NONCONVEX INTERACTIONS IN MEAN-FIELD SPIN GLASSES

JEAN-CHRISTOPHE MOURRAT

ABSTRACT. We propose a conjecture for the limit free energy of mean-field spin glasses with a bipartite structure, and show that the conjectured limit is an upper bound. The conjectured limit is described in terms of the solution to an infinite-dimensional Hamilton-Jacobi equation. A fundamental difficulty of the problem is that the nonlinearity in this equation is not convex. We also question the possibility to characterize this conjectured limit in terms of a saddle-point problem.

1. INTRODUCTION

Let $(J_{ij})_{i,j \geqslant 1}$ be independent standard Gaussian random variables, and, for every $\sigma = (\sigma_{1,1}, \ldots, \sigma_{1,N}, \sigma_{2,1}, \ldots, \sigma_{2,N}) \in \mathbb{R}^{2N}$, let

$$(1.1) \qquad H_N(\sigma) := N^{-\frac{1}{2}} \sum_{i,j=1}^{N} J_{ij} \, \sigma_{1,i} \, \sigma_{2,j}.$$

The main goal of this paper is to study the large-N behavior of the free energy

$$(1.2) \qquad \frac{1}{N} \mathbb{E} \log \int_{\mathbb{R}^{2N}} \exp\left(\beta H_N(\sigma)\right) \mathrm{d}P_N(\sigma),$$

where $\beta \geqslant 0$ and P_N is a "simple" probability measure over $\mathbb{R}^{2N}$. For convenience, we assume that there exist two probability measures π_1 and π_2 on $\mathbb{R}$ with compact support such that, for every $N \geqslant 1$,

$$(1.3) \qquad P_N = \pi_1^{\otimes N} \otimes \pi_2^{\otimes N}.$$

Without loss of generality, we assume that the supports of π_1 and π_2 are subsets of $[-1, 1]$. For every metric space E, we denote by $\mathcal{P}(E)$ the space of Borel probability measures on E, and, for every $p \in [1, \infty]$, by $\mathcal{P}_p(E)$ the subspace of $\mathcal{P}(E)$ of probability measures with finite p-th moment. We write δ_x for the Dirac probability measure at $x \in E$. For every $\nu \in \mathcal{P}(\mathbb{R}_+)$ and $r \in [0,1]$, we define

$$(1.4) \qquad F_\nu^{-1}(r) := \inf\{s \geqslant 0 \ : \ \nu([0,s]) \geqslant r\},$$

and, for U a uniform random variable over $[0,1]$, we write

$$(1.5) \qquad X_\nu := F_\nu^{-1}(U).$$

Recall that the law of X_ν is ν, and that this construction provides us with a joint coupling of all probability measures over $\mathbb{R}_+$. For every $\mu = (\mu_1, \mu_2) \in (\mathcal{P}(\mathbb{R}_+))^2$, we denote by $\widehat{\mu} \in \mathcal{P}(\mathbb{R}_+^2)$ the law of the pair (X_{μ_1}, X_{μ_2}). Here is the main result of this paper.

Theorem 1.1. *For every $t \geqslant 0$, we have*

$$(1.6) \qquad \liminf_{N \to \infty} -\frac{1}{N} \mathbb{E} \log \int \exp\left(\sqrt{2t} H_N(\sigma) - N^{-1} t |\sigma_1|^2 |\sigma_2|^2\right) \mathrm{d}P_N(\sigma) \geqslant f(t, (\delta_0, \delta_0)),$$

where $f = f(t, \mu) : \mathbb{R}_+ \times (\mathcal{P}_2(\mathbb{R}_+))^2 \to \mathbb{R}$ is the solution to

(1.7)
$$\begin{cases} \partial_t f - \int \partial_{\mu_1} f \, \partial_{\mu_2} f \, \mathrm{d}\widehat{\mu} = 0 & on \;\; \mathbb{R}_+ \times (\mathcal{P}_2(\mathbb{R}_+))^2, \\ f(0, \cdot) = \psi & on \;\; (\mathcal{P}_2(\mathbb{R}_+))^2, \end{cases}$$

and the initial condition ψ is defined below in (2.19).

We start by clarifying the meaning of the Hamilton-Jacobi equation in (1.7). Alternative expressions for the integral in (1.7) read

$$\int \partial_{\mu_1} f \, \partial_{\mu_2} f \, \mathrm{d}\widehat{\mu} = \int_{\mathbb{R}_+^2} \partial_{\mu_1} f(t, \mu, x_1) \, \partial_{\mu_2} f(t, \mu, x_2) \, \mathrm{d}\widehat{\mu}(x_1, x_2)$$
$$= \mathbb{E}\left[\partial_{\mu_1} f(t, \mu, X_{\mu_1}) \, \partial_{\mu_2} f(t, \mu, X_{\mu_2}) \right].$$

The notion of derivative at play here is not of Fréchet type (which would express the linear response to the addition of a small signed measure of zero total mass), but rather of transport type. Informally, for a "smooth" function $g = g(\nu) : \mathcal{P}_2(\mathbb{R}_+) \to \mathbb{R}$, the derivative $\partial_\nu g(\nu, \cdot) \in L^2(\mathbb{R}_+, \nu)$ is characterized by the first-order expansion

$$g(\nu') = g(\nu) + \mathbb{E}\left[\partial_\nu g(\nu, X_\nu)(X_{\nu'} - X_\nu) \right] + o\left(\mathbb{E}\left[(X_{\nu'} - X_\nu)^2 \right]^{\frac{1}{2}} \right).$$

More concretely, given some integer $k \geqslant 1$, and setting, for every $x = (x_1, \ldots, x_k) \in \mathbb{R}_+^k$ such that $x_1 \leqslant \cdots \leqslant x_k$,

$$g^{(k)}(x_1, \ldots, x_k) := g\left(\frac{1}{k} \sum_{\ell=1}^k \delta_{x_k} \right),$$

we have, for every $n \in \{1, \ldots, k\}$,

$$\partial_\nu g\left(\frac{1}{k} \sum_{\ell=1}^k \delta_{x_k}, x_n \right) = k \, \partial_{x_n} g^{(k)}(x).$$

This suggests natural finite-dimensional approximations of the equation (1.7). Denoting

$$\overline{U}_k := \left\{ q = (q_{1,1}, q_{1,2}, \ldots, q_{1,k}, q_{2,1}, \ldots, q_{2,k}) \in \mathbb{R}_+^{2k} \; : \; \forall a \in \{1, 2\}, \; q_{a,1} \leqslant \cdots \leqslant q_{a,k} \right\},$$

These approximations take the form

(1.8)
$$\partial_t f^{(k)} - k \sum_{\ell=1}^k \partial_{q_{1,\ell}} f^{(k)} \, \partial_{q_{2,\ell}} f^{(k)} = 0 \qquad on \;\; \mathbb{R}_+ \times \overline{U}_k.$$

We will define the solution to (1.7) as the limit of such finite-dimensional approximations.[1]

That there exists a connection between the free energy of spin glass models and certain infinite-dimensional Hamilton-Jacobi equations was first observed in the context of mixed p-spin models [40]. In these models, the energy function H_N is a centered Gaussian field such that the covariance between $H_N(\sigma)$ and $H_N(\tau)$ is proportional to $\xi(\sigma \cdot \tau / N)$, where the function ξ is fixed and can be written in the form $\xi(r) = \sum_{p \geqslant 2} \beta_p r^p$, for some family of coefficients $\beta_p \geqslant 0$ that decays sufficiently fast. (The constraint $\beta_p \geqslant 0$ is necessary and sufficient in order for ξ to define a covariance kernel for every N [51].) For these models, the corresponding Hamilton-Jacobi equation takes the form

(1.9)
$$\partial_t f - \int \xi(\partial_\mu f) \, \mathrm{d}\mu = 0 \qquad on \;\; \mathbb{R}_+ \times \mathcal{P}_2(\mathbb{R}_+).$$

---

[1]It may seem somewhat contrived to impose the ordering of the variables $q_{a,1} \leqslant \cdots \leqslant q_{a,k}$. However, in the proof of Theorem 1.1, this formulation will allow for a clearer treatment of the boundary condition on the "diagonal part", i.e. whenever $q_{a,\ell} = q_{a,\ell+1}$ for some $\ell \in \{1, \ldots, k-1\}$. (This point was overlooked in a preliminary version of the paper.) Moreover, in more general models, the relevant variables are matrix-valued, and there is no simple "symmetrization" of an ordered tuple of symmetric matrices, so there is no way around working with a set of the form of $\overline{U}_k$ in this more general setting.

With this in mind, it is natural to distinguish between three increasingly large classes of models. The first is the class of models for which the mapping ξ is convex over $\mathbb{R}$; roughly speaking, these are the models whose limit free energy can be identified using the methods of [32, 53, 56, 57] (in fact, the precise condition is slightly more restrictive, see [57, (14.101)]). An extension of this approach, developed in [43, 44], allows to cover all mixed p-spin models. The convexity property, once properly understood, is still fundamental in this setting. More precisely, one can check that the relevant solution to (1.9) satisfies $\partial_\mu f \geqslant 0$. On the other hand, in view of the form of ξ, this function is convex over $\mathbb{R}_+$. In other words, we can redefine the function ξ to be $+\infty$ over $(-\infty, 0)$; with this new definition, the relevant Hamilton-Jacobi equation is still (1.9), and now the convexity of the nonlinearity has been restored. This convexity is crucial to the validity of a Hopf-Lax formula for the solution, and this variational formula forms the basis of the arguments for identifying the limit free energy in these approaches.

The third class of models corresponds to situations in which the nonlinearity in the Hamilton-Jacobi equation may be genuinely nonconvex; a representative example in this class is the focus of the present paper. In this case, it is unclear whether the limit free energy can be described as a (reasonable) variational problem. The classical Hopf-Lax variational formula requires that the nonlinearity in the equation be convex (or concave), which it is clearly not in our setting. Alternatively, irrespectively of the structure of the nonlinearity, the solution to a Hamilton-Jacobi equation can always be written as a saddle-point problem, provided that the initial condition is concave (or convex) [33, 7, 35]. This motivates to study the concavity of the initial condition in (1.7), that is, the function ψ in (1.7). In the context of mixed p-spin models, the main result of [4] implies the concavity of this function. I do not know whether this argument can be generalized to cover the bipartite model investigated here. But in any case, this does not seem to be the appropriate notion of concavity to guarantee the validity of a saddle-point formulation for the solution to (1.7). In order for this to work, we would need instead that the function ψ be *transport-concave* (one may also say "displacement-concave"); but we will see that this is not so in general. At present, my impression is that it is *not* possible to express the limit free energy as a saddle-point problem in general, and that it would be very difficult to circumvent a description of this limit involving Hamilton-Jacobi equations.

We now discuss the intuition behind Theorem 1.1. The simplest setting in which to explain the idea is that of the Curie-Weiss model, see for instance [38]. The main point is to enrich the model to include "non-interacting" terms in the energy function, with the hope that, if these simpler terms are sufficiently "expressive", then certain asymptotic relations between the derivatives of the free energy will have to be satisfied. In our context, a first attempt is to try to compare $\sum J_{ij}\sigma_{1,i}\sigma_{2,j}$ with a linear combination of $z_1 \cdot \sigma_1$ and $z_2 \cdot \sigma_2$, where $z = (z_1, z_2) = (z_{1,1}, \ldots, z_{1,N}, z_{2,1}, \ldots, z_{2,N})$ is a vector of independent standard Gaussians. In other words, we consider, for every $t, p_1, p_2 \geqslant 0$, the free energy

$$(1.10) \quad G_N(t, p_1, p_2) :=$$
$$-\frac{1}{N}\mathbb{E}\log \int \exp\left(\sqrt{2t}H_N(\sigma) - N^{-1}t|\sigma_1|^2|\sigma_2|^2 + \sum_{a=1}^{2}\left(\sqrt{2p_a}\,z_a \cdot \sigma_a - p_a|\sigma_a|^2\right)\right)\mathrm{d}P_N(\sigma).$$

(Parametrizations of the form $\sqrt{t}X$ where X is a Gaussian random variable are of course natural: think of Brownian motion. Each random variable in the exponential comes with a compensating term, so that the expectation of the exponential is equal to 1.) Denoting by $\langle \cdot \rangle$ the expectation with respect to the Gibbs measure proportional to $\exp(\cdots)\mathrm{d}P_N(\sigma)$,

one can check that

$$\partial_t G_N = N^{-2} \, \mathbb{E} \left\langle (\sigma_1 \cdot \sigma_1')(\sigma_2 \cdot \sigma_2') \right\rangle,$$

where σ' denotes an independent copy of σ under $\langle \cdot \rangle$. On the other hand,

$$\partial_{p_a} G_N = N^{-1} \mathbb{E} \left\langle \sigma_a \cdot \sigma_a' \right\rangle \qquad (a \in \{1, 2\}),$$

so that

$$(1.11) \qquad \partial_t G_N - \partial_{p_1} G_N \, \partial_{p_2} G_N = N^{-2} \, \mathbb{E} \left\langle \left(\sigma_1 \cdot \sigma_1' - \mathbb{E} \left\langle \sigma_1 \cdot \sigma_1' \right\rangle \right) \left(\sigma_2 \cdot \sigma_2' - \mathbb{E} \left\langle \sigma_2 \cdot \sigma_2' \right\rangle \right) \right\rangle.$$

Hence, if the overlaps $\sigma_a \cdot \sigma_a'$ were concentrated, we would then infer that G_N converges to $g = g(t, p_1, p_2) : \mathbb{R}_+^3 \to \mathbb{R}$ solution to

$$(1.12) \qquad\qquad\qquad\qquad \partial_t g - \partial_{p_1} g \, \partial_{p_2} g = 0.$$

However, as is well-known, the concentration of the overlaps is only valid in a high-temperature (that is, small t) region; a more refined enriched system is necessary to "close the equation" in general. The formal manipulation allowing to obtain the true equation from the "naive" (or replica-symmetric) one given in (1.12) consists simply in replacing the variables (p_1, p_2) encoding the strength of the extraneous random magnetic field by probability measures on $\mathbb{R}_+$, thus leading to the equation in (1.7). Intuitively, the reason why this makes sense is as follows. In the term $\sqrt{2 p_a} z_a \cdot \sigma_a$, the magnetic field acting on σ_a has a "trivial" structure. However, we need to have access to a richer term that allows to represent extraneous magnetic fields with an ultrametric structure, and this structure is described by its overlap distribution, a probability measure on $\mathbb{R}_+$. This construction, explained precisely below, defines an enriched free energy $\overline{F}_N = \overline{F}_N(t, \mu_1, \mu_2) : \mathbb{R}_+ \times (\mathcal{P}(\mathbb{R}_+))^2 \to \mathbb{R}$, and we will show that this enriched free energy is asymptotically bounded from below by the solution to (1.7); see Theorem 2.7 for a precise statement. As will be seen in the next section, the corresponding enriched Gibbs measure features extraneous variables, denoted α, which are in correspondence with the overlap structure of the random magnetic fields. A crucial step of the argument consists in showing that "typically", the overlaps $\sigma_a \cdot \sigma_a'$ can be inferred from the knowledge of the overlap between α and α'.

We now discuss related works. Fundamental insights on spin glasses, most notably the ultrametricity property, were first identified in the physics literature [48, 49, 36], where variational formulas for limit free energies were predicted. These predictions were then proved rigorously in [32, 53, 56, 57] in the setting of mixed p-spin models discussed above, under the assumption that the function ξ is convex over $\mathbb{R}$. The extension to the case of general ξ was achieved in [43, 44], and relies in particular on the justification that "typical" Gibbs measures are indeed organized along an asymptotically ultrametric structure. Further studies of particular relevance to the current paper concern the synchronization property, for models with multiple types of spins, or vector-valued spins [45, 46, 47]. Earlier works on spin-glass models with spins of multiple types include [55, 13, 12, 9, 3, 5].

Heuristic connections between limit free energies and partial differential equations were first pointed out in [31, 11, 1, 10], under a replica-symmetric or one-step replica symmetry breaking assumption. A rigorous identification of limit free energies of disordered systems in terms of Hamilton-Jacobi equations was obtained in [38, 39, 17, 18], in the context of problems of statistical inference. In this latter context, particular properties of the models allow to "close the equation" using only a finite number of additional variables; in other words, the Hamilton-Jacobi equations appearing there are finite-dimensional. The relevant partial differential equation for mixed p-spin models, namely (1.9), was then identified in [40]; an extension of this convergence, valid for the relevant enriched free energy, was conjectured there, and then proved in [41]. This last reference also describes

how to "remove" compensating terms such as the term $N^{-1}t|\sigma_1|^2|\sigma_2|^2$ appearing in (1.6), so that we can indeed end up with an upper bound on the limit of (1.2).

The rest of the paper is organized as follows. In Section 2, we define the enriched free energy, record some of its basic properties, and state a generalized version of Theorem 1.1, see Theorem 2.7. In Section 3, we define the precise notion of viscosity solution for (1.8), and define the solution to (1.7) as the limit of such finite-dimensional solutions. In Section 4, we show that if we restrict the free energy to measures that are sums of k Dirac masses with equal weights, then the function we obtain is a supersolution to (1.8), up to an error that goes to 0 as k goes to infinity; this allows us to conclude the proof of Theorem 2.7 (and thus also of Theorem 1.1). A crucial ingredient used in Section 4 is the fact that overlaps synchronize, and the justification of this is deferred to Section 5. In this section, we revisit the synchronization results of [45], emphasizing the notion of monotone couplings, and giving a "finitary" version of the statement of asymptotic synchronization. Finally, in Section 6, we discuss possible attempts at writing the solution to (1.7) as a saddle-point problem, and show that these tentative formulas are invalid. The appendix collects a handful of basic results on Gaussian integrals.

2. Definitions and basic properties

We write $\mathbb{N} = \{0, 1, \ldots\}$ to denote the set of natural numbers, $\mathbb{N}_* := \mathbb{N} \smallsetminus \{0\}$, and $\mathbb{R}_+ := [0, \infty)$. For every $x, y \in \mathbb{R}^N$, we write

$$x \cdot y := \sum_{i=1}^N x_i y_i, \qquad |x|^2 = x \cdot x.$$

We always implicitly understand that a vector $\sigma \in \mathbb{R}^{2N}$ is indexed according to $\sigma = (\sigma_{1,1}, \ldots, \sigma_{1,N}, \sigma_{2,1}, \ldots, \sigma_{2,N})$. We recall that $H_N(\sigma)$ was defined in (1.1), and notice that, for every $\sigma, \sigma' \in \mathbb{R}^{2N}$,

$$\mathbb{E}\left[H_N(\sigma)H_N(\sigma')\right] = N^{-1}\sum_{i,j=1}^N \sigma_{1,i}\sigma'_{1,i}\sigma_{2,j}\sigma'_{2,j}$$

$$(2.1) \qquad = N^{-1}(\sigma_1 \cdot \sigma'_1)(\sigma_2 \cdot \sigma'_2).$$

For every $t \geqslant 0$, we define

$$(2.2) \qquad H_N^t(\sigma) := \sqrt{2t}\, H_N(\sigma) - N^{-1}t\,|\sigma_1|^2|\sigma_2|^2.$$

We are now going to introduce another energy function, parametrized by $\mu = (\mu_1, \mu_2) \in (\mathcal{P}(\mathbb{R}_+))^2$. It is much more convenient to describe and to work with this object in the case when the measures are discrete, and then simply argue by continuity. We therefore give ourselves an integer $k \geqslant 0$, and parameters

$$(2.3) \qquad 0 = \zeta_0 < \zeta_1 \leqslant \zeta_2 \leqslant \cdots \leqslant \zeta_{k-1} \leqslant \zeta_k < \zeta_{k+1} = 1,$$

$$(2.4) \qquad 0 = q_{a,-1} \leqslant q_{a,0} \leqslant q_{a,1} \leqslant \cdots \leqslant q_{a,k} < q_{a,k+1} = \infty \qquad (a \in \{1, 2\}),$$

and we set, for every $a \in \{1, 2\}$,

$$(2.5) \qquad \mu_a = \sum_{\ell=0}^k (\zeta_{\ell+1} - \zeta_\ell)\delta_{q_{a,\ell}}.$$

These measures will serve to parametrize certain ultrametric structures with a prescribed overlap distribution. We instantiate the rooted tree with (countably) infinite degree and depth k as

$$(2.6) \qquad \mathcal{A} := \mathbb{N}^0 \cup \mathbb{N} \cup \mathbb{N}^2 \cup \cdots \cup \mathbb{N}^k,$$

where $\mathbb{N}^0 = \{\varnothing\}$, and $\varnothing$ represents the root of the tree. For every $\alpha \in \mathbb{N}^\ell$, we write $|\alpha| := \ell$ to denote the depth of the vertex α in the tree $\mathcal{A}$. For every leaf $\alpha = (n_1, \ldots, n_k) \in \mathbb{N}^k$ and $\ell \in \{0, \ldots, k\}$, we write

$$\alpha_{|\ell} := (n_1, \ldots, n_\ell),$$

with the understanding that $\alpha_{|0} = \varnothing$. We also give ourselves a family $(z_{\alpha,a,i})_{\alpha \in \mathcal{A}, a \in \{1,2\}, 1 \leqslant i \leqslant N}$ of independent standard Gaussians, independent of H_N, and we let $(v_\alpha)_{\alpha \in \mathbb{N}^k}$ be a Poisson-Dirichlet cascade with weights given by the family $(\zeta_\ell)_{1 \leqslant \ell \leqslant k}$. We refer to [44, (2.46)] for a precise definition, and only mention here a few important points. First, in the case $k = 0$, we simply set $v_\varnothing = 1$. Second, in the case $k = 1$, the weights $(v_\alpha)_{\alpha \in \mathbb{N}}$ are obtained by normalizing a Poisson point process on $(0, \infty)$ with intensity measure $\zeta_1 x^{-1-\zeta_1} \, \mathrm{d}x$ so that $\sum_\alpha v_\alpha = 1$. Third, for general $k \geqslant 1$, the progeny of each non-leaf vertex at level $\ell \in \{0, \ldots, k-1\}$ is decorated with the values of an independent Poisson point process of intensity measure $\zeta_{\ell+1} x^{-1-\zeta_{\ell+1}} \, \mathrm{d}x$, then the weight of a given leaf $\alpha \in \mathbb{N}^k$ is calculated by taking the product of the "decorations" attached to each parent vertex, including the leaf vertex itself (but excluding the root, which has no assigned "decoration"), and finally, these weights over leaves are normalized so that their total sum is 1. We take this Poisson-Dirichlet cascade $(v_\alpha)_{\alpha \in \mathbb{N}^k}$ to be independent of H_N and of the random variables $(z_{\alpha,a,i})_{\alpha \in \mathcal{A}, a \in \{1,2\}, 1 \leqslant i \leqslant N}$. For every $\sigma \in \mathbb{R}^{2N}$ and $\alpha \in \mathbb{N}^k$, we set

$$(2.7) \qquad H_N^\mu(\sigma, \alpha) := \sum_{a=1}^{2} \left(\sum_{\ell=0}^{k} \left(2q_{a,\ell} - 2q_{a,\ell-1} \right)^{\frac{1}{2}} z_{\alpha_{|\ell},a} \cdot \sigma_a - q_{a,k} |\sigma_a|^2 \right),$$

where we write $z_{\alpha_{|\ell},a} \cdot \sigma_a = \sum_{i=1}^{N} z_{\alpha_{|\ell},a,i} \sigma_{a,i}$. The random variables $(H_N^\mu(\sigma, \alpha))_{\sigma \in \mathbb{R}^{2N}, \alpha \in \mathbb{N}^k}$ form a Gaussian family which is independent of $(H_N(\sigma))_{\sigma \in \mathbb{R}^{2N}}$. We understand that the symbol $\mathbb{E}$ stands for the expectation with respect to (J_{ij}), $(z_\alpha)_{\alpha \in \mathcal{A}}$ and $(v_\alpha)_{\alpha \in \mathbb{N}^k}$. Notice that, for each fixed choice of $\alpha, \alpha' \in \mathbb{N}^k$, we have

$$(2.8) \qquad \frac{1}{N} \mathbb{E}\left[\left(\sum_{\ell=0}^{k} \left(2q_{a,\ell} - 2q_{a,\ell-1} \right)^{\frac{1}{2}} z_{\alpha_{|\ell},a} \right) \cdot \left(\sum_{\ell=0}^{k} \left(2q_{a,\ell} - 2q_{a,\ell-1} \right)^{\frac{1}{2}} z_{\alpha'_{|\ell},a} \right) \right] = 2q_{a,\alpha \wedge \alpha'},$$

where we write

$$(2.9) \qquad \alpha \wedge \alpha' := \sup\{\ell \leqslant k \ : \ \alpha_{|\ell} = \alpha'_{|\ell}\}.$$

The point of the construction in (2.7) is to provide with a more refined "external field" than that introduced in (1.10). Indeed, if we sample two independent copies $\alpha, \alpha' \in \mathbb{N}^k$ according to the weights $(v_\alpha)_{\alpha \in \mathbb{N}^k}$, then the law of overlap

$$\frac{1}{N} \left(\sum_{\ell=0}^{k} \left(2q_{a,\ell} - 2q_{a,\ell-1} \right)^{\frac{1}{2}} z_{\alpha_{|\ell},a} \right) \cdot \left(\sum_{\ell=0}^{k} \left(2q_{a,\ell} - 2q_{a,\ell-1} \right)^{\frac{1}{2}} z_{\alpha'_{|\ell},a} \right)$$

under the measure in which we average over (z_α) and (v_α) is μ_a (this can be inferred from Lemma 2.3 below or, more directly, from [44, (2.34)]). We define

$$(2.10) \qquad F_N(t, \mu) := -\frac{1}{N} \log \int \sum_{\alpha \in \mathbb{N}^k} \exp\left(H_N^t(\sigma) + H_N^\mu(\sigma, \alpha) \right) v_\alpha \mathrm{d}P_N(\sigma).$$

We also define the Gibbs measure $\langle \cdot \rangle$, with canonical random variable (σ, α) taking values in $\mathbb{R}^{2N} \times \mathbb{N}^k$, in such a way that, for any bounded measurable function f,

$$(2.11)$$
$$\langle f(\sigma, \alpha) \rangle := \exp\left(N F_N(t, \mu) \right) \int \sum_{\alpha \in \mathbb{N}^k} f(\sigma, \alpha) \exp\left(H_N^t(\sigma) + H_N^\mu(\sigma, \alpha) \right) v_\alpha \mathrm{d}P_N(\sigma),$$

We also allow ourselves to consider multiple independent copies, or "replicas", of the random variable (σ, α), which we may denote by (σ', α'), (σ'', α''), and so on. Alternatively, in situations where many independent replicas need to be considered, we also

denote these replicas by $(\sigma^\ell, \alpha^\ell)_{\ell \geqslant 1}$. Recall that the measure $\langle \cdot \rangle$ is itself random; while the replicas are independent under $\langle \cdot \rangle$, conditionally on the randomness "extraneous" to the measure, they are no longer independent after we average further.

We denote by $\widetilde{F}_N$ the average of F_N with respect to the random variables (z_α) and (v_α). Since the only additional source of randomness in the problem comes from the J's in the definition of H_N, and since these are independent random variables, we can write

$$(2.12) \qquad \widetilde{F}_N(t, \mu) = \mathbb{E}\left[F_N(t, \mu) \,\middle|\, (H_N(\sigma))_{\sigma \in \mathbb{R}^{2N}} \right].$$

We also define the fully averaged free energy

$$(2.13) \qquad \overline{F}_N(t, \mu) \coloneqq \mathbb{E}\left[F_N(t, \mu) \right].$$

The notation just introduced suggests that these quantities depend on the parameters ζ and q in (2.3) and (2.4) only insofar as they affect the measures μ_1 and μ_2. The next proposition states that this is indeed the case, at least as far as the quantities $\widetilde{F}_N(t, \mu)$ and $\overline{F}_N(t, \mu)$ are concerned. (It would make more sense to speak of distributional identities for H_N^μ and $F_N(t, \mu)$; since such considerations will not play any role in this paper, we simply accept a slightly abusive notation for these latter two quantities.) It also states that $\widetilde{F}_N(t, \mu)$, and therefore also $\overline{F}_N(t, \mu)$, satisfy a Lipschitz estimate in their dependence in μ. Recall that the random variables of X_ν appearing in the statement were defined in (1.5).

Proposition 2.1 (Lipschitz continuity of $\widetilde{F}_N$). *The functions $\widetilde{F}_N(t, \mu)$ and $\overline{F}_N(t, \mu)$ depend in the parameters ζ and q in (2.3) and (2.4) only through their effect on the measures (μ_1, μ_2) in (2.5). Moreover, for every $t \geqslant 0$ and any two pairs $\mu, \mu' \in (\mathcal{P}(\mathbb{R}_+))^2$ of measures of finite support, we have*

$$(2.14) \qquad |\widetilde{F}_N(t, \mu) - \widetilde{F}_N(t, \mu')| \leqslant \sum_{a=1}^{2} \mathbb{E}\left[|X_{\mu_a} - X_{\mu_a'}| \right],$$

and the same inequality also holds with $\widetilde{F}_N$ replaced by $\overline{F}_N$. In particular, $\widetilde{F}_N$ and $\overline{F}_N$ can be extended by continuity to $\mathbb{R}_+ \times (\mathcal{P}_1(\mathbb{R}_+))^2$.

One possible way to prove Proposition 2.1 is to rely on the following two results. The first one describes a relatively concrete procedure for computing averages over Poisson-Dirichlet cascades; see [44, Theorem 2.9] for a proof.

Proposition 2.2 (Integration of Poisson-Dirichlet cascades). *Let $(\omega_\alpha)_{\alpha \in \mathcal{A}}$ be independent and identically distributed random variables taking values in some measurable space E, independent of the Poisson-Dirichlet cascade $(v_\alpha)_{\alpha \in \mathbb{N}^k}$. Let $X_k : E^r \to \mathbb{R}$ be a measurable function, and denote*

$$X_{-1} \coloneqq \mathbb{E} \log \sum_{\alpha \in \mathbb{N}^k} \exp\left(X_k(\omega_{\alpha|0}, \ldots, \omega_{\alpha|k}) \right) v_\alpha.$$

In the expression above, the expectation $\mathbb{E}$ is with respect to the law of $(\omega_\alpha)_{\alpha \in \mathcal{A}}$ and $(v_\alpha)_{\alpha \in \mathbb{N}^k}$. Define recursively, for every $\ell \in \{1, \ldots, k\}$, the measurable function $X_{\ell-1} : E^{\ell-1} \to \mathbb{R} \cup \{+\infty\}$ given by

$$X_{\ell-1}(\omega_0, \ldots, \omega_{\ell-1}) \coloneqq \zeta_\ell^{-1} \log \mathbb{E}_{\omega_\ell} \exp\left(\zeta_\ell X_\ell(\omega_0, \ldots, \omega_\ell) \right),$$

where, for every $\ell \in \{0, \ldots, k\}$, we write $\mathbb{E}_{\omega_\ell}$ to denote the integration of the variable ω_ℓ along the law of any of the variables $(\omega_\alpha)_{\alpha \in \mathcal{A}}$. We have

$$X_{-1} = \mathbb{E}_{\omega_0}\left[X_0(\omega_0) \right].$$

In the statement above, the random variables under each expectation are implicitly assumed to be integrable. In our context, we can apply this lemma in the following way: we set $\omega_\alpha := z_\alpha$ and, for every $y_0 = (y_{0,a,i})_{a \in \{1,2\}, 1 \leqslant i \leqslant N}, \ldots, y_k = (y_{k,a,i})_{a \in \{1,2\}, 1 \leqslant i \leqslant N} \in \mathbb{R}^{2N}$,

$$(2.15) \quad X_k(y_0, \ldots, y_k)$$
$$:= \log \int \exp\left(H_N^t(\sigma) + \sum_{a=1}^{2} \left(\sum_{\ell=0}^{k} \left(2q_{a,\ell} - 2q_{a,\ell-1} \right)^{\frac{1}{2}} y_{\ell,a} \cdot \sigma_a - q_{a,k} |\sigma_a|^2 \right) \right) \, \mathrm{d}P_N(\sigma).$$

We then define recursively, for every $\ell \in \{1, \ldots, k\}$,

$$(2.16) \qquad X_{\ell-1}(y_0, \ldots, y_{\ell-1}) := \zeta_\ell^{-1} \log \mathbb{E}_{y_\ell} \exp\left(\zeta_\ell X_\ell(y_0, \ldots, y_\ell) \right),$$

where, for every $\ell \in \{0, \ldots, k\}$, we write $\mathbb{E}_{y_\ell}$ to denote the integration of the variable $y_\ell \in \mathbb{R}^{2N}$ along the standard Gaussian measure. Proposition 2.2 then ensures that

$$-N \, \widetilde{F}_N(t, \mu) = \mathbb{E}_{y_0}\left[X_0(y_0) \right].$$

(A more careful argument would start by using Proposition 2.2 to verify that $|F_N(t, \mu)|$ is indeed integrable.)

The next lemma identifies the law of the overlap $\alpha \wedge \alpha'$ under the averaged measure $\mathbb{E}\langle \cdot \rangle$. The proof can be found for instance in [44, (2.82)] or [40, Lemma 2.3].

Lemma 2.3 (overlaps of Poisson-Dirichlet cascades). *For every $\ell \in \{0, \ldots, k\}$, we have*

$$\mathbb{E}\left\langle \mathbb{1}_{\{\alpha \wedge \alpha' = \ell\}} \right\rangle = \zeta_{\ell+1} - \zeta_\ell.$$

The combination of Proposition 2.2 and Lemma 2.3 allows to prove Proposition 2.1, see for instance [40, Proposition 2.1].

While we usually think of $\widetilde{F}_N$ and $\overline{F}_N$ as functions of the pair of measures μ, we also allow ourselves to speak of $\partial_{q_{a,\ell}} \widetilde{F}_N$ and $\partial_{q_{a,\ell}} \overline{F}_N$; this is meant to refer to the point of view in which these are seen as functions of the families of parameters q and ζ in (2.4) and (2.3). Another consequence of Proposition 2.2, which can be found for instance in [57, Proposition 14.3.2] or [40, Lemma 2.4], is that the derivatives of $\widetilde{F}_N$ with respect to each of the parameters $q_{a,\ell}$ in (2.4) are nonnegative, and they increase with ℓ after suitable normalization. The precise statement is as follows.

Lemma 2.4. *For every $a \in \{1,2\}$ and $\ell \leqslant \ell' \in \{0, \ldots, k\}$, we have*

$$(2.17) \qquad \partial_{q_{a,\ell}} \widetilde{F}_N \geqslant 0,$$

and

$$(2.18) \qquad (\zeta_{\ell+1} - \zeta_\ell)^{-1} \partial_{q_{a,\ell}} \widetilde{F}_N \leqslant (\zeta_{\ell'+1} - \zeta_{\ell'})^{-1} \partial_{q_{a,\ell'}} \widetilde{F}_N.$$

Remark 2.5. Clearly, the statement of Lemma 2.4 is also valid with $\widetilde{F}_N$ replaced by $\overline{F}_N$. It is part of the statement of this lemma that the quantity $(\zeta_{\ell+1} - \zeta_\ell)^{-1} \partial_{q_{a,\ell}} \widetilde{F}_N$ can be defined even when $\zeta_{\ell+1} = \zeta_\ell$, by continuity.

Yet another consequence of Proposition 2.2 concerns the "initial condition" for $\overline{F}_N$. Under the assumption of (1.3), the verification that $\overline{F}_N(0, \mu)$ converges as N tends to infinity is particularly simple.

Lemma 2.6 (Initial condition for product measures). *Recall that we assume* (1.3). *For every $N \geqslant 1$ and $\mu \in (\mathcal{P}_1(\mathbb{R}_+))^2$, we have*

$$\overline{F}_N(0, \mu) = \overline{F}_1(0, \mu).$$

Proof. The argument can be found for instance in [44, (2.60)]; we present it briefly here for the reader's convenience. When $t = 0$, and under the assumption of (1.3), the definition of X_k given in (2.15) can be rewritten as

$$X_k(y_0, \ldots, y_k) = \sum_{i=1}^{N} \log \int \exp \left(\sum_{a=1}^{2} \left(\sum_{\ell=0}^{k} \left(2q_{a,\ell} - 2q_{a,\ell-1} \right)^{\frac{1}{2}} y_{\ell,a,i} \sigma_a - q_{a,k} \sigma_a^2 \right) \right) \mathrm{d}P_1(\sigma).$$

Recall that P_1 is a probability measure over $\mathbb{R}^2$, so in the integral above, the variable σ takes the form $\sigma = (\sigma_a)_{a \in \{1,2\}} \in \mathbb{R}^2$. In particular, we have written X_k as a sum of independent and identically distributed random variables. Moreover, the law of each of these random variables does not depend on N. These properties are preserved as we go along the recursive procedure described in (2.16). As we reach X_{-1}, all randomness has been integrated out, and the result is thus N times some constant, as desired. $\qquad \square$

With an eye towards the initial condition in (1.7), we set, for every $\mu \in (\mathcal{P}_1(\mathbb{R}_+))^2$,

$$(2.19) \qquad \psi(\mu) := \overline{F}_1(0, \mu).$$

It is worth keeping in mind that the relatively simple definition of the initial condition in (2.19) is possible only because we made the assumption in (1.3) that the underlying measure has a product structure. In general, we only want to ascertain that for every $\mu \in (P_1(\mathbb{R}_+))^2$,

$$\lim_{N \to \infty} \overline{F}_N(0, \mu) \quad \text{exists};$$

and in this case, we call the limit $\psi(\mu)$. (We also use in the course of the proof that the support of P_N lies in a ball of fixed radius.) Other choices of reference measure are thus possible: for instance, one may replace P_N by the uniform measure on the product of two N-dimensional spheres of radius $\sqrt{N}$. See for instance [40, part (2) of Proposition 3.1] for a similar calcuation in this case (which itself borrows from [52]).

We now state the extended version of Theorem 1.1 that will be the main focus of the rest of the paper.

Theorem 2.7. *For every $t \geqslant 0$ and $\mu \in (\mathcal{P}_2(\mathbb{R}_+))^2$, we have*

$$\liminf_{N \to \infty} \overline{F}_N(t, \mu) \geqslant f(t, \mu),$$

where $f : \mathbb{R}_+ \times (\mathcal{P}_2(\mathbb{R}_+))^2 \to \mathbb{R}$ is the solution to (1.7).

The statement of Theorem 1.1 corresponds to the case $\mu = (\delta_0, \delta_0)$ in Theorem 2.7. We now discuss why one should expect that $\overline{F}_N$ indeed converges to the solution to (1.7). We first observe that

$$(2.20) \qquad \partial_t F_N = \frac{1}{N} \left\langle \frac{1}{\sqrt{2t}} H_N(\sigma) - \frac{1}{N} |\sigma_1|^2 |\sigma_2|^2 \right\rangle.$$

Taking the expectation, recalling (2.1), and using a Gaussian integration by parts, see (A.2), we obtain that

$$(2.21) \qquad \partial_t \overline{F}_N = \frac{1}{N^2} \mathbb{E} \left\langle (\sigma_1 \cdot \sigma_1')(\sigma_2 \cdot \sigma_2') \right\rangle.$$

By the same reasoning (or see for instance [40, (2.17)]), we have

$$(2.22) \qquad \partial_{q_{a,\ell}} \overline{F}_N = \frac{1}{N} \mathbb{E} \left\langle \sigma_a \cdot \sigma_a' \mathbb{1}_{\{\alpha \wedge \alpha' = \ell\}} \right\rangle.$$

Using Lemma 2.3, we can rewrite this identity as

$$(\zeta_{\ell+1} - \zeta_\ell)^{-1} \partial_{q_{a,\ell}} \overline{F}_N = \frac{1}{N} \mathbb{E} \left\langle \sigma_a \cdot \sigma_a' \,\middle|\, \alpha \wedge \alpha' = \ell \right\rangle,$$

where the conditional expectation is understood with respect to the measure $\mathbb{E}\langle\cdot\rangle$. We deduce that

$$\int \partial_{\mu_1}\overline{F}_N\,\partial_{\mu_2}\overline{F}_N\;\mathrm{d}\widehat{\mu} = \sum_{\ell=1}^{k}(\zeta_{\ell+1}-\zeta_\ell)^{-1}\partial_{q_{1,\ell}}\overline{F}_N\,\partial_{q_{2,\ell}}\overline{F}_N$$

$$= \frac{1}{N^2}\sum_{\ell=1}^{k}\mathbb{E}\left\langle\mathbb{1}_{\{\alpha\wedge\alpha'=\ell\}}\right\rangle\mathbb{E}\left\langle\sigma_1\cdot\sigma_1'\,\big|\,\alpha\wedge\alpha'=\ell\right\rangle\mathbb{E}\left\langle\sigma_2\cdot\sigma_2'\,\big|\,\alpha\wedge\alpha'=\ell\right\rangle$$

$$= \frac{1}{N^2}\mathbb{E}\left\langle\mathbb{E}\left\langle\sigma_1\cdot\sigma_1'\,\big|\,\alpha\wedge\alpha'\right\rangle\mathbb{E}\left\langle\sigma_2\cdot\sigma_2'\,\big|\,\alpha\wedge\alpha'\right\rangle\right\rangle.$$

We can now compare this expression with (2.21), and also with the situation encountered in the more naive attempt leading to (1.11). In the naive attempt, we could only hope to close the equation in situations for which the overlaps $\sigma_a\cdot\sigma_a'$ are concentrated. In our current more refined attempt, we have instead

$$\partial_t\overline{F}_N - \int \partial_{\mu_1}\overline{F}_N\,\partial_{\mu_2}\overline{F}_N\;\mathrm{d}\widehat{\mu}$$

$$= \frac{1}{N^2}\mathbb{E}\left\langle\left(\sigma_1\cdot\sigma_1' - \mathbb{E}\left\langle\sigma_1\cdot\sigma_1'\,\big|\,\alpha\wedge\alpha'\right\rangle\right)\left(\sigma_2\cdot\sigma_2' - \mathbb{E}\left\langle\sigma_2\cdot\sigma_2'\,\big|\,\alpha\wedge\alpha'\right\rangle\right)\right\rangle,$$

and in particular,

$$(2.23)\qquad \left|\partial_t\overline{F}_N - \int \partial_{\mu_1}\overline{F}_N\,\partial_{\mu_2}\overline{F}_N\;\mathrm{d}\widehat{\mu}\right| \leqslant \frac{1}{N^2}\sum_{a\in\{1,2\}}\mathbb{E}\left\langle\left(\sigma_a\cdot\sigma_a' - \mathbb{E}\left\langle\sigma_a\cdot\sigma_a'\,\big|\,\alpha\wedge\alpha'\right\rangle\right)^2\right\rangle.$$

In other words, we need to argue that the *conditional variance* of the overlaps $\sigma_a\cdot\sigma_a'$, given the overlap $\alpha\wedge\alpha'$, is small. This is precisely what the synchronization property should give us. (Moreover, there is some flexibility in that we do not need that this conditional variance be small for any single choice of the parameters.) From this point of view, the synchronization property becomes central even for models with a single type, since the point is to monitor synchronization with the extraneous random variables provided by the Poisson-Dirichlet cascade.

3. Viscosity solutions

The first goal of this section is to clarify the exact notion of solution for finite-dimensional approximations to (1.7), and show comparison principles for these finite-dimensional problems. The second goal is to show that as we increase the dimension, the sequence of finite-dimensional solutions converges to some limit. We then interpret the limit as the solution to (1.7).

In this context, the convex cone $\mathbb{R}_+^2$ plays a fundamental role. In more general models of mean-field spin glasses, this convex cone would have to be replaced by the set of positive semidefinite matrices. In the setting of the bipartite model, matrices are not obviously showing up because, in some sense, we are only looking at the diagonal entries of a 2-by-2 matrix: observe that there is no term of the form $\sigma_1\cdot\sigma_2'$ on the right side of (2.1). In order to avoid future repetitions, I found it useful to write this section so that it covers the two settings at once. Throughout this section, we keep the integer $D\geqslant 1$ fixed, and denote by S^D, S_+^D, and S_{++}^D the set of symmetric D-by-D matrices, and the subsets of positive semidefinite and positive definite matrices respectively. We define

$$(3.1)\qquad\qquad\qquad \mathcal{C} := (0,\infty)^D \qquad\text{or}\qquad \mathcal{C} := S_{++}^D,$$

its closure, given respectively by

$$(3.2)\qquad\qquad\qquad \overline{\mathcal{C}} := \mathbb{R}_+^D \qquad\text{or}\qquad \overline{\mathcal{C}} := S_+^D,$$

and observe that $\overline{\mathcal{C}}$ is a convex cone within its natural ambient vector space, namely

(3.3) $$\mathcal{E} := \mathbb{R}^D \qquad \text{or} \qquad \mathcal{E} := S^D.$$

When $\mathcal{E} = S^D$, we interpret the scalar product between two matrices $a = (a_{dd'})_{1 \leqslant d,d' \leqslant D}$ $b = (b_{dd'})_{1 \leqslant d,d' \leqslant D} \in S^D$ according to

$$a \cdot b := \sum_{d,d'=1}^D a_{dd'} b_{dd'} = \operatorname{tr}(a^* b),$$

with a^* denoting the transpose of the matrix a. We also write $|a| := (a \cdot a)^{\frac{1}{2}}$. In both settings, the convex cone $\overline{\mathcal{C}}$ defines a partial order: for every $x, y \in \mathcal{E}$, we write $x \leqslant y$ whenever $y - x \in \overline{\mathcal{C}}$. We could also use the notation $x < y$ whenever $y - x \in \mathcal{C}$, however I will refrain from doing so, in order to avoid possible confusions that would arise from the fact that the conjunction of $x \leqslant y$ and $x \neq y$ does not imply $x < y$.

Let $K \geqslant 1$ be an integer. We define the open set

(3.4) $$U_K := \left\{ x = (x_1, \ldots, x_K) \in \mathcal{C}^K \ : \ \forall k \in \{1, \ldots, K-1\}, \ x_{k+1} - x_k \in \mathcal{C} \right\},$$

and its closure

(3.5) $$\overline{U}_K := \left\{ x = (x_1, \ldots, x_K) \in \overline{\mathcal{C}}^K \ : \ x_1 \leqslant \cdots \leqslant x_K \right\}.$$

The first goal of this section is to study the existence and uniqueness of solutions to the equation

(3.6) $$\partial_t f - \mathsf{H}(\nabla f) = 0 \qquad \text{in } (0, T) \times U_K,$$

for a given locally Lipschitz function $\mathsf{H} : \mathcal{E}^K \to \mathbb{R}$, $T \in (0, \infty]$, and with a prescribed initial condition at $t = 0$. In the expression above, we use the notation, with the understanding that $f = f(t, x)$ with $x = (x_1, \ldots, x_K)$,

$$\nabla f := (\partial_{x_1} f, \ldots, \partial_{x_K} f),$$

where in this expression, each $\partial_{x_k} f$ takes values in the set $\mathcal{E}$. We will also impose a Neumann boundary condition on ∂U_K for solutions to (3.6). Since the domain $\overline{U}_K$ has corners, we define the outer normal to a point $x \in \partial U_K$ as the set

$$\mathbf{n}(x) := \left\{ \nu \in \mathcal{E}^K \ : |\nu| = 1 \ \text{ and } \ \forall y \in \overline{U}_K, \ (y - x) \cdot \nu \leqslant 0 \right\}.$$

(This definition would have to be modified for non-convex domains.) To display the Neumann boundary condition, we write the equation formally as

(3.7) $$\begin{cases} \partial_t f - \mathsf{H}(\nabla f) = 0 & \text{in } (0, T) \times U_K, \\ \mathbf{n} \cdot \nabla f = 0 & \text{on } (0, T) \times \partial U_K. \end{cases}$$

In order to study the equation (3.7), we rely on the notion of viscosity solutions. Although the techniques used here to handle the equation (3.7) do not differ much from classical arguments, I could not find results in the literature that would prove the well-posedness of viscosity solutions in non-smooth domains such as $\overline{U}_K$. The best result I could find is [23], where the authors consider the case where the domain is the intersection of a finite number of open sets with a smooth boundary that satisfy certain conditions.

Our main interest for studying solutions to (3.7) resides in the fact that we will then define the solution to (1.7) as the limit of solutions to such finite-dimensional problems. Several other works have also considered Hamilton-Jacobi equations posed on spaces of probability measures or other infinite-dimensional spaces [20, 21, 22, 25, 26, 15, 16, 29, 30, 2, 14]. However, I am not aware of results that show the well-posedness of equations of the type of (1.7); or that include the handling of a boundary condition; or that discuss the convergence of finite-dimensional approximations. These aspects will be covered here.

The remainder of this section is made of two parts. We first study finite-dimensional equations of the form (3.7); and then show how to pass to the limit and identify the solution to (1.7).

3.1. **Analysis of finite-dimensional equations.** The precise definition of solution to the equation (3.7) reads as follows.

Definition 3.1. We say that a function $f \in C([0,T] \times \overline{U}_K)$ is a *viscosity subsolution* to (3.7) if for every $(t,x) \in (0,T) \times \overline{U}_K$ and $\phi \in C^\infty((0,T) \times \overline{U}_K)$ such that (t,x) is a local maximum of $f - \phi$, we have

$$(3.8) \qquad (\partial_t \phi - \mathsf{H}(\nabla \phi))(t,x) \leqslant 0 \quad \text{if } x \in U_K,$$

while, if $x \in \partial U_K$,

$$(3.9) \qquad \min\left(\inf_{\nu \in \mathbf{n}(x)} \nabla \phi \cdot \nu, \partial_t \phi - \mathsf{H}(\nabla \phi) \right)(t,x) \leqslant 0.$$

We say that a function $f \in C([0,T] \times \overline{U}_K)$ is a *viscosity supersolution* to (3.7) if for every $(t,x) \in (0,T) \times \overline{U}_K$ and $\phi \in C^\infty((0,T) \times \overline{U}_K)$ such that (t,x) is a local minimum of $f - \phi$, we have

$$(\partial_t \phi - \mathsf{H}(\nabla \phi))(t,x) \geqslant 0 \quad \text{if } x \in U_K,$$

while, if $x \in \partial U_K$,

$$(3.10) \qquad \max\left(\sup_{\nu \in \mathbf{n}(x)} \nabla \phi \cdot \nu, \partial_t \phi - \mathsf{H}(\nabla \phi) \right)(t,x) \geqslant 0.$$

We say that a function $f \in C([0,T] \times \overline{U}_K)$ is a *viscosity solution* to (3.7) if it is both a viscosity subsolution and a viscosity supersolution to (3.7).

We may drop the qualifier *viscosity* and simply talk about subsolutions, supersolutions, and solutions to (3.7). We say that a function $f \in C([0,T] \times \overline{U}_K)$ is a *solution* to

$$(3.11) \qquad \begin{cases} \partial_t f - \mathsf{H}(\nabla f) \leqslant 0 & \text{in } (0,T) \times U_K, \\ \mathbf{n} \cdot \nabla f \leqslant 0 & \text{on } (0,T) \times \partial U_K, \end{cases}$$

whenever it is a subsolution to (3.7); and similarly with the inequalities reversed for supersolutions.

Historically, the notion of viscosity solutions emerged from the following construction of solutions: for a small parameter $\varepsilon > 0$, one considers the solution of the partial differential equation

$$\begin{cases} \partial_t f_\varepsilon - \mathsf{H}(\nabla f_\varepsilon) = \varepsilon \Delta f_\varepsilon, & \text{in } (0,T) \times U_K, \\ \mathbf{n} \cdot \nabla f_\varepsilon = 0 & \text{on } (0,T) \times \partial U_K, \end{cases}$$

and then one identifies the viscosity solution to (3.7) as the limit of f_ε as ε tends to zero. As will be seen below, the limit satisfies a form of maximum principle, as each of these approximations do. One can also consult [24, Section III.10.1] for more intuition concerning the definition of viscosity solutions.

The most useful result concerning solutions to (3.7) for our purposes is a comparison principle. For every $x \in \mathcal{E}^K$, we write

$$|x| := \left(\sum_{k=1}^K |x_k|^2 \right)^{\frac{1}{2}},$$

where $|x_k|$ stands for the standard Euclidean norm in $\mathcal{E}$, and, for every $r \in \mathbb{R}$, we write $r_+ := \max(r,0)$.

Proposition 3.2 (Comparison principle). *Let $T \in (0, \infty)$, and let u and v be respectively a sub- and a super-solution to (3.7) that are both uniformly Lipschitz continuous in the x variable. We have*

$$(3.12) \qquad \sup_{[0,T) \times \overline{U}_K} (u - v) = \sup_{\{0\} \times \overline{U}_K} (u - v).$$

More precisely, let

$$(3.13) \qquad L := \max(\| \, |\nabla u| \, \|_{L^\infty([0,T) \times U_K)}, \| \, |\nabla v| \, \|_{L^\infty([0,T) \times U_K)}),$$

and, for some arbitrary $\delta > 0$, let

$$(3.14) \qquad V := \sup \left\{ \frac{|\mathsf{H}(p') - \mathsf{H}(p)|}{|p' - p|} \; : \; |p|, |p'| \leqslant L + \delta \right\}.$$

For every $R, M \in \mathbb{R}$ such that

$$(3.15) \qquad M > \sup_{0 \leqslant t < T, x \in \overline{U}_K} \frac{u(t,x) - v(t,x)}{1 + |x|},$$

the mapping

$$(3.16) \qquad (t, x) \mapsto u(t,x) - v(t,x) - M\left(|x| + Vt - R\right)_+$$

achieves its supremum at a point in $\{0\} \times \overline{U}_K$.

Before turning to the proof of this proposition, it will be useful to identify the cone dual to the convex cone $\overline{U}_K$.

Lemma 3.3 (Dual cone to $\overline{U}_K$). *Let $\overline{U}_K^*$ denote the cone dual to $\overline{U}_K$, that is,*

$$(3.17) \qquad \overline{U}_K^* := \left\{ x \in \mathcal{E}^K \; : \; \forall v \in \overline{U}_K, \; x \cdot v \geqslant 0 \right\}.$$

We have

$$(3.18) \qquad \overline{U}_K^* = \left\{ x \in \mathcal{E}^K \; : \; \forall k \in \{1, \dots, K\}, \; \sum_{\ell=k}^{K} x_\ell \geqslant 0 \right\},$$

and

$$(3.19) \qquad \overline{U}_K = \left\{ v \in \mathcal{E}^K \; : \; \forall x \in \overline{U}_K^*, \; x \cdot v \geqslant 0 \right\}.$$

Proof. For concreteness, we write the proof in the case when $\mathcal{C} = S_{++}^D$. Let $x, v \in (S^D)^K$. Setting $v_0 := 0$, we have

$$(3.20) \qquad x \cdot v = \sum_{k=1}^{K} x_k \cdot v_k = \sum_{k=1}^{K} \sum_{\ell=k}^{K} x_\ell \cdot (v_k - v_{k-1}).$$

It is therefore clear that the set on the right side of (3.18) is contained in $\overline{U}_K^*$ (recall that if $a, b \in S_+^D$, then $a \cdot b = |\sqrt{a}\sqrt{b}|^2 \geqslant 0$). Conversely, if $x \in (S^D)^K$ does not belong to the set on the right side of (3.18), then there exists $k \in \{1, \dots, K\}$ such that

$$\sum_{\ell=k}^{K} x_\ell \notin S_+^D.$$

Letting $\mathfrak{p}$ denote the orthogonal projection onto the eigenspaces with negative eigenvalues of the matrix on the left side of the display above, and setting

$$v = \Big(\underbrace{0, \dots, 0}_{k-1 \text{ terms}}, \mathfrak{p}, \dots, \mathfrak{p} \Big) \in \overline{U}_K,$$

we find that $x \cdot v < 0$, so $x \notin \overline{U}_K^*$. This shows (3.18).

The proof of (3.19) could be derived from a general statement concerning the bidual of closed convex cones; see Step 1 of the proof of Proposition 3.6 below. We rather provide with a more elementary and explicit argument. It is clear from (3.20) that $\overline{U}_K$ is contained in the set on the right side of (3.19). Conversely, recall that a matrix $a \in S^D$ belongs to S^D_+ if and only if, for every $b \in S^D_+$, we have $a \cdot b \geqslant 0$ (see also [39, Lemma 2.2]). Let v belong to the set on the right side of (3.19). To see that $v \in \overline{U}_K$, it thus suffices to show that for every $b_1, \ldots, b_K \in S^D_+$, one can find $x_1, \ldots, x_K \in S^D$ such that, for every $k \in \{1, \ldots, K\}$, we have

$$\sum_{\ell=k}^K x_\ell = b_k.$$

It suffices to set, for every $k \in \{1, \ldots, K\}$, $x_k := b_k - b_{k+1}$, with the notation $b_{K+1} := 0$. $\quad\square$

Proof of Proposition 3.2. For concreteness, we write the proof in the case when $\mathcal{C} = S^D_{++}$. The case of $\mathcal{C} = (0, \infty)^D$ is only easier.

Since the second part of the statement implies the first part, we focus on the former. Without loss of generality, we assume that the functions u and v are continuous on $[0, T] \times \overline{U}_K$ (once the result is proved in this case, we can obtain the general case by approximating T with a sequence that converges to T increasingly). We argue by contradiction, and assume that the mapping

$$(t, x) \mapsto u(t, x) - v(t, x) - M \left(|x| + Vt - R \right)_+$$

does not achieve its supremum on $\{0\} \times \overline{U}_K$. Let $\varepsilon_0 > 0$, let $\theta \in C^\infty(\mathbb{R})$ be an increasing smooth function such that

$$\forall r \in \mathbb{R}, \quad (r - \varepsilon_0)_+ \leqslant \theta(r) \leqslant r_+,$$

and consider

$$(3.21) \qquad \Phi(t, x) := M \theta \left(\left(\varepsilon_0 + \sum_{k=1}^K |x_k|^2 \right)^{\frac{1}{2}} + Vt - R \right).$$

We fix $\varepsilon_0 \in (0, 1]$ sufficiently close to 0 so that

$$(3.22) \qquad \sup_{[0,T) \times \overline{U}_K} (u - v - \Phi) > \sup_{\{0\} \times \overline{U}_K} (u - v - \Phi).$$

For every $k \in \{1, \ldots, K\}$, we have

$$(3.23) \qquad \partial_{x_k} \Phi(t, x) = \frac{M x_k}{\left(\varepsilon_0 + \sum_{\ell=1}^K |x_\ell|^2 \right)^{\frac{1}{2}}} \theta' \left(\left(\varepsilon_0 + \sum_{\ell=1}^K |x_\ell|^2 \right)^{\frac{1}{2}} + Vt - R \right)$$

In particular, we see that

$$(3.24) \qquad \partial_t \Phi \geqslant V |\nabla \Phi|.$$

We also record for future use that for every $t \geqslant 0$ and $x \in \overline{U}_K$,

$$(3.25) \qquad \Phi(t, x) \geqslant M \left(|x| + Vt - R - 1 \right)_+.$$

We set $C_0 := 2V|\vec{\imath}| + 1$, where the vector $\vec{\imath}$ is explicitly defined below in (3.32). For some constant $\varepsilon > 0$ to be determined, we define the function

$$\chi(t, x, x') := \Phi(t, x) + C_0 \varepsilon t + \frac{\varepsilon}{T - t} - \varepsilon \sum_{k=1}^K k \, \mathrm{Id} \cdot (x_k + x'_k),$$

and set $\widetilde{\chi}(t, x) \coloneqq \chi(t, x, x)$. In view of (3.22), we can choose $\varepsilon > 0$ sufficiently small that

$$(3.26) \qquad \sup_{[0,T) \times \overline{U}_K} (u - v - \widetilde{\chi}) > \sup_{\{0\} \times \overline{U}_K} (u - v - \widetilde{\chi}).$$

For later purposes, we also impose that $2\varepsilon|\vec{\iota}| \leqslant \delta$. We now introduce, for every $\alpha \geqslant 1$, $t \in [0, T)$, $t' \in [0, T]$, and $x, x' \in \overline{U}_K$, the function

$$\Psi_\alpha(t, x, t', x') \coloneqq u(t, x) - v(t', x') - \frac{\alpha}{2}\left(|t - t'|^2 + |x - x'|^2\right) - \chi(t, x, x').$$

By the definitions of M, χ, (3.25), and the fact that the functions u and v are uniformly Lipschitz, we see that the supremum of Ψ_α is achieved, at a point which we denote by $(t_\alpha, x_\alpha, t'_\alpha, x'_\alpha)$. We also see that this maximizing quadruple stays in a bounded region as α tends to infinity, and thus that the quantity

$$(3.27) \qquad \alpha\left(|t_\alpha - t'_\alpha|^2 + |x_\alpha - x'_\alpha|^2\right)$$

must remain bounded as α tends to infinity. We infer that, up to the extraction of a subsequence, there exist $t_0 \in [0, T)$ and $x_0 \in \overline{U}_K$ such that, as α tends to infinity, we have $t_\alpha \to t_0$, $t'_\alpha \to t_0$, $x_\alpha \to x_0$, $x'_\alpha \to x_0$. Since

$$\Psi_\alpha(t_\alpha, x_\alpha, t'_\alpha, x'_\alpha) \leqslant u(t_\alpha, x_\alpha) - v(t'_\alpha, x'_\alpha) - \chi(t, x_\alpha, x'_\alpha),$$

and

$$\Psi_\alpha(t_\alpha, x_\alpha, t'_\alpha, x'_\alpha) \geqslant \sup_{[0,T) \times \overline{U}_K}(u - v - \widetilde{\chi}) \geqslant (u - v - \widetilde{\chi})(t_0, x_0),$$

we deduce, by continuity of u, v, and $\widetilde{\chi}$, that

$$(u - v - \widetilde{\chi})(t_0, x_0) = \sup_{[0,T) \times \overline{U}_K}(u - v - \widetilde{\chi}).$$

In particular, by (3.26), we must have $t_0 > 0$, and thus $t_\alpha > 0$ and $t'_\alpha > 0$ for every α sufficiently large. By construction, the function

$$(3.28) \qquad (t, x) \mapsto u(t, x) - v(t'_\alpha, x'_\alpha) - \frac{\alpha}{2}\left(|t - t'_\alpha|^2 + |x - x'_\alpha|^2\right) - \chi(t, x, x'_\alpha)$$

reaches its maximum at (t_α, x_α). Since u is a subsolution, at least one of the following two statements hold:

$$(3.29) \qquad \alpha(t_\alpha - t'_\alpha) + C_0\varepsilon + \partial_t\Phi(t_\alpha, x_\alpha) - \mathsf{H}\left(\alpha(x_\alpha - x'_\alpha) + \nabla_x\chi(t_\alpha, x_\alpha, x'_\alpha)\right) \leqslant 0,$$

$$(3.30) \qquad x_\alpha \in \partial U_K \quad \text{and} \quad \inf_{\nu \in \mathbf{n}(x_\alpha)}\left(\alpha(x_\alpha - x'_\alpha) + \nabla_x\chi(t_\alpha, x_\alpha, x'_\alpha)\right) \cdot \nu \leqslant 0.$$

In (3.29), we dropped an additional term of $\frac{\varepsilon}{(T-t)^2}$ on the left side; this is possible since this term is nonnegative. Notice also that we somewhat reorganized the set of two conditions in (3.8)-(3.9), so that we also allow for the possibility that $x_\alpha \in \partial U_K$ in (3.29). We now argue that (3.30) cannot hold. By definition of $\mathbf{n}(x_\alpha)$, for every $\nu \in \mathbf{n}(x_\alpha)$, we have $\nu \cdot (x_\alpha - x'_\alpha) \geqslant 0$. We observe that

$$(3.31) \qquad \nabla_x\chi = \nabla\Phi - \varepsilon\vec{\iota},$$

where we have set

$$(3.32) \qquad \vec{\iota} \coloneqq \begin{pmatrix} \mathrm{Id} \\ 2\,\mathrm{Id} \\ \vdots \\ K\,\mathrm{Id} \end{pmatrix} \in (S^D)^K.$$

Moreover, $\nabla\Phi(t, x)$ is a vector proportional to x. We now see that, for every $x \in \partial U_K$ and $\nu \in \mathbf{n}(x)$, we have $x \cdot \nu = 0$. Indeed, since $\overline{U}_K$ is a cone, we have that $\lambda x \in \overline{U}_K$ for every $\lambda \geqslant 0$. In particular, we must have that $(\lambda x - x) \cdot \nu \leqslant 0$ for every $\lambda \geqslant 0$. This can

only happen if $x \cdot \nu = 0$. Finally, we show that there exists a constant $c > 0$ such that for every $x \in \partial U_K$ and $\nu \in \mathbf{n}(x)$, we have

(3.33) $$-\vec{\imath} \cdot \nu \geqslant c.$$

By (3.20), we have

$$\vec{\imath} \cdot \nu = \sum_{k=1}^{K} \mathrm{Id} \cdot \sum_{\ell=k}^{K} \nu_\ell.$$

Notice also that $-\nu \in \overline{U}_K^*$. By Lemma 3.3, each matrix $-\sum_{\ell=k}^{K} \nu_\ell$ therefore belongs to S_+^D. Moreover, for every $a \in S_+^D$, we have

$$\mathrm{Id} \cdot a \geqslant |a|,$$

(the left side is the ℓ^1 norm of the eigenvalues of a, the right side the ℓ^2 norm), and thus

$$-\vec{\imath} \cdot \nu \geqslant \sum_{k=1}^{K} \left| \sum_{\ell=k}^{K} \nu_\ell \right|.$$

The right side of the inequality above, as a function of ν, defines a norm on $(S^D)^K$. Using the equivalence of norms and that $|\nu| = 1$, we conclude that (3.33) holds. Combining the preceding observations, we conclude that (3.30) cannot be valid, so (3.29) holds instead.

Similarly, since the function

$$(t', x') \mapsto v(t', x') - u(t_\alpha, x_\alpha) + \frac{\alpha}{2} \left(|t' - t_\alpha|^2 + |x' - x_\alpha|^2 \right) + \chi(t_\alpha, x_\alpha, x')$$

has a local minimum at (t'_α, x'_α), and since v is a supersolution, at least one of the following two statements must be valid:

(3.34) $$\alpha(t_\alpha - t'_\alpha) - \mathsf{H}\left(\alpha(x_\alpha - x'_\alpha) + \varepsilon\vec{\imath} \right) \geqslant 0,$$

(3.35) $$x'_\alpha \in \partial U_K \quad \text{and} \quad \sup_{\nu \in \mathbf{n}(x'_\alpha)} \left(\alpha(x_\alpha - x'_\alpha) + \varepsilon\vec{\imath} \right) \cdot \nu \geqslant 0.$$

In view of (3.33), we see that (3.35) cannot hold, and therefore (3.34) is valid.

We now show that (3.29) and (3.34) cannot hold simultaneously, thereby reaching the desired contradiction. We temporarily admit that the vectors

(3.36) $$\alpha(x_\alpha - x'_\alpha) + \nabla_x \chi(t_\alpha, x_\alpha, x'_\alpha) \quad \text{and} \quad \alpha(x_\alpha - x'_\alpha) + \varepsilon\vec{\imath}$$

are both of norm smaller than $L + \delta$. Admitting this, we use (3.31), the Lipschitz property of H, the fact that $C_0 = 2V|\vec{\imath}| + 1$, and (3.24), to deduce that (3.29) implies

$$\alpha(t_\alpha - t'_\alpha) + \varepsilon - \mathsf{H}\left(\alpha(x_\alpha - x'_\alpha) + \varepsilon\vec{\imath} \right) \leqslant 0,$$

in contradiction with (3.34).

There remains to verify that the vectors in (3.36) are bounded by $L + \delta$. For convenience, we rewrite the mapping in (3.28) as

(3.37) $$(t, x) \mapsto u(t, x) - \psi(t, x).$$

We recall that this mapping achieves its maximum at (t_α, x_α), and we aim to show that $|\nabla\psi(t_\alpha, x_\alpha)| \leqslant L + \delta$. Since u is L-Lipschitz, we have, for every $y \in \overline{U}_K$,

$$\psi(t_\alpha, y) - \psi(t_\alpha, x_\alpha) \geqslant u(t_\alpha, y) - u(t_\alpha, x_\alpha) \geqslant -L|y - x_\alpha|.$$

If $x_\alpha \in U_K$, the desired conclusion follows. Otherwise, we can only infer that, for every v in the set

$$\mathfrak{C} := \left\{ \lambda(y - x_\alpha) \ : \ \lambda \in [0, \infty), \ y \in \overline{U}_K \right\},$$

we have

$$v \cdot \nabla\psi(t_\alpha, x_\alpha) \geqslant -L|v|.$$

We now recall that

$$\nabla\psi(t_\alpha, x_\alpha) = \alpha(x_\alpha - x'_\alpha) + \nabla\Phi(t_\alpha, x_\alpha) - \varepsilon\vec{\imath}.$$

Moreover, we have that $\nabla\Phi(t_\alpha, x_\alpha)$ is proportional to x_α, say $\nabla\Phi(t_\alpha, x_\alpha) = \beta x_\alpha$, for some $\beta \geqslant 0$. Since

$$\alpha(x'_\alpha - x_\alpha) - \beta x_\alpha = (\alpha + \beta)\left(\frac{\alpha}{\alpha + \beta}x'_\alpha - x_\alpha\right) \in \mathfrak{C},$$

we deduce that

$$\left(\nabla\psi(t_\alpha, x_\alpha) + \varepsilon\vec{\imath}\right) \cdot \nabla\psi(t_\alpha, x_\alpha) \leqslant L|\nabla\psi(t_\alpha, x_\alpha) + \varepsilon\vec{\imath}|.$$

This yields

$$|\nabla\psi(t_\alpha, x_\alpha) + \varepsilon\vec{\imath}| \leqslant L + \varepsilon|\vec{\imath}|,$$

and thus

$$|\nabla\psi(t_\alpha, x_\alpha)| \leqslant L + 2\varepsilon|\vec{\imath}|.$$

This is the desired result, since we have chosen $\varepsilon > 0$ sufficiently small that $2\varepsilon|\vec{\imath}| \leqslant \delta$. The argument for the second vector in (3.36) is similar. $\qquad\square$

We next provide with the following result on existence of solutions.

Proposition 3.4 (Existence of solutions)**.** *For every uniformly Lipschitz initial condition $f_0 : U_K \to \mathbb{R}$, there exists a viscosity solution f to (3.7) that satisfies $f(0, \cdot) = f_0$. Moreover, the function f is Lipschitz continuous, and we have*

$$(3.38) \qquad \||\nabla f|\|_{L^\infty(\mathbb{R}_+ \times U_K)} = \||\nabla f_0|\|_{L^\infty(U_K)}.$$

Proof of Proposition 3.4. We will prove below that the proposition is valid if we assume furthermore that H is uniformly Lipschitz and that the initial condition is bounded. We first explain why this is sufficient. Denote the right side of (3.38) by L. The proof of Proposition 3.2 makes it clear that, if u and v are solutions to (3.7) with the same L-Lipschitz initial condition and with the nonlinearity H replaced by H_1 and H_2 respectively, and if H_1 and H_2 coincide on a ball of radius $L + 1$, then $u = v$. It follows that, in order to build a solution to (3.7), we may as well replace H by a globally Lipschitz nonlinearity that coincides with H on the ball of radius $L + 1$. Finally, once this is done, we can use the property of finite speed of propagation proved in Proposition 3.2 to remove the constraint that the initial condition is bounded.

The argument for the existence of a solution is as in [19] or [8, Theorem 7.1] (in the latter, the initial condition is not assumed continuous, but this additional assumption allows to conclude that the solution is continuous as well). For bounded initial conditions, this construction provides with bounded solutions.

We now turn to the proof of the fact that the solution thus constructed, which we denote by f, is Lipschitz, and that the identity (3.38) holds. Again we fix $\mathcal{C} = S_{++}^D$ for concreteness, the case $\mathcal{C} = (0, \infty)^D$ being only easier. We argue by contradiction, assuming instead that

$$\||\nabla f|\|_{L^\infty(\mathbb{R}_+ \times U_K)} > \||\nabla f_0|\|_{L^\infty(U_K)} = L.$$

We recall that we assume here that H is uniformly Lipschitz; we denote its Lipschitz constant by V. For a constant $R \in \mathbb{R}$ to be chosen, and $M = 1$, we define the function Φ as in (3.21). We then set, for constants $\varepsilon > 0$, and $T < \infty$ to be chosen, and every $t \in [0, T)$,

$$u_\varepsilon(t, x) := f(t, x) - \Phi(t, x) - \frac{\varepsilon}{T - t},$$

as well as, for every $t \geqslant 0$ and $x \in \overline{U}_K$,

$$v(t, x) := f(t, x) + \Phi(t, x).$$

Proceeding as in the proof of Proposition 3.2, we verify that u_ε and v are a sub- and a supersolution to (3.7) respectively. We have, for every $t \in [0, T)$ and $x \in \overline{U}_K$,

$$(3.39) \qquad u_\varepsilon(t, x) - v(t, x) = -2\Phi(t, x) - \frac{\varepsilon}{T - t}.$$

We then choose $\varepsilon > 0$ sufficiently small and T and $R \in \mathbb{R}$ sufficiently large that

$$(3.40) \quad \sup_{\substack{0 \leqslant t < T \\ x, x' \in \overline{U}_K}} \left(u_\varepsilon(t, x) - v(t, x') - L|x - x'| \right) > \sup_{x, x' \in \overline{U}_K} \left(u_\varepsilon(0, x) - v(0, x') - L|x - x'| \right).$$

We denote by $\eta > 0$ the difference between the left side and the right side of this inequality. We also remark that, by (3.39),

$$(3.41) \qquad \sup_{0 \leqslant t < T, x \in \overline{U}_K} \left(u_\varepsilon(t, x) - v(t, x) \right) = \sup_{x, x' \in \overline{U}_K} \left(u_\varepsilon(0, x) - v(0, x) \right).$$

We now let $\delta' \leqslant \delta \in (0, 1]$ to be chosen (we will first fix δ in terms of η, and then δ' in terms of δ, η and moduli of continuity of u_ε, v, and H), and for every $\alpha \in [1, \infty)$, $t < T$, $t' \geqslant 0$, and $x, x' \in \overline{U}_K$, we consider

$$\Psi_\alpha(t, x, t', x') := u_\varepsilon(t, x) - v(t', x') - (L + \delta t)|x - x'| - \frac{\alpha}{2}|t_\alpha - t'_\alpha|^2 + \delta' \sum_{k=1}^{K} k \operatorname{Id} \cdot (x_k + x'_k).$$

Since we assume f to be bounded, the maximum of Ψ_α is achieved at a point, which we denote by $(t_\alpha, x_\alpha, t'_\alpha, x'_\alpha)$, and this point remains in a bounded region as α tends to infinity (this bounded region can be chosen irrespectively of our choice of δ and δ' sufficiently small). Extracting a subsequence if necessary, we can further assume that $t_\alpha \to t_0$, $t'_\alpha \to t_0$, $x_\alpha \to x_0$, and $x'_\alpha \to x'_0$ (the limits of t_α and t'_α must be the same, since $|t_\alpha - t'_\alpha|^2 = O(\alpha^{-1})$). We also have that, for some constant $C < \infty$,

$$\Psi_\alpha(t_0, x_0, t_0, x'_0) \geqslant -C\delta + \sup_{\substack{0 \leqslant t < T \\ x, x' \in \overline{U}_K}} \left(u_\varepsilon(t, x) - v(t, x') - L|x - x'| \right),$$

while

$$\Psi_\alpha(0, x_0, 0, x'_0) \leqslant C\delta + \sup_{x, x' \in \overline{U}_K} \left(u_\varepsilon(0, x) - v(0, x') - L|x - x'| \right),$$

and, using (3.41),

$$\Psi_\alpha(t_0, x_0, t_0, x_0) \leqslant C\delta + \sup_{x, x' \in \overline{U}_K} \left(u_\varepsilon(0, x) - v(0, x') - L|x - x'| \right).$$

Choosing $\delta > 0$ such that $4C\delta \leqslant \eta$, we can thus guarantee that $t_0 \neq 0$ and $x_0 \neq x'_0$. More precisely, with this choice of $\delta > 0$, and using the continuity of u_ε and v, we can ensure that there exists $\gamma > 0$, not depending on δ', such that $|x_0 - x'_0| \geqslant \gamma$. As a consequence, we have $t_\alpha > 0$, and $|x_\alpha - x'_\alpha| \geqslant \gamma/2$ for every α sufficiently large. We use again the notation $\vec{\imath}$ from (3.32). Since u_ε is a subsolution, at least one of the following statements holds:

$$(3.42) \qquad \alpha(t_\alpha - t'_\alpha) + \delta|x_\alpha - x'_\alpha| - \mathsf{H}\left((L + \delta t_\alpha)\frac{x_\alpha - x'_\alpha}{|x_\alpha - x'_\alpha|} - \delta'\vec{\imath} \right) \leqslant 0,$$

$$(3.43) \qquad x_\alpha \in \partial U_K \quad \text{and} \quad \inf_{\nu \in \mathbf{n}(x_\alpha)} \nu \cdot \left((L + \delta t_\alpha)\frac{x_\alpha - x'_\alpha}{|x_\alpha - x'_\alpha|} - \delta'\vec{\imath} \right) \leqslant 0.$$

By (3.33) and the definition of $\mathbf{n}(x_\alpha)$, the statement in (3.43) cannot hold, and therefore (3.42) is valid. Conversely, since v is a supersolution, at least one of the following statements holds:

$$(3.44) \qquad \alpha(t_\alpha - t'_\alpha) - \delta|x_\alpha - x'_\alpha| - \mathsf{H}\left((L + \delta t_\alpha)\frac{x_\alpha - x'_\alpha}{|x_\alpha - x'_\alpha|} + \delta'\vec{\imath} \right) \geqslant 0,$$

$$(3.45) \qquad x'_\alpha \in \partial U_K \quad \text{and} \quad \sup_{\nu \in \mathbf{n}(x'_\alpha)} \nu \cdot \left((L + \delta t_\alpha) \frac{x_\alpha - x'_\alpha}{|x_\alpha - x'_\alpha|} + \delta' \vec{\imath} \right) \geqslant 0.$$

As above, we see that (3.45) cannot hold. We thus conclude that (3.42) and (3.44) are both valid. But, since $|x_\alpha - x'_\alpha|$ is bounded away from zero by a quantity not depending on δ', and since H is Lipschitz, we reach a contradiction by selecting δ' sufficiently small. $\quad\square$

We now point out a convenient way to verify that certain functions satisfy the boundary condition for being a subsolution to (3.7). The condition is a sort of monotonicity property, which we call being "tilted", and is inspired by Lemma 2.4. Let V be a subset of $\mathcal{E}^K$, and $f : V \to \mathbb{R}$. We say that the function f is *tilted* if, for every $x, y \in V$, we have

$$y - x \in \overline{U}_K^* \quad \Longrightarrow \quad f(x) \leqslant f(y),$$

where we recall that $\overline{U}_K^*$ was defined in Lemma 3.3. We may also consider functions f defined on $\mathbb{R}_+ \times V$ (or with $\mathbb{R}_+$ replaced by a subinterval); in this case, we say that the function f is tilted if the function $f(t, \cdot)$ is tilted for every fixed $t \geqslant 0$. The next lemma provides with a simple characterization of being tilted for Lipschitz functions.

Lemma 3.5 (Characterization of tilted functions)**.** *Let V be an open subset of $\mathcal{E}^K$, and let $f : V \to \mathbb{R}$ be a Lipschitz function. The function f is tilted if and only if $\nabla f \in \overline{U}_K$ almost everywhere in V.*

Proof. We decompose the proof into two steps.

Step 1. We assume that f is tilted, and show that $\nabla f \in \overline{U}_K$ almost everywhere. By Rademacher's theorem, the function f is differentiable almost everywhere. Let $z \in V$ be a point of differentiability of f, and $x \in \overline{U}_K^*$. Since V is open and f is tilted, we have, for every $\varepsilon > 0$ sufficiently small,

$$f(z + \varepsilon x) - f(z) \geqslant 0.$$

Dividing by ε and letting $\varepsilon > 0$ tend to zero, we conclude that $x \cdot \nabla f(z) \geqslant 0$. By Lemma 3.3, this means that $\nabla f(z) \in \overline{U}_K$.

Step 2. We assume that $\nabla f \in \overline{U}_K$ almost everywhere, and show that f is tilted. By Fubini's theorem, the set

$$\left\{ (x, y) \in V^2 \ : \ |\{ s \in [0,1] \ : \ f \text{ is differentiable at } sy + (1-s)x \}| = 1 \right\}$$

has full measure (for $I \subseteq [0,1]$, the notation $|I|$ above denotes its Lebesgue measure). We fix a pair (x, y) in this set. Since the mapping $s \mapsto f(sy + (1-s)x)$ is Lipschitz, we have

$$f(y) - f(x) = \int_0^1 (y - x) \cdot \nabla f(sy + (1-s)x) \, \mathrm{d}s.$$

The result then follows using Lemma 3.3 once more. $\quad\square$

Notice that, by Lemmas 2.4 and 3.5, the function

$$(t, q) \mapsto \widetilde{F}_N \left(t, \left(\frac{1}{k} \sum_{\ell=1}^k \delta_{q_{a,\ell}} \right)_{a \in \{1,2\}} \right)$$

is tilted. As announced, the next proposition states that a tilted function automatically satisfies the boundary condition (3.9).

Proposition 3.6 (Boundary condition for subsolution)**.** *Let $f \in C([0,T] \times \overline{U}_K)$ be a tilted function, $(t, x) \in (0,T) \times \partial U_K$, and $\phi \in C^\infty((0,T) \times \overline{U}_K)$ be such that (t, x) is a local maximum of $f - \phi$. We have*

$$\inf_{\nu \in \mathbf{n}(x)} \nu \cdot \nabla \phi(t, x) \leqslant 0.$$

Proof. We decompose the proof into three steps.

Step 1. In this step, we prove a general (and classical) statement concerning the bidual of a closed convex cone. Let $\mathfrak{C}$ be a closed convex cone, which for simplicity we assume to be in some Euclidean space E. Let $\mathfrak{C}'$ be, up to a sign, the cone dual to $\mathfrak{C}$:

$$(3.46) \qquad \mathfrak{C}' := \{ y \in E \ : \ \forall x \in \mathfrak{C}, \ x \cdot y \leqslant 0 \},$$

and let

$$(3.47) \qquad \mathfrak{C}'' := \{ x \in E \ : \ \forall y \in \mathfrak{C}', \ x \cdot y \leqslant 0 \}.$$

In this step, we show that $\mathfrak{C}'' = \mathfrak{C}$. Let $f : E \to E$ be such that $f = 0$ on $\mathfrak{C}$ and $f = +\infty$ otherwise. Its convex dual f^* is such that, for every $y \in E$,

$$f^*(y) = \sup_{x \in E} (x \cdot y - f(x)) = \sup_{x \in \mathfrak{C}} x \cdot y = \left|\begin{array}{ll} 0 & \text{if } y \in \mathfrak{C}', \\ +\infty & \text{otherwise.} \end{array}\right.$$

In the same way, using that $\mathfrak{C}'$ is a cone, we see that the bidual f^{**} is such that $f^{**} = 0$ on $\mathfrak{C}''$, and $f^{**} = +\infty$ otherwise. Since f is convex and lower semicontinuous, it is equal to its bidual. This shows that $\mathfrak{C}'' = \mathfrak{C}$.

Step 2. We now prove another general (and possibly less classical) statement about closed convex cones. Let E be some Euclidean space, and for any $A \subseteq E$, let $\mathsf{int}(A)$ and $\mathsf{conv}(A)$ denote the interior and the convex hull of A respectively. Let $\mathfrak{C} \subseteq E$ be a closed convex cone, and $\mathfrak{C}'$ be as in (3.46). Our aim is to show that if the interior of $\mathfrak{C}$ is not empty, then

$$(3.48) \qquad \mathsf{int}(\mathfrak{C}) \cap (-\mathfrak{C}') \neq \varnothing,$$

where we write $-\mathfrak{C}' := \{ -z \ : \ z \in \mathfrak{C}' \}$. Without loss of generality, we may assume that $\mathfrak{C}' \neq \{0\}$ (otherwise we have $\mathfrak{C} = E$, by the result of the previous step, and 0 belongs to the set on the left side of (3.48)). We first show that

$$(3.49) \qquad \mathfrak{C}' \cap (-\mathfrak{C}') = \{0\}.$$

Indeed, if $z \in \mathfrak{C}' \cap (-\mathfrak{C}')$, then by the result of the previous step, we must have that $y \cdot z = 0$ for every $y \in \mathfrak{C}$. Since we assume that the interior of $\mathfrak{C}$ is not empty, this is only possible if $z = 0$. Let

$$\mathbf{m} := \{ z \in \mathfrak{C}' \ : \ |z| = 1 \}.$$

We now show that

$$(3.50) \qquad 0 \notin \mathsf{conv}(\mathbf{m}).$$

Assume instead that $0 \in \mathsf{conv}(\mathbf{m})$. By Carathéodory's theorem, the point 0 can then be represented as the barycenter of a finite number of points in $\mathbf{m}$. Since $\mathbf{m} \subseteq \mathfrak{C}'$ and $\mathfrak{C}'$ is convex and contains the origin, this allows us to contradict (3.49). Using Carathéodory's theorem once more, we can also verify that $\mathsf{conv}(\mathbf{m})$ is compact.

For every $\varepsilon > 0$, we define

$$\mathbf{m}_\varepsilon := \{ z \in E \ : \ \mathsf{dist}(z, \mathsf{conv}(\mathbf{m})) \leqslant \varepsilon \},$$

and aim to show that there exists $\varepsilon > 0$ such that

$$(3.51) \qquad \mathfrak{C} \cap \{ \lambda \nu \ : \ \lambda > 0, \ \nu \in \mathbf{m}_\varepsilon \} = \varnothing.$$

Assume the contrary: for every $\varepsilon > 0$, we could then find $\lambda_\varepsilon > 0$ and $\nu_\varepsilon \in \mathbf{m}_\varepsilon$ such that $\lambda_\varepsilon \nu_\varepsilon \in \mathfrak{C}$. Since $\mathfrak{C}$ is a cone, the latter condition means that $\nu_\varepsilon \in \mathfrak{C}$. Since the sets $(\mathbf{m}_\varepsilon)_{\varepsilon > 0}$ are compact and nested, and since $\mathfrak{C}$ is closed, we can find a limit point $\nu \in \mathbf{m}_0 = \mathsf{conv}(\mathbf{m})$ such that $\nu \in \mathfrak{C}$. We have in particular that $\nu \in \mathfrak{C}'$, but by (3.46), we have $\mathfrak{C} \cap \mathfrak{C}' = \{0\}$. This implies that $\nu = 0$. But since $\nu \in \mathsf{conv}(\mathbf{m})$, this contradicts (3.50).

Notice next that the set $\{\lambda\nu \ : \ \lambda > 0, \ \nu \in \mathbf{m}_\varepsilon\}$ is convex: indeed, for every $\lambda, \lambda' > 0$, $\nu, \nu' \in \mathbf{m}_\varepsilon$, and $\alpha \in (0, 1)$, we have

$$\alpha\lambda\nu + (1 - \alpha)\lambda'\nu' = (\alpha\lambda + (1 - \alpha)\lambda')\left(\frac{\alpha\lambda}{\alpha\lambda + (1 - \alpha)\lambda'}\nu + \frac{(1 - \alpha)\lambda'}{\alpha\lambda + (1 - \alpha)\lambda'}\nu'\right),$$

and the quantity between parentheses on the right side belongs to $\mathbf{m}_\varepsilon$, since this set is convex. Since $\mathfrak{C}$ is also convex, we can find a hyperplane that separates the two disjoint sets appearing in (3.51): there exists $v \in E$, which we may assume to be of unit norm, such that

$$\forall z \in \mathfrak{C}, \ z \cdot v \geqslant 0 \quad \text{and} \quad \forall z \in \mathbf{m}_\varepsilon, \ v \cdot z \leqslant 0.$$

The first property in the previous display yields that $-v \in \mathfrak{C}'$. We will now see that $v \in \mathsf{int}(\mathfrak{C})$, which will complete the proof of (3.48). For every $v' \in E$ satisfies $|v' - v| \leqslant \varepsilon$ and $z \in \mathbf{m}$, we have

$$v' \cdot z \leqslant v \cdot z + \varepsilon = v \cdot (z + \varepsilon v) \leqslant 0,$$

where we used that $z + \varepsilon v \in \mathbf{m}_\varepsilon$ in the last inequality. Using the notation in (3.47), this shows that every such v' belongs to $\mathfrak{C}''$. By the result of the previous step, we have $\mathfrak{C}'' = \mathfrak{C}$, and we have thus verified that $v \in \mathsf{int}(\mathfrak{C})$.

Step 3. We fix (t, x), ϕ as in the statement of Proposition 3.6, let

(3.52) $$\mathfrak{C}_0 := \{\lambda(y - x) \ : \ \lambda > 0, \ y \in U_K\},$$

and let $\mathfrak{C}$ denote its closure. Since $\mathfrak{C}_0$ is open, we have $\mathsf{int}(\mathfrak{C}) = \mathfrak{C}_0$. For every $\lambda, \lambda' \in (0, \infty)$ and $y, y' \in U_K$, we have

$$\lambda(y - x) + \lambda'(y' - x) = (\lambda + \lambda')\left(\frac{\lambda}{\lambda + \lambda'}y + \frac{\lambda'}{\lambda + \lambda'}y' - x\right) \in \mathfrak{C}_0.$$

It follows that $\mathfrak{C}_0$ is convex, and thus that $\mathfrak{C}$ is a closed convex cone. Let $\mathfrak{C}'$ be defined by (3.46), with $E = \mathcal{E}^K$. Since $\mathfrak{C}$ has nonempty interior, we can apply the result of the previous step to infer that

$$\mathfrak{C}_0 \cap (-\mathfrak{C}') \neq \varnothing.$$

Let v denote an element of this set; without loss of generality, we may assume that v is of unit norm. By definition of $\mathfrak{C}_0$, there exists $\lambda > 0$ and $y \in U_K$ such that $v = \lambda(y - x)$. Since $-v \in \mathfrak{C}'$ and is of unit norm, we also have that $-v \in \mathbf{n}(x)$. Since the set $\mathfrak{C}$ contains $\overline{U}_K$, we have that $-\mathfrak{C}' \subseteq \overline{U}_K^*$, and thus $v \in \overline{U}_K^*$. By convexity of $\overline{U}_K$, for every $\varepsilon \in [0, \lambda^{-1}]$, we have that

$$x + \varepsilon v = (1 - \varepsilon\lambda)x + \varepsilon\lambda y \in \overline{U}_K.$$

By the assumption that (t, x) is a local maximum of $f - \phi$, for every $\varepsilon > 0$ sufficiently small, we have

$$(f - \phi)(t, x + \varepsilon v) \leqslant (f - \phi)(t, x).$$

Since f is tilted and $v \in \overline{U}_K^*$, we deduce that $\phi(t, x + \varepsilon v) - \phi(t, x) \geqslant 0$. Dividing by $\varepsilon > 0$ and letting it tend to zero, we obtain that

$$v \cdot \nabla\phi(t, x) \geqslant 0.$$

Since $-v \in \mathbf{n}(x)$, this is the desired result. $\qquad\square$

3.2. **Convergence of finite-dimensional approximations.** We now turn to the identification of the solution to (1.7), which we define to be the limit of the solutions to suitable finite-dimensional approximations. From now on, we specialize the results of the previous subsection to the case of

$$\mathcal{C} := (0, \infty)^2.$$

We aim to approximate each measure in a given pair $(\mu_1, \mu_2) \in (\mathcal{P}(\mathbb{R}_+))^2$ by a measure of the form

$$(3.53) \qquad\qquad \frac{1}{K} \sum_{k=1}^{K} \delta_{x_{k,a}}, \qquad (a \in \{1, 2\}),$$

for some (ultimately large) integer K and some $x \in \overline{U}_K$, where we set $x_k = (x_{k,1}, x_{k,2})$, and $x = (x_1, \ldots, x_K)$. We can clearly map any element of $\overline{U}_K$ to a pair of probability measures in $(\mathcal{P}(\mathbb{R}_+))^2$ through the mapping defined in (3.53). We can also define a converse operation, from a given pair of measures in $(\mathcal{P}(\mathbb{R}_+))^2$ to an element of $\overline{U}_K$. Fixing $K \geqslant 1$, we set, for every $\mu = (\mu_1, \mu_2) \in (\mathcal{P}(\mathbb{R}_+))^2$ and $k \in \{1, \ldots, K\}$,

$$(3.54) \qquad x_k^{(K)}(\mu) = (x_{k,1}^{(K)}(\mu), x_{k,2}^{(K)}(\mu)) := K \int_{\frac{k-1}{K}}^{\frac{k}{K}} (F_{\mu_1}^{-1}(u), F_{\mu_2}^{-1}(u)) \, \mathrm{d}u \in \mathbb{R}_+^2,$$

where we recall that the functions $F_{\mu_a}^{-1}$ were introduced in (1.4). This defines a mapping $\mu \mapsto x^{(K)}(\mu)$ from $(\mathcal{P}(\mathbb{R}_+))^2$ to $\overline{U}_K$. Notice that if we map an element x of $\overline{U}_K$ to a pair of measures according to (3.53), and then back into an element of $\overline{U}_K$ through the mapping above, we recover x (but obviously, some information is lost when we go from a pair of measures to an element of $\overline{U}_K$ and then back). We also use the notation

$$\mu^{(K)} = (\mu_1^{(K)}, \mu_2^{(K)}) := \left(\frac{1}{K} \sum_{k=1}^{K} \delta_{x_{k,a}^{(K)}(\mu)} \right)_{a \in \{1, 2\}}.$$

The mapping $\mu \mapsto \mu^{(K)}$ thus takes an element μ of $(\mathcal{P}(\mathbb{R}_+))^2$, and returns a pair in $(\mathcal{P}(\mathbb{R}_+))^2$, made of two measures with K atoms of equal masses (the latter is in some sense the "representative" of $x^{(K)}(\mu) \in \overline{U}_K$ within the set $(\mathcal{P}(\mathbb{R}_+))^2$).

The following proposition is the main result of this subsection.

Proposition 3.7 (Convergence of finite-dimensional approximations). *Let ψ be the function defined in (2.19), and for each integer $K \geqslant 1$, let $f^{(K)} : \mathbb{R}_+ \times \overline{U}_K \to \mathbb{R}$ be the viscosity solution to*

$$(3.55) \qquad \begin{cases} \partial_t f^{(K)} - K \sum_{k=1}^{K} \partial_{x_{k,1}} f^{(K)} \, \partial_{x_{k,2}} f^{(K)} = 0 & \text{on } (0, \infty) \times \overline{U}_K, \\ \mathbf{n} \cdot \nabla f^{(K)} = 0 & \text{on } (0, \infty) \times \partial U_K, \end{cases}$$

with initial condition given, for every $x \in \overline{U}_K$, by

$$(3.56) \qquad\qquad f^{(K)}(0, x) = \psi \left(\frac{1}{K} \sum_{k=1}^{K} \delta_{x_{k,1}}, \frac{1}{K} \sum_{k=1}^{K} \delta_{x_{k,2}} \right).$$

For every $t \geqslant 0$ and $\mu \in (\mathcal{P}_2(\mathbb{R}_+))^2$, the following limit exists and is finite:

$$(3.57) \qquad\qquad f(t, \mu) := \lim_{K \to \infty} f^{(K)} \left(t, x^{(K)}(\mu) \right),$$

where on the right side, we use the notation defined in (3.54). By definition, we interpret this limit as the solution to (1.7). Moreover, there exists a constant $C < \infty$ such that, for

every integer $K \geqslant 1$, $t \geqslant 0$, and $\mu, \nu \in (\mathcal{P}_2(\mathbb{R}_+))^2$, we have

$$(3.58) \qquad \left| f(t, \mu) - f^{(K)}(t, x^{(K)}(\mu)) \right| \leqslant \frac{C}{\sqrt{K}} \left(t + \left(\mathbb{E}\left[X_{\mu_1}^2 + X_{\mu_2}^2 \right] \right)^{\frac{1}{2}} \right),$$

as well as

$$(3.59) \qquad |f(t, \mu) - f(t, \nu)| \leqslant \left(\mathbb{E}\left[|X_{\mu_1} - X_{\nu_1}|^2 + |X_{\mu_2} - X_{\nu_2}|^2 \right] \right)^{\frac{1}{2}}.$$

Before turning to the proof of Proposition 3.7, we introduce some notation for norms that are rescaled to be consistent with Wasserstein-type distances on the space of probability measures, according to the correspondences discussed at the beginning of this subsection. For every $\rho \in [1, \infty]$ and $x \in \mathcal{E}^K = (\mathbb{R}^2)^K$, we write

$$(3.60) \qquad |x|_\rho := \left(\frac{1}{K} \sum_{k=1}^{K} |x_k|^\rho \right)^{\frac{1}{\rho}},$$

with the usual interpretation as a supremum if $\rho = \infty$. We also define the norm dual to $|\cdot|_\rho$ by setting, for $\tau \in [1, \infty]$ such that $\frac{1}{\rho} + \frac{1}{\tau} = 1$ and every $x \in \mathcal{E}^K = (\mathbb{R}^2)^K$,

$$(3.61) \qquad |x|_{\tau*} := K^{\frac{1}{\rho}} \left(\sum_{k=1}^{K} |x_k|^\tau \right)^{\frac{1}{\tau}} = \left(\frac{1}{K} \sum_{k=1}^{K} (K|x_k|)^\tau \right)^{\frac{1}{\tau}}.$$

We also observe that, by a simple rescaling, the statement of Proposition 3.2 also holds if we replace the displays (3.13)-(3.16), by, respectively,

$$L := \max\left(\big\| \, |\nabla u|_{2*} \, \big\|_{L^\infty([0,T] \times U_K)}, \big\| \, |\nabla v|_{2*} \, \big\|_{L^\infty([0,T] \times U_K)} \right),$$

$$(3.62) \qquad V := \sup\left\{ \frac{|\mathsf{H}(p') - \mathsf{H}(p)|}{|p' - p|_{2*}} \ : \ |p|_{2*}, |p'|_{2*} \leqslant L + \delta \right\},$$

$$M > \sup_{0 \leqslant t < T, x \in \overline{U}_K} \frac{u(t,x) - v(t,x)}{1 + |x|_2},$$

and

$$(t, x) \mapsto u(t, x) - v(t, x) - M\left(|x|_2 + Vt - R \right)_+ .$$

Proof of Proposition 3.7. We decompose the proof into three steps.

Step 1. We start with a simple but crucial observation regarding the relationship between the $f^{(K)}$'s for different values of K. For all integers $K, R \geqslant 1$, we set $K' := RK$, and for every $x \in \overline{U}_{K'}$, we define

$$x^{(K,K')} := \left(\frac{1}{R} \sum_{r=1}^{R} x_r, \frac{1}{R} \sum_{r=1}^{R} x_{R+r}, \ldots, \frac{1}{R} \sum_{r=1}^{R} x_{(K-1)R+r} \right) \in \overline{U}_K,$$

as well as

$$f^{(K,K')}(t, x) := f^{(K)}(t, x^{(K,K')}).$$

In some sense, the function $f^{(K,K')}$ is a "lifting" of the function $f^{(K)}$ to the space $\overline{U}_{K'}$ (and this "lifting" is consistent with the identification between measures and elements of $\overline{U}_K$ discussed at the beginning of this section). Formally, we have for every $k \in \{1, \ldots, K\}$ and $r \in \{1, \ldots, R\}$ that

$$\partial_{x_{(k-1)R+r}} f^{(K,K')}(t, x) = \frac{1}{R} \partial_{x_k} f^{(K)}(t, x^{(K,K')}),$$

and thus, on a formal level, the function $f^{(K,K')}$ solves the same equation as $f^{(K')}$ does, but with a different initial condition. It is not difficult to justify rigorously that $f^{(K,K')}$ indeed solves this equation in the viscosity sense. In a few words, for instance to verify

that $f^{(K,K')}$ is a subsolution: suppose that (t,x) is a local maximum of $f^{(K,K')} - \phi$ for some smooth function ϕ. Then we can build $\widetilde{\phi} \in C^\infty((0,\infty) \times \overline{U}_K)$ by setting, for every $x' \in \overline{U}_K$,

$$\widetilde{\phi}(t,x') := \phi\left(t, x + \left(x'_1 - x_1^{(K,K')}, \ldots, x'_1 - x_1^{(K,K')}, \ldots, x'_K - x_K^{(K,K')}, \ldots, x'_K - x_K^{(K,K')}\right)\right),$$

where each coordinate in the inner parenthesis above is repeated R times. This ensures that $f^{(K)} - \widetilde{\phi}$ has a local maximum at $(t, x^{(K,K')})$. We then use that $f^{(K)}$ is a subsolution, and the simple relationship between the deriatives of $\widetilde{\phi}$ and those of ϕ, to conclude.

Step 2. We next leverage on this observation to evaluate the difference between $f^{(K,K')}$ and $f^{(K')}$, using Proposition 3.2. Precisely, we will show that there exists a constant $C < \infty$ such that for every $t \geqslant 0$ and $x \in \overline{U}_{K'}$,

$$(3.63) \qquad \left| f^{(K,K')}(t,x) - f^{(K')}(t,x) \right| \leqslant \frac{C}{\sqrt{K}} \left(|x|_2 + t \right).$$

By the definition of ψ in (2.19) and Proposition 2.1, we have that, for every $x, y \in \overline{U}_K$,

$$(3.64) \qquad \left| f^{(K)}(0,y) - f^{(K)}(0,x) \right| \leqslant |x - y|_1.$$

By Jensen's inequality, we also have

$$(3.65) \qquad \left| f^{(K,K')}(0,y) - f^{(K,K')}(0,x) \right| \leqslant |x^{(K,K')} - y^{(K,K')}|_1 \leqslant |x - y|_1.$$

Since $|\cdot|_1 \leqslant |\cdot|_2$, we can appeal to Proposition 3.4 to infer that

$$(3.66) \qquad \| |\nabla f^{(K)}|_{2*} \|_{L^\infty(\mathbb{R}_+ \times U_K)} \leqslant 1,$$

and

$$(3.67) \qquad \| |\nabla f^{(K,K')}|_{2*} \|_{L^\infty(\mathbb{R}_+ \times U_K)} \leqslant 1.$$

In view of the observation preceding this proof, it is thus legitimate to apply Proposition 3.2 with u and v replaced by $f^{(K,K')}$ and $f^{(K')}$, and with the choice of $L = 1$. We also observe that, for every $p, p' \in \mathcal{E}^K = (\mathbb{R}^2)^K$,

$$\left| K \sum_{k=1}^K p_{k,1} p_{k,2} - K \sum_{k=1}^K p'_{k,1} p'_{k,2} \right| = \left| K \sum_{k=1}^K \left(p_{k,1} \left(p_{k,2} - p'_{k,2} \right) + (p_{k,1} - p'_{k,1}) p'_{k,2} \right) \right|$$

$$\leqslant \left(K \sum_{k=1}^K p_{k,1}^2 \right)^{\frac{1}{2}} \left(K \sum_{k=1}^K (p_{k,2} - p'_{k,2})^2 \right)^{\frac{1}{2}}$$

$$+ \left(K \sum_{\ell=1}^K (p'_{k,2})^2 \right)^{\frac{1}{2}} \left(K \sum_{k=1}^K (p_{k,1} - p'_{k,2})^2 \right)^{\frac{1}{2}},$$

$$\leqslant (|p|_{2*} + |p'|_{2*}) |p - p'|_{2*}.$$

We can thus for instance choose $V = 3$ when applying Proposition 3.2 to our current setting. For convenience, we also fix $M := 2L + 1 = 3$, which clearly satisfies (3.15) for such choices of u and v. We thus deduce that, for every $R \in \mathbb{R}$, the mapping

$$(3.68) \qquad (t,x) \mapsto f^{(K,K')}(t,x) - f^{(K')}(t,x) - 3 \left(|x|_2 + 3t - R \right)_+$$

achives its supremum on $\{0\} \times \overline{U}_{K'}$. We now derive two different bounds on this supremum, the first one being simple and convenient for large $|x|_2$, the second one covering the case of more moderate values of this quantity. The first bound is a consequence of the estimates (3.66) and (3.67), and of the fact that $f^{(K,K')}(0,0) = f^{(K')}(0,0)$: we have

$$(3.69) \qquad f^{(K,K')}(0,x) - f^{(K')}(0,x) - 3 \left(|x|_2 - R \right)_+ \leqslant 3R - |x|_2.$$

For the second bound, we first rewrite the supremum of (3.68) over $\{0\} \times \overline{U}_{K'}$ as

$$(3.70) \qquad \sup_{x \in \overline{U}_{K'}} \left\{ \left| \psi\left(\frac{1}{K'} \sum_{k=1}^{K'} \delta_{x_k} \right) - \psi\left(\frac{1}{K} \sum_{k=1}^{K} \delta_{x_k^{(K,K')}} \right) \right| - 3\left(|x|_2 - R \right)_+ \right\}.$$

By (3.64), the difference of ψ's in the supremum above can be bounded by

$$(3.71) \qquad \frac{1}{K'} \sum_{k=1}^{K} \sum_{r=1}^{R} \left| x_{(k-1)R+r} - x_k^{(K,K')} \right| \leqslant \frac{1}{K} \sum_{k=1}^{K} \frac{1}{R^2} \sum_{r,r'=1}^{R} \left| x_{(k-1)R+r} - x_{(k-1)R+r'} \right|.$$

For every $B \in (0, \infty)$, we have

$$(3.72) \qquad \frac{1}{K'} \sum_{k=1}^{K'} |x_k| \mathbb{1}_{\{|x_k| \geqslant B\}} \leqslant \frac{1}{BK'} \sum_{k=1}^{K'} |x_k|^2 = \frac{|x|_2^2}{B}.$$

On the other hand,

$$(3.73) \qquad \frac{1}{K} \sum_{k=1}^{K} \frac{2}{R^2} \sum_{1 \leqslant r' < r \leqslant R} \left| x_{(k-1)R+r} - x_{(k-1)R+r'} \right| \mathbb{1}_{\{|x_{(k-1)R+r}| < B\}}$$

$$\leqslant \frac{2}{R^2} \sum_{1 \leqslant r' < r \leqslant R} \frac{1}{K} \sum_{k=1}^{K} \left| x_{(k-1)R+r} - x_{(k-1)R+r'} \right| \mathbb{1}_{\{|x_{(K-1)R+r}| < B\}}.$$

By equivalence of norms over $\mathbb{R}^2$, up to a constant factor, we can replace the Euclidean norm in $|x_{(k-1)R+r} - x_{(k-1)R+r'}|$ by the ℓ^1 norm; and in this case, since $x_{(k-1)R+r'} \leqslant x_{(k-1)R+r}$, the sum above becomes telescopic. We thus have that, for some absolute constant $C < \infty$,

$$\frac{2}{R^2} \sum_{1 \leqslant r' < r \leqslant R} \frac{1}{K} \sum_{k=1}^{K} \left| x_{(k-1)R+r} - x_{(k-1)R+r'} \right| \mathbb{1}_{\{|x_{(K-1)R+r}| < B\}} \leqslant \frac{CB}{K}.$$

Summarizing, we have shown that, for every $x \in \overline{U}_{K'}$ amd $B \in (0, \infty)$,

$$\left| \psi\left(\frac{1}{K'} \sum_{k=1}^{K'} \delta_{x_k} \right) - \psi\left(\frac{1}{K} \sum_{k=1}^{K} \delta_{x_k^{(K,K')}} \right) \right| \leqslant \frac{2|x|_2^2}{B} + \frac{CB}{K}.$$

For $B = \sqrt{K} |x|_2$, this becomes, up to a redefinition of $C < \infty$,

$$\left| \psi\left(\frac{1}{K'} \sum_{k=1}^{K'} \delta_{x_k} \right) - \psi\left(\frac{1}{K} \sum_{k=1}^{K} \delta_{x_k^{(K,K')}} \right) \right| \leqslant \frac{C}{\sqrt{K}} |x|_2.$$

Summarizing, we have thus shown that the quantity inside the supremum in (3.70) is bounded by

$$\frac{C}{\sqrt{K}} |x|_2 - 3\left(|x|_2 - R \right)_+.$$

Notice that the bound in (3.69) is already negative for $|x|_2 \geqslant 3R$. On the complementary event, the quantity above is clearly bounded by $\frac{3CR}{\sqrt{K}}$. Up to a redefinition of $C < \infty$, we have thus shown that, for every $R > 0$,

$$\sup_{t \geqslant 0, x \in \overline{U}_{K'}} \left\{ f^{(K,K')}(t, x) - f^{(K')}(t, x) - 3\left(|x|_2 + 3t - R \right)_+ \right\} \leqslant \frac{CR}{\sqrt{K}}.$$

Choosing $R = 3|x|_2 + 3t$ then yields one bound for (3.63). The converse bound is obtained in the same way.

Step 3. We complete the proof, by showing that there exists a constant C such that for every $t \geqslant 0$, $\mu = (\mu_1, \mu_2) \in (\mathcal{P}_2(\mathbb{R}_+))^2$, and $1 \leqslant K_1 \leqslant K_2$, we have

$$(3.74) \qquad \left| f^{(K_1)}(t, x^{(K_1)}(\mu)) - f^{(K_2)}(t, x^{(K_2)}(\mu)) \right| \leqslant \frac{C}{\sqrt{K_1}} \left(\left(\mathbb{E}\left[X_{\mu_1}^2 + X_{\mu_2}^2 \right] \right)^{\frac{1}{2}} + t \right).$$

In order to show (3.74), it suffices to verify that, for all integers $K, R \geqslant 1$, and with $K' := RK$, we have

$$(3.75) \qquad \left| f^{(K)}(t, x^{(K)}(\mu)) - f^{(K')}(t, x^{(K')}(\mu)) \right| \leqslant \frac{C}{\sqrt{K}} \left(\left(\mathbb{E} \left[X_{\mu_1}^2 + X_{\mu_2}^2 \right] \right)^{\frac{1}{2}} + t \right).$$

Indeed, once (3.75) is proved, we can apply it with (K, K') replaced by $(K_1, K_1 K_2)$ and $(K_2, K_1 K_2)$ and obtain (3.74) by the triangle inequality. But (3.75) is almost identical to (3.63): indeed, the latter identity states that the left side of (3.75) is bounded by

$$\frac{C}{\sqrt{K}} \left(|x^{(K')}(\mu)|_2 + t \right),$$

and, by Jensen's inequality,

$$(3.76) \qquad |x^{(K')}(\mu)|_2 = \left(\frac{1}{K'} \sum_{k=1}^{K'} \left| K' \int_{\frac{k-1}{K'}}^{\frac{k}{K'}} (F_{\mu_1}^{-1}(u), F_{\mu_2}^{-1}(u)) \, \mathrm{d}u \right|^2 \right)^{\frac{1}{2}} \leqslant \left(\mathbb{E} \left[X_{\mu_1}^2 + X_{\mu_2}^2 \right] \right)^{\frac{1}{2}}.$$

Now, it is clear that (3.74) guarantees the existence of the limit in (3.57). It also yields the estimate (3.58), by letting K_2 tend to infinity. To show the Lipschitz estimate (3.59), we start from (3.66), which can be rewritten as, for every $t \geqslant 0$ and $\mu, \nu \in (\mathcal{P}_2(\mathbb{R}_+))^2$,

$$\left| f^{(K)}(t, x^{(K)}(\mu)) - f^{(K)}(t, x^{(K)}(\nu)) \right| \leqslant |x^{(K)}(\mu) - x^{(K)}(\nu)|_2.$$

As in (3.76), we can then bound the right side above by

$$\left(\mathbb{E} \left[|X_{\mu_1} - X_{\nu_1}|^2 + |X_{\mu_2} - X_{\nu_2}|^2 \right] \right)^{\frac{1}{2}}.$$

The estimate (3.59) then follows by letting K tend to infinity. $\qquad \square$

4. The free energy is a supersolution

The main goal of this section is to show that finite-dimensional approximations of $\overline{F}_N$ are supersolutions to the finite-dimensional approximations of (1.7), up to a small error. Compared with the previous section, we change the indexing convention and write, for every integer $k \geqslant 1$,

$$(4.1) \quad U_k := \left\{ q = (q_{1,1}, \ldots, q_{1,k}, q_{2,1}, \ldots, q_{2,k}) \in (0, \infty)^{2k} \; : \right.$$
$$\left. \forall a \in \{1, 2\}, \forall \ell \in \{1, \ldots, k-1\}, \; q_{a,\ell} < q_{a,\ell+1} \right\},$$

and we denote the closure of U_k by $\overline{U}_k$.

Theorem 4.1 (approximate HJ equation)**.** *For each integer $k \geqslant 1$, $t \geqslant 0$, and $q \in \overline{U}_k$ indexed as $q = (q_{1,1}, \ldots, q_{1,k}, q_{2,1}, \ldots, q_{2,k})$, denote*

$$(4.2) \qquad \overline{F}_N^{(k)}(t, q) := \overline{F}_N \left(t, \frac{1}{k} \sum_{\ell=1}^{k} \delta_{q_{1,\ell}}, \frac{1}{k} \sum_{\ell=1}^{k} \delta_{q_{2,\ell}} \right),$$

and let f be any subsequential limit of $\overline{F}_N^{(k)}$ as N tends to infinity. We have, in the sense of viscosity solutions,

$$(4.3) \qquad \begin{cases} \partial_t f - k \sum_{\ell=1}^{k} \partial_{q_{1,\ell}} f \, \partial_{q_{2,\ell}} f \geqslant -\dfrac{13}{k} & \text{on } (0, \infty) \times U_k, \\[2mm] \mathbf{n} \cdot \nabla f \geqslant 0 & \text{on } (0, \infty) \times \partial U_k. \end{cases}$$

In the statement above, we understand the notion of subsequential limit in the sense of locally uniform convergence. (By Proposition 2.1, the functions involved are uniformly Lipschitz, and by Lemma 2.6, the initial condition does not depend on N, so the existence of converging subsequences is clear.) Once Theorem 4.1 is proved, we will combine it with the results of the previous section to obtain a proof of Theorem 1.1.

As was announced in Section 2, see in particular (2.23), we need to show that the overlaps $\sigma_a \cdot \sigma_a'$ are "typically" synchronized with the overlap $\alpha \wedge \alpha'$. The argument for achieving this relies on the fact that, possibly after a small perturbation of the energy function, we can ensure that the structure of the Gibbs measure is ultrametric [43]. That the ultrametricity can be used to infer synchronization was first observed in [45]; we revisit the argument in Section 5 below to provide us with a "finitary" version of the statement of synchronization, which is more adapted to the needs of the proof of Theorem 4.1. The small perturbations of the energy function are meant to ensure the validity of the Ghirlanda-Guerra identities. The reader may want to have a brief look at Section 5 to understand better the motivation behind the introduction of such perturbations.

The bird's eye view proposed above is that "typically", the overlaps synchronize. One subtle point is to uncover what "typically" should actually mean. We cannot hope for this synchronization property to hold for *every* choice of the parameters. Conversely, knowing that synchronization occurs for almost every choice of the parameters is not sufficient. Indeed, this would boil down to considering a limit free energy that satisfies the partial differential equation (3.7) at almost every point. But such a property is unfortunately *not* sufficient to identify the solution to (3.7) uniquely; and this is the reason why the more involved notion of viscosity solutions is introduced. Roughly speaking, we will be able to show that synchronization occurs at any contact point appearing in the definition of supersolution. More precisely, in the notation of Definition 3.1, at a point where $f - \phi$ is minimal, we will be able to leverage on the fact that the Hessian of f must be bounded from below to deduce the validity of the Ghirlanda-Guerra identities, and therefore the synchronization of the overlaps.

We now introduce the small perturbations of the energy function alluded to above. We fix $(\lambda_n)_{n \geqslant 1}$ an enumeration of the set of rational numbers in $[0, 1]$. For every integer triple $h = (h_1, h_2, h_3) \in \mathbb{N}_*^3$ and $a \in \{1, 2\}$, we define the random energy $(H_N^{a,h}(\sigma, \alpha))_{\sigma \in \mathbb{R}^{2N}, \alpha \in \mathbb{N}^k}$, which is a centered Gaussian field with covariance given, for every $\sigma, \sigma' \in \mathbb{R}^{2N}$ and $\alpha, \alpha' \in \mathbb{N}^k$, by

$$(4.4) \qquad \mathbb{E}\left[H_N^{a,h}(\sigma, \alpha) H_N^{a,h}(\sigma', \alpha') \right] = N \left(\lambda_{h_1} \frac{\sigma_a \cdot \sigma_a'}{N} + \lambda_{h_2} \frac{\alpha \wedge \alpha'}{k} \right)^{h_3}.$$

The fact that such a Gaussian random field exists is shown in Lemma A.2 of the appendix. (It is also seen there that the variables α can be embedded into a Hilbert space in such a way that $\alpha \wedge \alpha'$ becomes the scalar product of the "embedded" variables. Strictly speaking, this observation is required to use the results of Section 5 with these variables.) We impose the fields $(H_N^{a,h})_{a \in \{1,2\}, h \in \mathbb{N}_*^3}$ to be independent, and to be independent of the other random variables in the problem. Enlarging the probability space if necessary, we assume that these additional random fields are defined on the probability space with measure $\mathbb{P}$. Let h_+ be an integer that will be chosen sufficiently large (in terms of k) in the course of the argument. For convenience, we understand that every element $x \in \mathbb{R}^{2+2h_+^3}$ is indexed according to

$$x = (x_1, (x_{1,h})_{h \in \{1,\ldots,h_+\}^3}, x_2, (x_{2,h})_{h \in \{1,\ldots,h_+\}^3}).$$

With this understanding, we set

$$H_N^x(\sigma, \alpha) := N^{-\frac{1}{16}} \sum_{a \in \{1,2\}} \left(x_a |\sigma_a|^2 + \sum_{h \in \{1,\dots,h_+\}^3} x_{a,h} H_N^{a,h}(\sigma, \alpha) \right).$$

The prefactor $N^{-\frac{1}{16}}$ is meant to ensure that H_N^x will not contribute to the limit free energy, see (4.5) and (4.8) below. The exponent $\frac{1}{16}$ is relatively arbitrary; one could replace it by any exponent in the interval $(0, \frac{1}{8})$. We now define a new free energy that includes the perturbative terms: for every $t \geqslant 0$, $\mu \in (\mathcal{P}(\mathbb{R}_+))^2$ of the form (2.5), and $x \in \mathbb{R}^{2+2h_+^3}$, we set, with H_N^t defined in (2.2) and H_N^μ defined in (2.7),

(4.5) $$F_N(t, \mu, x) := -\frac{1}{N} \log \int \sum_{\alpha \in \mathbb{N}^k} \exp \left(H_N^t(\sigma) + H_N^\mu(\sigma, \alpha) + H_N^x(\sigma, \alpha) \right) v_\alpha \mathrm{d}P_N(\sigma),$$

as well as

$$\overline{F}_N(t, \mu, x) := \mathbb{E}\left[F_N(t, \mu, x) \right].$$

In the last two displays, we slightly abuse notation in that we keep denoting the free energy by F_N (or $\overline{F}_N$ for its average), although there are now additional variables compared to the quantity defined in (2.10). This abuse of notation does not seem to risk causing much confusion. Indeed, every identity we have seen so far is still valid if $F_N(t, \mu)$ is replaced by $F_N(t, \mu, x)$, provided that we redefine the Gibbs measure in (2.11) to include the perturbation terms. Moreover, whenever a risk of confusion arises, we can always write the variables explicitly to dispel it.

We now record a few identities involving the derivatives of F_N and $\overline{F}_N$ with respect to this new variable x. We have

(4.6) $$\partial_{x_{a,h}} F_N = -N^{-1-\frac{1}{16}} \left\langle H_N^{a,h}(\sigma, \alpha) \right\rangle,$$

and, by (4.4) and Gaussian integration by parts, see (A.2),

$$\partial_{x_{a,h}} \overline{F}_N = -N^{-1-\frac{1}{16}} \mathbb{E}\left\langle H_N^{a,h}(\sigma, \alpha) \right\rangle$$

(4.7) $$= N^{-\frac{1}{8}} x_{a,h} \mathbb{E}\left\langle \left(\lambda_{h_1} \frac{\sigma_a \cdot \sigma_a'}{N} + \lambda_{h_2} \frac{\alpha \wedge \alpha'}{k} \right)^{h_3} - \left(\lambda_{h_1} \frac{|\sigma_a|^2}{N} + \lambda_{h_2} \right)^{h_3} \right\rangle.$$

In particular, recalling that $\lambda_{h_1}, \lambda_{h_2} \in [0, 1]$, we have

(4.8) $$\left| \partial_{x_{a,h}} \overline{F}_N \right| \leqslant 2^{h_3+2} N^{-\frac{1}{8}} |x_{a,h}|.$$

Similarly, for every $a \in \{1, 2\}$,

(4.9) $$\partial_{x_a} F_N = -N^{-1-\frac{1}{16}} \left\langle |\sigma_a|^2 \right\rangle,$$

and in particular,

(4.10) $$\left| \partial_{x_a} \overline{F}_N \right| \leqslant N^{-\frac{1}{16}}.$$

We also have

(4.11) $$\partial^2_{x_{a,h}} \overline{F}_N = -N^{-1-\frac{1}{8}} \mathbb{E}\left[\left\langle \left(H_N^{a,h}(\sigma, \alpha) \right)^2 \right\rangle - \left\langle H_N^{a,h}(\sigma, \alpha) \right\rangle^2 \right],$$

and

(4.12) $$\partial^2_{x_a} \overline{F}_N = -N^{-1-\frac{1}{8}} \mathbb{E}\left[\left\langle |\sigma_a|^4 \right\rangle - \left\langle |\sigma_a|^2 \right\rangle^2 \right].$$

Before we turn to the proof of Theorem 4.1, we record a useful concentration estimate for the function F_N.

Proposition 4.2 (Concentration of F_N). *Let $k, h_+ \in \mathbb{N}_*$ and, for every $(t, q, x) \in \mathbb{R}_+ \times \overline{U}_k \times \mathbb{R}^{2+h_+^3}$, let*

$$(4.13) \qquad F_N^{(k)}(t, q, x) := F_N\left(t, \frac{1}{k}\sum_{\ell=1}^{k}\delta_{q_{1,\ell}}, \frac{1}{k}\sum_{\ell=1}^{k}\delta_{q_{2,\ell}}, x\right),$$

as well as

$$(4.14) \qquad \overline{F}_N^{(k)}(t, q, x) := \mathbb{E}\left[F_N^{(k)}(t, q, x)\right].$$

For every $M < \infty$, $p \in [1, \infty)$ and $\varepsilon > 0$, there exists $C < \infty$ such that for every $N \geqslant 1$,

$$\mathbb{E}\left[\sup_{B_M}\left|F_N^{(k)} - \overline{F}_N^{(k)}\right|^p\right]^{\frac{1}{p}} \leqslant CN^{-\frac{1}{2}+\varepsilon},$$

where B_M denotes the set of $(t, q, x) \in \mathbb{R}_+ \times \overline{U}_k \times \mathbb{R}^{2+h_+^3}$ for which each coordinate is contained in $[-M, M]$ (in other words, B_M is the intersection of $\mathbb{R}_+ \times \overline{U}_k \times \mathbb{R}^{2+h_+^3}$ with the $(3 + 2k + h_+^3)$-dimensional L^∞ ball of radius M).

Proof. By [44, Theorem 1.2], there exists $C < \infty$ such that for every $(t, q, x) \in B_M$ and $a \geqslant 0$,

$$(4.15) \qquad \mathbb{P}\left[\left|(F_N^{(k)} - \overline{F}_N^{(k)})(t, q, x)\right|^2 \geqslant \frac{a}{N}\right] \leqslant 2\exp\left(-\frac{a}{C}\right).$$

In order to conclude, we need some estimate on the modulus of continuity of $F_N^{(k)}$. We denote

$$X := 1 + \frac{|\langle H_N(\sigma)\rangle|}{N} + \frac{1}{N}\sum_{a \in \{1,2\}}\left(\sum_{\ell=0}^{k}\left|\langle z_{\alpha_{|\ell},a} \cdot \sigma_a\rangle\right| + \sum_{h \in \{1,\ldots,h_+\}^3}\left|\langle H_N^{a,h}(\sigma, \alpha)\rangle\right|\right).$$

By integration of (2.20), we see that for every $t, t' \in [0, M]$, $q \in \overline{U}_k$, and $x \in \mathbb{R}^{2+h_+^3}$,

$$\left|F_N^{(k)}(t', q, x) - F_N^{(k)}(t, q, x)\right| \leqslant C\left(1 + \frac{|\langle H_N(\sigma)\rangle|}{N}\right)|t' - t|^{\frac{1}{2}}$$

$$\leqslant CX|t' - t|^{\frac{1}{2}}.$$

Similarly, we can compute, for every $\ell \in \{1, \ldots, k-1\}$,

$$\partial_{q_{a,\ell}}F_N^{(k)} = -\frac{1}{N}\left\langle (2q_{a,\ell} - 2q_{a,\ell-1})^{-\frac{1}{2}}z_{\alpha_{|\ell},a} \cdot \sigma_a - (2q_{a,\ell+1} - 2q_{a,\ell})^{-\frac{1}{2}}z_{\alpha_{|\ell+1},a} \cdot \sigma_a\right\rangle,$$

with the understanding that $q_{a,0} = 0$ here, and, in the case $\ell = k$,

$$\partial_{q_{a,k}}F_N^{(k)} = -\frac{1}{N}\left\langle (2q_{a,k} - 2q_{a,k-1})^{-\frac{1}{2}}z_{\alpha_{|k},a} \cdot \sigma_a\right\rangle.$$

By integration, we find that, for every $q, q' \in \overline{U}_k$, $t \geqslant 0$, and $x \in \mathbb{R}^{2+h_+^3}$,

$$\left|F_N^{(k)}(t, q', x) - F_N^{(k)}(t, q, x)\right| \leqslant CX\sum_{a \in \{1,2\}}\sum_{\ell=1}^{k}|q'_{a,\ell} - q_{a,\ell}|^{\frac{1}{2}}.$$

Finally, by (4.6) and (4.9), we also have that

$$\left|F_N^{(k)}(t, q, x') - F_N^{(k)}(t, q, x)\right| \leqslant CN^{-\frac{1}{16}}X|x' - x|.$$

On the other hand, it follows from (2.21), (2.22), (4.8), and (4.10), that the function $\overline{F}_N^{(k)}$ is Lipschitz continuous (globally in t and q, and locally in x). For every $\varepsilon \in (0, 1]$, we denote

$$A_\varepsilon := \left(\varepsilon\mathbb{Z}^{3+2k+h_+^3}\right) \cap B_M.$$

The previous estimates imply that

$$\sup_{B_M} \left| F_N^{(k)} - \overline{F}_N^{(k)} \right| \leqslant \sup_{A_\varepsilon} \left| F_N^{(k)} - \overline{F}_N^{(k)} \right| + CX\sqrt{\varepsilon},$$

and therefore, for every $p \geqslant 1$,

$$\mathbb{E}\left[\sup_{B_M} \left| F_N^{(k)} - \overline{F}_N^{(k)} \right|^p \right] \leqslant C\,\mathbb{E}\left[\sup_{A_\varepsilon} \left| F_N^{(k)} - \overline{F}_N^{(k)} \right|^p \right] + C\varepsilon^{\frac{p}{2}}\mathbb{E}[X^p],$$

with a constant $C < \infty$ that may depend on p (in addition to k, h_+, and M). We bound the supremum over A_ε by the sum over A_ε and use (4.15) to get

$$\mathbb{E}\left[\sup_{A_\varepsilon} \left| F_N^{(k)} - \overline{F}_N^{(k)} \right|^p \right] \leqslant C|A_\varepsilon| N^{-\frac{p}{2}} = C\varepsilon^{-\left(3+2k+h_+^3\right)} N^{-\frac{p}{2}}.$$

Using (A.4), we see that, for every $(t, q, x) \in B_M$,

$$\mathbb{E}\left\langle (H_N(\sigma))^{2p} \right\rangle \leqslant CN^{2p},$$

and similarly, for every $a \in \{1, 2\}$ and $\ell \in \{0, \ldots, k\}$,

$$\mathbb{E}\left\langle (z_{\alpha_{|\ell,a}} \cdot \sigma_a)^{2p} \right\rangle \leqslant CN^{2p},$$

as well as, for every $h \in \{1, \ldots, h_+\}^3$,

$$\mathbb{E}\left\langle (H_N^{a,h}(\sigma, \alpha))^{2p} \right\rangle \leqslant CN^{2p}.$$

By Jensen's inequality, this implies that $\mathbb{E}[X^p] \leqslant C$ (in other words, $\mathbb{E}[X^p]$ is bounded uniformly over N). We have thus shown that, with $\alpha := 3 + 2k + h_+^3$,

$$\mathbb{E}\left[\sup_{B_M} \left| F_N^{(k)} - \overline{F}_N^{(k)} \right|^p \right]^{\frac{1}{p}} \leqslant C\varepsilon^{-\frac{\alpha}{p}} N^{-\frac{1}{2}} + C\varepsilon^{\frac{1}{2}}.$$

Choosing $\varepsilon = N^{-\frac{p}{p+2\alpha}}$, we can bound the right side above by $CN^{-\frac{p}{2(p+2\alpha)}}$. By taking p sufficiently large, we can bring the exponent $\frac{p}{2(p+2\alpha)}$ as close to $\frac{1}{2}$ as desired. By Jensen's inequality, this proves the claim. $\qquad\square$

Proof of Theorem 4.1. We fix the integer h_+ sufficiently large that, with the choice of $\varepsilon = k^{-4}$, the statement of Proposition 5.5 holds for some $\delta \geqslant h_+^{-1}$. Recall that we slightly abuse notation and write $\overline{F}_N^{(k)}$ both to denote the function in (4.2) and the function in (4.14). We can dispel the confusion by writing $(t, q) \mapsto \overline{F}_N^{(k)}(t, q)$ for the former and $(t, q, x) \mapsto \overline{F}_N^{(k)}(t, q, x)$ for the latter. In order to lighten the notation, we drop the superscript $^{(k)}$ and simply write $\overline{F}_N$ in place of $\overline{F}_N^{(k)}$ throughout (and similarly for $F_N^{(k)}$). Let f be a subsequential limit of the mapping $(t, q) \mapsto \overline{F}_N(t, q)$. For convenience, we omit to denote the particular subsequence along which the convergence of $(t, q) \mapsto \overline{F}_N(t, q)$ to f holds.

Let $(t_\infty, q_\infty) \in (0, \infty) \times \overline{U}_k$ and $\phi \in C^\infty((0, \infty) \times \overline{U}_k)$ be such that $f - \phi$ has a local minimum at (t_∞, q_∞). Without loss of generality, we may assume that

$$(4.16) \qquad q_\infty \in \partial\overline{U}_k \quad\Longrightarrow\quad \max_{\nu \in \mathbf{n}(q_\infty)} \nu \cdot \nabla\phi(t_\infty, q_\infty) < 0,$$

since in the complementary event, the Neumann boundary condition is satisfied. Under this condition, we will show that

$$(4.17) \qquad \left(\partial_t \phi - k \sum_{\ell=1}^{k} \partial_{q_{1,\ell}} \phi\, \partial_{q_{2,\ell}} \phi \right)(t_\infty, q_\infty) \geqslant -\frac{13}{k}.$$

Throughout the rest of this proof, we denote by $C < \infty$ a constant whose value may change from one occurence to another, and is allowed to depend on k, h_+ (which itself has already been fixed in terms of k), t_∞, q_∞, and the function ϕ. We write

$$(4.18) \qquad x_\infty := (1, \ldots, 1) \in \mathbb{R}^{2+2h_+^3}.$$

For every $(t, q) \in (0, \infty) \times \overline{U}_k$, and $x \in \mathbb{R}^{2+2h_+^3}$, we set

$$(4.19) \qquad \widetilde{\phi}(t, q, x) := \phi(t, q) - (t - t_\infty)^2 - |q - q_\infty|^2 - |x - x_\infty|^2.$$

The mapping $(t, q, x) \mapsto f(t, q) - \widetilde{\phi}(t, q, x)$ has a strict local minimum at $(t_\infty, q_\infty, x_\infty)$. In view of (4.8) and (4.10), the mapping $(t, q, x) \mapsto \overline{F}_N(t, q, x)$ converges to the mapping $(t, q, x) \mapsto f(t, q)$ locally uniformly. We deduce that there exist $(t_N, q_N, x_N) \in (0, \infty) \times \overline{U}_k \times \mathbb{R}^{2+2h_+^3}$ satisfying

$$(4.20) \qquad \lim_{N \to \infty} (t_N, q_N, x_N) = (t_\infty, q_\infty, x_\infty)$$

and such that, for every N sufficiently large, the function $\overline{F}_N - \widetilde{\phi}$ has a local minimum at (t_N, q_N, x_N); more precisely, for every N sufficiently large,

$$(4.21) \quad (\overline{F}_N - \widetilde{\phi})(t_N, q_N, x_N)$$
$$= \inf \left\{ (\overline{F}_N - \widetilde{\phi})(t, q, x) : \ |t - t_N| + |q - q_N| + |x - x_N| \leqslant C^{-1} \right\}.$$

(In the infimum above, we also have the implicit restriction $q \in \overline{U}_k$, and we may choose C sufficiently large that the condition $|t - t_N| \leqslant C^{-1}$ implies that $t > 0$.) In particular,

$$(4.22) \qquad \partial_t (\overline{F}_N - \widetilde{\phi})(t_N, q_N, x_N) = 0,$$

$$(4.23) \qquad q_N \in U_k \quad \Longrightarrow \quad \nabla_q (\overline{F}_N - \widetilde{\phi})(t_N, q_N, x_N) = 0,$$

and

$$(4.24) \qquad \nabla_x (\overline{F}_N - \widetilde{\phi})(t_N, q_N, x_N) = 0.$$

We decompose the rest of the proof into five steps.

Step 1. In this step, we show that for every N sufficiently large and $|x| \leqslant C^{-1}$,

$$(4.25) \qquad -C|x|^2 \leqslant \overline{F}_N(t_N, q_N, x_N + x) - \overline{F}_N(t_N, q_N, x_N) - x \cdot \nabla_x \overline{F}_N(t_N, q_N, x_N) \leqslant 0.$$

The second inequality follows from the fact that $\overline{F}_N$ is a concave function of x (it is classical to verify that the function F_N itself is concave in x, since the Hessian of this function is a covariance matrix, up to a minus sign). To show the first inequality in (4.25), we start by writing Taylor's formula:

$$(4.26) \quad \overline{F}_N(t_N, q_N, x_N + x) - \overline{F}_N(t_N, q_N, x_N)$$
$$= x \cdot \nabla_x \overline{F}_N(t_N, q_N, x_N) + \int_0^1 (1 - s) x \cdot \nabla_x^2 \overline{F}_N(t_N, q_N, x_N + sx) x \, \mathrm{d}s,$$

where $\nabla_x^2 \overline{F}_N$ denotes the Hessian of the function $\overline{F}_N$ in the x variable. Naturally, the formula above is also valid if we replace $\overline{F}_N$ by $\widetilde{\phi}$. By (4.21), we have that for every $|x| \leqslant C^{-1}$,

$$\overline{F}_N(t_N, q_N, x_N + x) - \overline{F}_N(t_N, q_N, x_N) \geqslant \widetilde{\phi}(t_N, q_N, x_N + x) - \widetilde{\phi}(t_N, q_N, x_N).$$

Using also (4.24), we obtain that

$$\int_0^1 (1 - s) x \cdot \nabla_x^2 \overline{F}_N(t_N, q_N, x_N + sx) x \, \mathrm{d}s$$
$$\geqslant \int_0^1 (1 - s) x \cdot \nabla_x^2 \widetilde{\phi}(t_N, q_N, x_N + sx) x \, \mathrm{d}s \geqslant -C|x|^2.$$

Combining this with (4.26) yields (4.25).

Step 2. We show that, for every $\varepsilon > 0$,

$$(4.27) \qquad \mathbb{E}\left[\left|\nabla_x (F_N - \overline{F}_N)(t_N, q_N, x_N)\right|^2\right] \leqslant CN^{-\frac{1}{2}+\varepsilon},$$

where now we also allow the constant $C < \infty$ to depend on the choice of $\varepsilon > 0$. As observed in the previous step, the function F_N is concave in the x variable. We thus have, for every $x \in \mathbb{R}^{2+2h_+^3}$,

$$F_N(t_N, q_N, x_N + x) \leqslant F_N(t_N, q_N, x_N) + x \cdot \nabla_x F_N(t_N, q_N, x_N).$$

By (4.25), we also have, for every $|x| \leqslant C^{-1}$,

$$\overline{F}_N(t_N, q_N, x_N + x) \geqslant \overline{F}_N(t_N, q_N, x_N) + x \cdot \nabla_x \overline{F}_N(t_N, q_N, x_N) - C|x|^2.$$

For a (deterministic) parameter $\lambda \in [0, C^{-1}]$ to be determined in the course of the argument, we combine the two inequalities above and fix

$$x = \lambda \frac{\nabla_x (\overline{F}_N - F_N)(t_N, q_N, x_N)}{|\nabla_x (\overline{F}_N - F_N)(t_N, q_N, x_N)|},$$

so that $|x| \leqslant C^{-1}$, and, for this choice of x,

$$\lambda \left|\nabla_x (\overline{F}_N - F_N)(t_N, q_N, x_N)\right|$$
$$\leqslant (\overline{F}_N - F_N)(t_N, q_N, x_N + x) - (\overline{F}_N - F_N)(t_N, q_N, x_N) + C\lambda^2.$$

By Proposition 4.2, we infer that

$$\lambda^2 \mathbb{E}\left[\left|\nabla_x (\overline{F}_N - F_N)(t_N, q_N, x_N)\right|^2\right] \leqslant CN^{-1+2\varepsilon} + C\lambda^4.$$

Choosing $\lambda = N^{-\frac{1}{4}+\frac{\varepsilon}{2}}$ yields (4.27).

Step 3. We show that the Gibbs measure associated with the choice of parameters (t_N, q_N, x_N) satisfies approximate Ghirlanda-Guerra identities, in the following sense. Recall that we denote by $(\sigma^\ell, \alpha^\ell)_{\ell \geqslant 1}$ a family of independent copies of (σ, α) under $\langle \cdot \rangle$. For each $\ell, \ell' \in \mathbb{N}_*$ and $a \in \{1, 2\}$, we write

$$R_0^{\ell,\ell'} := \frac{\alpha^\ell \wedge \alpha^{\ell'}}{k}, \quad R_a^{\ell,\ell'} := \frac{\sigma_a^\ell \cdot \sigma_a^{\ell'}}{N},$$

and, for each $n \in \mathbb{N}_*$, we denote by $R^{\leqslant n}$ the array

$$R^{\leqslant n} := \left(R_a^{\ell,\ell'}\right)_{a\in\{0,1,2\},\ell,\ell'\in\{1,\dots,n\}}.$$

In this step, we show that, for every $\varepsilon > 0$, $a \in \{1, 2\}$, $n, h_1, h_2, h_3 \in \{1, \dots, h_+\}$, and $g \in C(\mathbb{R}^{3n^2})$ satisfying $\|g\|_{L^\infty} \leqslant 1$, we have

$$(4.28) \quad \left| \mathbb{E}\left\langle f(R^{\leqslant n})\left(\lambda_{h_1} R_a^{1,n+1} + \lambda_{h_2} R_0^{1,n+1}\right)^{h_3}\right\rangle - \frac{1}{n}\mathbb{E}\left\langle g(R^{\leqslant n})\right\rangle \mathbb{E}\left\langle \left(\lambda_{h_1} R_a^{1,2} + \lambda_{h_2} R_0^{1,2}\right)^{h_3}\right\rangle \right.$$

$$\left. - \frac{1}{n}\sum_{\ell=2}^{n}\mathbb{E}\left\langle g(R^{\leqslant n})\left(\lambda_{h_1} R_a^{1,\ell} + \lambda_{h_2} R_0^{1,\ell}\right)^{h_3}\right\rangle \right| \leqslant CN^{-\frac{1}{8}+\varepsilon},$$

where we understand that the Gibbs measure $\langle \cdot \rangle$ is with the parameters (t_N, q_N, x_N). It follows from (4.25) that

$$-C \leqslant \nabla_x^2 \overline{F}_N(t_N, q_N, x_N) \leqslant 0.$$

In particular, by (4.11), for every $a \in \{1, 2\}$ and $h = (h_1, h_2, h_3) \in \{1, \dots, h_+\}^3$, we have

$$\mathbb{E}\left[\left\langle \left(H_N^{a,h}(\sigma)\right)^2\right\rangle - \left\langle H_N^{a,h}(\sigma)\right\rangle^2\right] \leqslant CN^{1+\frac{1}{8}},$$

and similarly, by (4.12),

$$\mathbb{E}\left[\left\langle|\sigma_a|^4\right\rangle - \left\langle|\sigma_a|^2\right\rangle^2\right] \leqslant CN^{1+\frac{1}{8}}.$$

By (4.6) and (4.27), we also have

$$\mathbb{E}\left[\left(\left\langle H_N^{a,h}(\sigma)\right\rangle - \mathbb{E}\left\langle H_N^{a,h}(\sigma)\right\rangle\right)^2\right] \leqslant CN^{\frac{3}{2}+\frac{1}{8}+\varepsilon},$$

and similarly, by (4.9) and (4.27),

$$\mathbb{E}\left[\left(\left\langle|\sigma_a|^2\right\rangle - \mathbb{E}\left\langle|\sigma_a|\right\rangle\right)^2\right] \leqslant CN^{\frac{3}{2}+\frac{1}{8}+\varepsilon}.$$

Since, for any random variable X, we have the variance decomposition

$$\mathbb{E}\left\langle(X - \mathbb{E}\left\langle X\right\rangle)^2\right\rangle = \mathbb{E}\left\langle(X - \left\langle X\right\rangle)^2\right\rangle + \mathbb{E}\left[(\left\langle X\right\rangle - \mathbb{E}\left\langle X\right\rangle)^2\right],$$

we deduce that

(4.29) $$\mathbb{E}\left\langle\left(H_N^{a,h}(\sigma) - \mathbb{E}\left\langle H_N^{a,h}(\sigma)\right\rangle\right)^2\right\rangle \leqslant CN^{\frac{3}{2}+\frac{1}{8}+\varepsilon},$$

and

(4.30) $$\mathbb{E}\left\langle\left(|\sigma_a|^2 - \mathbb{E}\left\langle|\sigma_a|^2\right\rangle\right)^2\right\rangle \leqslant CN^{\frac{3}{2}+\frac{1}{8}+\varepsilon}.$$

It follows from (4.29) that

$$\left|\mathbb{E}\left\langle g(R^{\leqslant n})H_N^{a,h}(\sigma^1,\alpha^1)\right\rangle - \mathbb{E}\left\langle g(R^{\leqslant n})\right\rangle\mathbb{E}\left\langle H_N^{a,h}(\sigma^1,\alpha^1)\right\rangle\right| \leqslant CN^{\frac{3}{4}+\frac{1}{16}+\varepsilon}.$$

Recall the expression for $\mathbb{E}\left\langle H_N^{a,h}(\sigma,\alpha)\right\rangle$ in (4.7). By Gaussian integration by parts, see (A.2), we also have

$$N^{-1+\frac{1}{16}}\mathbb{E}\left\langle g(R^{\leqslant n})H_N^{a,h}(\sigma^1,\alpha^1)\right\rangle = \sum_{\ell=1}^{n} x_{a,h}\mathbb{E}\left\langle g(R^{\leqslant n})\left(\lambda_{h_1}\frac{\sigma_a^1\cdot\sigma_a^\ell}{N} + \lambda_{h_2}\frac{\alpha^1\wedge\alpha^\ell}{k}\right)^{h_3}\right\rangle$$
$$- nx_{a,h}\mathbb{E}\left\langle g(R^{\leqslant n})\left(\lambda_{h_1}\frac{\sigma_a^1\cdot\sigma_a^{n+1}}{N} + \lambda_{h_2}\frac{\alpha^1\wedge\alpha^{n+1}}{k}\right)^{h_3}\right\rangle,$$

where we dropped the dependence on N and simply wrote $x_{a,h}$ for the (a,h) coordinate of the vector x_N. Recall that $x_N \to x_\infty$ with x_∞ defined in (4.18), so that for N sufficiently large, we have $x_{a,h} \geqslant \frac{1}{2}$ (this is the point of defining x_∞ in this way, as opposed to setting $x_\infty = 0$). Matching the term indexed by $\ell = 1$ in the sum above with the last term in (4.7), we would like to show that the difference

$$\mathbb{E}\left\langle g(R^{\leqslant n})\left(\lambda_{h_1}\frac{|\sigma_a^1|^2}{N} + \lambda_{h_2}\right)^{h_3}\right\rangle - \mathbb{E}\left\langle g(R^{\leqslant n})\right\rangle\mathbb{E}\left\langle\left(\lambda_{h_1}\frac{|\sigma_a^1|^2}{N} + \lambda_{h_2}\right)^{h_3}\right\rangle$$

is small. Using (4.30) and the fact that the mapping $r \mapsto r^{h_3}$ is Lipschitz over $[0,2]$, we can bound this difference (in absolute value) by $CN^{-\frac{1}{4}+\frac{1}{16}+\varepsilon}$. Collecting the terms, we obtain (4.28).

Step 4. We now use the synchronization result of Section 5. Recall from (2.23) that

$$\left|\partial_t\overline{F}_N - k\sum_{\ell=1}^{k}\partial_{q_{1,\ell}}\overline{F}_N\,\partial_{q_{2,\ell}}\overline{F}_N\right| \leqslant \frac{1}{N^2}\sum_{a\in\{1,2\}}\mathbb{E}\left\langle\left(\sigma_a\cdot\sigma_a' - \mathbb{E}\left\langle\sigma_a\cdot\sigma_a'\,\middle|\,\alpha\wedge\alpha'\right\rangle\right)^2\right\rangle.$$

By (4.28), Proposition 5.5 (with the quantities R_1 and R_2 appearing there being substituted by R_a and R_0 in our current notation), and our choice of h_+, we infer that for every N sufficiently large,

(4.31) $$\left|\partial_t\overline{F}_N - k\sum_{\ell=1}^{k}\partial_{q_{1,\ell}}\overline{F}_N\,\partial_{q_{2,\ell}}\overline{F}_N\right|(t_N,q_N,x_N) \leqslant \frac{13}{k}.$$

We will argue in the next step that, for every N sufficiently large,

$$(4.32) \qquad \left(\sum_{\ell=1}^{k} \partial_{q_{1,\ell}} \overline{F}_N \, \partial_{q_{2,\ell}} \overline{F}_N - \sum_{\ell=1}^{k} \partial_{q_{1,\ell}} \widetilde{\phi} \, \partial_{q_{2,\ell}} \widetilde{\phi} \right) (t_N, q_N, x_N) \geqslant 0.$$

Temporarily assuming this, and using also (4.22), we thus infer that

$$(4.33) \qquad \left(\partial_t \widetilde{\phi} - k \sum_{\ell=1}^{k} \partial_{q_{1,\ell}} \widetilde{\phi} \, \partial_{q_{2,\ell}} \widetilde{\phi} \right) (t_N, q_N, x_N) \geqslant -\frac{13}{k}.$$

Using (4.20) and the fact that $\widetilde{\phi}$ is a smooth function, we deduce that the statement (4.33) also holds at $(t_\infty, q_\infty, x_\infty)$. Recalling also the definition of $\widetilde{\phi}$, see (4.19), we conclude that (4.17) holds.

Step 5. There only remains to show that (4.32) holds. If $q_N \in U_k$, then this is immediate, by (4.23). In particular, since q_N converges to q_∞ as N tends to infinity, we know that (4.32) holds for every N sufficiently large whenever $q_\infty \in U_k$. From now on, we assume that $q_\infty \in \partial U_k$. Using that (t_N, q_N, x_N) tends to $(t_\infty, q_\infty, x_\infty)$, the smoothness of the function ϕ, and the definition of $\widetilde{\phi}$ in (4.19), we can infer from (4.16) that, for every N sufficiently large,

$$(4.34) \qquad \max_{\nu \in \mathbf{n}(q_\infty)} \nu \cdot \nabla \widetilde{\phi}(t_N, q_N, x_N) \leqslant 0.$$

We now argue that

$$(4.35) \qquad q_N \in \partial \overline{U}_k \quad \Longrightarrow \quad \max_{\nu \in \mathbf{n}(q_N)} \nu \cdot \nabla \widetilde{\phi}(t_N, q_N, x_N) \leqslant 0.$$

The set $\overline{U}_k$ can be written as the intersection of $2k$ half-spaces, and the condition that $q_\infty \in \partial U_k$ is equivalent to the statement that q_∞ lies on the boundary of some of those half-spaces, say $\mathcal{D}_1, \ldots, \mathcal{D}_\ell$. For N sufficiently large, we know that q_N will not be on the boundary of any other half-space than those $\mathcal{D}_1, \ldots, \mathcal{D}_\ell$. It thus follows that, for N sufficiently large, and whenever $q_N \in \overline{U}_k$, we have that $\mathbf{n}(q_N)$ is a subset of $\mathbf{n}(q_\infty)$. This yields that (4.35) indeed follows from (4.34).

We now argue that

$$(4.36) \qquad \nabla \widetilde{\phi}(t_N, q_N, x_N) \in \overline{U}_k.$$

Since the arguments t_N and x_N will be kept fixed, we omit them from the notation. We denote by $(e_{a,\ell})_{a \in \{1,2\}, \ell \in \{1, \ldots k\}}$ the canonical basis of $\mathbb{R}^{2k}$, using our indexing convention. We also write $q_N = (q_{a,\ell})_{a \in \{1,2\}, 1 \leqslant \ell \leqslant k}$, dropping the dependence on N when writing q_N in coordinates. We fix $a \in \{1, 2\}$, and first show that $\partial_{q_{a,1}} \widetilde{\phi}(q_N) \geqslant 0$. If $q_{a,1} = 0$, then this follows from (4.35), since $-e_{a,1} \in \mathbf{n}(q_N)$ in this case. Otherwise, since the function $\overline{F}_N - \widetilde{\phi}$ has a local minimum at q_N, we have, for every $\varepsilon > 0$ sufficiently small,

$$\widetilde{\phi}(q_N - \varepsilon e_{a,1}) - \widetilde{\phi}(q_N) \leqslant \overline{F}_N(q_N - \varepsilon e_{a,1}) - \overline{F}_N(q_N).$$

Passing to the limit $\varepsilon \to 0$, and using Lemma 2.4, we obtain that $\partial_{q_{q,1}} \widetilde{\phi}(q_N) \geqslant 0$. We now fix $\ell \in \{1, \ldots, k-1\}$, and argue that $\partial_{q_{a,\ell}} \widetilde{\phi}(q_N) \leqslant \partial_{q_{a,\ell+1}} \widetilde{\phi}(q_N)$. If $q_{a,\ell} = q_{a,\ell+1}$, then this follows from (4.35), since $2^{-\frac{1}{2}}(e_{a,\ell} - e_{a,\ell+1}) \in \mathbf{n}(q_N)$ in this case. Otherwise, we have, for every $\varepsilon > 0$ sufficiently small,

$$\widetilde{\phi}(q_N + \varepsilon e_{a,\ell} - \varepsilon e_{a,\ell+1}) - \widetilde{\phi}(q_N) \leqslant \overline{F}_N(q_N + \varepsilon e_{a,\ell} - \varepsilon e_{a,\ell+1}) - \overline{F}_N(q_N).$$

Passing to the limit, and using Lemma 2.4, we obtain indeed that $\partial_{q_{a,\ell}} \widetilde{\phi}(q_N) \leqslant \partial_{q_{a,\ell+1}} \widetilde{\phi}(q_N)$.

We are now ready to conclude. For every $p = (p_{a,\ell})_{a\in\{1,2\},\ell\in\{1,\dots,k\}} \in \mathbb{R}^{2k}$, denote

$$\mathsf{H}(p) := \sum_{\ell=1}^{k} p_{1,\ell}\, p_{2,\ell}.$$

Our aim is to show that

$$(4.37) \qquad\qquad \mathsf{H}(\nabla_q \overline{F}_N) - \mathsf{H}(\nabla_q \widetilde{\phi}) \geqslant 0,$$

where we kept implicit that the gradients are evaluated at (t_N, q_N, x_N). We rewrite the left side of (4.37) in the form

$$(4.38) \qquad\qquad \int_0^1 (\nabla_q \overline{F}_N - \nabla_q \widetilde{\phi}) \cdot \nabla \mathsf{H}(s\nabla_q \overline{F}_N + (1-s)\nabla_q \widetilde{\phi})\, \mathrm{d}s.$$

(Evaluation at (t_N, q_N, x_N) is still kept implicit here.) Since $\overline{F}_N - \widetilde{\phi}$ has a local minimum at q_N, we must have that

$$(4.39) \qquad\qquad \forall y \in \overline{U}_k, \qquad (y - q_N) \cdot \nabla_q(\overline{F}_N - \widetilde{\phi})(t_N, q_N, x_N) \geqslant 0.$$

Recalling the notation $\overline{U}_k^*$ from (3.17), we infer from (4.39) that

$$(4.40) \qquad\qquad \nabla_q(\overline{F}_N - \widetilde{\phi})(t_N, q_N, x_N) \in \overline{U}_k^*.$$

(Indeed, for every $z \in \overline{U}_k$, we can choose $y = q_N + z \in \overline{U}_k$ in (4.39).) By (4.36) and Lemma 2.4, we have that, for every $s \in [0, 1]$,

$$\left(s\nabla_q \overline{F}_N + (1-s)\nabla_q \widetilde{\phi}\right)(t_N, q_N, x_N) \in \overline{U}_k.$$

Finally, notice that, for each $p = (p_{a,\ell})_{a\in\{1,2\},\ell\in\{1,\dots,k\}}$, the vector $\nabla \mathsf{H}(p)$ is obtained by a simple interchange of the index $a \in \{1, 2\}$. In particular, it is clear that $\nabla \mathsf{H}$ maps $\overline{U}_k$ into itself. Combining this with the previous display, we infer that the quantity $\nabla \mathsf{H}(\cdots)$ appearing in (4.38) belongs to $\overline{U}_k$. This and (4.40) yield (4.37), as desired. $\qquad\square$

We can now prove the main theorem of the paper.

Proof of Theorem 2.7. The argument consists in combining the results of Proposition 3.7 and Theorem 4.1. We denote by $(f^{(k)})_{k\geqslant 1}$ and f the functions appearing in the statement of Proposition 3.7. By Theorem 4.1 and the comparison principle (3.12), we have, for every integer $k \geqslant 1$, $t \geqslant 0$ and $q \in \mathbb{R}^{2k}$,

$$(4.41) \qquad\qquad \liminf_{N\to\infty} \overline{F}_N\left(t, \frac{1}{k}\sum_{\ell=1}^{k}\delta_{q_{1,\ell}}, \frac{1}{k}\sum_{\ell=1}^{k}\delta_{q_{2,\ell}}\right) + \frac{13t}{k} \geqslant f^{(k)}(t, q).$$

Let $\mu = (\mu_1, \mu_2) \in (\mathcal{P}_2(\mathbb{R}_+))^2$, and, for every integer $k \geqslant 1$, $a \in \{1, 2\}$, and $\ell \in \{1, \dots, k\}$, denote

$$q_{a,\ell}^{(k)} := k \int_{\frac{\ell-1}{k}}^{\frac{\ell}{k}} F_{\mu_a}^{-1}(u)\, \mathrm{d}u, \quad \text{and} \quad \mu_a^{(k)} := \sum_{\ell=1}^{k} \delta_{q_{a,\ell}}.$$

Recall from (3.58) that

$$(4.42) \qquad\qquad \left|f^{(k)}(t, q^{(k)}) - f(t, \mu)\right| \leqslant \frac{C}{\sqrt{k}}\left(t + \left(\mathbb{E}\left[X_{\mu_1}^2 + X_{\mu_2}^2\right]\right)^{\frac{1}{2}}\right).$$

On the other hand, we have from Proposition 2.1 that

$$\left|\overline{F}_N(t, \mu) - \overline{F}_N(t, \mu^{(k)})\right| \leqslant \sum_{a=1}^{2} \mathbb{E}\left[\left|X_{\mu_a} - X_{\mu_a^{(k)}}\right|\right]$$

$$= \sum_{a=1}^{2} \mathbb{E}\left[\left|F_{\mu_a}^{-1}(U) - k\int_{\frac{\lfloor kU\rfloor}{k}}^{\frac{\lfloor kU\rfloor+1}{k}} F_{\mu_a}^{-1}(u)\, \mathrm{d}u\right|\right].$$

We can bound this term by arguing as in the paragraph starting with (3.71). Indeed, this is the same argument, with the understanding that K' is now infinite. Explicitly,

$$\mathbb{E}\left[\left|F_{\mu_a}^{-1}(U) - k\int_{\frac{\lfloor kU\rfloor}{k}}^{\frac{\lfloor kU\rfloor+1}{k}} F_{\mu_a}^{-1}(u)\,\mathrm{d}u\right|\right] = \sum_{\ell=1}^{k}\int_{\frac{\ell-1}{k}}^{\frac{\ell}{k}}\left|F_{\mu_a}^{-1}(r) - k\int_{\frac{\ell-1}{k}}^{\frac{\ell}{k}} F_{\mu_a}^{-1}(u)\,\mathrm{d}u\right|\,\mathrm{d}r$$

$$\leqslant k\sum_{\ell=1}^{k}\int_{\frac{\ell-1}{k}}^{\frac{\ell}{k}}\int_{\frac{\ell-1}{k}}^{\frac{\ell}{k}}\left|F_{\mu_a}^{-1}(r) - F_{\mu_a}^{-1}(u)\right|\,\mathrm{d}u\,\mathrm{d}r.$$

Paralleling (3.72), we write, for a cutoff value $B \in (0,\infty)$ to be determined,

$$\int_0^1 |F_{\mu_a}^{-1}(u)|\mathbb{1}_{\{F_{\mu_a}^{-1}(u)\geqslant B\}}\,\mathrm{d}u \leqslant \frac{1}{B}\int |F_{\mu_a}^{-1}(u)|^2\,\mathrm{d}u = \frac{\mathbb{E}[X_{\mu_a}^2]}{B},$$

while, as in (3.73),

$$2k\sum_{\ell=1}^{k}\int_{\frac{\ell-1}{k}}^{\frac{\ell}{k}}\int_{\frac{\ell-1}{k}}^{\frac{\ell}{k}}\left|F_{\mu_a}^{-1}(r) - F_{\mu_a}^{-1}(u)\right|\mathbb{1}_{\{u\leqslant r, F_{\mu_a}^{-1}(r)\leqslant B\}}\,\mathrm{d}u\,\mathrm{d}r$$

$$= 2k\sum_{\ell=1}^{k}\int_{\left[0,\frac{1}{k}\right]^2}\left(F_{\mu_a}^{-1}\left(\frac{\ell-1}{k}+r\right) - F_{\mu_a}^{-1}\left(\frac{\ell-1}{k}+u\right)\right)\mathbb{1}_{\{u\leqslant r, F_{\mu_a}^{-1}\left(\frac{\ell-1}{k}+r\right)\leqslant B\}}\,\mathrm{d}u\,\mathrm{d}r$$

$$\leqslant \frac{2B}{k}.$$

Combining the displays above, and choosing $B^2 = k(\mathbb{E}[X_{\mu_1}^2 + X_{\mu_2}^2])$, we arrive at

(4.43) $$\left|\overline{F}_N(t,\mu) - \overline{F}_N(t,\mu^{(k)})\right| \leqslant \frac{6}{\sqrt{k}}\left(\mathbb{E}[X_{\mu_1}^2 + X_{\mu_2}^2]\right)^{\frac{1}{2}}.$$

Combining (4.41), (4.42), and (4.43), we deduce that

$$\liminf_{N\to\infty}\overline{F}_N(t,\mu) \geqslant f(t,\mu) - \frac{13t}{k} - \frac{C}{\sqrt{k}}\left(t + \left(\mathbb{E}[X_{\mu_1}^2 + X_{\mu_2}^2]\right)^{\frac{1}{2}}\right).$$

Letting the integer $k \geqslant 1$ tend to infinity, we obtain the desired result. $\square$

5. Synchronization

In this section, we revisit the synchronization result of [45], see also [46, 47]. The structure of the reasoning presented here is similar to that in [45], and emphasizes the fundamental importance of the ultrametric structure of the Gibbs measure. There are a few differences though: one of them is that we state "finitary" versions of the statements; that is, the statements provide approximate criteria that the Gibbs measure may satisfy for large but finite values of N and k; the conclusion is then that we have "synchronization up to a small error". A second difference between the treatment presented here and [45] is in the phrasing of the synchronization property itself. In [45], this is stated as the existence of Lipschitz functions that each map the sum of the overlaps of the different species to one of the single-species overlaps. In the present section, we instead choose to phrase the synchronization of different overlaps as the statement that they are monotonically coupled.

As said above, the main powerhouse behind the synchronization result comes from the possibility to enforce the ultrametricity of the Gibbs measure. The fundamental result of [43] is that the ultrametricity property is valid as soon as the Ghirlanda-Guerra identities hold; see also the preface to [44] for a review of the series of works that preceded this final result. Moreover, as is well-known and was seen again in Section 4, these identities are valid as soon as certain random energy functions become concentrated, a

property that one can "build into the measure" by means of a small perturbation of the energy function.

In order to emphasize that the underlying constants in the statements below do not depend on the specific Gibbs measure under consideration, we will state them for rather general measures. We start by stating a finitary version of the statement from [43] that "Ghirlanda-Guerra identities imply ultrametricity".

Theorem 5.1 (GG implies ultrametricity [43])**.** *For every $\varepsilon > 0$, there exists $\delta > 0$ such that the following holds. Let G be a random probability measure supported on the unit ball of an arbitrary Hilbert space; denote by $\langle \cdot \rangle$ the expectation associated with the measure $G^{\otimes \mathbb{N}}$, with canonical random variables $(\sigma^\ell)_{\ell \geqslant 1}$, and define, for every $\ell, \ell', n \geqslant 1$,*

$$R^{\ell,\ell'} := \sigma^\ell \cdot \sigma^{\ell'}, \quad and \quad R^{\leqslant n} := \left(R^{\ell,\ell'} \right)_{1 \leqslant \ell, \ell' \leqslant n}.$$

Finally, recalling that $\langle \cdot \rangle$ is itself random, denote by $\mathbb{E}$ the expectation with respect to this additional source of randomness. Assume that, for every $n, p \in \{1, \ldots, \lfloor \delta^{-1} \rfloor\}$ and $f \in C(\mathbb{R}^{n \times n})$ satisfying $\|f\|_{L^\infty} \leqslant 1$,

$$(5.1) \quad \left| \mathbb{E} \left\langle f(R^{\leqslant n})(R^{1,n+1})^p \right\rangle - \frac{1}{n} \mathbb{E} \left\langle f(R^{\leqslant n}) \right\rangle \mathbb{E} \left\langle (R^{1,2})^p \right\rangle - \frac{1}{n} \sum_{\ell=2}^n \mathbb{E} \left\langle f(R^{\leqslant n})(R^{1,\ell})^p \right\rangle \right| \leqslant \delta.$$

Then

$$(5.2) \quad \mathbb{E} \left\langle \mathbb{1}_{\{R^{1,2} \geqslant \min(R^{1,3}, R^{2,3}) - \varepsilon\}} \right\rangle \geqslant 1 - \varepsilon.$$

Proof. We argue by contradiction. Denote by $R := (R^{\ell,\ell'})_{\ell,\ell' \geqslant 1}$ the entire overlap array, and assume that Theorem 5.1 is false: there exists $\varepsilon > 0$ and, for each $\delta > 0$ no matter how small, a random probability measure G such that (5.1) holds but (5.2) is violated. Since each entry of R takes values in $[-1, 1]$, up to extraction of a subsequence, we can find a random array $\mathsf{R} = (\mathsf{R}^{\ell,\ell'})_{\ell,\ell' \geqslant 1}$ defined with respect to a certain probability measure $\mathbb{M}$ such that, for each integer $n \geqslant 1$, the law of the array $\mathsf{R}^k := (\mathsf{R}^{\ell,\ell'})_{1 \leqslant \ell, \ell' \leqslant n}$ under $\mathbb{M}$ is obtained as the limit law of a subsequence of overlap arrays, each violating (5.2) but satisfying (5.1) for a sequence of values of δ that tends to zero. In other words, the array R satisfies, for every integers $n, p \geqslant 1$ and $f \in C(\mathbb{R}^{n \times n})$,

$$\mathbb{M} \left[f(\mathsf{R}^{\leqslant n})(\mathsf{R}^{1,n+1})^p \right] = \frac{1}{n} \mathbb{M} \left[f(\mathsf{R}^{\leqslant n}) \right] \mathbb{M} \left[(\mathsf{R}^{1,2})^p \right] + \frac{1}{n} \sum_{\ell=2}^n \mathbb{M} \left[f(\mathsf{R}^{\leqslant n})(\mathsf{R}^{1,\ell})^p \right],$$

as well as

$$\mathbb{M} \left[R^{1,2} \leqslant \min \left(R^{1,3}, R^{2,3} \right) - \varepsilon \right] \geqslant \varepsilon.$$

This was shown to be impossible in [43], see also [44, Theorem 2.14]. $\qquad \square$

In order to prepare the ground for synchronization statements, we clarify the notion of monotone coupling in the next proposition.

Proposition 5.2 (Monotone coupling)**.** *Let (X, Y) be a random vector taking values in $\mathbb{R}^2$, and let (X', Y') be an independent copy of this vector, defined under the probability measure $\mathbb{P}$. The following three statements are equivalent.*

(1) We have

$$(5.3) \quad \mathbb{P} \left[X < X' \text{ and } Y' < Y \right] = 0.$$

(2) For every $x, y \in \mathbb{R}$, we have

$$(5.4) \quad \mathbb{P} \left[X \leqslant x \text{ and } Y \leqslant y \right] = \min \left(\mathbb{P} \left[X \leqslant x \right], \mathbb{P} \left[Y \leqslant y \right] \right).$$

(3) The law of (X,Y) is

$$(F_X^{-1}, F_Y^{-1})\left(\mathrm{Leb}_{[0,1]}\right),$$

that is, the law of (X,Y) is the image of the Lebesgue measure over $[0,1]$ under the mapping $r \mapsto (F_X^{-1}(r), F_Y^{-1}(r))$, where, for every $r \in [0,1]$,

(5.5) $$F_X^{-1}(r) := \inf\left\{s \in \mathbb{R} \ : \ \mathbb{P}\left[X \leqslant s\right] \geqslant r\right\},$$

and similarly with X replaced by Y.

Whenever any of the conditions (1-3) appearing in Proposition 5.2 holds, we say that the random variables X and Y are *monotonically coupled.*

Proof. We first show that (1) implies (2). The statement (5.4) with the equality sign replaced by "$\leqslant$" is clear. To show the converse inequality, we argue by contradiction and assume that there exist $x, y \in \mathbb{R}$ such that

$$\mathbb{P}\left[X \leqslant x \text{ and } Y \leqslant y\right] < \min\left(\mathbb{P}\left[X \leqslant x\right], \mathbb{P}\left[Y \leqslant y\right]\right).$$

It follows that

$$\mathbb{P}\left[X \leqslant x \text{ and } Y > y\right] = \mathbb{P}\left[X \leqslant x\right] - \mathbb{P}\left[X \leqslant x \text{ and } Y \leqslant y\right] > 0,$$

and similarly,

$$\mathbb{P}\left[X > x \text{ and } Y \leqslant y\right] = \mathbb{P}\left[Y \leqslant y\right] - \mathbb{P}\left[X \leqslant x \text{ and } Y \leqslant y\right] > 0.$$

In particular,

$$\mathbb{P}\left[X \leqslant x < X' \text{ and } Y' \leqslant y < Y\right] > 0.$$

This contradicts (5.3).

We now show that (3) implies (1). Let U and U' be two independent random variables distributed uniformly over $[0,1]$. We can realize (X,Y) and (X',Y') by setting

$$(X,Y) = (F_X^{-1}, F_Y^{-1})(U) \quad \text{and} \quad (X',Y') = (F_X^{-1}, F_Y^{-1})(U').$$

For definiteness, suppose that $U \leqslant U'$. Since F_X^{-1} and F_Y^{-1} are increasing (in the sense of wide inequalities), it then implies that $X \leqslant X'$ and $Y \leqslant Y'$. This shows that property (1) holds.

Summarizing, we have shown that (1) implies (2) and (3) implies (1). In particular, (3) implies (2). Since there is at most one joint law for (X,Y) that satisfies (2), we deduce that (2) and (3) are equivalent. The proof is thus complete. $\qquad\square$

We now turn to our variant of the main result of [45], which states that approximate Ghirlanda-Guerra identities imply approximate synchronization, in the sense of monotone couplings between overlaps.

Theorem 5.3 (Synchronization). *Let $(\lambda_n)_{n \geqslant 1}$ be an enumeration of the set of rational numbers in $[0,1]$. For every $\varepsilon > 0$, there exists $\delta > 0$ such that the following holds. Let G be a random probability measure supported on the Cartesian product of the unit balls of two arbitrary Hilbert spaces; denote by $\langle \cdot \rangle$ the expectation associated with the measure $G^{\otimes \mathbb{N}}$, with canonical random variables $\left(\sigma^\ell = (\sigma_1^\ell, \sigma_2^\ell)\right)_{\ell \geqslant 1}$, and define, for every $a \in \{1,2\}$ and $\ell, \ell', n \geqslant 1$,*

$$R_a^{\ell,\ell'} := \sigma_a^\ell \cdot \sigma_a^{\ell'}, \quad \text{and} \quad R^{\leqslant n} := \left(R_a^{\ell,\ell'}\right)_{a \in \{1,2\}, 1 \leqslant \ell, \ell' \leqslant n}.$$

Finally, recalling that $\langle \cdot \rangle$ is itself random, denote by $\mathbb{E}$ the expectation with respect to this additional source of randomness. Assume that, for every $n, h_1, h_2, p \in \{1, \ldots, \lfloor \delta^{-1} \rfloor\}$ and $f \in C(\mathbb{R}^{2 \times n \times n})$ satisfying $\|f\|_{L^\infty} \leqslant 1$,

$$(5.6) \quad \left| \mathbb{E} \left\langle f(R^{\leqslant n}) \left(\lambda_{h_1} R_1^{1,n+1} + \lambda_{h_2} R_2^{1,n+1} \right)^p \right\rangle \right.$$

$$\left. - \frac{1}{n} \mathbb{E} \left\langle f(R^{\leqslant n}) \right\rangle \mathbb{E} \left\langle \left(\lambda_{h_1} R_1^{1,2} + \lambda_{h_2} R_2^{1,2} \right)^p \right\rangle - \frac{1}{n} \sum_{\ell=2}^{n} \mathbb{E} \left\langle f(R^{\leqslant n}) \left(\lambda_{h_1} R_1^{1,\ell} + \lambda_{h_2} R_2^{1,\ell} \right)^p \right\rangle \right| \leqslant \delta.$$

Then, for every $f \in C^\infty(\mathbb{R}^2)$,

$$(5.7) \quad \left| \mathbb{E} \left\langle f(R_1^{1,2}, R_2^{1,2}) \right\rangle - \mathbb{E} \left[f \left(F_1^{-1}(U), F_2^{-1}(U) \right) \right] \right| \leqslant \varepsilon \left(\|f\|_{L^\infty} + \|\nabla f\|_{L^\infty} \right),$$

where U stands for a uniform random variable over $[0,1]$, and, for every $a \in \{1, 2\}$ and $r \in [0,1]$, we write

$$(5.8) \quad F_a^{-1}(r) := \inf \left\{ s \in \mathbb{R} \ : \ \mathbb{E} \left\langle \mathbb{1}_{\{R_a^{1,2} \leqslant s\}} \right\rangle \geqslant r \right\}.$$

The proof of Theorem 5.3 makes use of the following lemma, asserting that if two sequences of random variables converge in law separately, then their monotone coupling converges in law as well.

Lemma 5.4 (continuity of monotone coupling). *Let $(X_n), (Y_n)$ be two sequences of random variables which converge in law to X and Y respectively. Then the associated monotone couplings converge: using the notation in (5.5), and with U a uniform random variable over $[0,1]$, we have*

$$(5.9) \quad \left(F_{X_n}^{-1}(U), F_{Y_n}^{-1}(U) \right) \xrightarrow[n \to \infty]{(law)} \left(F_X^{-1}(U), F_Y^{-1}(U) \right).$$

Proof. Since the law of $F_{X_n}^{-1}(U)$ is that of X_n, it is clear that the convergence in (5.9) holds for each coordinate separately. Up to the extraction of a subsequence, we can assume that $\left(F_{X_n}^{-1}(U), F_{Y_n}^{-1}(U) \right)$ converges in law to some random vector (A, B); we denote by (A', B') an independent copy of this vector. By classical properties of convergence in law and Proposition 5.2, we infer that

$$\mathbb{P} \left[A < A' \text{ and } B' < B \right] = 0.$$

We conclude using Proposition 5.2 once more. □

Proof of Theorem 5.3. Step 1. For any two probability measures μ, ν on $[-1, 1]^2$, we define

$$\|\mu - \nu\| := \sup \left\{ \int f \, \mathrm{d}\mu - \int f \, \mathrm{d}\nu \ : \ \|f\|_{L^\infty} + \|\nabla f\|_{L^\infty} \leqslant 1 \right\}.$$

In this step, we show that the quantity above, as a function of (μ, ν), is continuous for the topology of weak convergence. In other words, if a sequence of probability measures μ_n over $[-1, 1]^2$ converges weakly to μ, then $\|\mu_n - \nu\|$ converges to $\|\mu - \nu\|$. For every integer $k \geqslant 1$ and $x \in \mathbb{R}^2$, define

$$P_k(x) := c_k \left(1 - \frac{|x|^2}{16} \right)^k,$$

where the constant c_k is such that $\int_{[-2,2]^2} P_k = 1$. Let $f \in C^\infty([-1, 1]^2)$ be such that $\|f\|_{L^\infty} + \|\nabla f\|_{L^\infty} \leqslant 1$. We may extend f to a Lipschitz function on $\mathbb{R}^2$ such that $\|f\|_{L^\infty} +$

$\|\nabla f\|_{L^\infty} \leqslant 2$. Denoting the spatial convolution by $*$, we have, for every $x \in [-1,1]^2$,

$$(f - f * P_k)(x) = \int_{[-2,2]^2} (f(x) - f(x-y)) P_k(y) \, \mathrm{d}y$$

$$= \int_{[-2,2]^2} \int_0^1 y \cdot \nabla f(x - ty) P_k(y) \, \mathrm{d}t \, \mathrm{d}y,$$

so

$$\|f - f * P_k\|_{L^\infty} \leqslant 2 \int_{[-2,2]^2} |y| \, P_k(y) \, \mathrm{d}y,$$

and the latter quantity tends to 0 as k tends to infinity (uniformly over f). On the other hand, $f * P_k$ is a polynomial of degree at most $2k$ and, and for each fixed k, the coefficients of this polynomial can be bounded in terms of $\|f\|_{L^\infty}$. In particular, for each fixed k, we have

$$\sup \left\{ \int f * P_k \, \mathrm{d}\mu_n - \int f * P_k \, \mathrm{d}\mu \; : \; \|f\|_{L^\infty} + \|\nabla f\|_{L^\infty} \leqslant 1 \right\} \xrightarrow[n \to \infty]{} 0.$$

Combining these two facts gives the announced continuity result.

Step 2. We need to show that, provided that $\delta > 0$ is chosen sufficiently small in terms of ε, we have

$$(5.10) \qquad \left\| \mathrm{Law}(R_1^{1,2}, R_2^{1,2}) - (F_1^{-1}, F_2^{-1}) \left(\mathrm{Leb}_{[0,1]} \right) \right\| \leqslant \varepsilon.$$

In the expression above, we denote by $\mathrm{Law}(R_1^{1,2}, R_2^{1,2})$ the law of $(R_1^{1,2}, R_2^{1,2})$ under the measure $\mathbb{E}\langle \cdot \rangle$. Assuming the contrary, there exist $\varepsilon > 0$ and, for $\delta > 0$ as small as desired, an overlap distribution satisfying (5.6) but not (5.10). Up to extraction of a subsequence, we can assume that the overlap array converges in law to a limit random overlap R, whose law we denote by $\mathbb{M}$. By Lemma 5.4 and the result of the previous step, we infer that

$$\left\| \mathrm{Law}(\mathsf{R}_1^{1,2}, \mathsf{R}_2^{1,2}) - (F_1^{-1}, F_2^{-1}) \left(\mathrm{Leb}_{[0,1]} \right) \right\| \geqslant \varepsilon,$$

where in the expression above, F_1^{-1} and F_2^{-1} now stand for the inverse cumulative distribution functions of $\mathsf{R}_1^{1,2}$ and $\mathsf{R}_2^{1,2}$ respectively (that is, we replace $\mathbb{E}\left\langle \mathbb{1}_{\{R_a^{1,2} \leqslant s\}} \right\rangle$ by $\mathbb{M}\left[\mathbb{1}_{\{\mathsf{R}_a^{1,2} \leqslant s\}} \right]$ in (5.8)). In particular, the random variables $\mathsf{R}_1^{1,2}$ and $\mathsf{R}_2^{1,2}$ are not monotonically coupled.

Step 3. We now show that $\mathsf{R}_1^{1,2}$ and $\mathsf{R}_2^{1,2}$ are in fact monotonically coupled, thereby reaching a contradiction. Denote by $\widetilde{\mathsf{R}}^{1,2}$ an independent copy of $\mathsf{R}^{1,2}$. (Notice that this is with respect to the "averaged" measure $\mathbb{M}$, so $\mathsf{R}^{3,4}$ would not qualify as an independent copy of $\mathsf{R}^{1,2}$ in this sense.) By Proposition 5.2, we need to show that

$$(5.11) \qquad \mathbb{M}\left[\mathsf{R}_1^{1,2} < \widetilde{\mathsf{R}}_1^{1,2} \text{ and } \widetilde{\mathsf{R}}_2^{1,2} < \mathsf{R}_2^{1,2} \right] = 0.$$

We first observe that, by the construction of R, we have that for all integers $n, h_1, h_2, p \geqslant 1$ and $f \in C(\mathbb{R}^{2 \times n \times n})$,

$$\mathbb{M}\left[f(\mathsf{R}^{\leqslant n}) \left(\lambda_{h_1} \mathsf{R}_1^{1,n+1} + \lambda_{h_2} \mathsf{R}_2^{1,n+1} \right)^p \right]$$

$$= \frac{1}{n} \mathbb{M}\left[f(\mathsf{R}^{\leqslant n}) \right] \mathbb{M}\left[\left(\lambda_{h_1} \mathsf{R}_1^{1,2} + \lambda_{h_2} \mathsf{R}_2^{1,2} \right)^p \right] + \frac{1}{n} \sum_{\ell=2}^n \mathbb{M}\left[f(\mathsf{R}^{\leqslant n}) \left(\lambda_{h_1} \mathsf{R}_1^{1,\ell} + \lambda_{h_2} \mathsf{R}_2^{1,\ell} \right)^p \right].$$

Since every continuous function can be uniformly approximated by a polynomial on compact sets, and using the Cramér-Wold theorem, we deduce that conditionally on $\mathsf{R}^{\leqslant n}$, the law of $\mathsf{R}^{1,n+1}$ is

$$\frac{1}{n} \mathrm{Law}(\mathsf{R}^{1,2}) + \frac{1}{n} \sum_{\ell=2}^n \delta_{\mathsf{R}^{1,\ell}},$$

where $\mathrm{Law}(\mathsf{R}^{1,2})$ denotes the law of $\mathsf{R}^{1,2}$ under $\mathbb{M}$, and $\delta_{\mathsf{R}^{1,\ell}}$ is the Dirac mass at $\mathsf{R}^{1,\ell}$. In particular,

$$2\mathbb{M}\left[\mathsf{R}_1^{1,2} < \mathsf{R}_1^{1,3} \text{ and } \mathsf{R}_2^{1,3} < \mathsf{R}_2^{1,2}\right]$$
$$= \mathbb{M}\left[\mathsf{R}_1^{1,2} < \widetilde{\mathsf{R}}_1^{1,2} \text{ and } \widetilde{\mathsf{R}}_2^{1,2} < \mathsf{R}_2^{1,2}\right] + \mathbb{M}\left[\mathsf{R}_1^{1,2} < \mathsf{R}_1^{1,2} \text{ and } \mathsf{R}_2^{1,2} < \mathsf{R}_2^{1,2}\right]$$
$$= \mathbb{M}\left[\mathsf{R}_1^{1,2} < \widetilde{\mathsf{R}}_1^{1,2} \text{ and } \widetilde{\mathsf{R}}_2^{1,2} < \mathsf{R}_2^{1,2}\right].$$

The statement (5.11) we aim to show is thus equivalent to

$$(5.12) \qquad\qquad \mathbb{M}\left[\mathsf{R}_1^{1,2} < \mathsf{R}_1^{1,3} \text{ and } \mathsf{R}_2^{1,3} < \mathsf{R}_2^{1,2}\right] = 0.$$

The validity of (5.12) now follows from the fact that R_1, R_2, and $\mathsf{R}_1 + \mathsf{R}_2$ are ultrametric, which itself is a consequence of Theorem 5.1. Indeed, by ultrametricity, we have

$$\mathsf{R}_1^{1,2} < \mathsf{R}_1^{1,3} \implies \mathsf{R}_1^{2,3} = \mathsf{R}_1^{1,2},$$

and

$$\mathsf{R}_2^{1,3} < \mathsf{R}_2^{1,2} \implies \mathsf{R}_2^{2,3} = \mathsf{R}_2^{1,3},$$

so that

$$\mathsf{R}_1^{1,2} < \mathsf{R}_1^{1,3} \text{ and } \mathsf{R}_2^{1,3} < \mathsf{R}_2^{1,2} \implies \mathsf{R}_1^{2,3} + \mathsf{R}_2^{2,3} < \min\left(\mathsf{R}_1^{1,2} + \mathsf{R}_2^{1,2}, \mathsf{R}_1^{1,3} + \mathsf{R}_2^{1,3}\right),$$

and the latter statement contradicts the ultrametricity of $\mathsf{R}_1 + \mathsf{R}_2$. This completes the proof of (5.12), and therefore of Theorem 5.3. $\qquad\square$

As was apparent in (2.23), what we ultimately want to use is not only that two overlaps asymptotically become monotonically coupled, but rather that one of the overlaps can essentially be inferred by observing the other. Even if the two overlaps were perfectly synchronized, this can only be true if the law of the observed overlap is sufficiently "spread out": in an extreme example, if the observed overlap is deterministic, then the statement of monotone coupling is uninformative, and the conditional variances in (2.23) boil down to regular variances, which need not be small. In the next proposition, we give a precise statement to this effect. That is, we show that if the law of one of the overlaps is sufficiently spread out, then the conditional variance of the other overlap is small. The usefulness of writing finitary versions of the statements of ultrametricity and synchronization appears most clearly here.

Proposition 5.5 (Control of conditional variance). *For every $\varepsilon > 0$, there exists $\delta > 0$ such that the following holds. Let $\mathbb{E}$, $\langle\cdot\rangle$, $(R_a^{\ell,\ell'})$ be as in the statement of Theorem 5.3, and assume that (5.6) holds for every $n, h_1, h_2, p \in \{1, \ldots, \lfloor\delta^{-1}\rfloor\}$ and $f \in C(\mathbb{R}^{2 \times n \times n})$ satisfying $\|f\|_{L^\infty} \leqslant 1$. Assume furthermore that the law of $R_2^{1,2}$ is of the form*

$$\frac{1}{k}\sum_{\ell=1}^{k}\delta_{q_\ell},$$

for some integer $k \geqslant 1$ and parameters $-1 = q_0 < q_1 < \cdots < q_k \leqslant 1$. We then have

$$(5.13) \qquad \mathbb{E}\left(\left(R_1^{1,2} - \mathbb{E}\left\langle R_1^{1,2} \,\middle|\, R_2^{1,2}\right\rangle\right)^2\right) \leqslant \frac{12}{k} + \varepsilon k^2 \sup_{\ell \in \{0,\ldots,k-1\}} (q_{\ell+1} - q_\ell)^{-1}.$$

Proof of Proposition 5.5. By Theorem 5.3, we can choose $\delta > 0$ sufficiently small that (5.7) holds for every Lipschitz function $f : \mathbb{R}^2 \to \mathbb{R}$. We set

$$\eta := \inf_{\ell \in \{0,\ldots,k-1\}} (q_{\ell+1} - q_\ell),$$

and, for every $a \in \{1, 2\}$ and $s \in \mathbb{R}$,

$$F_a(s) := \mathbb{E}\left\langle \mathbb{1}_{\{R_a^{1,2} \leqslant s\}}\right\rangle,$$

with F_a^{-1} defined as in (5.8). By assumption, the function F_2 is piecewise constant, with discontinuities at $q_1, \ldots, q_k$. Let $\widetilde{F}_2$ denote the function which coincides with F_2 on the set $(-\infty, q_0] \cup \{q_0, \ldots, q_k\} \cup [q_k, +\infty)$, and is affine on each interval $[q_\ell, q_{\ell+1}]$, $\ell \in \{0, \ldots, k-1\}$. The function $\widetilde{F}_2$ satisfies $\|\widetilde{F}_2\|_{L^\infty} \leq 1$ and $\|\nabla \widetilde{F}_2\|_{L^\infty} \leq \eta^{-1}$. Notice that, for every $u \in [0, 1]$,

$$(5.14) \qquad\qquad \widetilde{F}_2(F_2^{-1}(u)) = F_2(F_2^{-1}(u)) = k^{-1}\lceil ku \rceil.$$

We define, for every $x \in \mathbb{R}$,

$$\rho_k(x) := k \max(1 - |kx|, 0),$$

and observe that $\int \rho_k = 1$ and that the convolution $F_1^{-1} * \rho_k$ is a Lipschitz function, with Lipschitz constant bounded by k^2. For every $x, y \in [-1, 1]$, we set

$$f(x, y) = (x - (F_1^{-1} * \rho_k)(\widetilde{F}_2(y)))^2.$$

The Lipschitz constant of this function is bounded by $2\eta^{-1}k^2$, and thus

$$\left| \mathbb{E}\left\langle f(R_1^{1,2}, R_2^{1,2}) \right\rangle - \mathbb{E}\left[f(F_1^{-1}(U), F_2^{-1}(U)) \right] \right| \leq 3\varepsilon \eta^{-1} k^2.$$

Using (5.14) and the fact that F_1^{-1} takes values in $[-1, 1]$ and is monotone, we can estimate the second term on the left side above by

$$\mathbb{E}\left[\left(F_1^{-1}(U) - (F_1^{-1} * \rho_k)(k^{-1}\lceil kU \rceil) \right)^2 \right] \leq 2\mathbb{E}\left[\left| F_1^{-1}(U) - (F_1^{-1} * \rho_k)(k^{-1}\lceil kU \rceil) \right| \right]$$

$$= 2 \sum_{\ell=0}^{k-1} \int_{\frac{\ell}{k}}^{\frac{\ell+1}{k}} \left| F_1^{-1}(u) - (F_1^{-1} * \rho_k)(k^{-1}\lceil ku \rceil) \right| \, du$$

$$\leq \frac{2}{k} \sum_{\ell=0}^{k-1} \left(F_1^{-1}\left(\frac{\ell+2}{k}\right) - F_1^{-1}\left(\frac{\ell-1}{k}\right) \right)$$

$$\leq \frac{12}{k}.$$

We have thus shown that

$$\mathbb{E}\left\langle \left(R_1^{1,2} - (F_1^{-1} * \rho_k)(\widetilde{F}_2(R_2^{1,2})) \right)^2 \right\rangle \leq 3\varepsilon \eta^{-1} k^2 + 12k^{-1},$$

and thus in particular, since the conditional expectation is an L^2 projection,

$$\mathbb{E}\left\langle \left(R_1^{1,2} - \mathbb{E}\left\langle R_1^{1,2} \mid R_2^{1,2} \right\rangle \right)^2 \right\rangle \leq 3\varepsilon \eta^{-1} k^2 + 12k^{-1}.$$

Up to a redefinition of ε, this is (5.13). $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

6. The free energy as a saddle-point problem?

This final section has a more speculative flavor, and concerns the possibility to rewrite the limit free energy of models such as the one investigated here in the form of a saddle-point problem, a possibility discussed for instance in [54]. A strong indication in favor of this possibility comes from the study of certain models of statistical inference. The statistical-inference problem most similar to the spin-glass model studied here is probably that of estimating a non-symmetric rank-one matrix. This problem was investigated in [37, 6, 50, 17], and it was found there that the free energy could indeed be conveniently represented in the form of a saddle-point problem.

Of course, any quantity can be written as a saddle-point problem, so the relevant question is whether there is some natural way for doing so. The point of view provided by Hamilton-Jacobi equations suggests two natural routes for finding variational formulations of the limit free energy. The first one, available only when the nonlinearity in the equation is convex, consists in writing the Hopf-Lax formula for the solution, see for instance [40]. As was already emphasized, the main feature of the model under consideration here is

that the nonlinearity in the equation is *not* convex (nor concave). The second possible route is based on the fact that, irrespectively of the structure of the nonlinearity, it is also possible to write the solution of a Hamilton-Jacobi equation as a saddle-point problem, *provided that the initial condition is concave (or convex)*, as was suggested also by Hopf in [33] and then confirmed rigorously using the notion of viscosity solutions in [7] (see also [35]).

This second possibility can be applied to good effect in the context of the model of statistical inference studied in [37, 6, 50, 17]: as was shown in [17], the relevant Hamilton-Jacobi equation is a finite-dimensional version of (1.7), and the initial condition is convex, thereby allowing to recover the saddle-point formulas obtained in [37, 6, 50].

However, perhaps surprisingly, this strategy does not seem to work in the context of the model under consideration in this paper, and it is the aim of this section to explore this more precisely.

This point hides an important subtelty, which requires that we introduce more precise language to speak about concavity properties of the initial condition. Indeed, one can endow the set of probability measures with *two* different geometric structures. Perhaps the more immediate one is to think of it as an affine subspace of the space of signed measures. In this point of view, the natural "straight line" between the measures μ and ν is given by $t \mapsto (1-t)\mu + t\nu$. The second relevant geometric structure on the space of probability measures is that given by optimal transport. In this second point of view, the natural "straight line" between the measures μ and ν can be seen as the set of laws of the random variables $(1-t)X_\mu + tX_\nu$, with t varying in $[0,1]$, and where the law of (X_μ, X_ν) is an optimal coupling between the measures μ and ν (since we are only concerned with one-dimensional measures here, the coupling given by (1.5) is optimal).

These two points of view give rise to two different notions of convexity, which we will call "affine convexity" and "transport convexity" respectively. (The notion of "transport convexity" is sometimes also called "displacement convexity".) The subtelty here is that, at least in the simpler setting of mixed p-spin models, the initial condition in (1.7) is affine-concave, as was shown in [4]; but, whether for these p-spin models or for the bipartite model investigated here, this initial condition is *not* transport-concave (nor transport-convex). And, since the derivatives in (1.7) are transport-type derivatives, it is the notion of transport concavity (or convexity) that would have been required to guarantee saddle-point formulas by the general mechanism described above.

In the remainder of this section, we examine more precisely what natural attempts at writing saddle-point formulas for the solution to (1.7) may look like, and explain why these attempts fail in general (although we do not exclude the possibility that they be valid for some specific choices of the measures π_1 and π_2 in (1.3)).

6.1. **Attempts based on the Hopf formula.** We start by arguing that the initial condition ψ in (1.7) is neither transport-concave nor transport-convex in general. This observation is also valid for models with a single type such as mixed p-spin models. The transport concavity (or convexity) of the mapping $\mu \mapsto \psi(\mu)$ would imply in particular that the mapping

$$(6.1) \qquad \chi : \left\{ \begin{array}{ccc} \mathbb{R}_+ & \to & \mathbb{R} \\ h & \mapsto & \psi\left((\delta_h, \delta_0)\right) \end{array} \right.$$

is concave (or convex). Recall from (2.19) that

$$\chi(h) = \psi((\delta_h, \delta_0)) = -\mathbb{E} \log \int \exp\left((2h)^{\frac{1}{2}} z_1 \sigma_1 - h(\sigma_1)^2\right) \mathrm{d}\pi_1(\sigma_1),$$

where here σ_1 is real-valued, and z_1 is a standard one-dimensional Gaussian random variable. Dropping the subscript "1" on z_1 and σ_1 to lighten the notation, and denoting by $\langle \cdot \rangle$ the corresponding Gibbs measure, we have

$$\partial_h \chi = \mathbb{E}\left\langle \sigma \sigma' \right\rangle,$$

and

$$
\begin{aligned}
\partial_h^2 \chi &= \mathbb{E}\left\langle \sigma \sigma' \left((2h)^{-\frac{1}{2}} z(\sigma + \sigma') - \sigma^2 - (\sigma')^2 \right) \right\rangle - 2\mathbb{E}\left\langle \sigma \sigma' \left((2h)^{-\frac{1}{2}} z\sigma'' - (\sigma'')^2 \right) \right\rangle \\
&= \mathbb{E}\left\langle \sigma \sigma' \left((\sigma + \sigma')(\sigma + \sigma' - 2\sigma'') - \sigma^2 - (\sigma')^2 \right) \right\rangle \\
&\quad - 2\mathbb{E}\left\langle \sigma \sigma' \left(\sigma''(\sigma + \sigma' + \sigma'' - 3\sigma''') - (\sigma'')^2 \right) \right\rangle \\
&= 2\mathbb{E}\left[\left\langle \sigma^2 \right\rangle^2 - 4\left\langle \sigma^2 \right\rangle \left\langle \sigma \right\rangle^2 + 3\left\langle \sigma \right\rangle^4 \right] \\
&= 2\mathbb{E}\left[\left(\left\langle \sigma^2 \right\rangle - \left\langle \sigma \right\rangle^2 \right) \left(\left\langle \sigma^2 \right\rangle - 3\left\langle \sigma \right\rangle^2 \right) \right].
\end{aligned}
$$

Recall also that when $h = 0$, the Gibbs measure simplifies into being the measure π_1. It is therefore clear that we can choose the measure π_1 in such a way that $\partial_h^2 \chi(h = 0)$ has any desired sign: for instance, if π_1 is the uniform measure on $\{-1, 1\}$, then $\langle \sigma \rangle = 0$ at $h = 0$, so $\partial_h^2 \chi > 0$; but if we choose π_1 to be the probability measure on $\{-1, 1\}$ such that $\left\langle \sigma \right\rangle^2 = \frac{1}{2}$ at $h = 0$, then we have $\left\langle \sigma^2 \right\rangle = 1 < 3\left\langle \sigma \right\rangle^2 = \frac{3}{2}$, and thus $\partial_h^2 \chi < 0$ at $h = 0$. In both examples, we also have that $\partial_h \chi$ tends to 1 as h tends to infinity. In the case with $\left\langle \sigma \right\rangle^2 = \frac{1}{2}$ at $h = 0$, the derivative $\partial_h \chi$ at $h = 0$ is $\frac{1}{2}$ and then decreases, but must then tend to 1. In particular, the function $\partial_h \chi$ is not monotone: that is, the function χ is neither concave nor convex.

In a possibly confusing twist, for the most studied case in which π_1 is the uniform measure on $\{-1, 1\}$, one can show that the function χ is in fact convex. This implies that, at least for the model with a single type, the replica-symmetric solution for this specific choice of measure can in fact be written as a saddle-point problem. But, as is argued here, this is an accident rather than the rule. My understanding is that the solution proposed in [34, 27, 28] is based on this coincidence. As a side note, it is also worth mentioning that the de Ameida-Thouless-type stability criterion employed there is known to be invalid in general, even for models with a single type [42].

As was recalled in (1.9) (see also [40]), the limit free energy of mixed p-spin models can be expressed in terms of the solution $f_1 = f_1(t, \mu) : \mathbb{R}_+ \times \mathcal{P}_2(\mathbb{R}_+) \to \mathbb{R}$ of the equation

$$
(6.2) \qquad
\begin{cases}
\partial_t f_1 - \int \xi(\partial_\mu f_1)\, d\mu = 0 & \text{on } \mathbb{R}_+ \times \mathcal{P}_2(\mathbb{R}_+), \\
f_1(0, \cdot) = \psi_1 & \text{on } \mathcal{P}_2(\mathbb{R}_+),
\end{cases}
$$

and this solution can be written in variational form using the Hopf-Lax formula: we have

$$
(6.3) \qquad f_1(t, \mu) = \sup_{\nu \in \mathcal{P}_2(\mathbb{R}_+)} \left(\psi_1(\nu) - t\mathbb{E}\left[\xi^*\left(\frac{X_\nu - X_\mu}{t} \right) \right] \right),
$$

where X_μ, X_ν are defined according to (1.5), and

$$\xi^*(s) := \sup_{r \geqslant 0} \left(rs - \xi(r) \right).$$

We can rewrite this formula as

$$
(6.4) \qquad f_1(t, \mu) = \sup_{\nu \in \mathcal{P}_2(\mathbb{R}_+)} \left(\psi_1(\nu) - \mathbb{E}\left[\sup_{r \geqslant 0} \{ r(X_\nu - X_\mu) - t\xi(r) \} \right] \right)
$$

$$
(6.5) \qquad = \sup_{\nu \in \mathcal{P}_2(\mathbb{R}_+)} \inf_{f \in L^2([0,1];\mathbb{R}_+)} \left(\psi_1(\nu) - \mathbb{E}\left[f(U)(X_\nu - X_\mu) - t\xi(f(U)) \right] \right).
$$

One may wonder whether supremum and infimum can be interchanged in the expression above. If this were the case, it would imply in particular that

$$\psi_1(\mu) = \inf_{f \in L^2([0,1];\mathbb{R}_+)} \sup_{\nu \in \mathcal{P}_2(\mathbb{R}_+)} \left(\psi_1(\nu) - \mathbb{E}\left[f(U)(X_\nu - X_\mu) \right] \right).$$

But notice that the supremum over ν above is an affine function of X_μ; taking the infimum, we find that this would imply the transport concavity of ψ_1. But we have argued above that this is not so in general. Similarly, replacing $\inf_f \sup_\nu$ by $\sup_f \inf_\nu$ in the expression above would lead to the conclusion that ψ_1 is transport-convex, which has also been excluded in general. Conversely, if ψ_1 were actually transport-concave, then interchanging the supremum and the infimum in (6.5) would be valid; and in general, what we find after the interchange is the solution to the same Hamilton-Jacobi equation, but with the initial condition replaced by its transport-concave envelope.

We now come back to the bipartite model investigated in the present paper. The considerations above raise the question of whether the solution to (1.7) can be written as a saddle-point, with respect to the variables $f = (f_1, f_2) \in (L^2([0,1];\mathbb{R}_+))^2$ and $\nu = (\nu_1, \nu_2) \in (\mathcal{P}_2(\mathbb{R}_+))^2$, of the functional

$$\psi(\nu) - \mathbb{E}\left[\sum_{a=1}^2 f_a(U)(X_{\nu_a} - X_{\mu_a}) - t f_1(U) f_2(U) \right].$$

But any possible arrangement of inf's and sup's leads to a contradiction. If we aim for optimizing first over f and then over ν, in analogy with (6.5), then this amounts to trying to write down a Hopf-Lax formula although the nonlinearity in the equation is neither convex nor concave. The convex (or concave) dual of the mapping $(x, y) \mapsto xy$ is so degenerate that it is easy to rule out this possibility. On the other hand, if we try to optimize first over ν and then over f, then we face the same situation as above: each possibity would imply either that ψ is transport-convex, or that it is transport-concave, and both have been ruled out in general.

6.2. **A related attempt.** A related attempt at generating a candidate variational formula for the limit free energy of the bipartite model is as follows. In the papers [9, 45], the authors investigate a large class of models covering in particular the situation in which the definition of $H_N(\sigma)$ in (1.1) is replaced by

$$N^{-\frac{1}{2}} \sum_{i,j=1}^N J_{ij}\, \sigma_{1,i}\, \sigma_{2,j} + N^{-\frac{1}{2}} \sum_{a \in \{1,2\}} \sum_{i,j=1}^N J_{ij}^{(a)}\, \sigma_{a,i}\, \sigma_{a,j},$$

where $(J_{ij}^{(a)})_{a \in \{1,2\}, 1 \leqslant i,j \leqslant N}$ are independent centered Gaussian random variables with a fixed variance, independent of (J_{ij}). Assuming that the matrix

$$A := \begin{pmatrix} 2\mathbb{E}[(J_{11}^{(1)})^2] & \mathbb{E}[(J_{11})^2] \\ \mathbb{E}[(J_{11})^2] & 2\mathbb{E}[(J_{11}^{(2)})^2] \end{pmatrix}$$

is positive definite, they derive a variational formula for the free energy of the model.

One may wonder whether the formula obtained by ignoring the assumption of positive definiteness of the matrix A necessary for their proofs actually matches the prediction given by the Hamilton-Jacobi equation. We will argue here that this is not so. We have already seen in the previous subsection that writing up a naive Hopf-Lax formula for (1.7) would clearly lead to an invalid prediction. However, the formula given in [9, 45], while equivalent to the Hopf-Lax formula in the case when the matrix A is positive definite, is

actually different in outlook, and extends to a different expression in the setting when the matrix A is taken to be the matrix of interest to us here, namely

$$(6.6) \qquad \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Assuming that the matrix A is positive definite, we first explain the derivation of the formula in [9, 45] starting from the point of view provided by Hamilton-Jacobi equations. One can check (at least formally, and probably rigorously by combining the arguments of [9, 45] with those of [41]) that the relevant Hamilton-Jacobi equation for this model is given by

$$\partial_t f - \frac{1}{2} \int \partial_\mu f \cdot A \partial_\mu f \, \mathrm{d}\widehat{\mu},$$

with the same initial condition ψ defined in (2.19) (that is, ψ does not depend on A), and where we used the vector notation $\partial_\mu f \coloneqq (\partial_{\mu_1} f, \partial_{\mu_2} f)$. Since we assume that A is positive definite, we can write down the Hopf-Lax formula

$$f(t, \mu) = \sup_{\nu \in (\mathcal{P}_2(\mathbb{R}_+))^2} \left(\psi(\nu) - \frac{1}{2t} \mathbb{E}\left[(X_\nu - X_\mu) \cdot A^{-1} (X_\nu - X_\mu) \right] \right),$$

with the notation $X_\nu = (X_{\nu_1}, X_{\nu_2})$. From here, we could replace A by the matrix in (6.6), but it is easy to see that with this choice the supremum is infinite. However the formula of [9, 45] (for positive definite A) has an additional restriction on the support of the pair of measures ν. Under the assumption that A is positive definite, we can indeed write

$$(6.7) \qquad f(t, (\delta_0, \delta_0)) = \sup_\nu \left(\psi(\nu) - \frac{1}{2t} \mathbb{E}\left[X_\nu \cdot A^{-1} X_\nu \right] \right),$$

where the supremum is taken over every pair of measures $\nu = (\nu_a)_{a \in \{1,2\}} \in (\mathcal{P}(\mathbb{R}_+))^2$ with the restriction that, denoting by q_a the top of the support of ν_a,

$$A^{-1} \begin{pmatrix} q_1 \\ q_2 \end{pmatrix} \leqslant \begin{pmatrix} 2t \\ 2t \end{pmatrix},$$

in the sense that the inequality holds component by component. That this additional restriction does not change the value of the supremum in (6.7) in the case when A is positive definite can be derived from the Lipschitz estimate on ψ guaranteed by Proposition 2.1, and arguing as in [40, Step 4]. Blindly replacing the matrix A by that in (6.6) thus leads to the formula

$$(6.8) \qquad \sup_{\nu \in \mathcal{P}([0.2t])^2} \left(\psi(\nu) - \frac{1}{t} \mathbb{E}\left[X_{\nu_1} X_{\nu_2} \right] \right).$$

One can verify that this formula does not match the solution to the equation (1.7) evaluated at $\mu = (\delta_0, \delta_0)$. For instance, recalling the notation in (1.3), we may take π_1 to be the uniform measure on $\{-1, 1\}$, π_2 to be a non-uniform measure on $\{-1, 1\}$, and verify that in this case the solution to the equation satisfies $\partial_t f(0, (\delta_0, \delta_0)) = 0$, but that this property is not satisfied by the expression in (6.8). Assuming that the overlaps are concentrated for small t, the limit free energy should be described by the equation (1.12) in this region, and this would imply that indeed $\partial_t f(0, (\delta_0, \delta_0)) = 0$.

Appendix A. Gaussian integrals

A.1. **Gaussian integration by parts.** Let μ be the law of a d-dimensional centered Gaussian vector, with covariance matrix $\mathsf{C} \in \mathbb{R}^{d \times d}$. We assume (temporarily) that C is

invertible. In this case, the measure μ has a density with respect to the Lebesgue measure on $\mathbb{R}^d$, which is proportional to

$$\exp\left(-\frac{1}{2}x \cdot \mathsf{C}^{-1}x\right).$$

For every bounded and smooth function $F \in C^\infty(\mathbb{R}^d; \mathbb{R}^d)$, we thus have, by integration by parts,

$$\int \mathsf{C}^{-1}x \cdot F(x)\,\mathrm{d}\mu(x) = \int \nabla \cdot F(x)\,\mathrm{d}\mu(x),$$

or equivalently,

$$\int x \cdot F(x)\,\mathrm{d}\mu(x) = \int \nabla \cdot (\mathsf{C}F)(x)\,\mathrm{d}\mu(x).$$

This last identity remains valid when C is not invertible, by approximation. In particular, for every bounded and smooth $f \in C^\infty(\mathbb{R}^d; \mathbb{R})$,

(A.1) $$\int x_1 f(x)\,\mathrm{d}\mu(x) = \sum_{\ell=1}^d \mathbb{E}\left[x_1 x_\ell\right]\int \partial_{x_\ell} f(x)\,\mathrm{d}\mu(x).$$

One consequence of this observation is the following result. (At least the first part of it is very classical; the last part is certainly also well-known, but I could not find a precise reference).

Lemma A.1. *Let Σ be a finite set, let $(x_1(\sigma), x_2(\sigma))_{\sigma\in\Sigma}$ be a centered Gaussian random field with respect to the probability measure $\mathbb{P}$ (with expectation $\mathbb{E}$), and let P be a probability measure on Σ. For every $a, b \in \{1, 2\}$ and $\sigma, \sigma' \in \Sigma$, we write*

$$\mathsf{C}_{ab}(\sigma, \sigma') := \mathbb{E}\left[x_a(\sigma)x_b(\sigma')\right].$$

We denote by $\langle\cdot\rangle$ the Gibbs measure built from $(x_2(\sigma))$, so that for every $f : \Sigma \to \mathbb{R}$,

$$\langle f(\sigma)\rangle := \frac{\int f(\sigma)\exp\left(x_2(\sigma)\right)\,\mathrm{d}P(\sigma)}{\int \exp\left(x_2(\sigma)\right)\,\mathrm{d}P(\sigma)},$$

and write σ', σ'' for independent copies of the random variable σ under $\langle\cdot\rangle$. We have

(A.2) $$\mathbb{E}\langle x_1(\sigma)\rangle = \mathbb{E}\left\langle\mathsf{C}_{12}(\sigma, \sigma) - \mathsf{C}_{12}(\sigma, \sigma')\right\rangle,$$

and

(A.3) $$\mathbb{E}\left\langle x_1^2(\sigma)\right\rangle = \mathbb{E}\left\langle\mathsf{C}_{11}(\sigma, \sigma)\right\rangle + \mathbb{E}\Big\langle\left(\mathsf{C}_{12}(\sigma, \sigma) - \mathsf{C}_{12}(\sigma, \sigma')\right)$$
$$\left(\mathsf{C}_{12}(\sigma, \sigma) + \mathsf{C}_{12}(\sigma, \sigma') - 2\mathsf{C}_{12}(\sigma, \sigma'')\right)\Big\rangle.$$

More generally, we write $(\sigma^\ell)_{\ell\geqslant 1}$ for a sequence of independent copies of the random variable σ under $\langle\cdot\rangle$. For every $p \geqslant 1$, there exists a polynomial P_p (which does not depend on any parameter in the problem) taking as inputs the variables $(\mathsf{C}_{ab}(\sigma^k, \sigma^\ell))_{a,b\in\{1,2\},k,\ell\geqslant 1}$ such that

(A.4) $$\mathbb{E}\left\langle x_1^p(\sigma)\right\rangle = \mathbb{E}\left\langle P_p\left(\left(\mathsf{C}_{ab}(\sigma^k, \sigma^\ell)\right)_{a,b\in\{1,2\},k,\ell\geqslant 1}\right)\right\rangle.$$

Moreover, the polynomial only depends on $(\mathsf{C}_{11}(\sigma^k, \sigma^\ell))_{k,\ell\geqslant 1}$ and $(\mathsf{C}_{12}(\sigma^k, \sigma^\ell))_{k,\ell\geqslant 1}$, and is homogeneous of degree p provided that we count each occurrence of a variable $\mathsf{C}_{11}(\sigma^k, \sigma^\ell)$ as having degree 2.

Proof. We start by writing

$$\mathbb{E}\langle x_1(\sigma)\rangle = \int \mathbb{E}\left[x_1(\sigma)\frac{\exp(x_2(\sigma))}{\int \exp(x_2(\sigma'))\,\mathrm{d}P(\sigma')}\right]\mathrm{d}P(\sigma).$$

We then apply (A.1) to rewrite the inner expectation as

$$\mathsf{C}_{12}(\sigma,\sigma)\mathbb{E}\left[\frac{\exp\left(x_2(\sigma)\right)}{\int\exp(x_2(\sigma'))\,\mathrm{d}P(\sigma')}\right]-\int\mathsf{C}_{12}(\sigma,\sigma')\mathbb{E}\left[\frac{\exp(x_2(\sigma)+x_2(\sigma'))}{\left(\int\exp(x_2(\sigma''))\,\mathrm{d}P(\sigma'')\right)^2}\right]\mathrm{d}P(\sigma').$$

Combining the two previous displays leads to (A.2). The argument for (A.3) is similar, except that we now need to compute

$$(\text{A.5})\qquad\qquad\mathbb{E}\left[(x_1(\sigma))^2\frac{\exp(x_2(\sigma))}{\int\exp(x_2(\sigma'))\,\mathrm{d}P(\sigma')}\right].$$

In order to apply (A.1) in this case, we split the square of $x_1(\sigma)$ into two parts, one of them being incorporated into the function "f" in (A.1). We thus find that the quantity in (A.5) equals

$$\mathsf{C}_{11}(\sigma,\sigma)\mathbb{E}\left[\frac{\exp\left(x_2(\sigma)\right)}{\int\exp(x_2(\sigma'))\,\mathrm{d}P(\sigma')}\right]+\mathsf{C}_{12}(\sigma,\sigma)\mathbb{E}\left[\frac{x_1(\sigma)\exp\left(x_2(\sigma)\right)}{\int\exp(x_2(\sigma'))\,\mathrm{d}P(\sigma')}\right]$$
$$-\int\mathsf{C}_{12}(\sigma,\sigma')\mathbb{E}\left[\frac{x_1(\sigma)\exp(x_2(\sigma)+x_2(\sigma'))}{\left(\int\exp(x_2(\sigma''))\,\mathrm{d}P(\sigma'')\right)^2}\right]\mathrm{d}P(\sigma').$$

This shows that

$$\mathbb{E}\left\langle(x_1(\sigma))^2\right\rangle=\mathbb{E}\left\langle\mathsf{C}_{11}(\sigma,\sigma)\right\rangle+\mathbb{E}\left\langle x_1(\sigma)\left(\mathsf{C}_{12}(\sigma,\sigma)-\mathsf{C}_{12}(\sigma,\sigma')\right)\right\rangle.$$

For every $\sigma,\sigma'\in\Sigma$, we define

$$\widetilde{x}_1(\sigma,\sigma'):=x_1(\sigma)\left(\mathsf{C}_{12}(\sigma,\sigma)-\mathsf{C}_{12}(\sigma,\sigma')\right).$$

The variables $(\widetilde{x}_1(\sigma,\sigma'),x_2(\sigma)+x_2(\sigma'))_{\sigma,\sigma'\in\Sigma}$ form a centered Gaussian field, with

$$\mathbb{E}\left[\widetilde{x}_1(\sigma,\sigma')(x_2(\sigma'')+x_2(\sigma'''))\right]=\left(\mathsf{C}_{12}(\sigma,\sigma)-\mathsf{C}_{12}(\sigma,\sigma')\right)\left(\mathsf{C}_{12}(\sigma,\sigma'')+\mathsf{C}_{12}(\sigma,\sigma''')\right).$$

Applying (A.2), we deduce that

$$\mathbb{E}\left\langle x_1(\sigma)\left(\mathsf{C}_{12}(\sigma,\sigma)-\mathsf{C}_{12}(\sigma,\sigma')\right)\right\rangle=\mathbb{E}\big\langle\left(\mathsf{C}_{12}(\sigma,\sigma)-\mathsf{C}_{12}(\sigma,\sigma')\right)$$
$$\left(\mathsf{C}_{12}(\sigma,\sigma)+\mathsf{C}_{12}(\sigma,\sigma')-\mathsf{C}_{12}(\sigma,\sigma'')-\mathsf{C}_{12}(\sigma,\sigma''')\right)\big\rangle,$$

and replacing σ''' by σ'' in the expression above does not change its value. This completes the proof of (A.3). For (A.4), we apply (A.1) again to rewrite

$$\mathbb{E}\left[x_1^p(\sigma)\frac{\exp(x_2(\sigma))}{\int\exp(x_2(\sigma'))\,\mathrm{d}P(\sigma')}\right]$$

as

$$\mathsf{C}_{11}(\sigma,\sigma)(p-1)\mathbb{E}\left[\frac{x_1^{p-2}(\sigma)\exp\left(x_2(\sigma)\right)}{\int\exp(x_2(\sigma'))\,\mathrm{d}P(\sigma')}\right]+\mathsf{C}_{12}(\sigma,\sigma)\mathbb{E}\left[\frac{x_1^{p-1}(\sigma)\exp\left(x_2(\sigma)\right)}{\int\exp(x_2(\sigma'))\,\mathrm{d}P(\sigma')}\right]$$
$$-\int\mathsf{C}_{12}(\sigma,\sigma')\mathbb{E}\left[\frac{x_1^{p-1}(\sigma)\exp(x_2(\sigma)+x_2(\sigma'))}{\left(\int\exp(x_2(\sigma''))\,\mathrm{d}P(\sigma'')\right)^2}\right]\mathrm{d}P(\sigma').$$

We can then obtain (A.4) by induction on p. $\qquad\qquad\qquad\square$

A.2. **Existence of Gaussian process.** The next lemma serves to guarantee that the Gaussian random field introduced in (4.4) indeed exists.

Lemma A.2. *Let $p, k, N \geqslant 1$ be integers, and $\lambda_1, \lambda_2 \geqslant 0$. There exists a centered Gaussian field $(X(\sigma, \alpha))_{\sigma \in \mathbb{R}^N, \alpha \in \mathbb{N}^k}$ such that, for every $\sigma, \sigma' \in \mathbb{R}^N$ and $\alpha, \alpha' \in \mathbb{N}^k$,*

$$(A.6) \qquad \mathbb{E}\left[X(\sigma, \alpha)X(\sigma', \alpha')\right] = \left(\lambda_1\, \sigma \cdot \sigma' + \lambda_2\, \alpha \wedge \alpha'\right)^p,$$

where we recall that the notation $\alpha \wedge \alpha'$ was introduced in (2.9).

Proof. Recall the definition of the tree $\mathcal{A}$ in (2.6). For each $n \in \mathbb{N}$, we define the finite approximation

$$\mathcal{A}_n := \{0, \ldots, n\}^0 \cup \cdots \cup \{0, \ldots, n\}^k,$$

again with the understanding that $\{0, \ldots, n\}^0 = \{\varnothing\}$, and we denote the set of leaves by $\mathcal{L}_n := \{0, \ldots, n\}^k$. By Kolmogorov's extension theorem, it suffices to construct a Gaussian process $(X_n(\sigma, \alpha))_{\sigma \in \mathbb{R}^N, \alpha \in \mathcal{L}_n}$ such that (A.6) holds for every $\sigma, \sigma' \in \mathbb{R}^N$ and $\alpha, \alpha' \in \mathcal{L}_n$. Let $(f_\alpha)_{\alpha \in \mathcal{A}_n}$ be an orthonormal basis of $\mathbb{R}^{|\mathcal{A}_n|}$, and for each $\alpha \in \mathcal{L}_n$, let

$$g_\alpha := \sum_{\ell=1}^{k} f_{\alpha_{|\ell}},$$

so that for every $\alpha, \alpha' \in \mathcal{L}_n$,

$$(A.7) \qquad g_\alpha \cdot g_{\alpha'} = \alpha \wedge \alpha'.$$

Viewing $(\sqrt{\lambda_1}\sigma, \sqrt{\lambda_2}g_\alpha)$ as a vector in $\mathbb{R}^N \times \mathbb{R}^{|\mathcal{A}_n|}$, we consider the p-fold tensor product

$$\left(\sqrt{\lambda_1}\sigma, \sqrt{\lambda_2}g_\alpha\right)^{\otimes p} \in \left(\mathbb{R}^N \times \mathbb{R}^{|\mathcal{A}_n|}\right)^{\otimes p}.$$

Recall that, if we denote by $(e_i)_{i \in \{1, \ldots, N\}}$ an orthonormal basis of $\mathbb{R}^N$, then an orthonormal basis of the tensor product $(\mathbb{R}^N \times \mathbb{R}^{|\mathcal{A}_n|})^{\otimes p}$ is given by

$$\mathcal{B} := \left\{v_1 \otimes \cdots \otimes v_p \; : \; v_1, \ldots v_p \in \{e_i, i \in \{1, \ldots, N\}\} \cup \{f_\alpha, \alpha \in \mathcal{A}_n\}\right\}.$$

We now give ourselves a standard Gaussian vector W taking values in $\mathbb{R}^{|\mathcal{B}|}$, and define

$$X(\sigma, \alpha) := W \cdot \left(\sqrt{\lambda_1}\sigma, \sqrt{\lambda_2}e'_\alpha\right)^{\otimes p},$$

so that for every $\sigma, \sigma' \in \mathbb{R}^N$ and $\alpha, \alpha' \in \mathcal{L}_n$,

$$\begin{aligned}
\mathbb{E}\left[X(\sigma, \alpha)X(\sigma', \alpha')\right] &= \left(\sqrt{\lambda_1}\sigma, \sqrt{\lambda_2}g_\alpha\right)^{\otimes p} \cdot \left(\sqrt{\lambda_1}\sigma', \sqrt{\lambda_2}g_{\alpha'}\right)^{\otimes p} \\
&= \left(\left(\sqrt{\lambda_1}\sigma, \sqrt{\lambda_2}g_\alpha\right) \cdot \left(\sqrt{\lambda_1}\sigma', \sqrt{\lambda_2}g_{\alpha'}\right)\right)^p \\
&= \left(\lambda_1\, \sigma \cdot \sigma' + \lambda_2\, \alpha \wedge \alpha'\right)^p,
\end{aligned}$$

where we used (A.7) in the last step. $\qquad \square$

REFERENCES

[1] E. Agliari, A. Barra, R. Burioni, and A. Di Biasio. Notes on the p-spin glass studied via Hamilton-Jacobi and smooth-cavity techniques. *J. Math. Phys.*, 53(6):063304, 29, 2012.

[2] L. Ambrosio and J. Feng. On a class of first order Hamilton-Jacobi equations in metric spaces. *J. Differential Equations*, 256(7):2194–2245, 2014.

[3] A. Auffinger and W.-K. Chen. Free energy and complexity of spherical bipartite models. *J. Stat. Phys.*, 157(1):40–59, 2014.

[4] A. Auffinger and W.-K. Chen. The Parisi formula has a unique minimizer. *Comm. Math. Phys.*, 335(3):1429–1444, 2015.

[5] J. Baik and J. O. Lee. Free energy of bipartite spherical Sherrington–Kirkpatrick model. Preprint, arXiv:1711.06364.

[6] 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.

[7] M. Bardi and L. C. Evans. On Hopf's formulas for solutions of Hamilton-Jacobi equations. *Nonlinear Anal.*, 8(11):1373–1381, 1984.

[8] 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*, volume 2074 of *Lecture Notes in Math.*, pages 49–109. Springer, Heidelberg, 2013.

[9] A. Barra, P. Contucci, E. Mingione, and D. Tantari. Multi-species mean field spin glasses. Rigorous results. *Ann. Henri Poincaré*, 16(3):691–708, 2015.

[10] A. Barra, G. Del Ferraro, and D. Tantari. Mean field spin glasses treated with PDE techniques. *Eur. Phys. J. B*, 86(7):Art. 332, 10, 2013.

[11] A. Barra, A. Di Biasio, and F. Guerra. Replica symmetry breaking in mean-field spin glasses through the Hamilton-Jacobi technique. *J. Stat. Mech. Theory Exp.*, (9):P09006, 22, 2010.

[12] A. Barra, G. Genovese, and F. Guerra. Equilibrium statistical mechanics of bipartite spin systems. *J. Phys. A*, 44(24):245002, 22, 2011.

[13] A. Bovier and A. Klimovsky. The Aizenman-Sims-Starr and Guerra's schemes for the SK model with multidimensional spins. *Electron. J. Probab.*, 14:no. 8, 161–241, 2009.

[14] P. Cardaliaguet. Notes on mean field games. Technical report, 2010.

[15] P. Cardaliaguet and M. Quincampoix. Deterministic differential games under probability knowledge of initial condition. *Int. Game Theory Rev.*, 10(1):1–16, 2008.

[16] P. Cardaliaguet and A. Souquière. A differential game with a blind player. *SIAM J. Control Optim.*, 50(4):2090–2116, 2012.

[17] H.-B. Chen. Hamilton-Jacobi equations for nonsymmetric matrix inference. Preprint, arXiv:2006.05328.

[18] H.-B. Chen and J. Xia. Hamilton-Jacobi equations for inference of matrix tensor products. Preprint, arXiv:2009.01678.

[19] M. G. Crandall, H. Ishii, and P.-L. Lions. User's guide to viscosity solutions of second order partial differential equations. *Bull. Amer. Math. Soc. (N.S.)*, 27(1):1–67, 1992.

[20] M. G. Crandall and P.-L. Lions. Hamilton-Jacobi equations in infinite dimensions. I. Uniqueness of viscosity solutions. *J. Funct. Anal.*, 62(3):379–396, 1985.

[21] M. G. Crandall and P.-L. Lions. Hamilton-Jacobi equations in infinite dimensions. II. Existence of viscosity solutions. *J. Funct. Anal.*, 65(3):368–405, 1986.

[22] M. G. Crandall and P.-L. Lions. Hamilton-Jacobi equations in infinite dimensions. III. *J. Funct. Anal.*, 68(2):214–247, 1986.

[23] P. Dupuis and H. Ishii. On oblique derivative problems for fully nonlinear second-order elliptic PDEs on domains with corners. *Hokkaido Math. J.*, 20(1):135–164, 1991.

[24] L. C. Evans. *Partial differential equations*, volume 19 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, second edition, 2010.

[25] J. Feng and M. Katsoulakis. A comparison principle for Hamilton-Jacobi equations related to controlled gradient flows in infinite dimensions. *Arch. Ration. Mech. Anal.*, 192(2):275–310, 2009.

[26] J. Feng and T. G. Kurtz. *Large deviations for stochastic processes*, volume 131 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2006.

[27] Y. V. Fyodorov, I. Y. Korenblit, and E. Shender. Antiferromagnetic ising spin glass. *Journal of Physics C: Solid State Physics*, 20(12):1835, 1987.

[28] Y. V. Fyodorov, I. Y. Korenblit, and E. Shender. Phase transitions in frustrated metamagnets. *EPL (Europhysics Letters)*, 4(7):827, 1987.

[29] W. Gangbo, T. Nguyen, and A. Tudorascu. Hamilton-Jacobi equations in the Wasserstein space. *Methods Appl. Anal.*, 15(2):155–183, 2008.

[30] W. Gangbo and A. Święch. Optimal transport and large number of particles. *Discrete Contin. Dyn. Syst.*, 34(4):1397–1441, 2014.

[31] F. Guerra. Sum rules for the free energy in the mean field spin glass model. *Fields Institute Communications*, 30:161, 2001.

[32] F. Guerra. Broken replica symmetry bounds in the mean field spin glass model. *Comm. Math. Phys.*, 233(1):1–12, 2003.

[33] E. Hopf. Generalized solutions of non-linear equations of first order. *J. Math. Mech.*, 14:951–973, 1965.

[34] I. Y. Korenblit and E. Shender. Spin glass in an lsing two-sublattice magnet. *Zh. Eksp. Teor. Fiz*, 89:1785–1795, 1985.

[35] P.-L. Lions and J.-C. Rochet. Hopf formula and multitime Hamilton-Jacobi equations. *Proc. Amer. Math. Soc.*, 96(1):79–84, 1986.

[36] M. Mézard, G. Parisi, and M. Virasoro. *Spin glass theory and beyond: an introduction to the replica method and its applications*, volume 9. World Scientific Publishing Company, 1987.

[37] L. Miolane. Fundamental limits of low-rank matrix estimation: the non-symmetric case. Preprint, arXiv:1702.00473.

[38] J.-C. Mourrat. Hamilton-Jacobi equations for mean-field disordered systems. *Ann. Henri Lebesgue*, to appear.

[39] J.-C. Mourrat. Hamilton-Jacobi equations for finite-rank matrix inference. *Ann. Appl. Probab.*, to appear.

[40] J.-C. Mourrat. Parisi's formula is a Hamilton-Jacobi equation in Wasserstein space. Preprint, arXiv:1906.08471.

[41] J.-C. Mourrat and D. Panchenko. Extending the Parisi formula along a Hamilton-Jacobi equation. *Electron. J. Probab.*, 25:Paper No. 23, 17, 2020.

[42] D. Panchenko. Free energy in the generalized Sherrington-Kirkpatrick mean field model. *Rev. Math. Phys.*, 17(7):793–857, 2005.

[43] D. Panchenko. The Parisi ultrametricity conjecture. *Ann. of Math. (2)*, 177(1):383–393, 2013.

[44] D. Panchenko. *The Sherrington-Kirkpatrick model*. Springer Monographs in Mathematics. Springer, New York, 2013.

[45] D. Panchenko. The free energy in a multi-species Sherrington-Kirkpatrick model. *Ann. Probab.*, 43(6):3494–3513, 2015.

[46] D. Panchenko. Free energy in the Potts spin glass. *Ann. Probab.*, 46(2):829–864, 2018.

[47] D. Panchenko. Free energy in the mixed *p*-spin models with vector spins. *Ann. Probab.*, 46(2):865–896, 2018.

[48] G. Parisi. Infinite number of order parameters for spin-glasses. *Phys. Rev. Lett.*, 43(23):1754, 1979.

[49] G. Parisi. A sequence of approximated solutions to the SK model for spin glasses. *J. Phys. A*, 13(4):L115–L121, 1980.

[50] G. Reeves. Information-theoretic limits for the matrix tensor product, Preprint, arXiv:2005.11273.

[51] I. J. Schoenberg. Positive definite functions on spheres. *Duke Math. J.*, 9:96–108, 1942.

[52] M. Talagrand. Free energy of the spherical mean field model. *Probab. Theory Related Fields*, 134(3):339–382, 2006.

[53] M. Talagrand. The Parisi formula. *Ann. of Math. (2)*, 163(1):221–263, 2006.

[54] M. Talagrand. Mean field models for spin glasses: some obnoxious problems. In *Spin glasses*, volume 1900 of *Lecture Notes in Math.*, pages 63–80. Springer, Berlin, 2007.

[55] M. Talagrand. A general form of certain mean field models for spin glasses. *Probab. Theory Related Fields*, 143(1-2):97–111, 2009.

[56] M. Talagrand. *Mean field models for spin glasses. Volume I*, volume 54 of *Ergebnisse der Mathematik und ihrer Grenzgebiete*. Springer-Verlag, Berlin, 2011.

[57] M. Talagrand. *Mean field models for spin glasses. Volume II*, volume 55 of *Ergebnisse der Mathematik und ihrer Grenzgebiete*. Springer, Heidelberg, 2011.

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

E-mail address: jcm777@nyu.edu