Springer, New York, second edition, 2011.