SYMMETRIZED PHASE RETRIEVAL: INFINITE VERSUS FINITE-SIZE DYNAMICS IN LEARNING

Anonymous authors
Paper under double-blind review

ABSTRACT

Even in its simplest setting, phase retrieval is a difficult optimization problem with a rough, high-dimensional loss landscape. In the standard problem, glassy dynamics and finite-size effects obscure the thermodynamic learning transition. Here, we introduce a symmetrized variant, analyze its finite-size behavior numerically, and solve its infinite dimensional gradient-flow dynamics using dynamical mean-field theory (DMFT). We show that sample complexity is linear in the input dimension and identify two thresholds: a dynamical threshold, above which gradient flow optimization typically generalizes under random initialization; and a landscape-trivialization threshold, at which nongeneralizing minima lose stability. At finite dimensions, the former is significantly lower and learning requires much less data than is necessary for landscape trivialization. In the infinite dimensional limit the thresholds coincide at a limiting value of ≈ 26.86 samples per dimension. We understand this gap in terms of the loss landscape geometry: reflection symmetry confines nongeneralizing minima to the zero-overlap manifold, and retrieval occurs when the initial teacher overlap R_0 exceeds a critical value R_c, which depends on the sample-to-dimension ratio α. The competition between the initial overlap – which scales with the input dimension N as $N^{-1/2}$ – and $R_c(\alpha)$ – which decreases much more strongly – allows retrieval well before trivialization at finite N. In addition, we show that, remarkably, R_0 determines not only the learning success but also the dynamical timescale. An approximate exponential-growth description captures the observed dynamics and yields a learning time logarithmic in $|R_0|$. Thus, one scalar links basin geometry, retrieval probability, and learning timescale. Extensive finite-size simulations verify the predicted thresholds, retrieval and failure trajectories, and learning-time statistics.

1 INTRODUCTION

Estimating the sample complexity of learning problems, i.e. the number of samples required for successful training, is a central and challenging problem in machine learning, due to the non-convex landscape on which optimization is performed. It is often estimated by analyzing the stationary points of the loss landscape: with enough samples, the spurious minima that can trap the dynamics lose stability, and gradient descent succeeds from almost any initialization (Auffinger et al., 2013; Fyodorov & Le Doussal, 2014; Sun et al., 2018). This landscape-trivialization threshold is, however, only an upper bound on the sample complexity of gradient descent. Even with less data, gradient descent fails only from initial conditions in the basins of attraction of the spurious minima, so the actual threshold is set by the size of these basins relative to the distribution of initial conditions (Sarao Mannelli et al., 2019b;a; 2020b;a; Maillard et al., 2026). This is a property of the dynamics that cannot be determined by studying the stability of spurious minima. Moreover, the distinction depends on the dimension: since the size of the basin of spurious minima does not generically scale with dimension in the same manner as the initialization, the fraction of typical initializations that fall outside the spurious basins can differ substantially with dimension.

Phase retrieval provides a simple setting for studying how the amount of training data shapes the loss landscape and learning dynamics. Originally motivated by imaging applications, the problem consists of recovering an unknown signal from squared linear measurements, and can equivalently be viewed as learning a single quadratic neuron from examples generated by a teacher (Fienup, 1982; Shechtman et al., 2015; Fannjiang & Strohmer, 2020; Sarao Mannelli et al., 2020a). A substantial literature has

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

characterized the statistical limits of recovery (Balan et al., 2006; Mondelli & Montanari, 2019; Barbier et al., 2019; Maillard et al., 2020b), including extensions with structured generative priors (Aubin et al., 2020). Efficient algorithms have been developed using convex relaxations (Candès et al., 2013a;b; Candès & Li, 2014), spectral initialization (Lu & Li, 2020; Maillard et al., 2022), alternating minimization (Netrapalli et al., 2013), and gradient-based optimization (Candès et al., 2015; Chen & Candès, 2017; Wang et al., 2018). For sufficiently many Gaussian measurements, with sample complexity of order $N\ \mathrm{polylog}(N)$, rigorous results establish the absence of spurious local minima (Sun et al., 2018) and convergence of gradient descent from Gaussian initialization (Chen et al., 2019), while trajectory-based analyses reveal the role of implicit regularization (Ma et al., 2020). These guarantees involve logarithmic oversampling beyond the proportional regime $M = \alpha N$ considered here, although proportional-sample guarantees are available for other recovery methods (Candès & Li, 2014; Chen & Candès, 2017; Wang et al., 2018). When the number of measurements is proportional to the input dimension, however, the empirical loss can exhibit a rough landscape in which gradient descent spends long periods near nongeneralizing states (Maillard et al., 2020a; Sarao Mannelli et al., 2020a; Mignacco et al., 2021). This behavior connects phase retrieval to broader studies of landscape complexity, topological trivialization, and glassy barriers in inference (Auffinger et al., 2013; Fyodorov & Le Doussal, 2014; Antenucci et al., 2019; Sarao Mannelli et al., 2020b). Drawing on ideas from glassy dynamics, Sarao Mannelli et al. (2020a) proposed that the instability of nongeneralizing threshold states in the teacher direction could provide a criterion for recovery. Whether this stability criterion determines the actual dynamical transition, however, remains unresolved, particularly given the recovery observed below the predicted instability threshold at finite dimension.

Numerical evidence suggests that successful retrieval may precede the disappearance of spurious minima. This distinction also arises in spiked matrix-tensor models, where gradient flow can recover the signal even in the presence of exponentially many spurious minima (Sarao Mannelli et al., 2019b;a). In phase retrieval, gradient descent has been observed to recover the teacher with substantially fewer samples than predicted by stability analyses of nongeneralizing states (Sarao Mannelli et al., 2020a; Bonnaire et al., 2025). These observations reveal a gap between successful optimization and the destabilization of the states that can trap the dynamics. Analyses of the time-dependent Hessian and landscape topology have provided insight into this gap, highlighting the roles of initialization, transient instabilities, and the geometry of nongeneralizing states (Bonnaire et al., 2025; Maillard et al., 2026). Nevertheless, a complete quantitative account of when trajectories avoid these states, and how the resulting retrieval threshold depends on dimension, remains lacking. In particular, strong finite-size effects complicate the extrapolation of numerical observations to the infinite dimensional limit (Sarao Mannelli et al., 2020a; Bonnaire et al., 2025). General results on saddle avoidance and escape do not resolve this issue, since they do not exclude convergence to spurious local minima (Lee et al., 2016; Jin et al., 2017). Understanding the relation between retrieval and landscape trivialization therefore requires characterizing not only the existence and stability of spurious minima, but also the attraction basins that determine whether training reaches them.

In this work, we introduce and study a symmetrized version of the one-layer phase retrieval problem in a teacher-student setup. Each data point in this setting is reflected in a label-preserving manner across the teacher direction. This augmentation imposes an exact reflection symmetry, which is trivially present in the population loss, on the empirical loss for any sample size. In contrast to the usual role of data augmentation, which exploits a known symmetry to improve generalization and sample efficiency (Cohen & Welling, 2016; Bloem-Reddy & Teh, 2020; Chen et al., 2020; Dao et al., 2019; Lyle et al., 2020; Elesedy & Zaidi, 2021; Mei et al., 2021), here the symmetry is an analytical device: a controlled modification of the data distribution, in the spirit of high-dimensional models with structured inputs (Chung et al., 2018; Goldt et al., 2020; Gerace et al., 2020; Loureiro et al., 2022; Maillard et al., 2020a; Tsironis & Moustakas, 2025). The resulting symmetrized model allows us to analyze rigorously both the retrieval dynamics and the landscape trivialization in a controlled setting.

In particular, we find that typical gradient descent (GD) dynamics generalize well before spurious minima lose stability, and analyze this dynamical threshold. Combining numerical experiments with dynamical mean-field theory, we characterize how initialization and basin geometry govern retrieval and learning times, and show why the retrieval and trivialization thresholds are separated at finite dimension but coincide in the infinite dimensional limit.

Our main contributions are:

- We introduce a symmetrized phase-retrieval model, where the empirical loss at finite sample size satisfies exactly the reflection symmetry of the full population loss. We show analytically that in this model all spurious local minima have zero overlap with the teacher, i.e. local minima are confined to the symmetry plane.
- We demonstrate that retrieval and dynamics are largely controlled by the initial overlap R_0 between the teacher and student weights, and identify a critical initial overlap R_c that separates retrieval from convergence to a nongeneralizing state. This value depends on the input dimension N and number of samples M through the ratio $\alpha = M/N$. Comparing $R_c(\alpha)$ with the typical overlap from random-initialization, which scales as $N^{-1/2}$, defines the finite-size dynamical sample complexity.
- At larger ratios we find that the critical overlap vanishes and almost all trajectories generalize. This marks the landscape-trivialization threshold. We show that the two thresholds are widely separated at finite N but converge to the same limit as $N \to \infty$, and determine their common limit, $\alpha_\infty \approx 26.86$, obtained exactly using dynamical mean-field theory (DMFT).
- We solve the infinite dimensional gradient-flow dynamics using DMFT and show that in this limit the initial overlap controls both the learning outcome and its time scale. The dominant stage of successful learning is governed by an approximately exponential growth of $R(t)$, yielding a generalization time logarithmic in $|R_0|$ and a specific shape of the generalization time distribution at finite dimension.

2 Model setup

We consider a student–teacher phase-retrieval model in N dimensions. The teacher direction is a fixed vector $\boldsymbol{w}^\star$ with norm $\|\boldsymbol{w}^\star\| = \sqrt{N}$, and each input $\boldsymbol{x}_\mu$ is drawn independently from a standard Gaussian distribution $\mathcal{N}(\mathbf{0}, I_N)$, with a total of M independent training samples. Each sample is reflected through the teacher axis by

$$S(\boldsymbol{x}_\mu) = \frac{2}{\|\boldsymbol{w}^\star\|^2} (\boldsymbol{x}_\mu \cdot \boldsymbol{w}^\star)\, \boldsymbol{w}^\star - \boldsymbol{x}_\mu. \tag{1}$$

That is, $S(\boldsymbol{x}_\mu)$ reverses the component transverse to the teacher and leaves the parallel component unchanged, $\boldsymbol{x}_\mu \cdot \boldsymbol{w}^\star = S(\boldsymbol{x}_\mu) \cdot \boldsymbol{w}^\star$ (see Figure 1(a)). Repeating this procedure for every sample produces a symmetrized dataset containing M pairs. The phase retrieval loss is defined as the pair-averaged MSE between the squares of the teacher and student outputs:

$$L(\boldsymbol{w}) = \frac{1}{2MN^2} \sum_{\mu=1}^{M} \left[\left((\boldsymbol{x}_\mu \cdot \boldsymbol{w}^\star)^2 - (\boldsymbol{x}_\mu \cdot \boldsymbol{w})^2 \right)^2 + \left((S(\boldsymbol{x}_\mu) \cdot \boldsymbol{w}^\star)^2 - (S(\boldsymbol{x}_\mu) \cdot \boldsymbol{w})^2 \right)^2 \right]. \tag{2}$$

Importantly, this loss is invariant under the symmetry action $L(S(\boldsymbol{w})) = L(\boldsymbol{w})$. This is a symmetry of the full population loss that the finite size loss without symmetrization only respects approximately.

It will be useful to decompose the student weights into their teacher-parallel and teacher-orthogonal components,

$$\boldsymbol{w}(t) = R(t)\boldsymbol{w}^\star + \sqrt{N} s(t) \widehat{\boldsymbol{u}}(t), \qquad \|\widehat{\boldsymbol{u}}(t)\| = 1, \qquad \widehat{\boldsymbol{u}}(t) \cdot \boldsymbol{w}^\star = 0. \tag{3}$$

Here $R(t) = \boldsymbol{w}(t) \cdot \boldsymbol{w}^\star / N$ is the signed teacher-student overlap, $s(t) \geq 0$ is the normalized transverse norm, and $\widehat{\boldsymbol{u}}(t)$ is a unit vector pointing in the transverse direction. The weight norm is $\|\boldsymbol{w}(t)\|^2 = N\left(R^2(t) + s^2(t)\right)$. The two teacher solutions are $\boldsymbol{w} = \pm \boldsymbol{w}^\star$, corresponding to $(R, s) = (\pm 1, 0)$.

The student weights are drawn from a standard Gaussian distribution, $\boldsymbol{w}_0 \sim \mathcal{N}(\mathbf{0}, I_N)$, and evolve via full-batch gradient dynamics

$$\frac{d\boldsymbol{w}}{dt} = -\frac{\partial L(\boldsymbol{w})}{\partial \boldsymbol{w}}\,. \tag{4}$$

Numerically, we implement gradient descent with learning rate η and measure time as $t = \eta k$, where k is the number of gradient-descent steps.

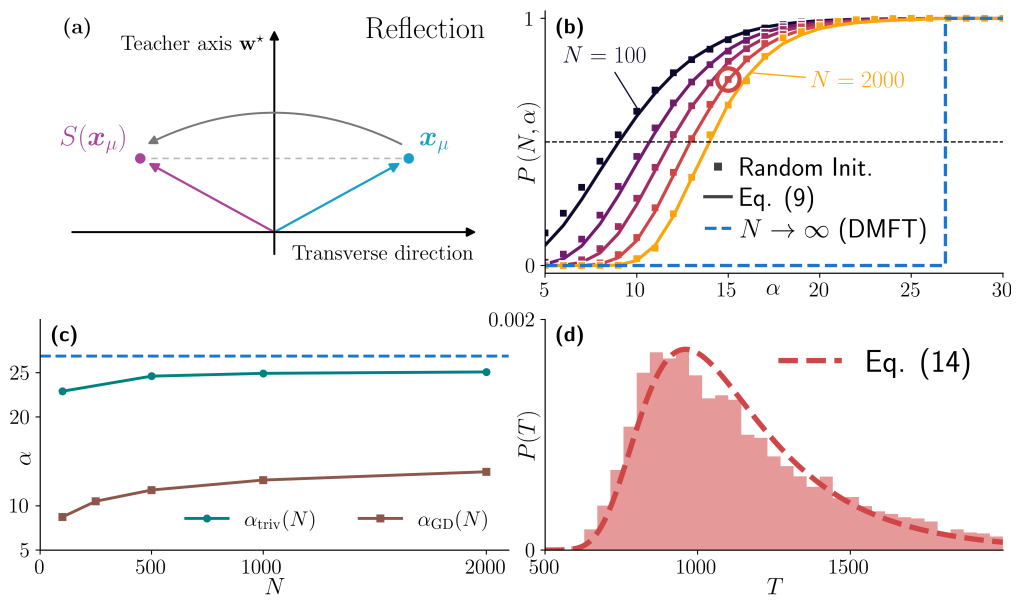


Figure 1: (a) **Symmetrization**: Each sample point is reflected across the teacher axis, reversing its transverse component while leaving its teacher-parallel component unchanged. (b) **Success probability**: $P(N, \alpha)$, the probability of a trajectory with random initialization to generalize. Empirical estimation (squares, each averages 10^4 realizations per point) matches theoretical prediction (solid curves, Eq. (9)). The dashed blue line corresponds to the infinite dimensional limit, obtained by DMFT analysis, Section 2.1. The horizontal dashed line marks the midpoint used to define $\alpha_{\rm GD}(N)$. (c) **Retrieval precedes landscape trivialization**: The dynamical threshold $\alpha_{\rm GD}(N)$ remains well below the landscape-trivialization threshold $\alpha_{\rm triv}(N)$ at finite N. Dashed blue lines show the analytical $N \to \infty$ limit, with the common thermodynamic threshold $\alpha_\infty \approx 26.86$ in (b). (d) **Retrieval-time distribution**: Histogram of the retrieval time T at $N = 1000, \alpha = 15$ for 10^4 random initializations (the location of this parameter choice in panel (b) is marked by a circle). The distribution is well described by the theory, Eq. (14).

2.1 Dynamical mean-field theory

Throughout the paper, we use dynamical mean-field theory (DMFT) (Mignacco et al., 2020; 2021; Bordelon & Pehlevan, 2022) to obtain analytic results for the thermodynamic limit. When the limit is taken proportionally, $N, M \to \infty$ at fixed $\alpha = M/N$, the macroscopic dynamics self-average and admit a closed DMFT description. The procedure produces a deterministic equation for the teacher overlap $R(t)$, given in terms of deterministic correlation and response kernels, while the many transverse coordinates are represented by a single stochastic process with colored Gaussian fluctuations and retarded memory. The full derivation, with numerical validation, is given in Appendix B, and the main results are briefly presented here.

Writing the theory in the natural dimensionless time scale $\tau = t/N$, the effective scalar process H associated with a representative sample obeys

$$H(\tau) = \zeta(\tau) - \frac{4}{\alpha}\int_0^\tau du\, G(\tau, u)\phi(u),$$
$$\phi(\tau) = H(\tau)\left[H^2(\tau) + \left(3R^2(\tau) - 1\right)Z^2\right], \qquad \mathbb{E}\left[\zeta(\tau)\zeta(\tau')\right] = C(\tau, \tau'). \tag{5}$$

The corresponding deterministic equation for the teacher overlap is

$$\frac{dR}{d\tau} = -12R(\tau)\left[R^2(\tau) - 1 + \mathbb{E}_{Z,\zeta}\left[Z^2H^2(\tau)\right]\right]. \tag{6}$$

Here $Z \sim \mathcal{N}(0, 1)$ is the teacher component of a representative sample, $\zeta(\tau)$ is a centered Gaussian process with covariance C, $G(\tau, u)$ is the retarded response kernel, and ϕ is the nonlinear force

generated by the loss. The kernels C and G are determined self-consistently by the representative transverse process, with

$$s^2(\tau) = C(\tau, \tau). \tag{7}$$

The resulting equations predict the stability of the $R = 0$ states, the retrieval and failure trajectories, the critical overlap separating them and the location of stationary points in the (R, s) plane.

3 RESULTS

We begin by examining the model under standard random initialization, with the initial student weights drawn as $\boldsymbol{w}_0 \sim \mathcal{N}(\mathbf{0}, I_N)$. We denote by $P(N, \alpha)$ the probability that a random trajectory successfully retrieves the teacher. The empirical estimate of $P(N, \alpha)$ is obtained by averaging over 10^4 independent trials, as shown by the square markers in Figure 1(b). It is seen that the success probability increases smoothly over a broad range of α, while at higher dimensions more samples are required for generalization and the transition becomes narrower. We define the finite-size dynamical threshold $\alpha_{\rm GD}(N)$ by the relation $P(N, \alpha_{\rm GD}(N)) = \frac{1}{2}$, i.e. the sample size above which trajectories are more likely to generalize than not. Figure 1(c) shows that $\alpha_{\rm GD}(N)$ grows slowly with the dimension.

Figure 1(c) also compares $\alpha_{\rm GD}$ to the landscape-trivialization threshold $\alpha_{\rm triv}(N)$, defined as the value of α above which the landscape has no spurious minima; see Section 3.1 for details. Interestingly, even at α well above $\alpha_{\rm GD}$ – that is, with enough sample points so that typical trajectories generalize with high probability – spurious minima still exist, but they do not trap the dynamics. Lastly, Figure 1(d) shows that the retrieval time T varies widely between trajectories for a fixed N and α.

These observations naturally raise a few interesting questions: what determines whether a given trajectory retrieves the teacher? Why does retrieval succeed at finite N well below $\alpha_{\rm triv}(N)$ even though spurious minima are still present? What sets the distribution of retrieval times? And what happens in the thermodynamic limit $N \to \infty$?

The last question was recently studied for the non-symmetrized problem with a spherical constraint (Sarao Mannelli et al., 2020a; Bonnaire et al., 2025), where strong finite-size effects prevent a conclusive extrapolation of the numerics to $N \to \infty$. In the following sections we answer the first three theoretically: Section 3.2 identifies a critical initial overlap that separates retrieval from failure, Section 3.3 derives $P(N, \alpha)$ and the gap between $\alpha_{\rm GD}(N)$ and $\alpha_{\rm triv}(N)$ from it, and Section 3.4 derives the distribution of T.

Lastly, the big advantage of the symmetrized model is that the thermodynamic limit $N \to \infty$ can be rigorously solved with DMFT. The analysis is carried out fully in Appendix B and summarized in Section 2.1. It describes the gradient-flow dynamics as $N, M \to \infty$ at fixed $\alpha = M/N$. The results are indicated by blue dashed lines in the figures above, which also show that the two thresholds are well separated at finite N. As we show in Section 3.3, they approach the same limiting value, $\alpha_\infty \approx 26.86$, as $N \to \infty$ (Figure 1(b,c)): in the thermodynamic limit, gradient flow retrieves the teacher if and only if the landscape is trivial.

3.1 CONFINEMENT AND STABILITY OF SPURIOUS MINIMA

The symmetrization forces the loss to be an even function of R (as defined in Eq. (3)), making the manifold $R = 0$ invariant under gradient flow. The same symmetry trivially holds in the full-population landscape (Appendix C), which is symmetric by definition. In addition, this symmetry constrains the location of all spurious minima.

Proposition 1 (Confinement of spurious minima)**.** *If $M \geq N$, almost surely, the only local minima of the symmetrized loss with $R \neq 0$ are the two teacher solutions $(R, s) = (\pm 1, 0)$. All other local minima lie on $R = 0$.*

An elementary proof is given in Appendix A. Figure 2(a,b) illustrates how symmetrization confines all spurious minima to $R = 0$, leaving the teacher solutions as the only local minima at nonzero overlap.

Since all spurious minima lie on the $R = 0$ manifold, retrieval can fail only if these zero-overlap states are stable. We probe their stability by drawing a standard Gaussian initial student, projecting it

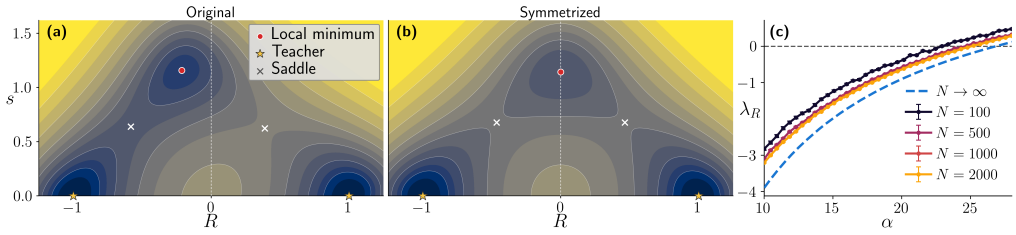


Figure 2: (a,b) Loss contours in (R, s)-plane with and without symmetrization. These are obtained by training for 10^4 training steps with $N = 1000$ and $\alpha = 3$ and fixing the resulting $\widehat{\boldsymbol{u}}$. The symmetrized loss is even in R, and spurious minima (red circles) are confined to $R = 0$. Stationary points at $R \neq 0$ are either saddles (crosses) or the global teacher minima (stars). (c) Average growth rate λ_R of stationary states on the symmetry manifold $R = 0$, for 100 trajectories per α and N (see text for explanation). The zero crossing $\lambda_R = 0$ defines $\alpha_{\text{triv}}(N)$.

onto the symmetry plane $R = 0$, and performing gradient flow dynamics constrained to the $R = 0$ plane, allowing the transverse coordinates to relax. The trajectories converge to a stationary point on the symmetry plane. We then quantify the growth rate in the teacher direction, denoted by λ_R, by numerically evaluating the negative curvature of the loss along the teacher direction (the first derivative vanishes by construction, from symmetry). Figure 2(c) shows λ_R, averaged over 10^2 independent datasets and initial conditions. The value of α at which λ_R changes sign from negative (stable) to positive (unstable) defines the landscape-trivialization threshold $\alpha_{\text{triv}}(N)$, above which the stationary points on the symmetry plane lose stability and the loss has no spurious minima.

In the thermodynamic limit, the stability of $R = 0$ minima can be inferred from stationary replica-symmetric DMFT equations, as detailed in Section B.2. Within this closure, the nontrivial $R = 0$ stationary branch appears to be characterized by a unique self-consistent transverse norm $s_0(\alpha)$, and linearizing the DMFT dynamics around this branch gives the thermodynamic growth rate λ_R. Its zero crossing gives the limiting trivialization threshold $\alpha_\infty \simeq 26.86$, shown together with the finite-N results in Figure 2(c).

3.2 Retrieval basins and the critical overlap

As seen above, trajectories can generalize even when spurious minima remain in the landscape. This is most evident in the gap between $\alpha_{\text{triv}}(N)$ and $\alpha_{\text{GD}}(N)$ (Figure 1(c)), and the slow transition of $P(N, \alpha)$ from 0 to 1 over a wide range of α (Figure 1(b)). It is interesting to ask what determines whether a trajectory generalizes, and a natural candidate is to look at the overlap $R(t)$. To study this, we fix $R(0) = R_0$ and draw the other directions isotropically at random while maintaining $s(0) = \sqrt{1 - R_0^2}$, so that $\|\boldsymbol{w}(0)\| = \sqrt{N}$. An example of the resulting dynamics for two choices of R_0 is shown in Figure 3(a), projected to the (R, s) plane. It is seen that trajectories starting with small $|R_0|$ flow to spurious minima at $R = 0$, while larger initial overlaps lead to retrieval. The DMFT trajectories closely track the typical path of each ensemble, motivating their use to estimate the retrieval boundary in the thermodynamic limit.

Repeating this for different R_0 and N, we denote the retrieval probability at fixed R_0 by $P(N, \alpha \mid R_0)$, averaged over the training data and the random transverse direction of initialization. Figure 3(b) shows $P(N, \alpha \mid R_0)$ at $\alpha = 10$ for multiple values of N. It is seen that as N increases the transition sharpens, and Figure 3(c) shows that this sharpening shows a $\sqrt{N}$ scaling, under which functions for different values of N collapse.

We define the critical overlap $R_c(\alpha, N)$ by the relation $P\left(N, \alpha \,|\, R_c(\alpha, N)\right) = \frac{1}{2}$, and empirically measure its dependence on α and N. The results are shown in panel (d), demonstrating a very weak N-dependence, which allows us to write loosely $R_c(\alpha, N) = R_c(\alpha)$. It is seen that R_c decreases rapidly with α, so that, as expected, larger amounts of data require smaller initial overlaps for successful retrieval.

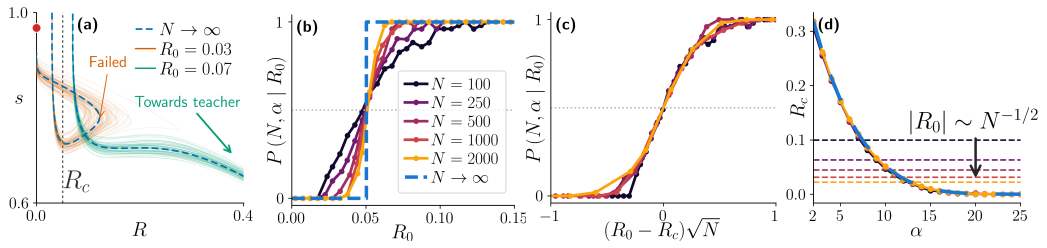


Figure 3: (a) 100 trajectories in the (R, s) plane for $R_0 = 0.03$ and 0.07, with $N = 1000$ and $\alpha = 10$. For these parameters, $R_c \approx 0.052$. Initial overlaps below R_c (orange) lead to failure, while those above R_c (green) lead to success. DMFT predictions for $N \to \infty$ (dashed blue) describe well the average of each ensemble. (b) Retrieval probability versus initial overlap R_0 at $\alpha = 10$, sharpening with increasing N toward the DMFT step at R_c. For each N, we ran $2 \cdot 10^4$ realizations over the range of overlaps studied. (c) The same curves collapse when plotted against $(R_0 - R_c)\sqrt{N}$, indicating a transition width proportional to $N^{-1/2}$. (d) Empirical estimates of $R_c(\alpha)$ from the curves of panel (b). R_c decreases with α, is largely independent of N and is in excellent agreement with the DMFT prediction. Horizontal dashed lines mark the typical random-initialization overlap $|R_0| \sim N^{-1/2}$. When the typical value of $|R_0|$ is above R_c, retrieval is likely to succeed.

Together with the sharpening of the crossover in Figure 3(c), these results suggest a sharp transition at $R_c(\alpha)$ in the thermodynamic limit

$$\lim_{N\to\infty} P(N, \alpha \mid R_0) = \begin{cases} 0, & |R_0| < R_c(\alpha), \\ 1, & |R_0| > R_c(\alpha). \end{cases} \tag{8}$$

Thus, in the thermodynamic limit, the initial overlap alone determines retrieval at fixed α. DMFT predicts this sharp transition, with the threshold obtained through binary search agreeing quantitatively with the limiting value at $\alpha = 10$ in Figure 3(b) and for all studied range of α in Figure 3(d).

3.3 Finite-size analysis of retrieval probability

The last discussion allows a quantitative prediction of the success probability, in terms of the competition between the typical initial overlap and the retrieval threshold $R_c(\alpha)$. At a given N, α, the probability of successful retrieval is a function of the initial overlap R_0 alone, and successful retrieval requires $|R_0| \sim N^{-1/2} > R_c(\alpha, N)$. Since R_0 is Gaussian with distribution $R_0 \simeq \mathcal{N}(0, N^{-1})$, the success probability can be approximated by the probability that $|R_0| > R_c(\alpha, N)$, which in turn can be approximated by the sharp DMFT boundary Eq. (8), since it depends only weakly on N. This gives

$$P(N, \alpha) \approx \text{Prob}\left(|R_0| > R_c(\alpha)\right) = \text{erfc}\left(\sqrt{\frac{N}{2}}\, R_c(\alpha)\right). \tag{9}$$

With $R_c(\alpha)$ obtained from DMFT, Eq. (9) provides a theoretical prediction with no fitting parameters, which closely agrees with experiments, as demonstrated in Figure 1(b). The shift of the $P(N, \alpha)$ curves in Figure 1(b) to higher α as N increases reflects an increase in $\alpha_{\text{GD}}(N)$, since the typical $|R_0|$ becomes smaller on average.

We stress that this highlights the gap between $\alpha_{\text{GD}}(N)$ and $\alpha_{\text{triv}}(N)$ at finite N: successful retrieval can occur even when $R_c(\alpha) > 0$, i.e. while stable spurious minima still exist. In the limit $N \to \infty$ these two thresholds coalesce. This happens because in this limit the initial overlap vanishes, and retrieval fails if and only if $R_c(\alpha) > 0$, i.e. if the zero-overlap minima are stable. The limiting value can be obtained from the stationary DMFT equations (Section B.2), which give

$$\alpha_{\text{GD}}(N),\ \alpha_{\text{triv}}(N) \xrightarrow[N\to\infty]{} \alpha_\infty \approx 26.86. \tag{10}$$

This common limit, marked by the blue dashed lines in Figure 1(b,c), establishes a thermodynamic sample complexity linear in the input dimension.

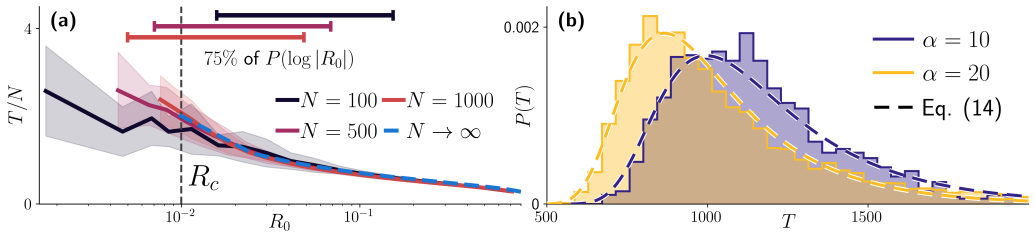


Figure 4: (a): T/N versus R_0 at $\alpha = 15$ for several dimensions. The solid lines show the average over 100 successful trajectories and the shaded region shows one standard deviation. The vertical dashed line marks the basin boundary R_c. Blue dashed curves show the $N \to \infty$ DMFT prediction. The horizontal bars span the central 75% of the distribution of R_0 under random initialization. (b) Histogram of the retrieval time T for $N = 1000$, with $\alpha = 10$ (purple) and $\alpha = 20$ (yellow), together with the density predicted by Eq. (14), estimating c_α by matching the median retrieval time from controlled-initialization simulations.

3.4 The initial overlap sets the retrieval time

Lastly, we show that the initial overlap determines not only the probability of retrieval, but also how long it takes. To analyze it, we examine the gradient-flow equation for the teacher overlap, which reads

$$\frac{dR}{dt} = \frac{4}{N} R \left[V_t (1 - R^2) - 3 V_m s^2 \right], \tag{11}$$

where $V_t = \left\langle (\boldsymbol{x}_\mu \cdot \boldsymbol{w}^\star / \sqrt{N})^4 \right\rangle$ and $V_m = \left\langle (\boldsymbol{x}_\mu \cdot \boldsymbol{w}^\star / \sqrt{N})^2 (\boldsymbol{x}_\mu \cdot \widehat{\boldsymbol{u}})^2 \right\rangle$ are empirical moments of the sample projections onto the teacher and the transverse direction, with $\langle \cdot \rangle$ denoting the average over the M samples. The derivation is given in Appendix A.

As argued above, the initial stages of training determine whether a trajectory will succeed or not. Since at short times $R \sim N^{-1/2}$ and $s \sim 1$, we posit that Eq. (11) can be approximated as

$$\frac{dR}{dt} \simeq c_\alpha R \qquad \Longrightarrow \qquad R(t) \simeq R_0 e^{c_\alpha t}, \tag{12}$$

where $c_\alpha = \mathcal{O}(N^{-1})$. In other words, we posit that the mechanism that determines the retrieval time is the exponential escape from the symmetry plane at short times.

To test this quantitatively, we define the retrieval time T as the time $|R(t)|$ reaches a fixed threshold $R_{\text{ret}} = 0.99$. Eq. (12) predicts that

$$T(R_0) \simeq \frac{1}{c_\alpha} \log \left(\frac{R_{\text{ret}}}{|R_0|} \right). \tag{13}$$

Indeed, Figure 4(a) shows that T has a simple, approximately logarithmic dependence on $|R_0|$, and that curves for different dimensions collapse after rescaling by N. In addition, T closely follows the DMFT prediction, indicating that retrieval time grows proportionally to N for a fixed initial overlap.

Moreover, most random initializations sample only a narrow range of R_0: the horizontal bars in Figure 4(a) mark the central 75% of this distribution for each N. The weak logarithmic dependence of T on R_0, together with the small relevant range of R_0, motivates using a single effective positive rate c_α to describe the escape time. Combining Eq. (13) with $R_0 \simeq \mathcal{N}(0, N^{-1})$ gives an approximate retrieval-time density. Absorbing the fixed retrieval threshold into the time origin yields

$$p(T) = c_\alpha \sqrt{\frac{2N}{\pi}} \, e^{-c_\alpha T} \exp \left[-\frac{N}{2} e^{-2 c_\alpha T} \right]. \tag{14}$$

For each α, we estimate c_α by matching the predicted median retrieval time to that of controlled-initialization simulations reweighted to represent random initialization. With this independently estimated rate, the predicted density closely matches the histograms at $N = 1000$ for $\alpha = 10$, 15 and 20 (Figure 4(b) and Figure 1(d)). It captures the asymmetric peak, its shift with α, and the long right tail.

As N increases, the distribution shifts toward smaller initial overlaps and therefore longer retrieval times. Deviations are largest near R_c, where trajectories linger near the basin boundary and the constant-rate approximation becomes less accurate. Finite-size fluctuations also allow some realizations with $|R_0| < R_c$ to retrieve the teacher. We also note in passing that the exact full-population solution yields the same logarithmic dependence on $|R_0|$ (Appendix C), providing a limiting check on this approximation.

More generally, retrieval-time distributions may help characterize learning dynamics even when no scalar coordinate such as R is known to organize the trajectories. Our calculation suggests that the shape of the distribution can provide evidence for an underlying multiplicative-growth mechanism and an estimate of its effective rate. Such statistics could therefore offer a way to study more complex systems where the relevant coordinate is difficult to identify or measure.

4 Discussion

We introduced symmetrized phase retrieval as a controlled setting for studying how loss-landscape geometry relates to learning dynamics. We showed that reflection symmetry confines all spurious minima to the zero-overlap manifold, and that the initial overlap with the teacher largely determines both whether a trajectory retrieves the teacher and the dynamical time scale of retrieval. All this is explained in terms of the competition between the critical overlap $R_c(\alpha)$ and the initialization scale $N^{-1/2}$, which also quantitatively explains why retrieval precedes landscape trivialization: At finite dimension, some initializations lie outside the failure basin even while spurious minima remain stable. As the dimension grows, the initial overlap decreases, and our analysis predicts that the dynamical and trivialization thresholds approach a common limit, $\alpha_\infty \approx 26.86$, obtained within the stationary replica-symmetric DMFT closure.

These findings provide an explicit mechanism for learning in the presence of spurious minima and address, within the symmetrized model, the distinction between successful retrieval and the instability of nongeneralizing states raised in studies of standard phase retrieval (Sarao Mannelli et al., 2020a; Bonnaire et al., 2025; Maillard et al., 2026). They also show why retrieval observed at finite dimensions can substantially underestimate the sample ratio required in the thermodynamic limit.

Lastly, we also found that in the symmetrized problem the initial overlap determines the retrieval time through an approximate exponential-growth model. This model, combined with the known Gaussian initialization statistics, gives excellent predictions of the retrieval-time distribution. This agreement arises because the retrieval time depends only logarithmically on the initial overlap, which is narrowly distributed under standard initialization. It is interesting to ask whether a similar situation exist in more complicated learning problems where a proper order parameter is not known. If so, learning-time distributions might serve as a useful diagnostic for a common mechanism that determines the dynamical statistics. This question we will attempt to answer in future work.

Our study relies on exact symmetrization which requires knowing the teacher direction, and therefore provides a theoretical setting for isolating these mechanisms rather than a practical procedure. An important next step is to determine whether the mechanism identified here is also relevant for standard phase retrieval, or whether additional features of its basin geometry are essential. Imperfect pairing, noisy labels, and stochastic training would allow us to test the robustness of our conclusions. More broadly, identifying data symmetries that simplify learning without requiring knowledge of the teacher could extend this approach to more general learning problems.

AI use statement

In this work, we used generative AI tools to assist with research ideation and execution. Additionally, we used generative AI tools to aid and polish the writing, including improving clarity, grammar, and phrasing. We have reviewed all AI-assisted work. We take responsibility for the final content of this work, including text, claims, or artifacts produced with the aid of generative AI.

References

Fabrizio Antenucci, Silvio Franz, Pierfrancesco Urbani, and Lenka Zdeborová. Glassy Nature of the Hard Phase in Inference Problems. *Physical Review X*, 9(1):011020, 2019. doi: 10.1103/physrevx.

9.011020. URL https://doi.org/10.1103/physrevx.9.011020.

Benjamin Aubin, Bruno Loureiro, Antoine Baker, Florent Krzakala, and Lenka Zdeborová. Exact Asymptotics for Phase Retrieval and Compressed Sensing with Random Generative Priors. In *Proceedings of the First Mathematical and Scientific Machine Learning Conference*, volume 107 of *Proceedings of Machine Learning Research*, pp. 55–73. PMLR, 2020. URL https://proceedings.mlr.press/v107/aubin20a.html.

Antonio Auffinger, Gérard Ben Arous, and Jiří Černý. Random Matrices and Complexity of Spin Glasses. *Communications on Pure and Applied Mathematics*, 66(2):165–201, 2013. doi: 10.1002/cpa.21422. URL https://doi.org/10.1002/cpa.21422.

Radu Balan, Pete Casazza, and Dan Edidin. On signal reconstruction without phase. *Applied and Computational Harmonic Analysis*, 20(3):345–356, 2006. doi: 10.1016/j.acha.2005.07.001. URL https://doi.org/10.1016/j.acha.2005.07.001.

Jean Barbier, Florent Krzakala, Nicolas Macris, Léo Miolane, and Lenka Zdeborová. Optimal Errors and Phase Transitions in High-Dimensional Generalized Linear Models. *Proceedings of the National Academy of Sciences*, 116(12):5451–5460, 2019. doi: 10.1073/pnas.1802705116. URL https://doi.org/10.1073/pnas.1802705116.

Benjamin Bloem-Reddy and Yee Whye Teh. Probabilistic Symmetries and Invariant Neural Networks. *Journal of Machine Learning Research*, 21(90):1–61, 2020. URL https://jmlr.org/papers/v21/19-322.html.

Tony Bonnaire, Giulio Biroli, and Chiara Cammarota. The Role of the Time-Dependent Hessian in High-Dimensional Optimization. *Journal of Statistical Mechanics: Theory and Experiment*, 2025 (8):083401, 2025. doi: 10.1088/1742-5468/adf4bc. URL https://doi.org/10.1088/1742-5468/adf4bc.

Blake Bordelon and Cengiz Pehlevan. Self-Consistent Dynamical Field Theory of Kernel Evolution in Wide Neural Networks. In *Advances in Neural Information Processing Systems*, volume 35, pp. 32240–32256. Curran Associates, Inc., 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/hash/d027a5c93d484a4312cc486d399c62c1-Abstract-Conference.html.

Emmanuel J. Candès and Xiaodong Li. Solving Quadratic Equations via PhaseLift When There Are About as Many Equations as Unknowns. *Foundations of Computational Mathematics*, 14(5):1017–1026, 2014. doi: 10.1007/s10208-013-9162-z. URL https://doi.org/10.1007/s10208-013-9162-z.

Emmanuel J. Candès, Yonina C. Eldar, Thomas Strohmer, and Vladislav Voroninski. Phase Retrieval via Matrix Completion. *SIAM Journal on Imaging Sciences*, 6(1):199–225, 2013a. doi: 10.1137/110848074. URL https://doi.org/10.1137/110848074.

Emmanuel J. Candès, Thomas Strohmer, and Vladislav Voroninski. PhaseLift: Exact and Stable Signal Recovery from Magnitude Measurements via Convex Programming. *Communications on Pure and Applied Mathematics*, 66(8):1241–1274, 2013b. doi: 10.1002/cpa.21432. URL https://doi.org/10.1002/cpa.21432.

Emmanuel J. Candès, Xiaodong Li, and Mahdi Soltanolkotabi. Phase Retrieval via Wirtinger Flow: Theory and Algorithms. *IEEE Transactions on Information Theory*, 61(4):1985–2007, 2015. doi: 10.1109/TIT.2015.2399924. URL https://doi.org/10.1109/TIT.2015.2399924.

Shuxiao Chen, Edgar Dobriban, and Jane H. Lee. A Group-Theoretic Framework for Data Augmentation. *Journal of Machine Learning Research*, 21(245):1–71, 2020. URL https://jmlr.org/papers/v21/20-163.html.

Yuxin Chen and Emmanuel J. Candès. Solving Random Quadratic Systems of Equations Is Nearly as Easy as Solving Linear Systems. *Communications on Pure and Applied Mathematics*, 70(5):822–883, 2017. doi: 10.1002/cpa.21638. URL https://doi.org/10.1002/cpa.21638.

Yuxin Chen, Yuejie Chi, Jianqing Fan, and Cong Ma. Gradient Descent with Random Initialization: Fast Global Convergence for Nonconvex Phase Retrieval. *Mathematical Programming*, 176:5–37, 2019. doi: 10.1007/s10107-019-01363-6. URL https://doi.org/10.1007/s10107-019-01363-6.

SueYeon Chung, Daniel D. Lee, and Haim Sompolinsky. Classification and Geometry of General Perceptual Manifolds. *Physical Review X*, 8:031003, 2018. doi: 10.1103/PhysRevX.8.031003. URL https://doi.org/10.1103/PhysRevX.8.031003.

Taco Cohen and Max Welling. Group Equivariant Convolutional Networks. In *Proceedings of The 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pp. 2990–2999. PMLR, 2016. URL https://proceedings.mlr.press/v48/cohenc16.html.

Tri Dao, Albert Gu, Alexander Ratner, Virginia Smith, Chris De Sa, and Christopher Ré. A Kernel Theory of Modern Data Augmentation. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 1528–1537. PMLR, 2019. URL https://proceedings.mlr.press/v97/dao19b.html.

Bryn Elesedy and Sheheryar Zaidi. Provably Strict Generalisation Benefit for Equivariant Models. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 2959–2969. PMLR, 2021. URL https://proceedings.mlr.press/v139/elesedy21a.html.

Albert Fannjiang and Thomas Strohmer. The numerics of phase retrieval. *Acta Numerica*, 29: 125–228, 2020. doi: 10.1017/s0962492920000069. URL https://doi.org/10.1017/s0962492920000069.

J. R. Fienup. Phase retrieval algorithms: a comparison. *Applied Optics*, 21(15):2758–2769, 1982. doi: 10.1364/ao.21.002758. URL https://doi.org/10.1364/ao.21.002758.

Yan V. Fyodorov and Pierre Le Doussal. Topology Trivialization and Large Deviations for the Minimum in the Simplest Random Optimization. *Journal of Statistical Physics*, 154(1-2): 466–490, 2014. doi: 10.1007/s10955-013-0838-1. URL https://doi.org/10.1007/s10955-013-0838-1.

Federica Gerace, Bruno Loureiro, Florent Krzakala, Marc Mézard, and Lenka Zdeborová. Generalisation Error in Learning with Random Features and the Hidden Manifold Model. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 3452–3462. PMLR, 2020. URL https://proceedings.mlr.press/v119/gerace20a.html.

Sebastian Goldt, Marc Mézard, Florent Krzakala, and Lenka Zdeborová. Modeling the Influence of Data Structure on Learning in Neural Networks: The Hidden Manifold Model. *Physical Review X*, 10:041044, 2020. doi: 10.1103/PhysRevX.10.041044. URL https://doi.org/10.1103/PhysRevX.10.041044.

Chi Jin, Rong Ge, Praneeth Netrapalli, Sham M. Kakade, and Michael I. Jordan. How to Escape Saddle Points Efficiently. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pp. 1724–1732. PMLR, 2017. URL https://proceedings.mlr.press/v70/jin17a.html.

Jason D. Lee, Max Simchowitz, Michael I. Jordan, and Benjamin Recht. Gradient Descent Only Converges to Minimizers. In *29th Annual Conference on Learning Theory*, volume 49 of *Proceedings of Machine Learning Research*, pp. 1246–1257. PMLR, 2016. URL https://proceedings.mlr.press/v49/lee16.html.

Bruno Loureiro, Cédric Gerbelot, Hugo Cui, Sebastian Goldt, Florent Krzakala, Marc Mézard, and Lenka Zdeborová. Learning curves of generic features maps for realistic datasets with a teacher-student model. *Journal of Statistical Mechanics: Theory and Experiment*, 2022(11):114001, 2022. doi: 10.1088/1742-5468/ac9825. URL https://doi.org/10.1088/1742-5468/ac9825.

Yue M. Lu and Gen Li. Phase Transitions of Spectral Initialization for High-Dimensional Non-Convex Estimation. *Information and Inference: A Journal of the IMA*, 9(3):507–541, 2020. doi: 10.1093/imaiai/iaz020. URL https://doi.org/10.1093/imaiai/iaz020.

Clare Lyle, Mark van der Wilk, Marta Kwiatkowska, Yarin Gal, and Benjamin Bloem-Reddy. On the Benefits of Invariance in Neural Networks. *arXiv preprint arXiv:2005.00178*, 2020. URL https://arxiv.org/abs/2005.00178.

Cong Ma, Kaizheng Wang, Yuejie Chi, and Yuxin Chen. Implicit Regularization in Nonconvex Statistical Estimation: Gradient Descent Converges Linearly for Phase Retrieval, Matrix Completion, and Blind Deconvolution. *Foundations of Computational Mathematics*, 20:451–632, 2020. doi: 10.1007/s10208-019-09429-9. URL https://doi.org/10.1007/s10208-019-09429-9.

Antoine Maillard, Gérard Ben Arous, and Giulio Biroli. Landscape Complexity for the Empirical Risk of Generalized Linear Models. In *Proceedings of the First Mathematical and Scientific Machine Learning Conference*, volume 107 of *Proceedings of Machine Learning Research*, pp. 287–327. PMLR, 2020a. URL https://proceedings.mlr.press/v107/maillard20a.html.

Antoine Maillard, Bruno Loureiro, Florent Krzakala, and Lenka Zdeborová. Phase Retrieval in High Dimensions: Statistical and Computational Phase Transitions. In *Advances in Neural Information Processing Systems*, volume 33, pp. 11071–11082. Curran Associates, Inc., 2020b. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/7ec0dbeee45813422897e04ad8424a5e-Paper.pdf.

Antoine Maillard, Florent Krzakala, Yue M. Lu, and Lenka Zdeborová. Construction of Optimal Spectral Methods in Phase Retrieval. In *Proceedings of the 2nd Mathematical and Scientific Machine Learning Conference*, volume 145 of *Proceedings of Machine Learning Research*, pp. 693–720. PMLR, 2022. URL https://proceedings.mlr.press/v145/maillard22a.html.

Antoine Maillard, Tony Bonnaire, and Giulio Biroli. Topological Exploration of High-Dimensional Empirical Risk Landscapes: General Approach, and Applications to Phase Retrieval. *arXiv preprint arXiv:2602.17779*, 2026. URL https://arxiv.org/abs/2602.17779.

Song Mei, Theodor Misiakiewicz, and Andrea Montanari. Learning with invariances in random features and kernel models. In *Proceedings of Thirty Fourth Conference on Learning Theory*, volume 134 of *Proceedings of Machine Learning Research*, pp. 3351–3418. PMLR, 2021. URL https://proceedings.mlr.press/v134/mei21a.html.

Francesca Mignacco, Florent Krzakala, Pierfrancesco Urbani, and Lenka Zdeborová. Dynamical Mean-Field Theory for Stochastic Gradient Descent in Gaussian Mixture Classification. In *Advances in Neural Information Processing Systems*, volume 33, pp. 9540–9550. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/6c81c83c4bd0b58850495f603ab45a93-Paper.pdf.

Francesca Mignacco, Pierfrancesco Urbani, and Lenka Zdeborová. Stochasticity Helps to Navigate Rough Landscapes: Comparing Gradient-Descent-Based Algorithms in the Phase Retrieval Problem. *Machine Learning: Science and Technology*, 2(3):035029, 2021. doi: 10.1088/2632-2153/ac0615. URL https://doi.org/10.1088/2632-2153/ac0615.

Marco Mondelli and Andrea Montanari. Fundamental Limits of Weak Recovery with Applications to Phase Retrieval. *Foundations of Computational Mathematics*, 19:703–773, 2019. doi: 10.1007/s10208-018-9395-y. URL https://doi.org/10.1007/s10208-018-9395-y.

Praneeth Netrapalli, Prateek Jain, and Sujay Sanghavi. Phase Retrieval using Alternating Minimization. In *Advances in Neural Information Processing Systems*, volume 26. Curran Associates, Inc., 2013. URL https://proceedings.neurips.cc/paper_files/paper/2013/file/242c100dc94f871b6d7215b868a875f8-Paper.pdf.

Stefano Sarao Mannelli, Giulio Biroli, Chiara Cammarota, Florent Krzakala, and Lenka Zdeborová. Who is Afraid of Big Bad Minima? Analysis of gradient-flow in spiked matrix-tensor models. In *Advances in Neural Information Processing Systems*, volume 32. Curran

Associates, Inc., 2019a. URL https://proceedings.neurips.cc/paper_files/paper/2019/file/fbad540b2f3b5638a9be9aa6a4d8e450-Paper.pdf.

Stefano Sarao Mannelli, Florent Krzakala, Pierfrancesco Urbani, and Lenka Zdeborová. Passed & Spurious: Descent Algorithms and Local Minima in Spiked Matrix-Tensor Models. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 4333–4342. PMLR, 2019b. URL https://proceedings.mlr.press/v97/mannelli19a.html.

Stefano Sarao Mannelli, Giulio Biroli, Chiara Cammarota, Florent Krzakala, Pierfrancesco Urbani, and Lenka Zdeborová. Complex Dynamics in Simple Neural Networks: Understanding Gradient Flow in Phase Retrieval. In *Advances in Neural Information Processing Systems*, volume 33, pp. 3265–3274. Curran Associates, Inc., 2020a. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/2172fde49301047270b2897085e4319d-Paper.pdf.

Stefano Sarao Mannelli, Giulio Biroli, Chiara Cammarota, Florent Krzakala, Pierfrancesco Urbani, and Lenka Zdeborová. Marvels and Pitfalls of the Langevin Algorithm in Noisy High-Dimensional Inference. *Physical Review X*, 10(1):011057, 2020b. doi: 10.1103/physrevx.10.011057. URL https://doi.org/10.1103/physrevx.10.011057.

Yoav Shechtman, Yonina C. Eldar, Oren Cohen, Henry N. Chapman, Jianwei Miao, and Mordechai Segev. Phase Retrieval with Application to Optical Imaging: A Contemporary Overview. *IEEE Signal Processing Magazine*, 32(3):87–109, 2015. doi: 10.1109/MSP.2014.2352673. URL https://doi.org/10.1109/MSP.2014.2352673.

Ju Sun, Qing Qu, and John Wright. A Geometric Analysis of Phase Retrieval. *Foundations of Computational Mathematics*, 18(5):1131–1198, 2018. doi: 10.1007/s10208-017-9365-9. URL https://doi.org/10.1007/s10208-017-9365-9.

Theodoros G. Tsironis and Aris L. Moustakas. Landscape Complexity for the Empirical Risk of Generalized Linear Models: Discrimination between Structured Data. *Physical Review E*, 112: 065307, 2025. doi: 10.1103/3mbj-xkgk. URL https://doi.org/10.1103/3mbj-xkgk.

Gang Wang, Georgios B. Giannakis, and Yonina C. Eldar. Solving Systems of Random Quadratic Equations via Truncated Amplitude Flow. *IEEE Transactions on Information Theory*, 64(2): 773–794, 2018. doi: 10.1109/tit.2017.2756858. URL https://doi.org/10.1109/tit.2017.2756858.

A Landscape reduction and confinement of spurious minima

We first express the symmetrized loss at fixed transverse direction, then prove the confinement of spurious minima using its quadratic dependence on the squared amplitudes. We also record the gradient-flow and stability identities used in the main text.

Reduction of the symmetrized loss. Fix a unit transverse direction $\widehat{\mathbf{u}}$ satisfying $\widehat{\mathbf{u}} \cdot \mathbf{w}^\star = 0$, and define

$$Z_\mu = \frac{1}{\sqrt{N}} \mathbf{x}_\mu \cdot \mathbf{w}^\star, \qquad z_\mu = \mathbf{x}_\mu \cdot \widehat{\mathbf{u}}. \tag{15}$$

In the normalized coordinates (R, s), the student predictions on the two members of a reflected pair are

$$\widehat{y}_\mu^{\pm} = (RZ_\mu \pm sz_\mu)^2, \tag{16}$$

while their common label is Z_μ^2. Writing $\langle f_\mu \rangle = M^{-1} \sum_{\mu=1}^{M} f_\mu$, the pair-averaged loss is

$$\begin{aligned} L(R, s; \widehat{\mathbf{u}}) &= \frac{1}{2} \left\langle \left[Z_\mu^2 - (RZ_\mu + sz_\mu)^2\right]^2 + \left[Z_\mu^2 - (RZ_\mu - sz_\mu)^2\right]^2 \right\rangle \\ &= V_t (R^2 - 1)^2 + 2V_m (3R^2 - 1)s^2 + V_\perp s^4, \end{aligned} \tag{17}$$

where

$$V_t = \langle Z_\mu^4 \rangle, \qquad V_m = \langle Z_\mu^2 z_\mu^2 \rangle, \qquad V_\perp = \langle z_\mu^4 \rangle. \tag{18}$$

The dependence on $\hat{\mathbf{u}}$ is contained in V_m and $V_\perp$; V_t is independent of the transverse direction. Pair averaging cancels all terms odd in z_μ, making the loss even in R.

Proof of Proposition 1. At fixed $\hat{\mathbf{u}}$, introduce the squared amplitudes $a = R^2$ and $b = s^2$. The reduced loss becomes

$$F(a,b) = V_t(a-1)^2 + 2V_m(3a-1)b + V_\perp b^2. \tag{19}$$

This is a polynomial of degree at most two in (a,b), and both teacher solutions correspond to $(a,b) = (1,0)$, where $F(1,0) = 0$.

Let $\overline{\mathbf{w}}$ be a local minimum with $\overline{R} \neq 0$, and write $\bar{a} = \overline{R}^2 > 0$ and $\bar{b} = \bar{s}^2 \geq 0$. Fix its transverse direction $\hat{\mathbf{u}}$; if $\bar{s} = 0$, this direction may be chosen arbitrarily. Consider the straight path in squared-amplitude coordinates

$$a(t) = (1-t)\bar{a} + t, \qquad b(t) = (1-t)\bar{b}, \qquad 0 \leq t \leq 1, \tag{20}$$

and its lift to weight space,

$$\mathbf{w}(t) = \operatorname{sgn}(\overline{R})\sqrt{a(t)}\,\mathbf{w}^\star + \sqrt{N}\sqrt{b(t)}\,\hat{\mathbf{u}}. \tag{21}$$

This path starts at $\overline{\mathbf{w}}$ and ends at the teacher solution with the same sign of overlap.

Crucially, the path also extends to an open neighborhood of $t = 0$: since $\bar{a} > 0$, we have $a(t) > 0$ for sufficiently small $|t|$, while $b(t) = (1-t)\bar{b} \geq 0$ there. This remains true when $\bar{b} = 0$, in which case $b(t)$ vanishes identically. Thus $\mathbf{w}(t)$ is continuous near $t = 0$, and local minimality of $\overline{\mathbf{w}}$ implies that $t = 0$ is an interior local minimum of

$$g(t) = L(\mathbf{w}(t)) = F(a(t), b(t)). \tag{22}$$

Because F is quadratic and $a(t), b(t)$ are affine, g is a polynomial of degree at most two. Such a polynomial attains its global minimum at any interior local minimum. Consequently,

$$0 \leq L(\overline{\mathbf{w}}) = g(0) \leq g(1) = F(1,0) = 0. \tag{23}$$

Hence every local minimum with nonzero overlap has zero loss.

It remains to identify the zero-loss solutions with $R \neq 0$. For $M \geq N$, Gaussian samples almost surely span $\mathbb{R}^N$, and $Z_\mu \neq 0$ for every μ. Since the loss is an average of nonnegative squared errors, zero loss implies

$$(\overline{R}Z_\mu + \bar{s}z_\mu)^2 = Z_\mu^2, \qquad (\overline{R}Z_\mu - \bar{s}z_\mu)^2 = Z_\mu^2 \qquad \text{for every } \mu. \tag{24}$$

Subtracting these equations gives

$$4\overline{R}\,\bar{s}\,Z_\mu z_\mu = 0. \tag{25}$$

Because $\overline{R} \neq 0$ and $Z_\mu \neq 0$, it follows that

$$\mathbf{x}_\mu \cdot (\bar{s}\,\hat{\mathbf{u}}) = 0 \qquad \text{for every } \mu. \tag{26}$$

The samples span $\mathbb{R}^N$, so $\bar{s}\,\hat{\mathbf{u}} = 0$, and therefore $\bar{s} = 0$. The zero-loss equations then reduce to $\overline{R}^2 Z_\mu^2 = Z_\mu^2$, giving $\overline{R}^2 = 1$.

Thus the only local minima with nonzero overlap are the two teacher solutions $(R,s) = (\pm 1, 0)$. All other local minima must lie on $R = 0$. □

Gradient-flow identities. Differentiating Eq. (17) at fixed $\hat{\mathbf{u}}$ gives

$$\begin{aligned} \frac{\partial L}{\partial R} &= 4R\big[V_t(R^2-1) + 3V_m s^2\big], \\ \frac{\partial L}{\partial s} &= 4s\big[V_m(3R^2-1) + V_\perp s^2\big]. \end{aligned} \tag{27}$$

With the learning rate absorbed into time, the corresponding overlap and transverse-norm dynamics are

$$\begin{aligned} \frac{dR}{dt} &= \frac{4}{N}R\big[V_t(1-R^2) - 3V_m s^2\big], \\ \frac{ds}{dt} &= \frac{4}{N}s\big[V_m(1-3R^2) - V_\perp s^2\big]. \end{aligned} \tag{28}$$

These identities hold along the full gradient flow, with V_m and $V_\perp$ evaluated at the evolving transverse direction. The first equation is the overlap equation used in Section 3.4 and shows that $R = 0$ is invariant.

Stability on the zero-overlap manifold. At $R = 0$, radial stationarity with $s_* > 0$ requires

$$s_*^2 = \frac{V_m}{V_\perp}. \tag{29}$$

Define

$$V_\Sigma = 3V_m^2 - V_t V_\perp. \tag{30}$$

The curvature in the overlap direction at such a point is

$$\left.\frac{\partial^2 L}{\partial R^2}\right|_{R=0,\, s=s_*} = -4V_t + \frac{12V_m^2}{V_\perp} = \frac{4V_\Sigma}{V_\perp}. \tag{31}$$

Accordingly, the linearized overlap dynamics obeys

$$\frac{d\,\delta R}{dt} = -\frac{4V_\Sigma}{NV_\perp}\,\delta R. \tag{32}$$

Thus $V_\Sigma > 0$ gives stability in the teacher direction, whereas $V_\Sigma < 0$ gives instability; $V_\Sigma = 0$ is marginal at linear order. These conditions concern the teacher direction only. A local minimum must additionally satisfy stationarity and stability with respect to variations of the transverse direction.

B DMFT DERIVATION, VALIDATION, AND STATIONARY GEOMETRY

We derive the dynamical mean-field equations used in the main text and their stationary replica-symmetric reduction. Throughout this appendix, time is measured by $\tau = t/N$, and a dot denotes differentiation with respect to τ.

B.1 DYNAMICAL MSRJD SADDLE

Writing the transverse component in Eq. (3) in an orthonormal basis of the teacher-orthogonal subspace, let $U_i(\tau) = \sqrt{N}s(\tau)\widehat{u}_i(\tau)$, for $i = 1, \ldots, N-1$, so that $s^2(\tau) = N^{-1}\sum_i U_i^2(\tau)$. Let ξ_i^μ denote the transverse components of the Gaussian input and $Z_\mu = N^{-1/2}\boldsymbol{x}_\mu \cdot \boldsymbol{w}^\star$ its teacher component. These variables are independent standard Gaussians. Define

$$\begin{aligned} A_{\mu i} &= \frac{\xi_i^\mu}{\sqrt{N}}, \qquad h_\mu(\tau) = \sum_{i=1}^{N-1} A_{\mu i} U_i(\tau), \\ \phi_\mu(\tau) &= h_\mu(\tau)\left[h_\mu^2(\tau) + (3R^2(\tau) - 1)Z_\mu^2\right]. \end{aligned} \tag{33}$$

The normalized student predictions on a reflected pair are $(RZ_\mu \pm h_\mu)^2$. Pair averaging therefore gives

$$L = \frac{1}{M}\sum_{\mu=1}^{M}\left[(R^2-1)^2 Z_\mu^4 + 2(3R^2-1)Z_\mu^2 h_\mu^2 + h_\mu^4\right].$$

Since $\partial L/\partial h_\mu = 4\phi_\mu/M$, the transverse gradient-flow dynamics in the time variable τ is

$$\dot{U}_i(\tau) = -\frac{4}{\alpha}\sum_{\mu=1}^{M} A_{\mu i}\phi_\mu(\tau). \tag{34}$$

We introduce sources $\theta_i(\tau)$ and $j_\mu(\tau)$ in the coordinate and local-field equations to generate response functions. Response fields $\widehat{U}_i$ and $\widehat{h}_\mu$ impose these equations through the Martin–Siggia–Rose–Janssen–de Dominicis (MSRJD) action

$$\begin{aligned} S = i\sum_i \int d\tau\, \widehat{U}_i\left[\dot{U}_i + \frac{4}{\alpha}\sum_\mu A_{\mu i}\phi_\mu - \theta_i\right] \\ + i\sum_\mu \int d\tau\, \widehat{h}_\mu\left[h_\mu - \sum_i A_{\mu i}U_i - j_\mu\right]. \end{aligned} \tag{35}$$

The physical dynamics is recovered by setting the sources to zero. A causal time discretization makes the MSRJD Jacobian a source-independent constant and fixes the strictly retarded convention used below.

The dependence on the transverse design matrix is linear in $A_{\mu i}$. Writing

$$K_{\mu i} = \frac{4}{\alpha}\widehat{U}_i\phi_\mu - \widehat{h}_\mu U_i,$$

the Gaussian average gives

$$\mathbb{E}_A \exp\left[i\sum_{\mu i} A_{\mu i}\int d\tau\, K_{\mu i}(\tau)\right] = \exp\left[-\frac{1}{2N}\sum_{\mu i}\int d\tau\, d\tau'\, K_{\mu i}(\tau)K_{\mu i}(\tau')\right]. \tag{36}$$

The resulting quadratic terms are expressed through four macroscopic kernels,

$$\begin{aligned} C(\tau,\tau') &= \frac{1}{N}\sum_i U_i(\tau)U_i(\tau'), \qquad G(\tau,\tau') = -\frac{i}{N}\sum_i U_i(\tau)\widehat{U}_i(\tau'), \\ B(\tau,\tau') &= \frac{1}{M}\sum_\mu \phi_\mu(\tau)\phi_\mu(\tau'), \qquad X(\tau,\tau') = -\frac{i}{M}\sum_\mu \phi_\mu(\tau)\widehat{h}_\mu(\tau'). \end{aligned} \tag{37}$$

Here C and B are correlation kernels, while G and X become causal response kernels. Introducing their definitions with functional delta functions separates the coordinate and sample indices, giving

$$\overline{\mathcal{Z}} = \int \mathcal{D}Q\,\mathcal{D}\widehat{Q}\,\exp\left[N\mathcal{F}[Q,\widehat{Q}]\right], \qquad Q = \{C,G,B,X\}. \tag{38}$$

Conditional on the macroscopic kernels, the path integral factorizes into one-coordinate and one-sample contributions. In the proportional large-N limit, the macroscopic integral is evaluated at its saddle. Varying the action with respect to the kernels and their conjugates, followed by Hubbard–Stratonovich transformations of the quadratic response-field terms, gives the effective processes below.

The representative sample process is

$$\begin{cases} H(\tau) = \zeta(\tau) - \dfrac{4}{\alpha}\displaystyle\int_0^{\tau^-} du\, G(\tau,u)\phi(u), \\ \phi(\tau) = H(\tau)\left[H^2(\tau) + (3R^2(\tau)-1)Z^2\right], \\ Z\sim\mathcal{N}(0,1), \qquad \mathbb{E}[\zeta(\tau)\zeta(\tau')] = C(\tau,\tau'), \qquad Z\perp\zeta. \end{cases} \tag{39}$$

Here ζ is a centered Gaussian process. This reproduces the effective field equation in Eq. (5).

To close the response sector, add the source $j(\tau)$ to the right-hand side of the sample equation and define

$$D(\tau,\tau') = \left.\frac{\delta H(\tau)}{\delta j(\tau')}\right|_{j=0}, \qquad \kappa(\tau) = 3H^2(\tau) + (3R^2(\tau)-1)Z^2. \tag{40}$$

Differentiation at fixed macroscopic kernels gives

$$D(\tau,\tau') = \delta_{\mathrm{D}}(\tau-\tau') - \frac{4}{\alpha}\int_0^{\tau^-} du\, G(\tau,u)\kappa(u)D(u,\tau'). \tag{41}$$

The corresponding force kernels are

$$\begin{aligned} B(\tau,\tau') &= \mathbb{E}_{Z,\zeta}[\phi(\tau)\phi(\tau')], \qquad X(\tau,\tau') = \mathbb{E}_{Z,\zeta}[\kappa(\tau)D(\tau,\tau')], \\ X(\tau,\tau') &= \chi(\tau)\delta_{\mathrm{D}}(\tau-\tau') + X_{\mathrm{ret}}(\tau,\tau'), \qquad \chi(\tau) = \mathbb{E}_{Z,\zeta}[\kappa(\tau)], \end{aligned} \tag{42}$$

where $X_{\mathrm{ret}}(\tau,\tau') = 0$ for $\tau \le \tau'$.

The one-coordinate saddle gives a representative transverse coordinate $\mathcal{U}$ satisfying

$$\begin{cases} \dot{\mathcal{U}}(\tau) = -4\chi(\tau)\mathcal{U}(\tau) - 4\displaystyle\int_0^{\tau^-} du\, X_{\mathrm{ret}}(\tau,u)\mathcal{U}(u) + \omega(\tau) + \theta(\tau), \\ \mathbb{E}[\omega(\tau)\omega(\tau')] = \dfrac{16}{\alpha}B(\tau,\tau'), \qquad \mathcal{U}(0)\sim\mathcal{N}(0,s^2(0)). \end{cases} \tag{43}$$

Here ω is a centered Gaussian process. Its covariance contains the square of the microscopic force coefficient. The correlation and response kernels are fixed self-consistently by

$$C(\tau,\tau') = \mathbb{E}[\mathcal{U}(\tau)\mathcal{U}(\tau')], \qquad G(\tau,\tau') = \left.\frac{\delta \mathbb{E}[\mathcal{U}(\tau)]}{\delta\theta(\tau')}\right|_{\theta=0}. \tag{44}$$

In particular, $s^2(\tau) = C(\tau,\tau)$.

Finally, projecting the gradient flow onto the teacher direction gives

$$\dot{R} = -4R\left[(R^2-1)\frac{1}{M}\sum_\mu Z_\mu^4 + 3\frac{1}{M}\sum_\mu Z_\mu^2 h_\mu^2\right]. \tag{45}$$

At the large-N saddle, $M^{-1}\sum_\mu Z_\mu^4 \to \mathbb{E}[Z^4] = 3$ and $M^{-1}\sum_\mu Z_\mu^2 h_\mu^2 \to \mathbb{E}_{Z,\zeta}[Z^2H^2]$. Therefore,

$$\dot{R}(\tau) = -12R(\tau)\left[R^2(\tau) - 1 + \mathbb{E}_{Z,\zeta}[Z^2H^2(\tau)]\right]. \tag{46}$$

Equations (39), (41), (42), (43), (44), and (46) form the closed dynamical DMFT.

B.2 Stationary replica-symmetric reduction

The dynamical derivation above uses no replicas. We now adopt a replica-symmetric (RS) stationary closure, assuming convergence to a single fixed point with time-translation-invariant response kernels and finite integrated response. No equilibrium fluctuation–dissipation relation is required.

At a fixed point, $C(\tau,\tau') \to s^2$, so the Gaussian process ζ reduces to a static Gaussian variable $a = sg$, where $g \sim \mathcal{N}(0,1)$ is independent of Z. Define

$$\Gamma = \int_0^\infty du\, G(u), \qquad \mathcal{X} = \int_0^\infty du\, X(u), \qquad \Delta = \frac{4}{\alpha}\Gamma. \tag{47}$$

The integrals use the same one-sided causal convention as the dynamical equations, with the instantaneous contribution to X included in $\mathcal{X}$. The long-time sample equation becomes

$$a = H + \Delta\phi, \qquad \phi = H\left[H^2 + (3R^2-1)Z^2\right]. \tag{48}$$

On a smooth branch, differentiation at fixed R, Δ, and Z gives

$$\frac{\partial H}{\partial a} = \frac{1}{1+\Delta\kappa}, \qquad \kappa = 3H^2 + (3R^2-1)Z^2. \tag{49}$$

Within the stationary closure, the integrated force response is

$$\mathcal{X} = \mathbb{E}_{g,Z}\left[\frac{\partial\phi}{\partial a}\right] = \frac{1}{\Delta}\left[1 - \mathbb{E}_{g,Z}\left[\frac{\partial H}{\partial a}\right]\right]. \tag{50}$$

On smooth branches, $\partial\phi/\partial a = \kappa\,\partial H/\partial a$. The zero-frequency coordinate-response equation gives $4\Gamma\mathcal{X} = 1$. Combining this identity with Eqs. (47) and (50) yields

$$\mathbb{E}_{g,Z}\left[\frac{\partial H}{\partial a}\right] = 1 - \frac{1}{\alpha}, \qquad \mathcal{X} = \frac{1}{\alpha\Delta}. \tag{51}$$

At stationarity, $B(\tau,\tau') \to B_\star = \mathbb{E}_{g,Z}[\phi^2]$. The coordinate noise ω becomes a static Gaussian variable with variance $16B_\star/\alpha$, and the stationary coordinate equation gives $\mathcal{U} = \omega/(4\mathcal{X})$. Consequently,

$$s^2 = \frac{B_\star}{\alpha\mathcal{X}^2} = \alpha\Delta^2\mathbb{E}_{g,Z}\left[H^2\left[H^2 + (3R^2-1)Z^2\right]^2\right]. \tag{52}$$

Stationarity of the signal equation requires

$$0 = R\left[R^2 - 1 + \mathbb{E}_{g,Z}[Z^2H^2]\right]. \tag{53}$$

The scalar relation in Eq. (48) may have several roots. For the stationary results reported here, we select a stable minimizing branch of

$$V_{\text{eff}}(h;a,Z) = \frac{1}{2}(h-a)^2 + \Delta\left[\frac{1}{4}h^4 + \frac{1}{2}(3R^2-1)Z^2h^2\right]. \tag{54}$$

If the selected branch changes discontinuously, derivatives of H and ϕ with respect to a are interpreted distributionally and include the corresponding jump contributions. In particular, the smooth-branch expression in Eq. (49) must then be supplemented by those contributions when taking averages.

Collecting these relations, the regular RS stationary macrostates satisfy

$$\begin{cases} a = sg = H + \Delta H \left[H^2 + (3R^2 - 1)Z^2\right], \\ \mathbb{E}_{g,Z}\left[\dfrac{\partial H}{\partial a}\right] = 1 - \dfrac{1}{\alpha}, \\ s^2 = \alpha\Delta^2 \mathbb{E}_{g,Z}\left[H^2\left[H^2 + (3R^2 - 1)Z^2\right]^2\right], \\ 0 = R\left[R^2 - 1 + \mathbb{E}_{g,Z}[Z^2 H^2]\right]. \end{cases} \tag{55}$$

Here $g, Z \sim \mathcal{N}(0,1)$ are independent, Δ is the static response parameter defined in Eq. (47), and H is the selected branch described by Eq. (54).

Zero-overlap branch and signal stability. At $R = 0$, the final equation in Eq. (55) is satisfied identically. Excluding the trivial stationary point at the origin, the zero-overlap branch is determined by

$$\begin{cases} a = s_0 g = H_0 + \Delta_0 H_0 (H_0^2 - Z^2), \\ \mathbb{E}_{g,Z}\left[\dfrac{\partial H_0}{\partial a}\right] = 1 - \dfrac{1}{\alpha}, \\ s_0^2 = \alpha\Delta_0^2 \mathbb{E}_{g,Z}\left[H_0^2 (H_0^2 - Z^2)^2\right]. \end{cases} \tag{56}$$

Here the subscript 0 denotes the stationary zero-overlap branch. Numerically, the selected-branch system in Eq. (56) appears to yield a single nontrivial regular-RS solution $(s_0(\alpha), \Delta_0(\alpha))$ for each α in the range studied. This statement concerns the selected RS macrostate and does not exclude replica-symmetry-broken or microscopically distinct states outside this closure.

Because the transverse dynamics depends on R through R^2, an infinitesimal signal perturbation about $R = 0$ does not change the transverse process to first order. Linearizing Eq. (46) therefore gives

$$\dot{\delta R} = \lambda_R \, \delta R, \qquad \lambda_R = 12\left[1 - \mathbb{E}_{g,Z}[Z^2 H_0^2]\right]. \tag{57}$$

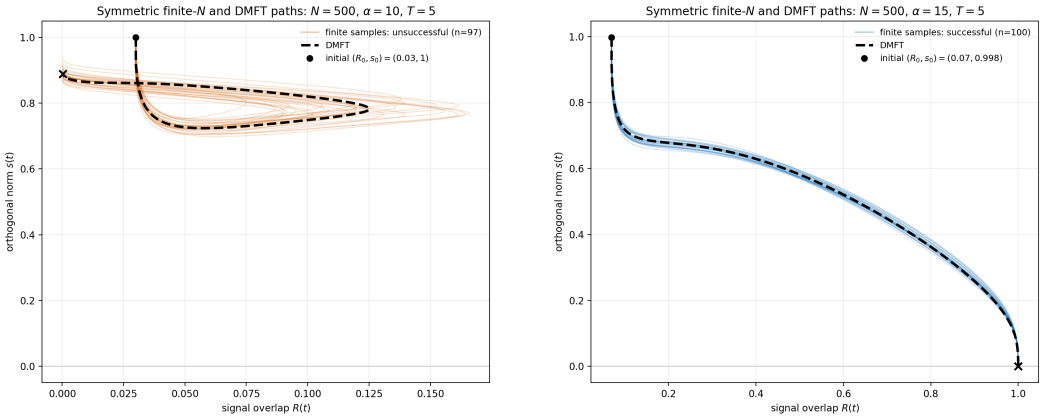
The zero-overlap state is stable in the teacher direction for $\lambda_R < 0$ and unstable for $\lambda_R > 0$. Within this stationary RS closure, the signal-instability threshold is obtained by solving Eq. (56) together with $\lambda_R = 0$, giving $\alpha_\infty \approx 26.86$. This is the stationary RS estimate of the thermodynamic trivialization threshold used in the main text.

B.3 Numerical validation and stationary geometry

We solve the dynamical equations by iterating between the sample and coordinate processes until the self-consistency conditions in Eq. (44) are satisfied.

Figure 5 compares the DMFT trajectories with finite-N paths in the (R, s) plane. For $\alpha = 10$, the displayed trajectory returns to the zero-overlap branch. For $\alpha = 15$ and sufficiently large initial overlap, it follows the retrieval branch toward the teacher. The comparison also resolves the non-monotone excursion on the failure branch.

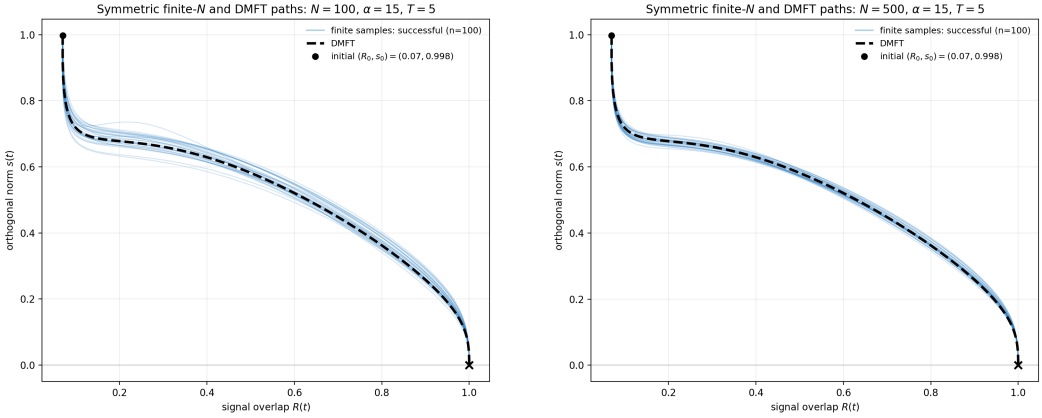
The dependence on dimension is shown in Figure 6. At $N = 100$, the sample paths fluctuate appreciably around the DMFT curve, whereas at $N = 500$ they form a narrower bundle. This concentration is consistent with the transverse self-averaging underlying the mean-field description.



(a) Failure branch at $\alpha = 10$.

(b) Retrieval branch at $\alpha = 15$.

Figure 5: Finite-N trajectories at $N = 500$ and the dashed DMFT predictions in the (R, s) plane, showing return to the symmetry manifold and successful retrieval for the two parameter choices.



(a) $N = 100$.

(b) $N = 500$.

Figure 6: Concentration of successful finite-N trajectories around the representative DMFT path at $\alpha = 15$. Sample-to-sample fluctuations shrink with increasing dimension.

The stationary projection in Figure 7 complements these dynamical comparisons. At each displayed value of α, the nontrivial point on $R = 0$ is the selected regular-RS solution $s_0(\alpha)$ described above. Alongside the two exact teacher minima and the maximum at the origin, the plot shows index-one saddle candidates obtained within the closure, whose positions vary with α. This macroscopic projection does not enumerate replica-symmetry-broken or microscopically distinct states with the same (R, s) coordinates.

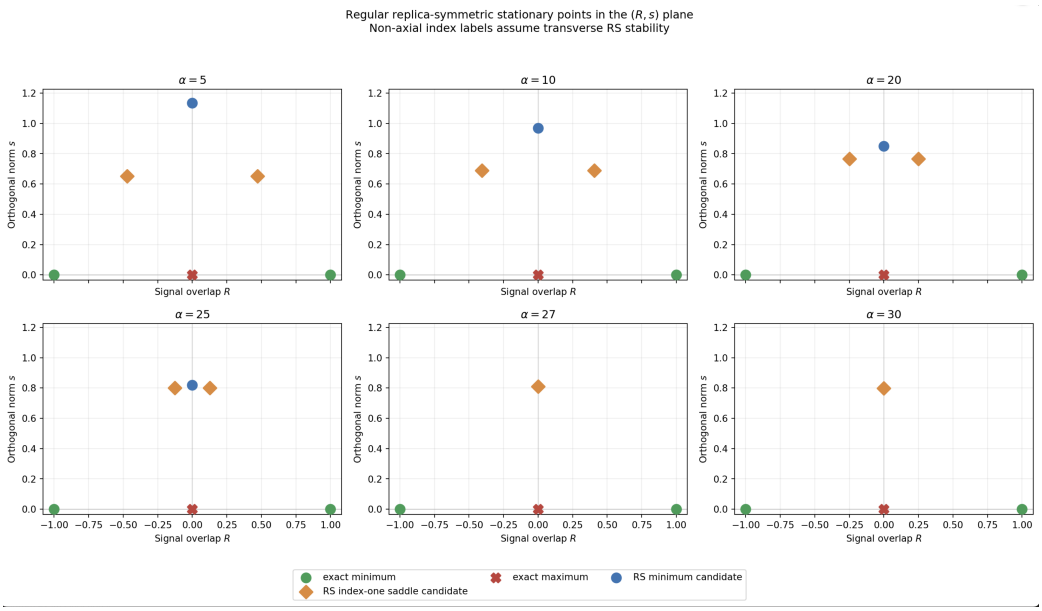


Figure 7: Regular RS stationary macrostates in the (R, s) plane for representative sample ratios. The nontrivial point on $R = 0$ is the selected branch $s_0(\alpha)$, while the motion of the index-one saddle candidates provides a static view of the changing retrieval geometry.

C Full-population dynamics and escape distribution

For a random transverse direction independent of the Gaussian data, $V_t \to 3$, $V_m \to 1$, and $V_\perp \to 3$ as the number of samples grows. The reduced loss therefore becomes

$$L_{\mathrm{FP}}(R, s) = 3(R^2 - 1)^2 + 2(3R^2 - 1)s^2 + 3s^4. \tag{58}$$

It is even in R, just like the finite-sample paired loss, so the reflection symmetry remains part of the full-population landscape.

Using the same R and s as in the main text, define $u = 8t/N$. The population flow is

$$\frac{dR}{du} = \frac{3}{2}(1 - R^2 - s^2)R, \qquad \frac{ds}{du} = \frac{1}{2}(1 - 3R^2 - 3s^2)s. \tag{59}$$

Its exact solution can be written as

$$R^2(u) = \frac{R_0^2 e^{3u}}{R_0^2 e^{3u} + 3s_0^2 e^u + 1 - R_0^2 - 3s_0^2}, \tag{60}$$

$$s^2(u) = \frac{s_0^2 e^u}{R_0^2 e^{3u} + 3s_0^2 e^u + 1 - R_0^2 - 3s_0^2}. \tag{61}$$

In particular,

$$\frac{R^2(u)}{s^2(u)} = \frac{R_0^2}{s_0^2} e^{2u}. \tag{62}$$

Thus $R = 0$ is invariant, while every nonzero overlap grows exponentially relative to the orthogonal component. Near the saddle $s^2 \simeq 1/3$, the exact solution reduces to $dR/dt = (8/N)R$. The population dynamics therefore produces the same logarithmic dependence on $|R_0|$ as the empirical approximation in Eq. (12).

For random initialization, $R_0 \sim \mathcal{N}(0, 1/N)$. If escape is measured at $|R| = R_{\mathrm{ret}}$, the large-N form of the exact solution in Eq. (60) gives

$$T_{\mathrm{FP}} \simeq \frac{1}{c_{\mathrm{FP}}} \log\left[\frac{\sqrt{3}\, R_{\mathrm{ret}}}{\sqrt{1 - (R_{\mathrm{ret}})^2}\, |R_0|}\right], \qquad c_{\mathrm{FP}} = \frac{8}{N}. \tag{63}$$

Applying this change of variables to the Gaussian distribution of R_0 gives the same functional form as Eq. (14). In the population problem, however, this distribution follows from the exact solution rather than from the empirical approximation $dR/dt \simeq c_\alpha R$.