On the Parisi PDE and the Overlap Gap Property in Ising Spin Glasses

Utkarsh Utkarsh
MIT
utkarsh5@mit.edu

Theodoros Xenakis
MIT
txenakis@mit.edu

Abstract

We study the overlap gap property (OGP) for the pure p-spin Ising spin glass through the lens of the zero-temperature Parisi variational principle. The ground-state energy is characterized by minimizing the Parisi functional over a class of nondecreasing functions γ, whose flat regions correspond to forbidden intervals in the limiting overlap distribution—the PDE-level signature of OGP. We prove that for every pure $p > 2$ model, no minimizer of the Parisi functional can be strictly increasing, establishing a PDE-level gap. The argument builds on the work of El Alaoui, Montanari, and Sellke [EAMS21] and reduces to a scaling incompatibility between a stationarity identity and the regularity of the Parisi PDE near $t = 0$. We also propose a conjectural upgrade to a finite-N binary OGP via constrained two-replica energies. On the numerical side, we introduce a practical simulation framework based on the Hopf–Cole formula and Gauss–Hermite quadrature, which avoids the exponential cost of nested Monte Carlo and reduces PDE solves to $O(KN_xQ)$ operations. Experiments for $p = 2, \ldots, 7$ confirm one-step replica symmetry breaking (1RSB) for $p \geq 3$ and full replica symmetry breaking (FRSB) for $p = 2$. Analysis of the Hessian loss landscape reveals that the condition number grows from ≈ 7.6 at $p = 2$ to $\sim 10^{13}$ at $p = 7$, providing a sharp numerical signature of the 1RSB structure.

1 Introduction

In recent years, high-dimensional optimization problems have gained renewed interest due to their connections with machine learning, statistics, and statistical physics [Gam21, CHM$^+$15]. Training a modern neural network, for instance, typically involves minimizing a highly non-convex loss function over a parameter space containing millions or even billions of parameters. Despite the empirical success of such optimization procedures, the geometric structure of near-optimal solutions remains poorly understood [CHL18, Gam21].

A recurring observation is that many high-dimensional random optimization problems exhibit very structured landscapes[CHL18, Gam21]. Rather than forming a single connected region, near-optimal solutions often gather into separated clusters [Gam21, GJS21]. Understanding this geometry has become a central question in both statistical physics and theoretical computer science, as it appears closely tied to the computational difficulty of finding optimal solutions [Gam21, GS14, GS17].

Spin glass models, stemming from statistical physics, provide a natural mathematical framework for studying such complex loss landscapes. Particularly, Choromanska et al. argues in [CHM$^+$15] that the loss landscapes of neural networks often exhibit similarities to spin glass models. This connection has motivated the use of spin glass theory to study geometric and algorithmic phenomena arising in high-dimensional optimization problems.

One of the most striking geometric phenomena observed in spin glasses is the overlap gap property (OGP)[Gam21]. Informally, OGP states that near-optimal configurations are either highly correlated,

or essentially unrelated; there is no middle ground. This implies a fragmented geometry in the space of near optimal solutions, and has been connected to barriers in the optimization algorithms used [GS14, GS17, HS22].

The mathematical description of this geometry is encoded in the Parisi variational principle, which characterizes the asymptotic ground-state energy of mean-field spin glasses through the solution of an infinite-dimensional, nonlinear partial differential equation (PDE), known as the Parisi PDE [Tal06, AC17]. Specifically, the asymptotic ground state energy is the minimizer of a functional of the solution of this PDE. The minimizer is typically referred to as the Parisi order parameter.

In this work, we study the zero-temperature Parisi variational problem for the pure p-spin Ising spin glass, with emphasis on $p = 3$. Our primary goal is to understand how overlap gap phenomena manifest at the level of the Parisi order parameter γ^*. A flat region of γ^* - that is, an interval (a, b) on which $d\gamma^*$ assigns no mass, is the PDE-level signature of OGP: overlap values in (a, b) are absent from the limiting overlap distribution of near-ground-state configurations.

We prove that for every pure $p > 2$ model, no minimizer γ^* can be strictly increasing (Theorem 5.1). The argument uses three lemmas from [EAMS21] and reduces to a single scaling observation: the stationarity identity $Q(t) = t$ is incompatible with the $O(t^{p-1})$ growth forced by the Parisi martingale whenever $p > 2$. We also discuss a conjectural upgrade of this PDE-level gap focusing on $p = 3$ case to a finite-N binary OGP statement via constrained two-replica energies (Section 6).

On the numerical side, we discretize γ as a K-step piecewise-constant function and solve the Parisi PDE using the Hopf–Cole formula with Gauss–Hermite quadrature, achieving $O(KN_xQ)$ complexity per evaluation. We optimize the Parisi functional with L-BFGS and forward-mode automatic differentiation. Our experiments confirm that the optimizer is step-like (1RSB) for $p \geq 3$ and smooth (FRSB) for $p = 2$. We additionally analyze the loss landscape of the Parisi optimization by projecting onto the top Hessian eigenvectors at optimality, finding that the Hessian condition number grows from ≈ 7.6 at $p = 2$ to $\sim 10^{13}$ at $p = 7$: a sharp numerical signature of the near-rank-1 structure of 1RSB minimizers.[1]

1.1 Outline

Section 3 introduces the overlap gap property. We define the OGP formally, explain its connection to algorithmic hardness via the stability argument, and establish the dictionary: a flat interval in the Parisi order parameter γ^* is the PDE-level signature of OGP, while a strictly increasing γ^* corresponds to no gap.

Section 4 introduces the Parisi PDE and the variational principle. We state the Parisi functional $P(\gamma)$ and recall the theorem of Auffinger–Chen [AC17] that $\inf_{\gamma \in \mathcal{U}} P(\gamma)$ gives the asymptotic ground-state energy.

Sections 5 and 6 contain the main theoretical contribution. We state and prove Theorem 5.1: for every pure $p > 2$, no minimizer γ^* of the Parisi functional can be strictly increasing. The proof uses the Parisi diffusion as a martingale, applies three lemmas from [EAMS21], and derives a contradiction from the incompatible scaling $Q(t) = t$ versus $Q(t) = O(t^{p-1})$. We also discuss why the $p = 2$ (SK) case is the critical borderline where the argument breaks down, and propose a conjectural upgrade to a full finite-N binary OGP via constrained two-replica energies.

Section 7 presents the numerical experiments. We describe the discretization of γ as a K-step function, the Hopf–Cole solution of the Parisi PDE, and the Gauss–Hermite quadrature scheme that reduces the exponential-in-K cost of nested expectations to $O(KN_xQ)$. We report the optimized Parisi functionals and order parameters for $p = 2, \ldots, 7$, confirming that γ^* is step-like, i.e., one-step replica symmetry breaking (1RSB [MPVT88]), for $p \geq 3$, and smooth, i.e., full replica symmetry breaking (FRSB [MPVT88]), for $p = 2$. We also analyze the loss landscape of the Parisi optimization by projecting onto the top two Hessian eigenvectors at the optimum following [LXT+18], observing that the condition number grows by ten orders of magnitude from $p = 2$ to $p = 7$.

[1] The code used for the numerical experiments is publicly available on https://github.com/xenakistheo/SpinGlasses

2 Physical Background

Spin glass models stem from different physical phenomena, of which the simplest one might be a set of N magnets arranged in a two-dimensional grid. Imagine each magnet having two possible arrangements: either with the negative pole or the positive pole facing rightwards. This configuration of the magnet is called the *spin* and can be represented by either -1 or $+1$. It follows that any given configuration of the magnets can be represented by the variable $\sigma \in \Sigma_N$. Where we define the hypercube $\Sigma_N = \{-1, +1\}^N$.

The main motivation of studying spin-glass models can be said to find the configuration that minimizes the total energy. To formalize this, define the *Hamiltonian*. That is, a function

$$H_N : \Sigma_N \to \mathbb{R}$$

that represents the (negative) *energy* of a system with configuration $\sigma \in \Sigma_N$. The overlying motivation is finding[2]

$$\sigma^* = \arg\max_{\sigma \in \Sigma_N} H_N(\sigma)$$

However, what characterizes spin glasses is that there is an inherent degree of stochasticity involved. One can think of this as arising from unknown external factors that affect the energy of the magnet-grid. For example, if the grid of magnets was placed in an external electromagnetic field, then the optimal energy-configuration would highly depend on this external field. In particular, there are some random variable weights $W \in Q$, so that the Hamiltonian is also a function of these random variables $H_N(\sigma) = H_N(\sigma; W)$. The space Q is typically some higher-order vector space.

A typical example of a Hamiltonian might be

$$H_N^{\text{elementary}}(\sigma) = \frac{1}{\sqrt{N}} \langle \sigma, W\sigma \rangle \tag{1}$$

where we define the symmetric matrix $W \in \mathbb{R}^{N \times N}$, with Gaussian entries $W_{ij} = W_{ji} \sim \mathcal{N}(0, 1)$ for all $1 \leq i \leq j \leq N$. These weights are sometimes referred to as the *disorder tensor*.

The problem one is then interested in asking is, *given a specific state of randomness, summarized by W: what is the optimal configuration σ that maximizes the Hamiltonian?*

2.1 Hardness

One can recognize that for the Ising spin glass model, there are only 2^N possible configurations for σ. In other words, there is always the choice of brute-forcing the problem in exponential time to find a configurations that deterministically is correct. It turns out, however, that deterministically finding this optimal configuration for a given set of weights is an NP-Hard problem [ABH+05]. We therefore turn our focus to instead leverage the structure and random distribution of the disorder tensor. If we can make certain assumptions on the weights we can instead opt for creating an algorithm that finds an optimal (or near optimal) solution *with high probability*. This also often the interest in the spin glass community today. Defining,

$$\text{OPT}_N := \frac{1}{N} \max\{H_N(\sigma) : \sigma \in \Sigma_N\} \tag{2}$$

we ask, given an $\epsilon > 0$, is there a polynomial time algorithm for finding $\hat{\sigma}$ so that

$$\frac{1}{N} H_N(\hat{\sigma}) \geq (1 - \epsilon)\text{OPT}_N \quad \textit{with high probability?}$$

Rather than attempting to characterize a single optimizer σ^*, we instead shift our focus to study the statistical structure of the energy landscape itself. In particular, we seek to understand how near-optimal configurations are organized geometrically.

[2] In the general problem, Σ_N need not be a binary hypercube. In many analyses σ instead has a restriction on its L_2-norm. Such a model is typically referred to as spherical spin glass model. Our focus is on the hypercube, which is referred to as Ising spin glass model.

As we will see, a remarkable feature of spin glass models is that the geometry of the landscape is encoded in the correlation structure of the Hamiltonian. Configurations with similar spin arrangements exhibit correlated energies, while sufficiently different configurations behave almost independently. This motivates studying the covariance structure of the Hamiltonian, which naturally leads to the idea of overlap between configurations.

2.2 Interactions, Mixtures, and the p-spin model

Expanding the inner product in (1), yields

$$H_N^{\text{elementary}}(\sigma) = \frac{1}{\sqrt{N}} \sum_{i,j} W_{ij}\sigma_i\sigma_j$$

In particular, the Hamiltonian contains only second-order interactions. That is, there is only a multiplication with two different elements of σ present. In the general case, one can have arbitrarily different interactions. Physically, higher-order interactions correspond to situations in which the energy of the system depends simultaneously on larger groups of spins rather than only on pairwise interactions.

In this work we adopt the notation and specific model defined in [EAMS21]. Namely, let the weights be a collection of tensors $W^{(2)}, W^{(3)}, ..., W^{(p)}$, where p is the highest order interaction (it is allowed for $p = \infty$). Further let $W^{(k)}$ be a standard symmetric Gaussian tensor of order k. That is, $W^{(k)}$ has entries $(W^{(k)}_{i_1,...,i_k})_{1\le i_1,...,i_k\le N}$, with

$$W^{(k)}_{i_1,...,i_k} = W^{(k)}_{\pi(i_1,...,i_k)} \sim \mathcal{N}\left(0, \frac{k!}{N^{k-1}}\right)$$

A key here is that for two sets of indices $(i_1, ..., i_p), (i'_1, ..., i'_p)$, if one is not a permutation of the other, then the entries in W characterized by those two sets of indices are independent.

$$(i_1, ..., i_p) \neq \pi(i'_1, ..., i'_p) \Rightarrow W^{(k)}_{i_1,...,i_k} \perp W^{(k)}_{i'_1,...,i'_k}$$

We define the Hamiltonian of this system as

$$H_N(\sigma) = \sum_{k=2}^{p} \frac{c_k}{k!} \langle W^{(k)}, \sigma^{\otimes k}\rangle \tag{3}$$

$$\langle W^{(k)}, \sigma^{\otimes k}\rangle = \sum_{1\le i_1,...,i_k\le N} W_{i_1,...,i_k}\sigma_{i_1} \cdot ... \cdot \sigma_{i_k} \tag{4}$$

For some constants $c_1, ..., c_p$. These constants define the so called *mixture* of the model, and are also used in the *mixture function*

$$\xi(x) := \sum_{k=2}^{\infty} c_k^2 x^k \tag{5}$$

The mixture function completely determines the covariance structure of the Hamiltonian.

We will focus on the third order pure spin model. That is $c_k = 0, \forall k \neq 3$. For simplicity, we set $c_3 = 1$. That is, we only have a single tensor $W = W^{(3)} \in \mathbb{R}^{N\times N\times N}$, and the mixture reduces to $\xi(x) = x^3$. Consequently, the Hamiltonian becomes

$$H_N(\sigma) = \frac{1}{3!} \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{N} W_{ijl}\sigma_i\sigma_j\sigma_l$$

With each entry of W being mean-zero gaussian with variance

$$\mathbb{E}[W_{ijl}^2] = \frac{6}{N^2}$$

2.2.1 Covariance

For two different configurations $\tilde{\sigma}, \hat{\sigma} \in \Sigma_N$, define their overlap by

$$R(\tilde{\sigma}, \hat{\sigma}) := \frac{1}{N}\langle \tilde{\sigma}, \hat{\sigma} \rangle \tag{6}$$

The overlap measures the normalized similarity between two spin configurations. It can be shown in the general case - for an arbitrary mixture function and two different spin configurations $\sigma, \hat{\sigma}$ that the correlation between the energy of the two different configurations depends only on this overlap quantity. Specifically,

$$\mathbb{E}[H_N(\tilde{\sigma}) \cdot H_N(\hat{\sigma})] = N\xi(R(\tilde{\sigma}, \hat{\sigma})) \tag{7}$$

We show it here for our specific model choice - $\xi(x) = x^3$.

$$\mathbb{E}[H_N(\sigma) \cdot H_N(\hat{\sigma})] = \frac{c^2}{6^2} \sum_{i,j,l} \sum_{i',j',l'} \mathbb{E}[W_{ijl} W_{i'j'l'}] \sigma_i \sigma_j \sigma_l \hat{\sigma}_{i'} \hat{\sigma}_{j'} \hat{\sigma}_{l'}$$

By the independence of the weight-entries, we the expectation is zero whenever $(i', j', l') \neq \pi(i, j, l)$. There are six different ways to permute those three indices, so we get that the expectation reduces down to

$$\begin{aligned}
\mathbb{E}[H_N(\sigma) \cdot H_N(\hat{\sigma})] &= 6 \cdot \frac{1}{6^2} \sum_{i,j,l} \mathbb{E}[W_{ijl}^2] \sigma_i \sigma_j \sigma_l \hat{\sigma}_i \hat{\sigma}_j \hat{\sigma}_l \\
&= \frac{1}{6} \frac{6}{N^2} \sum_{i,j,l} \sigma_i \sigma_j \sigma_l \hat{\sigma}_i \hat{\sigma}_j \hat{\sigma}_l = \frac{1}{N^2} \left(\sum_i \sigma_i \hat{\sigma}_i \right)^3 \\
&= \frac{1}{N^2} \langle \sigma, \hat{\sigma} \rangle^3 = N\xi(\langle \sigma, \hat{\sigma} \rangle / N)
\end{aligned}$$

This identity suggests that overlap plays a central role in the geometry of the landscape. A natural question to ask, therefore, is how the overlaps between near-optimal configurations are distributed. One might initially expect that near-optimal solutions form a connected family, allowing all intermediate overlap values to occur. However, this is often not the case with spin glass models.

3 Overlap Gap Property

The overlap gap property (OGP) is a way of describing the geometry of the near-optimal solution space of a random optimization problem. Recall from Section 1 the Ising p-spin model and the overlap $R(\sigma, \tau) = \frac{1}{N}\langle \sigma, \tau \rangle$ (equation (6)). If $R(\sigma, \tau) \approx 1$ the configurations are very close; if $R(\sigma, \tau) \approx 0$ they are essentially uncorrelated.

Informally, OGP says that the set of near-optimal configurations has a "forbidden" range of overlaps. That is, there exist constants $0 < a < b < 1$ such that two near-optimal configurations can have either small overlap, roughly below a, or large overlap, roughly above b, but cannot have overlap in the intermediate interval (a, b). Thus the near-optimal solution space is not connected in overlap space. It has a clustered geometry: solutions are either very close to each other or far apart, with no continuous path of near-optimal solutions connecting the two regimes.

This geometric picture is important because it suggests an algorithmic barrier. Many natural algorithms are stable: if the random instance is changed slightly, the output changes only slightly. If such an algorithm could always find near-optimal configurations, then under a continuous interpolation of the random instance, the overlaps between outputs should vary continuously from large to small values. The OGP forbids the intermediate overlap values that such a continuous path would have to pass through. In this sense, OGP provides a mechanism explaining why stable algorithms may fail to reach the true optimum.

In the Parisi PDE formulation (see Section 4), the object we study is not directly the finite N set of configurations, but rather the limiting order parameter. At positive temperature, this order parameter is a probability measure on the overlap coordinate, or equivalently its cumulative distribution function.

At zero temperature, following the [EAMS21] framework, the optimizer is a monotone integrable function

$$\gamma^* : [0,1) \to \mathbb{R}_{\geq 0}, \qquad \gamma^* \in \mathscr{U},$$

which should be viewed as the zero-temperature analogue of the Parisi CDF. It is not necessarily a bounded CDF, but it is still nondecreasing and therefore has an associated Stieltjes measure $d\gamma^*$.

The relationship between γ^* and OGP is that γ^* encodes the limiting overlap structure predicted by the Parisi variational problem. If γ^* is strictly increasing, then there is no interval of overlap values missing from the Parisi order parameter. This is the PDE-level analogue of a no-OGP situation. On the other hand, if γ^* is constant on a nontrivial interval (a, b), then

$$d\gamma^*((a,b)) = 0.$$

This means that the Parisi order parameter assigns no mass to the overlap interval (a, b). In the language of the project, this is precisely the PDE-level manifestation of an overlap gap.

Thus, in this project, we use the following dictionary:

$$\gamma^* \text{ strictly increasing} \quad \Longleftrightarrow \quad \text{no PDE-level OGP},$$

while

$$\gamma^* \text{ has a flat interval} \quad \Longleftrightarrow \quad \text{PDE-level OGP}.$$

The goal is therefore to show that for the pure $p = 3$ model, the zero-temperature Parisi minimizer γ^* cannot be strictly increasing. Since γ^* is nondecreasing, failure of strict increase implies that there exist $0 \leq a < b < 1$ such that

$$\gamma^*(a) = \gamma^*(b),$$

and hence, by monotonicity,

$$\gamma^*(t) = \gamma^*(a) \qquad \forall t \in [a, b].$$

Equivalently,

$$d\gamma^*((a,b)) = 0.$$

This is the gap in the PDE optimizer that corresponds to OGP in the sense of the project statement.

It is useful to emphasize the distinction between this PDE-level statement and a full finite-N OGP theorem. A full finite-N OGP theorem would assert that, with high probability, actual near-optimal configurations $\sigma, \tau \in \{\pm 1\}^N$ cannot have overlap in some interval (a, b). The PDE-level gap is the limiting variational signature of such a phenomenon, but upgrading it to a full finite-N statement generally requires additional machinery, such as coupled interpolation or constrained two-replica variational arguments. In the present proof, we focus on the PDE-level gap: showing that the Parisi minimizer γ^* has a flat interval.

4 The Parisi Variational Principle

Spin glasses exhibit an interesting property. While finding the exact optimizer of the Hamiltonian is computationally difficult, the asymptotic value (as $N \to \infty$) of the optimum can be characterized remarkably precisely through a variational descriptio [Tal06].

Specifically - at zero temperature - the problem of finding this asymptotic optimum, reduces to minimizing a nonlinear functional over a class of monotone functions. Namely, define the Parisi partial differential equation boundary value problem as

$$\partial_t \Phi_\gamma(t,x) + \frac{1}{2}\xi''(t)\left(\partial_x^2 \Phi_\gamma(t,x) + \gamma(t)(\partial_x \Phi_\gamma(t,x))^2\right) = 0 \tag{8}$$

$$\Phi_\gamma(1,x) = |x| \tag{9}$$

Here we define the function $\Phi_\gamma : [0,1] \times \mathbb{R} \to \mathbb{R}$ to be the solution of the PDE, for $\gamma \in \mathcal{U}$. Where we define the space of functions

$$\mathcal{U} := \left\{ \gamma : [0,1] \to \mathbb{R}_{\geq 0} : \quad \gamma \text{ non decreasing}, \int_0^1 \gamma(t)\, dt < \infty \right\}$$

Informally, the PDE propagates information about terminal energy landscape backward from $t = 1$ to $t = 0$. In practice this means that once we have solved the PDE, we can derive the Parisi functional

$$P(\gamma) := \Phi_\gamma(0,0) - \frac{1}{2}\int_0^1 t\xi''(t)\gamma(t)\,dt \tag{10}$$

The first term $\Phi_\gamma(0,0)$ represents the effective energy predicted by the PDE dynamics, while the second term penalizes the complexity of the overlap structure encoded by γ. The Parisi variational principle states that the asymptotic ground-state energy is obtained by balancing these two competing effects optimally.

In particular, Auffinger and Chen [AC17] have remarkably proved that the following limit holds almost surely.

$$\lim_{N\to\infty} \text{OPT}_N = \inf_{\gamma\in\mathcal{U}} P(\gamma) \tag{11}$$

The variational principle suggests that the structure of the optimizer γ^* that solves (11), encodes geometric information about the spin glass landscape.

5 A Short Proof of a PDE-Level Gap for Pure $p > 2$

5.1 Setup

We use the framework of Section 4: the order parameter space $\mathscr{U}$, the Parisi PDE (9), the functional P, and the pure p-spin specialization $\xi(t) = t^p$, $\xi''(t) = p(p-1)t^{p-2}$ are all as defined there. Let $\gamma^* \in \mathscr{U}$ be any minimizer of P and write $\Phi = \Phi_{\gamma^*}$. A gap in γ^* means it is constant on a nontrivial interval, equivalently $d\gamma^*$ assigns no mass to that interval. Our goal is to show that for every pure $p > 2$, no minimizer can be strictly increasing, which forces such a gap. We formalize this in the following theorem.

Theorem 5.1 (PDE-level gap for pure $p > 2$)**.** *For every pure $p > 2$, no zero-temperature Parisi minimizer $\gamma^* \in \mathscr{U}$ can be strictly increasing on $[0,1)$. Consequently, every minimizer has a nontrivial flat interval. That is, there exists $0 \le a < b \le 1$ such that*

$$d\gamma^*((a,b)) = 0.$$

In particular, the conclusion applies to $p = 3$.

In the terminology of the project, Theorem 5.1 is the PDE-level OGP statement: the optimal Parisi order parameter has a gap.

5.2 Why introduce the Parisi diffusion?

Before proving the theorem, it is useful to explain the role of the Parisi diffusion. Let

$$u(t,x) := \Phi_x(t,x).$$

Differentiating the Parisi PDE in x, we get

$$u_t + \frac{1}{2}\xi''(t)u_{xx} + \xi''(t)\gamma^*(t)uu_x = 0.$$

Thus u solves a viscous transport equation whose transport velocity is

$$\xi''(t)\gamma^*(t)u(t,x).$$

This motivates defining the stochastic characteristic

$$dX_t = \xi''(t)\gamma^*(t)\Phi_x(t,X_t)\,dt + \sqrt{\xi''(t)}\,dB_t, \qquad X_0 = 0.$$

With this choice of drift and diffusion coefficient, Itô's formula applied to

$$Y_t := \Phi_x(t,X_t)$$

gives

$$dY_t = \sqrt{\xi''(t)}\Phi_{xx}(t,X_t)\,dB_t.$$

The drift cancels exactly because Φ solves the Parisi PDE. Hence Y_t is a martingale, and by Itô's isometry,

$$Q(t) := \mathbb{E}[Y_t^2] = \int_0^t \xi''(s)\mathbb{E}[\Phi_{xx}(s, X_s)^2]\, ds.$$

The quantity $Q(t)$ is the bridge between the variational optimality condition and the PDE regularity estimate. Under the no-gap, or strict-increase, assumption, Lemma 6.15 of [EAMS21] gives $Q(t) = t$. On the other hand, the martingale identity and local boundedness of Φ_{xx} force $Q(t)$ to grow like t^{p-1} for a pure p-spin model. For $p > 2$, this is incompatible with $Q(t) = t$ near $t = 0$.

A more complete derivation of derivation between the Parisi PDE and the associated stochastic differential equation can be found in A

5.3 Inputs from [EAMS21]

We use three inputs from [EAMS21].

Lemma 5.2 (Lemma 6.15 of [EAMS21]: no-gap stationarity)**.** *Assume the no-overlap-gap condition holds, namely there exists a strictly increasing minimizer $\gamma^* \in \mathscr{U}$ of the zero-temperature Parisi functional. Let X_t be the Parisi diffusion associated with γ^*. Then*

$$\mathbb{E}\left[\Phi_x(t, X_t)^2\right] = t \qquad \forall t \in [0, 1).$$

Lemma 5.3 (Corollary 6.6 of [EAMS21]: martingale identity)**.** *Let X_t be the Parisi diffusion*

$$dX_t = \xi''(t)\gamma^*(t)\Phi_x(t, X_t)\, dt + \sqrt{\xi''(t)}\, dB_t, \qquad X_0 = 0.$$

Define

$$Q(t) := \mathbb{E}\left[\Phi_x(t, X_t)^2\right].$$

Then

$$Q(t) = \int_0^t \xi''(s)\mathbb{E}\left[\Phi_{xx}(s, X_s)^2\right]\, ds.$$

Lemma 5.4 (Lemma 6.4 of [EAMS21]: local regularity)**.** *For every $t_0 < 1$,*

$$\Phi_{xx} \in L^\infty([0, t_0]; L^\infty(\mathbb{R})).$$

Equivalently, there exists $C_{t_0} < \infty$ such that

$$|\Phi_{xx}(t, x)| \le C_{t_0}$$

for almost every $(t, x) \in [0, t_0] \times \mathbb{R}$. Consequently,

$$\mathbb{E}\left[\Phi_{xx}(t, X_t)^2\right] \le C_{t_0}^2$$

for almost every $t \in [0, t_0]$, which is sufficient in the integral identity below.

5.4 Proof of Theorem 5.1

Proof. We prove the theorem by contradiction. Suppose the no-overlap-gap condition holds. That is, suppose there exists a strictly increasing minimizer

$$\gamma^* \in \mathscr{U}.$$

Let X_t be the associated Parisi diffusion and define

$$Q(t) := \mathbb{E}\left[\Phi_x(t, X_t)^2\right].$$

By Lemma 5.2,

$$Q(t) = t \qquad \forall t \in [0, 1). \tag{1}$$

On the other hand, by Lemma 5.3,

$$Q(t) = \int_0^t \xi''(s)\mathbb{E}\left[\Phi_{xx}(s, X_s)^2\right]\, ds. \tag{2}$$

For the pure p-spin model,

$$\xi''(s) = p(p-1)s^{p-2}.$$

Fix $t_0 < 1$. By Lemma 5.4, there exists $C_{t_0} < \infty$ such that

$$\mathbb{E}\left[\Phi_{xx}(s, X_s)^2\right] \le C_{t_0}^2$$

for almost every $s \in [0, t_0]$. Therefore, for $0 < t \le t_0$,

$$Q(t) \le p(p-1)C_{t_0}^2 \int_0^t s^{p-2}\, ds.$$

Since

$$\int_0^t s^{p-2}\, ds = \frac{t^{p-1}}{p-1},$$

we obtain

$$Q(t) \le pC_{t_0}^2 t^{p-1}. \tag{3}$$

For $p > 2$, the right-hand side is $o(t)$ as $t \downarrow 0$. Combining (1) and (3), we get

$$t = Q(t) \le pC_{t_0}^2 t^{p-1}.$$

Dividing by $t > 0$,

$$1 \le pC_{t_0}^2 t^{p-2}.$$

This is impossible for sufficiently small $t > 0$, since $p > 2$.

Therefore no strictly increasing minimizer $\gamma^* \in \mathscr{U}$ can exist for any pure $p > 2$. In particular, for $p = 3$, there is no strictly increasing zero-temperature Parisi minimizer.

Since γ^* is monotone, if it is not strictly increasing then there exist $0 \le a < b < 1$ such that

$$\gamma^*(a) = \gamma^*(b).$$

Which implies

$$\gamma^*(t) = \gamma^*(a) \qquad \forall t \in [a, b].$$

Hence γ^* is constant on a nontrivial interval, or equivalently,

$$d\gamma^*((a, b)) = 0.$$

This proves the desired PDE-level gap. □

5.5 Remarks

Endpoint scaling is the key observation. Beyond the three inputs from [EAMS21], the only additional observation is the small-t scaling

$$\xi''(t) = p(p-1)t^{p-2}.$$

For $p > 2$, the martingale identity forces

$$Q(t) = O(t^{p-1}) = o(t),$$

which contradicts the no-gap stationarity condition $Q(t) = t$. The case $p = 2$, corresponding to the SK model, is exactly the critical case where this argument fails: the same estimate gives only $Q(t) = O(t)$, which is compatible with $Q(t) = t$.

Why strict increase matters: The proof of Lemma 6.15 in [EAMS21] uses strict increase to turn an integrated first-variation inequality into the pointwise identity $Q(t) = t$. The perturbation argument produces integrals weighted by terms of the form $\gamma^*(t) - \gamma^*(t_1)$. If γ^* is strictly increasing, this weight is positive for $t > t_1$, allowing one to conclude $Q(t) - t \le 0$ and $Q(t) - t \ge 0$ pointwise. If γ^* has flat regions, this argument no longer forces equality everywhere.

Some flat interval, not necessarily an initial one. The proof contradicts the identity $Q(t) = t$ at small t, but this identity was obtained from Lemma 6.15 only under the global strict-increase assumption. Thus the conclusion is that no globally strictly increasing minimizer exists. This yields a flat interval somewhere in $[0, 1)$, but does not by itself prove that the flat interval starts at 0. Locating the gap would require a more quantitative optimality condition.

5.6 Optional PDE intuition for the local Φ_{xx} bound

The main proof invokes Lemma 5.4 to obtain local boundedness of Φ_{xx} near $t = 0$. We briefly explain why this bound is natural from a parabolic PDE viewpoint.

For the pure p-spin model, define

$$\tau = \xi'(1) - \xi'(t).$$

Since $\xi'(t) = pt^{p-1}$, we have

$$\tau = p(1 - t^{p-1}).$$

Thus $t = 1$ corresponds to $\tau = 0$, while $t = 0$ corresponds to $\tau = p$. Define

$$U(\tau, x) := \Phi(t(\tau), x), \qquad a(\tau) := \gamma^*(t(\tau)).$$

Then the Parisi PDE becomes

$$U_\tau = \frac{1}{2}U_{xx} + \frac{1}{2}a(\tau)U_x^2, \qquad U(0, x) = |x|.$$

Let

$$V(\tau, x) := U_x(\tau, x).$$

Formally differentiating in x, we obtain

$$V_\tau = \frac{1}{2}V_{xx} + a(\tau)VV_x.$$

The maximum principle gives

$$\|V(\tau, \cdot)\|_\infty \le 1.$$

Indeed, at an interior positive maximum of V, one has $V_x = 0$ and $V_{xx} \le 0$, so $V_\tau \le 0$. At an interior negative minimum, one has $V_x = 0$ and $V_{xx} \ge 0$, so $V_\tau \ge 0$. Since the initial data satisfies

$$V(0, x) = \operatorname{sign}(x), \qquad |V(0, x)| \le 1,$$

the maximum cannot increase and the minimum cannot decrease. To make this argument fully rigorous, one can first replace $|x|$ by smooth approximations $g_\varepsilon(x)$ with $|g_\varepsilon'(x)| \le 1$, prove the estimate for the smooth solutions, and then pass to the limit.

Now rewrite the equation for V as

$$V_\tau = \frac{1}{2}V_{xx} + b(\tau, x)V_x, \qquad b(\tau, x) := a(\tau)V(\tau, x).$$

We do not need b to be bounded by 1. We only need it to be locally bounded. Since $\gamma^* \in \mathscr{U}$ is nondecreasing and integrable, γ^* is locally bounded on every interval $[0, \delta] \subset [0, 1)$. For original times $t \in [0, \delta]$, equivalently

$$\tau \in [p(1 - \delta^{p-1}), p],$$

we have

$$|b(\tau, x)| \le \sup_{0 \le s \le \delta} \gamma^*(s) < \infty.$$

Standard parabolic smoothing for linear equations with bounded drift then gives

$$\sup_{\tau \in [\tau_0, p]} \|V_x(\tau, \cdot)\|_\infty \le C(\tau_0, \delta, \gamma^*) \qquad \text{for every } \tau_0 > 0.$$

Since $V_x = U_{xx}$, this gives a uniform bound on U_{xx} for τ bounded away from 0. Translating back to the original time variable, we obtain

$$\sup_{t \in [0, \delta]} \sup_x |\Phi_{xx}(t, x)| < \infty.$$

This is precisely the type of estimate used in the main proof. In the rigorous argument above, we simply cite Lemma 6.4 of [EAMS21], which provides this regularity directly.

6 Toward the Finite-N OGP: A Conjectural Upgrade

Theorem 5.1 establishes a PDE-level gap: for pure $p > 2$, the zero-temperature Parisi minimizer γ^* cannot be strictly increasing. Equivalently, there exists a nontrivial interval $(a, b) \subset [0, 1)$ such that

$$d\gamma^*((a, b)) = 0.$$

In the terminology of the project, this is the OGP statement at the level of the Parisi PDE/order parameter, following the zero-temperature no-overlap-gap framework of [EAMS21]. The natural next question is whether this gap can be upgraded to a finite-N binary OGP statement for actual near-ground-state configurations.

A binary finite-N OGP statement (using the overlap $R(\sigma, \tau) = \frac{1}{N}\langle\sigma, \tau\rangle$ from (6)) asserts that there exist $0 < a < b < 1$ and $\varepsilon > 0$ such that, with high probability, there do not exist $\sigma, \tau \in \{\pm 1\}^N$ satisfying

$$H_N(\sigma) \geq N(\mathsf{OPT} - \varepsilon), \qquad H_N(\tau) \geq N(\mathsf{OPT} - \varepsilon),$$

and

$$a < |R(\sigma, \tau)| < b.$$

This is the standard finite-N formulation of the overlap gap property used in the spin-glass algorithmic barriers literature [GS14, GS17, GJS21]. Thus the finite-N OGP is stronger than the PDE-level gap: it rules out the existence of near-optimal pairs with intermediate overlap, rather than only showing that the limiting overlap order parameter assigns no mass to an interval.

6.1 A conditional route through constrained two-replica energies

The standard way to prove such a finite-N statement is to study the best energy achievable at a fixed overlap. This constrained two-replica perspective appears throughout the rigorous analysis of overlap structures, coupled replicas, and Parisi-type variational principles [Tal06, Pan14, CHL18, GJS21]. Define the constrained two-replica optimum

$$M_N(q, \delta) = \max_{\substack{\sigma,\tau\in\{\pm1\}^N \\ ||R(\sigma,\tau)|-q|\leq\delta}} \frac{H_N(\sigma) + H_N(\tau)}{2N}.$$

If this constrained optimum lies strictly below the ground-state energy for all q in some interval, then binary OGP follows immediately.

Proposition 6.1 (Constrained gap implies binary OGP)**.** *Suppose there exist $0 < a < b < 1$, $\eta > 0$, and $\delta_N \downarrow 0$ such that, with high probability,*

$$\sup_{q\in(a,b)} M_N(q, \delta_N) \leq \mathsf{OPT} - \eta.$$

Then for any $0 < \varepsilon < \eta$, with high probability there do not exist $\sigma, \tau \in \{\pm 1\}^N$ such that

$$H_N(\sigma) \geq N(\mathsf{OPT} - \varepsilon), \qquad H_N(\tau) \geq N(\mathsf{OPT} - \varepsilon),$$

and

$$a < |R(\sigma, \tau)| < b.$$

Proof. Suppose such σ, τ existed, and set

$$q := |R(\sigma, \tau)| \in (a, b).$$

Then (σ, τ) is feasible for $M_N(q, \delta_N)$, so

$$M_N(q, \delta_N) \geq \frac{H_N(\sigma) + H_N(\tau)}{2N} \geq \mathsf{OPT} - \varepsilon.$$

Since $\varepsilon < \eta$, this contradicts the assumed constrained-overlap upper bound. Hence no such pair can exist. □

Thus the finite-N OGP problem reduces to proving a strict constrained energy gap:

$$\limsup_{N\to\infty} M_N(q, \delta_N) < \mathsf{OPT} \qquad \text{for } q \in (a, b).$$

6.2 Conjectural bridge from the Parisi gap

Our PDE proof identifies a candidate forbidden interval: any flat interval of γ^*. This motivates the following conjectural bridge.

Conjecture 6.2 (Parisi gap implies constrained two-replica gap). Let (a, b) be a nontrivial flat interval of the zero-temperature Parisi minimizer γ^*, so that

$$d\gamma^*((a, b)) = 0.$$

Then there exists $\eta > 0$ such that for every compact subinterval $[a', b'] \subset (a, b)$,

$$\limsup_{N\to\infty} \sup_{q\in[a',b']} M_N(q, \delta_N) \le \mathsf{OPT} - \eta$$

for some $\delta_N \downarrow 0$.

The intuition is that γ^* encodes the zero-temperature limiting overlap structure of near-ground-state configurations, as described by the Parisi order parameter and its zero-temperature scaling [EAMS21]. If γ^* assigns no mass to an interval (a, b), then overlap values in that interval should be energetically unfavorable for pairs of near-ground states. In other words, forcing two replicas to have overlap $q \in (a, b)$ should lower the best achievable average energy below OPT.

A rigorous proof of this conjecture would likely require a coupled or two-replica Parisi variational principle. Such constrained overlap problems are typically approached through interpolation and coupled-replica variational arguments [Tal06, Pan14, CHL18]. One would need a formula for the limiting constrained energy

$$\lim_{N\to\infty} M_N(q, \delta_N),$$

and then show that the corresponding constrained variational problem has value strictly below OPT whenever q lies inside the flat interval of γ^*. This would upgrade the PDE-level gap to a full binary finite-N OGP theorem.

6.3 Numerical evidence and landscape interpretation

Our numerical experiments (Section 7) support this picture. In the discretized Parisi optimization problem, the $p = 2$ case shows curvature across multiple Hessian directions, consistent with a full-RSB-type optimizer. In contrast, for $p \ge 3$, the Hessian spectrum collapses: one dominant eigenvalue remains while the remaining directions are nearly flat. This suggests that the optimizer is effectively low-dimensional and step-like, consistent with the existence of a flat interval in γ^*.

This numerical rank collapse should not be interpreted as a proof of finite-N OGP. Rather, it is evidence for the PDE-level mechanism: the variational optimizer does not use a continuum of overlap levels. The conjectural finite-N upgrade is that the inactive interval in the Parisi order parameter corresponds to a genuine forbidden-overlap interval for near-ground-state spin configurations. This interpretation is consistent with the broader picture in which OGP acts as an obstruction to stable algorithmic procedures such as AMP and related local or Lipschitz algorithms [GS14, GS17, GJS21, HS22].

6.4 Status of the proof attempt

At this stage, the rigorous part of our argument is

$$\text{pure } p > 2 \quad\Longrightarrow\quad \gamma^* \text{ is not strictly increasing} \quad\Longrightarrow\quad \gamma^* \text{ has a flat interval.}$$

This proves the PDE-level gap.

The remaining finite-N step is

$$\gamma^* \text{ flat on } (a, b) \quad\Longrightarrow\quad M_N(q) < \mathsf{OPT} \text{ for } q \in (a, b).$$

Based on our current knowledge and expertise, this is a genuinely difficult upgrade for us. It would require additional machinery, most likely a constrained two-replica interpolation or a coupled Parisi variational formula. We therefore state the full binary OGP as a conjectural extension of the PDE-level result, and isolate the constrained-overlap energy gap as the precise missing ingredient.

7 Numerical Analysis

7.1 Numerical Solution

Numerically, to investigate OGP, we'd like to solve the Parisi PDE, and characterize the optimal solution $\gamma^*(t)$. To do this we would let $\gamma^*(t)$ be a step function, and investigate whether or not it contains a region of flatness. Formally, we'd like to verify whether there exists $0 \leq a < b \leq 1$ such that $\gamma^*(a) = \gamma^*(b)$. As γ^* is nondecreasing, it follows that this would imply that $\gamma^*(t) = \gamma^*(a)$ for all $t \in [a, b]$. We relax the class $\mathcal{U}$ so that instead γ is a piecewise step function. Define

$$\mathcal{U}_K := \left\{ \gamma_K \in \mathcal{U} : \gamma_K(t) = \sum_{i=1}^{K} g_i \mathbb{1}\{[t_{i-1}, t_i)\} \right\} \tag{12}$$

for some temporal discretization $0 = t_0 < t_1 < ... < t_K = 1$, and a set of coefficients $g_1, ..., g_K$. Since $\gamma_K \in \mathcal{U}$ requires nondecreasing nonnegative values, the coefficients must satisfy $0 \leq g_1 \leq g_2 \leq \cdots \leq g_K$. With this discretization, detecting a gap reduces to finding $i \neq j$ with $g_i = g_j$. The key idea that we are using to solve the Parisi PDE is that it has an analytical solution on the interval $[a, b]$ if $\gamma(t)$ is constant. This can be formalized in the following proposition.

Proposition 7.1. *If $\gamma(t) = m$, with $m \in \mathbb{R}$ a constant, on the time-interval $t \in [a, b]$, the solution of Parisi PDE* (9) *at the initial timepoint a is given by*

$$\Phi(a, x) = \frac{1}{m} \log \mathbb{E}\left[\exp\{ m\Phi(b, x + \sqrt{A(a, b)}Z\} \right] \tag{13}$$

where $Z \sim \mathcal{N}(0, 1)$ and $A(a, b) = \int_a^b \xi''(s)\, ds$

A derivation of 7.1 can be found in appendix B.

Given

$$\gamma_K(t) = \sum_{i=1}^{K} g_i \mathbb{1}\{[t_{i-1}, t_i)\}$$

for $0 = t_0 < t_1 < ... < t_K = 1$ we can then solve the Parisi PDE backwards in time starting with the terminal condition $\Phi(1, x) = \Phi(t_K, x) = |x|$. Specifically, we compute recursively

$$\Phi(t_{k-1}, x) = \frac{1}{g_{k-1}} \log \mathbb{E}\left[\exp\{ g_{k-1}\Phi(t_k, x + \sqrt{A(t_{k-1}, t_k)}Z\} \right] \tag{14}$$

For $k = K, ..., 1$. We then use the initial value $\Phi(0, 0)$ to compute the Parisi functional $P(\gamma_K)$ for a given set of coefficients.

The problem of finding γ^* then takes on the form of a standard optimization problem that we can approximate numerically using standard optimization algorithms, and accelerated using automatic differentiation.

7.1.1 Solving Equation (14)

Naively, one might try to approximate (14) using Monte-Carlo integration. Due to the nature of the recursive formula, if one attempts this with N_{MC} Monte-Carlo samples per evalutation of the expectation, one will end up with an computational complexity of $\mathcal{O}(N_{MC}^K)$ for solving the full PDE. This comes from the nested expectation, and the fact that Monte-Carlo simulation almost surely will give distinct spatial points in which to evaluate Φ on. To circumvent the exponential bottleneck (in K), one should instead approximate the expectation using some fixed spatial nodes. We propose using Gauss-Hermite quadrature [QSS07].

$$\mathbb{E}[f(Z)] \approx \frac{1}{\sqrt{\pi}} \sum_k w_k f(\sqrt{2}x_k) \tag{15}$$

where x_k, w_k are Hermite nodes and weights, respectively. Suppose N_x is the number of spatial grid points, and Q is the number of Gauss-Hermite nodes. The computational complexity of one solve of the Parisi PDE using (14) is $\mathcal{O}(KN_xQ)$.

7.1.2 Unconstrained Optimization

As previously discussed, finding the optimal γ^* in the discretized case corresponds to finding the coefficients $g_1, ..g_K$ such that $P(\gamma_K)$ is minimized. To uphold that $\gamma_K \in \mathcal{U}_K$ we must constrain that $0 \leq g_1 \leq g_2 \leq ... \leq g_K$. Letting $g = (g_i)_{1 \leq i \leq K}$, we can then consider the constrained optimization problem

$$\min_{g \in \mathbb{R}}^{K} P(\gamma_K(g)) \quad \text{subject to } 0 \leq g_1 \leq g_2 \leq ... \leq g_K \tag{16}$$

Another approach is to parametrize the coefficients g as a function of another set of parameters $\theta \in \mathbb{R}^K$, such that we instead consider an unconstrained optimization problem. Specifically, we can consider a mapping $T_{\text{sp}} : \mathbb{R}^{\mathbb{K}} \to \mathbb{R}^K$ such that

$$g_i = T_{\text{sp}}^{(i)}(\theta) := \sum_{k=1}^{i} \text{softplus}(\theta_k), \quad \text{softplus}(x) := \ln(1 + e^x) \tag{17}$$

By construction the constraint is upheld for this parametrization. Hence we can also consider the *un*constrained optimization problem

$$\min_{\theta \in \mathbb{R}^K} P(\gamma_K(T_{\text{sp}}(\theta))) \tag{18}$$

Similarly, if we would like to bound $\gamma_K \in [0, 1]$ we can use the softmax function to define the parametrization

$$g_i = T_{\text{sm}}^{(i)}(\theta) = \sum_{k=1}^{i} \frac{e^{\theta_k}}{\sum_{j=1}^{K} e^{\theta_j}} \tag{19}$$

7.2 Numerical Experiments

7.2.1 Optimum of Parisi Functional

To approximate γ^*, we computed the optimal set of coefficients for g using both the softplus and the softmax parametrization. The unconstrained optimization problem was solved using the L-BFGS algorithm [NW06], using forward-mode automatic differentiation to compute the gradients. We used a spatial discretization of $N_x = 1001$, and $Q = 40$ Gauss-Hermite nodes. We also tried optimizing directly over the coefficients g in two different manners. In the first, we added a penalty term in the objective function to enforce that $\gamma_K \in \mathcal{U}_K$. In the second approach we modelled the constrained optimization problem, and solved it using `JuMP.jl` [LDD+23].

7.3 Results

The hard-constraint optimization version using `JuMP.jl` did not converge in a reasonable amount of time, and was excluded from further analysis.

In plots 1 and 2 we see the optimal Parisi order parameter in the case that we constrain it between $[0, 1]$ using softmax, and the case where we let it be unbounded from above. Both plots show the order parameter in the case of $p = 2, 3, ..., 7$. Noticeably, we observe that in the unbounded case, the order parameter is strictly increasing for $p = 2$, while it has a flat region for $p > 3$. Note here that the "strictness" refers to the values at two different temporal nodes. For $p = 3$, we see - as we hypothesised - that the order parameter behaves as in the case $p > 3$. For the bounded case we practically see the same trend, although for $p = 2$, there is also an interval towards $t = 1$ on which the order parameter is seemingly constant.

We note that we did not expect the order parameters to be essentially flat in the whole interval $[0, 1]$. We tried to employ different numerical solvers, and change tolerances, but the result remained the same. A priori we had an expectation that - at least for the unbounded case - there would be a small but noticeable flat interval, and that otherwise the order parameter would be increasing. In contrast to our results where there is essentially just one "jump", for $p \geq 3$.

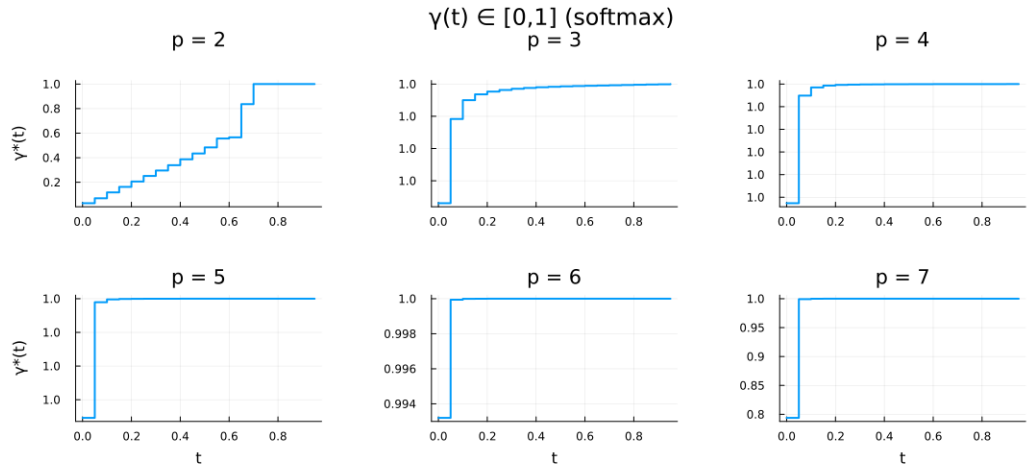


Figure 1: The numerical results show that in the bounded case the order parameter has an interval on which it is constant for all $p = 2, 3, ..., 7$

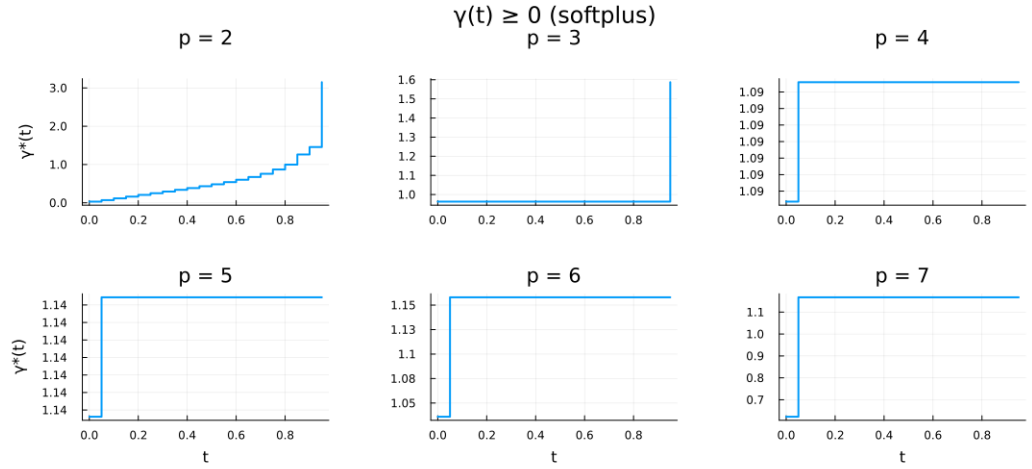


Figure 2: The numerical results find evidence that in the unbounded case the order parameter is strictly increasing when $p = 2$, and also that it remains constant on a nontrivial interval when $p \geq 3$

The coefficients computed from optimizing directly over g, with a penalty in the objective function, rendered essentially the same results as the softplus-coefficients, and therefore we do not include them here.

7.4 Loss Landscape of the Parisi Optimization

The optimization problem $\min_{\theta \in \mathbb{R}^K} P_K(\gamma_K(\theta))$ defines a loss landscape whose geometry encodes structural information about the optimal Parisi measure μ^*. In this section we analyze this landscape through the lens of Hessian spectral analysis, following the eigenvector-projection technique of [LXT+18]. The key finding is that the landscape geometry sharply distinguishes the full-RSB regime ($p = 2$) from the one-step RSB regime ($p \geq 3$), providing a numerical fingerprint consistent with the theoretical gap established in Section 5.

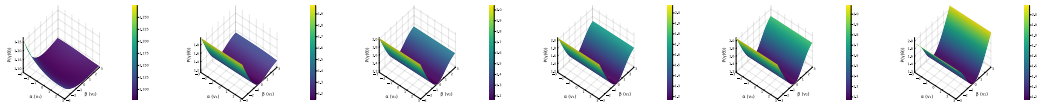
Method. Fix $K = 10$ intervals on $[0, 1]$ and the softplus reparametrization $g_i = \sum_{k=1}^{i} \mathrm{softplus}(\theta_k)$, which enforces $\gamma_K \in \mathcal{U}_K$. We optimize P_K using L-BFGS with automatic differentiation [NW06] to obtain $\theta^* \in \mathbb{R}^{10}$. The Hessian $H = \nabla^2_\theta P_K(\theta^*)$ is then computed exactly via forward-mode automatic differentiation [RLP16]. Let $v_1, v_2 \in \mathbb{R}^{10}$ denote the top two eigenvectors

of H (corresponding to the two largest eigenvalues). We evaluate P_K on the two-dimensional slice

$$(\alpha, \beta) \;\mapsto\; P_K(\gamma_K(\theta^* + \alpha v_1 + \beta v_2))\,, \qquad (\alpha, \beta) \in [-3, 3]^2,$$

and visualize the resulting surface. This projection isolates the directions of greatest curvature at the optimum.

Results. Figure 3 shows the loss surface projected onto the top two Hessian eigenvectors for $p = 2, \ldots, 7$. Table 1 reports the top four eigenvalues and condition number (ratio of the largest to second-largest eigenvalue) for each p.



(a) $p = 2$ (b) $p = 3$ (c) $p = 4$ (d) $p = 5$ (e) $p = 6$ (f) $p = 7$

Figure 3: P_K projected onto the top two Hessian eigenvectors at θ^* ($K = 10$, softplus). $p = 2$: bowl with curvature in both directions (FRSB). $p \geq 3$: ridge along v_1 (step-height direction), nearly flat along all remaining directions (1RSB).

p	λ_4	λ_3	λ_2	λ_1	λ_1/λ_2	RSB
2	6.0×10^{-5}	2.6×10^{-4}	9.7×10^{-4}	7.4×10^{-3}	7.6	FRSB
3	4.4×10^{-11}	9.8×10^{-9}	2.0×10^{-5}	1.85×10^{-1}	9.3×10^{3}	1RSB
4	≈ 0	7.6×10^{-156}	2.7×10^{-12}	2.7×10^{-1}	1.0×10^{11}	1RSB
5	≈ 0	8.3×10^{-130}	7.3×10^{-13}	3.2×10^{-1}	4.5×10^{11}	1RSB
6	2.9×10^{-215}	6.3×10^{-30}	3.7×10^{-14}	3.6×10^{-1}	9.6×10^{12}	1RSB
7	≈ 0	≈ 0	1.7×10^{-14}	3.8×10^{-1}	2.2×10^{13}	1RSB

Table 1: Hessian spectrum and condition number λ_1/λ_2 at θ^* ($K = 10$, softplus). The condition number for $p = 2$ is $O(1)$; for $p \geq 3$ it grows from $\sim 10^3$ to $\sim 10^{13}$, reflecting increasingly degenerate 1RSB structure as p increases.

The $p = 2$ case (FRSB). For $p = 2$ (the Sherrington–Kirkpatrick model [SK75]), the top four Hessian eigenvalues span roughly two orders of magnitude (6×10^{-5} to 7.4×10^{-3}), with a condition number of approximately 7.6. The projected surface shows meaningful curvature in both eigenvector directions, consistent with a well-conditioned optimization problem. Theoretically, the SK model has a full-RSB Parisi measure [Gue03, Tal06]: the optimal $\gamma^*(t)$ is smooth and strictly increasing, so all $K = 10$ parameters contribute independently to the objective. A discretization of a strictly increasing function requires all K degrees of freedom, yielding a full-rank Hessian.

The $p \geq 3$ case (1RSB). For $p = 3, \ldots, 7$ a single large eigenvalue ($\lambda_1 \sim 10^{-1}$) dwarfs the rest; the gap grows with p, from $\sim 10^3$ at $p = 3$ to $\sim 10^{13}$ at $p = 7$. The projected surface is a ridge: curved sharply along v_1 (the "step-height" direction) and nearly flat in all other directions. This near-rank-one structure is a numerical fingerprint of one-step RSB. Theoretically, the optimal Parisi measure for $p \geq 3$ is a single atom $\mu^* = c^* \delta_{q^*}$ [AC15, HS22], meaning the effective parameter space is at most two-dimensional (step height c^* and step location q^*). In the K-dimensional θ-parameterization this implies that the Hessian has at most two numerically significant eigenvalues: one for the height direction and one (much smaller, of order $\mathcal{O}(K^{-2})$ from finite-grid discretization) for the location direction. All remaining $K - 2$ eigendirections correspond to perturbations that rearrange γ within $[q^*, 1]$ while preserving the integrated measure, contributing negligible second-order cost.

Connection to the OGP gap. The landscape analysis is consistent with the theoretical result of Section 5. The existence of a nearly flat plateau in P_K (i.e., insensitivity to all perturbations except the step height) is a geometric manifestation of the fact that the minimizer γ^* cannot be strictly increasing for $p \geq 3$. A strictly increasing minimizer would require all K parameters to contribute equally, producing a full-rank Hessian as in the $p = 2$ case. The observed rank collapse for $p \geq 3$ confirms numerically that the optimizer resides on a low-dimensional manifold of step functions, which is precisely the OGP regime described by [EAMS21, HS22].

Comparison of parameterizations. We also compare the softplus parameterization (which allows $\gamma \in [0, \infty)$, matching the class $\mathcal{U}$) against the softmax parameterization (which constrains γ to a probability simplex CDF in $[0, 1]$). For $p = 3$, the softmax Hessian has all eigenvalues at the noise floor ($\sim 10^{-9}$), yielding a completely flat landscape. This degeneracy arises because the 1RSB optimizer wants a step to a large value $c^* \gg 1$, which lies outside the softmax constraint set. The optimizer saturates the boundary, and the Hessian reflects no useful curvature information. This confirms that the softplus parameterization is the appropriate one for studying the unconstrained Parisi problem in $\mathcal{U}$.

8 Conclusion

We have studied the zero-temperature Parisi variational problem for the pure p-spin Ising spin glass, focusing on the geometric structure of the optimizer γ^* and its connection to the overlap gap property. Our main theoretical contribution is Theorem 5.1: for every pure $p > 2$, no minimizer of the Parisi functional can be strictly increasing, and hence every minimizer must have a nontrivial flat interval. The proof is clean and self-contained, relying on three inputs from El Alaoui, Montanari, and Sellke [EAMS21] and a single scaling observation about $\xi''(t) = p(p-1)t^{p-2}$ near $t = 0$. The $p = 2$ (SK) case is the exact borderline where the argument breaks down, consistent with the known FRSB structure of the SK model.

On the numerical side, we developed a practical solver for the Parisi variational problem using the Hopf–Cole formula and Gauss–Hermite quadrature, reducing the cost from exponential in K to $O(KN_xQ)$ per evaluation. Combined with L-BFGS and automatic differentiation, this gives a reliable optimizer for γ^* across $p = 2, \ldots, 7$. The experiments confirm the theoretical picture: γ^* is step-like (1RSB) for $p \geq 3$ and smooth (FRSB) for $p = 2$. The Hessian landscape analysis further reveals that the condition number of the Parisi optimization grows from ≈ 7.6 at $p = 2$ to $\sim 10^{13}$ at $p = 7$, providing a sharp numerical signature that the 1RSB minimizer is effectively one-dimensional in the space of γ functions.

However, we also discuss several open questions which we encountered in exposition of this work and could not fully solve them yet. The most immediate open problem is to promote the PDE-level gap to a full binary OGP statement for actual spin configurations. As outlined in Section 6, a natural route is through constrained two-replica energies: one would need a formula for the limiting constrained optimum $M_N(q, \delta_N)$ and show it lies strictly below OPT for q in the flat interval of γ^*. This likely requires a coupled Parisi variational formula and interpolation argument. Our proof shows that some flat interval exists in $[0, 1)$ but does not determine where it starts. Proving that the flat interval begins at $t = 0$—an initial gap—would require a sharper optimality condition at the boundary, such as a quantitative version of the support condition in [EAMS21]. The argument extends immediately to any mixture ξ for which $\xi''(t) = O(t^\alpha)$ near $t = 0$ with $\alpha > 0$. Characterizing exactly which mixtures admit a PDE-level gap, and whether the gap interval can be explicitly computed from ξ, is an interesting direction.

Algorithmic implications. Connecting the PDE-level gap rigorously to barriers for specific algorithms such as AMP, Langevin dynamics, or low-degree polynomials remains an important open problem, and would strengthen the interpretation of our result as an obstruction to efficient optimization.

Acknowledgements

We acknowledge the use of Large Language Models (LLMs), including Claude Code and ChatGPT, throughout the development of this work. We employed these tools as experimental aids for initial brainstorming and as verification components of our proof-writing process. Specifically, LLMs assisted in organizing the logical structure of our proofs, suggesting appropriate lemmas and propositions, and evaluating the validity of certain underlying assumptions.

Additionally, generative AI tools were utilized to assist in writing and debugging code, as well as refining the linguistic clarity of this report. However, it is important to emphasize that the core proofs and technical arguments were written solely by the authors. We take full responsibility for the accuracy and integrity of the content presented herein.

References

[ABH+05] Sanjeev Arora, Eli Berger, Elad Hazan, Guy Kindler, and Muli Safra. On non-approximability for quadratic programs. In *Proceedings of the 46th Annual IEEE Symposium on Foundations of Computer Science*, FOCS '05, page 206–215, USA, 2005. IEEE Computer Society.

[AC15] Antonio Auffinger and Wei-Kuo Chen. The parisi formula has a unique minimizer. *Communications in Mathematical Physics*, 335(3):1429–1444, 2015.

[AC17] Antonio Auffinger and Wei-Kuo Chen. Parisi formula for the ground state energy in the mixed p-spin model. *The Annals of Probability*, 45(6B):4617 – 4631, 2017.

[CHL18] Wei-Kuo Chen, Madeline Handschy, and Gilad Lerman. On the energy landscape of the mixed even p-spin model. *Probability Theory and Related Fields*, 171(1):53–95, 2018.

[CHM+15] Anna Choromanska, MIkael Henaff, Michael Mathieu, Gerard Ben Arous, and Yann LeCun. The Loss Surfaces of Multilayer Networks. In Guy Lebanon and S. V. N. Vishwanathan, editors, *Proceedings of the Eighteenth International Conference on Artificial Intelligence and Statistics*, volume 38 of *Proceedings of Machine Learning Research*, pages 192–204, San Diego, California, USA, 09–12 May 2015. PMLR.

[EAMS21] Ahmed El Alaoui, Andrea Montanari, and Mark Sellke. Optimization of mean-field spin glasses. *The Annals of Probability*, 49(6):2922–2960, 2021.

[FS06] Wendell H Fleming and H Mete Soner. *Controlled Markov processes and viscosity solutions*. Springer, 2006.

[Gam21] David Gamarnik. The overlap gap property: A topological barrier to optimizing over random structures. *Proceedings of the National Academy of Sciences*, 118(41):e2108492118, 2021.

[GJS21] David Gamarnik, Aukosh Jagannath, and Subhabrata Sen. The overlap gap property in principal submatrix recovery. *Probability Theory and Related Fields*, 181(4):757–814, 2021.

[GS14] David Gamarnik and Madhu Sudan. Limits of local algorithms over sparse random graphs. In *Proceedings of the 5th conference on Innovations in theoretical computer science*, pages 369–376, 2014.

[GS17] David Gamarnik and Madhu Sudan. Performance of sequential local algorithms for the random nae-k-sat problem. *SIAM Journal on Computing*, 46(2):590–619, 2017.

[Gue03] Francesco Guerra. Broken replica symmetry bounds in the mean field spin glass model. *Communications in mathematical physics*, 233(1):1–12, 2003.

[HS22] Brice Huang and Mark Sellke. Tight lipschitz hardness for optimizing mean field spin glasses. 2022.

[Kle12] Fima C Klebaner. *Introduction to Stochastic Calculus with Applications*. Imperial College Press, 2012.

[LDD+23] Miles Lubin, Oscar Dowson, Joaquim Dias Garcia, Joey Huchette, Benoît Legat, and Juan Pablo Vielma. JuMP 1.0: Recent improvements to a modeling language for mathematical optimization. *Mathematical Programming Computation*, 15:581–589, 2023.

[LXT+18] Hao Li, Zheng Xu, Gavin Taylor, Christoph Studer, and Tom Goldstein. Visualizing the loss landscape of neural nets. *Advances in neural information processing systems*, 31, 2018.

[MPVT88] Marc Mézard, Giorgio Parisi, Miguel Angel Virasoro, and David J Thouless. Spin glass theory and beyond, 1988.

[NW06] J. Nocedal and S. Wright. *Numerical Optimization*. Springer, 2006.

[Pan14] Dmitry Panchenko. The parisi formula for mixed p-spin models. 2014.

[QSS07] A. Quarteroni, R. Sacco, and F. Saleri. *Numerical Mathematics*. Springer, 2007.

[RLP16] Jarrett Revels, Miles Lubin, and Theodore Papamarkou. Forward-mode automatic differentiation in julia. *arXiv preprint arXiv:1607.07892*, 2016.

[SK75] David Sherrington and Scott Kirkpatrick. Solvable model of a spin-glass. *Physical review letters*, 35(26):1792, 1975.

[Tal06] Michel Talagrand. The parisi formula. *Annals of mathematics*, pages 221–263, 2006.

A Representation of Parisi PDE as SDE

There is a close relationship between stochastic differential equations (SDEs) and partial differential equations (PDEs). Roughly speaking, an SDE describes the random evolution of a single trajectory, while a PDE describes the evolution of quantities averaged over many realizations of the process. As an example, think about how a diffusion process can be described on different scales. On a microscopic level, one can look at a single particle and model it as some form of random walk. On an intermediate level this random walk can better be described with a stochastic differential equation. On a macroscopic level, one might be more interested in the mean-field evolution. At this scale we can instead consider a deterministic diffusion process described by a PDE.

Consider a stochastic differential equation (SDE) of the form

$$dX_t = b(X_t, t)\, dt + \sigma(X_t, t)\, dB_t$$

where B_t is a standard Brownian motion. For a function $f(x,t)$, we define its associated *generator* as the operator

$$L_t f(x,t) = \frac{1}{2}\sigma^2(x,t) f_{xx}(x,t) + b(x,t) f_x(x,t) \tag{20}$$

From [Kle12] we know that for f smooth enough, the stochastic analogue of the chain rule is

$$df(X_t, t) = \left(L_t f(X_t, t) + f_t(X_t, t)\right)\, dt + f_x(X_t, t)\, \sigma(X_t, t)\, dB_t$$

Taking expectations removes the stochastic integral term, since it has expectation zero. This leads to a connection between diffusion processes and PDEs; see for example [FS06].

A typical example is the pure Brownian motion process

$$dX_t = dB_t$$

In this case, the generator simplifies to

$$L_t = \frac{1}{2}\partial_{xx}$$

If we define

$$u(t,x) = \mathbb{E}[g(X_T) \mid X_t = x],$$

then u solves the backward heat equation

$$\partial_t u + \frac{1}{2}\partial_{xx} u = 0, \qquad u(T,x) = g(x).$$

This illustrates how stochastic dynamics can be represented through PDEs.

How do we go the opposite way? Specifically, how do we characterize the Parisi PDE with a terminal condition as a stochastic differential equation? The Parisi PDE admits a similar interpretation, but with an important difference: the equation is nonlinear. This nonlinearity can be understood through stochastic control theory.

A.1 Hamilton-Jacobi-Bellman Formulation

Consider the Parisi PDE (9).

The quadratic term can be rewritten using the Legendre transform identity

$$\frac{1}{2}\gamma(t)p^2 = \sup_{a\in\mathbb{R}} \left\{ ap - \frac{a^2}{2\gamma(t)} \right\}.$$

Substituting $p = \partial_x \Phi_\gamma$ gives

$$\frac{1}{2}\xi''(t)\gamma(t)(\partial_x \Phi_\gamma)^2 = \sup_{a\in\mathbb{R}} \left\{ \xi''(t) a\, \partial_x \Phi_\gamma - \frac{1}{2}\xi''(t)\frac{a^2}{\gamma(t)} \right\}.$$

The Parisi PDE can therefore be rewritten as

$$\partial_t \Phi_\gamma + \sup_{a\in\mathbb{R}} \left\{ \frac{1}{2}\xi''(t)\partial_{xx}\Phi_\gamma + \xi''(t) a\, \partial_x \Phi_\gamma - \frac{1}{2}\xi''(t)\frac{a^2}{\gamma(t)} \right\} = 0.$$

This is a Hamilton-Jacobi-Bellman (HJB) equation associated with the controlled diffusion [FS06]

$$dX_t = \xi''(t)a_t\,dt + \sqrt{\xi''(t)}\,dB_t.$$

The optimal feedback control is obtained by maximizing the expression inside the supremum:

$$a_t^* = \gamma(t)\partial_x\Phi_\gamma(t, X_t).$$

Substituting this optimal control into the diffusion yields the *Parisi SDE*

$$dX_t = \xi''(t)\gamma(t)\partial_x\Phi_\gamma(t, X_t)\,dt + \sqrt{\xi''(t)}\,dB_t.$$

Thus, the Parisi SDE arises naturally as the optimal controlled diffusion associated with the Parisi PDE.

B Proof of Proposition 7.1

If $\gamma(t) = m$ is constant on some interval $[a, b]$, the Parisi PDE (9) reduces to

$$\partial_t\Phi + \frac{1}{2}\xi''(t)\left(\partial_x^2\Phi + m(\partial_x\Phi^2\right) = 0 \tag{21}$$

For $p = 3$, we have $\xi''(t) = 6t$, and so

$$\Phi_t + 3t\left(\Phi_{xx} + m\Phi_x^2\right) = 0 \tag{22}$$

Define the Cole-Hopf transformation

$$u(t, x) = \exp m\Phi_\gamma(t, x)$$

then

$$\Phi = \frac{1}{m}\log u$$

Taking derivatives

$$\Phi_x = \frac{u_x}{mu}$$
$$\Phi_{xx} = \frac{u_{xx}}{mu} - \frac{u_x^2}{mu^2}$$

so

$$\Phi_{xx} + m\Phi_x^2 = \frac{u_{xx}}{mu}$$

The PDE then becomes

$$\partial_t\left(\frac{1}{m}\log u\right) + \frac{1}{2}\xi''(t)\frac{u_{xx}}{mu} = 0$$

which simplifies to

$$u_t + \frac{1}{2}\xi''(t)u_{xx} = 0 \tag{23}$$

This is the heat equation, which has a known analytical solution based on Gaussian convolution.

B.1 Gaussian Convolution Solution to Heat Equation

We want to show that for $t \in (a, b)$, with terminal condition $u(b, x) = f(x)$, the heat equation (23) has a solution at $t = a$

$$u(a, x) = \mathbb{E}[f(x + \sqrt{A(a, b)}Z]$$

with $A(a, b) = \int_a^b \xi''(t)\,dt$, $Z \sim \mathcal{N}(0, 1)$.

To prove this, define

$$\tau(t) = \xi'(b) - \xi'(t)$$

and let

$$u(t, x) = v(\tau(t), x)$$

for some function $v : \mathbb{R}^2 \to \mathbb{R}$. Obviously, $\tau'(t) = -\xi''(t)$, and by the chain rule we then have

$$u_t = -\xi''(t) v_\tau$$
$$u_{xx} = v_{xx}$$

substituting this into the heat equation we get

$$-\xi''(t) v_\tau + \frac{1}{2}\xi''(t) v_{xx} = 0$$

Assuming $\xi''(t) \neq 0$ we have

$$v_\tau = \frac{1}{2} v_{xx} \tag{24}$$

with *initial* condition $v(0, x) = f(x)$. The analytic solution to this heat equation can be derived using Fourier analysis. Let $\hat{v}(\tau, k) = \mathcal{F}(v(\tau, x))$. We take the following three results/facts for granted

$$\mathcal{F}(v_{xx}) = -k^2 \hat{v}$$
$$\mathcal{F}^{-1}(\hat{f} \cdot \hat{g}) = \frac{1}{\sqrt{2\pi}}(f * g)$$
$$\mathcal{F}^{-1}(e^{-k^2\tau/2}) = \frac{1}{\sqrt{2\pi\tau}} e^{\frac{-x^2}{2\tau}}$$

From the first result we get that the PDE in the Fourier domain is an ODE

$$\hat{v}_\tau = -\frac{1}{2} k^2 \hat{v}, \quad \hat{v}(0, k) = \mathcal{F}(v(0, x)) = \mathcal{F}(f(x)) = \hat{f}(k)$$

with solution

$$\hat{v}(\tau, k) = e^{-k^2\tau/2} \hat{f}(k)$$

From the second and third results we get that

$$v(\tau, x) = \mathcal{F}^{-1}(\hat{v}(\tau, k)) = \mathcal{F}^{-1}\left(e^{-k^2\tau/2} \hat{f}(k)\right)$$

and therefore

$$v(\tau, x) = \int_{\mathbb{R}} f(x - y) \frac{1}{\sqrt{2\pi\tau}} e^{-y^2/(2\tau)}\, dy.$$

Making the substitution $y = \sqrt{\tau} z$, we obtain

$$\begin{aligned} v(\tau, x) &= \int_{\mathbb{R}} f(x - \sqrt{\tau} z) \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz \\ &= \mathbb{E}[f(x - \sqrt{\tau} Z)], \end{aligned}$$

where $Z \sim \mathcal{N}(0, 1)$. Since Z is symmetric,

$$\mathbb{E}[f(x - \sqrt{\tau} Z)] = \mathbb{E}[f(x + \sqrt{\tau} Z)].$$

Thus,

$$u(a, x) = \mathbb{E}[f(x + \sqrt{A(a, b)} Z)].$$

Which is our desired result.