THE CONVEX STRUCTURE OF THE PARISI FORMULA FOR MULTI-SPECIES SPIN GLASSES

HONG-BIN CHEN ⬤, VICTOR ISSA ⬤, AND JEAN-CHRISTOPHE MOURRAT ⬤

ABSTRACT. We study the free energy of mean-field multi-species spin glasses with convex covariance function. For such models with D species, the Parisi formula is known to be valid, and expresses the limit free energy as a supremum over monotone probability measures on $\mathbb{R}_+^D$. We show here that one can transform this representation into a supremum over *all* probability measures on $\mathbb{R}_+^D$ of a *concave* functional. We then deduce that the Parisi formula admits a unique maximizer. Using convex-duality arguments, we also obtain a new representation of the free energy as an infimum over martingales in a Wiener space.

KEYWORDS AND PHRASES: MEAN-FIELD SPIN GLASS, PARISI FORMULA, HAMILTON–JACOBI EQUATION

MSC 2020: 82D30, 82B44, 35D40

CONTENTS

(Hong-Bin Chen) INSTITUT DES HAUTES ÉTUDES SCIENTIFIQUES, BURES-SUR-YVETTE, FRANCE

(Victor Issa) DEPARTMENT OF MATHEMATICS, ENS LYON, LYON, FRANCE

(Jean-Christophe Mourrat) DEPARTMENT OF MATHEMATICS, ENS LYON AND CNRS, LYON, FRANCE

E-mail addresses: hbchen@ihes.fr, victor.issa@ens-lyon.fr, jean-christophe.mourrat@ens-lyon.fr.

1. Introduction

We study mean-field spin-glass models with multiple species. Fixing an integer $D \geqslant 1$, we write $[D] = \{1, \ldots, D\}$ to denote the set of different species of spins. For each $N \in \mathbb{N}$, we give ourselves $(I_{N,d})_{d \in [D]}$ a partition of $\{1, \ldots, N\}$, and interpret each $I_{N,d}$ as the set of indices that belong to the d-th species. For any two configurations $\sigma, \sigma' \in \mathbb{R}^N$ and $d \in [D]$, we define the overlap for the d-th species as

$$(1.1) \qquad R_{N,d}(\sigma, \sigma') = \frac{1}{N} \sum_{n \in I_{N,d}} \sigma_n \sigma'_n,$$

and we set

$$(1.2) \qquad R_N(\sigma, \sigma') = (R_{N,d}(\sigma, \sigma'))_{d \in [D]}.$$

We give ourselves a function $\xi : \mathbb{R}^D \to \mathbb{R}$ and a centered Gaussian field $(H_N(\sigma))_{\sigma \in [-1,+1]^N}$ with covariance

$$(1.3) \qquad \mathbb{E}\left[H_N(\sigma)H_N(\sigma')\right] = N\xi\left(R_N(\sigma, \sigma')\right).$$

Throughout, we assume that the function ξ admits an absolutely convergent power-series expansion and satisfies $\xi(0) = 0$, and we make the key assumption that *the function ξ is convex on $\mathbb{R}_+^D$*. The definitions of the free energy and Gibbs measure of the spin glass require that we also choose a reference probability measure. For each species $d \in [D]$, we fix a probability measure π_d with support in $[-1, 1]$ that is not a Dirac mass, and let

$$(1.4) \qquad \mathrm{d}P_N(\sigma) = \bigotimes_{d \in [D]} \bigotimes_{n \in I_{N,d}} \mathrm{d}\pi_d(\sigma_n).$$

In words, when $\sigma = (\sigma_1, \ldots, \sigma_N)$ is sampled according to P_N, the coordinates of σ are independent, and the law of σ_n is π_d if n belongs to $I_{N,d}$. For each $d \in [D]$, the proportion of spins that belong to the d-th species is given by

$$(1.5) \qquad \lambda_{N,d} = |I_{N,d}|/N, \quad \text{and we set } \lambda_N = (\lambda_{N,d})_{d \in [D]}.$$

We assume throughout that for some $\lambda_\infty = (\lambda_{\infty,d})_{d \in [D]} \in (0, 1)^D$, we have

$$(1.6) \qquad \lim_{N \to \infty} \lambda_N = \lambda_\infty.$$

The main object of study of this paper is the limiting value as $N \to +\infty$ of the free energy defined for every $t \geqslant 0$ by

$$(1.7) \quad \overline{F}_N(t, \delta_0) = -\frac{1}{N}\mathbb{E}\log \int \exp\left(\sqrt{2t}H_N(\sigma) - Nt\xi\left(R_N(\sigma, \sigma)\right)\right)\mathrm{d}P_N(\sigma).$$

The term $-Nt\xi\left(R_N(\sigma, \sigma)\right)$ is introduced as a convenience to simplify the expression of the limit. One possible way to remove it a posteriori is described in [43]; we also point out that this term is constant if each π_d is supported in $\{-1, +1\}$. In (1.7), the second argument of $\overline{F}_N$ is a Dirac mass at $0 \in \mathbb{R}^D$, and we will extend this function to more general arguments below.

In order to state the Parisi formula for the limit of (1.7), we introduce further notation. We write $\mathbb{R}_+ = [0, +\infty)$, and let $\mathcal{P}(\mathbb{R}_+^D)$ denote the set of probability measures on $\mathbb{R}_+^D$. We say that $\mu \in \mathcal{P}(\mathbb{R}_+^D)$ is *monotone* when there exists an increasing map $\mathsf{q} : [0, 1) \to \mathbb{R}_+^D$ such that $\mu = \text{Law}(\mathsf{q}(U))$ with U a uniform random variable in $[0, 1]$. Here and throughout, the word "increasing" (and likewise for "decreasing") is understood in the sense of non-strict inequalities, that is, we are asking here that $\mathsf{q}(t) - \mathsf{q}(s) \in \mathbb{R}_+^D$ for every $s \leqslant t \in [0, 1)$. We denote by $\mathcal{P}^\uparrow(\mathbb{R}_+^D)$ the collection of monotone probability measures on $\mathbb{R}_+^D$. For every $p \geqslant 1$, we denote by $\mathcal{P}_p(\mathbb{R}_+^D)$ the set of probability measures with finite p-th moment, by $\mathcal{P}_\infty(\mathbb{R}_+^D)$ the set of compactly supported probability measures on $\mathbb{R}_+^D$, and we set $\mathcal{P}_p^\uparrow(\mathbb{R}_+^D) = \mathcal{P}^\uparrow(\mathbb{R}_+^D) \cap \mathcal{P}_p(\mathbb{R}_+^D)$.

For all the models considered here, the Parisi formula is known to be valid [4, 9, 13, 27, 46, 47, 60]. We give two versions of the Parisi formula, the first one in (1.11) being the most classical, and the second one in (1.12) being the one which will best clarify its relationship with the sequel. For each $\nu \in \mathcal{P}(\mathbb{R}_+^D)$ and $d \in [D]$, we denote by ν_d the d-th marginal of ν; in other words, ν_d is the image of ν through the mapping $x \mapsto x_d$. The first term in the Parisi formula is given, for each $\nu \in \mathcal{P}_1^\uparrow(\mathbb{R}_+^D)$, by

$$(1.8) \qquad \psi(\nu) = \sum_{d=1}^{D} \lambda_{\infty,d} \psi_d(\nu_d),$$

where for each $d \in [D]$, the function ψ_d is the cascade transform of the reference measure $\boldsymbol{\pi}_d$ defined precisely in (2.16) (see also Lemma 3.1 for an alternative characterization). It follows from [2] that the functions $\psi_1, \ldots, \psi_D$ are each concave over $\mathcal{P}_1(\mathbb{R}_+)$. The formula (1.8) therefore allows us to extend ψ into a concave (but not strictly concave) function on all of $\mathcal{P}_1(\mathbb{R}_+^D)$. For every $y \in \mathbb{R}^D$, we set

$$(1.9) \qquad \theta(y) = y \cdot \nabla \xi(y) - \xi(y),$$

$$(1.10) \qquad \xi^*(y) = \sup_{x \in \mathbb{R}_+^D} \{x \cdot y - \xi(x)\},$$

and similarly for $(t\xi)^*$ for any $t \geqslant 0$. We note that for every $t > 0$, we have $(t\xi)^* = t\, \xi^* (\cdot/t)$.

Theorem 1.1 (Parisi formula)**.** *For every $t \geqslant 0$, we have*

$$(1.11) \qquad \lim_{N \to +\infty} \overline{F}_N(t, \delta_0) = \sup_{\mu \in \mathcal{P}_\infty^\uparrow(\mathbb{R}_+^D)} \left\{ \psi((t\nabla\xi)(\mu)) - t \int \theta \, \mathrm{d}\mu \right\}$$

$$(1.12) \qquad \qquad \qquad = \sup_{\nu \in \mathcal{P}_\infty^\uparrow(\mathbb{R}_+^D)} \left\{ \psi(\nu) - \int (t\xi)^* \, \mathrm{d}\nu \right\}$$

where $(t\nabla\xi)(\mu)$ denotes the image of the measure μ through the mapping $x \mapsto t\nabla\xi(x)$.

A measure $\mu \in \mathcal{P}_\infty^\uparrow(\mathbb{R}_+^D)$ that realizes the supremum in (1.11) is usually called a Parisi measure; this measure can often be interpreted as the asymptotic law of the overlap under the Gibbs measure (see [13, Corollary 8.7] or [9, Theorem 1.4] for precise statements). The existence of a Parisi measure can be obtained via coercivity and compactness arguments. For models with a single species ($D = 1$), it was proved in [2] that the Parisi measure is unique. Here we extend this result to all multi-species models.

Theorem 1.2 (Uniqueness of Parisi measure)**.** *For every $t \geqslant 0$, the supremum in* (1.12) *is achieved at exactly one measure, say $\overline{\nu} \in \mathcal{P}_\infty^\uparrow(\mathbb{R}_+^D)$; and the supremum in* (1.11) *is achieved at all measures $\overline{\mu} \in \mathcal{P}_\infty^\uparrow(\mathbb{R}_+^D)$ that satisfy $\overline{\nu} = (t\nabla\xi)(\overline{\mu})$, and only at those measures.*

As already noted, although the function ψ is initially only defined on $\mathcal{P}_1^\uparrow(\mathbb{R}_+^D)$, the formula in (1.8) readily allows us to extend it into a concave function defined on all of $\mathcal{P}_1(\mathbb{R}_+^D)$. In order to show Theorem 1.2, we aim to relax the maximization problem in (1.12) into one that is posed over all of $\mathcal{P}_\infty(\mathbb{R}_+^D)$. In our extension to this variational problem, the last term in there is replaced by the following optimal-transport cost. For every $\mu, \nu \in \mathcal{P}(\mathbb{R}_+^D)$, we denote by $\Pi(\mu, \nu)$ the set of all probability measures on $\mathbb{R}_+^D \times \mathbb{R}_+^D$ with first marginal μ and second marginal ν. For each $t \geqslant 0$ and $\mu, \nu \in \mathcal{P}(\mathbb{R}_+^D)$, we set

$$(1.13) \qquad \mathcal{T}_t(\mu, \nu) = \inf_{\pi \in \Pi(\mu, \nu)} \int (t\xi)^*(y - x) \mathrm{d}\pi(x, y).$$

In this introduction we have only defined $\overline{F}_N(t, \delta_0)$ in (1.7), but below we define $\overline{F}_N(t, \mu)$ for all $(t, \mu) \in \mathbb{R}_+ \times \mathcal{P}_1^\uparrow(\mathbb{R}_+^D)$, and we state our next results for arbitrary arguments.

Theorem 1.3. *For every $(t, \mu) \in \mathbb{R}_+ \times \mathcal{P}_\infty^\uparrow(\mathbb{R}_+^D)$, we have*

$$(1.14) \qquad \lim_{N \to +\infty} \overline{F}_N(t, \mu) = \sup_{\nu \in \mathcal{P}_\infty(\mathbb{R}_+^D)} \{\psi(\nu) - \mathcal{T}_t(\mu, \nu)\}.$$

Moreover, the functional inside the supremum in (1.14) *is concave in ν. Finally, there is exactly one measure $\nu \in \mathcal{P}_\infty(\mathbb{R}_+^D)$ that is monotone and realizes the supremum in* (1.14).

Using convex-duality arguments inspired by [31, 42], we then deduce from Theorem 1.3 a representation of the limit free energy as an infimum. We denote by $\mathfrak{X}$ the set of functions $\chi : \mathbb{R}_+^D \to \mathbb{R}$ of the form $\chi(x) = \sum_{d=1}^D \lambda_{\infty,d} \chi_d(x_d)$ where $\chi_1, \ldots, \chi_D : \mathbb{R}_+ \to \mathbb{R}$ are convex, 1-Lipschitz, increasing functions that vanish at the origin. For every $\chi \in \mathfrak{X}$, $x \in \mathbb{R}_+^D$ and $t \geqslant 0$, we let

$$(1.15) \qquad \psi_*(\chi) = \inf_{\nu \in \mathcal{P}_1^\uparrow(\mathbb{R}_+^D)} \left\{ \int \chi \mathrm{d}\nu - \psi(\nu) \right\}$$

and

$$(1.16) \qquad S_t\chi(x) = \sup_{p\in\mathbb{R}_+^D} \{p \cdot x + t\xi(p) - \chi^*(p)\}$$

$$(1.17) \qquad\qquad = \sup_{y\in\mathbb{R}_+^D} \{\chi(y) - (t\xi)^*(y - x)\}.$$

In (1.16), the definition of χ^* is as in (1.10). The equality between (1.16) and (1.17) will be explained below in Lemma 5.12.

Theorem 1.4. *For every $(t,\mu) \in \mathbb{R}_+ \times \mathcal{P}_\infty^\uparrow(\mathbb{R}_+^D)$, we have*

$$(1.18) \qquad \lim_{N\to+\infty} \overline{F}_N(t,\mu) = \inf_{\chi\in\mathfrak{X}}\left\{\int S_t\chi \, \mathrm{d}\mu - \psi_*(\chi)\right\}.$$

Finally, we describe a more concrete representation of the limit free energy as an infimum over martingales. Let $\mathscr{P} = (\Omega, \mathcal{F}, \mathbf{P})$ be a probability space with associated expectation $\mathbf{E}$, and let $(\mathcal{F}_1(t))_{t\geqslant 0}$, ..., $(\mathcal{F}_D(t))_{t\geqslant 0}$ be an independent family of filtrations over $\mathscr{P}$. We assume that all these σ-algebras are complete, that is, they all contain every subset of any null-measure set. We also assume that the probability space is rich enough to accommodate for the existence of independent Brownian motions $(B_1(t))_{t\geqslant 0}$, ..., $(B_D(t))_{t\geqslant 0}$ that are adapted to the filtrations $(\mathcal{F}_1(t))_{t\geqslant 0}$, ..., $(\mathcal{F}_D(t))_{t\geqslant 0}$ respectively. We denote by $\mathbf{Mart}_1$, ..., $\mathbf{Mart}_D$ the spaces of bounded martingales over $\mathscr{P}$ with respect to the filtrations $(\mathcal{F}_1(t))_{t\geqslant 0}$, ..., $(\mathcal{F}_D(t))_{t\geqslant 0}$ respectively, and we set

$$(1.19) \qquad \mathbf{Mart} = \prod_{d=1}^{D} \mathbf{Mart}_d.$$

For every $\alpha = (\alpha_1,\ldots,\alpha_D) \in \mathbf{Mart}$, we define $\chi_\alpha : \mathbb{R}_+^D \to \mathbb{R}$ such that for every $x = (x_1,\ldots,x_D) \in \mathbb{R}_+^D$,

$$(1.20) \qquad \chi_\alpha(x) = \sum_{d=1}^{D} \lambda_{\infty,d} \int_0^{x_d} \mathbf{E}\left[\alpha_d(t)^2\right] \mathrm{d}t.$$

We also define, for every $t \geqslant 0$ and $x \in \mathbb{R}$,

$$(1.21) \qquad \phi_d(t,x) = \log \int \exp\left(x\sigma - t|\sigma|^2\right) \mathrm{d}\boldsymbol{\pi}_d(x),$$

and set, for every $t \geqslant 0$ and $y \in \mathbb{R}$,

$$(1.22) \qquad \phi_d^*(t,y) = \sup_{x\in\mathbb{R}} \{x \cdot y - \phi_d(t,x)\}.$$

Theorem 1.5. *For every* $t \geqslant 0$, $\mu \in \mathcal{P}_\infty^\uparrow(\mathbb{R}_+^D)$, *and* $T = (T_1, \ldots, T_D) \in \mathbb{R}_+^D$ *such that* $\mathrm{supp}(\mu) + t\nabla\xi([0,1]^D) \subseteq \prod_{d=1}^D [0, T_d]$, *we have*

(1.23)

$$\lim_{N \to +\infty} \overline{F}_N(t, \mu) = \inf_{\alpha \in \mathbf{Mart}} \bigg\{ \sum_{d=1}^D \lambda_{\infty,d} \, \mathbf{E}\Big[\phi_d^*(T_d, \alpha_d(T_d)) - \sqrt{2}\alpha_d(T_d)B_d(T_d) \Big]$$
$$- \chi_\alpha(T) + \int S_t \chi_\alpha \mathrm{d}\mu \bigg\}.$$

In addition the infimum in (1.23) *is reached at a unique* $\overline{\alpha} \in \mathbf{Mart}$. *This martingale* $\overline{\alpha}$ *is such that for every* $d \in [D]$ *and almost every* $t \in [0, T_d]$, *we have* $\overline{\alpha}_d(t) = \partial_x \Phi_{\overline{\nu}_d}(t, X_d(t))$, *where* $\overline{\nu}$ *is the unique monotone measure that realizes the supremum in* (1.14), $\Phi_{\overline{\nu}_d}$ *is defined in* (3.2) *subject to the terminal condition* (3.3), *and* X_d *is defined in* (8.3).

Conjecture. We conjecture that Theorems 1.4 and 1.5 remain valid even when the function ξ is not assumed to be convex on $\mathbb{R}_+^D$, with the understanding that $S_t\chi$ is defined according to (1.16) (the equality of this quantity with (1.17) would be lost in this case). To explain why, we start by recalling that when ξ is convex over $\mathbb{R}_+^D$, the limit free energy is known [9, 13, 16, 39, 43] to converge to the unique viscosity solution to the Hamilton–Jacobi equation

(1.24)
$$\begin{cases} \partial_t f - \int \xi(\partial_\mu f) \, \mathrm{d}\mu = 0 & \text{on } \mathbb{R}_+ \times \mathcal{P}_2^\uparrow(\mathbb{R}_+^D), \\ f(0, \cdot) = \psi & \text{on } \mathcal{P}_2^\uparrow(\mathbb{R}_+^D). \end{cases}$$

The same result may also be valid in the case when ξ is non-convex, and partial results in this direction were obtained in [9, 13, 16, 38, 40]. As discussed more precisely in [31], the right-hand side of (1.18) can be interpreted as a generalized version of the Hopf variational representation of the solution to (1.24). In other words, there is a plausible heuristic argument connecting the Hamilton–Jacobi equation (1.24) with the variational representation in Theorem 1.4, and therefore also with that in Theorem 1.5; the idea being that since ψ is concave, it can be represented as an infimum of affine functions, and we expect the solution to the PDE to be the infimum of the solutions started with these enveloping affine functions. One consequence of the results presented here is a proof that when ξ is convex over $\mathbb{R}_+^D$, the solution to the Hamilton–Jacobi equation (1.24) can indeed be represented as the variational formula in (1.18). We expect this link between the solution to (1.24) and the variational representations in Theorems 1.4 and 1.5 to remain valid even when the function ξ is no longer assumed to be convex. In addition to being interesting on its own, this new representation could be helpful to complement the lower bound obtained in [38, 40] with the reverse bound.

Proof ideas. For the choice of $\mu = \delta_0$, and in view of (1.12), the task of showing (1.14) boils down to showing that

$$(1.25) \qquad \sup_{\nu \in \mathcal{P}^{\uparrow}_{\infty}(\mathbb{R}^D_+)} \left\{ \psi(\nu) - \int (t\xi)^* \, d\nu \right\} = \sup_{\nu \in \mathcal{P}_{\infty}(\mathbb{R}^D_+)} \left\{ \psi(\nu) - \int (t\xi)^* \, d\nu \right\}.$$

For each measure $\nu \in \mathcal{P}(\mathbb{R}^D_+)$, there is a unique measure $\nu^{\uparrow}$ that has the same marginals as ν and is a monotone measure. Due to the equality of the marginals and (1.8), we have that $\psi(\nu) = \psi(\nu^{\uparrow})$, so surely the most natural idea for showing (1.25) would be to assert that

$$(1.26) \qquad \int (t\xi)^* d\nu^{\uparrow} \leqslant \int (t\xi)^* d\nu.$$

We show in Proposition 4.1 that the inequality (1.26) is indeed valid when $D = 2$ and $\nabla\xi(0) = 0$. However, we also show in Remark 4.3 that the inequality (1.26) is false in general, so another strategy must be developed towards the proof of our main results.

Our approach consists instead of a careful analysis of the optimizers of the variational formula in (1.14). First, recall that the definition of $S_t\chi$ from (1.17) implies that for every $x, y \in \mathbb{R}^D_+$,

$$(1.27) \qquad \chi(y) - S_t\chi(x) \leqslant (t\xi)^*(y - x),$$

with equality when y is a maximizer in (1.17). From this, we check in Lemma 6.4 that

$$(1.28) \qquad \int S_t\chi \, d\mu = \sup_{\nu \in \mathcal{P}_{\infty}(\mathbb{R}^D_+)} \left\{ \int \chi \, d\nu - \mathcal{T}_t(\mu, \nu) \right\}.$$

Letting $\overline{\nu} \in \mathcal{P}_{\infty}(\mathbb{R}^D_+)$ denote a maximizer of (1.14), and χ denote the derivative of ψ at $\overline{\nu}$, we observe that, thanks to the convexity of the mapping $\nu \mapsto \mathcal{T}_t(\mu, \nu)$, we have that

$$(1.29) \qquad \int \chi \, d\overline{\nu} - \mathcal{T}_t(\mu, \overline{\nu}) = \sup_{\nu \in \mathcal{P}_{\infty}(\mathbb{R}^D_+)} \left\{ \int \chi \, d\nu - \mathcal{T}_t(\mu, \nu) \right\},$$

and thus, in view of (1.28), that

$$(1.30) \qquad \int \chi \, d\overline{\nu} - \int S_t\chi \, d\mu = \mathcal{T}_t(\mu, \overline{\nu}).$$

Denoting by π the optimal transport plan from μ to $\overline{\nu}$, the previous display tells us that the inequality (1.27) is in fact an equality for $d\pi$-almost every (x, y). Heuristically, we then want to exploit certain monotonicity properties of the mappings $\nabla\chi$ and $\nabla S_t\chi$ to deduce that if μ is monotone, then so is $\overline{\nu}$. The details are a bit more subtle, as it may well be that the measure $\overline{\nu}$ itself is actually not monotone, and we need to smear out the measure μ in order to effectuate this transfer of monotonicity, but we ultimately show that we can build a maximizing measure for (1.14) that is indeed monotone whenever μ itself is monotone. The bulk of the technical work consists in showing that $\nabla\chi$ and $\nabla S_t\chi$ satisfy the desired monotonicity properties. Once Theorem 1.3 is shown, we obtain Theorem 1.2 as a direct consequence, while the proofs of

Theorems 1.4 and 1.5 rely on convex-duality arguments that are inspired by [31, 42].

Related works. The Parisi formula was initially predicted using the non-rigorous replica method in the case of the Sherrington-Kirkpatrick model (i.e. for $D = 1$, $\xi(r) = r^2$, and $P_1 = \frac{1}{2}(\delta_1 + \delta_{-1})$) in [36, 50, 51, 52, 53], and was later proved rigorously in [27, 60]. A more conceptual proof based on the ultrametricity of the Gibbs measure was later developed [45, 46], and allowed all models with a single type of spins ($D = 1$) to be covered. This ultrametricity property was further leveraged in order to identify the limit free energy of spin glasses with multiple types and convex ξ [4, 47, 48, 49]. One key ingredient of the proof consists in showing that, up to a small perturbation of the energy function, the overlaps between the different types of spins must synchronize with one another; in other words, the joint law of the overlaps must be a monotone measure. Theorem 1.3 shows that one can in fact relax the final optimization problem into a "desynchronized" one, where the measure is no longer constrained to be monotone. Concurrent developments for spherical models include [17, 59] for models with a single type of spins, and [5] for multi-type models. For models with additional symmetries, the limit formula may in certain cases be simplified into an expression corresponding to an effective single-type model [6, 7, 10, 30, 41].

Models with non-convex ξ are less well-understood, and even the existence of the limit free energy is not proved in general. Yet, under the assumption that the limit free energy exists, its value has been identified in [57, 58] for all spherical models in the case when ξ is a monomial. Still for spherical models, the case of $D = 2$ and $\xi(x, y) = xy$ has been obtained unconditionally in [1, 3], and the cases of $\xi(x, y) = x^p y^q$ and $\xi(x_1, \ldots, x_D) = x_1 \cdots x_D$ have also been obtained unconditionally for particular choices of the parameter λ_∞ in [7, 19].

Outside of these cases, even stating an appropriate conjecture for the limit free energy is a non-trivial task. Indeed, several possible candidate variational formulas for the limit free energy, inspired by the classical Parisi formula, were shown to be invalid in [38, Section 6]. A candidate for the limit free energy, defined in terms of the solution to a Hamilton-Jacobi equation, was proposed in [16, 38, 39, 40], and this candidate was shown there to be a lower bound for the limit free energy. Under the assumption that the limit free energy exists, a representation of this limit in terms of a critical point of an explicit functional was also found in [9, 13]. This description is in the spirit of the cavity fixed-point equations appearing in the physics literature. It does not completely characterize the limit however, as there may be more than one critical point in general.

As discussed around (1.24), and despite the negative results of [38, Section 6], we now believe that there is in fact a plausible candidate variational formula for the limit free energy for non-convex ξ. The point is that it does

not take a form similar to the classical Parisi formula, but rather a form as in (1.23). Such alternative "un-inverted" formulas first appeared in the context of models with a single type of spins in [42], and were further developed in [31], with results close to Theorem 1.4 for vector spin glasses. A key ingredient for our analysis as well as in [42] is the fact that the function ψ is concave. This was first proved for models with a single type of spins in [2]; the uniqueness of the Parisi measure was also obtained there in this case. For spherical models, the uniqueness of the Parisi measure can be seen more directly, and a similar convex-duality analysis was performed in [33, 34]. We also mention that for vector spin glasses with convex ξ, the uniqueness of the Parisi measure was shown to be valid in [11] provided that the measure argument μ does not charge any single point.

These "un-inverted" formulas are in turn inspired by a series of works concerning the construction of efficient algorithms that aim to identify configurations σ in the support of P_N such that $H_N(\sigma)$ is as large as possible. For many models, an explicit quantity ALG was identified as a variational formula analogous to the Parisi formula, such that the following holds. On the one hand, there exists an efficient algorithm that outputs a configuration σ in the support of P_N such that $H_N(\sigma)/N$ is approximately equal to ALG. On the other hand, for any $\varepsilon > 0$, if an algorithm is Lipschitz continuous with respect to its entries, then it cannot reliably output a configuration σ in the support of P_N such that $H_N(\sigma)/N$ exceeds ALG $+ \varepsilon$. Whether or not this quantity ALG coincide with the limit of $\sup_{\sigma \in \text{supp} P_N} H_N(\sigma)$ depends on the choice of ξ. At least for models with a single type, this quantity ALG can also be written as an "un-inverted" formula very similar to that found in [42].

An algorithm that indeed achieves the correct value ALG was first put forward in [56] for spherical models with a single type of spins. The algorithm is based on a Hessian ascent. Using a different approach based on approximate message passing algorithms, this result was extended to models with ±1 spins in [21, 37, 54]. For models with ±1 spins, a Hessian-ascent algorithm was also constructed in [35]. The converse statement, that no Lipschitz algorithm can identify configurations with energy exceeding ALG, is based on the notion of overlap gap property [24, 25, 26, 29]. Remarkably, the identification of this threshold ALG has been extended to spherical models with multiple types, including for models with non-convex ξ for which the limit free energy is not currently known [28, 29].

Organization of the paper. In Section 2, we define the enriched free energy $\overline{F}_N(t,\mu)$ for arbitrary arguments $(t,\mu) \in \mathbb{R}_+ \times \mathcal{P}_1(\mathbb{R}_+)$. In Section 3, we study properties of $\psi(\mu) = \lim_{N \to \infty} \overline{F}_N(0,\mu)$, including its regularity and concavity and the decomposition in (1.8). In particular, we show that the derivative of ψ at any measure is an element of the set $\mathfrak{X}$ defined above (1.15). In Section 4, we focus on the two-species case ($D = 2$) with no external field

$(\nabla \xi(0) = 0)$, and show (1.26) in this particular case, leading to a simple proof of (1.25), with some long but unsurprising proof deferred to Appendix A. Section 5 is focused on the proof that $\nabla S_t \chi$ satisfies the desired monotonicity properties whenever $\chi \in \mathfrak{X}$. We show that the mapping $(t, x) \mapsto S_t \chi(x)$ can be interpreted as the solution to a Hamilton-Jacobi equation, and we rely on this property to obtain the desired result. Section 6 contains the proofs of (1.28) and (1.30). We also show that the variational problems appearing in (1.12) and (1.14) are equal. We complete the proofs of Theorems 1.1 to 1.4 in Section 7. Finally, we develop a more explicit dual representation of ψ and use it to prove Theorem 1.5 in Section 8.

2. Definition of the enriched free energy

The goal of this section is to give a precise definition of the enriched free energy $\overline{F}_N(t, \mu)$. We start by giving a definition of this quantity in the case when the measure $\mu \in \mathcal{P}^{\uparrow}(\mathbb{R}_+^D)$ has a finite support. For every $x, y \in \mathbb{R}^D$, we write $x \leqslant y$ when $y - x \in \mathbb{R}_+^D$. Let $K \in \mathbb{N}$ and let $\mu \in \mathcal{P}^{\uparrow}(\mathbb{R}_+^D)$ be of the form

$$(2.1) \qquad \mu = \sum_{k=0}^{K} (\zeta_{k+1} - \zeta_k) \delta_{q_k} \quad \text{with marginals} \quad \mu_d = \sum_{k=0}^{K} (\zeta_{k+1} - \zeta_k) \delta_{q_{k,d}}$$

for each $d \in [D]$, where ζ_k and $q_k = (q_{k,d})_{d \in [D]}$ satisfy

$$(2.2) \qquad\qquad 0 = q_{-1,d} \leqslant q_{0,d} < q_{1,d} < \cdots < q_{K,d},$$

$$(2.3) \qquad\qquad 0 = \zeta_0 < \zeta_1 < \cdots < \zeta_K < \zeta_{K+1} = 1.$$

The construction of $\overline{F}_N(t, \mu)$ involves a random probability measure with ultrametric properties called the Poisson–Dirichlet cascade. We briefly introduce this object and refer to [46, Section 2.3] or [20, Section 5.6] for a more detailed explanation. We define

$$(2.4) \qquad\qquad \mathcal{A} = \mathbb{N}^0 \cup \mathbb{N}^1 \cup \cdots \cup \mathbb{N}^K$$

with the understanding that $\mathbb{N}^0 = \{\varnothing\}$. We think of $\mathcal{A}$ as a tree rooted at $\varnothing$ such that each vertex of depth $k < K$ has countably many children. For each $k < K$ and $\alpha = (n_1, \ldots, n_k) \in \mathbb{N}^k$, the children of α are the vertices of the form

$$(2.5) \qquad\qquad \alpha n = (n_1, \ldots, n_k, n) \in \mathbb{N}^{k+1}.$$

The depth of $\alpha = (n_1, \ldots, n_k)$ is denoted by $|\alpha| = k$ and for every $l \leqslant k$, we write

$$(2.6) \qquad\qquad \alpha_{|l} = (n_1, \ldots, n_l)$$

to denote the ancestor of α at depth l. Given two leaves $\alpha, \beta \in \mathbb{N}^K$, we denote by $\alpha \wedge \beta$ the generation of the most recent common ancestor of α and β, that is

$$(2.7) \qquad\qquad \alpha \wedge \beta = \sup\{k \leqslant K : \alpha_{|k} = \beta_{|k}\}.$$

We attach an independent Poisson process to each non-leaf vertex $\alpha \in \mathcal{A}$ with intensity measure

$$(2.8) \qquad x^{-1-\zeta_{|\alpha|+1}}\mathrm{d}x.$$

We order increasingly the points of those Poisson processes and denote them by $u_{\alpha 1} \geqslant u_{\alpha 2} \geqslant \dots$. For every $\alpha \in \mathbb{N}^K$, we set $w_\alpha = \prod_{k=1}^K u_{\alpha_{|k}}$ and define

$$(2.9) \qquad v_\alpha = \frac{w_\alpha}{\sum_{\beta \in \mathbb{N}^K} w_\beta}.$$

The Poisson–Dirichlet cascade associated to $(\zeta_k)_{1 \leqslant K+1}$ in (2.3) is the random probability measure on $\mathbb{N}^K$ (the leaves of the tree $\mathcal{A}$) whose weights are given by $(v_\alpha)_{\alpha \in \mathbb{N}^K}$.

Now, let $(v_\alpha)_{\alpha \in \mathbb{N}^K}$ be the Poisson–Dirichlet cascade associated to $(\zeta_k)_{1 \leqslant K+1}$ in (2.3), chosen to be independent of H_N. Let $(z_{\beta,d})_{\beta \in \mathcal{A}, d \in [D]}$ be a family of independent standard Gaussian variables. We choose $(z_{\beta,d})_{\beta \in \mathcal{A}}$ independent of $(v_\alpha)_{\alpha \in \mathbb{N}^K}$ and H_N. For every $\alpha \in \mathbb{N}^K$, we set

$$(2.10) \qquad w^{\mu_d}(\alpha) = \sum_{k=0}^K (q_{k,d} - q_{k-1,d})^{1/2} z_{\alpha_{|k},d}$$

with $(q_{k,d})_{0 \leqslant k \leqslant K}$ given in (2.2). The centered Gaussian process $(w^{\mu_d}(\alpha))_{\alpha \in \mathbb{N}^K}$ has the following covariance structure

$$(2.11) \qquad \mathbb{E}\left[w^{\mu_d}(\alpha)w^{\mu_d}(\alpha')\right] = q_{\alpha \wedge \alpha',d}.$$

Let $(\omega_n^{\mu_d})_{n \in \mathbb{N}}$ be i.i.d. copies of ω^{μ_d}. For each $N \in \mathbb{N}$, we define

$$(2.12) \qquad W_N^\mu(\sigma,\alpha) = \sum_{d \in [D]} \sum_{n \in I_{N,d}} w_n^{\mu_d}(\alpha) \cdot \sigma_n$$

which is a centered Gaussian process with covariance

$$(2.13) \qquad \mathbb{E}\left[W_N^\mu(\sigma,\alpha)W_N^\mu(\sigma',\alpha')\right] = N q_{\alpha \wedge \alpha'} \cdot R_N(\sigma,\sigma').$$

For $t \in \mathbb{R}_+$ and μ given in (2.1), we define the enriched Hamiltonian

$$H_N^{t,\mu}(\sigma,\alpha) = \sqrt{2t}H_N(\sigma) - tN\xi\left(R_N(\sigma,\sigma)\right) + \sqrt{2}W_N^\mu(\sigma,\alpha) - Nq_K \cdot R_N(\sigma,\sigma),$$

where H_N is given as in (1.3). We define

$$(2.14) \qquad \overline{F}_N(t,\mu) = -\frac{1}{N}\mathbb{E}\log \int \sum_{\alpha \in \mathbb{N}^K} \exp\left(H_N^{t,\mu}(\sigma,\alpha)\right) v_\alpha \mathrm{d}P_N(\sigma).$$

The expression in the previous display is Lipschitz with respect to (t,μ). More precisely, for every $t, t' \in \mathbb{R}_+$ and every μ, μ' of the form in (2.1), we have

$$(2.15) \qquad |\overline{F}_N(t,\mu) - \overline{F}_N(t',\mu')| \leqslant \mathsf{W}_1(\mu,\mu') + |t-t'| \sup_{|a| \leqslant 1} |\xi(a)|,$$

where W_1 denotes the Wasserstein L^1 distance between μ and μ'. This is borrowed from [9, Proposition 4.1], where the term $\mathsf{W}_1(\mu,\mu')$ is written as $\int_0^1 |\mathsf{q}(s) - \mathsf{q}'(s)|\mathrm{d}s$ with $\mu \stackrel{\mathrm{d}}{=} \mathrm{Law}(\mathsf{q}(U))$ and $\mu' \stackrel{\mathrm{d}}{=} \mathrm{Law}(\mathsf{q}'(U))$ for a uniform

random variable U over $[0,1]$. The equality between the two expressions can be obtained from a straightforward modification of [40, Proposition 2.5].

As a consequence, the free energy $\overline{F}_N$ admits a unique Lipschitz extension to $\mathbb{R}_+ \times \mathcal{P}_1^\uparrow(\mathbb{R}_+^D)$.

Given w^{μ_d} in (2.10) and $\boldsymbol{\pi}_d$ in (1.4), each ψ_d appearing in (1.8) is given by

$$(2.16) \qquad \psi_d(\mu_d) = -\mathbb{E}\log \int \sum_{\alpha \in \mathbb{N}^K} \exp\left(\sqrt{2}w^{\mu_d}(\alpha)\tau - q_{K,d}\tau^2\right) v_\alpha \mathrm{d}\boldsymbol{\pi}_d(\tau)$$

at every μ of the form in (2.1). A similar property as in (2.15) holds for ψ_d and thus ψ_d can be extended to $\mathcal{P}_1^\uparrow(\mathbb{R}_+) = \mathcal{P}_1(\mathbb{R}_+)$. By an invariance property of $(\nu_\alpha)_{\alpha \in \mathbb{N}^K}$ (see e.g. [20, Corollary 5.26] and [9, Lemma 4.11]), we have

$$(2.17) \qquad \overline{F}_N(0,\mu) = \sum_{d \in [D]} \lambda_{N,d}\psi_d(\mu_d), \qquad \forall \mu \in \mathcal{P}_1^\uparrow(\mathbb{R}_+^D).$$

Therefore, due to (1.6) and (1.8), we have

$$(2.18) \qquad \lim_{N\to\infty} \overline{F}_N(0,\mu) = \psi(\mu), \qquad \forall \mu \in \mathcal{P}_1^\uparrow(\mathbb{R}_+^D).$$

In order to state the Parisi formula for $\overline{F}_N(t,\mu)$ at a general measure μ, we introduce the following notation. Let $\mathcal{Q}$ be the collection of càdlàg increasing maps $\mathsf{q}: [0,1) \to \mathbb{R}_+^D$, and let U denote a uniform random variable on $[0,1]$. For any $\mathsf{q} \in \mathcal{Q}$, we write

$$(2.19) \qquad \mathcal{L}_{\mathsf{q}} = \mathrm{Law}(\mathsf{q}(U)) \in \mathcal{P}^\uparrow(\mathbb{R}_+^D).$$

We also set $\mathcal{Q}_\infty = \mathcal{Q} \cap L^\infty([0,1);\mathbb{R}^D)$. It is clear that $\mathsf{q} \mapsto \mathcal{L}_{\mathsf{q}}$ is a bijection from $\mathcal{Q}$ to $\mathcal{P}^\uparrow(\mathbb{R}_+^D)$, and also from $\mathcal{Q}_\infty$ to $\mathcal{P}_\infty^\uparrow(\mathbb{R}_+^D)$. The following is a restatement of [9, Proposition 6.1] which adapts [13, Proposition 8.1] from the setting of vector spin glasses. It generalizes (1.11) in Theorem 1.1.

Proposition 2.1 (Parisi formula). *For every $t \geqslant 0$ and $\mathsf{q} \in \mathcal{Q}_\infty$, denoting $\mu = \mathcal{L}_{\mathsf{q}} \in \mathcal{P}_\infty^\uparrow(\mathbb{R}_+^D)$, we have*

$$(2.20) \qquad \lim_{N\to+\infty} \overline{F}_N(t,\mu) = \sup_{\mathsf{p}\in\mathcal{Q}_\infty} \left\{ \psi\left(\mathcal{L}_{\mathsf{q}+t\nabla\xi(\mathsf{p})}\right) - t\int \theta \,\mathrm{d}\mathcal{L}_{\mathsf{p}} \right\}.$$

In (2.20), the notation $\nabla\xi(\mathsf{p})$ stands for the path $s \mapsto \nabla\xi(\mathsf{p}(s))$, and $\int \theta \mathrm{d}\mathcal{L}_{\mathsf{p}}$ can be explicitly written as $\int_0^1 \theta(\mathsf{p}(s))\mathrm{d}s$. The right-hand side in (2.20) can be connected to the unique solution of a Hamilton–Jacobi equation [13, 38, 39, 40, 43].

3. Properties of the enriched free energy at infinite temperature

In this section, we derive useful properties of ψ. Due to the decomposition in (1.8), we mostly focus on the study of ψ_d for each $d \in [D]$.

3.1. **Properties of ψ_d.** We fix $d \in [D]$, and use the following notation in this subsection:

$$(3.1) \qquad \boldsymbol{\pi}_\circ = \boldsymbol{\pi}_d \qquad \text{and} \qquad \psi_\circ = \psi_d.$$

For $\nu \in \mathcal{P}(\mathbb{R}_+)$, we denote by $t \mapsto \nu(t) = \nu([0,t])$ its cumulative distribution function, which is a right-continuous increasing function taking values in $[0,1]$. When $\nu \in \mathcal{P}_\infty(\mathbb{R}_+)$, we denote by

$$\nu^{-1}(1) = \inf\{t \in \mathbb{R}_+ : 1 \leqslant \nu(t)\}$$

the right endpoint of the support of ν. For $\nu \in \mathcal{P}_\infty(\mathbb{R}_+)$, we consider the backward parabolic equation

$$(3.2) \qquad -\partial_t \Phi_\nu(t,x) = \partial_x^2 \Phi_\nu(t,x) + \nu(t)(\partial_x \Phi_\nu)^2(t,x)$$

with (t,x) ranging in $[0,\nu^{-1}(1)] \times \mathbb{R}$, and with the terminal condition

$$(3.3) \qquad \Phi_\nu(\nu^{-1}(1),x) = \log \int \exp\left(x\sigma - \nu^{-1}(1)\sigma^2\right) \mathrm{d}\boldsymbol{\pi}_\circ(\sigma), \quad \forall x \in \mathbb{R}.$$

We call the equation (3.2) the *Parisi PDE* associated with ν. The fact that the function Φ_ν is well-defined can be found in [32]. When the support of ν is finite, one can use the Cole-Hopf transform to solve (3.2) on each time sub-interval of $[0,\nu^{-1}(1))$ on which $\nu(\cdot)$ is constant; and then one can argue by continuity to build Φ_ν for arbitrary ν (see for instance [20, Section 6.5]). We extend Φ_ν to $\mathbb{R}_+ \times \mathbb{R}$ by setting

$$(3.4) \quad \Phi_\nu(t,x) = \log \int \exp\left(x\sigma - t\sigma^2\right) \mathrm{d}\boldsymbol{\pi}_\circ(\sigma), \quad \forall (t,x) \in [\nu^{-1}(1),\infty) \times \mathbb{R}.$$

A simple computation reveals that this extension also satisfies the equation (3.2) on $[\nu^{-1}(1),\infty) \times \mathbb{R}$. We call $\Phi_\nu : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}$ the *Parisi PDE solution* associated with $\nu \in \mathcal{P}_\infty(\mathbb{R}_+)$. We summarize the basic properties of the Parisi PDE solution in the following lemma.

Lemma 3.1 (Regularity of Parisi PDE). *For every $\nu \in \mathcal{P}_\infty(\mathbb{R}_+)$ and $k \in \mathbb{N} = \{1,2,\ldots\}$, the derivative $\partial_x^k \Phi_\nu$ exists everywhere and is bounded uniformly over ν, t, x. In particular, $|\partial_x \Phi_\nu| \leqslant 1$ everywhere. For every $k \in \mathbb{N} \cup \{0\}$, there is a constant C_k such that for every $\nu, \nu' \in \mathcal{P}_\infty(\mathbb{R}_+)$ and $(t,x) \in \mathbb{R}_+ \times \mathbb{R}$, we have*

$$(3.5) \qquad \left|\partial_x^k \Phi_\nu(t,x) - \partial_x^k \Phi_{\nu'}(t,x)\right| \leqslant C_k \int_t^\infty \left|\nu(\tau) - \nu'(\tau)\right| \mathrm{d}\tau.$$

For every $\nu \in \mathcal{P}_1(\mathbb{R}_+)$, we have

$$(3.6) \qquad \psi_\circ(\nu) = -\Phi_\nu(0,0).$$

Proof. First, assume $\nu, \nu' \in \mathcal{P}_\infty(\mathbb{R}_+)$. For each ν, there is an explicit expression for Φ_ν given in [12, Lemma 4.5 and Remark 4.6] (with $D = 1$ and $L(t) = t$ for $t \in \mathbb{R}_+$ therein, and with ν and P_1 substituted for α and μ therein). In particular, both (3.3) and (3.4) are satisfied. The relation to the initial condition as in (3.6) is given in [12, Lemma 5.4]. The fact that Φ_ν satisfies the equation and the regularity of Φ_ν are from [12, Proposition 4.7].

The particular case $|\partial_x \Phi_\nu| \leqslant 1$ follows from [12, Proposition 4.7 (3)] and the fact that $\boldsymbol{\pi}_\circ$ is supported in $[-1, 1]$. The inequality (3.5) then allows us to extend the results to every $\nu, \nu' \in \mathcal{P}_1(\mathbb{R}_+)$. To extend (3.6), we also use the continuity of $\psi_\circ$. $\qquad\qquad\square$

We now recall the concavity of $\psi_\circ$ originally due to [2].

Lemma 3.2 (Concavity of $\psi_\circ$)**.** *The function $\psi_\circ : \mathcal{P}_1(\mathbb{R}_+) \to \mathbb{R}$ is concave; that is, for every $\nu, \nu' \in \mathcal{P}_1(\mathbb{R}_+)$ and $\lambda \in [0, 1]$, we have $\psi_\circ((1 - \lambda)\nu + \lambda\nu') \geqslant (1 - \lambda)\psi_\circ(\nu) + \lambda\psi_\circ(\nu')$.*

The restriction of $\psi_\circ$ to $\mathcal{P}_\infty(\mathbb{R}_+)$ is strictly concave; that is, for every distinct $\nu, \nu' \in \mathcal{P}_\infty(\mathbb{R}_+)$ and $\lambda \in (0, 1)$, we have $\psi_\circ((1 - \lambda)\nu + \lambda\nu') > (1 - \lambda)\psi_\circ(\nu) + \lambda\psi_\circ(\nu')$.

Proof. By the Lipschitzness of $\psi_\circ$ in ν discussed in the previous section (or implied by (3.5) and (3.6)), it suffices to prove the strict concavity on $\mathcal{P}_\infty(\mathbb{R}_+)$. Fix any distinct ν and $\nu' \in \mathcal{P}_\infty(\mathbb{R}_+)$. By (3.4), we can fix any $T > \nu^{-1}(1), \nu'^{-1}(1)$ so that Φ_ν and $\Phi_{\nu'}$ both satisfy the Parisi PDE (3.2) on $[0, T] \times \mathbb{R}$ with the same terminal condition equal to $x \mapsto \log \int \exp(x \cdot \sigma - T\sigma^2) \mathrm{d}\boldsymbol{\pi}_\circ(\sigma)$. Then, the concavity follows from (3.6) and the strict convexity of $\nu \mapsto \Phi_\nu(0, 0)$ on the set of probability measures with support in $[0, T]$, which is available from a straightforward modification of [2, Theorem 4 (i)]. $\qquad\square$

In order to determine the derivative of $\psi_\circ$, we introduce an adjoint equation to the Parisi PDE, for each $\nu \in \mathcal{P}^1(\mathbb{R}_+)$. Let $P : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}$ be the one-dimensional heat kernel, namely

$$(3.7) \qquad\qquad P(t, x) = \frac{1}{\sqrt{4\pi t}} e^{-\frac{x^2}{4t}}.$$

Since $\partial_x \Phi_\nu$ is bounded due to Lemma 3.1, we can use a Picard fixed-point argument in a suitable function space to build a function $u_\nu : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}$ such that, for every $(t, x) \in \mathbb{R}_+ \times \mathbb{R}$,

$$u_\nu(t, x) = P(t, x) + \int_0^t \int_\mathbb{R} \partial_x P(t - s, x - y) u_\nu(s, y) \left(2\nu(s)\partial_x \Phi_\nu(s, y)\right) \mathrm{d}y \mathrm{d}s.$$

In other words, the function u_ν is a mild solution to

$$(3.8) \qquad \begin{cases} \partial_t u_\nu = \partial_x^2 u_\nu - 2\nu(t)\partial_x \left(u_\nu \partial_x \Phi_\nu\right), & \text{on } \mathbb{R}_+ \times \mathbb{R}, \\ u_\nu(0, \cdot) = \delta_0, & \text{on } \mathbb{R}, \end{cases}$$

where δ_0 is the Dirac mass at zero. One can verify that u_ν is also a weak solution to this equation. From the mild solution formulation, one can check that u_ν and its derivatives in x have exponential decay in x, which allows for integration by parts. We also recognize that (3.8) is a Kolmogorov equation and thus $u_\nu(t, \cdot)$ can be interpreted as the probability density function of a stochastic process at time t. Therefore, u_ν also satisfies

$$(3.9) \qquad u_\nu \geqslant 0 \qquad \text{and} \qquad \int u_\nu(t, x) \mathrm{d}x = 1, \quad \forall t \in \mathbb{R}_+.$$

We use the properties of u_ν to prove the next two lemmas.

Lemma 3.3 (Derivative of $\psi_\circ$). *Let $\nu, \nu' \in \mathcal{P}_1(\mathbb{R}_+)$ and let $\nu_\varepsilon = (1 - \varepsilon)\nu + \varepsilon\nu'$ for $\varepsilon \in [0, 1]$. We have*

$$(3.10) \qquad \frac{\mathrm{d}}{\mathrm{d}\varepsilon}\psi_\circ(\nu_\varepsilon)\big|_{\varepsilon=0} = \int_0^\infty \chi(t)\mathrm{d}(\nu' - \nu)$$

where

$$(3.11) \qquad \chi(t) = \int_0^t \int_\mathbb{R} \left(\partial_x \Phi_\nu(\tau, x)\right)^2 u_\nu(\tau, x)\mathrm{d}x\mathrm{d}\tau, \quad \forall t \in \mathbb{R}_+.$$

Notice that the right-hand side in (3.10) is well-defined and finite since the first moments of ν' and ν exist and $|\chi(t)| \leqslant t$ due to $\int u_\nu(\tau, x)\mathrm{d}x = 1$ and the boundedness of $|\partial_x \Phi_\nu| \leqslant 1$ given by Lemma 3.1.

Proof. For simplicity, we write $\Phi_\varepsilon = \Phi_{\nu_\varepsilon}$ and $u = u_\nu$. For any $T > 0$, we set $C_\varepsilon(T) = \int \Phi_\varepsilon(T, x)u(T, x)\mathrm{d}x$. Using $u(0, \cdot) = \delta_0$ and integrating by parts (or more precisely appealing to the weak formulation of (3.8)), we have

$$\Phi_\varepsilon(0, 0) = \int_\mathbb{R} \Phi_\varepsilon(0, x)u(0, x)\mathrm{d}x = C_\varepsilon(T) - \int_0^T \int_\mathbb{R} \frac{\mathrm{d}}{\mathrm{d}t}\left(\Phi_\varepsilon u\right)\mathrm{d}x\mathrm{d}t.$$

Using the equations satisfied by Φ_ε and u in (3.2) and (3.8), we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\Phi_\varepsilon u\right) = -u\partial_x^2\Phi_\varepsilon - \nu_\varepsilon u(\partial_x \Phi_\varepsilon)^2 + \Phi_\varepsilon \partial_x^2 u - 2\nu\Phi_\varepsilon \partial_x(\partial_x\Phi_0 u).$$

Integrating this in x and using integration by parts, we can cancel the second-order terms and get

$$\Phi_\varepsilon(0, 0) = C_\varepsilon(T) + \int_0^T \int_\mathbb{R} \left(\nu_\varepsilon a_\varepsilon^2 - 2\nu a_\varepsilon a_0\right) u$$

where we used the shorthand $a_\varepsilon = \partial_x \Phi_\varepsilon$. Then, we have

$$\Phi_\varepsilon(0, 0) - \Phi_0(0, 0) - (C_\varepsilon(T) - C_0(T)) = \int_0^T \int_\mathbb{R} \big(\underbrace{\nu_\varepsilon a_\varepsilon^2 - 2\nu a_\varepsilon a_0 + \nu a_0^2}_{\blacksquare}\big)u$$

We can rearrange $\blacksquare = (\nu_\varepsilon - \nu)a_0^2 - (\nu_\varepsilon - \nu)(a_\varepsilon^2 - a_0^2) + \nu_0(a_\varepsilon - a_0)^2$ and we argue that the last two terms are of order $O(\varepsilon^2)$ uniformly in ε, t, x. Indeed, this follows from the facts that $\nu_\varepsilon - \nu = O(\varepsilon)$ by definition, that $a_\varepsilon - a_0 = O(|\nu_\varepsilon - \nu|_{L^1}) = O(\varepsilon)$ due to (3.5), and that $a_\varepsilon, a_0 = O(1)$ by Lemma 3.1. Using this and $\nu_\varepsilon - \nu = \varepsilon(\nu' - \nu)$, we get $\blacksquare = \varepsilon(\nu' - \nu)a_0^2 + O(\varepsilon^2)$ uniformly in ε, t, x. Inserting this to the above display we get

$$\Phi_\varepsilon(0, 0) - \Phi_0(0, 0) - (C_\varepsilon(T) - C_0(T)) = \varepsilon \int_0^T \int_\mathbb{R} (\nu' - \nu)a_0^2 u + O(\varepsilon^2)$$

$$\overset{(3.11)}{=} \varepsilon \int_0^T (\nu' - \nu)\dot{\chi} + O(\varepsilon^2)$$

$$\overset{(\text{IBP})}{=} -\varepsilon \int_0^T \chi \mathrm{d}(\nu' - \nu) + \varepsilon\left(\nu'(T) - \nu(T)\right)\chi(T) + +O(\varepsilon^2).$$

In the last step, we used the fact that $\chi(0) = 0$. Using this relation and $\psi_\circ(\nu_\varepsilon) = -\Phi_\nu(0,0)$ (due to (3.6)), we can obtain (3.10) by first sending $T \to \infty$ and then $\varepsilon \to 0$ provided that we can show that, for any fixed ε,

$$
\text{(3.12)} \qquad \lim_{T\to\infty} \left(C_\varepsilon(T) - C_0(T) \right) = 0,
$$

$$
\text{(3.13)} \qquad \lim_{T\to\infty} \left(\nu'(T) - \nu(T) \right) \chi(T) = 0.
$$

Hence, it remains to verify them. Using the definition of $C_\varepsilon(T)$, the estimate in (3.5) (with $k = 0$), and $\int u(T,x)\mathrm{d}x = 1$, we have

$$
|C_\varepsilon(T) - C_0(T)| \leqslant c_0\varepsilon \int_T^\infty |\nu(\tau) - \nu'(\tau)|\mathrm{d}\tau
$$

for some constant c_0. Since $\int_0^\infty |\nu(\tau) - \nu'(\tau)|\mathrm{d}\tau$ is finite as it is equal to $\mathsf{W}_1(\nu, \nu')$, the right-hand side vanishes as $T \to \infty$, which verifies (3.12). To see (3.13), we start by recalling that $|\chi(t)| \leqslant t$ as explained below Lemma 3.3. Hence, we have

$$
\left| \left(\nu'(T) - \nu(T) \right) \chi(T) \right| \leqslant T(1 - \nu'(T)) + T(1 - \nu(T)).
$$

Let X be a random variable with law ν. We have $T(1 - \nu(T)) = T\mathbb{P}(X > T) \leqslant \mathbb{E}[X\mathbb{1}_{X>T}]$. Since ν has finite first moment, the dominated convergence theorem implies $\lim_{T\to\infty} T(1 - \nu'(T)) = 0$. The other term can be treated similarly and thus we obtain (3.13). As explained previously, now that (3.12) and (3.13) are verified, the proof is complete. $\qquad\square$

Lemma 3.4 (Properties of χ). *Let $\nu \in \mathcal{P}_1(\mathbb{R}_+)$ and let $\chi : \mathbb{R}_+ \to \mathbb{R}$ be given as in (3.11). Then, $\chi(0) = 0$ and χ is increasing, 1-Lipschitz, and convex.*

Proof. Due to $\dot\chi \geqslant 0$ evidently from (3.11), χ is increasing. By Lemma 3.1, $|\partial_x\Phi_\nu| \leqslant 1$ everywhere. Using this and (3.9), we have $|\dot\chi| \leqslant 1$ and thus χ is 1-Lipschitz. To show that χ is convex, we compute $\ddot\chi$. For brevity of notation, we write $\Phi = \Phi_\nu$, $u = u_\nu$, and $\int = \int_\mathbb{R} \mathrm{d}x$. Using the equations satisfied by Φ and u as in (3.2) and (3.8) and integration by parts (IBP), we can compute

$$
\ddot\chi = \frac{\mathrm{d}}{\mathrm{d}t} \int (\partial_x\Phi)^2 u = \int \left(2\partial_t\partial_x\Phi(u\partial_x\Phi) + (\partial_x\Phi_x)^2 \partial_t u \right)
$$

$$
\overset{\text{(IBP)}}{=} \int \left(-2\partial_t\Phi\partial_x(u\partial_x\Phi) + (\partial_x\Phi)^2 \partial_t u \right)
$$

$$
\overset{(3.2)(3.8)}{=} \int \left(2\left(\partial_x^2\Phi + \nu(\partial_x\Phi)^2\right)\partial_x(u\partial_x\Phi) + (\partial_x\Phi)^2 \left(\partial_x^2 u - 2\nu\partial_x(u\partial_x\Phi)\right) \right)
$$

$$
= \int \left(2\partial_x^2\Phi\partial_x(u\partial_x\Phi) + (\partial_x\Phi)^2\partial_x^2 u \right) \overset{\text{(IBP)}}{=} \int 2(\partial_x^2\Phi)^2 u \overset{(3.9)}{\geqslant} 0.
$$

Therefore, χ is convex, completing the proof. $\qquad\square$

3.2. **Properties of ψ.** Using the decomposition in (1.8), we obtain the Lipschitzness, concavity, and differentiability of ψ as stated below. For each $p \geqslant 1$, let W_p be the Wasserstein L^p metric over probability measures on a Euclidean space that will be clear from the context.

Recall that for any $\nu \in \mathcal{P}(\mathbb{R}_+^D)$ and $d \in [D]$, we denote by $\nu_d \in \mathcal{P}(\mathbb{R}_+)$ its d-th marginal. Due to (2.15) and (2.18), we have, for all $\nu, \nu' \in \mathcal{P}_1(\mathbb{R}_+^D)$,

$$(3.14) \qquad \left|\psi(\nu) - \psi(\nu')\right| \leqslant \mathsf{W}_1(\nu, \nu').$$

Proposition 3.5 (Concavity of ψ)**.** *The function ψ is concave on $\mathcal{P}_1(\mathbb{R}_+^D)$. Furthermore, for every $\nu, \nu' \in \mathcal{P}_1(\mathbb{R}_+^D)$, if there exists $d \in [D]$ such that $\nu_d \neq \nu'_d$ and $\nu_d, \nu'_d \in \mathcal{P}_\infty(\mathbb{R}_+)$, then we have, for every $\lambda \in (0,1)$,*

$$(3.15) \qquad \psi\left(\lambda\nu' + (1-\lambda)\nu\right) > \lambda\psi(\nu') + (1-\lambda)\psi(\nu).$$

Proof. This follows from the decomposition in (1.8) and Lemma 3.2 stating that each ψ_d is concave on $\mathcal{P}_1(\mathbb{R}_+)$ and strictly concave on $\mathcal{P}_\infty(\mathbb{R}_+)$. $\qquad\square$

Note that the function ψ is not strictly concave on $\mathcal{P}_\infty(\mathbb{R}_+^D)$ since it is possible for $\nu, \nu' \in \mathcal{P}_\infty(\mathbb{R}_+^D)$ to satisfy $\nu \neq \nu'$ while $\nu_d = \nu'_d$ for every $d \in \{1, \ldots, D\}$. However, the following is true.

Lemma 3.6 (Monotone measures characterized by marginals)**.** *Let $\nu, \nu' \in \mathcal{P}^\uparrow(\mathbb{R}_+^D)$. Then, $\nu = \nu'$ if and only if $\nu_d = \nu'_d$ for all $d \in [D]$.*

Proof. By the definition of monotone measures (see the paragraph below (1.7)), there are two increasing maps $\mathsf{q}, \mathsf{q}' : [0,1] \to \mathbb{R}_+^D$ such that $\nu = \mathrm{Law}(\mathsf{q}(U))$ and $\nu' = \mathrm{Law}(\mathsf{q}(U))$ where U is a uniform random variable in $[0,1]$. Then, $\nu_d = \nu'_d$ is equivalent to $\mathsf{q}_d = \mathsf{q}'_d$ a.e. on $[0,1)$. The desired result follows from the simple observation that $\mathsf{q} = \mathsf{q}'$ if and only if $\mathsf{q}_d = \mathsf{q}'_d$ for all $d \in [D]$. $\qquad\square$

The set $\mathcal{P}_\infty^\uparrow(\mathbb{R}_+^D)$ is not convex when $D > 1$, so it does not make sense to talk about concavity or strict concavity of ψ on $\mathcal{P}_\infty^\uparrow(\mathbb{R}_+^D)$. Yet we have the following result, which is an immediate corollary of Proposition 3.5 and Lemma 3.6.

Corollary 3.7 ("Strict concavity" of ψ on $\mathcal{P}_\infty^\uparrow(\mathbb{R}_+^D)$)**.** *For any two distinct $\nu, \nu' \in \mathcal{P}_\infty^\uparrow(\mathbb{R}_+^D)$ and $\lambda \in (0,1)$, we have*

$$(3.16) \qquad \psi\left(\lambda\nu' + (1-\lambda)\nu\right) > \lambda\psi(\nu') + (1-\lambda)\psi(\nu).$$

We define
$$(3.17)$$
$$\mathfrak{X} = \left\{ \chi : \mathbb{R}_+^D \to \mathbb{R} \;\middle|\; \begin{array}{l} \chi(x) = \sum_{d=1}^D \lambda_{\infty,d}\chi_d(x_d), \; \forall x \in \mathbb{R}_+^D, \text{ where } \chi_d(0) = 0, \\ \chi_d \text{ is increasing, 1-Lipschitz, and convex on } \mathbb{R}_+ \end{array} \right\}.$$

Proposition 3.8 (Differentiability of ψ)**.** *For every $\nu \in \mathcal{P}_1(\mathbb{R}_+^D)$, there is $\chi \in \mathfrak{X}$ such that, for every $\nu' \in \mathcal{P}_1(\mathbb{R}_+^D)$, we have*

$$(3.18) \qquad \frac{\mathrm{d}}{\mathrm{d}\varepsilon}\psi\left((1-\varepsilon)\nu + \varepsilon\nu'\right)\Big|_{\varepsilon=0} = \int_{\mathbb{R}_+^D} \chi\,\mathrm{d}\left(\nu' - \nu\right).$$

Proof. This follows from the decomposition in (1.8) and Lemma 3.3 for the differentiability of each ψ_d, with Lemma 3.4 ensuring the desired properties of each of the χ_d in the decomposition of χ. $\square$

4. Models with two species

In this section, we present relatively simple proofs of Theorems 1.3 and Theorem 1.4 in the special case of models with two species ($D = 2$) with no external field ($\nabla\xi(0) = 0$) and $\mu = \delta_0$. The proofs we give later of the general versions of these theorems do not make any reference to the results of this section, so one can skip this entire section if so enclined. Nevertheless, we find it interesting to also present this comparatively simpler argument, and also to explain why it does not generalize to models with more than two species.

Given $\mu \in \mathcal{P}(\mathbb{R}_+^2)$, recall that we write μ_1 and μ_2 to denote its marginals. Letting F_1 and F_2 be the cumulative distribution functions of μ_1 and μ_2 respectively, and U be a uniform random variable on $[0,1]$, we set $\mu^\uparrow = \text{Law}(F_1^{-1}(U), F_2^{-1}(U))$. This defines a mapping from $\mathcal{P}(\mathbb{R}_+^2)$ to $\mathcal{P}^\uparrow(\mathbb{R}_+^2)$, and we observe that $\mu = \mu^\uparrow$ if $\mu \in \mathcal{P}^\uparrow(\mathbb{R}_+^2)$. The fundamental idea is the following observation.

Proposition 4.1. *Suppose that $D = 2$ and that $\nabla\xi(0) = 0$. For every $\mu \in \mathcal{P}(\mathbb{R}_+^2)$, we have*

$$(4.1) \qquad \int \xi^* \mathrm{d}\mu^\uparrow \leqslant \int \xi^* \mathrm{d}\mu.$$

As will be explained in more details shortly, Proposition 4.1 allows us to rewrite the supremum in (1.12) as

$$(4.2) \quad \sup_{\nu \in \mathcal{P}_\infty^\uparrow(\mathbb{R}_+^2)} \left\{ \psi(\nu) - \int (t\xi)^* \mathrm{d}\nu \right\} = \sup_{\nu \in \mathcal{P}_\infty(\mathbb{R}_+^2)} \left\{ \psi(\nu) - \int (t\xi)^* \mathrm{d}\nu^\uparrow \right\}$$

$$= \sup_{\nu \in \mathcal{P}_\infty(\mathbb{R}_+^2)} \left\{ \psi(\nu) - \int (t\xi)^* \mathrm{d}\nu \right\},$$

which is key to establishing Theorems 1.3 and 1.4 in this particular case. The key ingredient of the proof of Proposition 4.1 is the following.

Proposition 4.2. *Suppose that $D = 2$ and that $\nabla\xi(0) = 0$. For all real numbers $a \leqslant a'$ and $b \leqslant b'$, we have*

$$(4.3) \qquad \xi^*(a,b) + \xi^*(a',b') \leqslant \xi^*(a',b) + \xi^*(a,b').$$

The condition (4.3) is an integrated version of the claim that the mixed second derivatives of ξ^* are non-positive, i.e. $\partial_a\partial_b\xi^*(a,b) \leqslant 0$. Heuristically, the mappings $\nabla\xi$ and $\nabla\xi^*$ are inverse of one another, and thus the Hessian of ξ^* at (a,b) should be the inverse of the matrix $A = \nabla\xi^2(\nabla\xi^*(a,b))$. Moreover, the assumption that ξ must allow for the existence of a Gaussian random field satisfying (1.3) imposes some constraints on ξ [40, Proposition 6.6].

In particular, the Hessian of ξ must be a symmetric positive semi-definite matrix with nonnegative coefficients, and thus we should have

$$\partial_{ab}^2 \xi^*(a, b) = -\det(A)^{-1} A_{12} \leqslant 0.$$

Relying on this heuristic, we give a rigorous proof of Proposition 4.2 in Appendix A.

Proof of Proposition 4.1. This follows from Proposition 4.2 and [40, Proposition 2.5] applied to $c(a, b) = \xi^*(a, b)$. (We point out that the sentence "Examples of functions satisfying the condition (2.16) include any convex function of $x - y$." in [40, Remark 2.5] is actually false, as explained below some convex functions fail to satisfy [40, (2.16)].) □

Remark 4.3. We now give an example with $D = 3$ of an admissible function $\xi : \mathbb{R}^3 \to \mathbb{R}$ for which Proposition 4.1 is invalid. Consider $\xi(x) = \frac{x \cdot Ax}{2}$ where

$$A = \begin{pmatrix} 6 & 2 & 1 \\ 2 & 5 & 3 \\ 1 & 3 & 4 \end{pmatrix}.$$

The function $\xi : \mathbb{R}^3 \to \mathbb{R}$ thus defined satisfies

$$\mathbb{E} H_N(\sigma) H_N(\tau) = \xi\left(\frac{\sigma_1 \cdot \tau_1}{N}, \frac{\sigma_2 \cdot \tau_2}{N}, \frac{\sigma_3 \cdot \tau_3}{N}\right)$$

provided that we set

$$H_N(\sigma) = \frac{1}{\sqrt{N}} \sum_{i,j=1}^{N} \sum_{d,d'=1}^{3} (A_{dd'})^{1/2} J_{i,j}^{d,d'} \sigma_{di} \sigma_{d'j},$$

where the $J_{i,j}^{d,d'}$ are independent standard Gaussian random variables. We observe that A is positive definite, as can be checked using Sylvester's criterion on the positivity of the leading principal minors, so the function ξ is convex on $\mathbb{R}^3$. We have

$$A^{-1} = \frac{1}{57} \begin{pmatrix} 11 & -5 & 1 \\ -5 & 23 & -16 \\ 1 & -16 & 26 \end{pmatrix}.$$

Note that $(A^{-1})_{13} > 0$. Let $\overline{V} = \nabla\xi(\mathbb{R}_+^3)$, we have that $\nabla\xi(x) = Ax$ so $\overline{V} = \{Ax \mid x \in \mathbb{R}_+^3\}$. Using that $\xi^*(\nabla\xi(x)) = x \cdot \nabla\xi(x) - \xi(x)$ (proved below in (5.7) in any dimension), we see that for every $y \in \overline{V}$,

$$\xi^*(y) = \frac{y \cdot A^{-1} y}{2}.$$

In particular ξ^* is differentiable on the interior of $\overline{V}$ and $\nabla^2 \xi^*(y) = A^{-1}$. Now let (e_1, e_2, e_3) denote the canonical basis of $\mathbb{R}^3$ and let us fix y_0 in the interior of $\overline{V}$. For $\varepsilon > 0$ small enough we have that the points $y_0 + \varepsilon e_1$, $y_0 + \varepsilon e_3$ and

$y_0 + \varepsilon e_1 + \varepsilon e_3$ all belong to the interior of $\overline{V}$. Let μ be the uniform measure on $\{y_0, y_0 + \varepsilon e_1, y_0 + \varepsilon e_3\}$, we have

$$\mu^{\uparrow} = \frac{1}{3}\left(2\delta_{y_0} + \delta_{y_0+\varepsilon e_1+\varepsilon e_3}\right).$$

In particular,

$$\int \xi^* d\mu^{\uparrow} - \int \xi^* d\mu$$
$$= \frac{1}{3}\left(\xi^*(y_0) + \xi^*(y_0 + \varepsilon e_1 + \varepsilon e_3) - \xi^*(y_0 + \varepsilon e_1) - \xi^*(y_0 + \varepsilon e_1)\right)$$
$$= \frac{1}{3}\int_0^1\int_0^1 \partial_{13}\xi^*(y_0 + t\varepsilon e_1 + s\varepsilon e_3)dtds.$$

Since $\partial_{13}\xi^* > 0$ on the interior of V, we obtain

$$\int \xi^* d\mu^{\uparrow} > \int \xi^* d\mu,$$

as announced. $\qquad\qquad\square$

We now prove Theorem 1.3 when $D = 2$, $\nabla\xi(0) = 0$ and $\mu = \delta_0$ by relying on Proposition 4.1. As a consequence, we will be able to deduce that Theorem 1.4 also holds at $D = 2$ and $\mu = \delta_0$.

Recall from (2.17) that every $\nu \in \mathcal{P}_1^{\uparrow}(\mathbb{R}_+^2)$, we have

$$(4.4) \qquad\qquad \psi(\nu) = \lambda_{\infty,1}\psi_1(\nu_1) + \lambda_{\infty,2}\psi_2(\nu_2),$$

where ψ_d is the Parisi functional associated to $\boldsymbol{\pi}_d$. The right-hand side of the previous display is well defined even when ν is not monotone, in particular this gives an extension of the functional ψ to $\mathcal{P}_1(\mathbb{R}_+^2)$.

Proof of Theorem 1.3 when $D = 2$, $\nabla\xi(0) = 0$ and $\mu = \delta_0$. Let $\nu \in \mathcal{P}_{\infty}(\mathbb{R}_+^2)$. Since ν and $\nu^{\uparrow}$ have the same marginals, we have $\psi(\nu) = \psi(\nu^{\uparrow})$. Thus according to Proposition 4.1 we have

$$\psi(\nu) - \int \xi^* d\nu \leqslant \psi(\nu^{\uparrow}) - \int \xi^* d\nu^{\uparrow}.$$

This implies

$$\sup_{\nu\in\mathcal{P}_{\infty}(\mathbb{R}_+^2)}\left\{\psi(\nu) - \int \xi^* d\nu\right\} \leqslant \sup_{\nu\in\mathcal{P}_{\infty}^{\uparrow}(\mathbb{R}_+^2)}\left\{\psi(\nu) - \int \xi^* d\nu\right\}.$$

Since the other bound is trivial,

$$\sup_{\nu\in\mathcal{P}_{\infty}(\mathbb{R}_+^2)}\left\{\psi(\nu) - \int \xi^* d\nu\right\} = \sup_{\nu\in\mathcal{P}_{\infty}^{\uparrow}(\mathbb{R}_+^2)}\left\{\psi(\nu) - \int \xi^* d\nu\right\}.$$

We use Theorem 1.1 to conclude that (1.14) holds at $\mu = \delta_0$.

From Lemma 3.2, we have that ψ_1 and ψ_2 are (strictly) concave on $\mathcal{P}_{\infty}(\mathbb{R}_+)$ and the map $\nu \mapsto \int \xi^* d\nu$ is affine, thus the function inside the supremum in (1.14) is concave in ν.

Finally, by contradiction, assume that there are two distinct monotone probability measures ν and ν' that reach the supremum in (1.14). Consider

$\nu'' = \frac{\nu+\nu'}{2} \in \mathcal{P}_\infty(\mathbb{R}_+^2)$ (note that ν'' may not be monotone). Since $\nu \neq \nu'$ and $\nu, \nu' \in \mathcal{P}^\uparrow(\mathbb{R}_+^2)$, we must have $\nu_1 \neq \nu_1'$ or $\nu_2 \neq \nu_2'$. Since ψ_1 and ψ_2 are strictly concave on $\mathcal{P}_\infty(\mathbb{R}_+)$, this yields

$$\psi(\nu'') = \psi_1\left(\frac{\nu_1 + \nu_1'}{2}\right) + \psi_2\left(\frac{\nu_2 + \nu_2'}{2}\right)$$
$$> \frac{\psi_1(\nu_1) + \psi_1(\nu_1')}{2} + \frac{\psi_2(\nu_2) + \psi_2(\nu_2')}{2}$$
$$= \frac{\psi(\nu) + \psi(\nu')}{2}.$$

By (1.14) at $\mu = \delta_0$, we thus have

$$\lim_{N\to+\infty} \overline{F}_N(t,\delta_0) = \frac{1}{2}\left(\psi(\nu) - \int \xi^* d\nu + \psi(\nu') - \int \xi^* d\nu'\right)$$
$$< \psi(\nu'') - \int \xi^* d\nu''$$
$$\leq \sup_{\rho \in \mathcal{P}_\infty(\mathbb{R}_+^2)} \left\{\psi(\rho) - \int \xi^* d\rho'\right\}$$
$$= \lim_{N\to+\infty} \overline{F}_N(t,\delta_0).$$

This is a contradiction. Therefore there is exactly one probability measure ν that reaches the supremum in (1.14). $\qquad\square$

Recall the definition of $\mathfrak{X}$ from (3.17). For every $\chi \in \mathfrak{X}$, we define $S_t\chi$ according to (1.17).

Proof of Theorem 1.4 when $D = 2$, $\nabla\xi(0) = 0$, and $\mu = \delta_0$. We assume, without loss of generality, that $t = 1$. According to [31, Lemma 4.1], we have that

(4.5) $$\psi_d(\nu_d) = \inf_{\chi_d}\left\{\int \chi_d d\nu_d - (\psi_d)_*(\chi_d)\right\},$$

where the infimum is taken over all $\chi_d : \mathbb{R}_+ \to \mathbb{R}$ that are 1-Lipschitz, increasing, convex functions that vanish at the origin and where we have defined

$$(\psi_d)_*(\chi_d) = \inf_{\nu_d \in \mathcal{P}_1(\mathbb{R}_+)}\left\{\int \chi_d d\nu_d - \psi_d(\nu_d)\right\}.$$

It then follows that

$$\psi(\nu) = \psi_1(\nu_1) + \psi_2(\nu_2)$$
$$= \inf_{\chi_1}\left\{\int \chi_1 d\nu_1 - (\psi_1)_*(\chi)\right\} + \inf_{\chi_2}\left\{\int \chi_2 d\nu_2 - (\psi_2)_*(\chi)\right\}$$
$$= \inf_{\chi \in \mathfrak{X}}\left\{\int \chi d\nu - \psi_*(\chi)\right\}.$$

Plugging this into the formula of Theorem 1.3 at $\mu = \delta_0$, we obtain

$$\lim_{N\to+\infty} \overline{F}_N(t,\delta_0) = \sup_{\mu\in\mathcal{P}_\infty(\mathbb{R}^2_+)} \inf_{\chi\in\mathfrak{X}} \left\{ \int \chi \mathrm{d}\nu - \psi_*(\chi) + \int \xi^* \mathrm{d}\nu \right\}.$$

The sets $\mathcal{P}_\infty(\mathbb{R}^2_+)$ and $\mathfrak{X}$ are both convex. In addition, $\mathfrak{X}$ is compact with respect to convergence in the local uniform topology. Furthermore, for every $\nu \in \mathcal{P}_\infty(\mathbb{R}^2_+)$, the map $\chi \mapsto \int (\chi - \xi^*)\mathrm{d}\nu - \psi_*(\chi)$ is convex on $\mathfrak{X}$ and for every $\chi \in \mathfrak{X}$, the map $\nu \mapsto \int (\chi - \xi^*)\mathrm{d}\nu - \psi_*(\chi)$ is concave on $\mathcal{P}_\infty(\mathbb{R}^2_+)$. According to Sion's min-max theorem [55], we obtain

$$\lim_{N\to+\infty} \overline{F}_N(t,\delta_0) = \inf_{\chi\in\mathfrak{X}} \sup_{\nu\in\mathcal{P}_\infty(\mathbb{R}^2_+)} \left\{ \int \chi \mathrm{d}\nu - \psi_*(\chi) + \int \xi^* \mathrm{d}\nu \right\}.$$

According to (1.17), we have

$$\sup_{\nu\in\mathcal{P}_\infty(\mathbb{R}^2_+)} \left\{ \int (\chi - \xi^*)\mathrm{d}\nu \right\} = \sup_{x\in\mathbb{R}^2_+} \{\chi(x) - \xi^*(x)\} = S_1\chi(0).$$

we obtain

$$\lim_{N\to+\infty} \overline{F}_N(t,\delta_0) = \inf_{\chi\in\mathfrak{X}} \{S_1\chi(0) - \psi_*(\chi)\},$$

as desired. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \square$

5. Stability of $\mathbb{R}^D_+$-convexity under HJ semigroup

We now return to the general setting where D and $\nabla\xi(0)$ are arbitrary, as is the case everywhere except in the previous section. For every $\chi : \mathbb{R}^D_+ \to \mathbb{R}$, $t \geqslant 0$ and $x \in \mathbb{R}^D_+$, we define

$$(5.1) \qquad S_t\chi(x) = \sup_{y\in\mathbb{R}^D_+} \{\chi(y) - (t\xi)^*(y-x)\},$$

as was announced in (1.17). The fact that this quantity is equal to that in (1.16) for every $\chi \in \mathfrak{X}$ will be a consequence of Lemma 5.12 proved below. The first result of this section is an analysis of the optimizers in (5.1).

Proposition 5.1. *Assume that ξ^* is differentiable on $\mathbb{R}^D_+$. Let $\chi : \mathbb{R}^D_+ \to \mathbb{R}$ be a Lipschitz function. Let $t > 0$ and $x, y \in \mathbb{R}^D_+$ be such that*

$$(5.2) \qquad \chi(y) - S_t\chi(x) - (t\xi)^*(y-x) = 0.$$

If $x \in (0, +\infty)^D$ and x is a differentiable point of $S_t\chi$, then we have

$$(5.3) \qquad \nabla S_t\chi(x) = \nabla(t\xi)^*(y-x).$$

If $y \in (0, +\infty)^D$ and y is a differentiable point of χ, then we have

$$(5.4) \qquad \nabla\chi(y) = \nabla(t\xi)^*(y-x).$$

In particular, if the two conditions hold at the same time, then we have

$$(5.5) \qquad \nabla S_t\chi(x) = \nabla\chi(y).$$

Before showing this, we derive in the next three lemmas basic properties of ξ and ξ^* that will be useful throughout the rest of the paper. We say that a function $\varsigma : \mathbb{R}^D \to \mathbb{R}$ is *superlinear on* $\mathbb{R}^D_+$ provided that, for every $M > 0$, there is $R > 0$ such that

$$(5.6) \qquad \inf_{x \in \mathbb{R}^D_+ : |x| \geqslant R} \frac{\varsigma(x)}{|x|} \geqslant M.$$

Lemma 5.2. *The function ξ^* is superlinear on $\mathbb{R}^D_+$. In addition, if we assume that ξ is superlinear on $\mathbb{R}^D_+$, then ξ^* is also continuous. The same holds for $(t\xi)^*$ for any $t > 0$.*

Proof. The function ξ is continuous so it is locally bounded. For every $r > 0$, let M_r denote the supremum of $|\xi|$ on the ball of radius r centered at 0. Letting $y \in \mathbb{R}^D_+$ and setting $x = ry/|y|$ in the supremum below, we get that

$$\xi^*(y) = \sup_{x \in \mathbb{R}^D_+} \{x \cdot y - \xi(x)\} \geqslant r|y| - M_r.$$

Then (5.6) follows from the last display. Let us now further assume that ξ is superlinear. In this case, we have that for every $y \in \mathbb{R}^D_+$,

$$\lim_{\substack{|x| \to +\infty \\ x \in \mathbb{R}^D_+}} (x \cdot y - \xi(x)) = -\infty.$$

This implies that $\xi^* < +\infty$ on $\mathbb{R}^D$. Hence, ξ^* is a convex function on $\mathbb{R}^D$ taking only finite values and thus ξ^* is continuous (see e.g. [20, Proposition 2.9]). $\square$

Lemma 5.3. *If ξ is superlinear on $\mathbb{R}^D_+$ and strictly convex on $\mathbb{R}^D_+$, then ξ^* is differentiable on $\mathbb{R}^D$.*

Proof. Let $M > 0$ be arbitrary and let $R > 0$ be such that (5.6) holds for ξ. Let $y \in \mathbb{R}^D$ be such that $|y| \leqslant M$. For every $|x| \geqslant R$, we have $x \cdot y - \xi(x) \leqslant |x|(|y| - M) \leqslant 0$. Hence,

$$\xi^*(y) = \sup_{|x| \leqslant R} \{x \cdot y - \xi(x)\}.$$

Since ξ is assumed to be strictly convex, the variational problem in the previous display admits exactly one maximizer. It then follows from the envelope theorem [20, Theorem 2.21] that ξ^* is differentiable on $\mathbb{R}^D$. $\square$

Lemma 5.4. *For every $x \in \mathbb{R}^D_+$, we have*

$$(5.7) \qquad \xi^*(\nabla\xi(x)) = x \cdot \nabla\xi(x) - \xi(x).$$

Furthermore, if ξ is strongly convex on $\mathbb{R}_+^D$, then we have for every $x, y \in \mathbb{R}_+^D$ that

$$(5.8) \qquad \nabla \xi^*(\nabla \xi(x)) = x,$$

$$(5.9) \qquad \nabla \xi(\nabla \xi^*(y)) - y \in \mathbb{R}_+^D,$$

$$(5.10) \qquad \nabla \xi^*(y) \cdot (\nabla \xi(\nabla \xi^*(y)) - y) = 0,$$

$$(5.11) \qquad \xi^*(\nabla \xi(\nabla \xi^*(y))) = \xi^*(y).$$

Remark 5.5. When ξ is strongly convex on $\mathbb{R}_+^D$, note that thanks to (5.8) we have in particular that for every $y \in \nabla \xi(\mathbb{R}_+^D)$,

$$(5.12) \qquad \nabla \xi(\nabla \xi^*(y)) = y.$$

Proof. Let $x \in \mathbb{R}_+^D$, we start by proving (5.7). By definition of ξ^*, it is clear that the left-hand side is larger than the right-hand side. To prove the other inequality, let $x' \in \mathbb{R}_+^D$, by convexity of ξ we have

$$\xi(\lambda x' + (1 - \lambda)x) \leqslant \lambda \xi(x') + (1 - \lambda)\xi(x).$$

Rearranging and letting $\lambda \to 0$ we get $\nabla \xi(x) \cdot (x' - x) \leqslant \xi(x') - \xi(x)$, in particular

$$x' \cdot \nabla \xi(x) - \xi(x') \leqslant x \cdot \nabla \xi(x) - \xi(x).$$

Taking the supremum over $x' \in \mathbb{R}_+^D$ in the previous display, we obtain $\xi^*(\nabla \xi(x)) \leqslant x \cdot \nabla \xi(x) - \xi(x)$, which thus completes the proof of (5.8).

Since ξ is assumed to be strongly on $\mathbb{R}_+^D$, it is superlinear on $\mathbb{R}_+^D$ and strictly convex on $\mathbb{R}_+^D$, so by Lemma 5.3, the function ξ^* is differentiable on $\mathbb{R}^D$. For $x \in \mathbb{R}_{++}^D$ we can differentiate (5.7) to obtain

$$\nabla^2 \xi(x) \nabla \xi^*(\nabla \xi(x)) = \nabla^2 \xi(x)x.$$

Since ξ is assumed to be strongly convex on $\mathbb{R}_+^D$, the matrix $\nabla^2 \xi(x)$ is positive definite and the previous display implies (5.8) for $x \in \mathbb{R}_{++}^D$, which we can then extend to $x \in \mathbb{R}_+^D$ by continuity.

We now turn to proving (5.9) and (5.10). Let $y \in \mathbb{R}_{++}^D$, by definition we have

$$\xi^*(y) = \sup_{x \in \mathbb{R}_+^D} \{x \cdot y - \xi(x)\}.$$

Let $M > 0$ and assume that $|y| \leqslant M$. Since ξ is assumed to be superlinear, there exists $R > 0$ such that (5.6) holds for ξ and we have that for $|x| > R$,

$$x \cdot y - \xi(x) \leqslant |x|(|y| - M) < 0.$$

Therefore,

$$\xi^*(y) = \sup_{\substack{x \in \mathbb{R}_+^D \\ |x| \leqslant R}} \{x \cdot y - \xi(x)\}.$$

In addition since ξ is assumed to be strongly convex on $\mathbb{R}_+^D$, there is a unique maximizer $x_{\mathrm{opt}}(y) \in \mathbb{R}_+^D$ in the variational formula of the above

display. According to the envelope theorem [20, Theorem 2.21] we have $\nabla \xi^*(y) = x_{\mathrm{opt}}(y)$. It follows that for every $x \in \mathbb{R}_+^D$,

$$\xi(x_{\mathrm{opt}}(y)) = x_{\mathrm{opt}}(y) \cdot y - \xi^*(y)$$
$$\leqslant x_{\mathrm{opt}}(y) \cdot y - (x \cdot y - \xi(x)).$$

Rearranging, we obtain

$$\xi(x) \geqslant \xi(x_{\mathrm{opt}}(y)) + y \cdot (x - x_{\mathrm{opt}}(y)).$$

Chosing $x = x_{\mathrm{opt}}(y) + t e_d$ for $t > 0$ and letting $t \to 0$ in the previous display, we obtain (5.9). Furthermore, if $x_{\mathrm{opt},d}(y) > 0$ for some $d \in [D]$, then we can also take $t < 0$ in the previous construction (provided that t is small enough) and this yields $\partial_{x_d} \xi(x_{\mathrm{opt}}(y)) = y_d$. Therefore for every $d \in [D]$, we have that the d-th coordinate of $\nabla \xi^*(y)$ is 0 or the d-th coordinate of $\nabla \xi(\nabla \xi^*(y)) - y$ is 0, and this yields (5.10) for $y \in \mathbb{R}_{++}^D$. By continuity, we obtain (5.9) and (5.10) as announced.

Finally, by a similar proof as for (5.7), we have for every $y \in \mathbb{R}_+^D$ that

$$\xi(\nabla \xi^*(y)) = y \cdot \nabla \xi^*(y) - \xi^*(y).$$

From this, (5.7) and (5.10), we deduce that

$$\begin{aligned}
\xi^*(\nabla \xi(\nabla \xi^*(y))) &= \nabla \xi^*(y) \cdot \nabla \xi(\nabla \xi^*(y)) - \xi(\nabla \xi^*(y)) \\
&= \nabla \xi^*(y) \cdot \nabla \xi(\nabla \xi^*(y)) - (y \cdot \nabla \xi^*(y) - \xi^*(y)) \\
&= \xi^*(y) + \nabla \xi^*(y) \cdot (\nabla \xi(\nabla \xi^*(y)) - y) \\
&= \xi^*(y),
\end{aligned}$$

and thus, (5.11) is proved.   □

In the proof of Proposition 5.1 as well as later on, we make use of the following simple fact:

$$(5.13) \qquad\qquad x, y \in \mathbb{R}_+^D \implies |y - x| - |(y - x)_+| \leqslant |x|.$$

Indeed, we have

$$|y - x| - |(y - x)_+| \leqslant |y - x - (y - x)_+| \leqslant \Big(\sum_{d=1}^D (y_d - x_d)^2 \mathbb{1}_{0 \leqslant y_d < x_d} \Big)^{\frac{1}{2}} \leqslant |x|.$$

We also need the following definition.

Definition 5.6 ($\mathbb{R}_+^D$-increasing)**.** For any $k \in \mathbb{N}$, we say that a function $\chi : \mathbb{R}_+^D \to \mathbb{R}^k$ is $\mathbb{R}_+^D$-*increasing* when for every $x, y \in \mathbb{R}_+^D$ we have

$$(5.14) \qquad\qquad\qquad \chi(x + y) \geqslant \chi(x).$$

The inequality (5.14) is interpreted in the sense that $\chi(x + y) - \chi(x) \in \mathbb{R}_+^k$. We may at times also make use of this notion for functions that are defined almost everywhere, in which case we ask that (5.14) be satisfied for almost every $x, y \in \mathbb{R}_+^D$.

We record the following last preparatory step, which will be used more crucially in the proof of Proposition 5.10 below. For every $y \in \mathbb{R}^D$, we set

$$(5.15) \qquad y_+ = (y_d \mathbb{1}_{y_d \geqslant 0})_{1 \leqslant d \leqslant D} \,.$$

Lemma 5.7. *Let* $\varsigma : \mathbb{R}_+^D \to \mathbb{R}$ *be an* $\mathbb{R}_+^D$*-increasing function, and for every* $y \in \mathbb{R}^D$, *let*

$$(5.16) \qquad \varsigma^*(y) = \sup_{x \in \mathbb{R}_+^D} \{x \cdot y - \varsigma(x)\} \,.$$

For every $y \in \mathbb{R}^D$, *we have* $\varsigma^*(y) = \varsigma^*(y_+)$.

Proof. Let $y \in \mathbb{R}^D$. Since the supremum in the definition of ς^* in (5.16) is taken over $\mathbb{R}_+^D$, we have that ς^* is $\mathbb{R}_+^D$-increasing and thus $\varsigma^*(y) \leqslant \varsigma^*(y_+)$. For each $x \in \mathbb{R}_+^D$, we set $\widehat{x} = (x_d \mathbb{1}_{y_d \geqslant 0})_{1 \leqslant d \leqslant D}$. We observe that $\widehat{x} \cdot y = x \cdot y_+$. Moreover, since ς is $\mathbb{R}_+^D$-increasing, we have $\varsigma(x) \geqslant \varsigma(\widehat{x})$. Hence, for every $x \in \mathbb{R}_+^D$, we have

$$x \cdot y_+ - \varsigma(x) \leqslant \widehat{x} \cdot y - \varsigma(\widehat{x}) \leqslant \varsigma^*(y),$$

which implies $\varsigma^*(y_+) \leqslant \varsigma^*(y)$. $\qquad\square$

Proof of Proposition 5.1. Assume that $x \in (0, +\infty)^D$ and that x is a differentiable point of $S_t \chi$. Using that $(t\xi)^*$ is superlinear on $\mathbb{R}_+^D$ (Lemma 5.2), Lemma 5.7, and (5.13), we can verify that, for each fixed $x \in \mathbb{R}_+^D$, $(t\xi)^*(\cdot - x)$ is superlinear on $\mathbb{R}_+^D$. Hence, for x' sufficiently close to x, we can see that the variational formula (5.1) for $S_t \chi(x')$ attains its maximum in a compact set. By (5.2), y is a maximizer of the formula for $S_t \chi(x)$. Hence, we can apply the envelope theorem (e.g. see [20, Theorem 2.21]) to see that $\nabla_x S_t \chi(x) = \nabla_x (\chi(y) - (t\xi)^*(y - x))$, which gives (5.3).

Next, assume that $y \in (0, +\infty)^D$ is a differentiable point of χ. Since y is a maximizer of the formula (5.1) for $S_t \chi(x)$, we have

$$\nabla_y (\chi(y) - (t\xi)^*(y - x)) = 0,$$

which gives (5.4). Lastly, (5.5) immediately follows from (5.3) and (5.4). $\quad\square$

We now turn to the main objective of this section. As discussed around (1.30), we will be interested in the case when χ is the derivative of ψ at some measure $\overline{\nu}$, and the main technical ingredients needed in the proofs of our main results involve certain monotonicity properties of $\nabla \chi$ and $\nabla S_t \chi$. More precisely, we are particularly interested in asserting that χ and $S_t \chi$ are $\mathbb{R}_+^D$-convex, where the notion of $\mathbb{R}_+^D$-convexity is defined as follows.

Definition 5.8 ($\mathbb{R}_+^D$-convexity). We say that a function $\chi : \mathbb{R}_+^D \to \mathbb{R}$ is $\mathbb{R}_+^D$-*convex* when for every $x, y, z \in \mathbb{R}_+^D$ we have

$$(5.17) \qquad \chi(x + y + z) - \chi(x + z) \geqslant \chi(x + y) - \chi(x).$$

When $\chi : \mathbb{R}_+^D \to \mathbb{R}$ is Lipschitz, one can check that χ is $\mathbb{R}_+^D$-convex if and only if $\nabla \chi$ is $\mathbb{R}_+^D$-increasing.

Recalling the definition of $\mathfrak{X}$ from (3.17), we already know from Proposition 3.8 that the derivative of ψ at any point in $\mathcal{P}_1(\mathbb{R}_+^D)$ is a function χ that belongs to $\mathfrak{X}$, and this implies in particular that χ is indeed $\mathbb{R}_+^D$-convex. The main result of this section is as follows.

Proposition 5.9. *For every $\chi \in \mathfrak{X}$ and $t \geqslant 0$, the function $S_t \chi$ is $\mathbb{R}_+^D$-convex.*

In order to prove Proposition 5.9, we will rely on the fact that the mapping $(t, x) \mapsto S_t \chi(x)$ can be interpreted as the solution to a Hamilton–Jacobi equation. We find it clearer to state some results in a slightly more general setting. Let $H : \mathbb{R}_+^D \to \mathbb{R}$ be a locally Lipschitz function and $\phi : \mathbb{R}_+^D \to \mathbb{R}$ be a Lipschitz function. Thanks to [15], we know that when ϕ and H are $\mathbb{R}_+^D$-increasing, the Hamilton–Jacobi equation

$$(5.18) \qquad \begin{cases} \partial_t u - H(\nabla u) = 0 \ \text{ on } (0, +\infty) \times \mathbb{R}_+^D \\ u(0, \cdot) = \phi \end{cases}$$

admits a unique viscosity solution. Moreover, by [15, Proposition 6.2], when H is further assumed to be convex and bounded from below, this unique viscosity solution can be represented as the following version of the Hopf-Lax formula

$$(5.19) \qquad u(t, x) = \sup_{y \in \mathbb{R}_+^D} \{ \phi(x + y) - (tH)^*(y) \},$$

where $(tH)^*(x) = \sup_{p \in \mathbb{R}_+^D} \{ x \cdot p - tH(p) \}$ denotes the convex dual of tH as defined in (5.16). We show that, under exactly the same assumptions on H and ϕ, one can represent the solution using another version of the Hopf-Lax representation.

Proposition 5.10. *Let $\phi : \mathbb{R}_+^D \to \mathbb{R}$ be an $\mathbb{R}_+^D$-increasing and Lipschitz function, and let $H : \mathbb{R}_+^D \to \mathbb{R}$ be a locally Lipschitz function that is $\mathbb{R}_+^D$-increasing, convex and bounded below. The unique $\mathbb{R}_+^D$-increasing and Lipschitz viscosity solution u of*

$$(5.20) \qquad \begin{cases} \partial_t u - H(\nabla u) = 0 \ \text{ on } (0, +\infty) \times \mathbb{R}_+^D, \\ u(0, \cdot) = \phi, \end{cases}$$

satisfies, for every $(t, x) \in [0, +\infty) \times \mathbb{R}_+^D$,

$$(5.21) \qquad u(t, x) = \sup_{x' \in \mathbb{R}_+^D} \{ \phi(x') - (tH)^*(x' - x) \}.$$

Proof. We let RHS denote the right-hand side in (5.21). It is clear from the standard Hopf-Lax representation in (5.19) that $u \leqslant RHS$. Let $x' \in \mathbb{R}_+^D$, we let $y \in \mathbb{R}_+^k$ denote the vector whose coordinates are positive parts of the coordinates of $x' - x$. We have $(tH)^*(y) = (tH)^*(x' - x)$ by Lemma 5.7, and

since $(x + y) - x' \in \mathbb{R}_+^D$, by monotonicity of ϕ we have $\phi(x') \leqslant \phi(x + y)$. This yields

$$u(t, x) \geqslant \phi(x + y) - (tH)^*(y) \geqslant \phi(x') + (tH)^*(x' - x).$$

Taking the supremum over $x' \in \mathbb{R}_+^D$ in the previous display, we obtain that $u(t, x) \geqslant RHS$, and thus the result. $\qquad\square$

Remark 5.11. The condition (1.3) puts some constraint on the function ξ. In fact, as proved in [40, Proposition 6.6], the coefficients in the power series expansion of ξ must be non-negative. This implies that both ξ and $\nabla\xi$ are $\mathbb{R}_+^D$-increasing on $\mathbb{R}_+^D$ and that ξ is $\mathbb{R}_+^D$-convex on $\mathbb{R}_+^D$. In particular, we can indeed apply Proposition 5.10 to $H = \xi_{|_{\mathbb{R}_+^D}}$ and $\phi = \chi$ for any $\chi \in \mathfrak{X}$. $\qquad\square$

A first consequence of this remark and the previous proposition is a simple proof of the equality between (1.16) and (1.17).

Lemma 5.12. *Let* $\chi : \mathbb{R}_+^D \to \mathbb{R}$ *be a Lipschitz, $\mathbb{R}_+^D$-increasing and convex function. We have*

$$(5.22) \qquad \sup_{y \in \mathbb{R}_+^D} \{\chi(y) - (t\xi)^*(y - x)\} = \sup_{p \in \mathbb{R}_+^D} \{p \cdot x + t\xi(p) - \chi^*(p)\}.$$

Proof. The Fenchel-Moreau theorem states that a function $h : \mathbb{R}^D \to \mathbb{R}$ is convex and lower semi-continuous if and only it can be written as the supremum of a family of affine functions. In [14], a similar result has been proved for functions on $\mathbb{R}_+^D$, more precisely a function $h : \mathbb{R}_+^D \to \mathbb{R}$ is convex, lower semi-continuous and $\mathbb{R}_+^D$-increasing if and only if it satisfies

$$h(x) = \sup_{p \in \mathbb{R}_+^D} \{x \cdot p - h^*(p)\},$$

where h^* is defined as in (5.16). From this, we have that

$$\xi(x) = \sup_{p \in \mathbb{R}_+^D} \{x \cdot p - \xi^*(p)\},$$

$$\chi(x) = \sup_{p \in \mathbb{R}_+^D} \{x \cdot p - \chi^*(p)\},$$

where again ξ^* and χ^* are defined as in (5.16). Thus, we can write

$$\sup_{p \in \mathbb{R}_+^D} \{p \cdot x + t\xi(p) - \chi^*(p)\}$$

$$= \sup_{p \in \mathbb{R}_+^D} \sup_{x' \in \mathbb{R}_+^D} \{p \cdot (x + x') - (t\xi)^*(x') - \chi^*(p)\}$$

$$= \sup_{x' \in \mathbb{R}_+^D} \sup_{p \in \mathbb{R}_+^D} \{p \cdot (x + x') - (t\xi)^*(x') - \chi^*(p)\}$$

$$= \sup_{x' \in \mathbb{R}_+^D} \{\chi(x + x') - (t\xi)^*(x')\}.$$

The last line is then equal to the left side of (5.22) thanks to Proposition 5.10 and (5.19). $\qquad\square$

Another useful consequence of Proposition 5.10 is the following result.

Lemma 5.13 (Preservation of Lipschitzness by S_t). *Let $\chi : \mathbb{R}_+^D \to \mathbb{R}$ be $\mathbb{R}_+^D$-increasing and Lipschitz. Let $(S_t)_{t \in \mathbb{R}_+}$ be the semigroup introduced in* (5.1) *and set $C = \sup_{z \in \mathbb{R}_+^D : |z| \leqslant \|\chi\|_{\mathrm{Lip}}} |\xi(z)|$. For every $t, t' \in \mathbb{R}_+$ and $x, x' \in \mathbb{R}_+^D$, we have*

$$(5.23) \qquad \left| S_t \chi(x) - S_{t'} \chi(x') \right| \leqslant C \left| t - t' \right| + \|\chi\|_{\mathrm{Lip}} \left| x - x' \right|.$$

Proof. By Proposition 5.10 and Remark 5.11, the mapping $(t, x) \mapsto S_t \chi(x)$ is the unique $\mathbb{R}_+^D$-increasing and Lipschitz viscosity solution of the equation described therein with the choices of $H = \xi\big|_{\mathbb{R}_+^D}$ and $\phi = \chi$. The desired result thus follows from [15, Theorem 1.2 (2a)]. $\qquad \square$

Note that when $D = 1$ the notion of $\mathbb{R}_+^D$-convexity coincides with the usual notion of convexity. In this case it is clear from the alternative representation of $S_t \chi$ obtained in Lemma 5.12 that, for every $\chi \in \mathfrak{X}$, we have that $S_t \chi$ is convex in the usual sense. In particular, when $D = 1$ Proposition 5.9 holds. More generally, the following characterization of $\mathbb{R}_+^D$-convexity holds.

Lemma 5.14. *Let $\chi : \mathbb{R}_+^D \to \mathbb{R}$ be a continuous function that is $\mathcal{C}^2$ on $(0, +\infty)^D$. We have that χ is $\mathbb{R}_+^D$-convex if and only if for every $x \in (0, +\infty)^D$, $\nabla^2 \chi(x) \in \mathbb{R}_+^{D \times D}$.*

Proof. Assume that χ is $\mathbb{R}_+^D$-convex. Applying (5.17) to (x, ty, sy) and letting $s \to 0$ and $t \to 0$, we obtain that for every $x \in (0, +\infty)^D$ and every $y, z \in \mathbb{R}_+^D$,

$$y \cdot \nabla^2 \chi(x) z \geqslant 0.$$

This exactly means that $\nabla^2 \chi(x) \in \mathbb{R}_+^{D \times D}$. Conversely, for $x \in (0, +\infty)^D$ and $y, z \in \mathbb{R}_+^D$ we have

$$(\chi(x + y + z) - \chi(x + z)) - (\chi(x + y) - \chi(x)) = \int_0^1 \int_0^1 y \cdot \nabla^2 \chi(x + ty + sz) z \, \mathrm{d}s \mathrm{d}t.$$

Thus, if $\nabla^2 \chi$ is $\mathbb{R}_+^{D \times D}$-valued, we obtain that the right-hand side of the previous display is nonnegative and thus (5.17) holds. We then obtain (5.17) for $x \in \mathbb{R}_+^D$ by continuity. $\qquad \square$

To prove Proposition 5.9 we will make use of the following idea. Assume that the map $u : (t, x) \mapsto S_t \chi(x)$ is smooth and let $w = \nabla^2 u$ denote its Hessian. We have

$$\begin{cases} \partial_t w = \nabla w \nabla \xi(\nabla u) + w \nabla^2 \xi(\nabla u) w \text{ on } (0, +\infty) \times \mathbb{R}_+^D \\ w(0, \cdot) = \nabla^2 \chi. \end{cases}$$

By Lemma 5.14, we have that $w(0, \cdot)$ is $\mathbb{R}_+^{D \times D}$-valued. Furthermore, it can be checked that the PDE in the previous display preserves positivity of the coefficients of its solutions. Indeed, on the right-hand side the first term is a plain transport term and the second term is of the form $f(t, x, w(t, x))$

with $f(t, x, z) \in \mathbb{R}_+^{D \times D}$ when $z \in \mathbb{R}_+^{D \times D}$, this is because the function ξ itself is $\mathbb{R}_+^D$-convex.

To make this observation rigorous we will have to pay attention to several details. First, we will extend ξ and χ in order to work with PDEs on $\mathbb{R}^D$ and avoid boundary problems. Second, we will add a regularizing term of the form $\varepsilon \Delta u$ on the right hand-side in the Hamilton–Jacobi equation defining $S_t \chi$, this will yield a family of approximations of $S_t \chi$ that are smooth enough to perform the computations sketched above.

Observe that every $\chi \in \mathfrak{X}$ is 1-Lipschitz, hence the function $S_t \chi$ is $\mathbb{R}_+^D$-increasing thanks to (1.16) and 1-Lipschitz thanks to Lemma 5.13. Thus, changing ξ outside the intersection of $\mathbb{R}_+^D$ with the ball of radius 1 in $\mathbb{R}^D$ does not affect $S_t \chi$, this is proved rigorously in [15, Theorem 1.2 (2) (c)]. Therefore, we have some freedom on choosing how to define ξ outside of $B(0, 1) \cap \mathbb{R}_+^D$, and it will be convenient for our purposes to use this freedom to modify ξ so that it becomes uniformly Lipschitz. Let L denote the largest value taken by $|\nabla \xi|$ on the set of $x \in \mathbb{R}_+^D$ such that $\sum_{d=1}^D x_d \leqslant 2D$. For every $x \in \mathbb{R}_+^D$, let

$$\xi_{\mathrm{reg}}(x) = \begin{cases} \max\{\xi(x), \xi(0) + 2L(\sum_{d=1}^D x_d - D)\} & \text{if } \sum_{d=1}^D x_d \leqslant 2D \\ \xi(0) + 2L(\sum_{d=1}^D x_d - D) & \text{otherwise.} \end{cases}$$

It can be checked (see [40, Proposition 6.8] and [16, Lemma 4.3]) that ξ_{reg} coincides with ξ on $B(0, 1) \cap \mathbb{R}_+^D$, that ξ_{reg} is convex and Lipschitz on $\mathbb{R}_+^D$, and that it is $\mathbb{R}_+^D$-convex and $\mathbb{R}_+^D$-increasing on $\mathbb{R}_+^D$ (this last property is referred to as "proper" in the cited references).

We define $\overline{\xi} : \mathbb{R}^D \to \mathbb{R}$ by setting

$$\overline{\xi}(x) = \xi_{\mathrm{reg}}(x_+).$$

It can be easily checked that $\overline{\xi}$ is locally Lipschitz and convex. We let $(\overline{S}_t)_{t \geqslant 0}$ denote the semigroup associated to the equation

$$\partial_t u - \overline{\xi}(\nabla u) = 0 \text{ on } (0+, \infty) \times \mathbb{R}^D.$$

According to [15] or [20], the semigroup $(\overline{S}_t)_{t \geqslant 0}$ is well-defined on the set of Lipschitz functions $\mathbb{R}^D \to \mathbb{R}$.

Lemma 5.15. *Let $\chi \in \mathfrak{X}$, define $\overline{\chi} : \mathbb{R}^D \to \mathbb{R}$ by setting $\overline{\chi}(x) = \chi(x_+)$. Then $\overline{\chi}$ is Lipschitz and convex on $\mathbb{R}^D$ and we have for every $(t, x) \in [0, +\infty) \times \mathbb{R}_+^D$,*

$$(5.24) \qquad \overline{S}_t \overline{\chi}(x) = S_t \chi(x).$$

Proof. In this proof, we let ξ_{reg}^* denote the convex dual of ξ_{reg} with respect to $\mathbb{R}_+^D$, more precisely

$$\xi_{\mathrm{reg}}^*(y) = \sup_{x \in \mathbb{R}_+^D} \{x \cdot y - \xi_{\mathrm{reg}}(x)\}.$$

We also let $\overline{\xi}^*$ denote the convex dual of $\overline{\xi}$ with respect to $\mathbb{R}^D$, more precisely

$$\overline{\xi}^*(y) = \sup_{x \in \mathbb{R}^D} \left\{ x \cdot y - \overline{\xi}(x) \right\}.$$

Step 1. We show that $\overline{\chi}$ is Lipschitz, convex and $\mathbb{R}_+^D$-increasing on $\mathbb{R}^D$.

Since $\chi \in \mathfrak{X}$ we have $\overline{\chi}(x) = \sum_{d=1}^D \overline{\chi}_d(x_d)$, where $\overline{\chi}_d$ is the function which coincides with χ_d on $\mathbb{R}_+$ and is constant taking the value $\chi_d(0) = 0$ otherwise. Therefore it suffices to show that $\overline{\chi}_d$ is Lipschitz, convex and increasing on $\mathbb{R}$. Let $\lambda_1 \in \mathbb{R}_+$ and $\lambda_2 \in \mathbb{R}_-$, we have

$$|\overline{\chi}_d(\lambda_1) - \overline{\chi}_d(\lambda_2)| = |\chi_d(\lambda_1) - \chi_d(0)| \leqslant |\lambda_1|.$$

Now observe that $|\lambda_1| = \lambda_1 \leqslant \lambda_1 - \lambda_2 = |\lambda_1 - \lambda_2|$, so

$$|\overline{\chi}_d(\lambda_1) - \overline{\chi}_d(\lambda_2)| \leqslant |\lambda_1 - \lambda_2|.$$

Now observe that the previous display clearly also holds when $\lambda_1, \lambda_2 \in \mathbb{R}_+$ or $\lambda_1, \lambda_2 \in \mathbb{R}_-$, thus $\overline{\chi}_d$ is Lipschitz. In addition, χ_d is increasing on $\mathbb{R}$ as a composition of increasing functions and $\overline{\chi}_d$ is convex on $\mathbb{R}$ as the composition of a convex function and a convex increasing function. In conclusion, $\overline{\chi}$ is Lipschitz, convex and $\mathbb{R}_+^D$-increasing on $\mathbb{R}^D$.

Step 2. We show that for every $x \in \mathbb{R}^D$, $\overline{\xi}^*(x) \geqslant \xi_{\mathrm{reg}}^*(x_+)$ with equality when $x \in \mathbb{R}_+^D$.

Let $x \in \mathbb{R}^D$, using Lemma 5.7 in the last line we have

$$\begin{aligned}
\overline{\xi}^*(x) &= \sup_{y \in \mathbb{R}^D} \left\{ x \cdot y - \xi_{\mathrm{reg}}(y_+) \right\} \\
&\geqslant \sup_{y \in \mathbb{R}_+^D} \left\{ x \cdot y - \xi_{\mathrm{reg}}(y) \right\} \\
&= \xi_{\mathrm{reg}}^*(x) \\
&= \xi_{\mathrm{reg}}^*(x_+).
\end{aligned}$$

This proves the first part of the statement. Now assume that $x \in \mathbb{R}_+^D$. Since $x \cdot y \leqslant x \cdot y_+$ we have

$$\overline{\xi}^*(x) = \sup_{y \in \mathbb{R}^D} \left\{ x \cdot y - \xi_{\mathrm{reg}}(y_+) \right\} \leqslant \sup_{y \in \mathbb{R}^D} \left\{ x \cdot y_+ - \xi_{\mathrm{reg}}(y_+) \right\} = \xi_{\mathrm{reg}}^*(x).$$

Step 3. We show that for every $x \in \mathbb{R}_+^D$, $\overline{S}_t \overline{\chi}(x) = S_t \chi(x)$.

Without loss of generality we assume that $t = 1$. Since ξ_{reg} and $\overline{\xi}$ are convex on $\mathbb{R}_+^D$ and $\mathbb{R}^D$, we are simply going to verify that the Hopf-Lax representations of $\overline{S}_1 \overline{\chi}(x)$ and $S_1 \chi(x)$ coincide. Let $x \in \mathbb{R}_+^D$, thanks to the

equality case in the inequality from Step 2 we have

$$\overline{S}_1\overline{\chi}(x) = \sup_{y \in \mathbb{R}^D} \left\{ \overline{\chi}(x+y) - \overline{\xi}^*(y) \right\}$$

$$\geqslant \sup_{y \in \mathbb{R}^D_+} \left\{ \overline{\chi}(x+y) - \overline{\xi}^*(y) \right\}$$

$$= \sup_{y \in \mathbb{R}^D_+} \left\{ \chi(x+y) - \xi^*_{\mathrm{reg}}(y) \right\}$$

$$= S_1\chi(x).$$

In addition, thanks to the sub-additivity of the positive part and the inequality from Step 2, we have for every $y \in \mathbb{R}^D$ that

$$x + y_+ - (x+y)_+ \in \mathbb{R}^D_+$$

$$-\overline{\xi}^*(y) \leqslant -\xi^*_{\mathrm{reg}}(y_+).$$

Since $\overline{\chi}$ is $\mathbb{R}^D_+$-increasing on $\mathbb{R}^D$ according to Step 1, we have

$$\overline{\chi}(x+y) - \overline{\xi}^*(y) \leqslant \chi(x+y_+) - \xi^*_{\mathrm{reg}}(y_+) \leqslant S_1\chi(x).$$

Taking the supremum over $y \in \mathbb{R}^D$ in the previous display, we obtain $\overline{S}_1\overline{\chi}(x) \leqslant S_1\chi(x)$, thus the result.   □

We can extend the notion of $\mathbb{R}^D_+$-convexity to functions on $\mathbb{R}^D$ by saying that such a function is $\mathbb{R}^D_+$-convex when (5.17) holds for every $x \in \mathbb{R}^D$ and $y, z \in \mathbb{R}^D_+$.

Definition 5.16 ($\mathbb{R}^D_+$-convexity on $\mathbb{R}^D$)**.** We say that $\overline{\chi} : \mathbb{R}^D \to \mathbb{R}$ is $\mathbb{R}^D_+$-convex when for every $x \in \mathbb{R}^D$ and $y, z \in \mathbb{R}^D_+$, we have

$$(5.25) \qquad \overline{\chi}(x+y+z) - \overline{\chi}(x+z) \geqslant \overline{\chi}(x+y) - \overline{\chi}(x).$$

Remark 5.17. Reproducing the proof of Lemma 5.14, one can check that a $\mathcal{C}^2$ function $\overline{\chi} : \mathbb{R}^D \to \mathbb{R}$ is $\mathbb{R}^D_+$-convex if and only for every $x \in \mathbb{R}^D$, $\nabla^2\overline{\chi}(x) \in \mathbb{R}^{D \times D}_+$.   □

Lemma 5.18. *Let* $H : \mathbb{R}^D \to \mathbb{R}$ *and* $g : \mathbb{R}^D \to \mathbb{R}$ *be* $\mathcal{C}^\infty$ *such that all the derivatives of order* $\geqslant 1$ *of* H *are in* L^∞ *and all the derivatives of order* $\geqslant 2$ *of* g *are in* L^∞. *Consider the Cauchy problem*

$$(5.26) \qquad \begin{cases} \partial_t u^\varepsilon - H(\nabla u^\varepsilon) = \varepsilon \Delta u^\varepsilon \ \ on \ (0, +\infty) \times \mathbb{R}^D \\ u^\varepsilon(0, \cdot) = g. \end{cases}$$

There is exactly one function u^ε *that is* $\mathcal{C}^\infty$ *smooth with bounded derivatives of order* $\geqslant 2$ *and that satisfies* (5.26). *Furthermore, if* H *and* g *are* $\mathbb{R}^D_+$-*convex, then for every* $t \geqslant 0$, *the function* $u^\varepsilon(t, \cdot)$ *is* $\mathbb{R}^D_+$-*convex.*

The well-posedness of (5.26) is well known in many contexts and is a basic result in the theory of viscosity solutions to Hamilton–Jacobi equations [22, 23]. For a pedagogical and detailed proof of this result under some slightly different boundedness assumptions, we refer to the lecture notes [8].

In the usual approach to well-posedness, the main ingredient is Duhamel's formula which allows to show that the solutions of (5.26) are the fixed points of a certain functional. For the sake of completeness, we explain in the proof below how to adapt the estimates of [8, Section 3] yielding existence of a fixed point.

Proof of Lemma 5.18. Up to modifying H and rescaling time, we can assume without loss of generality that $\varepsilon = 1$. In this proof we will thus not explicitly make the ε dependence of (5.26) apparent and simply write u instead of u^{ε}. We let P denote the D-dimensional heat kernel, more precisely

$$P(t,x) = \frac{1}{(4\pi t)^{\frac{D}{2}}} e^{-\frac{|x|^2}{4t}}.$$

Given two functions $a, b : \mathbb{R}^D \to \mathbb{R}$, we let $a * b$ denote their additive convolution, namely

$$a * b(x) = \int_{\mathbb{R}^D} a(y) b(x - y) \mathrm{d}y.$$

Step 1. Existence and uniqueness of solutions to (5.26).

Given a $\mathcal{C}^\infty$ function $u : \mathbb{R}_+ \times \mathbb{R}^D \to \mathbb{R}$, we let for every $t \geqslant 0$

$$S(u)(t,\cdot) = P(t,\cdot) * g + \int_0^t P(t-s,\cdot) * H(\nabla u(s,\cdot)) \mathrm{d}s.$$

We let $u^{(-1)} = 0$ and for every $k \geqslant 0$, $u^{(k)} = S(u^{(k-1)})$. By construction of S, for every $k \geqslant 0$, we have

(5.27) $$\begin{cases} \partial_t u^{(k)} - \Delta u^{(k)} = H(\nabla u^{(k-1)}) \text{ on } (0, +\infty) \times \mathbb{R}^D \\ u^{(k)}(0,\cdot) = g. \end{cases}$$

Therefore, it suffices to show that the sequence $(u^{(k)})_{k \geqslant 1}$ converges in an appropriate sense to a function u to get the existence of a solution for (5.26). To do so, let us fix $T > 0$. We will prove that for every $n \geqslant 2$, the sequence $(\nabla^n u^{(k)})_{k \geqslant 1}$ is uniformly bounded on $[0,T] \times \mathbb{R}^D$. From this we will deduce that the sequence $((\partial_t, \nabla)^2 u^{(k)})_{k \geqslant 1}$ is uniformly bounded on $[0,T] \times \mathbb{R}^D$. Since $S(u)(0,\cdot) = g$ and $\nabla S(u)(0,\cdot) = \nabla g$, thanks to the Arzelà–Ascoli theorem, those estimates will yield local uniform convergence up to extraction of the sequence $(\nabla^n u^{(k)})_{k \geqslant 1}$ for all $n \geqslant 0$, which will conlude the proof of existence of a solution with bounded derivatives of order $\geqslant 2$ for (5.26).

We define

$$L_n(k,t) = \|\nabla^n u^{(k)}(t,\cdot)\|_{L^\infty}.$$

We now explain how to prove bounds that are uniform in $k \in \mathbb{N}$ and $t \in [0,T]$ for $L_2(k,t)$ following the proof of [8, Claim 3.6], those estimates can easily

be generalized to any $n \geqslant 2$. Differentiating twice the recurrence relation $u^{(k)} = S(u^{(k-1)})$, we have

$$\partial_d \partial_{d'} u^{(k)}$$

$$= P(t) * \partial_d \partial_{d'} g + \int_0^t \partial_{d'} P(t-s) * \left(\nabla H(\nabla u^{(k-1)})(s,\cdot) \cdot \nabla \partial_d u^{(k-1)}(s,\cdot) \right) ds$$

There exists a constant $c > 0$ depending only on D such that $\|\nabla P(t)\|_{L^1} \leqslant ct^{-1/2}$. Therefore it follows from Young's inequality $\|a * b\|_{L^\infty} \leqslant \|a\|_{L^1} \|b\|_{L^\infty}$ and the previous display that

$$L_2(k,t) \leqslant \|\nabla^2 g\|_{L^\infty} + c|\nabla H|_\infty \int_0^t \frac{1}{\sqrt{t-s}} L_2(k-1,s) ds.$$

After further manipulations of this inequality and induction on k, we obtain the existence of a constant $\Lambda > 0$ depending only on D and $|\nabla H|_\infty$ such that

$$L_2(k,t) \leqslant 2\|\nabla^2 g\|_{L^\infty} e^{\Lambda t}.$$

More generally, by induction on n using a similar argument, we can show that for every $T > 0$ and every $n \geqslant 2$ there exists $C(n,T) > 0$ such that for every $k \in \mathbb{N}$ and every $t \in [0,T]$,

$$L_n(k,t) \leqslant C(n,T).$$

Those are the announced estimates for $\nabla^n u^{(k)}$. To convert this into estimates for $(\partial_t, \nabla)^2 u^{(k)}$ we use (5.27) to deduce that

$$\partial_t \nabla u^{(k)} = \nabla^2 u^{(k)} \nabla H(\nabla u^{(k)}) + \nabla \Delta u^{(k-1)},$$

$$\partial_t \partial_t u^{(k)} = \partial_t \nabla u^{(k)} \cdot \nabla H(\nabla u^{(k)}) + \partial_t \Delta u^{(k-1)}.$$

In addition, using (5.27) once more we have

$$\partial_t \Delta u^{(k-1)} = \sum_i \partial_i \nabla u^{(k-1)} \cdot \nabla^2 H(u^{(k-1)}) \partial_i \nabla u^{(k-1)} + \nabla H(\nabla u^{(k-1)}) \cdot \nabla \Delta u^{(k-1)}$$

$$+ \Delta \Delta u^{(k-2)}.$$

Since $\nabla H \in L^\infty$ and $\nabla^2 H \in L^\infty$, it follows from the previous two displays and our previous estimates that $(\partial_t \nabla u^{(k)})_{k \geqslant 1}$ and $(\partial_t \partial_t u^{(k)})_{k \geqslant 1}$ are uniformly bounded on $[0,T] \times \mathbb{R}^D$. Therefore, $((\partial_t, \nabla)^2 u^{(k)})_{k \geqslant 1}$ is uniformly bounded on $[0,T] \times \mathbb{R}^D$, so $((\partial_t, \nabla) u^{(k)})_{k \geqslant 1}$ is uniformly Lipschitz on $[0,T] \times \mathbb{R}^D$. Since at $t = 0$, $(\partial_t, \nabla) u^{(k)}(0,\cdot) = (\Delta g + H(\nabla g), \nabla g)$, we deduce that $((\partial_t, \nabla) u^{(k)})_{k \geqslant 1}$ converges uniformly on $[0,T] \times \mathbb{R}^D$ thanks to the Arzelà-Ascoli theorem. In addition, it follows from the uniform Lipschitzness that $((\partial_t, \nabla) u^{(k)})_{k \geqslant 1}$ is locally uniformly bounded on $[0,T] \times \mathbb{R}^D$, so $(u^{(k)})_{k \geqslant 1}$ is uniformly locally Lipschitz on $[0,T] \times \mathbb{R}^D$. Since $u^{(k)}(0,\cdot) = g$, we can apply the Arzelà-Ascoli theorem once more to obtain local uniform convergence of $(u^{(k)})_{k \geqslant 1}$ on $[0,T] \times \mathbb{R}^D$. This concludes the existence part of the proof. For the uniqueness part we can proceed as in the proof of [8, Theorem 3.2] but using weighted L^p norms instead of the L^∞ norm.

Step 2. Let $g_\alpha(x) = g(x) + \alpha \left(\sum_d^D x_d \right)^2$ and let u_α the unique solution of (5.26) with initial condition g_α. We show that for every $t \geqslant 0$, $\nabla^2 u_\alpha(t, \cdot) \to \nabla^2 u(t, \cdot)$ pointwise up to extraction as $\alpha \to 0$.

To prove this, we perform estimates as in Step 1, but this time for

$$\widetilde{L}_n(k, t) = \|\nabla^n u_\alpha^{(k)}(t, \cdot) - \nabla^n u^{(k)}(t, \cdot)\|_{L^p(w)},$$

with $n = 1, 2$, where $L^p(w)$ denotes the L^p space with weight $w(x) = e^{-|x|}$ and $\|f\|_{L^p(w)} = \|fw\|_{L^p}$. In this context we can replace the usual Young inequality by the following weighted version,

$$\|a * b\|_{L^p(w)} \leqslant \|a\|_{L^1(\overline{w})} \|b\|_{L^p(w)},$$

where $\overline{w}(x) = e^{|x|}$. This weighted inequality can be deduced from the usual Young inequality after observing that $w(x + y) \leqslant \overline{w}(x)w(y)$, we refer to [44, Theorem 2.4] for a full proof. As previously, this yields estimates that are uniform in $t \in [0, T]$ and $k \in \mathbb{N}$ and that vanish in the limit $\alpha \to 0$. From this, we obtain convergence of $\nabla^2 u_\alpha(t, \cdot) \to \nabla^2 u(t, \cdot)$ in $L^p(w)$ and thus we obtain pointwise convergence after extraction.

Step 3. We show that if H and g are $\mathbb{R}_+^D$-convex, then for every $t \geqslant 0$, $u(t, \cdot)$ is $\mathbb{R}_+^D$-convex.

According to Lemma 5.14 and Remark 5.17, it suffices to show that $w = \nabla^2 u$ is $\mathbb{R}_+^{D \times D}$-valued. To do so, observe that the coefficients of w satisfy a system of semilinear parabolic PDEs. More precisely, for every $d, d' \in \{1, \ldots, D\}$ we let $w_{dd'}$ denote the coefficient of index (d, d') of w. Then, w is a classical solution of

$$\begin{cases} \partial_t w_{dd'} = \Delta w_{dd'} + b(t, x) \cdot \nabla w_{dd'} + f_{dd'}(t, x, w) \text{ on } (0, +\infty) \times \mathbb{R}^D \\ w(0, \cdot) = \nabla^2 g, \end{cases}$$

where $b(t, x) = \nabla H(\nabla u(t, x))$ and $f_{dd'}(t, x, w) = \left(w \nabla^2 H(\nabla u) w \right)_{dd'}$ is the coefficient of index (d, d') of the matrix $w \nabla^2 H(\nabla u) w$. The key point here is that since H is $\mathbb{R}_+^D$-convex, for $z \in \mathbb{R}_+^{D \times D}$ we have $f_{dd'}(t, x, z) \geqslant 0$. Using this we are going to show that $w_{dd'}$ is nonnegative.

Let $\alpha > 0$ and let $w^\alpha = \nabla^2 u_\alpha$ where u^α is defined as in Step 2. The function w^α solves a system of semilinear parabolic PDEs as in the previous display but with different b and f depending on α and with initial condition $w^\alpha(0, \cdot) = \nabla^2 g + \alpha \mathbf{1}$, where $\mathbf{1}$ is the $D \times D$ matrix with all coefficients equal to 1. Let

$$T_\alpha = \sup \left\{ t \geqslant 0 \mid \forall x \in \mathbb{R}^D, \forall d, d' \leqslant D, w_{dd'}^\alpha(t, x) \geqslant 0 \right\}$$

denote the last time at which all the coefficients of $w^\alpha(t, \cdot)$ are all nonnegative. All the derivatives or order $\geqslant 2$ of u^α are bounded, thus for every $T > 0$, we have that $w^\alpha(\cdot, x)$ is uniformly Lipschitz in x on $[0, T]$. Thus, it follows from

the fact that $w_{dd'}^\alpha(0, x) \geqslant \alpha$, that $T_\alpha > 0$. Then $w_{dd'}^\alpha$ solves

$$\begin{cases} \partial_t w_{dd'}^\alpha \geqslant \Delta w_{dd'}^\alpha + b_\alpha(t, x) \cdot \nabla w_{dd'}^\alpha \text{ on } (0, T_\alpha) \times \mathbb{R}^D \\ w_{dd'}^\alpha \geqslant \alpha. \end{cases}$$

In particular, we are now working with a plain linear parabolic PDEs. By the comparison principle, we obtain that for every $(t, x) \in (0, T_\alpha) \times \mathbb{R}^D$, $w_{dd'}^\alpha(t, x) \geqslant \alpha$. Arguing by contradiction, we now assume that $T_\alpha < +\infty$. By continuity, we have $w_{dd'}^\alpha(T_\alpha, x) \geqslant \alpha$. Using the same argument that yielded $T_\alpha > 0$, we obtain that $w_{dd'}^\alpha$ remains nonnegative for a small time after T_α, and this contradicts the definition of T_α. In conclusion $T_\alpha = +\infty$ and for every $(t, x) \in (0, +\infty) \times \mathbb{R}^D$, we have $w_{dd'}^\alpha(t, x) \geqslant \alpha$. Finally, from Step 2, we have that as $\alpha \to 0$, $w_{dd'}^\alpha(t, x) \to w_{dd'}(t, x)$. Thus, w is $\mathbb{R}_+^{D \times D}$-valued, as desired. $\square$

Proof of Proposition 5.9. By Lemma 5.15 we know that $S_t \chi$ is the restriction to $\mathbb{R}_+^D$ of $\overline{S}_t \overline{\chi}$, therefore it suffices to prove that $\overline{S}_t \overline{\chi}$ is $\mathbb{R}_+^D$-convex. We will prove that this result is a consequence of Lemma 5.18 by performing several approximation procedures.

First, observe that for every $\delta > 0$ the mollifications $H_\delta = \overline{\xi} * \eta_\delta$ and $g_\delta = \overline{\chi} * \eta_\delta$ are $\mathbb{R}_+^D$-convex and $\mathcal{C}^\infty$. Since ξ_{reg} is linear outside a compact set, H_δ satisfies the hypotheses of Lemma 5.18. On the other hand, the function g_δ may have unbounded derivatives of higher order. This can be easily fixed by replacing χ by χ_R where, given $R > 0$, we let $\chi_{d,R}(\lambda) = \chi_d(\lambda)$ for $\lambda \leqslant R$ and $\chi_{d,R}(\lambda) = \chi_d(R) + \lambda - R$ otherwise and $\chi_R = \sum_d \chi_{d,R}$. This way we guarantee that outside of a large ball $\overline{\chi_R}$ is linear and thus its derivatives of all orders are bounded. Also note that since each χ_d is 1-Lipschitz and convex, the extensions $\chi_{d,R}$ are also 1-Lipschitz and convex. In particular $\overline{\chi}_R$ is $\mathbb{R}_+^D$-convex.

Step 1. Let H and g satisfying the hypotheses of Lemma 5.18 and such that g is Lipschitz. We show that for every $t \geqslant 0$, $u(t, \cdot)$ is $\mathbb{R}_+^D$-convex, where u is the unique vicosity solution of

$$\begin{cases} \partial_t u - H(\nabla u) = 0 \text{ on } (0, +\infty) \times \mathbb{R}^D \\ u(0, \cdot) = g. \end{cases}$$

It is classical [22] that in the limit $\varepsilon \to 0$, the sequence of functions $(u^\varepsilon)_\varepsilon$ from Lemma 5.18 converges to u. Since according to Lemma 5.18, for every $t \geqslant 0$, $u^\varepsilon(t, \cdot)$ is $\mathbb{R}_+^D$-convex, it follows that $u(t, \cdot)$ is $\mathbb{R}_+^D$-convex.

Step 2. We show that for every $t \geqslant 0$, $u_R(t, \cdot)$ is $\mathbb{R}_+^D$-convex, where u_R is the unique vicosity solution of

$$\begin{cases} \partial_t u_R - \overline{\xi}(\nabla u_R) = 0 \text{ on } (0, +\infty) \times \mathbb{R}^D \\ u_R(0, \cdot) = \overline{\chi_R}. \end{cases}$$

We can molify $\overline{\xi}$ and $\overline{\chi_R}$ into H_δ and $g_{R,\delta}$. According to Step 1, the unique viscosity solution $u_{R,\delta}$ of

$$\begin{cases} \partial_t u_{R,\delta} - H_\delta(\nabla u_{R,\delta}) = 0 \text{ on } (0,+\infty) \times \mathbb{R}^D \\ u_{R,\delta}(0,\cdot) = g_{R,\delta}, \end{cases}$$

is $\mathbb{R}_+^D$-convex at every fixed t. In addition $H_\delta \to \overline{\xi}$ and $g_{R,\delta} \to \overline{\chi}_R$ locally uniformly as $\delta \to 0$. Therefore, any locally uniform limit of the sequence $(u_{R,\delta})_\delta$ must be the unique viscosity solution of the equation defining u_R [18, Section 6]. Hence, since the sequence $(u_{R,\delta})_\delta$ is uniformly Lipschitz, it follows from the Arzelà–Ascoli theorem that $\delta \to 0$, $u_{R,\delta} \to u_R$ locally uniformly. The desired result follows.

Step 3. We show that $\overline{S}_t \overline{\chi}$ is $\mathbb{R}_+^D$-convex.

The initial condition $\overline{\chi}_R$ and $\overline{\chi}$ coincide on $B(0,R)$. Since the Hamilton–Jacobi equation has finite speed of propagation [20, Exercise 3.8 and solution], we have that $u_R(t,\cdot)$ and $\overline{S}_t\overline{\chi}$ coincide on $B(0, R-t/L)$ where L is a Lipschitz constant of H. Hence, we have that $u_R(t,\cdot) \to \overline{S}_t\overline{\chi}$ pointwise as $R \to +\infty$. This allows to deduce the desired result from Step 2. □

6. Optimal transport and convex duality

In this section, we prove the key observations (1.28) and (1.30), and use them with convex-duality arguments to show the identity between the variational problems in (1.14) and (1.18).

We recall that $S_t\chi$ was defined in (5.1), and that in (1.13) we have introduced the quantity

$$(6.1) \qquad \mathcal{T}_t(\mu,\nu) = \inf_{\pi \in \Pi(\mu,\nu)} \int (t\xi)^*(y-x) \mathrm{d}\pi(x,y),$$

which can be interpreted as the optimal transport cost from μ to ν with cost function $(x,y) \mapsto (t\xi)^*(y-x)$. As explained in [61, Chapter 1], Kantorovich duality yields the dual representation

$$(6.2) \qquad \mathcal{T}_t(\mu,\nu) = \sup_\chi \left\{ \int \chi d\nu - \int S_t\chi d\mu \right\},$$

where the supremum on the right-hand side is taken over all functions $\chi : \mathbb{R}_+^D \to \mathbb{R}$ that are in $L^1(\nu)$ and are such that $S_t\chi \in L^1(\mu)$. Observe that here $\mathcal{T}_t(\mu,\nu)$ is represented as a supremum of affine functions in (μ,ν), in particular this means that the mapping $(\mu,\nu) \mapsto \mathcal{T}_t(\mu,\nu)$ is convex. This dual representation motivates the following definition.

Definition 6.1 (Kantorovich potential). For every $\mu,\nu \in \mathcal{P}(\mathbb{R}_+^D)$, we say that a function $\chi : \mathbb{R}_+^D \to \mathbb{R}$ is a Kantorovich potential from μ to ν (for the cost function $(x,y) \mapsto (t\xi^*)(y-x)$) when $\chi \in L^1(\nu)$, $S_t\chi \in L^1(\mu)$, and

$$(6.3) \qquad \mathcal{T}_t(\mu,\nu) = \int \chi d\nu - \int S_t\chi d\mu.$$

We can thus rephrase the identity (1.30) as saying that the function χ is a Kantorovich potential. A key aspect of the proofs of our main results is that the function χ appearing in (1.30) and in the surrounding informal discussion is not only a Kantorovich potential, but also an element of the set $\mathfrak{X}$ defined in (3.17).

Recall the collection $\mathfrak{X}$ defined in (3.17). For ψ appearing in (1.8), we define

$$(6.4) \qquad \psi_*(\chi) = \inf_{\nu \in \mathcal{P}_1(\mathbb{R}^D_+)} \left\{ \int \chi \mathrm{d}\nu - \psi(\nu) \right\}, \quad \forall \chi \in \mathfrak{X}.$$

For every $(t, \mu) \in \mathbb{R}_+ \times \mathcal{P}_1(\mathbb{R}^D_+)$, we set

$$(6.5) \qquad g(t, \mu) = \inf_{\chi \in \mathfrak{X}} \left\{ \int S_t \chi \mathrm{d}\mu - \psi_*(\chi) \right\}$$

$$(6.6) \qquad h(t, \mu) = \sup_{\nu \in \mathcal{P}_1(\mathbb{R}^D_+)} \left\{ \psi(\nu) - \mathcal{T}_t(\mu, \nu) \right\}.$$

Proposition 6.2. *We have $h = g$ on $\mathbb{R}_+ \times \mathcal{P}_1(\mathbb{R}^D_+)$. Moreover, at every $(t, \mu) \in \mathbb{R}_+ \times \mathcal{P}_1(\mathbb{R}^D)$, there is a maximizer ν of $h(t, \mu)$ in (6.6) and a minimizer χ of $g(t, \mu)$ in (6.5) such that χ is the derivative of ψ at ν given by Proposition 3.8.*

To prove this, we start with some preliminary results.

Lemma 6.3 (Existence of maximizers in (6.6)). *For every $t \in \mathbb{R}_+$ and $\mu \in \mathcal{P}_1(\mathbb{R}^D_+)$, the variational formula of $h(t, \mu)$ in (6.6) achieves its maximum at some $\overline{\nu} \in \mathcal{P}_1(\mathbb{R}^D_+)$. Moreover, there is an optimal $\pi \in \Pi(\mu, \overline{\nu})$ such that $\mathcal{T}_t(\mu, \overline{\nu}) = \int (t\xi)^*(y - x) \mathrm{d}\pi(x, y)$.*

Proof. Step 1. We show that $\mathcal{T}_t(\mu, \nu)$ and $\mathsf{W}_1(\mu, \nu)$ must be bounded for near maximizers ν. Recall the notation y_+ in (5.15) for $y \in \mathbb{R}^D$. According to Lemma 5.2, we have that ξ^* is superlinear on $\mathbb{R}^D_+$. Since $(t\xi)^* = t\xi^*(\cdot/t)$, $(t\xi)^*$ is also superlinear on $\mathbb{R}^D_+$. Since $\mathcal{T}_t(\mu, \mu) = 0$, the supremum in (6.6) can be restricted to those measures $\nu \in \mathcal{P}_1(\mathbb{R}^D_+)$ such that

$$(6.7) \qquad \psi(\nu) - \mathcal{T}_t(\mu, \nu) \geq \psi(\mu).$$

We start by showing that for every $\nu \in \mathcal{P}_1(\mathbb{R}^D_+)$ satisfying (6.7), we have that $\mathsf{W}_1(\mu, \nu) \leq C$. The Lipschitzness of ψ in (3.14) together with (6.7) yields

$$(6.8) \qquad \mathcal{T}_t(\mu, \nu) \leq \mathsf{W}_1(\mu, \nu).$$

For $\varepsilon > 0$, let $\pi \in \Pi(\mu, \nu)$ be a nearly optimal coupling for $\mathcal{T}_t(\mu, \nu)$ such that $\int (t\xi)^*(y - x) \mathrm{d}\pi(x, y) \leq \mathcal{T}_t(\mu, \nu) + \varepsilon$. Let (X, Y) be a random variable with law π. Then, we can obtain from the above display that

$$(6.9) \qquad \mathbb{E}\left[(t\xi)^*(Y - X)\right] \leq \mathbb{E}\left[|X - Y|\right] + \varepsilon.$$

We choose $M > 1$ and let R be such that (5.6) holds with ς substituted with $(t\xi)^*$. By this and Lemma 5.7, we have

$$(6.10) \qquad (t\xi)^*(Y - X) \geqslant M|(Y - X)_+| \qquad \text{on the event } |(Y - X)_+| \geqslant R.$$

Also by (5.13), we have

$$(6.11) \qquad |X - Y| = |Y - X| - |(Y - X)_+| + |(Y - X)_+| \leqslant |X| + |(Y - X)_+|.$$

Using these and $|(Y - X)_+|\mathbb{1}_{|(Y-X)_+|<R} \leqslant R$, we get

$$M\mathbb{E}\left[|(Y - X)_+|\right] - MR \overset{(6.10)}{\leqslant} \mathbb{E}\left[(t\xi)^*(Y - X)\right] \overset{(6.9)}{\leqslant} \mathbb{E}\left[|X - Y|\right] + \varepsilon$$

$$\overset{(6.11)}{\leqslant} \mathbb{E}[|X|] + \mathbb{E}\left[|(Y - X)_+|\right] + \varepsilon.$$

In the first inequality in the display, we also used that $(t\xi)^* \geqslant -t\xi(0) = 0$. Rearranging, we arrive at

$$(6.12) \qquad \mathbb{E}\left[|(Y - X)_+|\right] \leqslant \frac{\mathbb{E}[|X|] + MR + \varepsilon}{M - 1}.$$

This along with (5.13) implies that $\mathbb{E}[|Y - X|]$ is bounded by some constant C. Hence, we conclude that for every ν satisfying (6.7), we have $\mathcal{T}_t(\mu, \nu) \leqslant \mathsf{W}_1(\mu, \nu) \leqslant C$, where we used (6.8) in the first inequality.

Step 2. Let $(\nu_n)_{n\in\mathbb{N}}$ be a maximizing sequence of (6.6), namely,

$$(6.13) \qquad \lim_{n\to\infty} (\psi(\nu_n) - \mathcal{T}_t(\mu, \nu_n)) = h(t, \mu).$$

Our goal is to extract a maximizer from its subsequential limits. We can choose the sequence $(\nu_n)_{n\in\mathbb{N}}$ such that (6.7) holds for every n. By the previous step, we have that $\mathcal{T}_t(\mu, \nu_n) \leqslant \mathsf{W}_1(\mu, \nu_n) \leqslant C$ for every n. Fix any $\varepsilon > 0$, since $\mathcal{T}_t(\mu, \nu_n)$ has to be finite, we can choose a nearly optimal coupling $\pi_n \in \Pi(\mu, \nu_n)$ with $\int (t\xi)^*(y - x)\mathrm{d}\pi_n(x, y) \leqslant \mathcal{T}_t(\mu, \nu_n) + \varepsilon$. Since $\mu \in \mathcal{P}_1(\mathbb{R}_+^D)$ and $\mathsf{W}_1(\mu, \nu_n) \leqslant C$, the first moments of ν_n are bounded uniformly in n. Hence, by passing to a subsequence and invoking Skorokhod's representation theorem, we may assume that there is a sequence $(X_n, Y_n)_{n\in\mathbb{N}}$ of random variables satisfying $\mathrm{Law}(X_n, Y_n) = \pi_n$ and that (X_n, Y_n) converges almost surely to some (X, Y). Clearly, we have $\mu = \mathrm{Law}(X)$ since $\mu = \mathrm{Law}(X_n)$ for every n. We need that $(X_n, Y_n)_{n\in\mathbb{N}}$ also converges to (X, Y) in L^1, which together with the above follows from the uniform integrability of the second marginal:

$$(6.14) \qquad \lim_{r\to\infty} \sup_n \mathbb{E}\left[|Y_n|\mathbb{1}_{|Y_n|\geqslant r}\right] = 0.$$

We postpone the proof of this fact and first use the L^1-convergence to find a maximizer.

Let $\overline{\nu} = \mathrm{Law}(Y)$ and consequently, we have $\overline{\nu} \in \mathcal{P}_1(\mathbb{R}_+^D)$. Then, the L^1-convergence together with the Lipschitzness of ψ in (3.14) implies

$$(6.15) \qquad \lim_{n\to\infty} \psi(\nu_n) = \psi(\overline{\nu}).$$

On the other hand, the function $(t\xi)^*$ is lower semi-continuous as a supremum of continuous functions, thus it follows from Fatou's lemma that (recall that $(t\xi)^* \geqslant -t\xi(0) = 0$ is bounded from below)

$$\mathbb{E}\left[(t\xi)^*(Y - X)\right] \leqslant \liminf_{n\to\infty} \mathbb{E}\left[(t\xi)^*(Y_n - X_n)\right],$$

which further yields

$$\mathcal{T}_t(\mu, \overline{\nu}) \leqslant \mathbb{E}\left[(t\xi)^*(Y - X)\right] \leqslant \liminf_{n\to\infty} \mathcal{T}_t(\mu, \nu_n) + \varepsilon.$$

This along with (6.15) gives

$$\psi(\overline{\nu}) - \mathcal{T}_t(\mu, \overline{\nu}) \geqslant \psi(\overline{\nu}) - \mathbb{E}\left[(t\xi)^*(Y - X)\right] \geqslant \limsup_{n\to\infty} \psi(\nu_n) - \mathcal{T}_t(\mu, \nu_n) - \varepsilon$$

$$\stackrel{(6.13)}{=} h(t, \mu) - \varepsilon \stackrel{(6.6)}{\geqslant} \psi(\overline{\nu}) - \mathcal{T}_t(\mu, \overline{\nu}) - \varepsilon.$$

Since ε is arbitrary, we conclude that $\overline{\nu}$ maximizes (6.6) and $\mathcal{T}_t(\mu, \overline{\nu}) = \mathbb{E}\left[(t\xi)^*(Y - X)\right]$, implying that π is an optimal coupling for $\mathcal{T}_t(\mu, \overline{\nu})$.

Step 3. It remains to verify (6.14). We need the following simple fact: for real numbers $r \geqslant 0$ and $a, b, c \geqslant 0$ satisfying $a \leqslant b + c$, we have

$$(6.16) \qquad\qquad a\mathbb{1}_{a \geqslant 4r} \leqslant 2b\mathbb{1}_{b \geqslant r} + 2c\mathbb{1}_{c \geqslant r},$$

which follows from the fact that

$$a\mathbb{1}_{a \geqslant 4r} \leqslant (b + c)\mathbb{1}_{a \geqslant 4r} = b\mathbb{1}_{a \geqslant 4r, \, b \geqslant r} + b\mathbb{1}_{a \geqslant 4r, \, b < r} + c\mathbb{1}_{a \geqslant 4r, \, c \geqslant r} + c\mathbb{1}_{a \geqslant 4r, \, c < r}$$

$$\leqslant b\mathbb{1}_{b \geqslant r} + \tfrac{a}{4}\mathbb{1}_{a \geqslant 4r} + c\mathbb{1}_{c \geqslant r} + \tfrac{a}{4}\mathbb{1}_{a \geqslant 4r}.$$

Simple applications of (6.16) are

$$|Y_n|\mathbb{1}_{|Y_n| \geqslant 4r} \leqslant 2|X_n|\mathbb{1}_{|X_n| \geqslant r} + 2|X_n - Y_n|\mathbb{1}_{|X_n - Y_n| \geqslant r},$$

$$|X_n - Y_n|\mathbb{1}_{|X_n - Y_n| \geqslant 4r} \leqslant 2|X_n|\mathbb{1}_{|X_n| \geqslant r} + 2|(Y_n - X_n)_+|\mathbb{1}_{|(Y_n - X_n)_+| \geqslant r},$$

where we used the triangle inequality in the first line and (5.13) in the second line. Due to $\mathrm{Law}(X_n) = \mu$ independent of n, from the above display, we can see that (6.14) follows if we can show

$$(6.17) \qquad\qquad \lim_{r\to\infty} \sup_n \mathbb{E}\left[|(Y_n - X_n)_+|\mathbb{1}_{|(Y_n - X_n)_+| \geqslant r}\right] = 0.$$

By the choice of ν_n and π_n, we have $\mathbb{E}\left[(t\xi)^*(Y_n - X_n)\right] \leqslant \mathcal{T}_t(\mu, \nu_n) + \varepsilon \leqslant C + \varepsilon$. Recall that, for any $M > 1$, we can find R such that (6.10) holds. Hence, for $r > R$, we have

$$\mathbb{E}\left[|(Y_n - X_n)_+|\mathbb{1}_{|(Y_n - X_n)_+| \geqslant r}\right] \leqslant M^{-1}\mathbb{E}\left[(t\xi)^*(Y_n - X_n)\right] \leqslant M^{-1}(C + \varepsilon)$$

uniformly in n. This implies (6.17) and completes the proof. $\qquad\square$

We now show (1.28).

Lemma 6.4. *Let $\chi : \mathbb{R}_+^D \to \mathbb{R}$ be a Lipschitz and $\mathbb{R}_+^D$-increasing function and let $(S_t)_{t \in \mathbb{R}_+}$ be given as in (5.1). For every $(t, \mu) \in \mathbb{R}_+ \times \mathcal{P}_1(\mathbb{R}_+^D)$, we have*

$$(6.18) \qquad\qquad \int S_t \chi \mathrm{d}\mu = \sup_{\nu \in \mathcal{P}_1(\mathbb{R}_+^D)} \left\{ \int \chi \mathrm{d}\nu - \mathcal{T}_t(\mu, \nu) \right\}.$$

Remark 6.5. When $\chi \in \mathfrak{X}$, we can also write $S_t\chi$ from (5.1) as

$$(6.19) \qquad S_t\chi(x) = \sup_{\substack{x' \in \mathbb{R}_+^D \\ x'-x \in t\nabla\xi([0,1]^D)}} \left\{ \chi(x') - (t\xi)^*(x' - x) \right\}.$$

To see this, observe first that if $p \in \mathbb{R}_+^D$ is a maximizer in (1.16), then $p \in [0,1]^D$. This is because $\chi^*(p) = +\infty$ when $p \in \mathbb{R}_+^D \smallsetminus [0,1]^D$ (recall that χ^* is defined as in (1.10)). Now given p a maximizer in (1.16), we have

$$\begin{aligned} S_t\chi(x) &= x \cdot p - \chi^*(p) + t\xi(p) \\ &= (x + t\nabla\xi(p)) \cdot p - \chi^*(p) - t(\nabla\xi(p) \cdot p - \xi(p)) \\ &= (x + t\nabla\xi(p)) \cdot p - \chi^*(p) - t\xi^*(\nabla\xi(p)) \\ &\leqslant \sup_q \left\{ (x + t\nabla\xi(p)) \cdot q - \chi^*(q) \right\} - t\xi^*(\nabla\xi(p)) \\ &= \chi(x + t\nabla\xi(p)) - (t\xi)^*(t\nabla\xi(p)). \end{aligned}$$

On the third line, we used that $\xi^*(\nabla\xi(p)) = \nabla\xi(p) \cdot p - \xi(p)$, which can be deduced from the definition of ξ^* in (1.10) (see (1.9) and (7.25)). This means that $y = x + t\nabla\xi(p)$ is a maximizer in (5.1), and this proves (6.19).

Using (6.19) in place of (5.1) in the proof of Lemma 6.4 below yields that (6.18) remains valid when the supremum is taken over $\nu \in \mathcal{P}_1(\mathbb{R}_+^D)$ with the added condition that ν is supported in $\mathrm{supp}(\mu) + t\nabla\xi([0,1]^D)$. $\qquad\square$

Proof of Lemma 6.4. Using the definition of S_t in (5.1), we can get

$$\frac{1}{K} \sum_{k=1}^K S_t\chi(x_k) = \sup_{y \in (\mathbb{R}_+^D)^K} \left\{ \frac{1}{K} \sum_{k=1}^K \chi(y_k) - \frac{1}{K} \sum_{k=1}^K (t\xi)^*(y_k - x_k) \right\}.$$

Let $\Pi^K(x, \cdot)$ be the set of probability measures π on $\mathbb{R}_+^D \times \mathbb{R}_+^D$ of the form $\pi = \frac{1}{K} \sum_{k=1}^K \delta_{(x_k, y_k)}$ with $y \in (\mathbb{R}_+^D)^K$. For $\pi \in \Pi^K(x, \cdot)$, let π_2 be the second marginal of π on $\mathbb{R}_+^D$. The previous display can be written as

$$(6.20) \qquad \frac{1}{K} \sum_{k=1}^K S_t\chi(x_k) = \sup_{\pi \in \Pi^K(x, \cdot)} \left\{ \int \chi \, d\pi_2 - \int (t\xi)^*(b - a) \mathrm{d}\pi(a, b) \right\}.$$

Now, we let $\mu \in \mathcal{P}_1(\mathbb{R}_+^D)$ and choose a sequence $(x^K)_{K \geqslant 1}$ in $(\mathbb{R}_+^D)^K$ such that μ is limit of $\frac{1}{K} \sum_{k=1}^K \delta_{x_k^K}$ in law. Setting $x = x^K$ in the previous display and letting $K \to +\infty$, we claim that we can get

$$(6.21) \qquad \int S_t\chi \mathrm{d}\mu = \sup_{\pi \in \Pi_1(\mu, \cdot)} \left\{ \int \chi \, d\pi_2 - \int (t\xi)^*(y - x) \mathrm{d}\pi(x, y) \right\},$$

where $\Pi_1(\mu, \cdot)$ denotes the set of probability measures π on $\mathbb{R}_+^D \times \mathbb{R}_+^D$ such that π has a finite first moment and its first marginal satisfies $\pi_1 = \mu$. Assuming the validity of this for now, writing $\Pi_1(\mu, \nu)$ for the set of couplings between μ and ν, and replacing "$\sup_{\pi \in \Pi_1(\mu, \cdot)}$" by "$\sup_{\nu \in \mathcal{P}_1(\mathbb{R}_+^D)} \sup_{\pi \in \Pi_1(\mu, \nu)}$" we obtain (6.18) from (6.21) as desired.

Now, we verify (6.21). Using the definition of $S_t\chi$ in (5.1), we can deduce that the inequality "$\geqslant$" in (6.21) holds. We focus on the other direction. Let x^K be given as above (6.21). Let $\pi^K \in \Pi^K(x^K, \cdot)$ be a near maximizer of (6.20) satisfying

$$(6.22) \qquad \int S_t\chi \mathrm{d}\pi_1^K \leqslant K^{-1} + \int \chi \mathrm{d}\pi_2^K - \int (t\xi)^*(y - x)\mathrm{d}\pi^K(x, y).$$

Let (X_K, Y_K) be a pair of $\mathbb{R}^D$-valued random variables with law equal to π^K. Due to the choice of x^K, the law of X_K converges to $\mu \in \mathcal{P}_1(\mathbb{R}_+^D)$. By choosing x^K suitably, we may assume that

$$(6.23) \qquad \sup_K \mathbb{E}[|X_K|] < \infty.$$

In the notation of random variables, we can rewrite (6.22) as

$$(6.24) \qquad \mathbb{E}\left[S_t\chi(X_K)\right] \leqslant K^{-1} + \mathbb{E}\left[\chi(Y_K)\right] - \mathbb{E}\left[(t\xi)^*(Y_K - X_K)\right].$$

Since $S_t\chi$ and χ are Lipschitz (see Lemma 5.13), there is a constant $C > 0$ such that, for every x, y,

$$|S_t\chi(x)| + |\chi(y)| \leqslant C(1 + |x| + |y - x|).$$

This along with (6.24) gives

$$\mathbb{E}\left[(t\xi)^*(Y_K - X_K)\right] \leqslant K^{-1} + C(1 + \mathbb{E}[|X_K|] + \mathbb{E}[|Y_K - X_K|]).$$

We can use this condition to substitute the one in (6.9) and follow the ensuing steps there (using the superlinearity of $(t\xi)^*$ on $\mathbb{R}_+^D$ by choosing M sufficiently large) to reach an upper bound on $\mathbb{E}[|(Y_K - X_K)_+|]$ similar to the one in (6.12). Using (6.23) and (5.13), we can deduce

$$(6.25) \quad \sup_K \mathbb{E}[|Y_K - X_K|] < \infty \qquad \text{and} \qquad \sup_K \mathbb{E}\left[(t\xi)^*(Y_K - X_K)\right] < \infty.$$

The first bound in (6.25) along with (6.23) gives the tightness of $(\pi^K)_{K \in \mathbb{N}}$. By passing to a subsequence and using Skorokhod's representation theorem, we may assume that (X_K, Y_K) converges a.s. to some (X, Y) with law π.

We argue that the convergence also takes place in L^1. It suffices to show that the families $(X_K)_K$ and $(Y_K)_K$ are uniformly integrable (as in (6.14)). By choosing x^K suitably, we can verify this for the former. For the latter, the second relation in (6.25) allows us to repeat the argument for (6.14) in Step 3 of the proof of Lemma 6.3. The extra input is that we need to use the uniform integrability of $(X_K)_K$ as their laws are no longer fixed. With this explained, we now have that (X_K, Y_K) converges in L^1 to (X, Y).

Due to the assumption on the first marginal, we have $\text{Law}(X) = \mu$ and thus $\pi \in \Pi_1(\mu, \cdot)$. Since $S_t\chi$ and χ are Lipschitz and thus have linear growths, we can use the dominated convergence theorem to get

$$\lim_{K \to \infty} \mathbb{E}[S_t\chi(X_K)] = \int S_t\chi \mathrm{d}\mu \quad \text{and} \quad \lim_{K \to \infty} \mathbb{E}[\chi(Y_K)] = \int \chi \mathrm{d}\pi_2.$$

The lower semi-continuity of $(t\xi)^*$ and Fatou's lemma (recall that $(t\xi)^* \geqslant -t\xi(0) = 0$) yield

$$\int (t\xi)^*(y-x)\mathrm{d}\pi(x,y) \leqslant \liminf_{K\to\infty} \mathbb{E}\left[(t\xi)^*(Y_K - X_K)\right].$$

Applying these convergences to (6.24), we get

$$\int S_t\chi\mathrm{d}\mu \leqslant \int \chi\mathrm{d}\pi_2 - \int (t\xi)^*(y-x)\mathrm{d}\pi(x,y).$$

This gives "$\leqslant$" in (6.21) and completes the proof. $\square$

We now show (1.30), which can be interpreted as saying that χ is a Kantorovich potential from μ to $\overline{\nu}$ (see Definition 6.1).

Lemma 6.6. *Let $(t,\mu) \in \mathbb{R}_+ \times \mathcal{P}_1(\mathbb{R}_+^D)$ and let $\overline{\nu}$ be a maximizer of $h(t,\mu)$ in (6.6) (which exists by Lemma 6.3). Let $\chi \in \mathfrak{X}$ be the derivative of ψ at $\overline{\nu}$ given as in Proposition 3.8. We have*

$$(6.26) \qquad \int \chi\mathrm{d}\overline{\nu} - \int S_t\chi\mathrm{d}\mu = \mathcal{T}_t(\mu,\overline{\nu}).$$

Proof. For every $\nu \in \mathcal{P}_1(\mathbb{R}_+^D)$ and $\lambda \in (0,1]$, the convexity of $\mathcal{T}_t(\mu, \cdot)$ implies

$$\mathcal{T}_t(\mu,\nu) - \mathcal{T}_t(\mu,\overline{\nu}) \geqslant \frac{1}{\lambda}\left(\mathcal{T}_t(\mu,\overline{\nu} + \lambda(\nu - \overline{\nu})) - \mathcal{T}_t(\mu,\overline{\nu})\right)$$

$$\geqslant \frac{1}{\lambda}\left(\psi(\overline{\nu} + \lambda(\nu - \overline{\nu})) - \psi(\overline{\nu})\right)$$

where the second inequality follows from maximality at $\overline{\nu}$ of the formula for $h(t,\mu)$ (see (6.6)). Using the differentiability of ψ in Proposition 3.8 and sending λ to zero, we get

$$\mathcal{T}_t(\mu,\nu) - \mathcal{T}_t(\mu,\overline{\nu}) \geqslant \int \chi\mathrm{d}(\nu - \overline{\nu})$$

for any $\nu \in \mathcal{P}_1(\mathbb{R}_+^D)$, which implies

$$\int \chi\mathrm{d}\overline{\nu} - \mathcal{T}_t(\mu,\overline{\nu}) = \sup_{\nu\in\mathcal{P}_1(\mathbb{R}_+^D)}\left\{\int \chi\mathrm{d}\nu - \mathcal{T}_t(\mu,\nu)\right\}.$$

By Lemma 6.4, we can recognize the right-hand side as $\int S_t\chi\mathrm{d}\mu$, which verifies (6.26). $\square$

We are now ready to show that the variational problem in (1.18) is dual to that in (1.14), that is, that $g = h$.

Proof of Proposition 6.2. We will proceed by double inequality.

First, we show $h \geqslant g$. Allowed by Lemma 6.3, let $\overline{\nu} \in \mathcal{P}_1(\mathbb{R}_+^D)$ be such that

$$(6.27) \qquad h(t,\mu) = \psi(\overline{\nu}) - \mathcal{T}_t(\mu,\overline{\nu}).$$

By Lemma 6.6, we have

$$\mathcal{T}_t(\mu,\overline{\nu}) = \int \chi\mathrm{d}\overline{\nu} - \int S_t\chi\mathrm{d}\mu$$

where χ is the derivative of ψ at $\overline{\nu}$ given as in Proposition 3.8. The above two displays together yield

(6.28) $$h(t,\mu) = \psi(\overline{\nu}) - \int \chi d\overline{\nu} + \int S_t\chi \mathrm{d}\mu.$$

On the other hand, the concavity of ψ given in Lemma 3.2 implies that, for any $\nu \in \mathcal{P}_1(\mathbb{R}_+^D)$ and $\lambda \in (0,1]$, we have

$$\frac{1}{\lambda}\left(\psi\left(\lambda\nu + (1-\lambda)\overline{\nu}\right) - \psi(\nu)\right) \geqslant \psi(\nu) - \psi(\overline{\nu}).$$

Sending $\lambda \to 0$ and using Proposition 3.8, for the same χ as above, we get $\int \chi \mathrm{d}(\nu - \overline{\nu}) \geqslant \psi(\nu) - \psi(\overline{\nu})$. Rearranging and taking the infimum over $\nu \in \mathcal{P}_1(\mathbb{R}_+^D)$, we obtain $\psi_*(\chi) = \int \chi d\overline{\nu} - \psi(\overline{\nu})$ due to its definition in (6.4). This along with (6.28) gives

(6.29) $$h(t,\mu) = \int S_t\chi \mathrm{d}\mu - \psi_*(\chi) \overset{(6.5)}{\geqslant} g(t,\mu).$$

Next, we show $g \geqslant h$. According to the Arzelà–Ascoli theorem, $\mathfrak{X}$ is compact for the topology of local uniform convergence. We can thus use Sion's min-max theorem to obtain that

$$\begin{aligned}
g(t,\mu) &= \inf_{\chi\in\mathfrak{X}}\left\{\int S_t\chi \mathrm{d}\mu - \psi_*(\chi)\right\} \\
&= \inf_{\chi\in\mathfrak{X}} \sup_{\nu\in\mathcal{P}_1(\mathbb{R}_+^D)}\left\{\int S_t\chi \mathrm{d}\mu - \int \chi \mathrm{d}\nu + \psi(\nu)\right\} \\
&= \sup_{\nu\in\mathcal{P}_1(\mathbb{R}_+^D)} \inf_{\chi\in\mathfrak{X}}\left\{\int S_t\chi \mathrm{d}\mu - \int \chi \mathrm{d}\nu + \psi(\nu)\right\} \\
&= \sup_{\nu\in\mathcal{P}_1(\mathbb{R}_+^D)}\left\{\psi(\nu) - T_t(\mu,\nu)\right\},
\end{aligned}$$

where $T_t(\mu,\nu) = \sup_{\chi\in\mathfrak{X}}\left\{\int \chi \mathrm{d}\nu - \int S_t\chi \mathrm{d}\mu\right\}$ is to be compared with $\mathcal{T}_t(\mu,\nu)$ defined in (1.13). The definition of $S_t\chi$ in (5.1) implies that $\chi(x) - S_t\chi(y) \leqslant (t\xi)^*(y-x)$ for every x,y. Then, the Kantorovich duality (6.2) implies $T_t(\mu,\nu) \leqslant \mathcal{T}_t(\mu,\nu)$ and thus

$$g(t,\mu) \geqslant \sup_{\nu\in\mathcal{P}_1(\mathbb{R}_+^D)}\left\{\psi(\nu) - \mathcal{T}_t(\mu,\nu)\right\} \overset{(6.6)}{=} h(t,\mu),$$

which completes the proof of $h = g$. The additional statements on optimizers follow from $h = g$, (6.27), and (6.29). $\qquad\square$

To close this section, we record the following useful result.

Lemma 6.7 (Lipschitzness of h)**.** *Let h be given as in* (6.6)*. For every $t,t' \in \mathbb{R}_+$ and $\mu,\mu' \in \mathcal{P}_1(\mathbb{R}_+^D)$, we have*

(6.30) $$\left|h(t,\mu) - h(t',\mu')\right| \leqslant C\left|t-t'\right| + \mathsf{W}_1(\mu,\mu'),$$

where $C = \sup_{z\in\mathbb{R}_+^D:\,|z|\leqslant 1}|\xi(z)|$.

Proof. By Proposition 6.2, we can work with g given by (6.5) instead of h. Fix any t, t', μ, μ'. By the same proposition, there is $\chi \in \mathfrak{X}$ such that $g(t, \mu) = \int S_t \chi \mathrm{d}\mu - \psi_*(\chi)$. From the definition of $\mathfrak{X}$ in (3.17), we see that $\|\chi\|_{\mathrm{Lip}} \leqslant 1$. Let $\pi \in \Pi(\mu, \mu')$ be an optimal coupling for $\mathsf{W}_1(\mu, \mu')$ and let (X, X') satisfy $\mathrm{Law}(X, X') = \pi$. We have

$$g(t', \mu') - g(t, \mu) \overset{(6.5)}{\leqslant} \int S_{t'}\chi \mathrm{d}\mu' - \int S_t \chi \mathrm{d}\mu \leqslant \mathbb{E}\left[\left|S_{t'}\chi(X') - S_t\chi(X)\right|\right]$$
$$\leqslant C|t - t'| + \mathbb{E}\left[|X' - X|\right] = C|t - t'| + \mathsf{W}_1(\mu, \mu'),$$

where the last inequality follows from Lemma 5.13. This implies the desired result. $\qquad\square$

7. Proofs of the main results

In this section, we complete the proofs of Theorems 1.2, 1.3, and 1.4. We also spell out the validity of Theorem 1.1. Recall from (6.6) that we have set, for every $(t, \mu) \in \mathbb{R}_+ \times \mathcal{P}_1(\mathbb{R}_+^D)$,

$$h(t, \mu) = \sup_{\nu \in \mathcal{P}_1(\mathbb{R}_+^D)} \{\psi(\nu) - \mathcal{T}_t(\mu, \nu)\}.$$

Recall also the notation $\mathcal{L}_\mathsf{q}$ from (2.19), and that for every $\mu \in \mathcal{P}^\uparrow(\mathbb{R}_+^D)$, there exists a unique $\mathsf{q} \in \mathcal{Q}$ such that $\mu = \mathcal{L}_\mathsf{q}$. The key step towards the derivation of our main results is as follows.

Proposition 7.1. *Assume that ξ is strongly convex on $\mathbb{R}_+^D$. Let h be given by (6.6). For every $t \in \mathbb{R}_+$, $\mathsf{q} \in \mathcal{Q}_\infty$, and for $\mu = \mathcal{L}_\mathsf{q} \in \mathcal{P}_\infty^\uparrow(\mathbb{R}_+^D)$, we have*

$$(7.1) \qquad h(t, \mu) = \sup_{\nu \in \mathcal{P}_\infty^\uparrow(\mathbb{R}_+^D)} \{\psi(\nu) - \mathcal{T}_t(\mu, \nu)\}$$

$$(7.2) \qquad = \sup_{\mathsf{p} \in \mathcal{Q}_\infty} \left\{\psi\left(\mathcal{L}_{\mathsf{q} + t\nabla\xi(\mathsf{p})}\right) - t \int_0^1 \theta \, \mathrm{d}\mathcal{L}_\mathsf{p}\right\}.$$

Moreover, the suprema in (7.1) *and* (7.2) *are achieved.*

As discussed below (1.30), the key point in showing (7.1) is to transfer the monotonicity of μ into the monotonicity of a maximizer in the definition of $h(t, \mu)$. In order to do so, we will smear out the measure μ so that it becomes absolutely continuous with respect to the Lebesgue measure. The smeared-out measure is however no longer monotone, only approximately so, and our first task is to develop convenient tools that allow us to track approximate and exact monotonicity of a measure.

Lemma 7.2 (Characterization of monotonicity). *Let $d \in \mathbb{N}$ and $\mu \in \mathcal{P}(\mathbb{R}_+^d)$. The measure μ is monotone if and only if, for every $i, j \in \{1, \ldots, d\}$,*

$$(7.3) \qquad \mu^{\otimes 2}\left(\left\{(y, y') \in \mathbb{R}_+^d \times \mathbb{R}_+^d : \ y_i \geqslant y_i' \ \ or \ \ y_j \leqslant y_j'\right\}\right) = 1.$$

Remark 7.3. Let Y and Y' two independent random variables with law μ. Then, (7.3) can be rewritten as $\mathbb{P}(Y_i \geqslant Y_i' \text{ or } Y_j \leqslant Y_j') = 1$. $\qquad\square$

Proof of Lemma 7.2. We proceed in two steps. First, we show that (7.3) is equivalent to a stronger formulation and then we apply results from [40, Section 2].

Step 1. We show that the condition specified in (7.3) is equivalent to

$$(7.4) \qquad \mu^{\otimes 2} \left(\left\{ (y, y') \in \mathbb{R}_+^d \times \mathbb{R}_+^d : \ a \cdot y \geqslant a \cdot y' \ \text{ or } \ b \cdot y \leqslant b \cdot y' \right\} \right) = 1$$

for every $a, b \in \mathbb{R}_+^d$. One direction is obvious and we focus on deducing (7.4) from (7.3).

It is more convenient to work with random variables and we start with some notation. For a pair (U, V) of real-valued random variables, we say that (U, V) is monotone if, given an independent copy (U', V'), we have almost surely that $\{U \geqslant U' \text{ or } V \leqslant V'\}$. It is clear that (U, V) is monotone if and only if (V, U) is so. We need the following property. Given three random variables U, V, W, we have, for every $s, t \geqslant 0$,

(7.5)   if both (U, W) and (V, W) are monotone, then so is $(sU + tV, W)$.

This immediately follows from the definition. Indeed, let (U', V', W') be an independent copy and there is nothing to show when $W \leqslant W'$. Otherwise, we must have $U \geqslant U'$ and $V \geqslant V'$, which implies $sU + tV \geqslant sU' + tV'$.

Let Y be a random vector in $\mathbb{R}^d$ with $\mu = \text{Law}(Y)$. The condition in (7.3) is equivalent to that (Y_i, Y_j) is monotone for every $1 \leqslant i, j \leqslant d$. Fix any $a, b \in \mathbb{R}_+^d$. For each j, iteratively applying (7.5) to pairs (Y_i, Y_j) for $1 \leqslant i \leqslant d$, we can get that $(a \cdot Y, Y_j)$ is monotone. Similarly, we can get that $(a \cdot Y, b \cdot Y)$ is monotone, which is equivalent to (7.4).

Step 2. We use results from [40, Section 2] to conclude. We start with some definitions. Let S^d be the linear space of real symmetric matrices endowed with the Frobenius inner product, namely, $a \cdot b = \sum_{i,j=1}^d a_{ij} b_{ij}$ for $a, b \in S^d$. Let S_+^d be the set of positive semi-definite matrices. Let $\overline{\mu}$ be the image of μ through the map $\mathbb{R}^d \ni x \mapsto \text{diag}(x_1, \ldots, x_d)$. Since $\overline{\mu}$ is supported on diagonal matrices and diagonal entries of $a, b \in S_+^d$ are nonnegative, we can see that (7.4) is equivalent to

$$(7.6) \qquad \overline{\mu}^{\otimes 2} \left(\left\{ (y, y') \in S_+^d \times S_+^d : \ a \cdot y \geqslant a \cdot y' \ \text{ or } \ b \cdot y \leqslant b \cdot y' \right\} \right) = 1$$

for every $a, b \in S_+^d$.

Monotone probability measures on S_+^d are defined in the paragraphs above [40, Proposition 2.3] and [40, (2.1)]. The condition in (7.6) exactly matches this definition and thus $\overline{\mu}$ is monotone on S_+^d. By [40, Proposition 2.4], this monotonicity is equivalent to the existence of an increasing path $\overline{\mathsf{q}} : [0, 1) \to S_+^d$ such that $\overline{\mu} = \text{Law}(\overline{\mathsf{q}}(U))$, where the monotonicity of $\overline{\mathsf{q}}$ is understood as $\overline{\mathsf{q}}(s') - \overline{\mathsf{q}}(s) \in S_+^d$ whenever $s' \geqslant s$. Since $\overline{\mu}$ is supported on diagonal matrices, $\overline{\mathsf{q}}$ must take values in diagonal matrices. Therefore, $\overline{\mu} = \text{Law}(\overline{\mathsf{q}}(U))$ is equivalent to $\mu = \text{Law}(\mathsf{q}(U))$ for some increasing

$\mathsf{q} : [0,1) \to \mathbb{R}^d$, where q and $\overline{\mathsf{q}}$ are related by $\mathrm{diag}(\overline{\mathsf{q}}) = \mathsf{q}$ a.e. on $[0,1)$. Combining these equivalences, we obtain the desired characterization. $\square$

The next lemma is tailored for application to our future context involving a smeared-out version of the measure $\mu \in \mathcal{P}^\uparrow_\infty(\mathbb{R}^D_+)$, where we later let the smearing-out parameter tend to zero. Recall that U is a random variable with uniform distribution over $[0,1]$.

Lemma 7.4. *Let $\mu = \mathcal{L}_\mathsf{q} \in \mathcal{P}^\uparrow_\infty(\mathbb{R}^D_+)$ (see (2.19)) with $\mathsf{q} \in \mathcal{Q}_\infty$ strictly increasing in the sense that $\mathsf{q}(s) - \mathsf{q}(s') \in (0,\infty)^D$ whenever $s > s'$. For each $n \in \mathbb{N}$, let $\varpi_n \in \mathcal{P}(\mathbb{R}^{2D}_+)$ and suppose that there is an $\mathbb{R}^D_+$-increasing measurable map $T_n : \mathbb{R}^D_+ \to \mathbb{R}^D_+$ such that*

$$(7.7) \qquad x' = T_n(x) \quad \text{for } \varpi_n\text{-a.e. } (x,x').$$

Assume that ϖ_n converges weakly to some $\varpi \in \mathcal{P}(\mathbb{R}^{2D}_+)$ whose projection on the first D coordinates is equal to μ. Then, we have $\varpi \in \mathcal{P}^\uparrow(\mathbb{R}^{2D}_+)$ and thus there is $\mathsf{p} \in \mathcal{Q}$ such that $\varpi = \mathrm{Law}(\mathsf{q}(U),\mathsf{p}(U))$.

Proof. In the following, we denote by (y,y') an element in $\mathbb{R}^{2D}_+ \times \mathbb{R}^{2D}_+$. For $i,j \in \{1,\dots,2D\}$, we consider the following events on $\mathbb{R}^{2D}_+ \times \mathbb{R}^{2D}_+$,

$$E_{i,j} = \left\{ y_i \geqslant y_i' \text{ or } y_j \leqslant y_j' \right\}.$$

We also set

$$E_i = \bigcap_{j=1}^{D} E_{i,j}, \text{ for } i \in \{1,\dots,D\} \quad \text{and} \quad E = \bigcap_{i=1}^{D} E_i = \bigcap_{i,j=1}^{D} E_{i,j}.$$

For each n, let $\mu_n \in \mathcal{P}(\mathbb{R}^D_+)$ be the projection of ϖ_n to the first D coordinates. For $i,j \leqslant D$, $E_{i,j}$ can be measured by μ_n. Then, we immediately have

$$\varpi_n^{\otimes 2}(E_{i,j}) = \mu_n^{\otimes 2}(E_{i,j}) \quad \text{if } i \leqslant j \leqslant D.$$

When $i \leqslant D < j$, by (7.7), we have $y_j = \big(T_n(y_{[D]})\big)_j$ for ϖ_n-a.e. y, where $y_{[D]} = (y_k)_{1 \leqslant k \leqslant D}$. Since T_n is assumed to be $\mathbb{R}^D_+$-increasing, for $\varpi_n^{\otimes 2}$-a.e. (y,y'), we have that $y_{[D]} \leqslant y_{[D]}'$ implies $y_j \leqslant y_j'$. On the event E_i we have $y_i \geqslant y_i'$ or $y_{[D]} \leqslant y_{[D]}'$. In the latter case, it follows from the previous observation that $y_j \leqslant y_j'$ $\varpi_n^{\otimes 2}$-almost surely. Hence, we have $E_i \subseteq E_{i,j}$ and thus

$$\varpi_n^{\otimes 2}(E_{i,j}) \geqslant \varpi_n^{\otimes 2}(E_i) = \mu_n^{\otimes 2}(E_i) \quad \text{if } i \leqslant D < j.$$

Similarly on the event E we have $y_{[D]} \leqslant y_{[D]}'$ or $y_{[D]} \geqslant y_{[D]}'$ and by (7.7), we can also deduce

$$\varpi_n^{\otimes 2}(E_{i,j}) \geqslant \mu_n^{\otimes 2}(E) \quad \text{if } D < i \leqslant j.$$

Using the above three displays and the symmetry $\varpi_n^{\otimes 2}(E_{i,j}) = \varpi_n^{\otimes 2}(E_{j,i})$, we obtain

$$(7.8) \qquad \varpi_n^{\otimes 2}(E_{i,j}) \geqslant \mu_n^{\otimes 2}(E), \quad \forall i,j \in \{1,\dots,2D\}.$$

Notice that $\partial E \subseteq \bigcup_{i,j=1}^{D} \partial E_{i,j} = \bigcup_{i=1}^{D}\{y_i = y_i'\}$. Since q as in $\mu = \mathcal{L}_\mathsf{q}$ is strictly increasing, we can get

$$\mu^{\otimes 2}\left(\{y_i = y_i'\}\right) = \iint_{[0,1]^2} \mathbb{1}_{\mathsf{q}_i(s)=\mathsf{q}_i(s')}\mathrm{d}s\mathrm{d}s' = 0.$$

Hence, E is a continuity set of $\mu^{\otimes 2}$. Since $\mu \in \mathcal{P}_\infty^\uparrow(\mathbb{R}_+^D)$ is monotone, by Lemma 7.2, we have that $\mu^{\otimes 2}(E) = 1$. Also, notice that $E_{i,j}$ is a closed set. Using these and the Portmanteau theorem, we pass to the limit in (7.8) to get

$$\varpi^{\otimes 2}(E_{i,j}) \geqslant \limsup_{n\to\infty} \varpi_n^{\otimes 2}(E_{i,j}) \geqslant \lim_{n\to\infty} \mu_n^{\otimes 2}(E) = \mu^{\otimes 2}(E) = 1$$

for every i, j. Therefore, by Lemma 7.2, ϖ is monotone. Since the first marginal of ϖ is $\mu = \mathcal{L}_\mathsf{q}$, we must have $\varpi = \mathrm{Law}(\mathsf{q}(U), \mathsf{p}(U))$ for some $\mathsf{p} \in \mathcal{Q}$. $\qquad\square$

We are now ready to prove Proposition 7.1.

Proof of Proposition 7.1. We need approximations of $\mu \in \mathcal{P}_\infty^\uparrow(\mathbb{R}_+^D)$ from the collection $\mathcal{P}_{\mathrm{ac}}(\mathbb{R}_+^D)$ of probability measures on $\mathbb{R}_+^D$ that are absolutely continuous with respect to the Lebesgue measure. Let $(\mu_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{P}_{\mathrm{ac}}(\mathbb{R}_+^D)$ that converges weakly to μ. Since μ is compactly supported, we may assume that the measures μ_n are all supported in some fixed ball (independent of n). Hence, by Lemma 6.7,

$$(7.9) \qquad\qquad \lim_{n\to\infty} h(t, \mu_n) = h(t, \mu).$$

Allowed by Lemma 6.3, for each $n \in \mathbb{N}$, let $\bar{\nu}_n$ be a maximizer of $h(t, \mu_n)$ in (6.6) and let π_n be an optimal coupling of $(\mu_n, \bar{\nu}_n)$ for $\mathcal{T}_t(\mu_n, \bar{\nu}_n)$. Hence, we have

$$(7.10) \qquad\qquad h(t, \mu_n) = \psi(\bar{\nu}_n) - \mathcal{T}_t(\mu_n, \bar{\nu}_n),$$

$$(7.11) \qquad\qquad \mathcal{T}_t(\mu_n, \bar{\nu}_n) = \int (t\xi)^*(y - x)d\pi_n(x, y).$$

The plan is to show that $\bar{\nu}_n$ is approximately monotone. Then, by taking limits, we expect that the formula for $h(t, \mu)$ maximizes over monotone measures. By Lemma 6.6, there is $\chi_n \in \mathfrak{X}$ (see (3.17)) such that

$$(7.12) \qquad\qquad \int \chi_n \mathrm{d}\bar{\nu}_n - \int S_t\chi_n \mathrm{d}\mu_n = \mathcal{T}_t(\mu_n, \bar{\nu}_n).$$

Note that according to Lemma 5.3, thanks to the assumption on ξ, the function ξ^* is differentiable on $\mathbb{R}^D$. We proceed in five steps.

Step 1. Let (X_n, Y_n) be a random variable with law π_n. We show that possibly up to a redefinition of π_n, we may assume without loss of generality that $Y_n - X_n \in \nabla\xi(\mathbb{R}_+^D)$ almost surely.

Define

$$Y_n' = \nabla\xi\big(\nabla\xi^*\big((Y_n - X_n)_+\big)\big) + X_n,$$

let $\overline{\nu}'_n$ denote the law of Y'_n and let π'_n denote the law of (X_n, Y'_n). Since $x_+ - x \in \mathbb{R}^D_+$, according to (5.9) in Lemma 5.4 we have that $Y'_n - Y_n \in \mathbb{R}^D_+$ almost surely. Hence, by monotonicity of ψ_d (which we can borrow from [13, Proposition 3.6] or rederive using Proposition 3.8 and the fact that functions in $\mathfrak{X}$ are increasing), we have $\psi(\overline{\nu}'_n) \geqslant \psi(\overline{\nu}_n)$. In addition, according to Lemma 5.7 and (5.11) in Lemma 5.4, we have $\mathbb{E}[(t\xi)^*(Y'_n - X_n)] = \mathbb{E}[(t\xi)^*(Y_n - X_n)]$, so

$$\mathcal{T}_t(\mu_n, \overline{\nu}'_n) \leqslant \int (t\xi)^*(y-x)\mathrm{d}\pi'_n(x,y) = \int (t\xi)^*(y-x)\mathrm{d}\pi_n(x,y) = \mathcal{T}_t(\mu_n, \overline{\nu}_n).$$

We deduce that

$$\begin{aligned}
\psi(\overline{\nu}'_n) - \mathcal{T}_t(\mu_n, \overline{\nu}'_n) &\geqslant \psi(\overline{\nu}'_n) - \int (t\xi)^*(y-x)\mathrm{d}\pi'_n(x,y) \\
&\geqslant \psi(\overline{\nu}_n) - \mathcal{T}_t(\mu_n, \overline{\nu}_n) \\
&= h(t, \mu_n) \\
&\geqslant \psi(\overline{\nu}'_n) - \mathcal{T}_t(\mu_n, \overline{\nu}'_n),
\end{aligned}$$

where the last inequality follows from (6.6). The above display implies that $\overline{\nu}'_n$ is a maximizer in the definition (6.6) of $h(t, \mu_n)$ and that π'_n is an optimal coupling for $\mathcal{T}_t(\mu_n, \overline{\nu}')$. In addition, if (X'_n, Y'_n) is a random variable with law π'_n, we have by definition that $Y'_n - X'_n \in \nabla\xi(\mathbb{R}^D_+)$ almost surely, this means that up to replacing $(\pi_n, \overline{\nu}_n)$ by $(\pi'_n, \overline{\nu}'_n)$ (and redefining χ_n accordingly), we may assume that $Y_n - X_n \in \nabla\xi(\mathbb{R}^D_+)$ almost surely.

Step 2. We now show that

(7.13) $$y = x + t\nabla\xi\left(\nabla S_t\chi_n(x)\right) \quad \text{for } \pi_n\text{-a.e. } (x,y).$$

This implicitly states that $S_t\chi_n$ is differentiable at such points. Using (7.11) and (7.12), we have

$$\int \left(\chi_n(y) - S_t\chi_n(x) - (t\xi)^*(y-x)\right)\mathrm{d}\pi_n(x,y) = 0.$$

By the formula for $S_t\chi_n$ in (5.1), the integrand in the above display is non-positive. Therefore, we must have

$$\chi_n(y) - S_t\chi_n(x) - (t\xi)^*(y-x) = 0, \quad \text{for } \pi_n\text{-a.e. } (x,y).$$

Since $S_t\chi_n$ is Lipschitz due to Lemma 5.13, $S_t\chi_n$ is differentiable almost everywhere with respect to the Lebesgue measure (see e.g. [20, Theorem 2.10]). Hence, due to $\mu_n \in \mathcal{P}_{\mathrm{ac}}(\mathbb{R}^D_+)$ (the first marginal of π_n), we have that for π_n-a.e. (x,y), $S_t\chi_n$ is differentiable at x. By (5.3) in Proposition 5.1, we get

$$\nabla S_t\chi_n(x) = \nabla(t\xi)^*(y-x), \quad \text{for } \pi_n\text{-a.e. } (x,y).$$

Recalling from Step 1 that $y - x \in \nabla\xi(\mathbb{R}^D_+)$ for π_n-a.e. (x,y), and using Remark 5.5 we can apply $t\nabla\xi$ to both sides in the above and obtain (7.13). Since $S_t\chi_n$ is $\mathbb{R}^D_+$-convex due to Proposition 5.9, we have that

(7.14) $$\nabla S_t\chi_n \text{ is } \mathbb{R}^D_+\text{-increasing}$$

which will be used later.

Step 3. We show that $(\pi_n)_{n\in\mathbb{N}}$ is tight and derive the limit of (7.10). Let random variables (X_n, Y_n, Z_n) satisfy

$$\text{(7.15)} \qquad \text{Law}(X_n, Y_n) = \pi_n; \qquad Z_n = \nabla S_t \chi_n(X_n).$$

By (7.13), we have

$$\text{(7.16)} \qquad Y_n = X_n + t\nabla\xi(Z_n) \quad \text{a.s.}$$

Recall that χ_n given by Proposition 3.8 belongs to the collection $\mathfrak{X}$ defined in (3.17), whose definition implies that χ_n is 1-Lipschitz. By Lemma 5.13, $S_t\chi_n$ is also 1-Lipschitz. Setting $C = t\sup_{|z|\leqslant 1}|\nabla\xi(z)|$, we have

$$\text{(7.17)} \qquad |Z_n| \leqslant 1 \quad \text{and} \quad |Y_n - X_n| \leqslant C \quad \text{a.s.}$$

Since $\mu_n = \text{Law}(X_n)$ and μ_n is assumed to be supported in a fixed ball, by (7.17), Y_n and Z_n take values in a fixed compact set. In particular, the family $(X_n, Y_n, Z_n)_{n\in\mathbb{N}}$ is tight. By passing to a subsequence and invoking Skorokhod's representation theorem, we may assume that (X_n, Y_n, Z_n) converges almost surely to some (X, Y, Z), and therefore also in L^1 since these random variables are bounded uniformly over n. Since Y_n is a continuous function of X_n and Z_n as in (7.16), we also have

$$\text{(7.18)} \qquad Y = X + t\nabla\xi(Z) \quad \text{a.s.}$$

Since $\mu_n = \text{Law}(X_n)$ and μ_n converges to μ in $\mathcal{P}_1(\mathbb{R}_+^D)$, we must have $\mu = \text{Law}(X)$. We set

$$\text{(7.19)} \qquad \overline{\nu} = \text{Law}(Y) \quad \text{and} \quad \pi = \text{Law}(X, Y) \in \Pi(\mu, \overline{\nu}).$$

By (7.11) and Fatou's lemma (recall that $(t\xi)^* \geqslant -t\xi(0) = 0$ and $(t\xi)^*$ is lower semi-continuous), we have

$$\mathcal{T}_t(\mu, \overline{\nu}) \leqslant \mathbb{E}\left[(t\xi)^*(Y - X)\right] \leqslant \liminf_{n\to\infty} \mathbb{E}\left[(t\xi)^*(Y_n - X_n)\right] = \liminf_{n\to\infty} \mathcal{T}_t(\mu_n, \overline{\nu}_n).$$

This along with the continuity of ψ in $\mathcal{P}_1(\mathbb{R}_+^D)$ in (3.14) implies that

$$\psi(\overline{\nu}) - \mathcal{T}_t(\mu, \overline{\nu}) \geqslant \psi(\overline{\nu}) - \mathbb{E}\left[(t\xi)^*(Y - X)\right]$$

$$\geqslant \liminf_{n\to\infty} \psi(\overline{\nu}_n) - \mathcal{T}_t(\mu_n, \overline{\nu}_n) \stackrel{(7.9)(7.10)}{=} h(t, \mu)$$

$$\stackrel{(6.6)}{\geqslant} \psi(\overline{\nu}) - \mathcal{T}_t(\mu, \overline{\nu}).$$

This together with (7.18) yields

$$\text{(7.20)} \qquad h(t, \mu) = \psi(\overline{\nu}) - \mathcal{T}_t(\mu, \overline{\nu})$$

$$\text{(7.21)} \qquad\qquad = \psi(\text{Law}(Y)) - \mathbb{E}\left[(t\xi^*)(t\nabla\xi(Z))\right].$$

Step 4. We show the desired results under the additional assumption that

$$\text{(7.22)} \qquad \mu = \mathcal{L}_{\mathsf{q}} \quad \text{with } \mathsf{q} \text{ strictly increasing}$$

in the sense that $\mathsf{q}(s) - \mathsf{q}(s') \in (0, \infty)^D$ whenever $s > s'$. Assume (7.22), we want to show that there is $\mathsf{p}_* \in \mathcal{Q}_\infty$ such that

$$(7.23) \qquad (X, Y, Z) \stackrel{\mathrm{d}}{=} (\mathsf{q}(U), \mathsf{q}(U) + t\nabla\xi(\mathsf{p}_*(U)), \mathsf{p}_*(U))$$

where U is uniform over $[0, 1]$. Due to (7.18), we only need to find the representation for (X, Z). The assumption in (7.22) together with (7.14) and (7.15) allow us to apply Lemma 7.4 with ϖ_n, T_n, ϖ substituted by $\mathrm{Law}(X_n, Z_n)$, $\nabla S_t \chi_n$, and $\mathrm{Law}(X, Z)$, respectively. Consequently, $\mathrm{Law}(X, Z)$ is monotone. Hence, there must be an increasing path $\mathsf{p}_* : [0, 1) \to \mathbb{R}_+^D$ such that $(X, Z) \stackrel{\mathrm{d}}{=} (\mathsf{q}(U), \mathsf{p}_*(U))$. Since Z is bounded (as a consequence of (7.17)), we have $\mathsf{p}_\star \in \mathcal{Q}_\infty$ and more precisely

$$(7.24) \qquad\qquad |\mathsf{p}_\star(s)| \leqslant 1, \quad \forall s \in [0, 1).$$

Due to $\overline{\nu} = \mathrm{Law}(Y)$ (see (7.19)) and (7.23), $\overline{\nu}$ is a monotone measure and has bounded support. In other words, $\overline{\nu} = \mathcal{L}_{\mathsf{q}+t\nabla\xi(\mathsf{p}_\star)} \in \mathcal{P}_1^\uparrow(\mathbb{R}_+^D)$. In view of the definition of h in (6.6) and (7.20), we can deduce (7.1) and that the supremum is achieved at $\mathcal{L}_{\mathsf{q}+t\nabla\xi(\mathsf{p}_\star)}$. From the definition of θ in (1.9) and (5.7) in Lemma 5.4, we have that for every $x \in \mathbb{R}_+^D$,

$$(7.25) \qquad\qquad t\theta(x) = (t\xi)^*(t\nabla\xi(x)).$$

Using (7.21), (7.23), and (7.25), we have

$$(7.26) \qquad\qquad h(t, \mu) = \psi\left(\mathcal{L}_{\mathsf{q}+t\nabla\xi(\mathsf{p}_\star)}\right) - \int \theta \, \mathrm{d}\mathcal{L}_{\mathsf{p}_\star}.$$

Due to (6.6) and the definition of the optimal transport problem $\mathcal{T}_t$ in (1.13), we see that $h(t, \mu)$ is always an upper bound for the term in (7.2). Then, the above display implies the equality in (7.2) and that the supremum is achieved at $\mathsf{p}_\star$. Hence, we have shown (7.1) and (7.2) under the assumption (7.22) and the suprema are achieved at $\mathcal{L}_{\mathsf{q}+t\nabla\xi(\mathsf{p}_\star)}$ and $\mathsf{p}_\star$, respectively.

Step 5. We show the results for general $\mu = \mathcal{L}_\mathsf{q}$ without assuming (7.22). We employ an approximation argument. For $k \in \mathbb{N}$, define a strictly increasing path q_k by setting $\mathsf{q}_k(s) = \mathsf{q}(s) + k^{-1}s\mathbf{1}$ for $s \in [0, 1)$, where $\mathbf{1} = (1, 1, \ldots, 1) \in \mathbb{R}_+^D$ consists of ones as entries. So, q_k converges to q pointwise. We take $\mu_k = \mathcal{L}_{\mathsf{q}_k} \in \mathcal{P}_\infty^\uparrow(\mathbb{R}_+^D)$ and thus μ_k converges weakly to μ. Let $\mathsf{p}_{\star,k}$ be the corresponding path appearing in (7.23). Due to (7.24), $(\mathsf{p}_{\star,k})_{k\in\mathbb{N}}$ is a family of increasing paths that are uniformly bounded. By [13, Lemma 3.4], this allows us to assume that $\mathsf{p}_{\star,k}$ converges to $\mathsf{p}_\star \in \mathcal{Q}_\infty$ a.e. on $[0, 1)$ after passing to a subsequence.

Then, (7.1) and (7.2) follow from the following convergences. We have that $h(t, \mu_k)$ converges to $h(t, \mu)$ as $k \to \infty$ by Lemma 6.7. For each $\nu \in \mathcal{P}_\infty^\uparrow(\mathbb{R}_+^D)$, we have $\mathcal{T}_t(\mu_k, \nu)$ converges to $\mathcal{T}_t(\mu, \nu)$ due to boundedness of the measures and the weak convergence of μ_k. By the continuity of ψ as in (3.14), $\psi\left(\mathcal{L}_{\mathsf{q}_k+t\nabla\xi(\mathsf{p})}\right)$ converges to $\psi\left(\mathcal{L}_{\mathsf{q}+t\nabla\xi(\mathsf{p})}\right)$ at any $\mathsf{p} \in \mathcal{Q}_\infty$.

Recall that, at each μ_k, the relevant suprema are achieved at $\mathcal{L}_{\mathsf{q}_k + t\nabla\xi(\mathsf{p}_{\star,k})}$ and $\mathsf{p}_{\star,k}$, respectively. By similar considerations, we can show that the suprema at μ are achieved at $\mathcal{L}_{\mathsf{q}+t\nabla\xi(\mathsf{p}_\star)}$ and $\mathsf{p}_\star$, respectively. $\qquad\square$

Proposition 7.1 requires ξ to be strongly convex on $\mathbb{R}_+^D$. Since this may not necessarily always be the case, we will introduce a small perturbation of ξ that enforces this property, and argue by continuity to obtain the main results. For every $\lambda \geq 0$, we set

$$(7.27) \qquad \xi_\lambda(x) = \xi(x) + \frac{\lambda}{2}|x|^2.$$

We clearly have that for every $\lambda > 0$, ξ_λ is strongly convex on $\mathbb{R}_+^D$. Setting

$$(7.28) \qquad H_N^\lambda(\sigma) = H_N(\sigma) + \frac{\sqrt{\frac{\lambda}{2}}}{\sqrt{N}} \sum_{d=1}^D \sum_{i,j\in I_{N,d}} J_{ij}^d \sigma_i \sigma_j,$$

where J_{ij}^d are independent $\mathcal{N}(0,1)$ random variables that are each independent of H_N, we have that

$$(7.29) \qquad \mathbb{E} H_N^\lambda(\sigma) H_N^\lambda(\sigma') = N \xi_\lambda(R_N(\sigma,\sigma')).$$

We let $\overline{F}_N^\lambda$, g^λ, and h^λ denote the quantities defined in (2.14), (6.5), and (6.6) with ξ and H_N replaced by ξ_λ and H_N^λ therein. At $\lambda = 0$, we have $\xi_0 = \xi$ and $H_N^0 = H_N$, therefore $\overline{F}_N^0 = \overline{F}_N$, $g^0 = g$, and $h^0 = h$. We now show the continuity of these quantities in λ.

Lemma 7.5. *For every* $(t,\mu) \in \mathbb{R}_+ \times \mathcal{P}_1(\mathbb{R}_+^D)$, *we have that the quantities* $\lim_{N\to+\infty} \overline{F}_N^\lambda(t,\mu)$, $g^\lambda(t,\mu)$, *and* $h^\lambda(t,\mu)$ *are continuous functions of* $\lambda \in \mathbb{R}_+$.

Proof. As in [31, Proposition 6.1], we can differentiate $\overline{F}_N^\lambda(t,\mu)$ with respect to λ and use Gaussian integration by parts to observe that the derivative is bounded uniformly in N. This yields Lipschitz continuity in λ of $\lim_{N\to+\infty} F_N^\lambda(t,\mu)$.

For every $\chi \in \mathfrak{X}$, we let $S_t^\lambda \chi$ denote the quantity obtained in (5.1) with ξ replaced by ξ_λ therein. By Lemma 5.12 and since $\chi^*(p) = +\infty$ when $p \in \mathbb{R}_+^D \smallsetminus [0,1]^D$, we have

$$S_t^\lambda \chi(x) = \sup_{p\in[0,1]^D} \{x \cdot p - \chi^*(p) + t\xi_\lambda(p)\}.$$

This implies that

$$|S_t^\lambda \chi(x) - S_t^{\lambda'}\chi(x)| \leq \frac{Dt}{2}|\lambda - \lambda'|,$$

and thus, for every $\mu \in \mathcal{P}_1(\mathbb{R}_+^D)$, we have

$$\left| \int S_t^\lambda \chi \mathrm{d}\mu - \int S_t^{\lambda'}\chi \mathrm{d}\mu \right| \leq \frac{Dt}{2}|\lambda - \lambda'|.$$

The quantity $g^\lambda(t,\mu)$ is therefore Lipschitz continuous in λ since using (6.5) it can be written as the infimum of a family of uniformly Lipschitz functions of λ. Finally, from Proposition 6.2 we have that $h^\lambda(t,\mu) = g^\lambda(t,\mu)$, so $h^\lambda(t,\mu)$ is also Lipschitz continuous in λ. $\qquad\square$

We have the following immediate corollary.

Corollary 7.6. *Let* $(t,\mu) \in \mathbb{R}_+ \times \mathcal{P}_\infty^\uparrow(\mathbb{R}_+^D)$ *and let* $h(t,\mu)$ *be given as in* (6.6). *We have* $\lim_{N\to\infty} \overline{F}_N(t,\mu) = h(t,\mu)$.

Proof. Let $\lambda > 0$, since ξ_λ is strongly convex on $\mathbb{R}_+^D$, it follows from Proposition 2.1 and (7.2) in Proposition 7.1 that

$$\lim_{N\to\infty} \overline{F}_N^\lambda(t,\mu) = h^\lambda(t,\mu).$$

We can then let $\lambda \to 0$ and use Lemma 7.5 to get the desired result. $\qquad\square$

We can now complete the proofs of Theorems 1.3 and 1.4.

Proof of Theorem 1.3. Let $\mu = \mathcal{L}_{\mathsf{q}} \in \mathcal{P}_\infty^\uparrow(\mathbb{R}_+)$. Due to $\mathcal{P}_\infty^\uparrow(\mathbb{R}_+^D) \subseteq \mathcal{P}_1(\mathbb{R}_+^D)$, the definition of h in (6.6) and (7.1) in Proposition 7.1 yield

$$(7.30) \qquad h(t,\mu) = \sup_{\nu \in \mathcal{P}_\infty(\mathbb{R}_+^D)} \{\psi(\nu) - \mathcal{T}_t(\mu,\nu)\}.$$

The formula (1.14) follows from (7.30) and Corollary 7.6. The concavity of the functional follows from Proposition 3.5 and (6.2). The last claim on the uniqueness of maximizers that are monotone follows from Corollary 3.7. $\quad\square$

Proof of Theorem 1.4. This is a consequence of Corollary 7.6, the identity $h = g$ given in Proposition 6.2, and the expression of g in (6.5). $\qquad\square$

Before proceeding with the proofs of Theorems 1.1 and 1.2, we record the following result.

Proposition 7.7 (Other forms of Parisi formula)**.** *For every* $t \geqslant 0$ *and* $\mu = \mathcal{L}_q \in \mathcal{P}_\infty^\uparrow(\mathbb{R}_+^D)$, *we have*

$$(7.31) \qquad \lim_{N\to+\infty} \overline{F}_N(t,\mu) = \sup_{\mathsf{q}'\in\mathcal{Q}_\infty} \left\{ \psi\left(\mathcal{L}_{\mathsf{q}+\mathsf{q}'}\right) - \int (t\xi)^* \, \mathrm{d}\mathcal{L}_{\mathsf{q}'} \right\}$$

$$(7.32) \qquad\qquad = \sup_{\mathsf{q}''\in\mathcal{Q}_\infty} \left\{ \psi\left(\mathcal{L}_{\mathsf{q}''}\right) - \int_0^1 (t\xi)^* \left(\mathsf{q}''(s) - \mathsf{q}(s)\right) \mathrm{d}s \right\}.$$

Proof. We denote by $\mathrm{v}(\mathrm{X})$ the variational formula in display (X) and by $\mathrm{v}^\lambda(\mathrm{X})$ the variational formula in display (X) where ξ is replaced by ξ_λ therein. We have

$$\mathrm{v}(7.2) \leqslant \mathrm{v}(7.31) \leqslant \mathrm{v}(7.32) \leqslant \mathrm{v}(7.1),$$

where the first inequality follows from the identity in (7.25) and the last inequality follows from $\int_0^1 (t\xi)^* (\mathsf{q}' - \mathsf{q}) \geqslant \mathcal{T}_t(\mathcal{L}_{\mathsf{q}}, \mathcal{L}_{\mathsf{q}'})$ (due to the definition of $\mathcal{T}_t$ in (1.13)). We now argue that $\mathrm{v}(7.1) \leqslant \mathrm{v}(7.2)$.

For every $\pi \in \mathcal{P}(\mathbb{R}_+^D \times \mathbb{R}_+^D)$ we have by monotone convergence

$$\lim_{\lambda \to 0} \int (t\xi_\lambda^*)(y-x)\mathrm{d}\pi(x,y) = \int (t\xi^*)(y-x)\mathrm{d}\pi(x,y)$$

In particular, writing

$$\mathrm{v}^\lambda(7.1) = \sup_{\pi \in \Pi(\mu,\cdot)} \left\{ \psi(\pi_2) - \int (t\xi_\lambda^*)(y-x)\mathrm{d}\pi(x,y) \right\},$$

we see that $\mathrm{v}^\lambda(7.1)$ is lower semi-continuous in λ as a supremum of continuous functions of λ. By Proposition 2.1 we know that for every $\lambda \geq 0$, $\lim_{N \to \infty} \overline{F}_N^\lambda(t,\mu) = \mathrm{v}^\lambda(7.2)$, so according to Lemma 7.5, $\mathrm{v}^\lambda(7.2)$ is a continuous function of $\lambda \in \mathbb{R}_+$. From Proposition 7.1 we have for $\lambda > 0$, $\mathrm{v}^\lambda(7.2) = \mathrm{v}^\lambda(7.1)$. Letting $\lambda \to 0$ in this we obtain $\mathrm{v}(7.1) \leq \mathrm{v}(7.2)$. This completes the proof since $\lim_{N \to \infty} \overline{F}_N(t,\mu) = \mathrm{v}(7.2)$ by Proposition 2.1. $\qquad\square$

Proof of Theorem 1.1. The formula in (1.11) is given by Proposition 2.1 at $\mu = \delta_0$. The formula (1.12) is given by (7.31) at $\mu = \delta_0$. $\qquad\square$

Proof of Theorem 1.2. We start by proving that the variational formula in (1.12) admits a maximizer $\overline{\nu}$. According to the refined version of the Parisi formula given in [9, Proposition 6.1], we know that the supremum in (7.2) can be taken over $\mathsf{p} \in \mathcal{Q}_\infty$ satisfying $|\mathsf{p}|_{L^\infty} \leq C$ for some fixed constant $C > 0$. By [13, Lemma 3.4], for every $p \in [1,+\infty)$, the set of $\mathsf{p} \in \mathcal{Q}_\infty$ satisfying $|\mathsf{p}|_{L^\infty} \leq C$ is compact for the convergence in L^p. Therefore, since by (2.15) the function ψ is Lipschitz continuous with respect to the topology of L^1 convergence, there exists a maximizer $\overline{\mathsf{p}} \in \mathcal{Q}_\infty$ for the variational formula in (7.2). Choosing $\overline{\mu} = \mathcal{L}_{\overline{\mathsf{p}}}$ and $\overline{\nu} = \nabla\xi(\overline{\mu})$, we have that $\overline{\mu}$ and $\overline{\nu}$ are maximizers of the variational formulas in (1.11) and (1.12) respectively.

Let $\overline{\mu}, \overline{\nu}$ be maximizers of formulas (1.11) and (1.12) respectively. Recall the expression of $t\theta$ in (7.25). By Theorem 1.3 at $\mu = \delta_0$ and the simple fact due to (1.13) that $\mathcal{T}_t(\delta_0, \nu) = \int (t\xi)^* \mathrm{d}\nu$ for any ν, we have that $(t\nabla\xi)(\overline{\mu})$ and $\overline{\nu}$ are maximizers of the variational formula in (1.14) at $\mu = \delta_0$. The uniqueness thus follows from the said theorem. Consequently, we also have $\overline{\nu} = (t\nabla\xi)(\overline{\mu})$. $\qquad\square$

We can in fact extend the statement concerning the uniqueness of Parisi measures in the following way.

Proposition 7.8 (Uniqueness of Parisi measure II)**.** *For every $t \geq 0$, the suprema in (7.31) and (7.32) are each achieved at exactly one path, say $\overline{\mathsf{q}}'$ and $\overline{\mathsf{q}}''$ respectively, and these paths satisfy $\mathsf{q} + \overline{\mathsf{q}}' = \overline{\mathsf{q}}''$. Moreover, the supremum in (2.20) is achieved at paths $\overline{\mathsf{p}}$ satisfying $t\nabla\xi(\overline{\mathsf{p}}) = \overline{\mathsf{q}}'$.*

Proof. For the existence of maximizers, as explained in the proof of Theorem 1.2 a simple compactness and continuity argument yields the existence of a maximizer $\overline{\mathsf{p}} \in \mathcal{Q}_\infty$ for the variational formula in (7.2). Then, the suprema in (7.31) and (7.32) are achieved at $\overline{\mathsf{q}}' = t\nabla\xi(\overline{\mathsf{p}})$ and $\overline{\mathsf{q}}'' = \mathsf{q} + \overline{\mathsf{q}}'$ respectively.

This verifies the existence part. For the uniqueness part, notice that it suffices to show that for (7.32). Let $\overline{\mathsf{q}}''$ be any maximizer of (7.32), and let us write $\overline{\nu} = \mathcal{L}_{\overline{\mathsf{q}}''} \in \mathcal{P}^{\uparrow}_{\infty}(\mathbb{R}^D_+)$ and $F = \lim_{N \to \infty} \overline{F}_N(t, \mu)$. Then, $\mathrm{Law}(\mathsf{q}(U), \overline{\mathsf{q}}''(U))$ is a coupling of $(\mu, \overline{\nu})$ and we have

$$F = \psi\left(\mathcal{L}_{\overline{\mathsf{q}}''}\right) - \int_0^1 (t\xi)^* \left(\overline{\mathsf{q}}''(s) - \mathsf{q}(s)\right) \mathrm{d}s \overset{(1.13)}{\leqslant} \psi\left(\overline{\nu}\right) - \mathcal{T}_t(\mu, \overline{\nu}) \overset{(1.14)}{\leqslant} F.$$

Hence, the supremum in (1.14) is achieved at $\overline{\nu}$. By Theorem 1.3, such $\overline{\nu}$ is unique in $\mathcal{P}^{\uparrow}_{\infty}(\mathbb{R}^D_+)$ and thus $\overline{\mathsf{q}}''$ has to be unique. Similarly if $\overline{\mathsf{q}}'$ is a maximizer in (7.31), then $\mathsf{q} + \overline{\mathsf{q}}'$ is a maximizer in (7.32). Thus $\overline{\mathsf{q}}'$ has to be unique and satisfy $\mathsf{q} + \overline{\mathsf{q}}' = \overline{\mathsf{q}}''$. $\qquad\square$

8. Explicit representation in terms of martingales

The goal of this section is to prove Theorem 1.5. One ingredient of the proof of Theorem 1.4 is a representation of the function ψ as, for every $\mu \in \mathcal{P}^{\uparrow}_1(\mathbb{R}^D_+)$,

$$(8.1) \qquad \psi(\mu) = \inf_{\chi \in \mathfrak{X}} \left\{ \int \chi \, \mathrm{d}\mu - \psi_*(\chi) \right\}.$$

In this representation, the set $\mathfrak{X}$ and function ψ_* are not uniquely determined. In order to derive Theorem 1.5, the main task compared with the proof of Theorem 1.4 is to derive a more explicit representation of this function ψ. To do this, we will rely heavily on ideas developed in [42], where very similar results are proved for models with scalar spins.

We recall that we give ourselves a probability space $\mathscr{P} = (\Omega, \mathcal{F}, \mathbf{P})$ with associated expectation $\mathbf{E}$, and $(\mathcal{F}_1(t))_{t \geqslant 0}, \ldots, (\mathcal{F}_D(t))_{t \geqslant 0}$ an independent family of complete filtrations over $\mathscr{P}$, which each comes with an adapted Brownian motion $(B_1(t))_{t \geqslant 0}, \ldots, (B_D(t))_{t \geqslant 0}$. We denote by $\mathbf{Mart}_1, \ldots,$ $\mathbf{Mart}_D$ the spaces of bounded martingales over $\mathscr{P}$ with respect to the filtrations $(\mathcal{F}_1(t))_{t \geqslant 0}, \ldots, (\mathcal{F}_D(t))_{t \geqslant 0}$ respectively, with $\mathbf{Mart} = \prod_{d=1}^{D} \mathbf{Mart}_d$. When we want to make the dependence on the underlying probability space explicit we write $\mathbf{Mart}(\mathscr{P})$ in place of $\mathbf{Mart}$.

We recall that the functions $\psi_d : \mathcal{P}_1(\mathbb{R}_+) \to \mathbb{R}$, $\phi_d : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}$ and $\phi_d^* : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}$ are defined in (2.16), (1.21), and (1.22) respectively. Since we have assumed that $\boldsymbol{\pi}_d$ is not a Dirac mass, it follows that for every $t \geqslant 0$, $\phi_d(t, \cdot)$ is strictly convex on $\mathbb{R}$. Adapting the proof of [42, Lemma 2.2], we obtain the following explicit representation for ψ_d.

Proposition 8.1. *For every* $d \in [D]$, $T_d \in \mathbb{R}_+$, *and* $\nu_d \in \mathcal{P}_{\infty}(\mathbb{R}_+)$ *supported in* $[0, T_d]$, *we have*

$$(8.2) \quad \psi_d(\nu_d) = \inf_{\alpha_d \in \mathbf{Mart}_d} \mathbf{E}\Bigg[\phi_d^*(T_d, \alpha_d(T_d)) - \sqrt{2}\alpha_d(T_d) \cdot B_d(T_d)$$
$$- \int_0^{T_d} \nu_d[0, t]\alpha_d^2(t) \, \mathrm{d}t \Bigg].$$

In addition, there is a unique minimizer $\alpha_d \in \mathbf{Mart}_d$ in (8.2) *and it must satisfy $\alpha_d(t) = \partial_x \Phi_{\nu_d}(t, X_d(t))$ a.s. for almost every $t \leqslant T_d$, where X_d is the unique strong solution to*

$$(8.3) \qquad \begin{cases} dX_d(t) = 2\nu_d[0,t]\, \partial_x \Phi_{\nu_d}(t, X_d(t))\, dt + \sqrt{2}\, dB_d(t), \\ X_d(0) = 0. \end{cases}$$

We next recall that, for each $\alpha \in \mathbf{Mart}$, we have defined $\chi_\alpha : \mathbb{R}_+^D \to \mathbb{R}$ in (1.20).

Proposition 8.2. *For every $\nu \in \mathcal{P}_\infty^\uparrow(\mathbb{R}_+^D)$ with marginals $\nu_1, \ldots, \nu_D$ and $T = (T_1, \ldots, T_D)$ such that for every $d \in [D]$, the measure ν_d is supported in $[0, T_d]$, we have*

$$(8.4) \quad \psi(\nu) = \inf_{\alpha \in \mathbf{Mart}} \left\{ \int \chi_\alpha \, d\nu - \chi_\alpha(T) \right. $$
$$\left. + \sum_{d=1}^{D} \lambda_{\infty,d} \mathbf{E}\left[\phi_d^*(T_d, \alpha_d(T_d)) - \sqrt{2}\alpha_d(T_d) \cdot B_d(T_d) \right] \right\}.$$

In addition, there is a unique minimizer $\alpha \in \mathbf{Mart}$ in (8.4) *and it must satisfy $\alpha_d(t) = \partial_x \Phi_{\nu_d}(t, X_d(t))$ a.s. for every $d \in [D]$ and almost every $t \leqslant T_d$, where X_d is the solution to* (8.3).

Proof. By integration by parts, we have for every $d \in [D]$ that

$$\int_0^{T_d} \nu_d[0,t]\mathbf{E}\left[\alpha_d^2(t)\right] dt = \int \int_s^{T_d} \mathbf{E}\left[\alpha_d^2(t)\right] dt \, d\nu_d(s),$$

and thus, due to (1.20),

$$\sum_{d=1}^{D} \lambda_{\infty,d} \int_0^{T_d} \nu_d[0,t]\mathbf{E}\left[\alpha_d^2(t)\right] dt = \chi_\alpha(T) - \int \chi_\alpha \, d\nu.$$

Using this, the decomposition of ψ into ψ_d as in (1.8), and the formula (8.2), we obtain (8.4). $\qquad \square$

Given $T = (T_1, \ldots, T_d)$, for every $\alpha \in \mathbf{Mart}$, we let

$$\varphi(\alpha) = -\chi_\alpha(T) + \sum_{d=1}^{D} \lambda_{\infty,d} \mathbf{E}\left[\phi_d^*(T_d, \alpha_d(T_d)) - \sqrt{2}\alpha_d(T_d) \cdot B_d(T_d) \right],$$

and

$$K_0 = \left\{ (\varphi(\alpha), (\mathbf{E}[\alpha_1^2(t)])_{t \leqslant T_1}, \ldots, (\mathbf{E}[\alpha_D^2(t)])_{t \leqslant T_D}) \, \middle| \, \alpha \in \mathbf{Mart}, \right.$$
$$\left. \mathbf{E}[\phi_d^*(T_d, \alpha_d(T_d))] < +\infty \right\}.$$

We also let K denote the closed convex hull of K_0. Since $\mathbf{Mart}$ depends on the underlying probability space $\mathscr{P}$, the sets K and K_0 also depend on $\mathscr{P}$. When we want to make this dependence explicit, we will write $K(\mathscr{P})$ and $K_0(\mathscr{P})$. For convenience we will also let $\mathcal{K}_T = \prod_d [0, T_d]$.

Proposition 8.3. *For every $\mu \in \mathcal{P}_\infty^\uparrow(\mathbb{R}_+^D)$ and $T = (T_1, \ldots, T_D)$ such that* $\mathrm{supp}(\mu) + t\nabla\xi([0,1]^D) \subseteq \prod_{d\in[D]}[0,T_d]$, *we have*

$$(8.5) \qquad \lim_{N\to+\infty} \overline{F}_N(t,\mu) = \inf_{(\varphi,\gamma)\in K} \left\{ \int S_t\chi_\gamma \mathrm{d}\mu - \varphi \right\},$$

where χ_γ is given by $\chi_\gamma(x) = \sum_{d=1}^D \lambda_{\infty,d} \int_0^{x_d} \gamma_d(s)\mathrm{d}s$.

Proof. According to Theorem 1.3, we have

$$\lim_{N\to+\infty} \overline{F}_N(t,\mu) = \sup_{\nu\in\mathcal{P}_\infty(\mathbb{R}_+^D)} \left\{ \psi(\nu) - \mathcal{T}_t(\mu,\nu) \right\}.$$

In view of the more precise statement of the Parisi formula for multi-species spin glasses given in [9, Proposition 6.1], the supremum in the previous display is reached at a probability measure supported on $\mathrm{supp}(\mu) + t\nabla\xi([0,1]^D)$, the assumption on T implies

$$\mathrm{supp}(\mu) + t\nabla\xi([0,1]^D) \subseteq \mathcal{K}_T.$$

So, the supremum in the first display of this proof can be restricted to $\nu \in \mathcal{P}(\mathcal{K}_T)$ and thus, thanks to Proposition 8.2, we have

$$\lim_{N\to+\infty} \overline{F}_N(t,\mu) = \sup_{\nu\in\mathcal{P}(\mathcal{K}_T)} \inf_{\alpha\in\mathbf{Mart}} \left\{ \int \chi_\alpha\, \mathrm{d}\nu - \chi_\alpha(T) \right.$$
$$\left. + \sum_{d=1}^D \lambda_{\infty,d}\mathbf{E}\left[\phi_d^*(T_d, \alpha_d(T_d)) - \sqrt{2}\alpha_d(T_d) \cdot B_d(T_d) \right] - \mathcal{T}_t(\mu,\nu) \right\}.$$

For $(\varphi,\gamma) \in K$, let $G_{t,\mu}(\nu;\varphi,\gamma) = \int \sum_d \lambda_{d,\infty}\Gamma_d\mathrm{d}\nu - \varphi - \mathcal{T}_t(\mu,\nu)$ where we have set $\Gamma_d(x) = \int_0^{x_d} \gamma_d(s)\mathrm{d}s$. The previous display can be rewritten as

$$\lim_{N\to+\infty} \overline{F}_N(t,\mu) = \sup_{\nu\in\mathcal{P}(\mathcal{K}_T)} \inf_{(\varphi,\gamma)\in K_0} G_{t,\mu}(\nu;\varphi,\gamma),$$

Since $G_{t,\mu}(\nu;\cdot)$ is affine, we can replace K_0 by K in the previous display, to obtain

$$\lim_{N\to+\infty} \overline{F}_N(t,\mu) = \sup_{\nu\in\mathcal{P}(\mathcal{K}_T)} \inf_{(\varphi,\gamma)\in K} G_{t,\mu}(\nu;\varphi,\gamma).$$

Observe that the function $G_{t,\mu}(\nu;\cdot)$ is continuous on $\mathbb{R} \times L^1$ and $G_{t,\mu}(\cdot;\varphi,\gamma)$ is lower semicontinuous with-respect to the topology of weak convergence. In addition, $G_{t,\mu}(\nu;\cdot)$ is convex and $G_{t,\mu}(\cdot;\varphi,\gamma)$ is concave because of (6.2). Therefore, we can apply Sion's min-max theorem and obtain

$$\lim_{N\to+\infty} \overline{F}_N(t,\mu) = \inf_{(\varphi,\gamma)\in K} \sup_{\nu\in\mathcal{P}(\mathcal{K}_T)} G_{t,\mu}(\nu;\varphi,\gamma).$$

From Lemma 6.4, Remark 6.5 and the fact that $\mathrm{supp}(\mu) + t\nabla\xi([0,1]^D) \subseteq \mathcal{K}_T$, we have that for every $(\varphi,\gamma) \in K$,

$$\sup_{\nu\in\mathcal{P}(\mathcal{K}_T)} G_{t,\mu}(\nu;\varphi,\gamma) = \int S_t\chi_\gamma\mathrm{d}\mu - \varphi,$$

where χ_γ is given by $\chi_\gamma(x) = \sum_{d=1}^D \lambda_{\infty,d} \int_0^{x_d} \gamma_d(s)\mathrm{d}s$. This yields (8.5). $\qquad\square$

Proposition 8.4. *Assume that there exists a uniform random variable* $U : \mathscr{P} \to [0,1]$ *that is* $\mathcal{F}_d(0)$*-measurable for every* $d \in [D]$*. Then* (1.23) *holds with* **Mart** = **Mart**$(\mathscr{P})$*.*

Proof. Since $K_0 \subseteq K$, Proposition 8.3 implies

$$(8.6) \qquad \lim_{N \to +\infty} \overline{F}_N(t,\mu) \leqslant \inf_{\alpha \in \mathbf{Mart}(\mathscr{P})} \left\{ \int S_t \chi_\alpha \mathrm{d}\mu - \varphi(\alpha) \right\}.$$

So we only need to prove the converse bound. Let $\mathscr{W}$ denote the D-dimensional Wiener space with the canonical random variables $(W(t))_{t \geqslant 0}$ equipped with D filtrations, each generated by a component of the canonical random variables. Let $K_1(\mathscr{W})$ be the convex hull of $K_0(\mathscr{W})$. We can apply Proposition 8.3 to the probability space $\mathscr{W}$, and obtain from (8.5) that (we can replace $K(\mathscr{W})$ by $K_1(\mathscr{W})$ using a continuity argument)

$$\lim_{N \to +\infty} \overline{F}_N(t,\mu) = \inf_{(\varphi,\gamma) \in K_1(\mathscr{W})} \left\{ \int S_t \chi_\gamma \mathrm{d}\mu - \varphi \right\}.$$

Now, let $(\varphi,\gamma) \in K_1(\mathscr{W})$. By the definition of $K_1(\mathscr{W})$ as the convex hull of $K_0(\mathscr{W})$, there is an integer $n \in \mathbb{N}$ and constants $c_1, \ldots, c_n \in [0,1]$ such that $\sum_i c_i = 1$ and $\alpha^{(1)}, \ldots, \alpha^{(n)} \in \mathbf{Mart}(\mathscr{W})$ such that

$$\varphi = \sum_{i=1}^n c_i \varphi(\alpha^{(i)}),$$

$$\gamma_d(t_d) = \sum_{i=1}^n c_i \mathbf{E}[(\alpha_d^{(i)}(t_d))^2].$$

For every $d \in [D]$, there is a canonical embedding from $\mathbf{Mart}_d(\mathscr{W})$ to the sets of martingales in $\mathbf{Mart}_d(\mathscr{P})$ that are independent of $\mathcal{F}_d(0)$. This is because we can identify W_d with the Brownian motion B_d. Thus, for every $i \leqslant n$ and $d \leqslant D$, we think of $\alpha_d^{(i)}$ as an element of $\mathbf{Mart}_d(\mathscr{P})$ that is independent of $\mathcal{F}_d(0)$. Now, by the assumption on $\mathscr{P}$ in the statement of Proposition 8.4, there exists a random variable $I : \mathscr{P} \to \{1, \ldots, n\}$ such that

$$\mathbb{P}(I = i) = c_i, \quad \forall 1 \leqslant i \leqslant n,$$

and such that I is $\mathcal{F}_d(0)$-measurable for every $d \in [D]$. We let

$$\beta_d(t) = \alpha_d^{(I)}(t)$$

Since I is $\mathcal{F}_d(0)$-measurable and $\alpha_d^{(i)} \in \mathbf{Mart}_d(\mathscr{P})$ is independent of $\mathcal{F}_d(0)$, we have that $\beta \in \mathbf{Mart}(\mathscr{P})$. Furthermore, it is easily verified that

$$\varphi = \varphi(\beta),$$

$$\gamma_d(t_d) = \mathbf{E}[(\beta_d(t_d))^2].$$

From this, it follows that

$$\int S_t \chi_\gamma \mathrm{d}\mu - \varphi = \int S_t \chi_\beta \mathrm{d}\mu - \varphi(\beta) \geqslant \inf_{\alpha \in \mathbf{Mart}(\mathscr{P})} \left\{ \int S_t \chi_\alpha \mathrm{d}\mu - \varphi(\alpha) \right\}.$$

and thus

$$\lim_{N \to +\infty} \overline{F}_N(t,\mu) = \inf_{(\varphi,\gamma) \in K_1(\mathscr{W})} \left\{ \int S_t \chi_\gamma \mathrm{d}\mu - \varphi \right\}$$

$$\geqslant \inf_{\alpha \in \mathbf{Mart}(\mathscr{P})} \left\{ \int S_t \chi_\alpha \mathrm{d}\mu - \varphi(\alpha) \right\}.$$

This along with (8.6) implies (1.23). □

Proposition 8.5. *There exists a probability space $\overline{\mathscr{P}}$ such that* (1.23) *holds with $\mathbf{Mart} = \mathbf{Mart}(\overline{\mathscr{P}})$ and the infimum appearing there is achieved.*

Proof. We let $\mathscr{P}$ be a probability space such that Proposition 8.4 holds and let $\alpha^{(n)} \in \mathbf{Mart}(\mathscr{P})$ be such that

$$\lim_{n \to +\infty} \int S_t \chi_{\alpha^{(n)}} \mathrm{d}\mu - \varphi(\alpha^{(n)}) = \lim_{N \to +\infty} \overline{F}_N(t,\mu).$$

Since Proposition 8.4 holds, for n large enough, in view of (1.23), we must have $\mathbf{E}[\phi_d^*(T_d, \alpha_d^{(n)}(T_d))] < +\infty$. The definition of ϕ_d in (1.21) and that of ϕ_d^* in (1.22) imply that, for each $d \in [D]$ and for $t \leqslant T$, $\alpha_d(t)$ is valued in $[-1,1]$ a.s. In particular, the sequence $(((\alpha_d^{(n)}(t))_{0 \leqslant t \leqslant T_d})_{d \in [D]})_{n \in \mathbb{N}}$ (indexed by n) is tight. Applying Prokhorov's theorem and Skorokhod's representation theorem, not relabeling the extraction, we obtain a probability space $\overline{\mathscr{P}} = (\overline{\Omega}, \overline{\mathcal{F}}, \overline{\mathbb{P}})$ and random variables $(B, (\alpha_1(t_1), \ldots, \alpha_D(t_D))_{t_d \in [0,T_d] \cap \mathbb{Q}})$ such that, for every integer $l \in \mathbb{N}$, every continuous bounded function $J : \mathbb{R} \times (\mathbb{R}^D)^l \to \mathbb{R}$, and $t^i = (t_d^i)_{d \in [D]} \in \prod_{d \in [D]}[0, T_d]$ for $1 \leqslant i \leqslant l$, we have

$$\lim_{n \to +\infty} \mathbf{E}J\left(B, \alpha^{(n)}(t^1), \ldots, \alpha^{(n)}(t^l)\right) = \overline{\mathbf{E}}J\left(B, \alpha(t^1), \ldots, \alpha(t^l)\right)$$

where $\alpha^{(n)}(t^i) = (\alpha_d^{(n)}(t_d^i))_{d \in [D]}$. For $t_d \leqslant T_d$, we let $\overline{\mathcal{F}}_d(t_d)$ denote the σ-algebra generated by $(B_d(s_d))_{s_d \in [0,t_d] \cap \mathbb{Q}}$ and $(\alpha_d(s_d))_{s_d \in [0,t_d] \cap \mathbb{Q}}$. We set $\alpha_d(t_d) = \overline{\mathbf{E}}[\alpha_d(T_d)|\overline{\mathcal{F}}_d(t_d)]$, by construction $\alpha_d \in \mathbf{Mart}_d(\overline{\mathscr{P}})$ with respect to the filtration $(\overline{\mathcal{F}}_d(t_d))_{t_d \leqslant T_d}$. Letting $n \to +\infty$, it can then checked as in [42, Lemma 2.5] that

$$\lim_{N \to +\infty} \overline{F}_N(t,\mu) = \int S_t \chi_\alpha \mathrm{d}\mu - \varphi(\alpha).$$

Therefore, we get

$$\lim_{N \to +\infty} \overline{F}_N(t,\mu) \geqslant \inf_{\alpha \in \mathbf{Mart}(\overline{\mathscr{P}})} \left\{ \int S_t \chi_\alpha \mathrm{d}\mu - \varphi(\alpha) \right\}.$$

Since the converse inequality holds thanks to Proposition 8.3, we obtain the desired result. □

We are now ready to prove Theorem 1.5. Note that as a consequence of the proof below, in addition to the announced result we will observe that a condition similar to condition (1) in [42, Theorem 2] (i.e. that the support of

the Parisi measure is contained in the set of maximizers of $\int_t^1 (s - \mathbf{E}[\overline{\alpha}_s^2])\mathrm{d}s)$
is satisfied by the optimal martingale $\overline{\alpha}$, see Remark 8.6 below.

Proof of Theorem 1.5. We let

$$\Gamma(\nu, \alpha) = \int \chi_\alpha \mathrm{d}\nu - \varphi(\alpha) - \mathcal{T}_t(\mu, \nu).$$

Let $\overline{\mathscr{P}}$ be a probability space such that Proposition 8.5 holds. This means
that we have

$$\lim_{N \to +\infty} \overline{F}_N(t, \mu) = \inf_{\alpha \in \mathbf{Mart}(\overline{\mathscr{P}})} \sup_{\nu \in \mathcal{P}(\mathcal{K}_T)} \Gamma(\nu, \alpha),$$

and that the infimum in the previous display is reached at some $\overline{\alpha}$. Now let $\overline{\nu}$
denote the unique maximizer in (1.14). As discussed at the beginning of the
proof of Proposition 8.3, thanks to [9, Proposition 6.1], we have $\overline{\nu} \in \mathcal{P}(\mathcal{K}_T)$,
hence

$$\lim_{N \to +\infty} \overline{F}_N(t, \mu) = \sup_{\nu \in \mathcal{P}(\mathcal{K}_T)} \inf_{\alpha \in \mathbf{Mart}(\overline{\mathscr{P}})} \Gamma(\nu, \alpha).$$

In particular, for every $\nu \in \mathcal{P}(\mathcal{K}_T)$ and $\alpha \in \mathbf{Mart}(\overline{\mathscr{P}})$, we have

$$\Gamma(\overline{\nu}, \alpha) \geqslant \lim_{N \to +\infty} \overline{F}_N(t, \mu) \geqslant \Gamma(\nu, \overline{\alpha}).$$

Choosing $\alpha = \overline{\alpha}$ and $\nu = \overline{\nu}$, we obtain $\lim_{N \to +\infty} \overline{F}_N(t, \mu) = \Gamma(\overline{\nu}, \overline{\alpha})$. In
particular the previous display reads

$$\Gamma(\overline{\nu}, \alpha) \geqslant \Gamma(\overline{\nu}, \overline{\alpha}) \geqslant \Gamma(\nu, \overline{\alpha}).$$

In particular $\overline{\alpha}$ is a minimizer of $\Gamma(\overline{\nu}, \cdot)$, this means that $\overline{\alpha}$ is a minimizer
in the variational formula (8.4) with $\nu = \overline{\nu}$. According to the optimality
condition in Proposition 8.2, we have that $\overline{\alpha}$ is uniquely characterized and
satisfies $\overline{\alpha}_d(t) = \partial_x \Phi_{\overline{\nu}_d}(t, X_d(t))$, where X_d solves (8.3). In particular, $\overline{\alpha}_d$
is measurable with respect to the filtration generated by B_d. This means
that $\overline{\alpha}$ lies in the image of $\mathbf{Mart}(\mathscr{W})$ in $\mathbf{Mart}(\overline{\mathscr{P}})$ through the canonical
injection (recall that $\mathscr{W}$ denotes the Wiener space and that the canonical
injection is obtained by replacing the canonical Brownian motion W by B).
From this, we obtain that for any probability space $\mathscr{P}$, there is a copy of $\overline{\alpha}$
in $\mathbf{Mart}(\mathscr{P})$ that we still denote $\overline{\alpha}$ for convenience. Thus we have

$$\lim_{N \to +\infty} \overline{F}_N(t, \mu) = \int S_t \chi_{\overline{\alpha}} \mathrm{d}\mu - \varphi(\alpha) \geqslant \inf_{\alpha \in \mathbf{Mart}(\mathscr{P})} \left\{ \int S_t \chi_\alpha \mathrm{d}\mu - \varphi(\alpha) \right\}.$$

The other bound follows immediately from Proposition 8.3 recalling that
$K_0 \subseteq K$. □

Remark 8.6. Also observe that from $\Gamma(\overline{\nu}, \overline{\alpha}) \geqslant \Gamma(\nu, \overline{\alpha})$, we have that

$$\int \chi_{\overline{\alpha}} \mathrm{d}\overline{\nu} - \mathcal{T}_t(\mu, \overline{\nu}) \geqslant \int \chi_{\overline{\alpha}} \mathrm{d}\nu - \mathcal{T}_t(\mu, \nu).$$

Taking the supremum over $\nu \in \mathcal{P}(\mathcal{K}_T)$ on the right-hand side, using Lemma 6.4 and rearranging, we obtain that

$$\int \chi_{\overline{\alpha}} \mathrm{d}\overline{\nu} - \int S_t \chi_{\overline{\alpha}} \mathrm{d}\mu \geqslant \mathcal{T}_t(\mu, \overline{\nu}).$$

This means that $\chi_{\overline{\alpha}}$ is a Kantorovich potential from μ to $\overline{\nu}$ (see Definition 6.1). This condition replaces condition (1) in [42, Theorem 2] (i.e. that the support of the Parisi measure is contained in the set of maximizers of $\int_t^1 (s - \mathbf{E}[\overline{\alpha}_s^2]) \mathrm{d}s$). $\qquad\square$

APPENDIX A. PROOF OF THE 2-DIMENSIONAL INEQUALITY FOR ξ^*

In this section, we give a proof of Proposition 4.2.

Proof of Proposition 4.2. Let us first show that without loss of generality we may assume that ξ is strongly convex on $\mathbb{R}_+^2$. Assume that Proposition 4.2 holds when ξ is further assumed to be strongly convex. Applying Proposition 4.2 to the function $\xi_\lambda = \xi + \frac{\lambda}{2} |\cdot|^2$, we obtain for all real numbers $a \leqslant a'$ and $b \leqslant b'$,

$$\xi_\lambda^*(a, b) + \xi_\lambda^*(a', b') \leqslant \xi_\lambda^*(a', b) + \xi_\lambda^*(a, b').$$

Thus, to show that (4.3) holds for ξ, it is enough to show that for every $y \in \mathbb{R}_+^2$, $\xi_\lambda^*(y) \to \xi^*(y)$ as $\lambda \to 0$. We fix $y \in \mathbb{R}_+^2$, the sequence $(\xi_\lambda^*(y))_\lambda$ increases as λ decreases to 0 and is upper-bounded by $\xi^*(y)$, we denote by $\ell \in [0, \xi^*(y)]$ its limit. In addition, for every $x \in \mathbb{R}_+^2$, we have $\xi_\lambda^*(y) \geqslant x \cdot y - \xi(x) - \frac{\lambda}{2} |x|^2$ letting $\lambda \to 0$ in this inequality yields

$$\ell = \lim_{\lambda \to 0} \xi_\lambda^*(y) \geqslant x \cdot y - \xi(x).$$

Taking the supremum over $x \in \mathbb{R}_+^2$ in the last display, we obtain $\xi^*(y) \leqslant \ell$. Thus $\lim_{\lambda \to 0} \xi_\lambda^*(y) = \xi^*(y)$, as desired.

In view of the previous argument, we will assume that ξ is λ-strongly convex for some $\lambda > 0$ for the rest of this proof.

Step 1. We show that for every $y \in \mathbb{R}_+^2$, the supremum in the definition of $\xi^*(y)$ is reached at a unique point $x_{\mathrm{opt}}(y) \in \mathbb{R}_+^2$ and that we have $|x_{\mathrm{opt}}(y)| \leqslant |y|/\lambda$.

We fix $y \in \mathbb{R}_+^2$. Since ξ is strongly convex, the map $x \mapsto x \cdot y - \xi(x)$ is strongly concave on $\mathbb{R}_+^2$; so it has at most one maximizer. In addition, we have for every $x \in \mathbb{R}_+^2$ that $\xi(x) \geqslant \frac{\lambda}{2} |x|^2$. Hence for $|x| > 2|y|/\lambda$, we have

$$x \cdot y - \xi(x) \leqslant |x||y| - \frac{\lambda}{2} |x|^2 \leqslant |x|(|y| - \frac{\lambda}{2} |x|) < 0 \leqslant \xi^*(y).$$

It follows that

$$\xi^*(y) = \sup_{\substack{x \in \mathbb{R}_+^2 \\ |x| \leqslant \frac{y}{\lambda}}} \{x \cdot y - \xi(x)\}.$$

The optimization problem in the last display has an optimizer $x_{\mathrm{opt}}(y) \in \mathbb{R}_+^2$ satisfying $|x_{\mathrm{opt}}(y)| \leqslant |y|/\lambda$.

Step 2. We show that $x_{\mathrm{opt}}(y) = 0$ if and only if $y = 0$.

If $y = 0$, the condition $|x_{\mathrm{opt}}(y)| \leqslant |y|/\lambda$ imposes $x_{\mathrm{opt}}(y) = 0$. Conversely, if we have $x_{\mathrm{opt}}(y) = 0$, then $\xi^*(y) = 0$. We will show that $\xi^*(y) > 0$ for $y \neq 0$. Let $r > 0$, the function ξ is $\mathcal{C}^{1,1}$ on the ball 0 of center and radius r in $\mathbb{R}^2_+$. In particular, since we have assumed that $\nabla\xi(0) = 0$, there exists $\theta = \theta(r) > 0$ such that for every $|x| < r$

$$\xi(x) \leqslant \xi(0) + \nabla\xi(0) \cdot x + \frac{\theta}{2}|x|^2 = \frac{\theta}{2}|x|^2.$$

For $|y| < \lambda r$, chosing $x = \varepsilon y$ for $\varepsilon > 0$ small enough, we have $|x| < r$ and $|x| \leqslant |y|/\lambda$, this yields

$$\xi^*(y) \geqslant x \cdot y - \xi(x) \geqslant \varepsilon|y|^2 - \frac{\theta\varepsilon^2}{2}|y|^2 = \varepsilon|y|^2\left(1 - \frac{\theta\varepsilon}{2}\right) > 0.$$

Step 3. We show that ξ^* is $\mathcal{C}^{1,1}$ on $\mathbb{R}^2$. In particular, ξ^* is differentiable on $\mathbb{R}^2$ and its gradient is a Lipschitz function which is itself differentiable almost everywhere.

We know that ξ^* is convex, so it is enough to show that the function $y \mapsto \xi^*(y) - \frac{1}{2\lambda}|y|^2$ is concave on $\mathbb{R}^2$. We have

$$\xi^*(y) - \frac{1}{2\lambda}|y|^2 = \sup_{x \in \mathbb{R}^2_+}\left\{x \cdot y - \frac{1}{2\lambda}|y|^2 - \xi(x)\right\}$$

$$= \sup_{x \in \mathbb{R}^2_+}\left\{-\frac{1}{2}\left|\frac{y}{\sqrt{\lambda}} - \sqrt{\lambda}x\right|^2 - \left(\xi(x) - \frac{\lambda}{2}|x|^2\right)\right\}.$$

Thus, $y \mapsto \xi^*(y) - \frac{1}{2\lambda}|y|^2$ is the supremum of a jointly concave functional on $\mathbb{R}^2 \times \mathbb{R}^2_+$ so it is concave on $\mathbb{R}^2$.

Step 4. We show that for $y \in \mathbb{R}^2_{++}$, $x_{\mathrm{opt}}(y) = \nabla\xi^*(y)$.

We fix $y \in \mathbb{R}^2_{++}$, we have for every $y' \in \mathbb{R}^2_{++}$ and $x \in \mathbb{R}^2_{++}$,

$$\xi^*(y') \geqslant x_{\mathrm{opt}}(y) \cdot y' - \xi(x_{\mathrm{opt}}(y))$$

$$= x_{\mathrm{opt}}(y) \cdot y - \xi(x_{\mathrm{opt}}(y)) + x_{\mathrm{opt}}(y) \cdot (y' - y)$$

$$= \xi^*(y) + x_{\mathrm{opt}}(y) \cdot (y' - y).$$

This means that for every $v \in \mathbb{R}^2$ such that $|v| = 1$, we have for $\varepsilon > 0$ small enough $y + \varepsilon v \in \mathbb{R}^2_{++}$ and

$$\frac{\xi^*(y + \varepsilon v) - \xi^*(y)}{\varepsilon} \geqslant x_{opt}(y) \cdot v.$$

Letting $\varepsilon \to 0$, we obtain

$$\nabla\xi^*(y) \cdot v \geqslant x_{\mathrm{opt}}(y) \cdot v.$$

Since the last display is valid for all v in the sphere of radius 1 in $\mathbb{R}^2$, it imposes that $\nabla\xi^*(y) = x_{\mathrm{opt}}(y)$.

Step 5. We show that $\partial_a\partial_b\xi^* \leqslant 0$ almost everywhere on $\mathbb{R}^2_+$.

Step 5.1. Letting $V = \nabla \xi(\mathbb{R}_{++}^2)$, we show that $\mathbb{R}_{++}^2 \smallsetminus \overline{V} = U_a \cup U_b$ where $U_a = \{y \in \mathbb{R}_{++}^2 \smallsetminus \overline{V} \mid x_{\mathrm{opt},a} = 0\}$ and $U_b = \{y \in \mathbb{R}_{++}^2 \smallsetminus \overline{V} \mid x_{\mathrm{opt},b} = 0\}$ are disjoint open sets.

Clearly $U_a \cup U_b \subseteq \mathbb{R}_{++}^2 \smallsetminus \overline{V}$. Conversely, let $y \in \mathbb{R}_{++}^2 \smallsetminus \overline{V}$, and by contradiction assume that $x_{\mathrm{opt}}(y) \in \mathbb{R}_{++}^2$. Perturbing slightly around $x_{\mathrm{opt}}(y)$, we get the first order condition and $\nabla \xi(x_{\mathrm{opt}}(y)) = y$, this is a contradiction. Hence, we have $x_{\mathrm{opt},a}(y) = 0$ or $x_{\mathrm{opt},b}(y) = 0$. This proves that $\mathbb{R}_{++}^2 \smallsetminus \overline{V} = U_a \cup U_b$. By Step 2, we must have $U_a \cap U_b = \varnothing$. Finally, let us show that U_a is open. Given $y \in U_a$, we have $x_{\mathrm{opt},a}(y) = 0$ and $x_{\mathrm{opt},b}(y) > 0$. By Step 4, $x_{\mathrm{opt},b}(y)$ is a continuous function of y so for every $y' \in B_r(y) \subseteq \mathbb{R}_{++}^2 \smallsetminus \overline{V}$ with $r > 0$ small we have $x_{\mathrm{opt},b}(y') > 0$. This imposes $x_{\mathrm{opt},a}(y') = 0$ as otherwise $y' \in V$ and thus $B_r(y) \subseteq U_a$. This proves that U_a is an open set and by symmetry we also have that U_b is an open set.

Step 5.2. We show that $\partial_a \partial_b \xi^* \leqslant 0$ on $V = \nabla \xi(\mathbb{R}_{++}^2)$.

According to Step 3, ξ^* is differentiable with Lipschitz gradient on $\mathbb{R}^2$. According to (5.7) in Lemma 5.4, we have that for every $x \in \mathbb{R}_{++}^2$,

$$\xi^*(\nabla \xi(x)) = x \cdot \nabla \xi(x) - \xi(x).$$

Differentiating the expression in the previous display, we obtain $\nabla \xi^*(\nabla \xi(x)) = x$. Now let $y \in V$, there exists $x \in \mathbb{R}_{++}^2$ such that $y = \nabla \xi(x)$, and we have

$$\nabla \xi(\nabla \xi^*(y)) = \nabla \xi(\nabla \xi^*(\nabla \xi(x))) = \nabla \xi(x) = y.$$

Finally, since ξ is strongly convex on $\mathbb{R}_+^2$, the function $\nabla \xi$ is continuous and injective on $\mathbb{R}_+^2$, so by invariance of domain, V is an open set. Differentiating the last display with respect to $y \in V$, we obtain that for almost all $y \in V$,

$$\nabla^2 \xi^*(y) = [\nabla^2 \xi(\nabla \xi^*(y))]^{-1}.$$

The 2×2 matrix $A = \nabla^2 \xi(\nabla \xi^*(y))$ is symmetric positive definite and has non-negative coefficients, this is because the condition (1.3) imposes some positivity constraint on the coefficients of the power series expansion of ξ as explained in Remark 5.11. Since $(A^{-1})_{ab} = -\det(A)^{-1} A_{ab}$, we obtain that $(A^{-1})_{ab} \leqslant 0$, as desired.

Step 5.3. We show that $\partial_a \partial_b \xi^* \leqslant 0$ almost everywhere on $U_a \cup U_b$.

Let $y \in U_a$, we have $x_{\mathrm{opt},a}(y) = 0$ and $x_{\mathrm{opt},b}(y) > 0$, so

$$\xi^*(y) = \sup_{x_b > 0} \{x_b y_b - \xi(0, x_b)\}.$$

The value of the unique optimizer in the previous display is $x_{\mathrm{opt},b}(y)$. In particular, $x_{\mathrm{opt},b}(y)$ is independent of the value of y_a and by Step 4, we have $\partial_b \xi^*(y) = x_{\mathrm{opt},b}(y)$. Thus, for almost all $y \in U_a$, $\partial_a \partial_b \xi^*(y) = 0$. By symmetry, this also holds for every $y \in U_b$.

Step 5.4. We show that the boundary of V in $\mathbb{R}_{++}^2$ is a Lebesgue negligible set.

Let us show that the boundary of V is $\nabla\xi(\mathbb{R}^2_+ \smallsetminus \mathbb{R}^2_{++})$. Since $\mathbb{R}^2_+ \smallsetminus \mathbb{R}^2_{++}$ is Lebesgue negligible and $\nabla\xi$ is a smooth function, the desired result will follow. Let $\overline{V}$ denote the closure of V in $\mathbb{R}^2_{++}$ and let $y \in \overline{V}$. There exists a sequence $(x_n)_n$ in $\mathbb{R}^2_{++}$ such that $\nabla\xi(x_n) \to y$. The sequence $(\nabla\xi(x_n))_n$ is bounded and $\nabla\xi(x_n) - \lambda x_n \in \mathbb{R}^2_+$ so the sequence $(x_n)_n$ is bounded. Let $x \in \mathbb{R}^2_+$ denote a subsequential limit of $(x_n)_n$, up to extraction we have $y = \lim_n \nabla\xi(x_n) = \nabla\xi(x)$. Hence $\overline{V} \subseteq \nabla\xi(\mathbb{R}^2_+)$. Conversely given $y \in \nabla\xi(\mathbb{R}^2_+)$, for some $x \in \mathbb{R}^2_+$, we have $y = \nabla\xi(x) = \lim_n \nabla\xi(x_a + \frac{1}{n}, x_b + \frac{1}{n})$, so $y \in \overline{V}$. Thus $\overline{V} = \nabla\xi(\mathbb{R}^2_+)$ and $\overline{V} \smallsetminus V = \nabla\xi(\mathbb{R}^2_+) \smallsetminus \nabla\xi(\mathbb{R}^2_{++})$. Since ξ is assumed to be strongly convex on $\mathbb{R}^2_+$, $\nabla\xi$ is injective on $\mathbb{R}^2_+$ and we have $\nabla\xi(\mathbb{R}^2_+) \smallsetminus \nabla\xi(\mathbb{R}^2_{++}) = \nabla\xi(\mathbb{R}^2_+ \smallsetminus \mathbb{R}^2_{++})$. Thus $\overline{V} \smallsetminus V = \nabla\xi(\mathbb{R}^2_+ \smallsetminus \mathbb{R}^2_{++})$ as announced.

Step 6. We show that (4.3) holds when ξ is λ-strongly convex.

We let $a \leqslant a'$, $b \leqslant b'$, such that $a \neq a'$ and $b \neq b'$. This way $a' = a + h$ and $b' = b + k$ for $h, k \in \mathbb{R}^2_{++}$. We then have

$$\left(\xi^*(a,b) - \xi^*(a',b)\right) - \left(\xi^*(a,b') - \xi^*(a',b')\right)$$
$$= \int_0^1 \int_0^1 hk\partial_a\partial_b\xi^*(a+th, b+sk)\mathrm{d}s\mathrm{d}t.$$

According to Step 5, the right-hand side in the previous display is the integral of an almost everywhere nonnegative function, so it is nonnegative and

$$\left(\xi^*(a,b) - \xi^*(a',b)\right) - \left(\xi^*(a,b') - \xi^*(a',b')\right) \geqslant 0.$$

Rearranging the terms in the previous display, we obtain (4.3) when $a \neq a'$ and $b \neq b'$. Finally, by continuity of ξ^*, the inequality (4.3) holds even when $a = a'$ or $b = b'$. $\qquad\square$

REFERENCES

[1] Antonio Auffinger and Wei-Kuo Chen. Free energy and complexity of spherical bipartite models. *J. Stat. Phys.*, 157(1):40–59, 2014.
[2] Antonio Auffinger and Wei-Kuo Chen. The Parisi formula has a unique minimizer. *Comm. Math. Phys.*, 335(3):1429–1444, 2015.
[3] Jinho Baik and Ji Oon Lee. Free energy of bipartite spherical Sherrington-Kirkpatrick model. *Ann. Inst. Henri Poincaré Probab. Stat.*, 56(4):2897–2934, 2020.
[4] Adriano Barra, Pierluigi Contucci, Emanuele Mingione, and Daniele Tantari. Multi-species mean field spin glasses. Rigorous results. *Ann. Henri Poincaré*, 16(3):691–708, 2015.
[5] Erik Bates and Youngtak Sohn. Free energy in multi-species mixed p-spin spherical models. *Electron. J. Probab.*, 27:Paper No. 52, 75, 2022.
[6] Erik Bates and Youngtak Sohn. Parisi formula for balanced Potts spin glass. *Comm. Math. Phys.*, 405(10):Paper No. 228, 68, 2024.
[7] Erik Bates and Youngtak Sohn. Balanced multi-species spin glasses. *Preprint, arXiv:2507.06522*, 2025.
[8] Matania Ben-Artzi. Lectures on viscous Hamilton-Jacobi equations. Lecture notes, Pacific Institute for the Mathematical Sciences, 2007.
[9] Hong-Bin Chen. On free energy of non-convex multi-species spin glasses. *Preprint, arXiv:2411.13342*, 2024.
[10] Hong-Bin Chen. On Parisi measures of Potts spin glasses with correction. *Electron. Commun. Probab.*, 29:Paper No. 35, 13, 2024.

[11] Hong-Bin Chen and Victor Issa. Uniqueness of Parisi measures for enriched convex vector spin glass. *Preprint, arXiv:2504.15818*, 2024.

[12] Hong-Bin Chen and Jean-Christophe Mourrat. Simultaneous replica-symmetry breaking for vector spin glasses. *Preprint, arXiv:2411.14105*, 2024.

[13] Hong-Bin Chen and Jean-Christophe Mourrat. On the free energy of vector spin glasses with nonconvex interactions. *Probab. Math. Phys.*, 6(1):1–80, 2025.

[14] Hong-Bin Chen and Jiaming Xia. Fenchel–Moreau identities on self-dual cones. *Ann. Fac. Sci. Toulouse*, 33(2):287–309, 2024.

[15] Hong-Bin Chen and Jiaming Xia. Hamilton–Jacobi equations with monotone nonlinearities on convex cones. *Ann. Fac. Sci. Toulouse*, 34(1):1–45, 2025.

[16] Hong-Bin Chen and Jiaming Xia. Hamilton-Jacobi equations from mean-field spin glasses. *Probab. Theory Related Fields*, 192(3):803–873, 2025.

[17] Wei-Kuo Chen. The Aizenman-Sims-Starr scheme and Parisi formula for mixed p-spin spherical models. *Electron. J. Probab.*, 18:no. 94, 14, 2013.

[18] Michael G. Crandall, Hitoshi Ishii, and Pierre-Louis Lions. User's guide to viscosity solutions of second order partial differential equations. *Bull. Amer. Math. Soc. (N.S.)*, 27(1):1–67, 1992.

[19] Stephane Dartois and Benjamin McKenna. Injective norm of real and complex random tensors I: From spin glasses to geometric entanglement. *Preprint, arXiv:2404.03627*, 2024.

[20] Tomas Dominguez and Jean-Christophe Mourrat. *Statistical mechanics of mean-field disordered systems: a Hamilton-Jacobi approach*. Zurich Lectures in Advanced Mathematics. EMS Press, 2024.

[21] Ahmed El Alaoui, Andrea Montanari, and Mark Sellke. Optimization of mean-field spin glasses. *Ann. Probab.*, 49(6):2922–2960, 2021.

[22] Lawrence C Evans. On solving certain nonlinear partial differential equations by accretive operator methods. *Israel J. Math.*, 36(3):225–247, 1980.

[23] Avner Friedman and Eberhard Hopf. The Cauchy problem for first order partial differential equations. *Indiana Univ. Math. J.*, 23(1):27–40, 1973.

[24] David Gamarnik. The overlap gap property: a topological barrier to optimizing over random structures. *Proc. Natl. Acad. Sci. USA*, 118(41):e2108492118, 2021.

[25] David Gamarnik and Aukosh Jagannath. The overlap gap property and approximate message passing algorithms for p-spin models. *Ann. Probab.*, 49(1):180–205, 2021.

[26] David Gamarnik, Cristopher Moore, and Lenka Zdeborová. Disordered systems insights on computational hardness. *J. Stat. Mech. Theory Exp.*, 2022(11):114015, 2022.

[27] Francesco Guerra. Broken replica symmetry bounds in the mean field spin glass model. *Comm. Math. Phys.*, 233(1):1–12, 2003.

[28] Brice Huang and Mark Sellke. Algorithmic threshold for multi-species spherical spin glasses. *Preprint, arXiv:2303.12172*, 2023.

[29] Brice Huang and Mark Sellke. Tight Lipschitz hardness for optimizing mean field spin glasses. *Comm. Pure Appl. Math.*, 78(1):60–119, 2025.

[30] Victor Issa. Existence and uniqueness of permutation-invariant optimizers for Parisi formula. *Preprint, arXiv:2407.13846*, 2024.

[31] Victor Issa. A Hopf-like formula for mean-field spin glass models. *Preprint, arXiv:2410.08754*, 2024.

[32] Aukosh Jagannath and Ian Tobasco. A dynamic programming approach to the Parisi functional. *Proc. Amer. Math. Soc.*, 144(7):3135–3150, 2016.

[33] Aukosh Jagannath and Ian Tobasco. Low temperature asymptotics of spherical mean field spin glasses. *Comm. Math. Phys.*, 352(3):979–1017, 2017.

[34] Aukosh Jagannath and Ian Tobasco. Bounds on the complexity of replica symmetry breaking for spherical spin glasses. *Proc. Amer. Math. Soc.*, 146(7):3127–3142, 2018.

[35] David Jekel, Juspreet Singh Sandhu, and Jonathan Shi. Potential Hessian ascent: the Sherrington-Kirkpatrick model. In *Proceedings of the 2025 Annual ACM-SIAM*

Symposium on Discrete Algorithms (SODA), pages 5307–5387. SIAM, Philadelphia, PA, 2025.

[36] Marc Mézard, Giorgio Parisi, and Miguel Virasoro. *Spin glass theory and beyond: an introduction to the replica method and its applications*, volume 9. World Scientific Publishing Company, 1987.

[37] Andrea Montanari. Optimization of the Sherrington–Kirkpatrick Hamiltonian. *SIAM Journal on Computing*, 0(0):FOCS19–1–FOCS19–38, 2021.

[38] Jean-Christophe Mourrat. Nonconvex interactions in mean-field spin glasses. *Probab. Math. Phys.*, 2(2):281–339, 2021.

[39] Jean-Christophe Mourrat. The Parisi formula is a Hamilton-Jacobi equation in Wasserstein space. *Canad. J. Math.*, 74(3):607–629, 2022.

[40] Jean-Christophe Mourrat. Free energy upper bound for mean-field vector spin glasses. *Ann. Inst. Henri Poincaré Probab. Stat.*, 59(3):1143–1182, 2023.

[41] Jean-Christophe Mourrat. Color symmetry breaking in the Potts spin glass. *ALEA Lat. Am. J. Probab. Math. Stat.*, 22(1):651–655, 2025.

[42] Jean-Christophe Mourrat. Un-inverting the Parisi formula. *Ann. Inst. H. Poincaré Probab. Statist.*, to appear.

[43] Jean-Christophe Mourrat and Dmitry Panchenko. Extending the Parisi formula along a Hamilton-Jacobi equation. *Electron. J. Probab.*, 25:Paper No. 23, 17, 2020.

[44] Jean-Christophe Mourrat and Hendrik Weber. Global well-posedness of the dynamic Φ^4 model in the plane. *Ann. Probab.*, 45(4):2398–2476, 2017.

[45] Dmitry Panchenko. The Parisi ultrametricity conjecture. *Ann. of Math. (2)*, 177(1):383–393, 2013.

[46] Dmitry Panchenko. *The Sherrington-Kirkpatrick model*. Springer Monographs in Mathematics. Springer, 2013.

[47] Dmitry Panchenko. The free energy in a multi-species Sherrington-Kirkpatrick model. *Ann. Probab.*, 43(6):3494–3513, 2015.

[48] Dmitry Panchenko. Free energy in the Potts spin glass. *Ann. Probab.*, 46(2):829–864, 2018.

[49] Dmitry Panchenko. Free energy in the mixed p-spin models with vector spins. *Ann. Probab.*, 46(2):865–896, 2018.

[50] Giorgio Parisi. Infinite number of order parameters for spin-glasses. *Phys. Rev. Lett.*, 43(23):1754, 1979.

[51] Giorgio Parisi. The order parameter for spin glasses: a function on the interval 0-1. *J. Phys. A*, 13(3):1101, 1980.

[52] Giorgio Parisi. A sequence of approximated solutions to the S-K model for spin glasses. *J. Phys. A*, 13(4):L115–L121, 1980.

[53] Giorgio Parisi. Order parameter for spin-glasses. *Phys. Rev. Lett.*, 50(24):1946, 1983.

[54] Mark Sellke. Optimizing mean field spin glasses with external field. *Electron. J. Probab.*, 29:Paper No. 4, 47, 2024.

[55] Maurice Sion. On general minimax theorems. *Pac. J. Math.*, 8:171–176, 1958.

[56] Eliran Subag. Following the ground states of full-RSB spherical spin glasses. *Comm. Pure Appl. Math.*, 74(5):1021–1044, 2021.

[57] Eliran Subag. TAP approach for multispecies spherical spin glasses II: The free energy of the pure models. *Ann. Probab.*, 51(3):1004–1024, 2023.

[58] Eliran Subag. TAP approach for multi-species spherical spin glasses I: General theory. *Electron. J. Probab.*, 30:Paper No. 87, 32, 2025.

[59] Michel Talagrand. Free energy of the spherical mean field model. *Probab. Theory Related Fields*, 134(3):339–382, 2006.

[60] Michel Talagrand. The Parisi formula. *Ann. of Math. (2)*, 163(1):221–263, 2006.

[61] Cédric Villani. *Topics in optimal transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2003.