ON HAMILTON–JACOBI EQUATIONS AND PERCEPTRON MODELS

HONG-BIN CHEN AND JEAN-CHRISTOPHE MOURRAT

Abstract. The solution to a finite-dimensional Hamilton–Jacobi equation is classically known to admit a variational representation in two situations: when the nonlinearity is convex or concave, by the Hopf–Lax formula; and when the initial condition is convex or concave, by the Hopf formula. We identify a third situation: the space variable is split into two groups of coordinates, the nonlinearity is the scalar product of the two corresponding partial gradients, and the initial condition is the sum of a convex function of the first group of coordinates and of an arbitrary Lipschitz function of the second group. This structure is precisely that of the Hamilton–Jacobi equations describing the limit free energy of perceptron models. We illustrate this point by identifying the limit free energy of a class of perceptron models that can be rewritten as bipartite spin glasses, with Ising spins in one layer and bounded soft spins in the other. We also discuss how to extend the approach to several other classes of models, including general Ising and spherical perceptrons, possibly with labelled patterns.

1. Introduction

1.1. **A variational formula for Hamilton–Jacobi equations.** Let $D \geqslant 1$ be an integer. We consider Hamilton–Jacobi equations of the form

$$\partial_t u - \mathsf{H}(\nabla u) = 0 \quad \text{on } (0,\infty)\times\mathbb{R}^D, \qquad u(0,\cdot) = \psi \quad \text{on } \mathbb{R}^D, \tag{1.1}$$

where the nonlinearity $\mathsf{H} : \mathbb{R}^D \to \mathbb{R}$ and the initial condition $\psi : \mathbb{R}^D \to \mathbb{R}$ are given, and where the unknown is $u : \mathbb{R}_+ \times \mathbb{R}^D \to \mathbb{R}$, with $\mathbb{R}_+ := [0,\infty)$. Throughout the paper, we write ∇ for the gradient with respect to the space variable, and ∂_t for the derivative with respect to the time variable. Solutions to (1.1) are understood in the viscosity sense; see Definition 2.1 below. When H is locally Lipschitz continuous and ψ is Lipschitz continuous, the equation (1.1) admits a unique Lipschitz continuous viscosity solution; see for instance [17, Corollary 3.7].

There are two classical situations in which this solution admits a variational representation. The first one is when the nonlinearity H is convex. In this case, the Hopf–Lax formula states that, for every $t > 0$ and $x \in \mathbb{R}^D$,

$$u(t,x) = \sup_{y\in\mathbb{R}^D} \inf_{p\in\mathbb{R}^D} \{\psi(y) + p\cdot(x-y) + t\mathsf{H}(p)\} = \sup_{y\in\mathbb{R}^D} \left\{ \psi(y) - t\mathsf{H}^*\left(\frac{y-x}{t}\right)\right\}, \tag{1.2}$$

where $\mathsf{H}^*(y) := \sup_{p\in\mathbb{R}^D}\{p\cdot y - \mathsf{H}(p)\}$ denotes the convex conjugate of H. The case of a concave nonlinearity reduces to this one by considering $-u$ in place of u, and yields a formula with $\inf_y \sup_p$ in place of $\sup_y \inf_p$. The second classical situation is when the initial condition ψ is convex. In this case, the Hopf formula states that, for every $t \in \mathbb{R}_+$ and $x \in \mathbb{R}^D$,

$$u(t,x) = \sup_{p\in\mathbb{R}^D} \inf_{y\in\mathbb{R}^D} \{\psi(y) + p\cdot(x-y) + t\mathsf{H}(p)\} = \sup_{p\in\mathbb{R}^D} \{p\cdot x + t\mathsf{H}(p) - \psi^*(p)\}, \tag{1.3}$$

and a symmetric formula, with $\inf_p \sup_y$ in place of $\sup_p \inf_y$, holds when ψ is concave. We refer to [5, 17, 19, 31] for proofs of these formulas.

The mechanism behind the Hopf formula (1.3) can be described as follows. If the initial condition is affine, say $\psi(x) = p\cdot x - c$ for some $p \in \mathbb{R}^D$ and $c \in \mathbb{R}$, then the solution to (1.1) is simply $u(t,x) = p\cdot x + t\mathsf{H}(p) - c$. A convex function is the supremum of the affine functions lying below it, namely $\psi(x) = \sup_{p\in\mathbb{R}^D}\{p\cdot x - \psi^*(p)\}$, and the Hopf formula asserts that the solution started from ψ is the supremum of the solutions started from these affine functions. Notice also that in order for (1.3) to be the solution to (1.1), the convexity of ψ is necessary, as one can see by evaluating (1.3) at $t = 0$.

In this paper, we present a new circumstance in which the solution to (1.1) admits a variational formula. We take $D = 2d$ for some integer $d \geqslant 1$, and we write $x = (x_1, x_2) \in \mathbb{R}^d \times \mathbb{R}^d$ for the space variable, and $\nabla = (\nabla_{x_1}, \nabla_{x_2})$ for the gradient. We consider the bilinear nonlinearity $\mathsf{H}(p_1, p_2) := p_1 \cdot p_2$, and initial conditions of the separable form $\psi(x_1, x_2) = \psi_1(x_1) + \psi_2(x_2)$. The equation (1.1) then becomes

$$\partial_t u - \nabla_{x_1} u \cdot \nabla_{x_2} u = 0 \quad \text{on } (0,\infty) \times \mathbb{R}^d \times \mathbb{R}^d, \qquad u(0, x_1, x_2) = \psi_1(x_1) + \psi_2(x_2). \tag{1.4}$$

We assume that ψ_1 is convex, but we make no convexity assumption on ψ_2. Since the nonlinearity H is neither convex nor concave, and since the initial condition ψ is in general neither convex nor concave, neither of the two classical formulas recalled above applies to this situation. Nevertheless, we have the following result.

Theorem 1.1 (Variational formula)**.** *Let $\psi_1, \psi_2 : \mathbb{R}^d \to \mathbb{R}$ be Lipschitz continuous, with Lipschitz constants L_1 and L_2 respectively, and assume that ψ_1 is convex. Let $\psi_1^* : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ denote the convex conjugate of ψ_1, that is, $\psi_1^*(p) := \sup_{y \in \mathbb{R}^d} \{p \cdot y - \psi_1(y)\}$. For every $(t, x_1, x_2) \in \mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{R}^d$, let*

$$u(t, x_1, x_2) := \sup_{p \in \mathbb{R}^d} \left\{ p \cdot x_1 + \psi_2(x_2 + tp) - \psi_1^*(p) \right\}. \tag{1.5}$$

The supremum in (1.5) is attained, and it can be restricted to the closed ball of radius L_1 centered at the origin. The function u is the unique Lipschitz continuous viscosity solution to (1.4). Moreover, for every $t, s \in \mathbb{R}_+$ and $x_1, x_2, y_1, y_2 \in \mathbb{R}^d$, we have

$$|u(t, x_1, x_2) - u(s, y_1, y_2)| \leqslant L_1 |x_1 - y_1| + L_2 |x_2 - y_2| + L_1 L_2 |t - s|. \tag{1.6}$$

The heuristic explanation for Theorem 1.1 is similar to that for the Hopf formula, but it only uses convexity in the first group of coordinates. Indeed, suppose first that ψ_1 is affine, say $\psi_1(x_1) = p \cdot x_1 - c$ for some $p \in \mathbb{R}^d$ and $c \in \mathbb{R}$, and let ψ_2 be an arbitrary Lipschitz continuous function. Then the solution to (1.4) is explicit:

$$u(t, x_1, x_2) = p \cdot x_1 + \psi_2(x_2 + tp) - c. \tag{1.7}$$

Indeed, the gradient of this function with respect to x_1 is constant equal to p, so that for this function, the equation in (1.4) reduces to the transport equation $\partial_t u - p \cdot \nabla_{x_2} u = 0$, which is solved by translation. Writing $\psi_1(x_1) = \sup_{p \in \mathbb{R}^d} \{p \cdot x_1 - \psi_1^*(p)\}$, we see that the formula (1.5) asserts that the solution started from $\psi_1(x_1) + \psi_2(x_2)$ is the supremum over $p \in \mathbb{R}^d$ of the solutions in (1.7) with $c = \psi_1^*(p)$. The fact that this supremum is a viscosity subsolution follows directly from this observation. The proof that it is a viscosity supersolution requires a bit more work, and also relies on the convexity of ψ_1. We prove Theorem 1.1 in Section 2.

Remark 1.2. The formula (1.5) can be rewritten as

$$u(t, x_1, x_2) = \sup_{p \in \mathbb{R}^d} \inf_{y \in \mathbb{R}^d} \left\{ \psi_1(y) + \psi_2(x_2 + tp) + p \cdot (x_1 - y) \right\}. \tag{1.8}$$

If ψ_2 is also convex, then the Hopf formula applies to (1.4), and gives

$$u(t, x_1, x_2) = \sup_{p_1, p_2 \in \mathbb{R}^d} \left\{ p_1 \cdot x_1 + p_2 \cdot x_2 + t\, p_1 \cdot p_2 - \psi_1^*(p_1) - \psi_2^*(p_2) \right\}. \tag{1.9}$$

Taking the supremum over p_2 first and using that $\psi_2^{**} = \psi_2$, we recover (1.5). The point of Theorem 1.1 is that (1.5) remains valid without any convexity assumption on ψ_2.

Remark 1.3. The roles of the two groups of coordinates in Theorem 1.1 can of course be exchanged. Moreover, if u solves (1.4), then the function $(t, x_1, x_2) \mapsto -u(t, x_1, -x_2)$ solves the same equation, with initial condition $-\psi_1(x_1) - \psi_2(-x_2)$. The conclusions of Theorem 1.1 can therefore be transposed to the case in which one of the two functions ψ_1 and ψ_2 is concave, and the other one is arbitrary. For instance, when ψ_1 is concave, the solution is given by (1.8) with $\inf_p \sup_y$ in place of $\sup_p \inf_y$.

For the application to perceptron models, we need a version of Theorem 1.1 in which the space $\mathbb{R}^d$ is replaced by a closed convex cone, with no boundary condition imposed, in the spirit of [15], and ultimately by an infinite-dimensional cone of increasing paths. These extensions are the objects of Sections 2 and 4 respectively.

1.2. **Perceptron models and bipartite spin glasses.** While the structural assumptions of Theorem 1.1 may seem rather narrow, they are in fact satisfied by the Hamilton–Jacobi equations that are relevant to the study of the free energy of perceptron models, as we now explain. Let $N, M \geqslant 1$ be integers, let $(g_{ij})_{i\leqslant N, j\leqslant M}$ be independent standard Gaussian random variables, let P_N be a probability measure on $\mathbb{R}^N$, and let $U : \mathbb{R} \to \mathbb{R}$ be a measurable function, which we call the activation function. The free energy of the associated perceptron model is

$$-\frac{1}{N}\mathbb{E}\log\int \exp\left(\sum_{j=1}^{M} U\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N} g_{ij}\sigma_i\right)\right) \mathrm{d}P_N(\sigma), \tag{1.10}$$

where $\mathbb{E}$ denotes the expectation with respect to the Gaussian random variables (g_{ij}). We are interested in the limit of this quantity as N tends to infinity, with M/N converging to some $\lambda \in (0, \infty)$. The most studied cases are the Ising perceptron, in which P_N is the uniform measure on $\{-1, 1\}^N$, and the spherical perceptron, in which P_N is the uniform measure on the sphere of radius $\sqrt{N}$.

In the interpretation of this model that goes back to [21, 22], the vector σ encodes the synaptic weights of a single neuron, and each vector $(g_{ij})_{i\leqslant N}$ is a pattern presented to this neuron, multiplied by the output in $\{-1, 1\}$ that the neuron is required to produce in response; by symmetry, this multiplication does not change the law of the pattern. The neuron responds with the sign of its pre-activation, so the quantity $N^{-1/2}\sum_i g_{ij}\sigma_i$ is positive if and only if the response to the j-th pattern is correct, and U assigns a reward to its value. For instance, the choice $U = -\beta\mathbf{1}_{(-\infty,\kappa)}$ penalizes each pattern that is not stored with margin κ, while the hard constraint $U = \log\mathbf{1}_{[\kappa,\infty)}$, which takes the value $-\infty$, forbids such patterns altogether; the storage capacity is the largest value of λ for which this constraint can typically be satisfied.

Predictions for the free energy based on the replica method were developed in [21, 22, 23, 24, 25, 30]; a cavity derivation of the replica-symmetric formulas was given in [32]. Related predictions in the context of empirical risk minimization were developed in [33]. Rigorous formulas for the Ising perceptron were first obtained in regimes of high temperature or small λ in [49, 50]. For the hard constraint $U = \log\mathbf{1}_{[\kappa,\infty)}$, the replica-symmetric formula at small λ was established in [51, Chapter 9]; a different proof, covering a broader class of activation functions, was subsequently given in [6]. For the spherical perceptron with the same hard constraint and $\kappa \geqslant 0$, the Gardner formula was proved in [45].

Much effort has also been devoted to determining the storage capacity and to understanding the geometry of the set of solutions; see for instance [29, 38, 48, 54] for the Ising perceptron, and [1, 4, 44] for its symmetric variants, such as $U = \log\mathbf{1}_{[-\kappa,\kappa]}$. For the zero-margin Ising perceptron, the capacity predicted in [30] was established in [16, 28, 46].

Variational formulas for the limit free energy, valid for every λ and at every temperature, were recently obtained for the Ising perceptron in [39] and for the spherical perceptron in [3, 40]. In [20], variational formulas are obtained for both Ising and spherical perceptrons with i.i.d. random activation functions, independent of the Gaussian patterns, under a common Lipschitz bound and an integrability assumption at the origin. We also refer to [8] for variational bounds on the free energy of perceptrons with log-concave spin priors and concave activation functions.

In this paper, we relate perceptron models and Hamilton–Jacobi equations by following the approach developed for spin glasses in [9, 11, 12, 13, 17, 34, 35, 36, 37]. That is, we enrich the model in (1.10) by adding to its Hamiltonian Gaussian fields indexed by the leaves of a Poisson–Dirichlet cascade. One of these fields acts on the spins σ_i, and another one is added to the pre-activations $N^{-1/2}\sum_i g_{ij}\sigma_i$ inside U; the covariance structure of each of these fields is encoded by an increasing path. We also introduce a time parameter t, which interpolates between the original pre-activations and independent Gaussian random variables. The enriched free energy is then a function of t and of the pair of paths. By Gaussian integration by parts, this function satisfies, up to an error term which vanishes if the overlaps concentrate conditionally on the overlap of the cascade, a Hamilton–Jacobi equation whose nonlinearity is the L^2 scalar product of the two partial derivatives with respect to the paths. This is the infinite-dimensional analogue of the bilinear nonlinearity in (1.4). At $t = 0$, the two groups of variables decouple,

so the initial condition is the sum of a term that depends only on the path associated with the spins, and of a term that depends only on the path associated with the pre-activations. This is the analogue of the separable initial condition in (1.4). The first term is the cascade transform of the reference measure for one spin, and it does not depend on the activation function U; for Ising spins, this term was shown to be convex in [11, Proposition 2.2] and in [26]; for spherical spins, this term was shown to be convex in [13, Lemma 2.9]. The second term depends on U, and in general it is not convex. This is exactly the structure of Theorem 1.1, and the infinite-dimensional analogue of the formula (1.5) is therefore expected to describe the limit free energy of perceptron models in full generality, the choice of the activation function U affecting only the non-convex part of the initial condition. We refer to Section 7 for a more detailed discussion of general Ising and spherical perceptrons, including the derivation of the candidate Hamilton–Jacobi equation and of the resulting candidate variational formula.

We illustrate this point with full rigor on a subclass of perceptron models which can conveniently be rewritten as bipartite spin glasses. We fix a probability measure ν with compact support, and consider activation functions of the form

$$u_{t,\nu}(x) := \log \int e^{\sqrt{2t}xv - tv^2} \, \mathrm{d}\nu(v), \qquad t \in \mathbb{R}_+. \tag{1.11}$$

The parameter t in (1.11) plays the role of an inverse temperature, and it will also be the time variable of the Hamilton–Jacobi equation. For such activation functions, the free energy in (1.10) can be rewritten as that of a model with an auxiliary layer of ν-distributed spins $v = (v_j)_{j \leqslant M}$, and with a Gaussian Hamiltonian which is bilinear in (σ, v); see (1.14) below. Related bipartite models arise in the study of restricted Boltzmann machines; see for instance [52]. This restriction is meant to simplify the presentation: it allows us to borrow cavity arguments and comparison results already developed for multi-species spin glasses [9, 11, 12, 34, 36], and to focus on the new ingredient, which is the variational formula. For general activation functions, the Hamiltonian in (1.10) is not a Gaussian process, and these arguments need to be redone.

Our rigorous results on perceptron models are less general than those presented in [20, 39, 40]. Our aim with this paper is to present a systematic and clear structure for the study of the free energy of perceptron and related models. While we only identify the limit free energy of a restricted class of models with full rigor, we discuss in Sections 6 and 7 how the approach can plausibly be generalized to the cases covered in [20, 39, 40] and more (the entirety of Section 6 and several results in Section 7 are fully rigorous, and the missing ingredients are clearly identified). To the best of our understanding, the approaches proposed in [20, 39, 40] do not match the one we present here: the definition of the enriched free energy and the relevant Hamilton–Jacobi equation in [39, 40] seem to be different from ours, while [20] does not seem to rely on Hamilton–Jacobi equations. The approach of [3] seems closer to ours.

1.3. **The bipartite model and the main rigorous result.** We now give the precise definition of the model that we study in full rigor. We fix $\lambda \in (0, \infty)$ and let $(M_N)_{N \in \mathbb{N}}$ be a sequence of positive integers satisfying $\lim_{N \to \infty} \frac{M_N}{N} = \lambda$. We fix a probability measure ν supported on $[-1, 1]$; see Remark 5.3 for the extension to an arbitrary compact support. Let $(g_{ij})_{i,j \in \mathbb{N}}$ be a collection of i.i.d. standard Gaussian random variables. For $t \in \mathbb{R}_+$, we consider the Ising perceptron model with Hamiltonian $H_N^{\mathrm{perc}} : \{-1, 1\}^N \to \mathbb{R}$ given by

$$H_N^{\mathrm{perc}}(\sigma) := \sum_{j=1}^{M_N} u_{t,\nu} \left(\frac{1}{\sqrt{N}} \sum_{i=1}^{N} g_{ij} \sigma_i \right),$$

where $u_{t,\nu}$ is defined in (1.11). The associated free energy is

$$\overline{F}_N^{\mathrm{perc}} := -\frac{1}{N} \mathbb{E} \log \sum_{\sigma \in \{-1,1\}^N} 2^{-N} \exp\left(H_N^{\mathrm{perc}}(\sigma)\right). \tag{1.12}$$

As announced, this model can be rewritten as a bipartite model with Ising spins in one layer and ν-distributed spins in the other. We define the bipartite Hamiltonian $H_N : \{-1, 1\}^N \times [-1, 1]^{M_N} \to$

$\mathbb{R}$ by

$$H_N(\sigma, v) := \frac{1}{\sqrt{N}} \sum_{\substack{i\in\{1,\dots,N\}\\ j\in\{1,\dots,M_N\}}} g_{ij}\sigma_i v_j, \tag{1.13}$$

where we have written $\sigma = (\sigma_1, \dots, \sigma_N)$ and $v = (v_1, \dots, v_{M_N})$. Its covariance is given by

$$\mathbb{E}\left[H_N(\sigma, v)H_N(\sigma', v')\right] = \frac{1}{N}(\sigma \cdot \sigma')(v \cdot v'),$$

and thus $\mathbb{E}\left[H_N(\sigma, v)^2\right] = |v|^2$. Using (1.11) and writing $\mu_{\mathcal{I}} := \frac{1}{2}(\delta_{-1} + \delta_1)$ for the Ising spin distribution, we have

$$\overline{F}_N(t, 0) := \overline{F}_N^{\mathrm{perc}} = -\frac{1}{N}\mathbb{E}\log \int \exp\left(\sqrt{2t}H_N(\sigma, v) - t|v|^2\right) \mathrm{d}\left(\mu_{\mathcal{I}}^{\otimes N} \otimes \nu^{\otimes M_N}\right)(\sigma, v). \tag{1.14}$$

Notice that the correction term $-t|v|^2$ is $-\frac{1}{2}$ times the variance of $\sqrt{2t}H_N(\sigma, v)$.

Enriched models. We let $\mathcal{Q}$ denote the set of right-continuous increasing paths $q : [0, 1) \to \mathbb{R}_+$ (here and throughout, we say that a path q is increasing provided that $q(s) \leqslant q(s')$ for every $0 \leqslant s \leqslant s' < 1$). For $r \in [1, \infty]$, we write $\mathcal{Q}_r := \mathcal{Q} \cap L^r[0, 1)$. For $q \in \mathcal{Q}_\infty$, we set $q(1) := \lim_{r\uparrow 1} q(r)$, which exists by monotonicity. We use $|\cdot|_{L^r}$ for the usual norm on $L^r([0, 1))$. For $c \geqslant 0$, we set

$$\mathcal{Q}_{\infty,\leqslant c} := \{q \in \mathcal{Q}_\infty : q(r) \leqslant c,\ \forall r \in [0, 1)\}.$$

Below we will extend $\overline{F}_N(t, \cdot)$ so that its second argument is any element of the set $\mathcal{Q}_2^2 = \mathcal{Q}_2 \times \mathcal{Q}_2$; the 0 appearing as the second argument of $\overline{F}_N(t, 0)$ in (1.14) is the pair of paths that are constant equal to zero.

Let $\mathfrak{R}$ be a continuous Poisson–Dirichlet cascade independent of (g_{ij}) whose overlap is uniformly distributed on $[0, 1]$. This is a random probability measure on the unit sphere of a separable Hilbert space, with almost surely ultrametric support; see [12, Section 4] for its construction. For $\alpha, \alpha' \in \operatorname{supp}\mathfrak{R}$, we write $\alpha \wedge \alpha'$ for their scalar product. Under $\mathbb{E}\mathfrak{R}^{\otimes 2}$, the overlap $\alpha \wedge \alpha'$ is uniformly distributed on $[0, 1]$.

For $q = (q_1, q_2) \in \mathcal{Q}_\infty^2$, conditionally on $\mathfrak{R}$, let W_N^q be a centered Gaussian process indexed by $(\sigma, v, \alpha) \in \{-1, 1\}^N \times [-1, 1]^{M_N} \times \operatorname{supp}\mathfrak{R}$, independent of H_N, with covariance

$$\mathbb{E}\left[W_N^q(\sigma, v, \alpha)W_N^q(\sigma', v', \alpha') \mid \mathfrak{R}\right] = q_1(\alpha \wedge \alpha')\sigma \cdot \sigma' + q_2(\alpha \wedge \alpha')v \cdot v'. \tag{1.15}$$

The field W_N^q is the sum of independent scalar cascade fields, one for each spin coordinate; see [12, Proposition 4.1 and (5.1)] for the construction. We set $P_N := \mu_{\mathcal{I}}^{\otimes N} \otimes \nu^{\otimes M_N}$. For $t \in \mathbb{R}_+$ and $q \in \mathcal{Q}_\infty^2$, we define $H_N^{t,q} : \{-1, 1\}^N \times [-1, 1]^{M_N} \times \operatorname{supp}\mathfrak{R} \to \mathbb{R}$ and $\overline{F}_N : \mathbb{R}_+ \times \mathcal{Q}_\infty^2 \to \mathbb{R}$ by

$$H_N^{t,q}(\sigma, v, \alpha) := \sqrt{2t}H_N(\sigma, v) + \sqrt{2}W_N^q(\sigma, v, \alpha) - Nq_1(1) - (t + q_2(1))|v|^2, \tag{1.16}$$

$$\overline{F}_N(t, q) := -\frac{1}{N}\mathbb{E}\log \iint \exp(H_N^{t,q}(\sigma, v, \alpha))\,\mathrm{d}P_N(\sigma, v)\,\mathrm{d}\mathfrak{R}(\alpha). \tag{1.17}$$

The terms subtracted in (1.16) equal one-half of the variance of the Gaussian part.

For a compactly supported probability measure ρ on $\mathbb{R}$, we define $\psi_\rho : \mathcal{Q}_\infty \to \mathbb{R}$ by

$$\psi_\rho(q) := -\mathbb{E}\log \iint \exp\left(\sqrt{2}w^q(\alpha)\tau - q(1)\tau^2\right) \mathrm{d}\rho(\tau)\,\mathrm{d}\mathfrak{R}(\alpha), \tag{1.18}$$

where, conditionally on $\mathfrak{R}$, the process w^q is centered Gaussian with covariance $\mathbb{E}[w^q(\alpha)w^q(\alpha') \mid \mathfrak{R}] = q(\alpha \wedge \alpha')$, as in (1.15). For piecewise-constant q, the quantity $\psi_\rho(q)$ can be computed by recursive Gaussian integration; see [17, Theorem 5.25]. An equivalent representation through the Parisi PDE is given in [17, Lemma 6.4 and Proposition 6.6]. For the fixed measures ν and $\mu_{\mathcal{I}}$, we can thus define ψ_ν, $\psi_{\mu_{\mathcal{I}}} : \mathcal{Q}_\infty \to \mathbb{R}$, which can be extended to $\mathcal{Q}_1$ by Lemma 3.2. We also define $\psi_{\mu_{\mathcal{I}}}^* : \mathcal{Q}_\infty \to \mathbb{R} \cup \{+\infty\}$ by

$$\psi_{\mu_{\mathcal{I}}}^*(p) := \sup_{q\in\mathcal{Q}_\infty} \left\{\langle p, q\rangle_{L^2} - \psi_{\mu_{\mathcal{I}}}(q)\right\}, \qquad \forall p \in \mathcal{Q}_\infty. \tag{1.19}$$

Here and throughout, $\langle p, q\rangle_{L^2} := \int_0^1 p(s)q(s)\mathrm{d}s$ denotes the pairing of two real-valued paths.

Theorem 1.4 (Free energy of the bipartite model)**.** *For every $t \geqslant 0$ and $q = (q_1, q_2) \in \mathcal{Q}_2^2$, we have*

$$\lim_{N\to+\infty} \overline{F}_N(t,q) = \sup_{p\in\mathcal{Q}_\infty} \left\{ \langle p, q_1 \rangle_{L^2} + \lambda\psi_\nu(q_2 + tp) - \psi^*_{\mu_{\mathcal{I}}}(p) \right\}. \tag{1.20}$$

Moreover, denoting this limit by $f(t,q)$, the function f is the unique Lipschitz viscosity solution, in the sense of Definition 4.1, to

$$\begin{cases} \partial_t f - \int_0^1 \partial_{q_1} f \, \partial_{q_2} f \, \mathrm{d}r = 0, & \text{on } (0,\infty) \times \mathcal{Q}_2^2, \\ f(0,q) = \psi_{\mu_{\mathcal{I}}}(q_1) + \lambda\psi_\nu(q_2), & \text{for } q \in \mathcal{Q}_2^2. \end{cases} \tag{1.21}$$

When f is differentiable at (t,q), its partial derivative $\partial_{q_i} f$ is an element of $L^2([0,1))$, represented by $r \mapsto \partial_{q_i} f(t,q,r)$. The nonlinearity in (1.21) is the L^2 scalar product of the two partial derivatives, namely $\int_0^1 \partial_{q_1} f(t,q,r) \partial_{q_2} f(t,q,r) \mathrm{d}r$, and is thus the infinite-dimensional analogue of the bilinear nonlinearity appearing in (1.4).

We recall from (1.14) that the limit of the free energy $\overline{F}_N^{\mathrm{perc}}$ defined in (1.12) is obtained by setting $q = 0$ in (1.20). As already mentioned, the limit free energy of perceptron models with arbitrary bounded measurable activation functions was already determined in [39].

Remark 1.5 (Comparison with the Hopf formula). The variational formula on the right-hand side of (1.20) can be rewritten as

$$\sup_{p_1\in\mathcal{Q}_\infty} \inf_{q_1'\in\mathcal{Q}_\infty} \left\{ \psi_{\mu_{\mathcal{I}}}(q_1') + \lambda\psi_\nu(q_2 + tp_1) + \langle p_1, q_1 - q_1' \rangle_{L^2} \right\}, \tag{1.22}$$

which is the analogue on the space of paths of the formula (1.8). When ψ_ν is convex, one could adapt the approach in [11, Theorem 1.1] to obtain the Hopf formula for the limit free energy, reading

$$\sup_{p\in\mathcal{Q}_\infty^2} \inf_{q'\in\mathcal{Q}_\infty^2} \left\{ \psi_{\mu_{\mathcal{I}}}(q_1') + \lambda\psi_\nu(q_2') + \langle p, q - q' \rangle_{L^2} + t \langle p_1, p_2 \rangle_{L^2} \right\}, \tag{1.23}$$

where $p = (p_1, p_2)$, $q' = (q_1', q_2')$, and $\langle p, q - q' \rangle_{L^2} = \sum_{i=1}^2 \langle p_i, q_i - q_i' \rangle_{L^2}$. As in Remark 1.2, the two formulas (1.22) and (1.23) formally coincide when ψ_ν is convex: optimizing first over q_2' and then over p_2 in (1.23) produces the term $\lambda\psi_\nu(q_2 + tp_1)$. However, for the formula (1.23) to be valid, the convexity of both $\psi_{\mu_{\mathcal{I}}}$ and ψ_ν is required, as can be seen by considering the case $t = 0$. In the present setting, the function ψ_ν need not be convex. For instance, convexity fails when ν is a probability measure on $\{-1, 1\}$ whose mean has absolute value in $(1/\sqrt{3}, 1)$; see [17, Exercise 6.7 and solution]. Theorem 1.4 shows that (1.22) gives the limit free energy and the viscosity solution of (1.21) regardless of whether ψ_ν is convex.

1.4. **Outline of the proof and organization of the paper.** To identify the limiting free energy of the bipartite model, we proceed in the following way. On the one hand, we use [34] to obtain that, for $t > 0$ and $q \in \mathcal{Q}_2^2$,

$$f(t,q) \leqslant \liminf_{N\to\infty} \overline{F}_N(t,q), \tag{1.24}$$

where f is the solution to (1.21). For the converse bound, we follow the approach developed in [11], which builds upon the cavity computations of [9, 12], and we show that

$$\limsup_{N\to\infty} \overline{F}_N(t,q) \leqslant \psi_{\mu_{\mathcal{I}}}(q_1 + t\overline{p}_2) + \lambda\psi_\nu(q_2 + t\overline{p}_1) - t \langle \overline{p}_1, \overline{p}_2 \rangle_{L^2}, \tag{1.25}$$

for some $\overline{p} \in \mathcal{Q}_\infty^2$ satisfying

$$\overline{p}_1 = \partial_q \psi_{\mu_{\mathcal{I}}}(q_1 + t\overline{p}_2) \qquad \text{and} \qquad \overline{p}_2 = \lambda\partial_q \psi_\nu(q_2 + t\overline{p}_1). \tag{1.26}$$

In order to relate the two bounds thus obtained, we show in Section 4 that the solution to (1.21) admits the variational representation

$$f(t,q) = \sup_{p\in\mathcal{Q}_\infty} \left\{ \langle p, q_1 \rangle_{L^2} + \lambda\psi_\nu(q_2 + tp) - \psi^*_{\mu_{\mathcal{I}}}(p) \right\}. \tag{1.27}$$

Using the first relation in (1.26) and the convexity of $\psi_{\mu_{\mathcal{I}}}$, we also observe that

$$\psi^*_{\mu_{\mathcal{I}}}(\overline{p}_1) = \langle \overline{p}_1, q_1 + t\overline{p}_2 \rangle_{L^2} - \psi_{\mu_{\mathcal{I}}}(q_1 + t\overline{p}_2), \tag{1.28}$$

and therefore

$$\text{r.h.s. of (1.25)} \overset{(1.28)}{=} \langle \overline{p}_1, q_1 \rangle_{L^2} + \lambda \psi_\nu(q_2 + t\overline{p}_1) - \psi^*_{\mu_{\mathcal{I}}}(\overline{p}_1) \overset{(1.27)}{\leqslant} f(t,q).$$

This and (1.24)–(1.25) imply the desired identity (1.20).

For multi-species models in which every species consists of centered Ising spins, a bound of the form (1.25) is proved in [11, Proposition 8.1]. The model we study here is not literally covered by this statement, since the spins of the second layer are sampled from ν. However, as we explain in Section 5, the argument given there applies with very little modification. Alternatively, one could first show that the free energy converges as N tends to infinity, by adapting the results of [10, 27] to our multi-species context, and then appeal to the critical-point representation of the limit obtained in [9, 12].

The rest of the paper is organized as follows. In Section 2, we prove Theorem 1.1, and we extend it to equations posed on finite-dimensional cones. Section 3 collects regularity properties of the enriched free energy and of the cascade transforms $\psi_{\mu_{\mathcal{I}}}$ and ψ_ν. In Section 4, we pass to the limit along dyadic discretizations, and obtain the variational formula on the space of paths. Section 5 completes the proof of Theorem 1.4, following the outline above. In Section 6, we show that the variational formula remains valid for equations posed on a domain of paths that shrinks as time increases. This is used in Section 7, where we discuss the extension of this approach to general Ising and spherical perceptrons, and to even more general models.

Statement of AI use. The ideas presented in this paper, except for the contents of Section 6, were fully developed in early September 2026, prior to the release of [3, 39, 40]. Our starting point was the conjecture in [24, 25, 33] for the spherical perceptron, which initially surprised us because we could not see how to reduce it to the classical cases under which we expected a variational formula to be valid. We observed that we could combine the arguments in [11] and in [13] to obtain a variational formula for the large-N limit of

$$\frac{1}{N}\mathbb{E}\log\int \exp\left(\beta H_N(\sigma,v)\right)\,\mathrm{d}\mu_{\mathcal{I}}^{\otimes N}(\sigma)\,\mathrm{d}\mu_{\mathbb{S}_{M_N}}(v),$$

where $H_N(\sigma,v)$ is as in (1.13), and $\mu_{\mathbb{S}_{M_N}}$ is the uniform measure on $\mathbb{S}_{M_N} := \{v \in \mathbb{R}^{M_N} \,:\, |v|^2 = M_N\}$. Dividing this limit variational formula by β, and formally sending β to infinity, we obtained a candidate formula for the large-N limit of

$$N^{-1}\mathbb{E}\max_{\sigma\in\{-1,1\}^N, v\in\mathbb{S}_{M_N}} H_N(\sigma,v) = N^{-\frac{3}{2}}\,\mathbb{E}\max_{v\in\mathbb{S}_{M_N}}\sum_{i=1}^N |g_{i\bullet}\cdot v|,$$

with the notation $g_{i\bullet} = (g_{ij})_{j\leqslant M_N} \in \mathbb{R}^{M_N}$. The paper [33] also proposes a variational formula for this limit (if we ignore the fact that the activation function is assumed bounded in [33]). With assistance from ChatGPT 5.6, we could observe that the variational formula obtained as explained above and the one from [33] are indeed equal. An inspection of the argument connecting the two formulas led us to wonder whether Theorem 1.1 could be correct. We then obtained a rigorous proof of Theorem 1.1 as well as the other results presented here, with assistance from ChatGPT 5.6. We have manually verified all statements and proofs, and we have polished the paper with assistance from ChatGPT 6 Astra and Fable 5.1.

Acknowledgments. We would like to thank Jeanne Boursier, Cédric Gerbelot and Victor Issa for stimulating discussions. HBC acknowledges funding from the NYU Shanghai Start-Up Fund and support from the NYU–ECNU Institute of Mathematical Sciences at NYU Shanghai. JCM acknowledges support from the MSCA grant SLOHD (101203974).

2. Variational formula in finite dimensions

In this section, we prove Theorem 1.1, and we then extend it to equations posed on finite-dimensional cones, in preparation for the passage to the space of paths in Section 4.

2.1. **Viscosity solutions.** We use a single notion of viscosity solution in finite and infinite dimensions, following [14, Definitions 1.2 and 1.4]. Let $\mathcal{H}$ be a separable Hilbert space, let $V \subseteq (0,\infty) \times \mathcal{H}$, and let $\mathsf{H} : \mathcal{H} \to \mathbb{R}$ be continuous. A function $\phi : V \to \mathbb{R}$ is called **smooth** if, for every $(t,x) \in V$, there is a unique element of $\mathbb{R} \times \mathcal{H}$, denoted by $(\partial_t \phi(t,x), \nabla_x \phi(t,x))$, such that

$$\phi(s,y) - \phi(t,x) = \partial_t \phi(t,x)(s-t) + \langle \nabla_x \phi(t,x), y - x \rangle_{\mathcal{H}} + O\left(|s-t|^2 + |y-x|_{\mathcal{H}}^2\right)$$

as $(s,y) \in V$ tends to (t,x), and the map $(t,x) \mapsto (\partial_t \phi(t,x), \nabla_x \phi(t,x))$ from V to $\mathbb{R} \times \mathcal{H}$ is continuous.

Definition 2.1 (Viscosity solutions)**.** A continuous function $u : V \to \mathbb{R}$ is a **viscosity subsolution** of $\partial_t u - \mathsf{H}(\nabla_x u) = 0$ if, for every smooth $\phi : V \to \mathbb{R}$ and every local maximum $(t,x) \in V$ of $u - \phi$, we have

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

A **viscosity supersolution** satisfies the reverse inequality at every local minimum of $u - \phi$. A **viscosity solution** is both a viscosity subsolution and a viscosity supersolution.

If V is open, then the differential of a smooth function is its Fréchet differential. If moreover $\mathcal{H}$ is finite-dimensional, then every C^2 function on V is smooth, and every smooth function is C^1. In this case, Definition 2.1 agrees with the classical definition of viscosity solutions, in which the test functions are C^1, since the latter can equivalently be formulated with C^∞ test functions; see for instance [17, Exercise 3.4 and solution]. If $V = (0,\infty) \times C$ for a closed convex cone $C \subseteq \mathcal{H}$ satisfying $\overline{C - C} = \mathcal{H}$, then the differential is unique also at boundary points; see [14, Remark 1.3].

2.2. **The case of the whole space.** Throughout this subsection, we place ourselves in the setting of Theorem 1.1, and we write $g := \psi_1^*$. We first record the properties of g that will be used. The function g is convex and lower semicontinuous, being a supremum of affine functions. It is proper: for every $p \in \mathbb{R}^d$, we have $g(p) \geqslant -\psi_1(0) > -\infty$, and if p_0 is a subgradient of the finite convex function ψ_1 at the origin, then $g(p_0) = -\psi_1(0) < +\infty$. Its effective domain $\operatorname{dom} g := \{p \in \mathbb{R}^d : g(p) < +\infty\}$ is bounded: for every $p \in \mathbb{R}^d$ and $a \geqslant 0$, the Lipschitz continuity of ψ_1 gives

$$g(p) \geqslant a|p|^2 - \psi_1(ap) \geqslant a|p|\big(|p| - L_1\big) - \psi_1(0),$$

and letting a tend to infinity shows that

$$\operatorname{dom} g \subseteq \{p \in \mathbb{R}^d : |p| \leqslant L_1\}. \tag{2.1}$$

In particular, the supremum in (1.5) can be restricted to the closed ball of radius L_1 centered at the origin. Finally, since ψ_1 is convex and continuous, the Fenchel–Moreau theorem gives

$$g^*(x) := \sup_{p \in \mathbb{R}^d} \{p \cdot x - g(p)\} = \psi_1(x), \qquad \forall x \in \mathbb{R}^d; \tag{2.2}$$

see for instance [17, Theorem 2.5]. For $p \in \operatorname{dom} g$ and $z = (t, x_1, x_2) \in \mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{R}^d$, we set

$$w_p(z) := p \cdot x_1 + \psi_2(x_2 + tp) - g(p), \tag{2.3}$$

so that w_p is the explicit solution (1.7) with $c = g(p)$, and the function u in (1.5) is $u = \sup_{p \in \operatorname{dom} g} w_p$. We use Definition 2.1 with ambient Hilbert space $\mathbb{R}^d \times \mathbb{R}^d$, with $V := (0,\infty) \times \mathbb{R}^d \times \mathbb{R}^d$, and with the Hamiltonian $\mathsf{H}(a_1, a_2) := a_1 \cdot a_2$. For a function ϕ of $z = (t, x_1, x_2) \in \mathbb{R} \times \mathbb{R}^d \times \mathbb{R}^d$, we write $\nabla_{t,x} \phi := (\partial_t \phi, \nabla_{x_1} \phi, \nabla_{x_2} \phi)$ for its full gradient, in contrast with $\nabla \phi = (\nabla_{x_1} \phi, \nabla_{x_2} \phi)$, which only involves the space variables.

Proof of Theorem 1.1. Step 1: Attainment, Lipschitz continuity, and the initial condition. We fix $z = (t, x_1, x_2) \in \mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{R}^d$. The function $p \mapsto p \cdot x_1 + \psi_2(x_2 + tp) - g(p)$, extended by $-\infty$ outside $\operatorname{dom} g$, is upper semicontinuous on the closed ball of radius L_1, and it is not identically $-\infty$ there. Hence $u(z)$ is finite, the supremum in (1.5) is attained, and the set of maximizers

$$\mathcal{A}(z) := \{p \in \operatorname{dom} g : w_p(z) = u(z)\}$$

is nonempty and compact. For every $p \in \operatorname{dom} g$, we have $|p| \leqslant L_1$ by (2.1), and thus, for every $(s, y_1, y_2) \in \mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{R}^d$,

$$|w_p(t, x_1, x_2) - w_p(s, y_1, y_2)| \leqslant L_1|x_1 - y_1| + L_2|x_2 - y_2| + L_1L_2|t - s|.$$

Taking the supremum over p yields (1.6). At $t = 0$, the formula (1.5) and the identity (2.2) give $u(0, x_1, x_2) = g^*(x_1) + \psi_2(x_2) = \psi_1(x_1) + \psi_2(x_2)$.

Step 2: The subsolution inequality. Let $z = (t, x_1, x_2) \in V$, and let ϕ be a smooth function such that $u - \phi$ has a local maximum at z. We fix $p \in \mathcal{A}(z)$. Since $w_p \leqslant u$ and $w_p(z) = u(z)$, the function $w_p - \phi$ also has a local maximum at z, that is,

$$\phi(z') - \phi(z) \geqslant w_p(z') - w_p(z) \tag{2.4}$$

for every z' in a neighborhood of z. The function w_p is affine in x_1 with slope p, and it is constant along the direction $(1, 0, -p)$: for every $y \in \mathbb{R}^d$ and every $h \in \mathbb{R}$ sufficiently small,

$$w_p(t, x_1 + hy, x_2) - w_p(z) = hp \cdot y \qquad \text{and} \qquad w_p(t + h, x_1, x_2 - hp) - w_p(z) = 0.$$

We apply (2.4) with $z' = (t, x_1 + hy, x_2)$ and with $z' = (t + h, x_1, x_2 - hp)$, divide by h, and let h tend to 0 from above and from below. This yields $\nabla_{x_1}\phi(z) \cdot y = p \cdot y$ for every $y \in \mathbb{R}^d$, and $\partial_t\phi(z) - p \cdot \nabla_{x_2}\phi(z) = 0$. Hence $\nabla_{x_1}\phi(z) = p$, and

$$\partial_t\phi(z) - \nabla_{x_1}\phi(z) \cdot \nabla_{x_2}\phi(z) = \partial_t\phi(z) - p \cdot \nabla_{x_2}\phi(z) = 0.$$

In particular, u is a viscosity subsolution of the equation in (1.4). We note that this step uses no regularity of ψ_2 beyond continuity.

Step 3: The supersolution inequality when $\psi_2 \in C^1(\mathbb{R}^d)$. In this step, we assume that ψ_2 is continuously differentiable. For every $p \in \operatorname{dom} g$, the function w_p is then continuously differentiable on V, with

$$\nabla_{t,x} w_p(t', x_1', x_2') = \big(p \cdot \nabla\psi_2(x_2' + t'p),\, p,\, \nabla\psi_2(x_2' + t'p)\big) \in \mathbb{R} \times \mathbb{R}^d \times \mathbb{R}^d, \tag{2.5}$$

and the right-hand side depends continuously on (p, t', x_1', x_2'). We fix $z = (t, x_1, x_2) \in V$, and for $p \in \mathcal{A}(z)$, we set $b_p := \nabla\psi_2(x_2 + tp)$ and $\gamma_p := \nabla_{t,x} w_p(z) = (p \cdot b_p, p, b_p)$.

We first derive the first-order optimality condition satisfied by $p \in \mathcal{A}(z)$, namely

$$g(r) - g(p) \geqslant (x_1 + tb_p) \cdot (r - p), \qquad \forall r \in \mathbb{R}^d, \tag{2.6}$$

which states that $x_1 + tb_p$ is a subgradient of g at p. There is nothing to prove if $g(r) = +\infty$. Otherwise, for every $\varepsilon \in (0, 1]$, the point $p_\varepsilon := p + \varepsilon(r - p)$ belongs to $\operatorname{dom} g$, by convexity, and $w_{p_\varepsilon}(z) \leqslant u(z) = w_p(z)$. Together with the convexity inequality $g(p_\varepsilon) \leqslant g(p) + \varepsilon(g(r) - g(p))$, this gives

$$\varepsilon\,(r - p) \cdot x_1 + \psi_2\big(x_2 + tp + \varepsilon t(r - p)\big) - \psi_2(x_2 + tp) \leqslant g(p_\varepsilon) - g(p) \leqslant \varepsilon\big(g(r) - g(p)\big).$$

Dividing by ε and letting $\varepsilon \downarrow 0$ yields (2.6). For $p, r \in \mathcal{A}(z)$, we write (2.6) as is and also with p and r interchanged, and add the two inequalities. Since $t > 0$, this gives

$$(p - r) \cdot (b_p - b_r) \geqslant 0, \qquad \forall p, r \in \mathcal{A}(z). \tag{2.7}$$

Let now ϕ be a smooth function such that $u - \phi$ has a local minimum at z. We claim that

$$\nabla_{t,x}\phi(z) \cdot v \leqslant \max_{p \in \mathcal{A}(z)} \gamma_p \cdot v, \qquad \forall v \in \mathbb{R} \times \mathbb{R}^d \times \mathbb{R}^d. \tag{2.8}$$

We fix $v \in \mathbb{R} \times \mathbb{R}^d \times \mathbb{R}^d$, and for every $h > 0$ small enough that $z + hv \in V$, we pick $p_h \in \mathcal{A}(z + hv)$. Since $u(z + hv) = w_{p_h}(z + hv)$ and $u(z) \geqslant w_{p_h}(z)$, the local minimum of $u - \phi$ at z and the mean value theorem give

$$\phi(z + hv) - \phi(z) \leqslant u(z + hv) - u(z) \leqslant w_{p_h}(z + hv) - w_{p_h}(z) = h\,\nabla_{t,x} w_{p_h}(z + \theta_h hv) \cdot v \tag{2.9}$$

for some $\theta_h \in (0, 1)$. By (2.1), there is a sequence $h_n \downarrow 0$ such that p_{h_n} converges to some $p \in \mathbb{R}^d$. We verify that $p \in \mathcal{A}(z)$. Writing $z_n := z + h_n v = (t_n, x_{1,n}, x_{2,n})$, the continuity of ψ_2 and u gives

$$g(p_{h_n}) = p_{h_n} \cdot x_{1,n} + \psi_2(x_{2,n} + t_n p_{h_n}) - u(z_n) \xrightarrow[n \to \infty]{} p \cdot x_1 + \psi_2(x_2 + tp) - u(z),$$

and the lower semicontinuity of g yields $g(p) \leqslant p \cdot x_1 + \psi_2(x_2 + tp) - u(z)$, that is, $w_p(z) \geqslant u(z)$. Hence $p \in \mathcal{A}(z)$. By the continuity of the right-hand side of (2.5), we have $\lim_{n\to\infty} \nabla_{t,x} w_{p_{h_n}}(z + \theta_{h_n} h_n v) = \gamma_p$. Dividing (2.9) by h_n and letting $n \to \infty$, we obtain $\nabla_{t,x}\phi(z) \cdot v \leqslant \gamma_p \cdot v$, which proves (2.8).

The set $\Gamma_z := \{\gamma_p : p \in \mathcal{A}(z)\}$ is compact, being the image of the compact set $\mathcal{A}(z)$ under a continuous map, so its convex hull $\operatorname{co}\Gamma_z$ is compact and convex, by Carathéodory's theorem. If $\nabla_{t,x}\phi(z)$ did not belong to $\operatorname{co}\Gamma_z$, then by the separating hyperplane theorem, there would exist $v \in \mathbb{R} \times \mathbb{R}^d \times \mathbb{R}^d$ such that $\nabla_{t,x}\phi(z) \cdot v > \max_{\gamma\in\Gamma_z} \gamma \cdot v$, contradicting (2.8). Hence there exist $m \in \mathbb{N}$, $p_1, \dots, p_m \in \mathcal{A}(z)$, and $\alpha_1, \dots, \alpha_m \geqslant 0$ with $\sum_{i=1}^m \alpha_i = 1$ such that

$$\nabla_{t,x}\phi(z) = \sum_{i=1}^m \alpha_i \gamma_{p_i} = \sum_{i=1}^m \alpha_i (p_i \cdot b_{p_i}, p_i, b_{p_i}).$$

Consequently, we have

$$\begin{aligned}\partial_t\phi(z) - \nabla_{x_1}\phi(z) \cdot \nabla_{x_2}\phi(z) &= \sum_{i=1}^m \alpha_i p_i \cdot b_{p_i} - \Big(\sum_{i=1}^m \alpha_i p_i\Big) \cdot \Big(\sum_{j=1}^m \alpha_j b_{p_j}\Big) \\ &= \frac{1}{2} \sum_{i,j=1}^m \alpha_i \alpha_j (p_i - p_j) \cdot (b_{p_i} - b_{p_j}) \overset{(2.7)}{\geqslant} 0.\end{aligned}$$

This is the viscosity supersolution inequality.

Step 4: Conclusion. We now remove the differentiability assumption on ψ_2. For $\varepsilon > 0$, let $\psi_{2,\varepsilon} := \psi_2 * \rho_\varepsilon$, where ρ_ε is a smooth probability density supported in the ball of radius ε. The function $\psi_{2,\varepsilon}$ is smooth and L_2-Lipschitz, and it satisfies $\sup_{\mathbb{R}^d} |\psi_{2,\varepsilon} - \psi_2| \leqslant L_2\varepsilon$. Let u_ε be defined by (1.5) with $\psi_{2,\varepsilon}$ in place of ψ_2. By Steps 1 to 3, the function u_ε is a viscosity solution to the equation in (1.4), and comparing (1.5) for ψ_2 and for $\psi_{2,\varepsilon}$ directly gives

$$\sup_{\mathbb{R}_+\times\mathbb{R}^d\times\mathbb{R}^d} |u_\varepsilon - u| \leqslant \sup_{\mathbb{R}^d} |\psi_{2,\varepsilon} - \psi_2| \leqslant L_2\varepsilon. \tag{2.10}$$

We conclude by a standard stability argument. Let ϕ be a smooth function such that $u - \phi$ has a local minimum at $z_0 \in V$. Replacing ϕ by $z \mapsto \phi(z) - |z - z_0|^2$, which does not change $\nabla_{t,x}\phi(z_0)$, we may assume that this minimum is strict, so that there is a closed ball $B \subseteq V$ centered at z_0 with $u - \phi > (u - \phi)(z_0)$ on $B \setminus \{z_0\}$. For every $\varepsilon > 0$, let z_ε be a minimum point of $u_\varepsilon - \phi$ on B. By (2.10), every limit point of $(z_\varepsilon)_{\varepsilon\downarrow 0}$ is a minimum point of $u - \phi$ on B, and thus equals z_0. Hence z_ε tends to z_0, so for ε small, z_ε lies in the interior of B, and the supersolution inequality for u_ε at z_ε gives $\partial_t\phi(z_\varepsilon) - \nabla_{x_1}\phi(z_\varepsilon) \cdot \nabla_{x_2}\phi(z_\varepsilon) \geqslant 0$. Letting $\varepsilon \downarrow 0$ and using the continuity of $\nabla_{t,x}\phi$, we obtain the supersolution inequality for u at z_0. Together with Steps 1 and 2, this shows that u is a Lipschitz continuous viscosity solution to (1.4). Finally, Lipschitz continuous viscosity solutions to (1.4) are unique, by the comparison principle for Hamilton–Jacobi equations with locally Lipschitz nonlinearities; see for instance [17, Corollary 3.7]. □

Remark 2.2. The proof above uses ψ_1 only through $g = \psi_1^*$, and the only properties of g that were used are that it is proper, lower semicontinuous, and convex, with bounded effective domain. Conversely, if $g : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ is any function with these properties, then $\psi_1 := g^*$ is finite, convex and Lipschitz continuous, with Lipschitz constant $\sup_{p\in\operatorname{dom} g} |p|$, and $\psi_1^* = g^{**} = g$, by the Fenchel–Moreau theorem. Theorem 1.1 can therefore be rephrased as follows: for every such g and every Lipschitz continuous $\psi_2 : \mathbb{R}^d \to \mathbb{R}$, all the conclusions of Theorem 1.1 hold, with $L_1 := \sup_{p\in\operatorname{dom} g} |p|$, for the function

$$(t, x_1, x_2) \mapsto \sup_{p\in\operatorname{dom} g} \big\{p \cdot x_1 + \psi_2(x_2 + tp) - g(p)\big\}, \tag{2.11}$$

which is thus the unique Lipschitz continuous viscosity solution to the equation in (1.4) with initial condition $g^*(x_1) + \psi_2(x_2)$. We use this formulation in the proof of Proposition 2.3 below, and in Section 7.2.

2.3. **Extension to cones.** Let E be a finite-dimensional Euclidean space. For a closed convex cone K in a Hilbert space $\mathcal{H}$, we define the **dual cone** $K^* := \{h \in \mathcal{H} : \langle h, x\rangle_{\mathcal{H}} \geqslant 0,\ \forall x \in K\}$, and we say that a function h defined on a subset S of $\mathcal{H}$ is K^***-increasing** if $h(y) \geqslant h(x)$ whenever $x, y \in S$ and $y - x \in K^*$. Let $C \subseteq E$ be a pointed closed convex cone with nonempty interior; we recall that a cone C is said to be pointed if $C \cap (-C) = \{0\}$. We have $(C^2)^* = (C^*)^2$, and we say that a function $v : \mathbb{R}_+ \times C^2 \to \mathbb{R}$ is $(C^*)^2$-increasing if $v(t, \cdot)$ is $(C^*)^2$-increasing for every $t \geqslant 0$. Suppose that $\psi_1, \psi_2 : C \to \mathbb{R}$ are C^*-increasing and Lipschitz, with constants L_1, L_2 respectively, and that ψ_1 is convex. We define $\psi_{1,C}^* : C \to \mathbb{R} \cup \{+\infty\}$ and $u : \mathbb{R}_+ \times C^2 \to \mathbb{R}$ by

$$\psi_{1,C}^*(p) := \sup_{y \in C} \{p \cdot y - \psi_1(y)\}, \tag{2.12}$$

$$u(t, x_1, x_2) := \sup_{p \in C} \{p \cdot x_1 + \psi_2(x_2 + tp) - \psi_{1,C}^*(p)\}. \tag{2.13}$$

Proposition 2.3. *The supremum in (2.13) is attained. The function u is $(C^*)^2$-increasing, and for every $t, s \geqslant 0$ and $x_1, x_2, y_1, y_2 \in C$, we have*

$$|u(t, x_1, x_2) - u(s, y_1, y_2)| \leqslant L_1|x_1 - y_1| + L_2|x_2 - y_2| + L_1L_2|t - s|. \tag{2.14}$$

Moreover, u is a viscosity solution of

$$\partial_t u - \nabla_{x_1} u \cdot \nabla_{x_2} u = 0 \qquad \text{on } (0, \infty) \times (\operatorname{int} C)^2, \tag{2.15}$$

with initial condition $u(0, x_1, x_2) = \psi_1(x_1) + \psi_2(x_2)$ for all $x_1, x_2 \in C$, and it is the only Lipschitz, $(C^*)^2$-increasing function $v : \mathbb{R}_+ \times C^2 \to \mathbb{R}$ with these two properties.

Proof. Let $g : E \to \mathbb{R} \cup \{+\infty\}$ be equal to $\psi_{1,C}^*$ on C and to $+\infty$ outside of C. As a supremum of affine functions on the closed convex set C, this function is convex and lower semicontinuous. For every $p \in C$ and $a \geqslant 0$, the choice of $y = ap$ in (2.12) and the Lipschitz continuity of ψ_1 give $g(p) \geqslant a|p|^2 - \psi_1(ap) \geqslant a|p|(|p| - L_1) - \psi_1(0)$. Letting $a \to \infty$ shows that

$$\operatorname{dom} g \subseteq \{p \in C : |p| \leqslant L_1\}. \tag{2.16}$$

We next show the Fenchel–Moreau identity on the cone,

$$\psi_1(x) = \sup_{p \in C} \{p \cdot x - \psi_{1,C}^*(p)\} = g^*(x), \qquad \forall x \in C, \tag{2.17}$$

which in particular shows that g is proper. The inequality $\geqslant$ in (2.17) follows from the definition (2.12). Conversely, let first $x \in \operatorname{int} C$, and let $p \in E$ be a subgradient of ψ_1 at x relative to C, meaning that $\psi_1(y) \geqslant \psi_1(x) + p \cdot (y - x)$ for every $y \in C$. Such a subgradient exists since ψ_1 is convex and finite on a neighborhood of x; see [17, Proposition 2.11]. For every $h \in C^*$ and $\varepsilon > 0$ small enough that $x - \varepsilon h \in C$, the monotonicity of ψ_1 and the subgradient inequality give

$$0 \leqslant \psi_1(x) - \psi_1(x - \varepsilon h) \leqslant \varepsilon\, p \cdot h.$$

Thus $p \in (C^*)^* = C$, by the bipolar identity for closed convex cones; see [15, Section 1.1.1]. The subgradient inequality also shows that the supremum in (2.12) is attained at $y = x$, that is, $\psi_{1,C}^*(p) = p \cdot x - \psi_1(x)$. Hence $\psi_1(x) = p \cdot x - \psi_{1,C}^*(p) \leqslant g^*(x)$, and (2.17) holds on $\operatorname{int} C$. Both sides of (2.17) are Lipschitz on C, the right-hand side because of (2.16), and $\operatorname{int} C$ is dense in C, so the identity extends to C.

We now extend ψ_2 to E by setting

$$\tilde{\psi}_2(x) := \inf_{y \in C} \big\{\psi_2(y) + L_2|x - y|\big\}, \qquad \forall x \in E,$$

which defines an L_2-Lipschitz function on E that coincides with ψ_2 on C. By Theorem 1.1 in the form of Remark 2.2, applied to g and $\tilde{\psi}_2$, and by (2.16), the function

$$\tilde{u}(t, x_1, x_2) := \sup_{p \in \operatorname{dom} g} \big\{p \cdot x_1 + \tilde{\psi}_2(x_2 + tp) - g(p)\big\}, \qquad (t, x_1, x_2) \in \mathbb{R}_+ \times E^2,$$

is a viscosity solution of $\partial_t \tilde{u} - \nabla_{x_1} \tilde{u} \cdot \nabla_{x_2} \tilde{u} = 0$ on $(0, \infty) \times E^2$, with initial condition $\tilde{u}(0, x_1, x_2) = g^*(x_1) + \tilde{\psi}_2(x_2)$; moreover, the supremum defining $\tilde{u}$ is attained, and $\tilde{u}$ satisfies the estimate (1.6). Since $\operatorname{dom} g \subseteq C$ and C is a convex cone, for every $x_2 \in C$, $t \geqslant 0$ and $p \in \operatorname{dom} g$, we have $x_2 + tp \in C$, and thus $\tilde{\psi}_2(x_2 + tp) = \psi_2(x_2 + tp)$. As $\psi_{1,C}^*(p) = +\infty$ for $p \in C \setminus \operatorname{dom} g$, comparing

with (2.13) shows that $\tilde{u} = u$ on $\mathbb{R}_+ \times C^2$. Hence the supremum in (2.13) is attained, the estimate (2.14) holds, and by (2.17), we have $u(0, x_1, x_2) = \psi_1(x_1) + \psi_2(x_2)$ for every $x_1, x_2 \in C$. Since the viscosity property is local, and $(0,\infty) \times (\operatorname{int} C)^2$ is an open subset of $(0,\infty) \times E^2$ on which u and $\tilde{u}$ coincide, the function u is a viscosity solution of (2.15).

If $x_1, x_2, y_1, y_2 \in C$ are such that $y_1 - x_1$ and $y_2 - x_2$ belong to C^*, then for every $p \in C$, we have $p \cdot y_1 \geqslant p \cdot x_1$ and $\psi_2(y_2 + tp) \geqslant \psi_2(x_2 + tp)$, so that $u(t, y_1, y_2) \geqslant u(t, x_1, x_2)$. Hence u is $(C^*)^2$-increasing.

Finally, the Hamiltonian $\mathsf{H} : E^2 \to \mathbb{R}$, $\mathsf{H}(a_1, a_2) := a_1 \cdot a_2$, is locally Lipschitz, and its restriction to C^2 is $(C^*)^2$-increasing: if $a, b \in C^2$ and $a - b \in (C^*)^2$, then

$$\mathsf{H}(a) - \mathsf{H}(b) = (a_1 - b_1) \cdot a_2 + b_1 \cdot (a_2 - b_2) \geqslant 0.$$

The comparison principle of [15, Theorem 1.2(1)–(2)], applied in the Hilbert space E^2 with the cone C^2, therefore gives the uniqueness of Lipschitz, $(C^*)^2$-increasing viscosity solutions of (2.15) with prescribed initial condition. □

3. Regularity properties of the enriched free energy

We return to the bipartite model of the introduction and collect the regularity and compactness properties used below, as well as the convexity of $\psi_{\mu_{\mathcal{I}}}$.

The N Ising spins and the M_N ν-distributed spins form two species with one-dimensional single-spin laws $\mu_{\mathcal{I}}$ and ν. The covariance of H_N is N times the product of their overlaps $N^{-1}\sigma \cdot \sigma'$ and $N^{-1} v \cdot v'$, whose absolute values are bounded by 1 and M_N/N, respectively. Thus the Gaussian interpolation estimates of [9, Proposition 4.1], with normalization by N instead of $N + M_N$, yield the following result.

Proposition 3.1 ([9, Proposition 4.1])**.** *For every $N \in \mathbb{N}$, $t, t' \in \mathbb{R}_+$, and $q, q' \in \mathcal{Q}_\infty^2$, we have*

$$|\overline{F}_N(t, q) - \overline{F}_N(t', q')| \leqslant \frac{M_N}{N}|t - t'| + |q_1 - q_1'|_{L^1} + \frac{M_N}{N}|q_2 - q_2'|_{L^1}. \tag{3.1}$$

Consequently, $\overline{F}_N$ extends uniquely to $\mathbb{R}_+ \times \mathcal{Q}_1^2$.

In the notation of Section 2.3, the dual cone of $\mathcal{Q}_2$ in $L^2([0,1))$ is

$$\mathcal{Q}_2^* := \{h \in L^2([0,1)) : \langle h, q\rangle_{L^2} \geqslant 0,\ \forall q \in \mathcal{Q}_2\}. \tag{3.2}$$

A function $g : \mathcal{Q}_2 \to \mathbb{R}$ is **$\mathcal{Q}_2^*$-increasing** if

$$g(q') \geqslant g(q) \quad \text{whenever } q, q' \in \mathcal{Q}_2 \text{ and } q' - q \in \mathcal{Q}_2^*. \tag{3.3}$$

We use the notion of Gateaux differentiability from [12, Definition 2.2]. For $q \in \mathcal{Q}_2$, we define the set of admissible directions by

$$\operatorname{Adm}(\mathcal{Q}_2, q) := \left\{e \in L^2([0,1)) : \exists a > 0,\ \forall a' \in [0, a],\ q + a'e \in \mathcal{Q}_2\right\}.$$

A function $g : \mathcal{Q}_2 \to \mathbb{R}$ is **Gateaux differentiable** at $q \in \mathcal{Q}_2$ if, for every $e \in \operatorname{Adm}(\mathcal{Q}_2, q)$, the limit

$$g'(q, e) := \lim_{a \downarrow 0} \frac{g(q + ae) - g(q)}{a}$$

exists and there is a unique $y \in L^2([0,1))$ such that

$$g'(q, e) = \langle y, e\rangle_{L^2}, \qquad \forall e \in \operatorname{Adm}(\mathcal{Q}_2, q).$$

We call y the **Gateaux derivative** of g at q and write $\partial_q g(q) := y$. We recall the regularity properties needed below.

Lemma 3.2 ([12, Corollary 5.2])**.** *Let ρ be a probability measure with support on $[-1, 1]$. The function ψ_ρ defined in (1.18) extends uniquely to $\mathcal{Q}_1$, with $\psi_\rho(0) = 0$, and*

$$|\psi_\rho(q) - \psi_\rho(q')| \leqslant |q - q'|_{L^1}, \qquad \forall q, q' \in \mathcal{Q}_1. \tag{3.4}$$

Its restriction to $\mathcal{Q}_2$ is Gateaux differentiable and $\mathcal{Q}_2^*$-increasing, with

$$\partial_q \psi_\rho(q) \in \mathcal{Q}_{\infty, \leqslant 1}, \qquad \forall q \in \mathcal{Q}_2. \tag{3.5}$$

The derivative extends continuously to $\mathcal{Q}_1$, and

$$|\partial_q \psi_\rho(q) - \partial_q \psi_\rho(q')|_{L^1} \leqslant 16|q - q'|_{L^1}, \qquad \forall q, q' \in \mathcal{Q}_1. \tag{3.6}$$

We will also use the following compactness consequence of [12, Lemma 3.4].

Lemma 3.3. *For every $c \geqslant 0$ and $r \in [1, \infty)$, the set $\mathcal{Q}_{\infty,\leqslant c}$ is compact in $L^r([0,1))$.*

The cascade recursion separates over independent spin coordinates; see [9, Lemma 4.11]. Thus, defining $\psi : \mathcal{Q}_1^2 \to \mathbb{R}$ by $\psi(q) := \psi_{\mu_\mathcal{I}}(q_1) + \lambda\psi_\nu(q_2)$, we have

$$\overline{F}_N(0,q) = \psi_{\mu_\mathcal{I}}(q_1) + \frac{M_N}{N}\psi_\nu(q_2) \qquad \text{and} \qquad \lim_{N\to\infty} \overline{F}_N(0,q) = \psi(q). \tag{3.7}$$

For our model, we use the convexity only of $\psi_{\mu_\mathcal{I}}$, provided by the following result.

Proposition 3.4 ([11, Proposition 2.2], [26])**.** *The function $\psi_{\mu_\mathcal{I}} : \mathcal{Q}_1 \to \mathbb{R}$ is convex: for every $q, q' \in \mathcal{Q}_1$ and $\theta \in [0,1]$, we have*

$$\psi_{\mu_\mathcal{I}}((1-\theta)q + \theta q') \leqslant (1-\theta)\psi_{\mu_\mathcal{I}}(q) + \theta\psi_{\mu_\mathcal{I}}(q').$$

4. Variational formula on the space of paths

In this section, we extend Theorem 1.1 to the space of paths, by passing to the limit in Proposition 2.3 along dyadic discretizations. The main result of this section is Theorem 4.2 below, which identifies the right-hand side of (1.20) with the viscosity solution of (1.21).

We apply Definition 2.1 to $\mathcal{Q}_2^2 \subseteq L^2([0,1);\mathbb{R}^2)$. The identity $\overline{\mathcal{Q}_2^2 - \mathcal{Q}_2^2} = L^2([0,1);\mathbb{R}^2)$ ensures uniqueness of the differential in Definition 2.1; see [14, Remark 1.3].

We set $\xi : \mathbb{R}^2 \to \mathbb{R}$ by $\xi(a) := a_1 a_2$, and choose a Lipschitz regularization $\overline{\xi} : \mathbb{R}^2_+ \to \mathbb{R}$ agreeing with ξ on the ball of radius $2 \vee (1+\lambda)$. We require $\overline{\xi}$ to be coordinatewise increasing and, for every $b \in \mathbb{R}^2_+$, the function $a \mapsto \overline{\xi}(a+b) - \overline{\xi}(a)$ to be coordinatewise increasing. Such a regularization exists by [14, Lemma 4.4 and Remark 4.5]. Let $\mathsf{H} : L^2([0,1);\mathbb{R}^2) \to \mathbb{R}$ be such that

$$\mathsf{H}(h) := \inf\left\{ \int_0^1 \overline{\xi}(r(u))\,\mathrm{d}u : r \in \mathcal{Q}_2^2,\ r - h \in (\mathcal{Q}_2^*)^2 \right\}. \tag{4.1}$$

By [14, Lemma 4.6], the map H is Lipschitz and $(\mathcal{Q}_2^*)^2$-increasing, and for every $r \in \mathcal{Q}_2^2$, we have $\mathsf{H}(r) = \int_0^1 \overline{\xi}(r(u))\,\mathrm{d}u$. We use the notion of viscosity solution from [14, Definition 4.3].

Definition 4.1. A continuous function $f : \mathbb{R}_+ \times \mathcal{Q}_2^2 \to \mathbb{R}$ is a **viscosity solution of** (1.21) if $f(0,\cdot) = \psi$ and it solves $\partial_t f - \mathsf{H}(\partial_q f) = 0$ in the sense of Definition 2.1 on $(0,\infty) \times \mathcal{Q}_2^2$, with the regularized Hamiltonian H from (4.1).

Existence, uniqueness in the Lipschitz class, and convergence of finite-dimensional approximations follow from [14, Theorem 4.7(1)]. In particular, the solution in Definition 4.1 does not depend on the choice of $\overline{\xi}$; see [14, Remark 4.10].[1] The scalar two-species setting is treated by the same argument; see also [11, Remark 7.7].

For $j \in \mathbb{N}$, we equip $E_j := \mathbb{R}^{2^j}$ with the scalar product $\langle x, y\rangle_j := 2^{-j}\sum_{k=1}^{2^j} x_k y_k$, and set $C_j := \{x \in E_j : 0 \leqslant x_1 \leqslant \cdots \leqslant x_{2^j}\}$. We define the lift $\iota_j : E_j \to L^2([0,1))$ and the averaging map $\pi_j : L^2([0,1)) \to E_j$ by

$$\iota_j x := \sum_{k=1}^{2^j} x_k \mathbf{1}_{[(k-1)2^{-j}, k2^{-j})} \qquad \text{and} \qquad (\pi_j q)_k := 2^j \int_{(k-1)2^{-j}}^{k2^{-j}} q(r)\,\mathrm{d}r, \quad k \in \{1,\dots,2^j\}. \tag{4.2}$$

These maps preserve the increasing cones, ι_j is an isometry, and

$$\lim_{j\to\infty} |\iota_j \pi_j q - q|_{L^2} = 0, \qquad \forall q \in L^2([0,1)). \tag{4.3}$$

[1] To use the unit Lipschitz normalization in [14, Theorem 4.7], it suffices to divide the unknown function and its initial condition by $1+\lambda$, and to replace ξ by $a \mapsto \xi((1+\lambda)a)/(1+\lambda)$. This is legitimate since (3.4) gives an L^1-Lipschitz constant at most $1+\lambda$ for ψ.

For $\rho \in \{\mu_{\mathcal{I}}, \nu\}$, we define $\psi_{\rho,j} : C_j \to \mathbb{R}$ by $\psi_{\rho,j}(x) := \psi_\rho(\iota_j x)$. Its cone conjugate for $\rho = \mu_{\mathcal{I}}$ is $\psi^*_{\mu_{\mathcal{I}},j} : C_j \to \mathbb{R} \cup \{+\infty\}$, defined by

$$\psi^*_{\mu_{\mathcal{I}},j}(p) := \sup_{x \in C_j} \{\langle p, x\rangle_j - \psi_{\mu_{\mathcal{I}}}(\iota_j x)\}. \tag{4.4}$$

We now consider the following finite-dimensional approximation of (1.21):

$$\begin{cases} \partial_t f_j - \langle \nabla_{x_1} f_j, \nabla_{x_2} f_j \rangle_j = 0, & \text{on } (0,\infty) \times (\operatorname{int} C_j)^2, \\ f_j(0, x_1, x_2) = \psi_{\mu_{\mathcal{I}},j}(x_1) + \lambda \psi_{\nu,j}(x_2), & \text{for } (x_1, x_2) \in C_j^2. \end{cases} \tag{4.5}$$

Here the gradients $\nabla_{x_i} f_j$ are taken with respect to $\langle \cdot, \cdot \rangle_j$: at differentiability points, the derivative in direction $h \in E_j$ is $\langle \nabla_{x_i} f_j, h\rangle_j$. This problem is of the form considered in Proposition 2.3, with the Euclidean space E_j, the cone C_j, and the functions $\psi_1 = \psi_{\mu_{\mathcal{I}},j}$ and $\psi_2 = \lambda\psi_{\nu,j}$. The assumptions of this proposition are satisfied: the functions $\psi_{\rho,j}$ are C_j^*-increasing because $\iota_j(C_j^*) \subseteq \mathcal{Q}_2^*$, as follows from the identity $\langle \iota_j h, r\rangle_{L^2} = \langle h, \pi_j r\rangle_j$; the function $\psi_{\mu_{\mathcal{I}},j}$ is convex by Proposition 3.4; and Lemma 3.2 gives the Lipschitz bounds. Proposition 2.3 therefore ensures that (4.5) admits a unique Lipschitz, $(C_j^*)^2$-increasing viscosity solution $f_j : \mathbb{R}_+ \times C_j^2 \to \mathbb{R}$, and that this solution is given by

$$f_j(t, x_1, x_2) = \sup_{p \in C_j} \{\langle p, x_1\rangle_j + \lambda \psi_{\nu,j}(x_2 + tp) - \psi^*_{\mu_{\mathcal{I}},j}(p)\}. \tag{4.6}$$

Theorem 4.2. *The right-hand side of (1.20) defines the unique Lipschitz viscosity solution $f : \mathbb{R}_+ \times \mathcal{Q}_2^2 \to \mathbb{R}$ of (1.21), in the sense of Definition 4.1. The supremum can be restricted to $\mathcal{Q}_{\infty,\leqslant 1}$ and is attained. For every $t, s \geqslant 0$ and $q, y \in \mathcal{Q}_2^2$, we have*

$$|f(t,q) - f(s,y)| \leqslant |q_1 - y_1|_{L^1} + \lambda |q_2 - y_2|_{L^1} + \lambda |t - s|. \tag{4.7}$$

Proof. Step 1: Restriction to $\mathcal{Q}_{\infty,\leqslant 1}$. Recall the conjugate $\psi^*_{\mu_{\mathcal{I}}}$ defined in (1.19). We first show that $\psi^*_{\mu_{\mathcal{I}}}(p) < +\infty$ implies $p \in \mathcal{Q}_{\infty,\leqslant 1}$. If $p(r) > 1$ for some $r \in [0,1)$, then (3.4), applied with $q_a := a\mathbf{1}_{[r,1)}$ for $a \geqslant 0$, gives

$$\psi^*_{\mu_{\mathcal{I}}}(p) \geqslant a \int_r^1 (p(u) - 1)\, \mathrm{d}u, \qquad \forall a \geqslant 0.$$

The integral is strictly positive by monotonicity of p, so $\psi^*_{\mu_{\mathcal{I}}}(p) = +\infty$. Similarly, using $x = (0, \dots, 0, a) \in C_j$ in (4.4), with $a \geqslant 0$, shows that $\psi^*_{\mu_{\mathcal{I}},j}(p) < +\infty$ implies $p_{2^j} \leqslant 1$. Finally, $\psi^*_{\mu_{\mathcal{I}}}(0) = \psi^*_{\mu_{\mathcal{I}},j}(0) = 0$, since $\psi_{\mu_{\mathcal{I}}}(0) = 0$ and (3.5) gives $\psi_{\mu_{\mathcal{I}}}(q) \geqslant 0$ for every $q \in \mathcal{Q}_\infty$.

Step 2: Identification of the limit of the dyadic solutions. By [14, Theorem 4.7(1)], the unique viscosity solution f of (1.21) satisfies

$$f(t,q) = \lim_{j \to \infty} f_j(t, \pi_j q_1, \pi_j q_2), \qquad \forall (t,q) \in \mathbb{R}_+ \times \mathcal{Q}_2^2. \tag{4.8}$$

We identify this limit with the right-hand side of (1.20). For every $p \in \mathcal{Q}_\infty$, the definitions (4.4) and (1.19), together with the identity $\langle \pi_j p, x\rangle_j = \langle p, \iota_j x\rangle_{L^2}$ give

$$\psi^*_{\mu_{\mathcal{I}},j}(\pi_j p) = \sup_{x \in C_j} \{\langle p, \iota_j x\rangle_{L^2} - \psi_{\mu_{\mathcal{I}}}(\iota_j x)\} \leqslant \psi^*_{\mu_{\mathcal{I}}}(p). \tag{4.9}$$

Using $\pi_j p \in C_j$ in (4.6) and applying (4.9), we obtain

$$f_j(t, \pi_j q_1, \pi_j q_2) \geqslant \langle \pi_j p, \pi_j q_1\rangle_j + \lambda \psi_\nu\big(\iota_j \pi_j q_2 + t \iota_j \pi_j p\big) - \psi^*_{\mu_{\mathcal{I}}}(p).$$

For p with $\psi^*_{\mu_{\mathcal{I}}}(p) < +\infty$, Step 1 gives $p \in \mathcal{Q}_{\infty,\leqslant 1}$. By (4.3), we have $\lim_{j\to\infty} \langle \pi_j p, \pi_j q_1\rangle_j = \langle p, q_1\rangle_{L^2}$, while (3.4) and (4.3) imply

$$\lim_{j \to \infty} \psi_\nu(\iota_j \pi_j q_2 + t \iota_j \pi_j p) = \psi_\nu(q_2 + tp).$$

Consequently, by (4.8), we obtain

$$f(t,q) \geqslant \langle p, q_1\rangle_{L^2} + \lambda \psi_\nu(q_2 + tp) - \psi^*_{\mu_{\mathcal{I}}}(p).$$

Taking the supremum over $p \in \mathcal{Q}_\infty$ gives $f(t,q) \geqslant$ r.h.s. of (1.20).

Conversely, Proposition 2.3 ensures that the supremum in (4.6) is attained. For each $j \in \mathbb{N}$, we choose a maximizer $p_j \in C_j$ at $(t, \pi_j q_1, \pi_j q_2)$. Step 1 gives $0 \leqslant (p_j)_1 \leqslant \cdots \leqslant (p_j)_{2^j} \leqslant 1$. Since the expression in braces in (4.6) at $p = p_j$ is at least its value at $p = 0$, and $\psi^*_{\mu_{\mathcal{I}},j}(0) = 0$, we have

$$\langle p_j, \pi_j q_1 \rangle_j + \lambda \psi_\nu(\iota_j \pi_j q_2 + t\iota_j p_j) - \psi^*_{\mu_{\mathcal{I}},j}(p_j) \geqslant \lambda \psi_\nu(\iota_j \pi_j q_2).$$

Thus, using (3.4), $0 \leqslant \iota_j p_j \leqslant 1$, and $q_1 \geqslant 0$, we obtain

$$0 \leqslant \psi^*_{\mu_{\mathcal{I}},j}(p_j) \leqslant \langle p_j, \pi_j q_1 \rangle_j + \lambda\Big(\psi_\nu(\iota_j \pi_j q_2 + t\iota_j p_j) - \psi_\nu(\iota_j \pi_j q_2)\Big) \leqslant |q_1|_{L^1} + \lambda t. \tag{4.10}$$

By Lemma 3.3, we may pass to a subsequence for which $\lim_{j\to\infty} |\iota_j p_j - p|_{L^2} = 0$ for some $p \in \mathcal{Q}_{\infty,\leqslant 1}$. For every $r \in \mathcal{Q}_\infty$, taking $x = \pi_j r$ in (4.4) gives

$$\psi^*_{\mu_{\mathcal{I}},j}(p_j) \geqslant \langle \iota_j p_j, \iota_j \pi_j r \rangle_{L^2} - \psi_{\mu_{\mathcal{I}}}(\iota_j \pi_j r).$$

Passing to the lower limit, using (4.3), (3.4), and the L^2 convergence of $\iota_j p_j$, and then taking the supremum over $r \in \mathcal{Q}_\infty$, we obtain

$$\liminf_{j\to\infty} \psi^*_{\mu_{\mathcal{I}},j}(p_j) \geqslant \psi^*_{\mu_{\mathcal{I}}}(p). \tag{4.11}$$

In particular, $\psi^*_{\mu_{\mathcal{I}}}(p) < +\infty$ by (4.10). By (4.3) and the L^2 convergence of $\iota_j p_j$, we have

$$\lim_{j\to\infty} \langle p_j, \pi_j q_1 \rangle_j = \langle p, q_1 \rangle_{L^2} \qquad \text{and} \qquad \lim_{j\to\infty} \psi_\nu(\iota_j \pi_j q_2 + t\iota_j p_j) = \psi_\nu(q_2 + tp),$$

where the second identity uses the L^1-Lipschitz continuity in (3.4). Passing to the limit in (4.6) therefore gives

$$f(t,q) \overset{(4.11)}{\leqslant} \langle p, q_1 \rangle_{L^2} + \lambda \psi_\nu(q_2 + tp) - \psi^*_{\mu_{\mathcal{I}}}(p) \leqslant \text{r.h.s. of (1.20)}.$$

Together with the reverse inequality, this identifies f with the right-hand side of (1.20).

Step 3: Attainment and the Lipschitz estimate. On $\mathcal{Q}_{\infty,\leqslant 1}$, the function $\psi^*_{\mu_{\mathcal{I}}}$ is lower semicontinuous in the L^2 topology, being a supremum of continuous affine functions. By (3.4), the map

$$p \mapsto \langle p, q_1 \rangle_{L^2} + \lambda \psi_\nu(q_2 + tp) - \psi^*_{\mu_{\mathcal{I}}}(p)$$

is therefore upper semicontinuous on $\mathcal{Q}_{\infty,\leqslant 1}$. Since this set is compact by Lemma 3.3 and the function is finite at $p = 0$, the supremum is attained.

Finally, if $p \in \mathcal{Q}_{\infty,\leqslant 1}$ and $\psi^*_{\mu_{\mathcal{I}}}(p) < +\infty$, then (3.4) gives

$$\big|\, \langle p, q_1 - y_1 \rangle_{L^2} + \lambda \psi_\nu(q_2 + tp) - \lambda \psi_\nu(y_2 + sp) \big| \leqslant |q_1 - y_1|_{L^1} + \lambda |q_2 - y_2|_{L^1} + \lambda |t - s|.$$

Since this bound is uniform in p, and $f(t,q)$ and $f(s,y)$ are suprema over the same set of p, we obtain (4.7). □

5. Proof of Theorem 1.4

In this section, we complete the proof of Theorem 1.4, following the outline given in Section 1.4. We first show in Proposition 5.1 that the enriched free energy is asymptotically bounded from below by the solution to (1.21). We then show in Proposition 5.2 that it is asymptotically bounded from above by the value of an explicit functional at one of its critical points, following the argument developed in [11] for multi-species models with centered Ising spins. Finally, we use the convexity of $\psi_{\mu_{\mathcal{I}}}$ and the variational formula of Theorem 4.2 to show that this critical value is itself bounded from above by the solution to (1.21), which completes the argument.

Proposition 5.1 (Lower bound)**.** *Let $f : \mathbb{R}_+ \times \mathcal{Q}_2^2 \to \mathbb{R}$ be the unique Lipschitz viscosity solution to (1.21), in the sense of Definition 4.1. For every $(t,q) \in \mathbb{R}_+ \times \mathcal{Q}_2^2$, we have*

$$f(t,q) \leqslant \liminf_{N\to\infty} \overline{F}_N(t,q). \tag{5.1}$$

Proof. A bound of this form is proved in [34, Theorem 2.7] for bipartite models in which the two layers have the same number of spins. There, the enriched free energy is parametrized by pairs of probability measures on $\mathbb{R}_+$ instead of pairs of paths; we pass from one description to the other by identifying a probability measure with its quantile function. In order to reduce our setting to that of [34], we pad the two layers with zeros. We set $K_N := N \vee M_N$ and $c_N := K_N/N$, so that c_N converges to $c := 1 \vee \lambda$ as N tends to infinity. We identify each $\sigma \in \{-1,1\}^N$ with the vector of $\mathbb{R}^{K_N}$ obtained by appending $K_N - N$ coordinates equal to zero, we proceed similarly for each $v \in [-1,1]^{M_N}$, and we denote by $\widehat{P}_N$ the image of the measure P_N under this identification, which is a probability measure on $\mathbb{R}^{K_N} \times \mathbb{R}^{K_N}$. We let $\widehat{F}_N : \mathbb{R}_+ \times \mathcal{Q}_2^2 \to \mathbb{R}$ be the enriched free energy of the bipartite model with K_N spins in each layer and with reference measure $\widehat{P}_N$, as defined in [34, Section 2]. In this model, the Gaussian Hamiltonian is normalized by $K_N^{-1/2}$ in place of $N^{-1/2}$, and the free energy by K_N^{-1} in place of N^{-1}. Since the additional coordinates vanish, the Gaussian Hamiltonian of this model is $c_N^{-1/2} H_N(\sigma, v)$, while the fields indexed by the cascade are the same as in (1.15). The enriched Hamiltonian of this model at (s,q) is therefore the Hamiltonian in (1.16) at $(s/c_N, q)$, and we obtain that, for every $(s,q) \in \mathbb{R}_+ \times \mathcal{Q}_2^2$,

$$\widehat{F}_N(s,q) = \frac{1}{c_N} \overline{F}_N\left(\frac{s}{c_N}, q\right). \tag{5.2}$$

The measures $\widehat{P}_N$ are not product measures of the form assumed in the statement of [34, Theorem 2.7]. However, as explained in [34, the paragraph following (2.19)], the proof of this theorem only uses two properties of the reference measures: that the squared norm of each layer of spins is bounded by the number of spins in the layer, and that the free energy at time zero converges as the number of spins tends to infinity. Both properties are valid here. Indeed, under $\widehat{P}_N$, the squared norms of the two layers are equal to N and bounded by M_N respectively, and they are thus both bounded by K_N. Moreover, by (5.2) and (3.7), we have for every $q \in \mathcal{Q}_2^2$ that

$$\lim_{N\to\infty} \widehat{F}_N(0,q) = \widehat{\psi}(q), \qquad \text{where} \quad \widehat{\psi}(q) := \frac{\psi(q)}{c} = \frac{\psi_{\mu_\mathcal{I}}(q_1) + \lambda \psi_\nu(q_2)}{1 \vee \lambda}.$$

The proof of [34, Theorem 2.7] therefore yields that, for every $(s,q) \in \mathbb{R}_+ \times \mathcal{Q}_2^2$,

$$\widehat{f}(s,q) \leqslant \liminf_{N\to\infty} \widehat{F}_N(s,q), \tag{5.3}$$

where $\widehat{f} : \mathbb{R}_+ \times \mathcal{Q}_2^2 \to \mathbb{R}$ is the solution to (1.21) with $\widehat{\psi}$ in place of ψ as the initial condition. In [34, Section 3], this solution is defined as the limit of the solutions to finite-dimensional approximations of the equation; by the arguments of [14, Proposition 4.13 and Theorem 4.7(1)] (see also [11, Remark 7.7]), it is the unique Lipschitz viscosity solution in the sense of Definition 4.1, with $\widehat{\psi}$ in place of ψ.

We now fix $(t,q) \in \mathbb{R}_+ \times \mathcal{Q}_2^2$. By (5.2) with $s = ct$, we have $\overline{F}_N(ct/c_N, q) = c_N \widehat{F}_N(ct, q)$, while the time regularity in (3.1) and the convergence of c_N to c imply that $\overline{F}_N(t,q) - \overline{F}_N(ct/c_N, q)$ tends to zero as N tends to infinity. Combining these two observations with (5.3), we obtain

$$\liminf_{N\to\infty} \overline{F}_N(t,q) = c \liminf_{N\to\infty} \widehat{F}_N(ct,q) \geqslant c\, \widehat{f}(ct,q).$$

It remains to identify the right-hand side of this inequality. Since the nonlinearity in (1.21) is homogeneous of degree two, we see directly from the definition of viscosity solutions that the function $(t',q') \mapsto c\, \widehat{f}(ct', q')$ is a Lipschitz viscosity solution to (1.21) with initial condition $c\,\widehat{\psi} = \psi$.[2] By uniqueness, this function is f; in particular, we have $c\,\widehat{f}(ct,q) = f(t,q)$, and the proof of (5.1) is complete. □

We now turn to the converse bound. As explained in Section 1.4, we show that the upper limit of the free energy is bounded from above by the value of an explicit functional at one of its critical points. In the course of the argument, we will need to vary the number of spins in the second layer independently of N. For all integers $N, M \geqslant 1$, we denote by $\overline{F}_{N,M} : \mathbb{R}_+ \times \mathcal{Q}_1^2 \to \mathbb{R}$

[2] In Definition 4.1, this rescaling replaces the regularization $\overline{\xi}$ by $a \mapsto c^2\, \overline{\xi}(a/c)$, which is also an admissible regularization since $c \geqslant 1$; we recall that the solution does not depend on this choice.

the function defined as $\overline{F}_N$ in (1.16)–(1.17), but with M in place of M_N throughout. In other words, $\overline{F}_{N,M}$ is the enriched free energy of the model with N Ising spins and M spins sampled from ν, still normalized by N. In particular, we have $\overline{F}_N = \overline{F}_{N,M_N}$, and the estimate (3.1) is valid for $\overline{F}_{N,M}$, with M in place of M_N.

Some of the results of [9] that we rely on are stated for multi-species models in which the overlaps and the free energy are normalized by the total number of spins. In order to pass from one convention to the other, we set $\widetilde{\xi}(r_1, r_2) := (1+\lambda) r_1 r_2$, and for all integers $N, M \geqslant 1$, we let $\widetilde{F}_{N,M} : \mathbb{R}_+ \times \mathcal{Q}_1^2 \to \mathbb{R}$ denote the enriched free energy of the two-species model of [9] with N spins sampled from $\mu_{\mathcal{I}}$, with M spins sampled from ν, and with covariance function $\widetilde{\xi}$. Explicitly, writing $K := N + M$, the Gaussian Hamiltonian of this model has covariance

$$K\,\widetilde{\xi}\left(\frac{\sigma\cdot\sigma'}{K}, \frac{v\cdot v'}{K}\right) = \frac{(1+\lambda)N}{K}\,\mathbb{E}\left[H_N(\sigma, v) H_N(\sigma', v')\right],$$

the fields indexed by the cascade are the same as in (1.15), and the free energy is normalized by K instead of N. The enriched Hamiltonian of this model at (t, q) thus has the same law as the Hamiltonian in (1.16) at (t', q), with $t' := (1+\lambda)Nt/K$, and we obtain that, for every $(t, q) \in \mathbb{R}_+ \times \mathcal{Q}_1^2$,

$$\widetilde{F}_{N,M}(t, q) = \frac{N}{N+M}\,\overline{F}_{N,M}\left(\frac{(1+\lambda)N}{N+M}\,t, q\right). \tag{5.4}$$

The factor $1+\lambda$ in the definition of $\widetilde{\xi}$ is chosen so that the rescaling of the time variable in (5.4) becomes trivial in the regime of interest, in which M/N is close to λ.

For every $t > 0$ and $q \in \mathcal{Q}_1^2$, we define $\mathscr{P}_{t,q} : \mathcal{Q}_\infty^2 \to \mathbb{R}$ by

$$\mathscr{P}_{t,q}(p) := \psi_{\mu_{\mathcal{I}}}(q_1 + tp_2) + \lambda\psi_\nu(q_2 + tp_1) - t\,\langle p_1, p_2\rangle_{L^2}, \tag{5.5}$$

and we say that $p \in \mathcal{Q}_\infty^2$ is a **critical point** associated with (t, q) if

$$p_1 = \partial_q\psi_{\mu_{\mathcal{I}}}(q_1 + tp_2) \qquad \text{and} \qquad p_2 = \lambda\partial_q\psi_\nu(q_2 + tp_1). \tag{5.6}$$

We recall from Lemma 3.2 that the derivatives appearing in (5.6) are well-defined on $\mathcal{Q}_1$. Formally, the relations in (5.6) express that the derivative of the functional $\mathscr{P}_{t,q}$ vanishes at p, which explains the terminology.

Proposition 5.2 (Upper bound by a critical value)**.** *For every $t > 0$ and $q \in \mathcal{Q}_1^2$, there exists a critical point $p \in \mathcal{Q}_{\infty,\leqslant 1} \times \mathcal{Q}_{\infty,\leqslant\lambda}$ associated with (t, q) such that*

$$\limsup_{N\to\infty} \overline{F}_N(t, q) \leqslant \mathscr{P}_{t,q}(p). \tag{5.7}$$

Proof. For multi-species models in which every species consists of centered Ising spins, a bound of this form is proved in [11, Proposition 8.1]. The argument given there applies with very little change to the present setting, in which the spins of the second layer are sampled from ν. It consists in first treating the case in which the ratio λ is rational, by grouping the spins into blocks and viewing the model as a vector spin glass, for which the bound is obtained by means of cavity computations; and in then deducing the general case by approximation. In this argument, the assumption that the spins are Ising spins is only used to ensure that the self-overlap is constant, which simplifies the cavity computations; we explain in Step 1 below the modification that is required in the absence of this property. For the reader's convenience, and since our normalization differs from that of [11], we give the details of the rest of the argument. We fix $t > 0$ throughout the proof.

Step 1: The case of rational λ and bounded paths. We assume that $\lambda = a/b$ for some integers $a, b \geqslant 1$ and that $q \in \mathcal{Q}_\infty^2$, and we show that the statement of the proposition holds in this case. To do so, we group the spins into blocks made of b Ising spins and of a spins sampled from ν, and we view the model as a vector spin glass whose spins are these blocks. We set $I_1 := \{1, \dots, b\}$, $I_2 := \{b+1, \dots, b+a\}$, and $\mu := \mu_{\mathcal{I}}^{\otimes b} \otimes \nu^{\otimes a}$, which is a probability measure on

$\{-1,1\}^b \times [-1,1]^a \subseteq \mathbb{R}^{a+b}$. We define $\xi^{\mathrm{vec}} : \mathbb{R}^{a+b} \to \mathbb{R}$ by

$$\xi^{\mathrm{vec}}(y) := \frac{1}{b}\left(\sum_{i\in I_1} y_i\right)\left(\sum_{i\in I_2} y_i\right), \tag{5.8}$$

and we let $L : \mathbb{R}^2 \to \mathbb{R}^{a+b}$ be the linear map such that $(Lx)_i = x_1$ for $i \in I_1$ and $(Lx)_i = x_2$ for $i \in I_2$. For a pair of paths $q = (q_1, q_2)$, we write Lq for the path $s \mapsto L(q(s))$. For every integer $n \geqslant 1$, we consider configurations $\tau = (\tau_{ik})_{i\in I_1\cup I_2,\, 1\leqslant k\leqslant n} \in (\{-1,1\}^b \times [-1,1]^a)^n$, sampled according to the product measure $\mu^{\otimes n}$. Given independent standard Gaussian random variables $(g_{ik,j\ell})_{i\in I_1,\, j\in I_2,\, 1\leqslant k,\ell\leqslant n}$, we set

$$H_n^{\mathrm{vec}}(\tau) := \frac{1}{\sqrt{bn}} \sum_{\substack{i\in I_1,\, j\in I_2\\ 1\leqslant k,\ell\leqslant n}} g_{ik,j\ell}\, \tau_{ik}\tau_{j\ell}. \tag{5.9}$$

For two configurations τ and τ', we denote by $R(\tau,\tau') \in \mathbb{R}^{a+b}$ the vector with coordinates $R_i(\tau,\tau') := n^{-1}\sum_{k=1}^n \tau_{ik}\tau'_{ik}$, so that

$$\mathbb{E}\left[H_n^{\mathrm{vec}}(\tau) H_n^{\mathrm{vec}}(\tau')\right] = n\,\xi^{\mathrm{vec}}\big(R(\tau,\tau')\big).$$

For every $q^{\mathrm{vec}} \in \mathcal{Q}_\infty^{a+b}$, conditionally on $\mathfrak{R}$, we let $W_n^{q^{\mathrm{vec}}}$ be a centered Gaussian process indexed by (τ,α), independent of H_n^{vec}, with covariance

$$\mathbb{E}\left[W_n^{q^{\mathrm{vec}}}(\tau,\alpha) W_n^{q^{\mathrm{vec}}}(\tau',\alpha') \mid \mathfrak{R}\right] = n \sum_{i\in I_1\cup I_2} q_i^{\mathrm{vec}}(\alpha\wedge\alpha') R_i(\tau,\tau').$$

In analogy with (1.16)–(1.17), we define

$$\begin{aligned}\overline{F}_n^{\mathrm{vec}}(t, q^{\mathrm{vec}}) := -\frac{1}{n}\mathbb{E}\log \iint \exp\Bigg(&\sqrt{2t} H_n^{\mathrm{vec}}(\tau) + \sqrt{2} W_n^{q^{\mathrm{vec}}}(\tau,\alpha)\\ &- nt\,\xi^{\mathrm{vec}}\big(R(\tau,\tau)\big) - n\sum_{i\in I_1\cup I_2} q_i^{\mathrm{vec}}(1) R_i(\tau,\tau)\Bigg)\,\mathrm{d}\mu^{\otimes n}(\tau)\,\mathrm{d}\mathfrak{R}(\alpha).\end{aligned} \tag{5.10}$$

This is the enriched free energy of a vector spin glass with spins in $\mathbb{R}^{a+b}$: the single-spin measure μ has compact support; the covariance of H_n^{vec} is n times a polynomial of the matrix of overlaps $(n^{-1}\sum_{k=1}^n \tau_{ik}\tau'_{jk})_{i,j\in I_1\cup I_2}$, which here only involves the diagonal entries of this matrix; and the fields indexed by the cascade are those associated with the diagonal matrix-valued path with diagonal entries q^{vec}.

We now state the bound that we need for this vector spin glass. For every $q^{\mathrm{vec}} = (q_i^{\mathrm{vec}})_{i\in I_1\cup I_2} \in \mathcal{Q}_\infty^{a+b}$, we set

$$\psi^{\mathrm{vec}}(q^{\mathrm{vec}}) := \sum_{i\in I_1} \psi_{\mu_{\mathcal{I}}}(q_i^{\mathrm{vec}}) + \sum_{i\in I_2} \psi_\nu(q_i^{\mathrm{vec}}), \tag{5.11}$$

and we denote by $\partial_q\psi^{\mathrm{vec}}(q^{\mathrm{vec}})$ the family of paths whose coordinate indexed by i is $\partial_q\psi_{\mu_{\mathcal{I}}}(q_i^{\mathrm{vec}})$ if $i \in I_1$, and $\partial_q\psi_\nu(q_i^{\mathrm{vec}})$ if $i \in I_2$. As for (3.7), the fact that the cascade recursion separates over independent spin coordinates implies that $\overline{F}_n^{\mathrm{vec}}(0,\cdot) = \psi^{\mathrm{vec}}$ for every $n \geqslant 1$. We also write $\nabla\xi^{\mathrm{vec}}$ for the gradient of ξ^{vec}, and we set $\theta^{\mathrm{vec}}(y) := y\cdot\nabla\xi^{\mathrm{vec}}(y) - \xi^{\mathrm{vec}}(y)$ for every $y \in \mathbb{R}^{a+b}$. We claim that there exists $p^{\mathrm{vec}} = (p_i^{\mathrm{vec}})_{i\in I_1\cup I_2} \in \mathcal{Q}_\infty^{a+b}$ such that

$$p^{\mathrm{vec}} = \partial_q\psi^{\mathrm{vec}}\left(Lq + t\nabla\xi^{\mathrm{vec}}(p^{\mathrm{vec}})\right) \tag{5.12}$$

and

$$\limsup_{n\to\infty} \overline{F}_n^{\mathrm{vec}}(t, Lq) \leqslant \psi^{\mathrm{vec}}\left(Lq + t\nabla\xi^{\mathrm{vec}}(p^{\mathrm{vec}})\right) - t\int_0^1 \theta^{\mathrm{vec}}(p^{\mathrm{vec}}(s))\,\mathrm{d}s. \tag{5.13}$$

If ν were equal to $\mu_{\mathcal{I}}$, then this vector spin glass would be a special case of the model studied in [11, Sections 4 to 6], and the existence of such a path p^{vec} would be [11, Theorem 6.1]. The proof of this theorem applies essentially verbatim to the present model. The only difference is that, since the coordinates of the spins indexed by I_2 are now sampled from ν, the self-overlap $R(\tau,\tau)$ is no longer constant; as in [12, Section 6], one then needs to add to the Hamiltonian a small perturbation that forces the self-overlap to concentrate, and the cavity computations are

otherwise unchanged. This modification is explained in [11, Appendix A], in the more general setting of vector spin glasses whose covariance may depend on the full matrix of overlaps, and whose enriched free energy is therefore parametrized by paths with values in the set of positive semidefinite matrices; see [11, Theorem A.3].

We now express (5.12) and (5.13) in terms of the bipartite model. We set

$$p_1 := \frac{1}{b}\sum_{i\in I_1} p_i^{\mathrm{vec}}, \qquad p_2 := \frac{1}{b}\sum_{i\in I_2} p_i^{\mathrm{vec}}, \qquad \text{and} \qquad p := (p_1, p_2) \in \mathcal{Q}_\infty^2.$$

By (5.8), the coordinates of $\nabla\xi^{\mathrm{vec}}(p^{\mathrm{vec}})$ indexed by I_1 are all equal to p_2, and those indexed by I_2 are all equal to p_1; in other words, we have $Lq + t\nabla\xi^{\mathrm{vec}}(p^{\mathrm{vec}}) = L(q_1 + tp_2, q_2 + tp_1)$. The identity (5.12) thus states that $p_i^{\mathrm{vec}} = \partial_q\psi_{\mu_{\mathcal{I}}}(q_1 + tp_2)$ for every $i \in I_1$, and that $p_i^{\mathrm{vec}} = \partial_q\psi_\nu(q_2 + tp_1)$ for every $i \in I_2$. In view of the definition of p_1 and p_2, and since $\lambda = a/b$, this implies that p satisfies (5.6). That is, the pair p is a critical point associated with (t, q), and by (3.5), it belongs to $\mathcal{Q}_{\infty,\leqslant 1} \times \mathcal{Q}_{\infty,\leqslant\lambda}$. Moreover, since ξ^{vec} is homogeneous of degree two, we have $\theta^{\mathrm{vec}}(p^{\mathrm{vec}}) = \xi^{\mathrm{vec}}(p^{\mathrm{vec}}) = b\,p_1p_2$, and the right-hand side of (5.13) is therefore equal to

$$b\,\psi_{\mu_{\mathcal{I}}}(q_1 + tp_2) + a\,\psi_\nu(q_2 + tp_1) - tb\,\langle p_1, p_2\rangle_{L^2} = b\,\mathscr{P}_{t,q}(p).$$

We have thus shown that

$$\limsup_{n\to\infty} \overline{F}_n^{\mathrm{vec}}(t, Lq) \leqslant b\,\mathscr{P}_{t,q}(p). \tag{5.14}$$

We now relate this vector spin glass to the bipartite model. When $N = bn$ and $M = an$, relabeling the coordinates of $(\sigma, v) \in \{-1, 1\}^{bn} \times [-1, 1]^{an}$ as $(\tau_{ik})_{i\in I_1\cup I_2,\, 1\leqslant k\leqslant n}$ identifies the Hamiltonian H_{bn} in (1.13) with the Hamiltonian H_n^{vec} in (5.9), and the field W_{bn}^q in (1.15) with W_n^{Lq}. Using that $R_i(\tau, \tau) = 1$ for every $i \in I_1$, one can check that the correction terms also coincide, and since the two free energies are normalized by bn and by n respectively, we get

$$\overline{F}_{bn,an}(t, q) = \frac{1}{b}\,\overline{F}_n^{\mathrm{vec}}(t, Lq). \tag{5.15}$$

For a general sequence $(M_N)_{N\in\mathbb{N}}$ such that M_N/N converges to a/b, the identity (5.15) is replaced by an approximate version, which we borrow from [9]. In the notation of that paper, the function $\widetilde{F}_{N,M_N}$ is the free energy of a two-species model with $N + M_N$ spins, whose species proportions converge to $\frac{b}{a+b}$ and $\frac{a}{a+b}$. Moreover, since

$$(a + b)\,\widetilde{\xi}\left(\frac{y_1}{a+b}, \frac{y_2}{a+b}\right) = \frac{y_1y_2}{b},$$

the vector spin glass associated with this model in [9, Section 3] has covariance function ξ^{vec} as in (5.8), and its enriched free energy at the diagonal path associated with q is $\overline{F}_n^{\mathrm{vec}}(t, Lq)$, with $\overline{F}_n^{\mathrm{vec}}$ as in (5.10). Setting $n_N := \lceil (N + M_N)/(a + b)\rceil$, we thus obtain from [9, Corollary 3.2] that

$$\lim_{N\to\infty} \left|\widetilde{F}_{N,M_N}(t, q) - \frac{1}{a+b}\overline{F}_{n_N}^{\mathrm{vec}}(t, Lq)\right| = 0. \tag{5.16}$$

Since n_N tends to infinity with N, we deduce from (5.16) and (5.14) that

$$\limsup_{N\to\infty} \widetilde{F}_{N,M_N}(t, q) \leqslant \frac{b}{a+b}\,\mathscr{P}_{t,q}(p) = \frac{1}{1+\lambda}\,\mathscr{P}_{t,q}(p).$$

By (5.4), we have

$$\overline{F}_N\left(\frac{(1+\lambda)N}{N+M_N}\,t, q\right) = \frac{N+M_N}{N}\,\widetilde{F}_{N,M_N}(t, q).$$

The ratio $(N + M_N)/N$ converges to $1 + \lambda$, and by (3.1), the right-hand side of this identity remains bounded, while the difference between its left-hand side and $\overline{F}_N(t, q)$ tends to zero as N tends to infinity. We thus obtain (5.7).

Step 2: Continuity in the number of spins in the second layer. We show that, for all integers $N, M, M' \geqslant 1$ and every $q \in \mathcal{Q}_\infty^2$,

$$\left|\overline{F}_{N,M}(t, q) - \overline{F}_{N,M'}(t, q)\right| \leqslant \frac{|M - M'|}{N}\,(t + |q_2|_{L^1})\,. \tag{5.17}$$

We assume without loss of generality that $M \leqslant M'$, and we use the Gaussian interpolation of [9, Corollary 2.2], as in the proofs of [9, Lemmas 2.3 and 2.4]. We interpolate between the Hamiltonian of the model with M' spins in the second layer, and the Hamiltonian obtained from it by replacing v_j by zero for every $j \in \{M+1, \dots, M'\}$. Since ν is a probability measure, the free energies associated with these two Hamiltonians are $\overline{F}_{N,M'}(t,q)$ and $\overline{F}_{N,M}(t,q)$ respectively. By [9, Corollary 2.2], the difference $N\overline{F}_{N,M'}(t,q) - N\overline{F}_{N,M}(t,q)$ is therefore the integral along the interpolation of the averaged Gibbs expectation of

$$\left(\frac{t}{N} \sigma \cdot \sigma' + q_2(\alpha \wedge \alpha') \right) \sum_{j=M+1}^{M'} v_j v_j',$$

where (σ, v, α) and (σ', v', α') denote two independent replicas. Since $|\sigma_i| \leqslant 1$ and $|v_j| \leqslant 1$, the absolute value of this quantity is at most $(M'-M)(t + q_2(\alpha \wedge \alpha'))$. Moreover, by the invariance property of the cascade, the overlap $\alpha \wedge \alpha'$ is uniformly distributed over $[0,1]$ under each of the averaged Gibbs measures along the interpolation; see [12, Proposition 4.8]. The averaged Gibbs expectation of $q_2(\alpha \wedge \alpha')$ is thus equal to $|q_2|_{L^1}$, and we obtain (5.17).

Step 3: The general case. Let $q \in \mathcal{Q}_1^2$. Let $(\lambda_k)_{k \in \mathbb{N}}$ be a sequence of positive rational numbers that converges to λ, and for every $k, N \in \mathbb{N}$, let $M_N^{(k)} := \max\{1, \lfloor \lambda_k N \rfloor\}$, so that $M_N^{(k)}/N$ converges to λ_k as N tends to infinity. For every $k \in \mathbb{N}$, we also let $q^{(k)} \in \mathcal{Q}_\infty^2$ be the pair of truncated paths defined by $q_i^{(k)}(s) := \min\{q_i(s), k\}$, for $i \in \{1,2\}$ and $s \in [0,1)$, so that $q^{(k)}$ converges to q in L^1 as k tends to infinity. By Step 1, applied with $(M_N^{(k)})_{N \in \mathbb{N}}$, λ_k and $q^{(k)}$ in place of $(M_N)_{N \in \mathbb{N}}$, λ and q, there exists $p^{(k)} \in \mathcal{Q}_{\infty, \leqslant 1} \times \mathcal{Q}_{\infty, \leqslant \lambda_k}$ such that

$$p_1^{(k)} = \partial_q \psi_{\mu_{\mathcal{I}}}\big(q_1^{(k)} + t p_2^{(k)}\big) \qquad \text{and} \qquad p_2^{(k)} = \lambda_k \, \partial_q \psi_\nu\big(q_2^{(k)} + t p_1^{(k)}\big), \tag{5.18}$$

and

$$\limsup_{N \to \infty} \overline{F}_{N, M_N^{(k)}}\big(t, q^{(k)}\big) \leqslant \psi_{\mu_{\mathcal{I}}}\big(q_1^{(k)} + t p_2^{(k)}\big) + \lambda_k \psi_\nu\big(q_2^{(k)} + t p_1^{(k)}\big) - t \langle p_1^{(k)}, p_2^{(k)} \rangle_{L^2}. \tag{5.19}$$

By (3.1) and (5.17), we have

$$\left| \overline{F}_N(t,q) - \overline{F}_{N, M_N^{(k)}}\big(t, q^{(k)}\big) \right| \leqslant \big|q_1 - q_1^{(k)}\big|_{L^1} + \frac{M_N}{N} \big|q_2 - q_2^{(k)}\big|_{L^1} + \frac{|M_N - M_N^{(k)}|}{N} \left(t + |q_2|_{L^1}\right).$$

Letting N tend to infinity, we obtain that $\limsup_{N \to \infty} \overline{F}_N(t,q)$ is bounded from above by the right-hand side of (5.19), up to an error that tends to zero as k tends to infinity. By Lemma 3.3, up to the extraction of a subsequence, we may assume that $p^{(k)}$ converges in L^1 to some $p \in \mathcal{Q}_\infty^2$. The paths $q_1^{(k)} + t p_2^{(k)}$ and $q_2^{(k)} + t p_1^{(k)}$ then converge in L^1 to $q_1 + t p_2$ and $q_2 + t p_1$ respectively, and the continuity estimate (3.6) allows us to pass to the limit in (5.18). Hence p is a critical point associated with (t,q), and by (3.5), it belongs to $\mathcal{Q}_{\infty, \leqslant 1} \times \mathcal{Q}_{\infty, \leqslant \lambda}$. Moreover, since the paths $p_1^{(k)}$ and $p_2^{(k)}$ are uniformly bounded, their convergence in L^1 implies that $\langle p_1^{(k)}, p_2^{(k)} \rangle_{L^2}$ converges to $\langle p_1, p_2 \rangle_{L^2}$. Using also the Lipschitz estimate (3.4), we see that the right-hand side of (5.19) converges to $\mathscr{P}_{t,q}(p)$ as k tends to infinity, and the proof of (5.7) is complete. □

We can now complete the proof of Theorem 1.4, by combining the lower bound of Proposition 5.1 with the upper bound of Proposition 5.2, the convexity of $\psi_{\mu_{\mathcal{I}}}$, and the variational formula of Theorem 4.2.

Proof of Theorem 1.4. Let $f : \mathbb{R}_+ \times \mathcal{Q}_2^2 \to \mathbb{R}$ be the unique Lipschitz viscosity solution to (1.21), in the sense of Definition 4.1. By Theorem 4.2, the function f is given by the right-hand side of (1.20). It therefore suffices to show that, for every $(t,q) \in \mathbb{R}_+ \times \mathcal{Q}_2^2$,

$$\lim_{N \to \infty} \overline{F}_N(t,q) = f(t,q). \tag{5.20}$$

For $t = 0$, this follows from (3.7), since $f(0, \cdot) = \psi$. We now fix $t > 0$ and $q \in \mathcal{Q}_2^2$. By Proposition 5.2, there exists a critical point $p \in \mathcal{Q}_{\infty, \leqslant 1} \times \mathcal{Q}_{\infty, \leqslant \lambda}$ associated with (t,q) such

that (5.7) holds. We set $q_1' := q_1 + tp_2$, which belongs to $\mathcal{Q}_2$, and we recall from (5.6) that $p_1 = \partial_q \psi_{\mu_\mathcal{I}}(q_1')$. The main step of the proof is to show the Fenchel identity

$$\psi^*_{\mu_\mathcal{I}}(p_1) = \langle p_1, q_1' \rangle_{L^2} - \psi_{\mu_\mathcal{I}}(q_1'). \tag{5.21}$$

Let $y \in \mathcal{Q}_2$. For every $\theta \in (0,1]$, the path $q_1' + \theta(y - q_1') = (1-\theta)q_1' + \theta y$ belongs to $\mathcal{Q}_2$, and the convexity of $\psi_{\mu_\mathcal{I}}$ given by Proposition 3.4 yields

$$\frac{\psi_{\mu_\mathcal{I}}\big(q_1' + \theta(y - q_1')\big) - \psi_{\mu_\mathcal{I}}(q_1')}{\theta} \leqslant \psi_{\mu_\mathcal{I}}(y) - \psi_{\mu_\mathcal{I}}(q_1').$$

Letting θ tend to zero, and using that $\psi_{\mu_\mathcal{I}}$ is Gateaux differentiable at q_1' with derivative p_1, we obtain

$$\psi_{\mu_\mathcal{I}}(y) \geqslant \psi_{\mu_\mathcal{I}}(q_1') + \langle p_1, y - q_1' \rangle_{L^2}, \qquad \forall y \in \mathcal{Q}_2. \tag{5.22}$$

Since $\mathcal{Q}_\infty \subseteq \mathcal{Q}_2$, we deduce from (5.22) and the definition of $\psi^*_{\mu_\mathcal{I}}$ in (1.19) that

$$\psi^*_{\mu_\mathcal{I}}(p_1) \leqslant \langle p_1, q_1' \rangle_{L^2} - \psi_{\mu_\mathcal{I}}(q_1').$$

For the converse inequality, we cannot directly use q_1' as a test path in (1.19), since q_1' may be unbounded. For every $j \in \mathbb{N}$, we thus consider the truncated path $y_j := q_1' \wedge j \in \mathcal{Q}_\infty$, for which (1.19) gives

$$\psi^*_{\mu_\mathcal{I}}(p_1) \geqslant \langle p_1, y_j \rangle_{L^2} - \psi_{\mu_\mathcal{I}}(y_j).$$

As j tends to infinity, the path y_j converges to q_1' in L^1. Since $0 \leqslant p_1 \leqslant 1$, and by (3.4), the right-hand side of the previous display converges to $\langle p_1, q_1' \rangle_{L^2} - \psi_{\mu_\mathcal{I}}(q_1')$. This completes the proof of (5.21).

Since $\langle p_1, q_1' \rangle_{L^2} = \langle p_1, q_1 \rangle_{L^2} + t \langle p_1, p_2 \rangle_{L^2}$, we can combine the definition of $\mathscr{P}_{t,q}$ in (5.5) with the identity (5.21) to get

$$\mathscr{P}_{t,q}(p) = \langle p_1, q_1 \rangle_{L^2} + \lambda \psi_\nu(q_2 + tp_1) - \psi^*_{\mu_\mathcal{I}}(p_1) \leqslant f(t,q),$$

where the inequality follows from the fact that p_1 belongs to $\mathcal{Q}_\infty$ and that $f(t,q)$ is the supremum on the right-hand side of (1.20). Using also Proposition 5.1, we conclude that

$$f(t,q) \overset{(5.1)}{\leqslant} \liminf_{N\to\infty} \overline{F}_N(t,q) \leqslant \limsup_{N\to\infty} \overline{F}_N(t,q) \overset{(5.7)}{\leqslant} \mathscr{P}_{t,q}(p) \leqslant f(t,q),$$

and thus that (5.20) holds. □

Remark 5.3 (Measures with general compact support). Theorem 1.4 remains valid if the support of ν is only assumed to be contained in $[-a,a]$, for some $a > 0$. To see this, we write $\overline{F}^\nu_N$ in place of $\overline{F}_N$ to display the dependence of the free energy on the measure ν, and we denote by $\widehat{\nu}$ the image of ν under the mapping $v \mapsto v/a$, which is a probability measure supported on $[-1,1]$. The change of variables $v \mapsto v/a$ in (1.17) and in (1.18) gives

$$\overline{F}^\nu_N(t,q_1,q_2) = \overline{F}^{\widehat{\nu}}_N(a^2 t, q_1, a^2 q_2) \qquad \text{and} \qquad \psi_\nu(q_2) = \psi_{\widehat{\nu}}(a^2 q_2).$$

Applying Theorem 1.4 to the measure $\widehat{\nu}$, we thus obtain that the formula (1.20) holds for the measure ν as well. Moreover, in the passage from $\widehat{\nu}$ to ν, the time derivative of the limit and its derivative with respect to the second path are both multiplied by a^2, so the equation in (1.21) is unchanged. The only difference is that the derivative of the limit with respect to the second path is now bounded by λa^2 instead of λ; accordingly, in Definition 4.1, the regularization $\overline{\xi}$ must be chosen to agree with ξ on a ball of radius at least $2 \vee (1 + \lambda a^2)$.

6. A shrinking domain of paths

In Section 7, we will enrich general perceptron models in such a way that the variance of the pre-activations remains equal to one. This forces the time variable t and the path q_2 associated with the pre-activations to satisfy the constraint $q_2 \leqslant 1 - t$; see (7.2). The enriched free energy is then expected to satisfy, in the limit of large system size, the same equation as in (1.21), but on a domain that shrinks as t increases, and which is therefore not of the product form $(0,\infty) \times \mathcal{Q}_2^2$ considered in Section 4. The goal of this section is to show that this equation satisfies a comparison principle on this domain, with no condition imposed on its moving boundary, and

that its solution is still given by the variational formula of Theorem 4.2; see Proposition 6.4 and Theorem 6.7 below.

To see why no boundary condition is needed on the boundary of the set $\{q_2 \leqslant 1-t\}$, let us denote by f a solution to this equation, let $(t,q) = (t,q_1,q_2) \in (0,\infty)\times\mathcal{Q}_2^2$ be such that $q_2 \leqslant 1-t$, and suppose that f is differentiable at (t,q), with $p = (p_1,p_2) := \partial_q f(t,q)$. The characteristic line from (t,q) that goes backwards in time is given by $s \mapsto (t-s, q_1+sp_2, q_2+sp_1)$, for $s \in [0,t]$; see [17, (3.94)]. In order for this characteristic line not to exit the domain, it suffices that $p_1 \leqslant 1$, for in this case we indeed have $q_2 + sp_1 \leqslant 1-(t-s)$ for all $s \in [0,t]$. In other words, as long as $\partial_{q_1} f \leqslant 1$, characteristic lines may leave the domain through this boundary, but they cannot enter it. In our setting, the bound $\partial_{q_1} f \leqslant 1$ indeed holds, since $\partial_{q_1} f$ represents an average overlap between replicas of the spins, which cannot take values greater than 1; see (3.5) or (7.4).

In order to turn this observation into a rigorous statement, we need a notion of solution that encodes it. As in Definition 4.1, we require the viscosity inequalities to hold at every point of the domain with $t \in (0,1)$, including the points at which $q_2(1) = 1-t$. Since the gradient of a test function at a boundary point is not constrained by the bounds satisfied by the derivatives of the solution, this requires extending the nonlinearity beyond the set in which these derivatives take their values. The extension used in Definition 4.1 was designed for the cone $\mathcal{Q}_2^2$, and it does not satisfy the inequality that we need at the moving boundary, namely (6.15) below. We therefore use a different extension, built so that the outflow property just described also holds at the level of test functions; see Lemma 6.3. With this notion of solution, the comparison principle follows from the usual doubling of variables, and the main task is to show that the variational formula satisfies the viscosity inequalities up to the boundary.

6.1. **The domain and the notion of solution.** Throughout this section, we write $\mathcal{H} := L^2([0,1))$, and for every $c \geqslant 0$, we set $K_c := \mathcal{Q}_{\infty,\leqslant c}$. Each of the sets K_c is convex, and it is compact in $\mathcal{H}$ by Lemma 3.3. We denote by $\omega : \mathcal{H} \to \mathbb{R}$ the support function of K_1, that is,

$$\omega(x) := \sup_{p\in K_1} \langle p, x\rangle_{L^2}, \qquad x \in \mathcal{H}. \tag{6.1}$$

As a supremum of linear functions, the function ω is positively homogeneous and subadditive. Since $0 \in K_1$, and since every $p \in K_1$ takes values in $[0,1]$, we also have

$$0 \leqslant \omega(x) \leqslant |x|_{L^1} \leqslant |x|_{L^2}, \qquad \forall x \in \mathcal{H}. \tag{6.2}$$

In particular, the function ω is Lipschitz continuous. Since $K_c = cK_1$, the support function of K_c is $c\omega$. We define

$$\widehat{\mathcal{D}} := \{(t,q_1,q_2) : t \in [0,1],\ q_1 \in \mathcal{Q}_2,\ q_2 \in K_{1-t}\} \qquad \text{and} \qquad \mathcal{D} := \widehat{\mathcal{D}} \cap (\mathbb{R}\times\mathcal{Q}_\infty^2). \tag{6.3}$$

The condition $q_2 \in K_{1-t}$ means that $q_2 \in \mathcal{Q}_\infty$ and $q_2 \leqslant 1-t$. The set $\widehat{\mathcal{D}}$ is convex, and it is closed in the Hilbert space $\mathbb{R}\times\mathcal{H}^2$, since $\mathcal{Q}_2$ is closed in $\mathcal{H}$ and the constraint $q_2 \leqslant 1-t$ is preserved by convergence in $\mathbb{R}\times\mathcal{H}$. We write $\langle\cdot,\cdot\rangle$ and $|\cdot|$ for the scalar product and the norm of this space, and we use the same notation for the space $\mathcal{H}^2$. We often write $z = (t,q)$ for a generic point of $\widehat{\mathcal{D}}$, with $q = (q_1,q_2)$. Unless otherwise specified, the Lipschitz continuity of a function defined on a subset of $\mathbb{R}\times\mathcal{H}^2$ is understood with respect to the norm $|\cdot|$. The set $\mathcal{D}$ is the one on which the enriched free energies of Section 7 are defined. Truncating q_1 shows that $\mathcal{D}$ is dense in $\widehat{\mathcal{D}}$, so every Lipschitz function on $\mathcal{D}$ admits a unique Lipschitz extension to $\widehat{\mathcal{D}}$.

We now fix $B \geqslant 0$. In what follows, the sets K_1 and K_B are those in which the derivatives of the solution with respect to q_1 and to q_2 take their values, respectively. We define $\mathsf{H}_B : \mathcal{H}^2 \to \mathbb{R}$ by

$$\mathsf{H}_B(a,b) := \min_{p\in K_1}\max_{r\in K_B}\left\{\langle a,r\rangle_{L^2} + \langle p,b\rangle_{L^2} - \langle p,r\rangle_{L^2}\right\} = \min_{p\in K_1}\left\{\langle p,b\rangle_{L^2} + B\,\omega(a-p)\right\}. \tag{6.4}$$

The two expressions coincide because $B\,\omega$ is the support function of K_B, and the extrema are attained because K_1 and K_B are compact and ω is continuous. The next lemma records the properties of H_B that we will use: this function extends the bilinear nonlinearity in (1.21) from $K_1\times K_B$ to $\mathcal{H}^2$, and its increments are bounded from above in terms of the support functions of K_B and K_1.

Lemma 6.1 (Properties of H_B)**.** *For every $a \in K_1$ and $b \in K_B$, we have*

$$\mathsf{H}_B(a,b) = \langle a,b\rangle_{L^2}\,. \tag{6.5}$$

For every $a,b,a',b' \in \mathcal{H}$, we have

$$\mathsf{H}_B(a+a',b+b') - \mathsf{H}_B(a,b) \leqslant B\,\omega(a') + \omega(b'), \tag{6.6}$$

and in particular,

$$|\mathsf{H}_B(a,b) - \mathsf{H}_B(a',b')| \leqslant B|a-a'|_{L^2} + |b-b'|_{L^2}. \tag{6.7}$$

Proof. Let $a \in K_1$ and $b \in K_B$. Choosing $p = a$ in the second expression in (6.4) gives $\mathsf{H}_B(a,b) \leqslant \langle a,b\rangle_{L^2}$. Conversely, since $b \in K_B$ and $B\,\omega$ is the support function of K_B, we have $B\,\omega(a-p) \geqslant \langle b, a-p\rangle_{L^2}$ for every $p \in K_1$, and thus $\mathsf{H}_B(a,b) \geqslant \langle a,b\rangle_{L^2}$. This proves (6.5). Let now $a,b,a',b' \in \mathcal{H}$. For every $p \in K_1$, the subadditivity of ω and the definition (6.1) give

$$\langle p, b+b'\rangle_{L^2} + B\,\omega(a+a'-p) \leqslant \langle p,b\rangle_{L^2} + B\,\omega(a-p) + \omega(b') + B\,\omega(a'),$$

and taking the minimum over $p \in K_1$ yields (6.6). The estimate (6.7) follows from (6.6), applied in both directions, and from (6.2). □

Definition 6.2 (Viscosity solutions on the shrinking domain)**.** Let $B \geqslant 0$, and let

$$V := \Big\{(t,q_1,q_2) \in \widehat{\mathcal{D}} : 0 < t < 1\Big\}. \tag{6.8}$$

A continuous function $u : \widehat{\mathcal{D}} \to \mathbb{R}$ is a **viscosity subsolution**, respectively a **viscosity supersolution**, of

$$\partial_t u - \langle \partial_{q_1} u, \partial_{q_2} u\rangle_{L^2} = 0 \tag{6.9}$$

with derivative bound B if its restriction to V is a viscosity subsolution, respectively a viscosity supersolution, of

$$\partial_t u - \mathsf{H}_B(\partial_{q_1} u, \partial_{q_2} u) = 0 \tag{6.10}$$

in the sense of Definition 2.1, with ambient Hilbert space $\mathcal{H}^2$. A **viscosity solution** of (6.9) with derivative bound B is a function that is both a viscosity subsolution and a viscosity supersolution of (6.9) with derivative bound B. For $u_0 : \mathcal{Q}_2 \times K_1 \to \mathbb{R}$, we say that such a function u has initial condition u_0 if $u(0,\cdot) = u_0$ on $\mathcal{Q}_2 \times K_1$.

Let us comment on this definition. The local extrema in there are understood relative to V: for instance, $u-\phi$ has a local maximum at $z \in V$ if $(u-\phi)(z') \leqslant (u-\phi)(z)$ for every $z' \in V$ sufficiently close to z. Definition 2.1 involves test functions that are smooth on V, and the differential of such a function is required to be unique. This requirement is automatically satisfied here. Indeed, if $\zeta \in \mathbb{R}\times\mathcal{H}^2$ is the difference between two candidate differentials at a point $z \in V$, then $\langle \zeta, z'-z\rangle = o(|z'-z|)$ as $z' \in V$ tends to z. Since V is convex, we can apply this along the segment joining z to any other point of V, and we obtain that ζ is orthogonal to $V-z$, and thus also to $V-V$. The linear span of $V-V$ contains the vectors $(1,0,0)$, $(0,\mathbf{1}_{[s,s')},0)$ and $(0,0,\mathbf{1}_{[s,s')})$ for every $0 \leqslant s < s' \leqslant 1$, so it is dense in $\mathbb{R}\times\mathcal{H}^2$, and therefore $\zeta = 0$.

For a smooth function $\phi : V \to \mathbb{R}$ and $z \in V$, we write

$$D\phi(z) := (\partial_t\phi(z), \partial_{q_1}\phi(z), \partial_{q_2}\phi(z)) \in \mathbb{R}\times\mathcal{H}^2$$

for the differential of ϕ at z, and we set

$$\mathscr{F}_B(\tau,a,b) := \tau - \mathsf{H}_B(a,b), \qquad (\tau,a,b) \in \mathbb{R}\times\mathcal{H}^2. \tag{6.11}$$

With this notation, a continuous function $u : \widehat{\mathcal{D}} \to \mathbb{R}$ is a viscosity subsolution of (6.9) with derivative bound B if and only if $\mathscr{F}_B(D\phi(z)) \leqslant 0$ whenever $\phi : V \to \mathbb{R}$ is smooth and $u-\phi$ has a local maximum at $z \in V$; for supersolutions, the condition is that $\mathscr{F}_B(D\phi(z)) \geqslant 0$ whenever $u-\phi$ has a local minimum at $z \in V$.

The key property of this notion of solution is a compatibility between the function $\mathscr{F}_B$ and the geometry of the domain, which we now state. For every $z \in \widehat{\mathcal{D}}$, we denote the normal cone to $\widehat{\mathcal{D}}$ at z by

$$N_{\widehat{\mathcal{D}}}(z) := \left\{ n \in \mathbb{R} \times \mathcal{H}^2 : \langle n, z' - z \rangle \leqslant 0 \ \text{ for every } z' \in \widehat{\mathcal{D}} \right\}. \tag{6.12}$$

Lemma 6.3 (Outflow property). *Let $B \geqslant 0$ and $z = (t, q_1, q_2) \in V$. For every $p \in K_1$, $r \in K_B$, and $h \in [0, t]$, we have*

$$z + h(-1, r, p) \in \widehat{\mathcal{D}}. \tag{6.13}$$

As a consequence, every $n = (n_t, n_1, n_2) \in N_{\widehat{\mathcal{D}}}(z)$ satisfies

$$n_t \geqslant B\,\omega(n_1) + \omega(n_2), \tag{6.14}$$

and for every $\zeta \in \mathbb{R} \times \mathcal{H}^2$ and $n \in N_{\widehat{\mathcal{D}}}(z)$, we have

$$\mathscr{F}_B(\zeta + n) \geqslant \mathscr{F}_B(\zeta). \tag{6.15}$$

Proof. The point in (6.13) is $(t - h, q_1 + hr, q_2 + hp)$. We have $t - h \in [0, 1]$ and $q_1 + hr \in \mathcal{Q}_2$, while the path $q_2 + hp$ belongs to $\mathcal{Q}_\infty$ and satisfies $q_2 + hp \leqslant (1 - t) + h = 1 - (t - h)$. This proves (6.13). Let now $n = (n_t, n_1, n_2) \in N_{\widehat{\mathcal{D}}}(z)$. Using the point in (6.13) with $h = t$ as the point z' in (6.12), and dividing by $t > 0$, we obtain that $-n_t + \langle n_1, r \rangle_{L^2} + \langle n_2, p \rangle_{L^2} \leqslant 0$. Taking the supremum over $r \in K_B$ and $p \in K_1$ yields (6.14). Finally, for every $\zeta = (\tau, a, b) \in \mathbb{R} \times \mathcal{H}^2$, the inequalities (6.6) and (6.14) give

$$\mathscr{F}_B(\zeta + n) - \mathscr{F}_B(\zeta) = n_t - (\mathsf{H}_B(a + n_1, b + n_2) - \mathsf{H}_B(a, b)) \geqslant n_t - B\,\omega(n_1) - \omega(n_2) \geqslant 0,$$

which is (6.15). □

The vector $(-1, r, p)$ in (6.13) is the direction of the characteristic line of the equation (6.9), followed backward in time, when the derivatives of the solution with respect to q_1 and q_2 are p and r respectively. The property (6.13) is thus the precise form of the observation made at the beginning of this section, and the inequality (6.15) is its counterpart for test functions. We will use (6.15) in the proof of Theorem 6.7, to verify both viscosity inequalities at boundary points.

6.2. **Comparison and stability.** Since the viscosity inequalities of Definition 6.2 are imposed at every point of V, the comparison principle can be proved by doubling the variables, as for equations posed on the whole space, and the boundary requires no specific treatment. We follow the proof of [14, Proposition 3.8], with one modification that we point out in Remark 6.5. To compensate for the lack of local compactness of $\mathcal{H}$, we use Stegall's variational principle [47], in the form stated in [14, Theorem 3.7]: if $\mathcal{E}$ is a nonempty, closed, bounded and convex subset of a separable Hilbert space $\mathcal{X}$, and if $J : \mathcal{E} \to \mathbb{R}$ is upper semicontinuous and bounded from above, then for every $\rho > 0$, there exists $\ell \in \mathcal{X}$ with $|\ell|_{\mathcal{X}} \leqslant \rho$ such that the function $J + \langle \ell, \cdot \rangle_{\mathcal{X}}$ attains its maximum over $\mathcal{E}$. The statement in [14] is for convex and weakly compact sets; a closed, bounded and convex subset of a Hilbert space is weakly compact.

Proposition 6.4 (Comparison principle). *Let $B \geqslant 0$, and let $u, v : \widehat{\mathcal{D}} \to \mathbb{R}$ be Lipschitz continuous functions such that u is a viscosity subsolution and v is a viscosity supersolution of (6.9) with derivative bound B, in the sense of Definition 6.2. If $u(0, q) \leqslant v(0, q)$ for every $q \in \mathcal{Q}_2 \times K_1$, then $u \leqslant v$ on $\widehat{\mathcal{D}}$.*

Proof. Let $L < \infty$ be a Lipschitz constant for both u and v. We argue by contradiction, and assume that $u > v$ at some point of $\widehat{\mathcal{D}}$. Every point of $\widehat{\mathcal{D}}$ of the form $(1, q_1, 0)$ is the limit of the points $(t, q_1, 0)$ as t increases to 1. Since u and v are continuous, and since $u \leqslant v$ at time $t = 0$, there exists $(t_*, q_*) \in V$ such that $u(t_*, q_*) > v(t_*, q_*)$. We fix such a point, and then $\eta, \kappa, c_0 \in (0, 1]$ such that

$$u(t_*, q_*) - v(t_*, q_*) - \eta t_* - \frac{2\kappa}{1 - t_*} \geqslant 3c_0. \tag{6.16}$$

For $\delta, \varepsilon \in (0,1]$ to be chosen below, and for every $(t,q),(s,y) \in \widehat{\mathcal{D}}$ with $t,s<1$, we set

$$J(t,q,s,y) := u(t,q) - v(s,y) - \frac{|t-s|^2+|q-y|^2}{2\varepsilon} - \delta\left(|q|^2+|y|^2\right) - \eta t - \frac{\kappa}{1-t} - \frac{\kappa}{1-s}. \tag{6.17}$$

We also consider linear perturbations of J. For $\rho \in (0,\delta]$ to be chosen below, and for $(\alpha,a,\beta,b) \in (\mathbb{R}\times\mathcal{H}^2)^2$ such that

$$|\alpha|,\ |a|,\ |\beta|,\ |b| \leqslant \rho, \tag{6.18}$$

we set $\ell(t,q,s,y) := \alpha t + \langle a,q\rangle + \beta s + \langle b,y\rangle$. Throughout the proof, we impose that

$$2\delta|q_*|^2 \leqslant c_0 \qquad \text{and} \qquad 2\rho(1+|q_*|) \leqslant c_0, \tag{6.19}$$

so that, by (6.16),

$$J(t_*,q_*,t_*,q_*) \geqslant 2c_0 \qquad \text{and} \qquad (J+\ell)(t_*,q_*,t_*,q_*) \geqslant c_0. \tag{6.20}$$

Step 1: A priori bounds. We set $C_L := L^2+2L+3$, and we show that, for every $(t,q),(s,y) \in \widehat{\mathcal{D}}$ with $t,s<1$ and every ℓ as above, if $(J+\ell)(t,q,s,y) \geqslant 0$, then

$$\frac{|t-s|^2+|q-y|^2}{4\varepsilon} + \frac{\delta}{2}\left(|q|^2+|y|^2\right) + \frac{\kappa}{1-t} + \frac{\kappa}{1-s} \leqslant C_L. \tag{6.21}$$

Since $q_2 \leqslant 1-t \leqslant 1$, the point $(0,q)$ belongs to $\widehat{\mathcal{D}}$. Using that $u(0,q) \leqslant v(0,q)$, and the Lipschitz continuity of u and v, we get

$$u(t,q) - v(s,y) \leqslant (u(t,q)-u(0,q)) + (v(0,q)-v(s,y)) \leqslant L\left(t+s+|q-y|\right). \tag{6.22}$$

Moreover, we have $L|q-y| \leqslant \frac{1}{4\varepsilon}|q-y|^2 + L^2\varepsilon$, and since $\rho \leqslant \delta \leqslant 1$, the bound (6.18) yields

$$|\ell(t,q,s,y)| \leqslant \rho\left(2+|q|+|y|\right) \leqslant 3 + \frac{\delta}{2}\left(|q|^2+|y|^2\right). \tag{6.23}$$

Combining these three estimates with the definition of J, and using that $t,s,\varepsilon \leqslant 1$, we obtain

$$(J+\ell)(t,q,s,y) \leqslant C_L - \frac{|t-s|^2+|q-y|^2}{4\varepsilon} - \frac{\delta}{2}\left(|q|^2+|y|^2\right) - \frac{\kappa}{1-t} - \frac{\kappa}{1-s},$$

and the announced implication follows.

Step 2: Construction of a maximizer. We set $R := 2\sqrt{C_L/\delta}$ and $\theta := \kappa/(2C_L)$, and we let

$$\mathcal{E} := \left\{(t,q,s,y) : (t,q),(s,y) \in \widehat{\mathcal{D}},\ t\vee s \leqslant 1-\theta,\ |q|\vee|y| \leqslant R\right\}.$$

This set is closed, bounded and convex in $(\mathbb{R}\times\mathcal{H}^2)^2$. Since u and v are Lipschitz continuous, the function J is continuous and bounded from above on $\mathcal{E}$. By (6.20) and Step 1 with $\ell=0$, the point (t_*,q_*,t_*,q_*) satisfies (6.21), and it therefore belongs to $\mathcal{E}$. By Stegall's variational principle, there exists (α,a,β,b) satisfying (6.18) such that, with ℓ defined as above, the function $J+\ell$ attains its maximum over $\mathcal{E}$ at some point $(\overline{t},\overline{q},\overline{s},\overline{y}) \in \mathcal{E}$. By (6.20), we have

$$(J+\ell)(\overline{t},\overline{q},\overline{s},\overline{y}) \geqslant c_0. \tag{6.24}$$

Step 1 therefore applies at this point, and (6.21) yields

$$|\overline{t}-\overline{s}| + |\overline{q}-\overline{y}| \leqslant 2\sqrt{2C_L\varepsilon}, \qquad |\overline{q}|+|\overline{y}| \leqslant 2\sqrt{C_L/\delta}, \qquad \text{and} \qquad \overline{t}\vee\overline{s} \leqslant 1 - \frac{\kappa}{C_L}. \tag{6.25}$$

More precisely, (6.21) gives $|\overline{q}|^2+|\overline{y}|^2 \leqslant 2C_L/\delta$, so that $|\overline{q}|\vee|\overline{y}| < R$, and we also have $\overline{t}\vee\overline{s} < 1-\theta$. That is, both inequalities introduced in the definition of $\mathcal{E}$ are strict at the maximizer.

Step 3: Positivity of the contact times. We show that $\overline{t}>0$ and $\overline{s}>0$, provided that

$$2L\sqrt{2C_L\varepsilon} \leqslant \frac{c_0}{3} \qquad \text{and} \qquad 2\rho\left(1+\sqrt{C_L/\delta}\right) \leqslant \frac{c_0}{3}. \tag{6.26}$$

If $\overline{t}=0$, then the ordering at time zero and the Lipschitz continuity of v give

$$u(0,\overline{q}) - v(\overline{s},\overline{y}) \leqslant v(0,\overline{q}) - v(\overline{s},\overline{y}) \leqslant L\left(\overline{s}+|\overline{q}-\overline{y}|\right).$$

If $\overline{s} = 0$, then we similarly have $u(\overline{t},\overline{q}) - v(0,\overline{y}) \leqslant u(\overline{t},\overline{q}) - u(0,\overline{y}) \leqslant L(\overline{t} + |\overline{q} - \overline{y}|)$. In both cases, the definition of J, the first inequality in (6.23), and the bounds (6.25) and (6.26) yield

$$(J+\ell)(\overline{t},\overline{q},\overline{s},\overline{y}) \leqslant L\left(|\overline{t}-\overline{s}| + |\overline{q}-\overline{y}|\right) + \rho\left(2 + |\overline{q}| + |\overline{y}|\right) \leqslant 2L\sqrt{2C_L\varepsilon} + 2\rho\left(1 + \sqrt{C_L/\delta}\right) \leqslant \frac{2c_0}{3},$$

in contradiction with (6.24).

Step 4: Conclusion. We assume that (6.26) holds. By Steps 2 and 3, the points $(\overline{t},\overline{q})$ and $(\overline{s},\overline{y})$ belong to V, and for every $(t,q) \in V$ sufficiently close to $(\overline{t},\overline{q})$, the point $(t,q,\overline{s},\overline{y})$ belongs to $\mathcal{E}$. The function $(t,q) \mapsto (J+\ell)(t,q,\overline{s},\overline{y})$ therefore has a local maximum at $(\overline{t},\overline{q})$ relative to V. Up to an additive constant, this function is $u - \phi_1$, where

$$\phi_1(t,q) := \frac{|t-\overline{s}|^2 + |q-\overline{y}|^2}{2\varepsilon} + \delta|q|^2 + \eta t + \frac{\kappa}{1-t} - \alpha t - \langle a, q\rangle$$

is a smooth function on V. Since u is a viscosity subsolution, we obtain that

$$\frac{\overline{t}-\overline{s}}{\varepsilon} + \eta + \frac{\kappa}{(1-\overline{t})^2} - \alpha - \mathsf{H}_B(P) \leqslant 0, \qquad \text{where} \quad P := \frac{\overline{q}-\overline{y}}{\varepsilon} + 2\delta\overline{q} - a. \tag{6.27}$$

Here and below, for $P = (P_1, P_2) \in \mathcal{H}^2$, we write $\mathsf{H}_B(P) := \mathsf{H}_B(P_1, P_2)$. Similarly, the function $(s,y) \mapsto -(J+\ell)(\overline{t},\overline{q},s,y)$ has a local minimum at $(\overline{s},\overline{y})$ relative to V, and up to an additive constant, it is equal to $v - \phi_2$, where

$$\phi_2(s,y) := -\frac{|\overline{t}-s|^2 + |\overline{q}-y|^2}{2\varepsilon} - \delta|y|^2 - \frac{\kappa}{1-s} + \beta s + \langle b, y\rangle$$

is a smooth function on V. Since v is a viscosity supersolution, we obtain that

$$\frac{\overline{t}-\overline{s}}{\varepsilon} - \frac{\kappa}{(1-\overline{s})^2} + \beta - \mathsf{H}_B(Q) \geqslant 0, \qquad \text{where} \quad Q := \frac{\overline{q}-\overline{y}}{\varepsilon} - 2\delta\overline{y} + b. \tag{6.28}$$

Subtracting (6.28) from (6.27), and using (6.7), (6.18), and (6.25), we get

$$\begin{aligned} \eta &\leqslant \mathsf{H}_B(P) - \mathsf{H}_B(Q) + \alpha + \beta \\ &\leqslant (B+1)\left(2\delta\left(|\overline{q}| + |\overline{y}|\right) + 2\rho\right) + 2\rho \leqslant 4(B+1)\sqrt{C_L\delta} + 2(B+2)\rho. \end{aligned} \tag{6.29}$$

We now choose the parameters. We first choose $\delta \in (0,1]$ such that $2\delta|q_*|^2 \leqslant c_0$ and $4(B+1)\sqrt{C_L\delta} \leqslant \eta/3$. We then choose $\rho \in (0,\delta]$ such that $2\rho(1+|q_*|) \leqslant c_0$, $2\rho(1+\sqrt{C_L/\delta}) \leqslant c_0/3$, and $2(B+2)\rho \leqslant \eta/3$. Finally, we choose $\varepsilon \in (0,1]$ such that $2L\sqrt{2C_L\varepsilon} \leqslant c_0/3$. With these choices, the conditions (6.19) and (6.26) are satisfied, and (6.29) gives $\eta \leqslant 2\eta/3$, which is absurd since $\eta > 0$. □

Remark 6.5. In the proof of [14, Proposition 3.8], the distance between the two contact points is estimated by comparing the value of the doubled function at the maximizer $(\overline{t},\overline{q},\overline{s},\overline{y})$ with its values at $(\overline{t},\overline{q},\overline{s},\overline{q})$ and at $(\overline{t},\overline{q},\overline{t},\overline{y})$. On the shrinking domain, this is not possible in general, since the points $(\overline{s},\overline{q})$ and $(\overline{t},\overline{y})$ need not belong to $\widehat{\mathcal{D}}$. In the proof above, we instead rely on the estimate (6.22), in which the only auxiliary point is $(0,q)$; this point always belongs to $\widehat{\mathcal{D}}$, because the domain shrinks as t increases. The resulting bound on the distance between the contact points in (6.25) is of order $\sqrt{\varepsilon}$ instead of ε, but this is still sufficient to conclude.

Proposition 6.6 (Stability)**.** *Let $B \geqslant 0$, and let $(u_j)_{j\in\mathbb{N}}$ be a sequence of viscosity subsolutions, respectively supersolutions, of (6.9) with derivative bound B. If u_j converges to a function $u : \widehat{\mathcal{D}} \to \mathbb{R}$ uniformly on every bounded subset of $\widehat{\mathcal{D}}$ as j tends to infinity, then u is a viscosity subsolution, respectively supersolution, of (6.9) with derivative bound B.*

Proof. The function u is continuous, since it is a locally uniform limit of continuous functions. We only consider the case of subsolutions, the case of supersolutions being similar. Let $\phi : V \to \mathbb{R}$ be a smooth function and $z = (t,q) \in V$ be such that $u - \phi$ has a local maximum at z. We set $\widetilde{\phi}(z') := \phi(z') + |z'-z|^2$, which defines a smooth function on V such that $D\widetilde{\phi}(z) = D\phi(z)$. We fix $\delta \in (0, t \wedge (1-t))$ sufficiently small that, denoting by $\mathcal{E}$ the intersection of $\widehat{\mathcal{D}}$ with the closed ball of radius δ centered at z, we have

$$(u - \widetilde{\phi})(z') \leqslant (u - \widetilde{\phi})(z) - |z'-z|^2, \qquad \forall z' \in \mathcal{E}. \tag{6.30}$$

The set $\mathcal{E}$ is closed, bounded and convex, and it is contained in V. We set $\varepsilon_j := \sup_{\mathcal{E}} |u_j - u|$, which tends to zero as j tends to infinity. For j sufficiently large, the continuous function $u_j - \widetilde{\phi}$ is bounded from above on $\mathcal{E}$, by (6.30), and Stegall's variational principle provides us with $\ell_j \in \mathbb{R} \times \mathcal{H}^2$ such that $|\ell_j| \leqslant 2^{-j}$ and such that the function $u_j - \widetilde{\phi} + \langle \ell_j, \cdot \rangle$ attains its maximum over $\mathcal{E}$ at some point $z_j \in \mathcal{E}$. Comparing the values of this function at z_j and at z, and using (6.30), we get

$$0 \leqslant (u_j - \widetilde{\phi})(z_j) - (u_j - \widetilde{\phi})(z) + \langle \ell_j, z_j - z \rangle \leqslant 2\varepsilon_j - |z_j - z|^2 + 2^{-j}\delta,$$

so that z_j converges to z as j tends to infinity. In particular, for j sufficiently large, the point z_j lies in the open ball of radius δ centered at z, and the function $u_j - (\widetilde{\phi} - \langle \ell_j, \cdot \rangle)$ therefore has a local maximum at z_j relative to V. Since u_j is a viscosity subsolution, we obtain that

$$\mathscr{F}_B \left(D\widetilde{\phi}(z_j) - \ell_j \right) \leqslant 0.$$

Letting j tend to infinity, and using that $\mathscr{F}_B$ and $D\widetilde{\phi}$ are continuous and that $D\widetilde{\phi}(z) = D\phi(z)$, we conclude that $\mathscr{F}_B(D\phi(z)) \leqslant 0$, as desired. □

6.3. **The variational formula.** We now show that the variational formula of Theorems 1.1 and 4.2 remains valid on the shrinking domain. We state the result in the formulation of Remark 2.2, in which the convex part of the initial condition is described through its convex conjugate. That is, we give ourselves a convex function g on $\mathcal{H}$ whose effective domain $\operatorname{dom} g := \{p \in \mathcal{H} : g(p) < +\infty\}$ is contained in K_1; we also give ourselves a function ψ_2 on K_1; and we consider the initial condition $(q_1, q_2) \mapsto g^*(q_1) + \psi_2(q_2)$. The function ψ_2 is not assumed to be convex. Our only assumption on ψ_2 is the one-sided bound (6.31) below, which is a way to state that the derivative of ψ_2 takes values in K_B without assuming that ψ_2 is differentiable. Indeed, if ψ_2 is the restriction to K_1 of a continuously differentiable function on $\mathcal{H}$ whose gradient takes values in K_B, then (6.31) follows by integrating along the segment from x to y, since $B\,\omega$ is the support function of K_B; conversely, we will see in the proof of Theorem 6.7 that every function satisfying (6.31) is a uniform limit of functions of this form. By (6.2), the condition (6.31) implies that ψ_2 is Lipschitz continuous with respect to the L^1 norm, with Lipschitz constant B. It also implies that $\psi_2(x) \leqslant \psi_2(y)$ whenever $y - x \in \mathcal{Q}_2^*$, since we have $\omega(x - y) = 0$ in this case; compare with the properties of the functions ψ_ρ stated in Lemma 3.2.

The assumption $\operatorname{dom} g \subseteq K_1$ is the counterpart, for the first path, of the assumption on ψ_2 discussed above: since g^* is a supremum of affine functions whose slopes lie in K_1, we have $g^*(y) - g^*(x) \leqslant \omega(y - x)$ for every $x, y \in \mathcal{H}$, which is a way to state that the derivative of g^* takes values in K_1. This bound is essential: as explained at the beginning of this section, it ensures that the characteristic lines move the variable q_2 at speed at most one. In contrast with the bound B on the derivative with respect to q_2, the bound 1 on the derivative with respect to q_1 cannot be relaxed, since it must match the speed at which the domain shrinks.

Theorem 6.7 (Variational formula on the shrinking domain)**.** *Let $B \geqslant 0$, let $g : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ be a proper, lower semicontinuous and convex function such that $\operatorname{dom} g \subseteq K_1$, and let $\psi_2 : K_1 \to \mathbb{R}$ be such that*

$$\psi_2(y) - \psi_2(x) \leqslant B\,\omega(y - x), \qquad \forall x, y \in K_1. \tag{6.31}$$

For every $(t, q_1, q_2) \in \widehat{\mathcal{D}}$, let

$$u(t, q_1, q_2) := \sup_{p \in K_1} \left\{ \langle p, q_1 \rangle_{L^2} + \psi_2(q_2 + tp) - g(p) \right\}. \tag{6.32}$$

The supremum in (6.32) is attained. The function u is the unique Lipschitz viscosity solution of (6.9) with derivative bound B, in the sense of Definition 6.2, such that

$$u(0, q_1, q_2) = g^*(q_1) + \psi_2(q_2), \qquad \forall (q_1, q_2) \in \mathcal{Q}_2 \times K_1, \tag{6.33}$$

where $g^*(q_1) := \sup_{p \in \mathcal{H}} \{\langle p, q_1 \rangle_{L^2} - g(p)\}$. Moreover, for every $(t, q), (s, y) \in \widehat{\mathcal{D}}$, we have

$$|u(t, q) - u(s, y)| \leqslant |q_1 - y_1|_{L^1} + B|q_2 - y_2|_{L^1} + B|t - s|. \tag{6.34}$$

The right-hand side of (6.32) does not involve B. Theorem 6.7 thus shows in particular that the viscosity solution with derivative bound B and initial condition (6.33) does not depend on the choice of B, as long as (6.31) holds.

Proof of Theorem 6.7. For every $z = (t, q_1, q_2) \in \widehat{\mathcal{D}}$ and $p \in K_1$, the path $q_2 + tp$ belongs to $\mathcal{Q}_\infty$ and is bounded by $(1-t) + t = 1$. It thus belongs to K_1, and the right-hand side of (6.32) is well-defined. For every $p \in \operatorname{dom} g$ and $z = (t, q_1, q_2) \in \widehat{\mathcal{D}}$, we set

$$w_p(z) := \langle p, q_1 \rangle_{L^2} + \psi_2(q_2 + tp) - g(p), \tag{6.35}$$

so that $u = \sup_{p \in \operatorname{dom} g} w_p$. The structure of the proof is the same as that of the proof of Theorem 1.1, with the following differences: the space is infinite-dimensional, which we handle using the compactness of K_1; the contact points can be on the boundary of the domain, which we handle using Lemma 6.3; and the approximation of ψ_2 by smooth functions must preserve the condition (6.31).

Step 1: Attainment, Lipschitz continuity, and the initial condition. By (6.31) and (6.2), we have

$$|\psi_2(y) - \psi_2(x)| \leqslant B|y - x|_{L^1}, \qquad \forall x, y \in K_1. \tag{6.36}$$

We fix $z = (t, q_1, q_2) \in \widehat{\mathcal{D}}$. By (6.36), the function $p \mapsto \langle p, q_1 \rangle_{L^2} + \psi_2(q_2 + tp)$ is continuous on K_1 for the topology of $\mathcal{H}$. Since g is lower semicontinuous and is not identically $+\infty$ on K_1, and since K_1 is compact, we deduce that $u(z)$ is finite, that the supremum in (6.32) is attained, and that the set of maximizers

$$\mathcal{A}(z) := \{p \in \operatorname{dom} g : w_p(z) = u(z)\}$$

is nonempty and compact. For every $p \in K_1$ and $(t, q), (s, y) \in \widehat{\mathcal{D}}$, we have $|\langle p, q_1 - y_1 \rangle_{L^2}| \leqslant |q_1 - y_1|_{L^1}$, and by (6.36),

$$|\psi_2(q_2 + tp) - \psi_2(y_2 + sp)| \leqslant B|q_2 - y_2|_{L^1} + B|t - s|.$$

Each of the functions w_p therefore satisfies the estimate (6.34), and taking the supremum over p yields (6.34) for u. In particular, the function u is Lipschitz continuous. Finally, since $\operatorname{dom} g \subseteq K_1$, the identity (6.33) follows from (6.32) at $t = 0$.

Step 2: The subsolution inequality when ψ_2 is smooth. In this step and the next one, we assume in addition that ψ_2 is the restriction to K_1 of a continuously differentiable function on $\mathcal{H}$, which we still denote by ψ_2, and that $\nabla\psi_2(x) \in K_B$ for every $x \in \mathcal{H}$. For every $p \in \operatorname{dom} g$, the formula (6.35) then defines a continuously differentiable function w_p on $\mathbb{R} \times \mathcal{H}^2$, whose differential at $z' = (t', q_1', q_2')$ is

$$Dw_p(z') = \left(\langle p, \nabla\psi_2(q_2' + t'p)\rangle_{L^2}, p, \nabla\psi_2(q_2' + t'p)\right) \in \mathbb{R} \times \mathcal{H}^2, \tag{6.37}$$

and the right-hand side of (6.37) depends continuously on $(p, z') \in K_1 \times \mathbb{R} \times \mathcal{H}^2$. We fix $z = (t, q_1, q_2) \in V$, and for every $p \in \mathcal{A}(z)$, we set

$$b_p := \nabla\psi_2(q_2 + tp) \in K_B \qquad \text{and} \qquad \gamma_p := Dw_p(z) = \left(\langle p, b_p \rangle_{L^2}, p, b_p\right).$$

Since $p \in K_1$ and $b_p \in K_B$, the identity (6.5) gives

$$\mathscr{F}_B(\gamma_p) = \langle p, b_p \rangle_{L^2} - \mathsf{H}_B(p, b_p) = 0, \qquad \forall p \in \mathcal{A}(z). \tag{6.38}$$

Let now $\phi : V \to \mathbb{R}$ be a smooth function such that $u - \phi$ has a local maximum at z, and let $p \in \mathcal{A}(z)$. Since $w_p \leqslant u$ and $w_p(z) = u(z)$, the function $w_p - \phi$ also has a local maximum at z relative to V. Let $z' \in \widehat{\mathcal{D}}$. Since $\widehat{\mathcal{D}}$ is convex and $t \in (0, 1)$, the point $z + \varepsilon(z' - z)$ belongs to V for every $\varepsilon \in [0, 1)$, and thus $(w_p - \phi)(z + \varepsilon(z' - z)) \leqslant (w_p - \phi)(z)$ for every $\varepsilon > 0$ sufficiently small. Dividing by ε and letting ε tend to zero, we obtain that $\langle \gamma_p - D\phi(z), z' - z \rangle \leqslant 0$. Since $z' \in \widehat{\mathcal{D}}$ is arbitrary, this means that $\gamma_p - D\phi(z) \in N_{\widehat{\mathcal{D}}}(z)$. By (6.15) and (6.38), we conclude that

$$\mathscr{F}_B(D\phi(z)) \leqslant \mathscr{F}_B(\gamma_p) = 0,$$

which is the subsolution inequality.

Step 3: The supersolution inequality when ψ_2 is smooth. We keep the additional assumption and the notation of Step 2. We first show that

$$\langle p - p', b_p - b_{p'} \rangle_{L^2} \geqslant 0, \qquad \forall p, p' \in \mathcal{A}(z). \tag{6.39}$$

Let $p \in \mathcal{A}(z)$ and $p' \in \operatorname{dom} g$. For every $\varepsilon \in (0,1]$, the point $p_\varepsilon := p + \varepsilon(p' - p)$ belongs to $\operatorname{dom} g$, by convexity, and $w_{p_\varepsilon}(z) \leqslant u(z) = w_p(z)$. Together with the convexity inequality $g(p_\varepsilon) \leqslant g(p) + \varepsilon(g(p') - g(p))$, this gives

$$\varepsilon \langle p' - p, q_1 \rangle_{L^2} + \psi_2\big(q_2 + tp + \varepsilon t(p' - p)\big) - \psi_2(q_2 + tp) \leqslant g(p_\varepsilon) - g(p) \leqslant \varepsilon\big(g(p') - g(p)\big).$$

Dividing by ε and letting ε tend to zero, we obtain the analogue of (2.6), namely

$$g(p') - g(p) \geqslant \langle q_1 + tb_p, p' - p \rangle_{L^2}. \tag{6.40}$$

For $p, p' \in \mathcal{A}(z)$, we add (6.40) to the same inequality with p and p' interchanged. Since $t > 0$, this yields (6.39).

We set $\Gamma_z := \{\gamma_p : p \in \mathcal{A}(z)\}$. This set is compact, since it is the image of the compact set $\mathcal{A}(z)$ under a continuous mapping. Its closed convex hull $\overline{\operatorname{co}}\,\Gamma_z$ is also compact. Indeed, it is closed, and it is totally bounded: for every $\varepsilon > 0$, we can cover Γ_z by finitely many balls of radius ε, and then every point of the convex hull of Γ_z is within distance ε of the convex hull of the centers of these balls, which is compact. We next show that

$$\mathscr{F}_B(\overline{\gamma}) \geqslant 0, \qquad \forall \overline{\gamma} \in \overline{\operatorname{co}}\,\Gamma_z. \tag{6.41}$$

Since $\mathscr{F}_B$ is continuous, it suffices to consider the case when $\overline{\gamma} = \sum_{i=1}^m \alpha_i \gamma_{p_i}$ for some $m \in \mathbb{N}$, $p_1, \dots, p_m \in \mathcal{A}(z)$, and $\alpha_1, \dots, \alpha_m \geqslant 0$ with $\sum_{i=1}^m \alpha_i = 1$. Since K_1 and K_B are convex, we have $\sum_{i=1}^m \alpha_i p_i \in K_1$ and $\sum_{i=1}^m \alpha_i b_{p_i} \in K_B$, and the identity (6.5) thus gives

$$\mathscr{F}_B(\overline{\gamma}) = \sum_{i=1}^m \alpha_i \langle p_i, b_{p_i} \rangle_{L^2} - \Big\langle \sum_{i=1}^m \alpha_i p_i, \sum_{j=1}^m \alpha_j b_{p_j} \Big\rangle_{L^2} = \frac{1}{2} \sum_{i,j=1}^m \alpha_i \alpha_j \langle p_i - p_j, b_{p_i} - b_{p_j} \rangle_{L^2} \overset{(6.39)}{\geqslant} 0.$$

Let now $\phi : V \to \mathbb{R}$ be a smooth function such that $u - \phi$ has a local minimum at z. We will show that

$$D\phi(z) \in \overline{\operatorname{co}}\,\Gamma_z + N_{\widehat{\mathcal{D}}}(z). \tag{6.42}$$

This implies the supersolution inequality. Indeed, if $D\phi(z) = \overline{\gamma} + n$ with $\overline{\gamma} \in \overline{\operatorname{co}}\,\Gamma_z$ and $n \in N_{\widehat{\mathcal{D}}}(z)$, then (6.15) and (6.41) give

$$\mathscr{F}_B(D\phi(z)) \geqslant \mathscr{F}_B(\overline{\gamma}) \geqslant 0.$$

As a first step towards (6.42), we show that

$$\langle D\phi(z), z' - z \rangle \leqslant \max_{\gamma \in \Gamma_z} \langle \gamma, z' - z \rangle, \qquad \forall z' \in \widehat{\mathcal{D}}. \tag{6.43}$$

We fix $z' \in \widehat{\mathcal{D}}$ and write $v := z' - z$. For every $h \in (0,1)$, the point $z + hv$ belongs to V, and we pick $p_h \in \mathcal{A}(z + hv)$. Since $u(z + hv) = w_{p_h}(z + hv)$ and $u(z) \geqslant w_{p_h}(z)$, the local minimum of $u - \phi$ at z and the mean value theorem give, for every $h > 0$ sufficiently small,

$$\phi(z + hv) - \phi(z) \leqslant u(z + hv) - u(z) \leqslant w_{p_h}(z + hv) - w_{p_h}(z) = h \langle Dw_{p_h}(z + \theta_h hv), v \rangle \tag{6.44}$$

for some $\theta_h \in (0,1)$. Since K_1 is compact, there is a sequence $h_k \downarrow 0$ such that p_{h_k} converges in $\mathcal{H}$ to some $p \in K_1$. We verify that $p \in \mathcal{A}(z)$. Writing $z + h_k v = (t_k, q_{1,k}, q_{2,k})$, the continuity of ψ_2 on K_1 and of u gives

$$g(p_{h_k}) = \langle p_{h_k}, q_{1,k} \rangle_{L^2} + \psi_2(q_{2,k} + t_k p_{h_k}) - u(z + h_k v) \xrightarrow[k \to \infty]{} \langle p, q_1 \rangle_{L^2} + \psi_2(q_2 + tp) - u(z),$$

and the lower semicontinuity of g yields $g(p) \leqslant \langle p, q_1 \rangle_{L^2} + \psi_2(q_2 + tp) - u(z)$. That is, we have $p \in \operatorname{dom} g$ and $w_p(z) \geqslant u(z)$, and thus $p \in \mathcal{A}(z)$. By the continuity of the right-hand side of (6.37), the vector $Dw_{p_{h_k}}(z + \theta_{h_k} h_k v)$ converges to γ_p as k tends to infinity. Dividing (6.44) by h_k and letting k tend to infinity, we obtain that $\langle D\phi(z), v \rangle \leqslant \langle \gamma_p, v \rangle$, which proves (6.43).

We now derive (6.42) from (6.43) by a separation argument. The set $\overline{\operatorname{co}}\,\Gamma_z + N_{\widehat{\mathcal{D}}}(z)$ is convex, and it is closed, as the sum of a compact set and a closed set. If $D\phi(z)$ did not belong to this set, then there would exist $v \in \mathbb{R} \times \mathcal{H}^2$ such that

$$\langle D\phi(z), v \rangle > \sup\big\{\langle \gamma + n, v \rangle : \gamma \in \overline{\operatorname{co}}\,\Gamma_z,\ n \in N_{\widehat{\mathcal{D}}}(z)\big\}.$$

Since $N_{\widehat{\mathcal{D}}}(z)$ is a cone, this would imply that

$$\langle n, v\rangle \leqslant 0 \quad \text{for every } n \in N_{\widehat{\mathcal{D}}}(z), \qquad \text{and} \qquad \langle D\phi(z), v\rangle > \max_{\gamma\in\Gamma_z} \langle \gamma, v\rangle\,. \tag{6.45}$$

Let $\mathcal{T}_z := \{h(z'-z) : z' \in \widehat{\mathcal{D}},\ h \geqslant 0\}$, which is a convex cone. By (6.43) and by homogeneity, we have $\langle D\phi(z), w\rangle \leqslant \max_{\gamma\in\Gamma_z}\langle\gamma, w\rangle$ for every $w \in \mathcal{T}_z$. Both sides of this inequality depend continuously on w, so the inequality remains valid for every w in the closure of $\mathcal{T}_z$. By the second part of (6.45), the vector v thus does not belong to the closure of $\mathcal{T}_z$. Since this closure is a closed convex cone, there would then exist $n \in \mathbb{R}\times\mathcal{H}^2$ such that $\langle n, v\rangle > 0$ and $\langle n, w\rangle \leqslant 0$ for every $w \in \mathcal{T}_z$. The latter property means that $n \in N_{\widehat{\mathcal{D}}}(z)$, and we have reached a contradiction with the first part of (6.45). This completes the proof of (6.42), and thus of the supersolution inequality.

Step 4: Conclusion. We now remove the additional assumption on ψ_2 made in Steps 2 and 3. Let $\psi_2 : K_1 \to \mathbb{R}$ be a function satisfying (6.31). We first extend it to $\mathcal{H}$, by setting

$$\widetilde{\psi}_2(x) := \inf_{y\in K_1} \{\psi_2(y) + B\,\omega(x-y)\}\,, \qquad x \in \mathcal{H}. \tag{6.46}$$

We fix $y_0 \in K_1$. For every $x \in \mathcal{H}$ and $y \in K_1$, the assumption (6.31) and the subadditivity of ω give

$$\psi_2(y) + B\,\omega(x-y) \geqslant \psi_2(y_0) - B\,\omega(y_0-y) + B\,\omega(x-y) \geqslant \psi_2(y_0) - B\,\omega(y_0-x),$$

so that $\widetilde{\psi}_2(x)$ is finite. Using (6.31) once more, we see that the infimum in (6.46) is attained at $y = x$ when $x \in K_1$, and thus that $\widetilde{\psi}_2 = \psi_2$ on K_1. The subadditivity of ω also implies that

$$\widetilde{\psi}_2(x+x') - \widetilde{\psi}_2(x) \leqslant B\,\omega(x'), \qquad \forall x, x' \in \mathcal{H}. \tag{6.47}$$

In particular, by (6.2), the function $\widetilde{\psi}_2$ is Lipschitz continuous on $\mathcal{H}$, with Lipschitz constant B.

We next regularize $\widetilde{\psi}_2$, using the mappings ι_j and π_j defined in (4.2). We recall that $\langle \iota_j y, x\rangle_{L^2} = \langle y, \pi_j x\rangle_j$ for every $y \in E_j$ and $x \in \mathcal{H}$, and we denote by $|\cdot|_j$ the norm associated with the scalar product $\langle\cdot,\cdot\rangle_j$, so that $|\iota_j y|_{L^2} = |y|_j$. For every $j \in \mathbb{N}$, let χ_j be a smooth probability density on E_j supported in $\{y \in E_j : |y|_j \leqslant 2^{-j}\}$, and let

$$\psi_{2,j}(x) := \int_{E_j} \widetilde{\psi}_2\big(\iota_j(\pi_j x - y)\big)\chi_j(y)\,\mathrm{d}y, \qquad x \in \mathcal{H}. \tag{6.48}$$

The function $\psi_{2,j}$ is the composition of the continuous linear mapping π_j with the convolution of the continuous function $\widetilde{\psi}_2 \circ \iota_j$ and the smooth density χ_j. It is therefore continuously differentiable on $\mathcal{H}$. We now show that its gradient takes values in K_B. The mapping $\iota_j\pi_j$ is self-adjoint, and it maps K_1 into itself, since averaging a path over dyadic intervals preserves its monotonicity and the bounds that it satisfies. For every $x' \in \mathcal{H}$, we thus have

$$\omega(\iota_j\pi_j x') = \sup_{p\in K_1} \langle \iota_j\pi_j p, x'\rangle_{L^2} \leqslant \omega(x').$$

Combining this with (6.47) and (6.48), we obtain that

$$\psi_{2,j}(x+x') - \psi_{2,j}(x) \leqslant B\,\omega(x'), \qquad \forall x, x' \in \mathcal{H}. \tag{6.49}$$

Replacing x' by hx' in (6.49), with $h > 0$, dividing by h, and letting h tend to zero, we get

$$\langle \nabla\psi_{2,j}(x), x'\rangle_{L^2} \leqslant B\,\omega(x') = \sup_{r\in K_B} \langle r, x'\rangle_{L^2}\,, \qquad \forall x, x' \in \mathcal{H}.$$

Since K_B is closed and convex, the separating hyperplane theorem then ensures that $\nabla\psi_{2,j}(x) \in K_B$ for every $x \in \mathcal{H}$. The function $\psi_{2,j}$ thus satisfies the additional assumption of Steps 2 and 3, and by (6.49), its restriction to K_1 satisfies (6.31). Moreover, since $\widetilde{\psi}_2$ is B-Lipschitz and coincides with ψ_2 on K_1, we have

$$\sup_{x\in K_1} |\psi_{2,j}(x) - \psi_2(x)| \leqslant B \sup_{x\in K_1} |\iota_j\pi_j x - x|_{L^2} + 2^{-j}B \xrightarrow[j\to\infty]{} 0. \tag{6.50}$$

The convergence to zero follows from (4.3) and the compactness of K_1, since the mappings $\iota_j\pi_j$ are contractions of $\mathcal{H}$.

For every $j \in \mathbb{N}$, let $u_j : \widehat{\mathcal{D}} \to \mathbb{R}$ be defined by (6.32) with $\psi_{2,j}$ in place of ψ_2. By Steps 1 to 3, the function u_j is a viscosity solution of (6.9) with derivative bound B. Since $q_2 + tp \in K_1$ for every $(t, q_1, q_2) \in \widehat{\mathcal{D}}$ and $p \in K_1$, comparing (6.32) for ψ_2 and for $\psi_{2,j}$ gives

$$\sup_{\widehat{\mathcal{D}}} |u_j - u| \leqslant \sup_{K_1} |\psi_{2,j} - \psi_2|,$$

which tends to zero as j tends to infinity, by (6.50). By Proposition 6.6, the function u is therefore a viscosity solution of (6.9) with derivative bound B. By Step 1, it is Lipschitz continuous and satisfies (6.33) and (6.34). Finally, the uniqueness of Lipschitz viscosity solutions with a given initial condition follows from Proposition 6.4. □

6.4. **Additional unconstrained variables.** In Section 7.4, we will use a version of Theorem 6.7 in which each of the two groups of variables contains an additional real variable that is not subject to any constraint. We now state this result precisely and explain how to adapt the arguments of this section. The two groups of variables are (q_1, x_1) and (q_2, x_2), with $(t, q_1, q_2) \in \widehat{\mathcal{D}}$ and $x_1, x_2 \in \mathbb{R}$. We set $E := \mathcal{H} \times \mathbb{R}$, and we denote by $\langle \cdot, \cdot \rangle_{L^2 \times \mathbb{R}}$ the scalar product on this space, that is, $\langle (a, \alpha), (b, \beta) \rangle_{L^2 \times \mathbb{R}} := \langle a, b \rangle_{L^2} + \alpha\beta$. Writing $\partial_{(q_i, x_i)} u := (\partial_{q_i} u, \partial_{x_i} u)$, we consider the equation

$$\partial_t u - \left\langle \partial_{(q_1, x_1)} u, \partial_{(q_2, x_2)} u \right\rangle_{L^2 \times \mathbb{R}} = 0. \tag{6.51}$$

We write (t, q, x) for a generic point of $\widehat{\mathcal{D}}^{\mathrm{ext}} := \widehat{\mathcal{D}} \times \mathbb{R}^2$, with $q = (q_1, q_2)$ and $x = (x_1, x_2)$, and we identify $\widehat{\mathcal{D}}^{\mathrm{ext}}$ with a closed convex subset of the Hilbert space $\mathbb{R} \times E^2$ through the mapping $(t, q, x) \mapsto (t, (q_1, x_1), (q_2, x_2))$. We also set $V^{\mathrm{ext}} := V \times \mathbb{R}^2$. We fix $B, R_1, R_2 \geqslant 0$, and we replace the sets K_1 and K_B by the compact and convex subsets of E defined by

$$K_1^{\mathrm{ext}} := K_1 \times [-R_1, R_1] \qquad \text{and} \qquad K_B^{\mathrm{ext}} := K_B \times [-R_2, R_2]. \tag{6.52}$$

The support functions of these two sets are, for every $(y, \xi) \in E$,

$$\omega_1^{\mathrm{ext}}(y, \xi) := \omega(y) + R_1|\xi| \qquad \text{and} \qquad \omega_B^{\mathrm{ext}}(y, \xi) := B\,\omega(y) + R_2|\xi|. \tag{6.53}$$

We define $\mathsf{H}^{\mathrm{ext}} : E^2 \to \mathbb{R}$ as in (6.4), that is,

$$\mathsf{H}^{\mathrm{ext}}(a, b) := \min_{p \in K_1^{\mathrm{ext}}} \max_{r \in K_B^{\mathrm{ext}}} \left\{ \langle a, r \rangle_{L^2 \times \mathbb{R}} + \langle p, b \rangle_{L^2 \times \mathbb{R}} - \langle p, r \rangle_{L^2 \times \mathbb{R}} \right\}. \tag{6.54}$$

In analogy with Definition 6.2, we say that a continuous function $u : \widehat{\mathcal{D}}^{\mathrm{ext}} \to \mathbb{R}$ is a viscosity subsolution, respectively supersolution, of (6.51) with derivative bounds (B, R_1, R_2) if its restriction to V^{ext} is a viscosity subsolution, respectively supersolution, of

$$\partial_t u - \mathsf{H}^{\mathrm{ext}}\left(\partial_{(q_1, x_1)} u, \partial_{(q_2, x_2)} u\right) = 0$$

in the sense of Definition 2.1, with ambient Hilbert space E^2. A viscosity solution is a function that is both a viscosity subsolution and a viscosity supersolution.

Proposition 6.8 (Variational formula with additional variables)**.** *Let $B, R_1, R_2 \geqslant 0$, and let $g : E \to \mathbb{R} \cup \{+\infty\}$ be a proper, lower semicontinuous and convex function such that* $\operatorname{dom} g \subseteq K_1^{\mathrm{ext}}$. *Let $\psi_2 : K_1 \times \mathbb{R} \to \mathbb{R}$ be such that, for every $(y, \xi), (y', \xi') \in K_1 \times \mathbb{R}$,*

$$\psi_2(y', \xi') - \psi_2(y, \xi) \leqslant B\,\omega(y' - y) + R_2|\xi' - \xi|. \tag{6.55}$$

For every $(t, q, x) \in \widehat{\mathcal{D}}^{\mathrm{ext}}$, let

$$u(t, q, x) := \sup_{(p, m) \in K_1^{\mathrm{ext}}} \left\{ \langle p, q_1 \rangle_{L^2} + m x_1 + \psi_2(q_2 + tp, x_2 + tm) - g(p, m) \right\}. \tag{6.56}$$

The supremum in (6.56) is attained. The function u is the unique Lipschitz viscosity solution of (6.51) with derivative bounds (B, R_1, R_2) such that, for every $(q, x) \in \mathcal{Q}_2 \times K_1 \times \mathbb{R}^2$,

$$u(0, q, x) = g^*(q_1, x_1) + \psi_2(q_2, x_2), \tag{6.57}$$

where $g^*(q_1, x_1) := \sup_{(p,m) \in E} \{ \langle p, q_1 \rangle_{L^2} + m x_1 - g(p, m) \}$. Moreover, for every (t, q, x) and (t', q', x') in $\widehat{\mathcal{D}}^{\mathrm{ext}}$, we have

$$\begin{aligned} |u(t, q, x) - u(t', q', x')| \leqslant{}& |q_1 - q_1'|_{L^1} + R_1|x_1 - x_1'| \\ &+ B|q_2 - q_2'|_{L^1} + R_2|x_2 - x_2'| + (B + R_1 R_2)|t - t'|. \end{aligned} \tag{6.58}$$

Proof. The proof is the same as that of Theorem 6.7, including the preliminary results on which it relies, up to the following modifications. Throughout, the space E replaces $\mathcal{H}$, the sets K_1^{ext} and K_B^{ext} replace K_1 and K_B, and the functions ω_1^{ext} and ω_B^{ext} replace ω and $B\,\omega$.

The proof of Lemma 6.1 only uses that K_1 and K_B are compact and convex, and that ω and $B\,\omega$ are their support functions. The function $\mathsf{H}^{\mathrm{ext}}$ therefore coincides with $\langle\cdot,\cdot\rangle_{L^2\times\mathbb{R}}$ on $K_1^{\mathrm{ext}}\times K_B^{\mathrm{ext}}$, and for every $a,b,a',b'\in E$, it satisfies

$$\mathsf{H}^{\mathrm{ext}}(a+a',b+b')-\mathsf{H}^{\mathrm{ext}}(a,b)\leqslant\omega_B^{\mathrm{ext}}(a')+\omega_1^{\mathrm{ext}}(b').$$

Since the sets K_1^{ext} and K_B^{ext} are bounded, their support functions are bounded by a constant multiple of the norm of E, and the function $\mathsf{H}^{\mathrm{ext}}$ is thus Lipschitz continuous. Concerning Lemma 6.3, since the variables x_1 and x_2 are not constrained, for every $(t,q,x)\in V^{\mathrm{ext}}$, $(p,m)\in K_1^{\mathrm{ext}}$, $(r,m')\in K_B^{\mathrm{ext}}$, and $h\in[0,t]$, we have

$$(t-h,q_1+hr,q_2+hp,x_1+hm',x_2+hm)\in\widehat{\mathcal{D}}^{\mathrm{ext}}.$$

The proof of Lemma 6.3 thus applies without change, and it shows that the analogue of (6.15) holds for the function $(\tau,a,b)\mapsto\tau-\mathsf{H}^{\mathrm{ext}}(a,b)$ and the normal cones to $\widehat{\mathcal{D}}^{\mathrm{ext}}$ at the points of V^{ext}. The proofs of Propositions 6.4 and 6.6 only use that the domain is closed and convex, that the time variable ranges in $[0,1]$, that the point obtained by replacing t by 0 in any given point of the domain also belongs to the domain, and that the nonlinearity is Lipschitz continuous. They therefore also apply in the present setting, provided that the norms of $\mathcal{H}^2$ and $\mathbb{R}\times\mathcal{H}^2$ are replaced throughout by those of E^2 and $\mathbb{R}\times E^2$, so that the quadratic terms in (6.17) also involve the variables x, and that the factor $B+1$ in (6.29) is replaced by a Lipschitz constant of $\mathsf{H}^{\mathrm{ext}}$.

We now turn to the proof of Theorem 6.7 itself. In Step 1, the estimate (6.36) is replaced by the fact that, by (6.55) and (6.2), the function ψ_2 is Lipschitz continuous on $K_1\times\mathbb{R}$, with Lipschitz constants B in the first variable, with respect to the L^1 norm, and R_2 in the second variable. Using also the compactness of K_1^{ext}, we obtain that the supremum in (6.56) is attained, and that (6.57) and (6.58) hold; the last term in (6.58) comes from the fact that $|m|\leqslant R_1$ for every $(p,m)\in K_1^{\mathrm{ext}}$. Steps 2 and 3 are unchanged, under the additional assumption that ψ_2 is the restriction to $K_1\times\mathbb{R}$ of a continuously differentiable function on E whose gradient takes values in K_B^{ext}. In Step 4, we extend ψ_2 to E by setting, for every $(y,\xi)\in E$,

$$\widetilde{\psi}_2(y,\xi):=\inf_{(y',\xi')\in K_1\times\mathbb{R}}\left\{\psi_2(y',\xi')+\omega_B^{\mathrm{ext}}(y-y',\xi-\xi')\right\}.$$

The argument given below (6.46) shows that this function is finite, that it coincides with ψ_2 on $K_1\times\mathbb{R}$, and that its increments are bounded from above by ω_B^{ext}, as in (6.47). We then set, for every $j\in\mathbb{N}$ and $(y,\xi)\in E$,

$$\psi_{2,j}(y,\xi):=\int_{E_j\times\mathbb{R}}\widetilde{\psi}_2\big(\iota_j(\pi_j y-y'),\xi-\xi'\big)\chi_j(y',\xi')\,\mathrm{d}y'\,\mathrm{d}\xi',$$

where χ_j is now a smooth probability density on $E_j\times\mathbb{R}$ supported in $\{(y',\xi'):|y'|_j^2+|\xi'|^2\leqslant 4^{-j}\}$. As in Step 4, the function $\psi_{2,j}$ is continuously differentiable on E, its gradient takes values in K_B^{ext}, and

$$\sup_{(y,\xi)\in K_1\times\mathbb{R}}|\psi_{2,j}(y,\xi)-\psi_2(y,\xi)|\leqslant B\sup_{y\in K_1}|\iota_j\pi_j y-y|_{L^2}+2^{-j}(B+R_2)\xrightarrow[j\to\infty]{}0.$$

The functions defined by (6.56) with $\psi_{2,j}$ in place of ψ_2 therefore converge to u uniformly on $\widehat{\mathcal{D}}^{\mathrm{ext}}$, and we conclude as before. □

7. Towards general perceptron models

We now return to the general perceptron models discussed in Section 1.2, and explain how the approach of this paper is expected to extend to them. We consider five models in turn: the Ising perceptron with a general activation function, as in [39], in Section 7.1; the spherical perceptron, as in [3, 40], in Section 7.2; Ising and spherical perceptrons with i.i.d. random activation functions, as in [20], in Section 7.3; an Ising perceptron in which the patterns carry labels and are sampled from a mixture of two Gaussian distributions, in Section 7.4; and the same model with spherical

spins, in Section 7.5. In the last two models, the patterns are no longer standard Gaussian vectors: each pattern is the sum of a standard Gaussian vector and of a random shift, whose direction is the same for all the patterns, and whose sign is given by the label of the pattern.

For each of the first four models, the discussion follows the same four steps. First, we define an enriched free energy, which depends on a time variable t and on a pair of paths $q = (q_1, q_2)$ ranging in the shrinking domain of Section 6, and which reduces to the free energy of interest at $t = 1$ and $q = 0$. Second, a computation based on Gaussian integration by parts shows that the enriched free energy satisfies the Hamilton–Jacobi equation (6.9), up to an error term that is small as soon as the overlap between two replicas concentrates conditionally on the overlap of the cascade. Third, we compute the enriched free energy at $t = 0$, and observe that it is the sum of a convex function of q_1, which only depends on the reference measure of the spins, and of a function of q_2, which only depends on the activation function, and which is not convex in general. Fourth, since this is exactly the structure required by Theorem 6.7, this theorem identifies the unique viscosity solution to the equation with this initial condition, and represents it by a variational formula. This solution is our candidate for the limit of the enriched free energy, and its value at $t = 1$ and $q = 0$ is our candidate for the limit free energy of the model. For the fourth model, the initial condition does not have the required convexity property, and an additional step is needed, in which the spin configurations are first sorted according to their magnetization. The fifth model is treated differently: sorting the spin configurations according to their projection onto the direction of the shifts, and using the rotational invariance of the sphere, we reduce it to the spherical perceptron with random activation functions of Section 7.3.

The status of these steps is as follows. For the first three models, the third and fourth steps are rigorous: the results of Section 6 provide the comparison principle and the variational formula for the equation on the shrinking domain, and we verify below that the initial conditions satisfy their assumptions. On the other hand, we do not prove that the enriched free energy converges to the candidate limit. This would require to show analogues of Propositions 5.1 and 5.2; in particular, the cavity arguments underlying the proof of Proposition 5.2 would need to be adjusted. Under various assumptions on the activation functions, the limit free energies of the Ising and spherical perceptrons were identified in [39, 3, 40], while the models with random activation functions considered in Section 7.3 were covered in [20]. Whenever the results of [3, 20, 39, 40] apply, the candidate formulas that we obtain coincide with the formulas proved there. For the fifth model, the reduction to the model of Section 7.3 is rigorous, so that the candidate formula for this model is a consequence of the one obtained there for the spherical perceptron with random activation functions.

Throughout this section, the parameter λ, the sequence $(M_N)_{N\in\mathbb{N}}$, the Gaussian random variables $(g_{ij})_{i,j\in\mathbb{N}}$ and the cascade $\mathfrak{R}$ are as in Section 1.3, and we use the notation of Section 6: we write $\mathcal{H} = L^2([0,1))$ and $K_c = \mathcal{Q}_{\infty,\leqslant c}$, the function ω is the support function of K_1 defined in (6.1), and the domains $\mathcal{D} \subseteq \widehat{\mathcal{D}}$ are defined in (6.3). In each of Sections 7.1 to 7.4, we denote by $\overline{F}_N$ the enriched free energy of the model under consideration, and by f the candidate for its limit. The meaning of these two symbols thus changes from one subsection to the next, and it also differs from the one they had in Sections 1 to 5, where they referred to the bipartite model; the relation between the two settings is explained in Remark 7.3.

7.1. **The Ising perceptron.** Let $U : \mathbb{R} \to \mathbb{R}$ be a twice continuously differentiable function with bounded first and second derivatives. We do not assume that U is bounded or concave; for instance, for every $s \geqslant 0$ and every probability measure ν with compact support, the activation function $u_{s,\nu}$ defined in (1.11) satisfies our assumption. We consider the Ising perceptron with activation function U, that is, the model in (1.10) with $P_N = \mu_{\mathcal{I}}^{\otimes N}$ and $M = M_N$, where we recall that $\mu_{\mathcal{I}} = \frac{1}{2}(\delta_{-1} + \delta_1)$. Its free energy is

$$-\frac{1}{N}\mathbb{E}\log\int \exp\left(\sum_{j=1}^{M_N} U\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N} g_{ij}\sigma_i\right)\right) \mathrm{d}\mu_{\mathcal{I}}^{\otimes N}(\sigma). \tag{7.1}$$

The enriched free energy. We recall that the domain $\mathcal{D}$ defined in (6.3) is the set of triples (t, q_1, q_2) with $t \in [0,1]$, $q_1 \in \mathcal{Q}_\infty$, and $q_2 \in \mathcal{Q}_\infty$ such that $q_2 \leqslant 1-t$, and we write $q = (q_1, q_2)$. Conditionally on $\mathfrak{R}$, let $(V_i^{q_1})_{i\in\mathbb{N}}$ and $(W_j^{q_2})_{j\in\mathbb{N}}$ be independent centered Gaussian processes indexed by $\operatorname{supp}\mathfrak{R}$, independent of (g_{ij}), with covariances

$$\mathbb{E}\left[V_i^{q_1}(\alpha)V_i^{q_1}(\alpha') \mid \mathfrak{R}\right] = q_1(\alpha\wedge\alpha') \qquad \text{and} \qquad \mathbb{E}\left[W_j^{q_2}(\alpha)W_j^{q_2}(\alpha') \mid \mathfrak{R}\right] = q_2(\alpha\wedge\alpha').$$

Let also $(z_j)_{j\in\mathbb{N}}$ be independent standard Gaussian random variables, independent of all the other random variables. For every $(t,q) \in \mathcal{D}$, $\sigma \in \{-1,1\}^N$, $\alpha \in \operatorname{supp}\mathfrak{R}$ and $j \in \mathbb{N}$, we set

$$S_j := \sqrt{\frac{t}{N}}\sum_{i=1}^N g_{ij}\sigma_i + W_j^{q_2}(\alpha) + \sqrt{1-t-q_2(1)}\, z_j. \tag{7.2}$$

The first term in (7.2) is the pre-activation of the original model multiplied by $\sqrt{t}$, and the second term is a field indexed by the cascade, as in the bipartite model. The third term compensates for the variations of the first two: conditionally on $\mathfrak{R}$, and for every fixed (σ,α), the random variable S_j is a centered Gaussian with variance $t+q_2(1)+(1-t-q_2(1)) = 1$, irrespective of the choice of $(t,q) \in \mathcal{D}$. The constraint $q_2 \leqslant 1-t$ in the definition of $\mathcal{D}$ is precisely what makes this possible. We define the enriched free energy $\overline{F}_N : \mathcal{D} \to \mathbb{R}$ by

$$\overline{F}_N(t,q) := -\frac{1}{N}\mathbb{E}\log\iint \mathbb{E}_z \exp\left(\sum_{j=1}^{M_N} U(S_j) + \sqrt{2}\sum_{i=1}^N V_i^{q_1}(\alpha)\sigma_i - Nq_1(1)\right) \mathrm{d}\mu_{\mathcal{I}}^{\otimes N}(\sigma)\,\mathrm{d}\mathfrak{R}(\alpha), \tag{7.3}$$

where $\mathbb{E}_z$ denotes the expectation with respect to the variables $(z_j)_{j\in\mathbb{N}}$ only, and $\mathbb{E}$ denotes the expectation with respect to all the other sources of randomness. The variables (z_j) are thus integrated inside the logarithm, on the same footing as σ and α. As in (1.16), the term $Nq_1(1)$ is one half of the variance of the Gaussian field acting on the spins. At $t=1$, the constraint $q_2 \leqslant 1-t$ imposes that $q_2 = 0$, and $\overline{F}_N(1,0)$ is the free energy in (7.1). At the other endpoint $t=0$, the quantities S_j no longer depend on the spins, and the enriched free energy can be computed explicitly; see (7.10) below.

The approximate Hamilton–Jacobi equation. We denote by $\langle\cdot\rangle$ the expectation with respect to the Gibbs measure associated with (7.3). This is a random probability measure on the set of triples (σ,α,z), with $z = (z_j)_{j\leqslant M_N}$, and we write (σ^1,α^1,z^1) and (σ^2,α^2,z^2) for two independent random variables sampled according to it, also called replicas. For $\ell \in \{1,2\}$, we write S_j^ℓ for the quantity in (7.2) evaluated at $(\sigma^\ell,\alpha^\ell,z^\ell)$, and we consider the three overlaps

$$R_{12}^\alpha := \alpha^1\wedge\alpha^2, \qquad R_{12}^\sigma := \frac{1}{N}\sigma^1\cdot\sigma^2, \qquad \text{and} \qquad R_{12}^{U'} := \frac{1}{N}\sum_{j=1}^{M_N} U'(S_j^1)U'(S_j^2).$$

Under the measure $\mathbb{E}\langle\cdot\rangle$, the overlap R_{12}^α is uniformly distributed on $[0,1]$; see [12, Proposition 4.8]. For every integrable random variable X, we write $\mathbb{E}\langle X \mid R_{12}^\alpha\rangle$ for the conditional expectation of X given R_{12}^α under this measure, and $\mathbb{E}\langle X \mid R_{12}^\alpha = s\rangle$ for its value at $s \in [0,1]$, so that

$$\mathbb{E}\langle X\rangle = \int_0^1 \mathbb{E}\langle X \mid R_{12}^\alpha = s\rangle\,\mathrm{d}s.$$

By Gaussian integration by parts, we have, for almost every $s \in [0,1]$,

$$\partial_{q_1}\overline{F}_N(s) = \mathbb{E}\langle R_{12}^\sigma \mid R_{12}^\alpha = s\rangle \qquad \text{and} \qquad \partial_{q_2}\overline{F}_N(s) = \frac{1}{2}\mathbb{E}\left\langle R_{12}^{U'} \mid R_{12}^\alpha = s\right\rangle, \tag{7.4}$$

as well as

$$\partial_t\overline{F}_N = \frac{1}{2}\mathbb{E}\left\langle R_{12}^{U'}R_{12}^\sigma\right\rangle. \tag{7.5}$$

Here, as in (1.21), the derivative $\partial_{q_i}\overline{F}_N$ at a given point (t,q) is an element of $L^2([0,1))$, and we write $\partial_{q_i}\overline{F}_N(s)$ for its value at s, keeping the dependence on (t,q) implicit. In the derivation of (7.4) and (7.5), the integrations by parts also produce terms that involve $U''(S_j)$ and $U'(S_j)^2$ evaluated at a single replica, and these terms cancel out. This cancellation comes from the fact

that the variance of S_j does not depend on (t,q), and it is the reason for the introduction of the variables (z_j) in (7.2). Since R_{12}^α is uniformly distributed, the identities in (7.4) yield

$$\big\langle \partial_{q_1}\overline{F}_N, \partial_{q_2}\overline{F}_N\big\rangle_{L^2} = \frac{1}{2}\mathbb{E}\left\langle R_{12}^{U'}\,\mathbb{E}\,\langle R_{12}^\sigma \mid R_{12}^\alpha\rangle\right\rangle,$$

and combining this with (7.5), we obtain

$$\partial_t\overline{F}_N - \big\langle \partial_{q_1}\overline{F}_N, \partial_{q_2}\overline{F}_N\big\rangle_{L^2} = \frac{1}{2}\mathbb{E}\left\langle R_{12}^{U'}\,(R_{12}^\sigma - \mathbb{E}\,\langle R_{12}^\sigma \mid R_{12}^\alpha\rangle)\right\rangle. \tag{7.6}$$

Since $|R_{12}^{U'}| \leqslant \frac{M_N}{N}\|U'\|_\infty^2$, the Cauchy–Schwarz inequality shows that the absolute value of the right-hand side of (7.6) is at most

$$\frac{M_N}{2N}\|U'\|_\infty^2\,\Delta_N^{1/2}, \qquad \text{where} \quad \Delta_N := \mathbb{E}\left\langle (R_{12}^\sigma - \mathbb{E}\,\langle R_{12}^\sigma \mid R_{12}^\alpha\rangle)^2\right\rangle. \tag{7.7}$$

The quantity Δ_N measures the fluctuations of the spin overlap R_{12}^σ conditionally on the overlap of the cascade R_{12}^α; cf. [17, (6.79)] or [34, (2.23)]. Due to the synchronization mechanism of [41, 42, 43], we expect this term to be small for generic q; see also [18]. The identity (7.6) thus leads us to expect that the function $\overline{F}_N$ converges, as N tends to infinity, to a solution f of the equation

$$\partial_t f - \langle \partial_{q_1} f, \partial_{q_2} f\rangle_{L^2} = 0. \tag{7.8}$$

This is the equation (6.9) studied in Section 6, and we interpret it in the sense of Definition 6.2. This definition involves a parameter B, which bounds the derivative of the solution with respect to q_2. The appropriate choice is suggested by (7.4): the derivatives $\partial_{q_1}\overline{F}_N$ and $\partial_{q_2}\overline{F}_N$ are bounded by 1 and by $\frac{M_N}{2N}\|U'\|_\infty^2$ respectively, and the latter quantity converges to

$$B := \frac{\lambda}{2}\,\|U'\|_\infty^2. \tag{7.9}$$

The initial condition. At $t=0$, the quantities S_j in (7.2) do not depend on σ, and the exponential in (7.3) becomes a product of N factors, each depending on one spin σ_i and on the field $V_i^{q_1}$, and of M_N factors, each depending on one variable z_j and on the field $W_j^{q_2}$. As for (3.7), the additivity property of the cascade [9, Lemma 4.11] then gives

$$\overline{F}_N(0,q) = \psi_{\mu_{\mathcal{I}}}(q_1) + \frac{M_N}{N}\mathcal{A}(q_2), \tag{7.10}$$

where $\psi_{\mu_{\mathcal{I}}}$ is the cascade transform of the Ising measure defined in (1.18), and where, for every $q_2 \in K_1$, we set

$$\mathcal{A}(q_2) := -\mathbb{E}\log\int \mathbb{E}_{z_1}\exp\left(U\left(W_1^{q_2}(\alpha) + \sqrt{1-q_2(1)}\,z_1\right)\right)\mathrm{d}\mathfrak{R}(\alpha). \tag{7.11}$$

We call $\mathcal{A}$ the pattern functional associated with U. In view of (7.10), the initial condition for the equation (7.8) must be

$$f(0,q) = \psi_{\mu_{\mathcal{I}}}(q_1) + \lambda\mathcal{A}(q_2). \tag{7.12}$$

The equation (7.8) with initial condition (7.12) has the same structure as the equation (1.21) of the bipartite model: the nonlinearity is bilinear, and the initial condition is the sum of a function of q_1 and of a function of q_2. The first function does not depend on the activation function; it is the same as in (1.21), and it is convex by Proposition 3.4. The second function is the only place where the activation function enters, and in general it is not convex. For instance, when $U = u_{s,\nu}$ for some $s \geqslant 0$, integrating over z_1 in (7.11) gives $\mathcal{A}(q_2) = \psi_\nu(sq_2)$, and we have seen in Remark 1.5 that ψ_ν need not be convex.

The only property of the pattern functional that we need is the following one-sided bound on its increments. As discussed before the statement of Theorem 6.7, this bound expresses that the derivative of $\mathcal{A}$ is an increasing path taking values in $[0, L^2/2]$, where L is the Lipschitz constant of U, without requiring that $\mathcal{A}$ be differentiable. This is consistent with the second identity in (7.4).

Lemma 7.1 (Increments of the pattern functional)**.** *Let $U : \mathbb{R} \to \mathbb{R}$ be a Lipschitz continuous function, with Lipschitz constant L, and let $\mathcal{A}$ be defined by (7.11). For every $q, q' \in K_1$, we have*

$$\mathcal{A}(q') - \mathcal{A}(q) \leqslant \frac{L^2}{2}\,\omega(q' - q), \tag{7.13}$$

where ω is the support function of K_1 defined in (6.1). In particular, the function $\psi_2 := \lambda\mathcal{A}$ satisfies the assumption (6.31) of Theorem 6.7 with $B = \lambda L^2/2$, and the function $\mathcal{A}$ is Lipschitz continuous with respect to the L^1 norm, with Lipschitz constant $L^2/2$.

Proof. The last assertion follows from (7.13) and (6.2). We decompose the proof of (7.13) into three steps.

Step 1: Step paths and smooth activation functions. We first assume that U is twice continuously differentiable with bounded first and second derivatives, so that $\|U'\|_\infty \leqslant L$, and we consider step paths. Let $k \geqslant 1$ be an integer, let $0 = s_0 < s_1 < \cdots < s_k = 1$ and $0 \leqslant q_0 \leqslant \cdots \leqslant q_{k-1} \leqslant 1$, and let $q := \sum_{i=0}^{k-1} q_i \mathbf{1}_{[s_i, s_{i+1})} \in K_1$. We also set $q_{-1} := 0$ and $q_k := 1$. For such a path, the right-hand side of (7.11) can be computed by a recursive Gaussian integration; see [17, Theorem 5.25]. To state this, we let Z be a standard Gaussian random variable, and for every $s \in [0,1]$, $v \geqslant 0$ and every Lipschitz function $h : \mathbb{R} \to \mathbb{R}$, we define $T_{s,v}h : \mathbb{R} \to \mathbb{R}$ by

$$T_{s,v}h(x) := \frac{1}{s}\log \mathbb{E}\exp\left(s\,h\big(x + \sqrt{v}\,Z\big)\right) \quad \text{if } s > 0, \qquad \text{and} \qquad T_{0,v}h(x) := \mathbb{E}\left[h\big(x + \sqrt{v}\,Z\big)\right].$$

Setting $X_{k+1} := U$ and then, successively for $i = k, k-1, \ldots, 0$,

$$X_i := T_{s_i,\, q_i - q_{i-1}} X_{i+1}, \tag{7.14}$$

we have $\mathcal{A}(q) = -X_0(0)$. The step $i = k$ of this recursion is the average over z_1 in (7.11), since $s_k = 1$ and $q_k - q_{k-1} = 1 - q(1)$, and the step $i = 0$ is a plain Gaussian average, since $s_0 = 0$.

We will use two classical properties of the operator $T_{s,v}$, valid for every function h that is twice continuously differentiable with bounded first and second derivatives. For every $x \in \mathbb{R}$, we denote by $\mathbb{E}_x^{s,v,h}$ the expectation with respect to the probability measure whose density with respect to the law of Z is proportional to $\exp(s\,h(x + \sqrt{v}\,Z))$. The first property is that the derivative of the mapping $h \mapsto T_{s,v}h(x)$ in the direction of a bounded continuous function η is $\mathbb{E}_x^{s,v,h}[\eta(x + \sqrt{v}\,Z)]$, and that, similarly,

$$(T_{s,v}h)'(x) = \mathbb{E}_x^{s,v,h}\left[h'\big(x + \sqrt{v}\,Z\big)\right]. \tag{7.15}$$

The second property is that, for every $v > 0$ and $x \in \mathbb{R}$,

$$\partial_v T_{s,v}h(x) = \frac{1}{2}\left((T_{s,v}h)'' + s\big((T_{s,v}h)'\big)^2\right)(x) = \frac{1}{2}\,\mathbb{E}_x^{s,v,h}\left[\big(h'' + s(h')^2\big)\big(x + \sqrt{v}\,Z\big)\right]. \tag{7.16}$$

Indeed, when $s > 0$, the function $w(v,x) := \exp(s\,T_{s,v}h(x)) = \mathbb{E}\left[\exp\left(s\,h\big(x + \sqrt{v}\,Z\big)\right)\right]$ solves the heat equation $\partial_v w = \frac{1}{2}\partial_x^2 w$, so each of the three quantities in (7.16) is equal to $\partial_x^2 w/(2s\,w)$ evaluated at (v,x): for the second one, we express the derivatives of w in terms of those of $T_{s,v}h$, and for the third one, we compute $\partial_x^2 w$ by differentiating twice under the expectation. When $s = 0$, the function $(v,x) \mapsto T_{0,v}h(x)$ itself solves the heat equation, and the argument is the same.

Let $(\xi_i)_{-1 \leqslant i \leqslant k}$ be the Markov chain started from $\xi_{-1} := 0$ and such that, for every $i \in \{0, \ldots, k\}$ and conditionally on ξ_{i-1}, the random variable ξ_i has the law of $\xi_{i-1} + \sqrt{q_i - q_{i-1}}\,Z$ under the probability measure associated with $\mathbb{E}_{\xi_{i-1}}^{s_i,\, q_i - q_{i-1},\, X_{i+1}}$. By (7.15) and (7.14), we have $X_i'(\xi_{i-1}) = \mathbb{E}[X_{i+1}'(\xi_i) \mid \xi_{i-1}]$ for every $i \in \{0, \ldots, k\}$. In other words, setting $m_i := X_{i+1}'(\xi_i)$, the sequence $(m_i)_{-1 \leqslant i \leqslant k}$ is a martingale. Since $m_k = U'(\xi_k)$, this martingale is bounded by L, and by Jensen's inequality,

$$0 \leqslant \mathbb{E}\left[m_0^2\right] \leqslant \mathbb{E}\left[m_1^2\right] \leqslant \cdots \leqslant \mathbb{E}\left[m_{k-1}^2\right] \leqslant L^2. \tag{7.17}$$

We now assume that

$$0 < q_0 < q_1 < \cdots < q_{k-1} < 1. \tag{7.18}$$

The set of vectors $(q_0, \ldots, q_{k-1})$ satisfying (7.18) is an open subset of $\mathbb{R}^k$, and on this set, all the variance parameters $q_j - q_{j-1}$, $j \in \{0, \ldots, k\}$, appearing in the recursion (7.14) are positive. Since

the heat kernel depends smoothly on its variance parameter away from zero, the quantity $\mathcal{A}(q)$ is a continuously differentiable function of $(q_0, \dots, q_{k-1})$ on this set. We fix $i \in \{0, \dots, k-1\}$ and compute the derivative of $\mathcal{A}(q)$ with respect to q_i. Increasing q_i increases the variance parameter $q_i - q_{i-1}$ of the step i of the recursion (7.14), and it decreases by the same amount the variance parameter $q_{i+1} - q_i$ of the step $i+1$. By the first equality in (7.16), the derivative of X_{i+1} with respect to the variance parameter of the step $i+1$ is $\frac{1}{2}(X''_{i+1} + s_{i+1}(X'_{i+1})^2)$. By the second equality in (7.16), applied with $h = X_{i+1}$, the derivative of $X_i(x)$ with respect to the variance parameter of the step i is

$$\mathbb{E}_x^{s_i,\, q_i - q_{i-1},\, X_{i+1}} \left[\eta\big(x + \sqrt{q_i - q_{i-1}}\, Z\big) \right], \qquad \text{with} \quad \eta := \frac{1}{2}\left(X''_{i+1} + s_i (X'_{i+1})^2\right),$$

and by the first of the two properties recalled above, this is also the derivative of $X_i(x)$ when X_{i+1} is perturbed in the direction of η. The combined effect of these two variations on X_i is therefore the same as that of perturbing X_{i+1} in the direction of

$$\frac{1}{2}\left(X''_{i+1} + s_i (X'_{i+1})^2\right) - \frac{1}{2}\left(X''_{i+1} + s_{i+1} (X'_{i+1})^2\right) = -\frac{s_{i+1} - s_i}{2} (X'_{i+1})^2.$$

The terms involving second derivatives have cancelled out. Propagating this perturbation through the steps $i, i-1, \dots, 0$ of the recursion, using the first of the two properties recalled above at each step, we obtain

$$\partial_{q_i} \mathcal{A}(q) = \frac{s_{i+1} - s_i}{2}\, \mathbb{E}\left[m_i^2\right]. \tag{7.19}$$

Let now q' be another step path defined on the same partition $(s_i)_{0 \leqslant i \leqslant k}$, and whose values $(q'_i)_{0 \leqslant i < k}$ also satisfy the strict inequalities in (7.18), that is, $0 < q'_0 < \cdots < q'_{k-1} < 1$. For every $a \in [0,1]$, the path $q(a) := q + a(q' - q)$ is a step path in K_1 defined on this partition, its values also satisfy the strict inequalities in (7.18), and we denote by $(m_i(a))_{-1 \leqslant i \leqslant k}$ the martingale constructed above for this path. By (7.19) and the chain rule, we have

$$\frac{\mathrm{d}}{\mathrm{d}a} \mathcal{A}(q(a)) = \sum_{i=0}^{k-1} (q'_i - q_i)\, \frac{s_{i+1} - s_i}{2}\, \mathbb{E}\left[m_i(a)^2\right] = \langle c(a), q' - q \rangle_{L^2},$$

where

$$c(a) := \frac{1}{2} \sum_{i=0}^{k-1} \mathbb{E}\left[m_i(a)^2\right] \mathbf{1}_{[s_i, s_{i+1})}.$$

By (7.17), the path $c(a)$ belongs to $K_{L^2/2}$. Since $\frac{L^2}{2}\omega$ is the support function of $K_{L^2/2}$, the derivative above is at most $\frac{L^2}{2}\omega(q' - q)$, and integrating over $a \in [0,1]$ yields (7.13) for every pair of step paths that are defined on a common partition and whose values satisfy the strict inequalities in (7.18).

Step 2: General paths. We keep the assumption that U is smooth, and let $q, q' \in K_1$. With the notation of (4.2), for every $j \in \mathbb{N}$, we let $e_j \in E_j$ be the vector with coordinates $(e_j)_k := (k - \frac{1}{2})2^{-j}$, $k \in \{1, \dots, 2^j\}$, and we set

$$q^{(j)} := \iota_j\left((1 - 2^{-j})\pi_j q + 2^{-j} e_j\right) \qquad \text{and} \qquad q'^{(j)} := \iota_j\left((1 - 2^{-j})\pi_j q' + 2^{-j} e_j\right).$$

The coordinates of $\pi_j q$ and of $\pi_j q'$ are increasing and take values in $[0,1]$, while those of e_j are strictly increasing and take values in $(0,1)$. The paths $q^{(j)}$ and $q'^{(j)}$ are therefore step paths defined on the same dyadic partition, their values satisfy the strict inequalities in (7.18), and by Step 1, the inequality (7.13) holds for this pair of paths. Moreover, since the dyadic averages $\iota_j \pi_j q$ and $\iota_j \pi_j q'$ converge to q and q' in L^1 as j tends to infinity, so do the paths $q^{(j)}$ and $q'^{(j)}$. By (6.2) and the subadditivity of ω, the function ω is continuous with respect to the L^1 norm. The functional $\mathcal{A}$ is also continuous on K_1 with respect to the L^1 norm; this can be seen by a Gaussian interpolation between the fields associated with two paths, as for the function ψ_ρ in Lemma 3.2. We can therefore pass to the limit in the inequality (7.13) for $q^{(j)}$ and $q'^{(j)}$, and we obtain that it also holds for q and q'.

Step 3: Lipschitz activation functions. Finally, let U be Lipschitz continuous with Lipschitz constant L. For every $\varepsilon > 0$, let U_ε be the convolution of U with a smooth probability density

supported in $[-\varepsilon,\varepsilon]$. The function U_ε is twice continuously differentiable with bounded first and second derivatives, it is Lipschitz continuous with Lipschitz constant L, and $\sup_{x\in\mathbb{R}}|U_\varepsilon(x)-U(x)|\leqslant L\varepsilon$. The quantity in (7.11) changes by at most $L\varepsilon$ when U is replaced by U_ε, uniformly over $q_2\in K_1$. By Steps 1 and 2, the inequality (7.13) holds for the pattern functional associated with U_ε, and letting ε tend to zero, we conclude that it also holds for $\mathcal{A}$. □

The candidate limit. We can now identify the solution of the equation (7.8) with initial condition (7.12). We recall that the function $\psi^*_{\mu_{\mathcal{I}}}$ is defined in (1.19).

Proposition 7.2 (Candidate limit for the Ising perceptron)**.** *Let B be as in (7.9). For every $(t,q)\in\widehat{\mathcal{D}}$, let*

$$f(t,q):=\sup_{p\in\mathcal{Q}_{\infty,\leqslant 1}}\left\{\langle p,q_1\rangle_{L^2}+\lambda\mathcal{A}(q_2+tp)-\psi^*_{\mu_{\mathcal{I}}}(p)\right\}. \tag{7.20}$$

The supremum in (7.20) is attained. The function f is the unique Lipschitz viscosity solution of (7.8) with derivative bound B, in the sense of Definition 6.2, such that $f(0,q)=\psi_{\mu_{\mathcal{I}}}(q_1)+\lambda\mathcal{A}(q_2)$ for every $q\in\mathcal{Q}_2\times K_1$. Moreover, for every $(t,q),(t',q')\in\widehat{\mathcal{D}}$, we have

$$|f(t,q)-f(t',q')|\leqslant|q_1-q_1'|_{L^1}+B|q_2-q_2'|_{L^1}+B|t-t'|. \tag{7.21}$$

We recall that, for every $(t,q)\in\widehat{\mathcal{D}}$ and $p\in\mathcal{Q}_{\infty,\leqslant 1}$, we have $q_2+tp\leqslant(1-t)+t=1$, so that the path q_2+tp belongs to K_1 and the right-hand side of (7.20) is well-defined.

Proof of Proposition 7.2. We apply Theorem 6.7 with $\psi_2:=\lambda\mathcal{A}$, and with the function $g:\mathcal{H}\to\mathbb{R}\cup\{+\infty\}$ defined by $g:=\psi^*_{\mu_{\mathcal{I}}}$ on K_1 and $g:=+\infty$ on $\mathcal{H}\setminus K_1$. By Lemma 7.1, the function ψ_2 satisfies the assumption (6.31). By the definition in (1.19), the function $\psi^*_{\mu_{\mathcal{I}}}$ is a supremum of functions of p that are affine and continuous for the topology of $\mathcal{H}$. Since K_1 is convex and closed in $\mathcal{H}$, the function g is convex and lower semicontinuous. It is also proper, since $\psi^*_{\mu_{\mathcal{I}}}(0)=0$, as observed in Step 1 of the proof of Theorem 4.2. Theorem 6.7 therefore applies, and it yields all the announced properties, provided that we identify the initial condition in (6.33); that is, we need to show that, for every $q_1\in\mathcal{Q}_2$,

$$\sup_{p\in\mathcal{Q}_{\infty,\leqslant 1}}\left\{\langle p,q_1\rangle_{L^2}-\psi^*_{\mu_{\mathcal{I}}}(p)\right\}=\psi_{\mu_{\mathcal{I}}}(q_1). \tag{7.22}$$

This identity is contained in Theorem 4.2. Indeed, this theorem states that the right-hand side of (1.20), in which the supremum can be restricted to $\mathcal{Q}_{\infty,\leqslant 1}$, is equal to $\psi_{\mu_{\mathcal{I}}}(q_1)+\lambda\psi_\nu(q_2)$ at $t=0$, and at $t=0$, the term $\lambda\psi_\nu(q_2)$ can be taken out of the supremum. □

In view of the formal derivation of the equation (7.8) and of the identity (7.10), we expect that, for every $(t,q)\in\mathcal{D}$,

$$\lim_{N\to\infty}\overline{F}_N(t,q)\stackrel{?}{=}f(t,q), \tag{7.23}$$

where f is the function defined in (7.20). Evaluating this at $(t,q)=(1,0)$, we obtain the following prediction for the limit of the free energy in (7.1):

$$\lim_{N\to\infty}\overline{F}_N(1,0)\stackrel{?}{=}f(1,0)=\sup_{p\in\mathcal{Q}_{\infty,\leqslant 1}}\left\{\lambda\mathcal{A}(p)-\psi^*_{\mu_{\mathcal{I}}}(p)\right\}. \tag{7.24}$$

Comparison with the formula of [39]. When U is bounded, the limit of the free energy in (7.1) was identified in [39, Theorem 2.3], and the result coincides with the right-hand side of (7.24). To see this, we first note that in [39], the free energy is defined without the minus sign in front of the logarithm. Accordingly, the variational formula obtained there is an infimum over paths $p\in\mathcal{Q}_{\infty,\leqslant 1}$, and the functional to be minimized is the sum of two terms. The first term is λ times the pattern functional of [39, Definition 2.1]. For step paths, this functional is defined by the Gaussian recursion (7.14), including the final Gaussian average with variance $1-p(1)$, which corresponds to the average over z_1 in (7.11); it is therefore equal to $-\mathcal{A}(p)$. The second term is the Ising entropy of [39, Definition 2.2], which is the supremum over increasing step paths h of $\ell(h)+\frac{1}{2}\langle h,p\rangle_{L^2}$, where, in our notation, $\ell(h)=-\psi_{\mu_{\mathcal{I}}}(h/2)$. The change of variables $h=2r$ shows that this entropy is equal to $\psi^*_{\mu_{\mathcal{I}}}(p)$. The recipe described here thus recovers the formula of [39],

and interprets it as the value at $(t,q)=(1,0)$ of the solution to a Hamilton–Jacobi equation with the structure of Theorem 1.1. The same formula was also obtained in [20] for Lipschitz activation functions; see Section 7.3.

Remark 7.3 (Back to the bipartite model). Let ν be a probability measure supported in $[-1,1]$, let $s \geqslant 0$, and suppose that $U = u_{s,\nu}$, with $u_{s,\nu}$ as in (1.11). For clarity, we denote here by $\overline{F}_N^{\mathrm{bip}}$ the enriched free energy of the bipartite model defined in (1.17), and by f^{bip} its limit, given by Theorem 1.4. Inserting the definition of $u_{s,\nu}$ into (7.3), integrating out the variables (z_j), and using that the process $\sqrt{s}\, W_j^{q_2}$ has the same law as $W_j^{sq_2}$, we find that, for every $(t,q) \in \mathcal{D}$,

$$\overline{F}_N(t,q_1,q_2) = \overline{F}_N^{\mathrm{bip}}(st,q_1,sq_2). \tag{7.25}$$

In this computation, the integration over the variables (z_j) turns the term $-s|v|^2$ coming from the definition of $u_{s,\nu}$ into $-s(t+q_2(1))|v|^2$, where v denotes the layer of ν-distributed spins of the bipartite model; this is the correction term in (1.16), evaluated at (st,sq_2) in place of (t,q_2). The variables (z_j) thus play here the role that this correction term plays in the bipartite model. Similarly, since $\mathcal{A}(q_2) = \psi_\nu(sq_2)$ in this case, comparing (7.20) with (1.20) gives $f(t,q_1,q_2) = f^{\mathrm{bip}}(st,q_1,sq_2)$. For these activation functions, the convergence (7.23) is therefore a consequence of Theorem 1.4. The identity (7.25) also clarifies the relation between the time variables of the two settings. In the bipartite model, the time variable is the parameter s of the activation function $u_{s,\nu}$, which plays the role of an inverse temperature. In the present section, the activation function is fixed, and the time variable $t \in [0,1]$ is the fraction of the variance of the quantity S_j in (7.2) that comes from the pre-activation of the original model.

7.2. **The spherical perceptron.** We now replace the Ising spins by spherical spins. We keep the activation function U of Section 7.1, and we denote by $\mu_{\mathbb{S}_N}$ the uniform probability measure on the sphere $\mathbb{S}_N := \{\sigma \in \mathbb{R}^N : |\sigma|^2 = N\}$. With S_j as in (7.2), we define the enriched free energy $\overline{F}_N : \mathcal{D} \to \mathbb{R}$ by

$$\overline{F}_N(t,q) := -\frac{1}{N}\mathbb{E}\log \iint \mathbb{E}_z \exp\left(\sum_{j=1}^{M_N} U(S_j) + \sqrt{2}\sum_{i=1}^{N} V_i^{q_1}(\alpha)\sigma_i - Nq_1(1)\right) \mathrm{d}\mu_{\mathbb{S}_N}(\sigma)\,\mathrm{d}\mathfrak{R}(\alpha). \tag{7.26}$$

This is the definition in (7.3) with $\mu_{\mathbb{S}_N}$ in place of $\mu_{\mathcal{I}}^{\otimes N}$, and $\overline{F}_N(1,0)$ is the free energy of the spherical perceptron with activation function U, that is, the quantity in (1.10) with $P_N = \mu_{\mathbb{S}_N}$ and $M = M_N$.

The approximate Hamilton–Jacobi equation. The derivation of the identity (7.6) in Section 7.1 only uses one property of the Ising spins, namely the identity $|\sigma|^2 = N$. This identity ensures that the variance of S_j is equal to one, and that the term $-Nq_1(1)$ in (7.3) is one half of the variance of the Gaussian field acting on the spins. Since it also holds on $\mathbb{S}_N$, the identities (7.4), (7.5) and (7.6) remain valid for the function $\overline{F}_N$ in (7.26), provided that the replicas are now sampled from the Gibbs measure associated with (7.26). The candidate limit f is therefore again expected to solve the equation (7.8), in the sense of Definition 6.2 with B as in (7.9), and the error term in the equation satisfied by $\overline{F}_N$ is still controlled by the quantity Δ_N in (7.7).

The initial condition. At $t=0$, the additivity property of the cascade still separates the contribution of the patterns from that of the spins, and we get

$$\overline{F}_N(0,q) = \psi_{\mathrm{sph},N}(q_1) + \frac{M_N}{N}\mathcal{A}(q_2), \tag{7.27}$$

where $\mathcal{A}$ is the pattern functional defined in (7.11), and where

$$\psi_{\mathrm{sph},N}(q_1) := -\frac{1}{N}\mathbb{E}\log \iint \exp\left(\sqrt{2}\sum_{i=1}^{N} V_i^{q_1}(\alpha)\sigma_i - Nq_1(1)\right) \mathrm{d}\mu_{\mathbb{S}_N}(\sigma)\,\mathrm{d}\mathfrak{R}(\alpha). \tag{7.28}$$

The part of the initial condition that depends on q_2 is thus the same as for Ising spins, and only the part that depends on q_1 is modified. Since $\mu_{\mathbb{S}_N}$ is not a product measure, the quantity

$\psi_{\mathrm{sph},N}(q_1)$ does not reduce to a computation involving a single spin, and it depends on N. Its limit

$$\psi_{\mathrm{sph}}(q_1) := \lim_{N\to\infty} \psi_{\mathrm{sph},N}(q_1)$$

exists, and it is a convex function of q_1; see [13, Proposition 2.11 and Lemma 2.9]. This is the spherical counterpart of Proposition 3.4. Moreover, as will be explained just below, we have for every $q_1 \in \mathcal{Q}_\infty$ that

$$\psi_{\mathrm{sph}}(q_1) = \sup_{p\in\mathcal{Q}_{\infty,\leqslant 1}} \left\{ \langle p, q_1 \rangle_{L^2} - \mathsf{S}(p) \right\}, \tag{7.29}$$

where the functional $\mathsf{S} : \mathcal{Q}_{\infty,\leqslant 1} \to [0,+\infty]$ is defined by

$$\mathsf{S}(p) := \frac{1}{2}\int_0^1 \left(\frac{1}{D_p(v)} - \frac{1}{1-v} \right) \mathrm{d}v, \qquad \text{with} \quad D_p(v) := 1 - v - \int_0^1 \big(p(u) - v\big)_+ \,\mathrm{d}u, \tag{7.30}$$

and with the understanding that $1/0 = +\infty$. Denoting by $\gamma_p(v) := |\{u \in [0,1) : p(u) \leqslant v\}|$ the distribution function of the path p, we can also write $D_p(v) = \int_v^1 \gamma_p(w)\,\mathrm{d}w$. The identity (7.29) is a consequence of [2, Lemmas 2.2–2.3]. Indeed, let ψ^*_{sph} denote the convex conjugate of ψ_{sph}, defined as in (1.19). By [2, Lemma 2.3], we have $\psi^*_{\mathrm{sph}}(p) = \mathsf{S}(p)$ for every $p \in \mathcal{Q}_{\infty,\leqslant 1}$ such that $p(1) < 1$. In view of the definition of ψ^*_{sph}, this implies that the inequality

$$\langle p, q_1 \rangle_{L^2} - \mathsf{S}(p) \leqslant \psi_{\mathrm{sph}}(q_1)$$

holds for every such p, and by [2, Lemma 2.2], it is an equality for at least one such p. It thus remains to verify that this inequality is also valid when $p(1) = 1$. For every $\theta \in (0,1)$, since $p \geqslant 0$, we have $D_{\theta p} \geqslant D_p$, and thus $\mathsf{S}(\theta p) \leqslant \mathsf{S}(p)$. Applying the inequality to the path θp therefore gives $\theta \langle p, q_1 \rangle_{L^2} - \mathsf{S}(p) \leqslant \psi_{\mathrm{sph}}(q_1)$, and we conclude by letting θ tend to 1.

The candidate limit. The representation (7.29) provides us with an explicit convex function whose convex conjugate is ψ_{sph}. We extend S to $\mathcal{H}$ by setting $\mathsf{S} := +\infty$ outside of K_1. For every $v \in [0,1)$, the mapping $p \mapsto D_p(v)$ is concave and continuous on K_1, and it takes values in $[0, 1-v]$. Since the function $x \mapsto 1/x$ is convex and decreasing on $[0,\infty)$, the integrand in (7.30) is a nonnegative, convex and lower semicontinuous function of p, and by Fatou's lemma, the functional S is convex and lower semicontinuous. It is also proper, since $\mathsf{S}(0) = 0$. We can therefore apply Theorem 6.7 with $g := \mathsf{S}$, with $\psi_2 := \lambda\mathcal{A}$, which satisfies the assumption (6.31) by Lemma 7.1, and with B as in (7.9). We obtain that the function $f : \widehat{\mathcal{D}} \to \mathbb{R}$ defined by

$$f(t,q) := \sup_{p\in\mathcal{Q}_{\infty,\leqslant 1}} \left\{ \langle p, q_1 \rangle_{L^2} + \lambda\mathcal{A}(q_2 + tp) - \mathsf{S}(p) \right\} \tag{7.31}$$

is the unique Lipschitz viscosity solution of (7.8) with derivative bound B such that, for every $q \in \mathcal{Q}_2 \times K_1$,

$$f(0,q) = \psi_{\mathrm{sph}}(q_1) + \lambda\mathcal{A}(q_2), \tag{7.32}$$

where we have extended the identity (7.29) to every $q_1 \in \mathcal{Q}_2$ by continuity. In view of (7.27), this function is our candidate for the limit of $\overline{F}_N$, and at $(t,q) = (1,0)$, it yields the prediction

$$\lim_{N\to\infty} \overline{F}_N(1,0) \overset{?}{=} \sup_{p\in\mathcal{Q}_{\infty,\leqslant 1}} \left\{ \lambda\mathcal{A}(p) - \mathsf{S}(p) \right\}. \tag{7.33}$$

Comparison with the formulas of [3, 40]. For activation functions satisfying the assumptions of this subsection and bounded above, the limit of $\overline{F}_N(1,0)$ was identified in [3, Part I, Theorem 1.1]. The opposite sign convention for the free energy is used there, and their formula is therefore an infimum. Their order parameter is the distribution function γ_p of our path p. Under this identification, their Crisanti–Sommers entropy equals $\mathsf{S}(p)$, and their pattern functional equals $-\mathcal{A}(p)$, so the functional they minimize is $\mathsf{S}(p) - \lambda\mathcal{A}(p)$. The restriction $p(1) < 1$ in their parametrization can be removed by approximation, and their formula thus agrees with (7.33). When U is bounded and continuous, the limit of $\overline{F}_N(1,0)$ was also identified in [40, Theorem 1]. The variational problem there is again parametrized by distribution functions γ_p, with spherical term $\mathsf{S}(p)$ and a pattern functional expressed through a stochastic control problem whose value

is $-\mathcal{A}(p)$. After accounting for the opposite sign convention, this gives the same formula as in (7.33).

7.3. **Random activation functions.** We now allow the activation function to vary from one pattern to the next, in an i.i.d. fashion, and we treat Ising and spherical spins simultaneously. This is the setting of [20]. Let π be a probability measure on $\mathbb{R}$, let $(\eta_j)_{j\in\mathbb{N}}$ be i.i.d. random variables with law π, independent of (g_{ij}), and let $(\eta, x) \mapsto U_\eta(x)$ be a measurable function from $\mathbb{R}^2$ to $\mathbb{R}$. We call η_j the mark of the j-th pattern. We set $P_N^{\mathrm{I}} := \mu_{\mathcal{I}}^{\otimes N}$ and $P_N^{\mathrm{S}} := \mu_{\mathbb{S}_N}$, and for $\kappa \in \{\mathrm{I}, \mathrm{S}\}$, we consider the model with free energy

$$-\frac{1}{N}\mathbb{E}\log\int \exp\left(\sum_{j=1}^{M_N} U_{\eta_j}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N} g_{ij}\sigma_i\right)\right) \mathrm{d}P_N^\kappa(\sigma), \tag{7.34}$$

where, in this subsection, the expectation $\mathbb{E}$ also averages over the marks (η_j). The models of Sections 7.1 and 7.2 correspond to the case in which π is a Dirac mass. We assume that there exists a constant $L < \infty$ such that, for π-almost every η, the function U_η is twice continuously differentiable with $\|U_\eta'\|_\infty \leqslant L$ and $\|U_\eta''\|_\infty \leqslant L$, and that $\int |U_\eta(0)|\,\mathrm{d}\pi(\eta) < \infty$. We explain at the end of this subsection how to remove the smoothness assumption.

The enriched free energy and the approximate Hamilton–Jacobi equation. The enriched free energy $\overline{F}_N : \mathcal{D} \to \mathbb{R}$ is defined as in (7.3), with the function U_{η_j} applied to S_j and with the measure P_N^κ on the spins, that is,

$$\overline{F}_N(t,q) := -\frac{1}{N}\mathbb{E}\log\iint \mathbb{E}_z \exp\left(\sum_{j=1}^{M_N} U_{\eta_j}(S_j) + \sqrt{2}\sum_{i=1}^{N} V_i^{q_1}(\alpha)\sigma_i - Nq_1(1)\right) \mathrm{d}P_N^\kappa(\sigma)\,\mathrm{d}\mathfrak{R}(\alpha), \tag{7.35}$$

with S_j as in (7.2). The quantity $\overline{F}_N(1,0)$ is the free energy in (7.34). The computations of Section 7.1 can be performed conditionally on the marks, and they show that the identities (7.4), (7.5) and (7.6) remain valid, with

$$R_{12}^{U'} := \frac{1}{N}\sum_{j=1}^{M_N} U_{\eta_j}'(S_j^1)U_{\eta_j}'(S_j^2), \qquad \text{so that} \quad |R_{12}^{U'}| \leqslant \frac{M_N}{N}L^2.$$

The candidate limit f is thus again expected to solve the equation (7.8), in the sense of Definition 6.2 with $B := \lambda L^2/2$, and the error term in the equation satisfied by $\overline{F}_N$ is still controlled by the quantity Δ_N in (7.7).

The initial condition and the candidate limit. For π-almost every η, we write $\mathcal{A}_{U_\eta}$ for the pattern functional defined in (7.11) with U_η in place of U, and for every $q_2 \in K_1$, we set

$$\mathcal{A}_\pi(q_2) := \int \mathcal{A}_{U_\eta}(q_2)\,\mathrm{d}\pi(\eta). \tag{7.36}$$

Notice that the average over the mark is taken outside of the logarithm in (7.11). Conditioning on the marks and using the additivity property of the cascade, we see that

$$\overline{F}_N(0,q) = \psi_{\kappa,N}(q_1) + \frac{M_N}{N}\mathcal{A}_\pi(q_2),$$

where $\psi_{\mathrm{I},N} := \psi_{\mu_{\mathcal{I}}}$ and $\psi_{\mathrm{S},N} := \psi_{\mathrm{sph},N}$. Setting $\psi_{\mathrm{I}} := \psi_{\mu_{\mathcal{I}}}$ and $\psi_{\mathrm{S}} := \psi_{\mathrm{sph}}$, the initial condition for the equation (7.8) is therefore

$$f(0,q) = \psi_\kappa(q_1) + \lambda\mathcal{A}_\pi(q_2). \tag{7.37}$$

Compared with the previous two subsections, only the part of the initial condition that depends on q_2 has changed. Averaging the inequality (7.13) over the marks, we see that the function $\psi_2 := \lambda\mathcal{A}_\pi$ satisfies the assumption (6.31) with $B = \lambda L^2/2$. We also set

$$\psi_{\mathrm{I}}^* := \psi_{\mu_{\mathcal{I}}}^* \qquad \text{and} \qquad \psi_{\mathrm{S}}^* := \mathsf{S},$$

and we recall from the proof of Proposition 7.2 and from Section 7.2 that, for each $\kappa \in \{\mathrm{I}, \mathrm{S}\}$, the function equal to ψ_κ^* on K_1 and to $+\infty$ elsewhere is proper, lower semicontinuous and convex, and that its convex conjugate is ψ_κ. By Theorem 6.7, the function $f : \widehat{\mathcal{D}} \to \mathbb{R}$ defined by

$$f(t,q) := \sup_{p \in \mathcal{Q}_{\infty,\leqslant 1}} \left\{ \langle p, q_1 \rangle_{L^2} + \lambda \mathcal{A}_\pi(q_2 + tp) - \psi_\kappa^*(p) \right\} \tag{7.38}$$

is therefore the unique Lipschitz viscosity solution of (7.8) with derivative bound B and with initial condition (7.37). This formula contains (7.20) and (7.31) as special cases. At $(t,q) = (1,0)$, it yields the prediction

$$\lim_{N\to\infty} \overline{F}_N(1,0) \stackrel{?}{=} \sup_{p \in \mathcal{Q}_{\infty,\leqslant 1}} \left\{ \lambda \mathcal{A}_\pi(p) - \psi_\kappa^*(p) \right\}. \tag{7.39}$$

Comparison with the formulas of [20]. The right-hand side of (7.39) coincides with the formula proved in [20, Theorem 1.1], as we now explain. For every $p \in \mathcal{Q}_{\infty,\leqslant 1}$, let γ_p be the distribution function of p introduced below (7.30). For every Lipschitz function $U : \mathbb{R} \to \mathbb{R}$, the channel functional of [20] is defined by $\Phi_U(p) := \varphi(0,0)$, where $\varphi : [0,1] \times \mathbb{R} \to \mathbb{R}$ is the solution of the Parisi equation

$$\partial_v \varphi + \frac{1}{2} \left(\partial_x^2 \varphi + \gamma_p(v) (\partial_x \varphi)^2 \right) = 0 \quad \text{on } [0,1] \times \mathbb{R}, \qquad \varphi(1,\cdot) = U. \tag{7.40}$$

When p is a step path, the function γ_p is piecewise constant, and on each interval on which γ_p is equal to a constant s, the equation (7.40) is solved by the operators $T_{s,v}$ appearing in (7.14); see (7.16). In particular, we have $\gamma_p = 1$ on the interval $[p(1), 1]$, and solving the equation on this interval produces the logarithm of the Gaussian average with variance $1 - p(1)$ in (7.11). Solving (7.40) backward in v thus reproduces the recursion (7.14), and we obtain that $\Phi_U = -\mathcal{A}_U$. We note that the paths considered in [20] are defined on $[0,1]$ and take the value 1 at the endpoint 1; our path p is the restriction of such a path to $[0,1)$, and $p(1)$ denotes its left limit at 1. The functionals ψ_I^* and ψ_S^* coincide with the spin costs of [20], which are denoted there by A_I and A_S respectively. Indeed, the Ising cost A_I is defined in the same way as the Ising entropy of [39] discussed in Section 7.1, and it is therefore equal to $\psi_{\mu_\mathcal{I}}^*$; and the spherical cost A_S is defined by the formula (7.30). Changing the sign in (7.39), we can therefore rewrite this prediction as

$$\lim_{N\to\infty} \frac{1}{N} \mathbb{E} \log \int \exp\left(\sum_{j=1}^{M_N} U_{\eta_j} \left(\frac{1}{\sqrt{N}} \sum_{i=1}^{N} g_{ij} \sigma_i \right) \right) \mathrm{d}P_N^\kappa(\sigma) \stackrel{?}{=} \inf_{p \in \mathcal{Q}_{\infty,\leqslant 1}} \left\{ A_\kappa(p) + \lambda \int \Phi_{U_\eta}(p) \, \mathrm{d}\pi(\eta) \right\}, \tag{7.41}$$

and the right-hand side of (7.41) is the formula of [20, Theorem 1.1], in which the ratio λ is denoted by α.

On the assumptions and the mode of convergence. The result of [20] does not require the activation functions to be smooth: it only assumes that, for π-almost every η, the function U_η is Lipschitz continuous with Lipschitz constant at most L, and that $\int |U_\eta(0)| \, \mathrm{d}\pi(\eta) < \infty$. The smoothness assumption was only used here to perform the integrations by parts leading to (7.6), and the prediction (7.39) extends to this more general setting by regularization. Indeed, Lemma 7.1 is valid for Lipschitz activation functions, so the function f in (7.38) is well-defined in this setting, and it is still the solution of the Hamilton–Jacobi equation. Moreover, if U_η^ε denotes the convolution of U_η with a smooth probability density supported in $[-\varepsilon, \varepsilon]$, then $\sup_{x\in\mathbb{R}} |U_\eta^\varepsilon(x) - U_\eta(x)| \leqslant L\varepsilon$. Replacing U_η by U_η^ε therefore changes the free energy in (7.34) by at most $\frac{M_N}{N} L\varepsilon$, and it changes the pattern functional $\mathcal{A}_\pi$ by at most $L\varepsilon$, uniformly over K_1. Hence, if the prediction (7.39) is valid for the regularized activation functions, then it is also valid for the original ones.

We also point out that [20, Theorem 1.1] concerns the random variable obtained by removing the expectation in (7.34), which is shown there to converge in L^1; this implies the convergence of its expectation. Conversely, the fluctuations of this random variable can be controlled independently of the identification of the limit of its expectation. The constants $U_{\eta_j}(0)$ contribute

the term $-N^{-1}\sum_{j=1}^{M_N} U_{\eta_j}(0)$, which obeys the law of large numbers. For the remaining part, resampling one pattern together with its mark changes the logarithm of the partition function by a quantity whose second moment is bounded in terms of L only, and the Efron–Stein inequality thus shows that its variance is of order N^{-1}. The convergence of the expectation therefore implies the convergence of the random variable in L^1, and also in L^2 if $\int |U_\eta(0)|^2 \,\mathrm{d}\pi(\eta) < \infty$. When the activation functions are uniformly bounded, the bounded differences inequality also gives the almost sure convergence, and the convergence in L^r for every $r < \infty$.

7.4. **Labelled patterns from a Gaussian mixture.** We next consider a model in which each pattern carries a label, and in which the patterns are no longer standard Gaussian vectors: they are sampled from a mixture of two Gaussian distributions, as in the continuous-spin perceptron of [8, Section 2]. We point out however that, in [8], the spins are real-valued with a log-concave prior, and the activation functions are assumed to be concave. Under these assumptions, the Gibbs measure is log-concave, which ensures that the overlaps concentrate; in other words, the model is replica-symmetric, and its free energy is described in terms of finitely many scalar parameters. Here, the spins are Ising spins and the activation functions are not assumed to be concave, so that replica symmetry breaking is expected in general, and the variational formula below involves paths rather than scalar parameters.

Let $(y_j)_{j\in\mathbb{N}}$ be i.i.d. random variables with values in $\{-1,1\}$, and let $(\theta_i)_{i\in\mathbb{N}}$ be i.i.d. random variables whose law μ is a compactly supported probability measure on $\mathbb{R}$. These two families are independent of each other and of (g_{ij}), and in this subsection, the expectation $\mathbb{E}$ also averages over them. For each $y \in \{-1,1\}$, let $U_y : \mathbb{R} \to \mathbb{R}$ be a twice continuously differentiable function with bounded first and second derivatives. We consider the model with Ising spins whose free energy is

$$-\frac{1}{N}\mathbb{E}\log \int \exp\left(\sum_{j=1}^{M_N} U_{y_j}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N} g_{ij}\sigma_i + \frac{y_j}{N}\sum_{i=1}^{N}\theta_i\sigma_i\right)\right) \mathrm{d}\mu_{\mathcal{I}}^{\otimes N}(\sigma). \tag{7.42}$$

The argument of U_{y_j} in (7.42) can be rewritten as $N^{-1/2}\sum_{i=1}^{N}(g_{ij} + N^{-1/2}y_j\theta_i)\sigma_i$. The j-th pattern is thus the vector $(g_{ij})_{i\leqslant N} + N^{-1/2}y_j\theta$, where $\theta := (\theta_1,\dots,\theta_N)$: it is obtained by shifting a standard Gaussian vector by $N^{-1/2}\theta$ or by $-N^{-1/2}\theta$, according to the value of its label. Conditionally on θ, the patterns are therefore independent samples from the mixture of the two Gaussian distributions with identity covariance and with means $N^{-1/2}\theta$ and $-N^{-1/2}\theta$, and the label y_j selects both the component of the mixture from which the j-th pattern is sampled and the activation function. When $\mu = \delta_0$, the patterns are not shifted, and we recover the model of Section 7.3 with Ising spins and with the marks $\eta_j = y_j$; if in addition $U_{-1} = U_1$, then we recover the model of Section 7.1.

The enriched free energy. In addition to the pair of paths $q = (q_1, q_2)$, the enrichment of this model involves a pair of real parameters $x = (x_1, x_2)$: the parameter x_1 is an external field acting on the weighted magnetization, and the parameter x_2 shifts the pre-activations in the direction of the labels. We set $\mathcal{D}^{\mathrm{ext}} := \mathcal{D} \times \mathbb{R}^2$. With the same cascade fields $(V_i^{q_1})$ and $(W_j^{q_2})$ and the same Gaussian random variables (z_j) as in Section 7.1, for every $(t,q,x) \in \mathcal{D}^{\mathrm{ext}}$, we replace the definition of S_j in (7.2) by

$$S_j := \sqrt{\frac{t}{N}}\sum_{i=1}^{N} g_{ij}\sigma_i + \frac{ty_j}{N}\sum_{i=1}^{N}\theta_i\sigma_i + y_j x_2 + W_j^{q_2}(\alpha) + \sqrt{1-t-q_2(1)}\, z_j, \tag{7.43}$$

and we define $\overline{F}_N : \mathcal{D}^{\mathrm{ext}} \to \mathbb{R}$ by

$$\overline{F}_N(t,q,x) := -\frac{1}{N}\mathbb{E}\log \iint \mathbb{E}_z \exp\left(\sum_{j=1}^{M_N} U_{y_j}(S_j) + \sqrt{2}\sum_{i=1}^{N} V_i^{q_1}(\alpha)\sigma_i - Nq_1(1) - x_1\sum_{i=1}^{N}\theta_i\sigma_i\right) \mathrm{d}\mu_{\mathcal{I}}^{\otimes N}(\sigma)\,\mathrm{d}\mathfrak{R}(\alpha). \tag{7.44}$$

The quantity $\overline{F}_N(1,0,0)$ is the free energy in (7.42). When $\mu = \delta_0$, the function $\overline{F}_N$ does not depend on x_1, and at $x_2 = 0$, it reduces to the enriched free energy in (7.35) with $\kappa = \mathrm{I}$ and $\eta_j = y_j$. Notice that in (7.43), the shift of the patterns is multiplied by t, while their Gaussian part is multiplied by $\sqrt{t}$. This choice ensures that the contribution of the shift to the time derivative of $\overline{F}_N$ is matched by the product of the derivatives of $\overline{F}_N$ with respect to x_1 and x_2, as we now explain.

The approximate Hamilton–Jacobi equation. Let $\langle\cdot\rangle$ denote the expectation with respect to the Gibbs measure associated with (7.44), and let the overlaps R_{12}^{α} and R_{12}^{σ} be as in Section 7.1. We also set

$$R_{12}^{U'} := \frac{1}{N}\sum_{j=1}^{M_N} U'_{y_j}(S_j^1)U'_{y_j}(S_j^2), \qquad R_1^{yU'} := \frac{1}{N}\sum_{j=1}^{M_N} y_j U'_{y_j}(S_j^1), \qquad \text{and} \qquad m_1^{\theta} := \frac{1}{N}\sum_{i=1}^{N}\theta_i\sigma_i^1.$$

The identities in (7.4) for the derivatives of $\overline{F}_N$ with respect to q_1 and q_2 remain valid, and we also have

$$\partial_{x_1}\overline{F}_N = \mathbb{E}\left\langle m_1^{\theta}\right\rangle \qquad \text{and} \qquad \partial_{x_2}\overline{F}_N = -\mathbb{E}\left\langle R_1^{yU'}\right\rangle, \tag{7.45}$$

while the time derivative now is

$$\partial_t\overline{F}_N = \frac{1}{2}\mathbb{E}\left\langle R_{12}^{U'}R_{12}^{\sigma}\right\rangle - \mathbb{E}\left\langle R_1^{yU'}m_1^{\theta}\right\rangle. \tag{7.46}$$

The second term on the right-hand side of (7.46) comes from the derivative with respect to t of the second term on the right-hand side of (7.43). Writing $\partial_{(q_i,x_i)} := (\partial_{q_i},\partial_{x_i})$, and denoting by $\langle\cdot,\cdot\rangle_{L^2\times\mathbb{R}}$ the scalar product on $L^2([0,1))\times\mathbb{R}$, as in Section 6.4, we obtain

$$\begin{aligned}\partial_t\overline{F}_N - \left\langle\partial_{(q_1,x_1)}\overline{F}_N, \partial_{(q_2,x_2)}\overline{F}_N\right\rangle_{L^2\times\mathbb{R}} \\ = \frac{1}{2}\mathbb{E}\left\langle R_{12}^{U'}\left(R_{12}^{\sigma} - \mathbb{E}\left\langle R_{12}^{\sigma} \mid R_{12}^{\alpha}\right\rangle\right)\right\rangle - \mathbb{E}\left\langle R_1^{yU'}\left(m_1^{\theta} - \mathbb{E}\left\langle m_1^{\theta}\right\rangle\right)\right\rangle.\end{aligned} \tag{7.47}$$

The first term on the right-hand side of (7.47) is the same as in (7.6). The second term is new, and it is small as soon as the weighted magnetization m_1^{θ} is concentrated around its mean. More precisely, setting $L := \max_{y\in\{-1,1\}}\|U'_y\|_\infty$, we have $|R_{12}^{U'}| \leqslant \frac{M_N}{N}L^2$ and $|R_1^{yU'}| \leqslant \frac{M_N}{N}L$, and by the Cauchy–Schwarz inequality, the absolute value of the right-hand side of (7.47) is at most $\frac{M_N}{N}(L^2+L)(\Delta_N^{\mathrm{ext}})^{1/2}$, where

$$\Delta_N^{\mathrm{ext}} := \mathbb{E}\left\langle\left(R_{12}^{\sigma} - \mathbb{E}\left\langle R_{12}^{\sigma} \mid R_{12}^{\alpha}\right\rangle\right)^2\right\rangle + \mathbb{E}\left\langle\left(m_1^{\theta} - \mathbb{E}\left\langle m_1^{\theta}\right\rangle\right)^2\right\rangle. \tag{7.48}$$

Whenever the quantity Δ_N^{ext} is small, the function $\overline{F}_N$ is thus an approximate solution of the equation

$$\partial_t f - \left\langle\partial_{(q_1,x_1)}f, \partial_{(q_2,x_2)}f\right\rangle_{L^2\times\mathbb{R}} = 0. \tag{7.49}$$

This is the equation (6.51) studied in Section 6.4: the nonlinearity is again bilinear, now with respect to the two groups of variables (q_1,x_1) and (q_2,x_2). By (7.4) and (7.45), the derivatives of $\overline{F}_N$ with respect to q_1, x_1, q_2 and x_2 are bounded by 1, $\|\theta_1\|_\infty$, $\frac{M_N}{2N}L^2$ and $\frac{M_N}{N}L$ respectively. We therefore interpret the equation (7.49) in the sense of Section 6.4, with the derivative bounds

$$B := \frac{\lambda}{2}L^2, \qquad R_1 := \|\theta_1\|_\infty, \qquad \text{and} \qquad R_2 := \lambda L. \tag{7.50}$$

The initial condition. The enriched free energy at $t = 0$ is computed as in the previous subsections. We define $\mathcal{A}^{\mathrm{ext}} : K_1\times\mathbb{R}\to\mathbb{R}$ and $\mathcal{B} : \mathcal{Q}_\infty\times\mathbb{R}\to\mathbb{R}$ by

$$\mathcal{A}^{\mathrm{ext}}(q_2,x_2) := -\mathbb{E}\log\int\mathbb{E}_{z_1}\exp\left(U_{y_1}\left(W_1^{q_2}(\alpha) + y_1x_2 + \sqrt{1-q_2(1)}\,z_1\right)\right)\mathrm{d}\mathfrak{R}(\alpha), \tag{7.51}$$

$$\mathcal{B}(q_1,b) := -\mathbb{E}\log\int e^{-q_1(1)}\cosh\left(\sqrt{2}\,V_1^{q_1}(\alpha) + b\right)\mathrm{d}\mathfrak{R}(\alpha). \tag{7.52}$$

The function $\mathcal{A}^{\mathrm{ext}}(\cdot,x_2)$ is the average over the label y_1 of the pattern functionals associated with the shifted activation functions $U_{y_1}(\cdot + y_1x_2)$. The function $\mathcal{B}(\cdot,b)$ is, up to the additive constant

$-\log\cosh b$, the cascade transform of the law of an Ising spin tilted by the external field b; in particular, we have $\mathcal{B}(q_1,0)=\psi_{\mu_{\mathcal{I}}}(q_1)$. The additivity property of the cascade gives

$$\overline{F}_N(0,q,x)=\mathbb{E}\left[\mathcal{B}(q_1,-x_1\theta_1)\right]+\frac{M_N}{N}\mathcal{A}^{\mathrm{ext}}(q_2,x_2). \tag{7.53}$$

As in the previous subsections, this is the sum of a function of the first group of variables (q_1,x_1) and of a function of the second group of variables (q_2,x_2). However, the first function is no longer convex: for instance, the function $b\mapsto\mathcal{B}(0,b)=-\log\cosh b$ is concave. Proposition 6.8 therefore does not apply directly to this initial condition.

Fixing the magnetization. To overcome this difficulty, we follow [7] and sort the spin configurations according to their magnetization. We call magnetization profile any element m of $L^\infty(\mu;[-1,1])$, that is, any measurable function from $\mathbb{R}$ to $[-1,1]$, two such functions being identified if they coincide μ-almost everywhere. Given a magnetization profile m, we consider the model obtained by restricting the integral over σ in (7.44) to the configurations with magnetization profile m; informally, these are the configurations such that, for every ϑ in the support of μ, the average of the spins σ_i over the indices i for which θ_i is close to ϑ is close to $m(\vartheta)$. For these configurations, the weighted magnetization $\frac{1}{N}\sum_{i=1}^N\theta_i\sigma_i$ is close to

$$\overline{m}:=\mathbb{E}\left[\theta_1 m(\theta_1)\right]. \tag{7.54}$$

The restriction to a magnetization profile has two effects. First, the derivation of the identity (7.47) does not depend on the set of spin configurations over which we integrate, so this identity remains valid for the restricted model. Since the weighted magnetization is essentially constant on the set of configurations with magnetization profile m, the second term on the right-hand side of (7.47) is then negligible, and we are left with the same concentration requirement as in the previous subsections. Second, the initial condition becomes convex with respect to (q_1,x_1). To see this, for every $a\in[-1,1]$, we define $\psi_{\mu_{\mathcal{I}},a}:\mathcal{Q}_\infty\to\mathbb{R}$ by

$$\psi_{\mu_{\mathcal{I}},a}(r):=\sup_{b\in\mathbb{R}}\left\{\mathcal{B}(r,b)+ba\right\}. \tag{7.55}$$

This function is the cascade transform of Ising spins whose magnetization is constrained to be equal to a; see [7, Proposition 7.5]. It is convex, and $\psi_{\mu_{\mathcal{I}},0}=\psi_{\mu_{\mathcal{I}}}$; see [7, Lemma 6.7 and Proposition 6.10]. For every magnetization profile m, we define $\Psi_m:\mathcal{Q}_\infty\to\mathbb{R}$ and $\Psi_m^*:\mathcal{Q}_\infty\to\mathbb{R}\cup\{+\infty\}$ by

$$\Psi_m(r):=\mathbb{E}\left[\psi_{\mu_{\mathcal{I}},m(\theta_1)}(r)\right]\qquad\text{and}\qquad\Psi_m^*(p):=\sup_{r\in\mathcal{Q}_\infty}\left\{\langle p,r\rangle_{L^2}-\Psi_m(r)\right\}. \tag{7.56}$$

The enriched free energy at $t=0$ of the model restricted to the magnetization profile m is then expected to converge to

$$\Psi_m(q_1)+x_1\overline{m}+\lambda\mathcal{A}^{\mathrm{ext}}(q_2,x_2). \tag{7.57}$$

The function $(q_1,x_1)\mapsto\Psi_m(q_1)+x_1\overline{m}$ is convex, since Ψ_m is an average of convex functions.

The candidate limit at fixed magnetization. We now verify that Proposition 6.8 applies to the initial condition in (7.57). We start with the part that depends on (q_2,x_2). For every fixed $x_2\in\mathbb{R}$, Lemma 7.1 applies to each of the two shifted activation functions $U_y(\cdot+yx_2)$, with a Lipschitz constant L that depends neither on x_2 nor on y. Moreover, changing x_2 into x_2' changes each of these two functions by at most $L|x_2'-x_2|$ in the supremum norm, and it thus changes $\mathcal{A}^{\mathrm{ext}}(q_2,\cdot)$ by at most the same amount. The function $\psi_2:=\lambda\mathcal{A}^{\mathrm{ext}}$ therefore satisfies the assumption (6.55), with B and R_2 as in (7.50). We now turn to the part that depends on (q_1,x_1). For every $b\in\mathbb{R}$, the function $\mathcal{B}(\cdot,b)$ is a cascade transform up to an additive constant, so by Lemma 3.2, it is $\mathcal{Q}_2^*$-increasing, and Lipschitz continuous with respect to the L^1 norm, with Lipschitz constant 1. These two properties are preserved by the supremum in (7.55) and by the average in (7.56). In particular, the function Ψ_m extends to $\mathcal{Q}_2$ by continuity. Since Ψ_m is also convex, the arguments of Sections 2.3 and 4 leading to the identity (7.22), which only rely on these three properties of $\psi_{\mu_{\mathcal{I}}}$, also show that Ψ_m^* is infinite outside of K_1, and that, for every $q_1\in\mathcal{Q}_2$,

$$\sup_{p\in\mathcal{Q}_{\infty,\leqslant 1}}\left\{\langle p,q_1\rangle_{L^2}-\Psi_m^*(p)\right\}=\Psi_m(q_1). \tag{7.58}$$

We can thus apply Proposition 6.8 with $\psi_2 := \lambda\mathcal{A}^{\mathrm{ext}}$, and with the function $g : \mathcal{H} \times \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ defined by $g(p,c) := \Psi_m^*(p)$ if $p \in K_1$ and $c = \overline{m}$, and $g(p,c) := +\infty$ otherwise. This function is proper, lower semicontinuous and convex. Its effective domain is contained in $K_1 \times [-R_1, R_1]$, since $|\overline{m}| \leqslant \|\theta_1\|_\infty = R_1$, and by (7.58), its convex conjugate is the function $(q_1, x_1) \mapsto \Psi_m(q_1) + x_1\overline{m}$. We obtain that the function $f_m : \widehat{\mathcal{D}} \times \mathbb{R}^2 \to \mathbb{R}$ defined by

$$f_m(t,q,x) := x_1\overline{m} + \sup_{p \in \mathcal{Q}_{\infty,\leqslant 1}} \left\{ \langle p, q_1 \rangle_{L^2} + \lambda\mathcal{A}^{\mathrm{ext}}\left(q_2 + tp,\, x_2 + t\overline{m}\right) - \Psi_m^*(p) \right\} \tag{7.59}$$

is the unique Lipschitz viscosity solution of (7.49) with derivative bounds (B, R_1, R_2) whose initial condition is the function in (7.57). This function is our candidate for the limit of the enriched free energy of the model restricted to the magnetization profile m. In the formula (7.59), the dual variable associated with x_1 is frozen at the value $\overline{m}$, because the initial condition is affine in x_1 with slope $\overline{m}$. Its only effect is to shift the variable x_2 by $t\overline{m}$, in agreement with the definition of S_j in (7.43), in which the labels are multiplied by the sum of x_2 and of t times the weighted magnetization.

The candidate limit. It remains to relate the restricted models to the original one. The integral over σ in (7.44) is the sum of its restrictions to the different magnetization profiles, and on the exponential scale, such a sum is governed by its largest term. Taking into account the minus sign in the definition of the free energy, and in analogy with [7, (1.14)], we thus expect the enriched free energy $\overline{F}_N$ to be asymptotically equal to the infimum over the magnetization profiles of the enriched free energies of the restricted models. This leads to the candidate

$$\lim_{N\to\infty} \overline{F}_N(t,q,x) \stackrel{?}{=} f(t,q,x) := \inf_{m \in L^\infty(\mu;[-1,1])} f_m(t,q,x), \tag{7.60}$$

and in particular, at $(t,q,x) = (1,0,0)$, to the following prediction for the limit of the free energy in (7.42):

$$\lim_{N\to\infty} \overline{F}_N(1,0,0) \stackrel{?}{=} \inf_{m \in L^\infty(\mu;[-1,1])} \sup_{p \in \mathcal{Q}_{\infty,\leqslant 1}} \left\{ \lambda\mathcal{A}^{\mathrm{ext}}\left(p, \mathbb{E}[\theta_1 m(\theta_1)]\right) - \Psi_m^*(p) \right\}. \tag{7.61}$$

We can perform two consistency checks on (7.60). First, at $t = 0$, the Legendre duality between the external field and the magnetization [7, Lemma 6.7] gives

$$\inf_{m \in L^\infty(\mu;[-1,1])} \left\{ \Psi_m(q_1) + x_1 \mathbb{E}[\theta_1 m(\theta_1)] \right\} = \mathbb{E}\left[\mathcal{B}(q_1, -x_1\theta_1)\right], \tag{7.62}$$

so that $f(0,\cdot)$ is indeed the limit of the right-hand side of (7.53). Second, when $\mu = \delta_0$, a magnetization profile reduces to a number $a \in [-1,1]$, and we have $\overline{m} = 0$ and $\Psi_m = \psi_{\mu_{\mathcal{I}},a} \geqslant \psi_{\mu_{\mathcal{I}},0} = \psi_{\mu_{\mathcal{I}}}$, so that $\Psi_m^* \leqslant \psi_{\mu_{\mathcal{I}}}^*$. The infimum in (7.60) is therefore attained at $m = 0$, and at $x_2 = 0$, we recover the candidate in (7.38), with $\kappa = \mathrm{I}$ and with π the law of y_1.

7.5. **Gaussian mixtures and spherical spins.** We now consider the model of Section 7.4 with spherical spins in place of Ising spins. For this model, the rotational invariance of the sphere allows us to reduce the model to the spherical perceptron with random activation functions of Section 7.3. The reduction consists in sorting the spin configurations according to their projection onto the vector θ; this is the spherical counterpart of the sorting of the Ising configurations according to their magnetization profile performed in Section 7.4. Once this projection is fixed, the remaining coordinates of the spin configuration range in a sphere of one dimension less, and they only interact with the components of the patterns that are orthogonal to θ, which are standard Gaussian vectors. The components of the patterns along θ then only enter the model as parameters of the activation functions, on the same footing as the labels. A similar reduction is used in the proof of [20, Theorem 1.6], in the context of a teacher–student model at zero temperature.

We keep the labels $(y_j)_{j\in\mathbb{N}}$, the random variables $(\theta_i)_{i\in\mathbb{N}}$ of law μ, and the vector $\theta = (\theta_1, \dots, \theta_N)$ of Section 7.4, and we recall that the expectation $\mathbb{E}$ also averages over these random variables. Concerning the activation functions, we only assume here that U_{-1} and U_1 are Lipschitz continuous, and we denote by L the larger of their two Lipschitz constants. We write

$g_j := (g_{ij})_{i \leqslant N}$ for the Gaussian part of the j-th pattern, and we consider the model whose free energy is $-\frac{1}{N}\mathbb{E}\log Z_N$, where

$$Z_N := \int \exp\left(\sum_{j=1}^{M_N} U_{y_j}\left(\frac{g_j \cdot \sigma}{\sqrt{N}} + \frac{y_j}{N}\,\theta \cdot \sigma\right)\right) \mathrm{d}\mu_{\mathbb{S}_N}(\sigma). \tag{7.63}$$

This is the free energy in (7.42) with the measure $\mu_{\mathbb{S}_N}$ in place of $\mu_{\mathcal{I}}^{\otimes N}$.

Rotational invariance. The law of the family $(g_j)_{j \leqslant M_N}$ and the measure $\mu_{\mathbb{S}_N}$ are both invariant under the orthogonal transformations of $\mathbb{R}^N$, and the Gaussian vectors $(g_j)_{j \leqslant M_N}$ are independent of θ and of the labels. The law of Z_N is therefore not modified if we replace θ by the vector $|\theta|\, e_1$ in (7.63), where e_1 denotes the first vector of the canonical basis of $\mathbb{R}^N$. After this replacement, the second term in the argument of U_{y_j} becomes $y_j\, \Theta_N\, \sigma_1/\sqrt{N}$, where $\Theta_N := N^{-1/2}|\theta|$. By the law of large numbers, the random variable Θ_N converges almost surely to

$$\Theta := \left(\mathbb{E}\left[\theta_1^2\right]\right)^{\frac{1}{2}} \tag{7.64}$$

as N tends to infinity. We therefore set

$$\widehat{Z}_N := \int \exp\left(\sum_{j=1}^{M_N} U_{y_j}\left(\frac{g_j \cdot \sigma}{\sqrt{N}} + y_j\, \Theta\, \frac{\sigma_1}{\sqrt{N}}\right)\right) \mathrm{d}\mu_{\mathbb{S}_N}(\sigma). \tag{7.65}$$

Since $|\sigma_1| \leqslant \sqrt{N}$ for every $\sigma \in \mathbb{S}_N$, and since the activation functions are Lipschitz continuous with Lipschitz constant L, we have

$$\left|\frac{1}{N}\mathbb{E}\log Z_N - \frac{1}{N}\mathbb{E}\log \widehat{Z}_N\right| \leqslant \frac{M_N}{N}\, L\, \mathbb{E}\left[|\Theta_N - \Theta|\right]. \tag{7.66}$$

Since $\Theta_N \leqslant \|\theta_1\|_\infty$, the dominated convergence theorem ensures that the right-hand side of (7.66) tends to zero as N tends to infinity. It thus suffices to study the quantity $\widehat{Z}_N$. In particular, the limit of the free energy can only depend on the law μ through its second moment, and the role played by the magnetization profile in Section 7.4 will be played here by a single real parameter.

The model at fixed projection. In the definition of $\widehat{Z}_N$, the vector θ has been replaced by a multiple of e_1, so the projection of a spin configuration onto the direction of the shifts is given by its first coordinate. We now sort the spin configurations according to the value of this coordinate. We assume that $N \geqslant 3$, we set $n := N - 1$, and we write every $\sigma \in \mathbb{S}_N$ such that $|\sigma_1| < \sqrt{N}$ in the form

$$\sigma = \left(\sqrt{N}\, a,\ \sqrt{\frac{N(1-a^2)}{n}}\, \tau\right), \qquad \text{with} \quad a \in (-1,1) \quad \text{and} \quad \tau \in \mathbb{S}_n. \tag{7.67}$$

Under the measure $\mu_{\mathbb{S}_N}$, the random variables a and τ are independent, the law of τ is $\mu_{\mathbb{S}_n}$, and the law of a has the density

$$\beta_N(a) := \frac{\Gamma(N/2)}{\sqrt{\pi}\,\Gamma((N-1)/2)}\,(1-a^2)^{\frac{N-3}{2}} \tag{7.68}$$

with respect to the Lebesgue measure on $(-1,1)$. We decompose the Gaussian vectors accordingly: for every j, we write $g_j = (g_{1j}, \widetilde{g}_j)$, with $\widetilde{g}_j := (g_{ij})_{2 \leqslant i \leqslant N} \in \mathbb{R}^n$. For σ as in (7.67), the argument of U_{y_j} in (7.65) can then be rewritten as

$$\frac{g_j \cdot \sigma}{\sqrt{N}} + y_j\, \Theta\, \frac{\sigma_1}{\sqrt{N}} = \sqrt{1-a^2}\, \frac{\widetilde{g}_j \cdot \tau}{\sqrt{n}} + a\, g_{1j} + a\, \Theta\, y_j. \tag{7.69}$$

For every $a \in (-1,1)$, $y \in \{-1,1\}$ and $\zeta \in \mathbb{R}$, we define the function $U_{y,\zeta}^a : \mathbb{R} \to \mathbb{R}$ by

$$U_{y,\zeta}^a(v) := U_y\left(\sqrt{1-a^2}\, v + a\, \zeta + a\, \Theta\, y\right), \tag{7.70}$$

and we set

$$Z_{N,a} := \int \exp\left(\sum_{j=1}^{M_N} U_{y_j, g_{1j}}^a\left(\frac{\widetilde{g}_j \cdot \tau}{\sqrt{n}}\right)\right) \mathrm{d}\mu_{\mathbb{S}_n}(\tau). \tag{7.71}$$

By (7.69) and the description of the measure $\mu_{\mathbb{S}_N}$ given above, we have

$$\widehat{Z}_N = \int_{-1}^{1} Z_{N,a}\, \beta_N(a)\, \mathrm{d}a. \tag{7.72}$$

For each fixed $a \in (-1,1)$, the quantity $Z_{N,a}$ is the partition function of a spherical perceptron with random activation functions, as considered in Section 7.3. This perceptron has n spins and M_N patterns; its patterns are the standard Gaussian vectors $(\widetilde{g}_j)_{j \leqslant M_N}$, and the activation function applied to the j-th pattern is $U^a_{y_j, g_{1j}}$. The mark of the j-th pattern is thus the pair (y_j, g_{1j}). These marks are i.i.d., and they are independent of the vectors $(\widetilde{g}_j)_{j \leqslant M_N}$. The fact that they are not real-valued, as was assumed in Section 7.3, is immaterial. We stress that the Gaussian random variable g_{1j}, which is a coordinate of the j-th pattern in the original model, has now become a parameter of the activation function, on the same footing as the label y_j. This perceptron falls within the setting considered at the end of Section 7.3, which is also that of [20, Theorem 1.1]. Indeed, the functions in (7.70) are Lipschitz continuous with Lipschitz constant at most L, and since

$$\left|U^a_{y,\zeta}(0)\right| \leqslant \max_{y' \in \{-1,1\}} |U_{y'}(0)| + L\left(|\zeta| + \Theta\right),$$

their values at the origin are integrable with respect to the law of the marks. Moreover, the ratio M_N/n converges to λ.

It remains to identify the pattern functional of this perceptron. In the notation of Section 7.3, it is the function $r \mapsto \mathbb{E}[\mathcal{A}_{U^a_{y_1,\zeta}}(r)]$, where ζ is a standard Gaussian random variable independent of y_1, and where the expectation is over y_1 and ζ. We claim that, for every $r \in K_1$,

$$\mathbb{E}\left[\mathcal{A}_{U^a_{y_1,\zeta}}(r)\right] = \mathcal{A}^{\mathrm{ext}}\left(a^2\mathbf{1} + (1-a^2)r,\, a\Theta\right), \tag{7.73}$$

where $\mathcal{A}^{\mathrm{ext}}$ is the functional defined in (7.51), and where $\mathbf{1}$ denotes the constant path equal to one. Indeed, by the definitions in (7.11) and (7.70), the left-hand side of (7.73) is

$$-\mathbb{E}\log \int \mathbb{E}_{z_1} \exp\left(U_{y_1}\left(a\,\zeta + \sqrt{1-a^2}\, W_1^r(\alpha) + a\,\Theta\, y_1 + \sqrt{(1-a^2)(1-r(1))}\, z_1\right)\right) \mathrm{d}\mathfrak{R}(\alpha),$$

where the expectation $\mathbb{E}$ now also averages over y_1 and ζ. Conditionally on $\mathfrak{R}$, the process $\alpha \mapsto a\,\zeta + \sqrt{1-a^2}\, W_1^r(\alpha)$ is centered Gaussian with covariance $a^2 + (1-a^2)r(\alpha \wedge \alpha')$. It thus has the same law as the process $W_1^{q_2}$ with $q_2 := a^2\mathbf{1} + (1-a^2)r$, and for this choice of q_2, we have $1 - q_2(1) = (1-a^2)(1-r(1))$. Comparing with (7.51), we obtain (7.73). The path $a^2\mathbf{1} + (1-a^2)r$ appearing in this identity has a simple interpretation in terms of overlaps: if σ^1 and σ^2 are two spin configurations of the form (7.67), with the same parameter a and with τ^1 and τ^2 in place of τ, then

$$\frac{\sigma^1 \cdot \sigma^2}{N} = a^2 + (1-a^2)\,\frac{\tau^1 \cdot \tau^2}{n}.$$

For every $a \in (-1,1)$, we set

$$\mathcal{F}(a) := \sup_{r \in K_1} \left\{\lambda\, \mathcal{A}^{\mathrm{ext}}\left(a^2\mathbf{1} + (1-a^2)r,\, a\Theta\right) - \mathsf{S}(r)\right\}, \tag{7.74}$$

where S is the functional defined in (7.30). By (7.73), the quantity $\mathcal{F}(a)$ is the right-hand side of (7.39) for the spherical perceptron with partition function $Z_{N,a}$. Since normalizing by N instead of n does not affect the limit, the discussion of Section 7.3 thus leads to the prediction that, for every $a \in (-1,1)$,

$$\lim_{N\to\infty} -\frac{1}{N}\mathbb{E}\log Z_{N,a} \stackrel{?}{=} \mathcal{F}(a). \tag{7.75}$$

As explained there, this prediction coincides with the formula of [20, Theorem 1.1].

The candidate limit. By (7.72), the quantity $\widehat{Z}_N$ is the integral over $a \in (-1,1)$ of $Z_{N,a}\,\beta_N(a)$, and by (7.68), we have, for every $a \in (-1,1)$,

$$\lim_{N\to\infty} -\frac{1}{N}\log \beta_N(a) = -\frac{1}{2}\log(1-a^2). \tag{7.76}$$

The right-hand side of (7.76) is the entropic cost of the constraint that the projection of the spin configuration be close to a. In view of (7.75), we expect that, on the exponential scale, the function $a \mapsto Z_{N,a}\,\beta_N(a)$ behaves like $\exp\left(-N\left(\mathcal{F}(a) - \frac{1}{2}\log(1-a^2)\right)\right)$, and that its integral is governed by its largest value. As in Section 7.4, and taking into account the minus sign in the definition of the free energy, we thus expect the free energy associated with $\widehat{Z}_N$ to converge to the infimum over a of the sum of this cost and of $\mathcal{F}(a)$. Recalling (7.66), we are led to the following prediction for the limit of the free energy of the model:

(7.77)
$$\lim_{N\to\infty} -\frac{1}{N}\mathbb{E}\log Z_N \stackrel{?}{=} \inf_{a\in(-1,1)}\left\{-\frac{1}{2}\log(1-a^2) + \sup_{r\in K_1}\left\{\lambda\,\mathcal{A}^{\mathrm{ext}}\left(a^2\mathbf{1} + (1-a^2)r,\, a\Theta\right) - \mathsf{S}(r)\right\}\right\}.$$

The passage from (7.75) to (7.77) can be justified as follows. Since the activation functions are Lipschitz continuous, the identity (7.69) and the Cauchy–Schwarz inequality yield that, for every $a, b \in (-1,1)$,

$$\left|\frac{1}{N}\log Z_{N,a} - \frac{1}{N}\log Z_{N,b}\right| \leqslant C_N\sqrt{|a-b|}, \tag{7.78}$$

where

$$C_N := \frac{\sqrt{2}\,L}{N}\left(\sqrt{M_N}\,\|\widetilde{G}\|_{\mathrm{op}} + \sum_{j=1}^{M_N}|g_{1j}| + M_N\Theta\right),$$

and where $\|\widetilde{G}\|_{\mathrm{op}}$ denotes the operator norm of the matrix with M_N rows and n columns whose rows are the vectors $(\widetilde{g}_j)_{j\leqslant M_N}$. Since $\mathbb{E}\|\widetilde{G}\|_{\mathrm{op}} \leqslant \sqrt{M_N} + \sqrt{n}$, see for instance [53, Theorem 7.3.1], the expectation of C_N is bounded uniformly over N. As explained at the end of Section 7.3, the prediction (7.75) is equivalent to the convergence of $-\frac{1}{N}\log Z_{N,a}$ to $\mathcal{F}(a)$ in L^1. If this convergence holds for every $a \in (-1,1)$, then the estimate (7.78) ensures that the function $\mathcal{F}$ is bounded and uniformly continuous, and that

$$\lim_{N\to\infty}\mathbb{E}\left[\sup_{a\in(-1,1)}\left|\frac{1}{N}\log Z_{N,a} + \mathcal{F}(a)\right|\right] = 0.$$

By (7.66), (7.72) and the Laplace method, this yields (7.77).

References

[1] Emmanuel Abbe, Shuangping Li, and Allan Sly. Proof of the contiguity conjecture and lognormal limit for the symmetric perceptron. In *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 327–338. IEEE Computer Society, 2022.

[2] Ayman Ali, Jeanne Boursier, Hong-Bin Chen, Fu-Hsuan Ho, Victor Issa, and Jean-Christophe Mourrat. Free energy of the Hopfield model. *Preprint, arXiv:2611.03306*, 2026.

[3] P. M. Aronow and Patrick Lopatto. The negative spherical perceptron: capacity, isostatic jamming, and full replica symmetry breaking. *Preprint, hexagon:2610.00019v1*, 2026.

[4] Benjamin Aubin, Will Perkins, and Lenka Zdeborová. Storage capacity in symmetric binary perceptrons. *J. Phys. A: Math. Theor.*, 52(29):294003, 2019.

[5] Martino Bardi and Lawrence C Evans. On Hopf's formulas for solutions of Hamilton–Jacobi equations. *Nonlinear Anal.*, 8(11):1373–1381, 1984.

[6] Erwin Bolthausen, Shuta Nakajima, Nike Sun, and Changji Xu. Gardner formula for Ising perceptron models at small densities. In *Proceedings of Thirty Fifth Conference on Learning Theory*, volume 178 of *Proceedings of Machine Learning Research*, pages 1787–1911, 2022.

[7] Jeanne Boursier and Jean-Christophe Mourrat. Free energy of non-convex multi-species Ising spin glasses with external fields. *Preprint, arXiv:2610.18350*, 2026.

[8] Francesco Camilli, Pierluigi Contucci, Federica Gerace, and Emanuele Mingione. Variational bounds for perceptron learning from structured data. *Preprint, arXiv:2608.04882*, 2026.

[9] Hong-Bin Chen. On free energy of non-convex multi-species spin glasses. *ALEA Lat. Am. J. Probab. Math. Stat.*, 23(1):429–473, 2026.

[10] Hong-Bin Chen, Fu-Hsuan Ho, and Victor Issa. A formula for the free energy of non-convex vector spin glasses. *Preprint, arXiv:2610.18744*, 2026.

[11] Hong-Bin Chen, Victor Issa, and Jean-Christophe Mourrat. Free energy of non-convex multi-species spin glasses with centered Ising spins. *Preprint, arXiv:2606.16636*, 2026.

[12] 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.

[13] Hong-Bin Chen and Jean-Christophe Mourrat. Free energy of non-convex multi-species spherical spin glasses. *Preprint, arXiv:2609.32710*, 2026.
[14] Hong-Bin Chen and Jiaming Xia. Hamilton–Jacobi equations from mean-field spin glasses. *Probab. Theory Related Fields*, 192(3):803–873, 2025.
[15] Hong-Bin Chen and Jiaming Xia. Hamilton–Jacobi equations with monotone nonlinearities on convex cones. *Ann. Fac. Sci. Toulouse Math. (6)*, 34(1):1–45, 2025.
[16] Jian Ding and Nike Sun. Capacity lower bound for the Ising perceptron. *Probab. Theory Related Fields*, 193:627–715, 2025.
[17] 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.
[18] Franziska Engler and Justin Ko. Synchronization of enriched spin glasses. *Preprint, arXiv:2611.03587*, 2026.
[19] Lawrence C Evans. *Partial Differential Equations*, volume 19. American Mathematical Soc., 2010.
[20] Daniel Fu and Youngtak Sohn. Parisi formula for Ising and spherical perceptrons. *Preprint, hexagon:2610.00154v1*, 2026.
[21] E. Gardner. The space of interactions in neural network models. *J. Phys. A*, 21(1):257–270, 1988.
[22] E. Gardner and B. Derrida. Optimal storage properties of neural network models. *J. Phys. A*, 21(1):271–284, 1988.
[23] Géza Györgyi. Techniques of replica symmetry breaking and the storage problem of the McCulloch–Pitts neuron. *Phys. Rep.*, 342(4-5):263–392, 2001.
[24] Géza Györgyi and Peter Reimann. Parisi phase in a neuron. *Phys. Rev. Lett.*, 79(14):2746–2749, 1997.
[25] Géza Györgyi and Peter Reimann. Beyond storage capacity in a single model neuron: continuous replica symmetry breaking. *J. Stat. Phys.*, 101(1-2):679–702, 2000.
[26] Fu-Hsuan Ho. A concavity theorem for the Parisi PDE. *Preprint, arXiv:2606.15432*, 2026.
[27] Fu-Hsuan Ho. Convergence of the free energy of vector spin glasses with non-convex interactions. *Preprint, arXiv:2609.39307*, 2026.
[28] Brice Huang. Capacity threshold for the Ising perceptron. In *2024 IEEE 65th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1126–1136, 2024.
[29] Jeong Han Kim and James R. Roche. Covering cubes by random half cubes, with applications to binary neural networks. *J. Comput. System Sci.*, 56(2):223–252, 1998.
[30] Werner Krauth and Marc Mézard. Storage capacity of memory networks with binary couplings. *J. Physique*, 50(20):3057–3066, 1989.
[31] P-L Lions and J-C Rochet. Hopf formula and multitime Hamilton–Jacobi equations. *Proc. Amer. Math. Soc.*, 96(1):79–84, 1986.
[32] Marc Mézard. The space of interactions in neural networks: Gardner's computation with the cavity method. *J. Phys. A: Math. Gen.*, 22(12):2181–2190, 1989.
[33] Andrea Montanari and Kangjie Zhou. Replica symmetry breaking and algorithmic thresholds in empirical risk minimization under multi-index model. *Preprint, arXiv:2606.28573*, 2026.
[34] Jean-Christophe Mourrat. Nonconvex interactions in mean-field spin glasses. *Probab. Math. Phys.*, 2(2):281–339, 2021.
[35] Jean-Christophe Mourrat. The Parisi formula is a Hamilton–Jacobi equation in Wasserstein space. *Canad. J. Math.*, 74(3):607–629, 2022.
[36] Jean-Christophe Mourrat. Free energy upper bound for mean-field vector spin glasses. *Ann. Inst. Henri Poincaré Probab. Stat.*, 59(3):1143–1182, 2023.
[37] Jean-Christophe Mourrat and Dmitry Panchenko. Extending the Parisi formula along a Hamilton–Jacobi equation. *Electron. J. Probab.*, 25:Paper No. 23, 17, 2020.
[38] Shuta Nakajima and Nike Sun. Sharp threshold sequence and universality for Ising perceptron models. In *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 638–674, 2023.
[39] OpenAI. The free energy of the Ising random perceptron. OpenAI Math Release preprint OAI:The-free-energy-of-the-Ising-random-perceptron-September-24-2026, 2026.
[40] OpenAI. The free energy of the spherical random perceptron. OpenAI Math Release preprint OAI:The-free-energy-of-the-spherical-random-perceptron-September-24-2026, 2026.
[41] Dmitry Panchenko. The free energy in a multi-species Sherrington–Kirkpatrick model. *Ann. Probab.*, 43(6):3494–3513, 2015.
[42] Dmitry Panchenko. Free energy in the Potts spin glass. *Ann. Probab.*, 46(2):829–864, 2018.
[43] Dmitry Panchenko. Free energy in the mixed p-spin models with vector spins. *Ann. Probab.*, 46(2):865–896, 2018.
[44] Will Perkins and Changji Xu. Frozen 1-RSB structure of the symmetric Ising perceptron. *Random Structures Algorithms*, 64(4):856–877, 2024.
[45] Mariya Shcherbina and Brunello Tirozzi. Rigorous solution of the Gardner problem. *Comm. Math. Phys.*, 234(3):383–422, 2003.
[46] Yitzchak Shmalo. The storage capacity of the Ising perceptron: verification of the outstanding numerical conditions. *Preprint, arXiv:2607.11010*, 2026.
[47] Charles Stegall. Optimization of functions on certain subsets of Banach spaces. *Math. Ann.*, 236(2):171–176, 1978.
[48] Michel Talagrand. Intersecting random half cubes. *Random Structures Algorithms*, 15(3-4):436–449, 1999.

[49] Michel Talagrand. Intersecting random half-spaces: toward the Gardner–Derrida formula. *Ann. Probab.*, 28(2):725–758, 2000.
[50] Michel Talagrand. On the Gaussian perceptron at high temperature. *Math. Phys. Anal. Geom.*, 5(1):77–99, 2002.
[51] Michel Talagrand. *Mean field models for spin glasses. Volume II*, volume 55 of *Ergebnisse der Mathematik und ihrer Grenzgebiete*. Springer, Heidelberg, 2011.
[52] Jérôme Tubiana and Rémi Monasson. Emergence of compositional representations in restricted Boltzmann machines. *Phys. Rev. Lett.*, 118(13):138301, 2017.
[53] Roman Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press, 2018.
[54] Changji Xu. Sharp threshold for the Ising perceptron model. *Ann. Probab.*, 49(5):2399–2415, 2021.

(Hong-Bin Chen) NYU-ECNU Institute of Mathematical Sciences, NYU Shanghai, China

(Jean-Christophe Mourrat) Department of Mathematics, ENS Lyon and CNRS, Lyon, France