Dynamic BBP for Multi-Index Phase Retrieval

Abstract

We study a quadratic multi-index teacher-student model and the empirical Hessian obsereved on an independent Gaussian sample. The population dynamics closes on the Gram summaries (Q, M) and gives an exact Riccati equation for the captured teacher subspace.

After a smooth truncation of the square loss, each Hessian block has the form $n^{-1}X^\top DX$, with diagonal weights depending only on finitely many projections. This places the Hessian in the framework of Ben Arous and gives deterministic matrices for the bulk, outliers, and eigenvector overlaps as functions of the summary statistics. The Dyson equations, Stieltjes edges, residues, and the three BBP transitions are then computed below. Table of notations at the endd.

1 Model and population summaries

Let p, k be fixed and $d \to \infty$. Let $\Theta = (\theta_1, \dots, \theta_k) \in \mathbb{R}^{d\times k}$ have orthonormal columns and let

$$\Lambda = \operatorname{diag}(\mu_1, \dots, \mu_k), \qquad \mu_i > 0.$$

For $W = (w_1, \dots, w_p) \in \mathbb{R}^{d\times p}$ set

$$f_W(x) = \frac{1}{p}\sum_{a=1}^{p}(w_a^\top x)^2, \qquad f_\star(x) = \sum_{i=1}^{k}\mu_i(\theta_i^\top x)^2.$$

Equivalentely,

$$f_W(x) = x^\top \Sigma_W x, \qquad \Sigma_W = \frac{1}{p}WW^\top, \qquad f_\star(x) = x^\top A_\star x, \qquad A_\star = \Theta\Lambda\Theta^\top.$$

For $x \sim N(0, I_d)$ define

$$E_W = \Sigma_W - A_\star, \qquad \tau_W = \operatorname{Tr} E_W.$$

The population square loss is

$$R(W) = \frac{1}{2}\mathbb{E}(f_W(x) - f_\star(x))^2 = \operatorname{Tr}(E_W^2) + \frac{1}{2}\tau_W^2. \tag{1.1}$$

The finite summaries are

$$Q = W^\top W, \qquad M = W^\top\Theta, \qquad \mathcal{G}(W) = \begin{pmatrix} Q & M \\ M^\top & I_k \end{pmatrix}. \tag{1.2}$$

When Q is invertible, put

$$C = M^\top Q^{-1}M, \qquad R_c = I_k - C. \tag{1.3}$$

then $C = \Theta^\top P_{\operatorname{span}(W)}\Theta$ is the part of the teacher subspace already contained in the student span.

Proposition 1.1 (Closed population dynamics)**.** *The population gradient is*

$$\nabla_W R(W) = \frac{2}{p}(2E_W + \tau_W I_d)W.$$

Hence the population gradient flow satisfies

$$\dot{W} = -\frac{2}{p}(2E_W + \tau_W I_d)W. \tag{1.4}$$

It induces the closed system

$$\dot{Q} = -\frac{4}{p}\left[2\left(\frac{1}{p}Q^2 - M\Lambda M^\top\right) + \tau Q\right], \tag{1.5}$$

$$\dot{M} = -\frac{2}{p}\left[2\left(\frac{1}{p}QM - M\Lambda\right) + \tau M\right], \tag{1.6}$$

where $\tau = p^{-1}\operatorname{Tr} Q - \operatorname{Tr}\Lambda$*. Morever,*

$$\dot{C} = \frac{4}{p}(\Lambda C + C\Lambda - 2C\Lambda C). \tag{1.7}$$

If $C(0)$ *is invertible, then*

$$C(t) = \left[I_k + e^{-4\Lambda t/p}(C(0)^{-1} - I_k)e^{-4\Lambda t/p}\right]^{-1}. \tag{1.8}$$

Proof. Differenciate the population loss. The equations for Q and M follow from $\dot{Q} = \dot{W}^\top W + W^\top \dot{W}$ and $\dot{M} = \dot{W}^\top \Theta$. Differentiate $C = M^\top Q^{-1} M$; using the closed equations cancels the common left multiplication terms and gives the Riccati equation for C. Setting $Y = C^{-1} - I_k$ gives $\dot{Y} = -(4/p)(\Lambda Y + Y\Lambda)$, hence the explicit solution. □

We can now work in a seperated one-dimensional window. This simplifies the computations and is verified experimentally with a power-law spectrum. In the general case one should instead keep the full Riccati matrix ODE and extract its eigenvalues at time t. In this scalar window the Riccati equation reduces to

$$\dot{r}_i = \frac{8\mu_i}{p} r_i(1 - r_i), \qquad r_i(t) = \frac{r_i(0)e^{8\mu_i t/p}}{1 - r_i(0) + r_i(0)e^{8\mu_i t/p}}. \tag{1.9}$$

If $r_i(0) \asymp d^{-1}$, then the time at which r_i reaches any fixed level in $(0, 1)$ is

$$T_i = \frac{p}{8\mu_i}\log d + O(1). \tag{1.10}$$

2 The Hessian in the effective spectral framework

Let $x_1, \ldots, x_n \sim N(0, I_d)$ be the independant sample, and let $X \in \mathbb{R}^{n\times d}$ have rows $x_\ell^\top$. Write

$$h = W^\top x \in \mathbb{R}^p, \qquad y = \Theta^\top x \in \mathbb{R}^k, \qquad s(h, y) = \frac{1}{p}\|h\|^2 - y^\top \Lambda y. \tag{2.1}$$

For the pure square loss $\ell(h, y) = s(h, y)^2/2$, the Hessian block with respect to $(w_1, \ldots, w_p)$ is

$$H_{ab}(W) = \frac{1}{n}X^\top D_{ab}(W)X, \tag{2.2}$$

where $D_{ab}(W)$ is diagonal and

$$D_{ab,\ell}(W) = \Phi_{ab}(W^\top x_\ell, \Theta^\top x_\ell), \tag{2.3}$$

$$\Phi_{ab}(h,y) = \frac{4}{p^2} h_a h_b + \frac{2}{p}\left(\frac{1}{p}\|h\|^2 - y^\top \Lambda y\right)\delta_{ab}. \tag{2.4}$$

Indeed, $\partial_{h_a} s = 2h_a/p$ and $\partial^2_{h_a h_b} s = 2\delta_{ab}/p$.

For a direct use of the effective spectral theorem 2.1 from Ben Arous et al, fix $R < \infty$ and choose a smooth loss $\ell_R \in C_b^m(\mathbb{R}^{p+k})$ such that

$$\ell_R(h,y) = \frac{1}{2} s(h,y)^2 \qquad \text{for } \|(h,y)\| \le R. \tag{2.5}$$

Set

$$\Phi_{ab}^{(R)}(h,y) = \partial^2_{h_a h_b} \ell_R(h,y), \qquad A_R(h,y) = (\Phi_{ab}^{(R)}(h,y))_{a,b=1}^p. \tag{2.6}$$

Then

$$H_{ab}^{(R)}(W) = \frac{1}{n} X^\top D_{ab}^{(R)}(W) X, \qquad D_{ab,\ell}^{(R)}(W) = \Phi_{ab}^{(R)}(W^\top x_\ell, \Theta^\top x_\ell). \tag{2.7}$$

Conditionally on W,

$$(h,y) \sim N(0, \mathcal{G}(W)). \tag{2.8}$$

Remark 2.1 (Outlier F matrix and empirical / deterministic disentangle)**.** *The effective spectral theorem of Ben Arous applies to bounded Hessian blocks. It gives deterministc limits for the bulk, outliers and eigenvector projections as functions of the finite Gram matrix $\mathcal{G}(W)$.*

This is the only imported RMT theorem, and the most difficult one. Without this theorem, one would have to prove the Dyson equation and the projected resolvent convergence directly. The proof removes the self-coupling between $D^{(R)}$ and X by splitting the data into the finite projection part and an independent Gaussian orthogonal noise.

The bulk is then an independent weighted Wishart type expansion, while the finite projection part produces the outlier equations. In the notation of that theorem, the finite matrix governing outliers is of the form

$$F_d(z) = R^\top D(I + S_d(z)D)^{-1} R. \tag{2.9}$$

The difficult step is the uniform convergence of $F_d(z)$ to its deterministic limit, including the control of possible poles $1 + S_d(z) D_\ell \simeq 0$.

To apply this theorem, we need to make the loss bounded or Huberized, as in Maillard [2], to have bounded derivatives.

Remark 2.2. *The paper is written in the regime where the time path of the weights is independent of the data: the path $W(t)$ is fixed before the Hessian sample is drawn. If the same data both train W and build the Hessian, one needs an additionnal leave-one-out or cavity argument.*

In order to apply their theorem for the Hessian we take

$$\alpha_d = \frac{n_d}{d} = (\log d)^2, \tag{2.10}$$

or $\alpha_d = (\log d)^2 \omega_d$ with $\omega_d \to \infty$ when an additional covarience margin is useful. This is not a training-time assumption: the trajectory $W_d(t)$ is generated first by the population ODE, and the Hessian sample is then drawn independently. The reason is the following standard Hanson–Wright estimate: if $x_\ell \sim N(0, I_d)$ and B is deterministic, then

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{\ell=1}^n x_\ell^\top B x_\ell - \operatorname{Tr} B\right| \ge t\right) \le 2\exp\left[-c \min\left(\frac{nt^2}{\|B\|_F^2}, \frac{nt}{\|B\|_{\mathrm{op}}}\right)\right].$$

For the full covariance operator, the natural scale is $O_{\mathbb{P}}(\alpha_d^{-1/2})$. However, the finite-dimensional covariance quantities entering the summary statistics and the Hessian weights only test functions of $(W^\top x, \Theta^\top x) \in \mathbb{R}^{p+k}$, with p, k fixed. Their effective sample size is therefore $n = d\alpha_d$.

3 Deterministic computations after applying Ben arous theorem

All expectations in this section are with respect to $(h, y) \sim N(0, \overline{\mathcal{G}}(t))$, where $\overline{\mathcal{G}}(t)$ is the deterministic summary path supplied by Section 1. For $S \in \mathrm{Sym}_p(\mathbb{C})$ define

$$T_S(A) = A\left(I_p + \alpha_d^{-1} S A\right)^{-1}. \tag{3.1}$$

From Ben arous paper Theorem 2.1, the matrix Stieltjes transform is the solution $S_R(z, t) \in \mathbb{C}^{p \times p}$ of

$$-S_R(z, t)^{-1} = z I_p - \mathbb{E}\left[T_{S_R(z,t)}(A_R(h, y))\right]. \tag{3.2}$$

The limiting density is

$$\rho_R(E, t) = \frac{1}{\pi} \lim_{\eta \downarrow 0} \Im \frac{1}{p} \operatorname{Tr} S_R(E + i\eta, t). \tag{3.3}$$

Let $x_-(t)$ and $x_+(t)$ be its left and right edges.

For a group of modes/eigenvectors $J \subset \{1, \ldots, k\}$ let P_J be coordinate restriction to J. Whith

$$C = M^\top Q^{-1} M, \qquad R_c = I_k - C, \tag{3.4}$$

define the residual and parallel teacher variables

$$\xi_J^-(h, y, t) = R_{c,J}^{-1/2} P_J(y - M^\top Q^{-1} h), \tag{3.5}$$

$$\xi_J^+(h, t) = C_J^{-1/2} P_J M^\top Q^{-1} h. \tag{3.6}$$

The finite matrices for left and right outliers are

$$K_{J,R}^-(z, t) = \mathbb{E}\left[\xi_J^- \xi_J^{-\top} \otimes T_{S_R(z,t)}(A_R(h, y))\right], \tag{3.7}$$

$$K_{J,R}^+(z, t) = \mathbb{E}\left[\xi_J^+ \xi_J^{+\top} \otimes T_{S_R(z,t)}(A_R(h, y))\right]. \tag{3.8}$$

The right side also contains rotations of the hidden representation, which are not informative for phase retrieval. Let Π_{amp} be the orthogonal projection removing the tangent space $\{W\Omega : \Omega^\top = -\Omega\}$. Set

$$K_{J,R,\mathrm{amp}}^+(z, t) = \Pi_{\mathrm{amp}} K_{J,R}^+(z, t) \Pi_{\mathrm{amp}}. \tag{3.9}$$

A left outlier is a root $\lambda < x_-(t)$ of

$$\det\left(\lambda I - K_{J,R}^-(\lambda, t)\right) = 0, \tag{3.10}$$

and a right outlier is a root $\lambda > x_+(t)$ of

$$\det\left(\lambda I - K_{J,R,\mathrm{amp}}^+(\lambda, t)\right) = 0. \tag{3.11}$$

The contact margins between outliers and edges are

$$\Delta_J^-(t) = x_-(t) - \lambda_{\min} K_{J,R}^-(x_-(t), t), \tag{3.12}$$

$$\Delta_J^+(t) = \lambda_{\max} K_{J,R,\mathrm{amp}}^+(x_+(t), t) - x_+(t). \tag{3.13}$$

If $\lambda_\star$ is a simple outlier and c is the normalised eigenvector of the corresponding finite matrix, then the empirical squared projection onto the associated teacher direction is given by the usual BBP residue formula; see also Eq. (58) of Cammarota [3]:

$$\Omega_J^\pm(t) = \frac{1}{c^\top\left(I - \partial_z K_{J,R}^\pm(\lambda_\star, t)\right) c} + \mathrm{err}_d, \qquad \mathrm{err}_d \to 0. \tag{3.14}$$

For the right side, K^+ is replaced by K_{amp}^+.

4 Expansion in large α without groups J

We then mimic the former calculations in our framework, but with the general Dyson equation and without groups J. We apply an asymptotic expension with respect to α.

First rewrite the notation of the paper in the scalar one-dimensional reduction. In the deterministic matrix formula, the Hessian weight $A_R(h, y)$ is tested on a fixed unit direction $u \in \mathbb{R}^p$ and is replaced by the scalar weight

$$\phi = \phi_u(h, y) = u^\top A_R(h, y)u.$$

Its law depends on time through the summary statistics $\overline{\mathcal{G}}(t)$, but we omit this dependence and write simply ϕ. For the untruncated square loss this reads

$$\phi_u(h, y) = \frac{4}{p^2}(u^\top h)^2 + \frac{2}{p}\left(\frac{1}{p}\|h\|^2 - y^\top \Lambda y\right).$$

The matrix Stieltjes variable S is replaced by a scalar g. Then

$$T_g(\phi) = \phi\left(1 + \alpha^{-1}g\phi\right)^{-1} = \frac{\alpha\phi}{\alpha + g\phi}.$$

For a fixed teacher mode i, the scalar teacher variable whose overlap is measured is one of the normalized coordinates

$$\xi_i^-(h, y, t) = \frac{y_i - e_i^\top M(t)^\top Q(t)^{-1}h}{\sqrt{1 - e_i^\top C(t)e_i}}, \qquad \xi_i^+(h, t) = \frac{e_i^\top M(t)^\top Q(t)^{-1}h}{\sqrt{e_i^\top C(t)e_i}},$$

where e_i is the i-th coordinate vector and $C(t) = M(t)^\top Q(t)^{-1}M(t)$. The left BBP branch uses $\xi = \xi_i^-$, while the right branch uses $\xi = \xi_i^+$. Define

$$z(g) = -\frac{1}{g} + \alpha\mathbb{E}\frac{\phi}{\alpha + g\phi}, \qquad K_\xi(z) = \alpha\mathbb{E}\frac{\xi^2\phi}{\alpha + g(z)\phi}. \tag{4.1}$$

A regular edge is a turning point of $z(g)$. If

$$m_\ell = \mathbb{E}[\phi^\ell], \qquad \theta_\ell = \mathbb{E}[\xi^2\phi^\ell], \tag{4.2}$$

Set $\epsilon = \alpha^{-1/2}$ and write $g = \epsilon^{-1}q + r + O(\epsilon)$. Since

$$\frac{1}{\alpha + g\phi} = \frac{1}{\alpha}\left(1 - \frac{g\phi}{\alpha} + \frac{g^2\phi^2}{\alpha^2} - \frac{g^3\phi^3}{\alpha^3} + \cdots\right), \tag{4.3}$$

the edge equation $\partial_g z = 0$ gives

$$q_\pm = \pm m_2^{-1/2}, \qquad r = \frac{m_3}{m_2^2}. \tag{4.4}$$

Substituion into $z(g)$ and $K_\xi(g)$ gives the edge and kernel expansions below.

then, for large α,

$$x_\pm = m_1 \pm 2\sqrt{m_2/\alpha} + \frac{m_3}{m_2\alpha} + O(\alpha^{-3/2}), \tag{4.5}$$

$$K^\pm(x_\pm) = \theta_1 \pm \frac{\theta_2}{\sqrt{m_2}\sqrt{\alpha}} + \frac{1}{\alpha}\left(\frac{\theta_3}{m_2} - \frac{m_3\theta_2}{m_2^2}\right) + O(\alpha^{-3/2}). \tag{4.6}$$

At a residual isotropic plateau J, the leading left BBP condition is

$$\mu_i^2 > \frac{B_J}{2\alpha}, \qquad B_J = S_{2,J} + (p_J + 8)\left(\frac{\rho_J}{p_J + 2}\right)^2, \tag{4.7}$$

where $p_J = |J|$, $\rho_J = \sum_{j\in J} \mu_j$ and $S_{2,J} = \sum_{j\in J} \mu_j^2$.

If $\mu_j = \mu_0 j^{-a}$ with $a > 1/2$, then $B_j \asymp \mu_0^2 j^{1-2a}$ and the visible front satisfies $j_{\mathrm{BBP}} \asymp \alpha$.

For an isolated mode obeying the scalar logistic equation, left reentry occurs when

$$\mu_i(1 - r_i(t^{\mathrm{in}}_{i,-})) = \frac{c_i^-}{\sqrt{\alpha}} + O(\alpha^{-1}), \tag{4.8}$$

so

$$t^{\mathrm{in}}_{i,-} = \frac{p}{8\mu_i}\log d + \frac{p}{16\mu_i}\log\alpha + O(1). \tag{4.9}$$

The isolated left eigenvector projection vanishes at contact:

$$\Omega_i^-(t) = A_i^-\left[\sqrt{\alpha}\mu_i(1 - r_i(t)) - c_i^-\right]_+ + \mathrm{err}_\alpha, \qquad \mathrm{err}_\alpha \to 0. \tag{4.10}$$

A right branch near $r_i = 1$ has

$$\Omega_i^{+,s}(t) = A_i^{+,s}\left[c_i^{+,s} - \sqrt{\alpha}\mu_i(1 - r_i(t))\right]_+ + \mathrm{err}_\alpha. \tag{4.11}$$

A Derivative in the scalar Stieltjes coordinate

Let ϕ and ξ be as in the scalar reduction above. Then

$$\partial_g z(g) = \frac{1}{g^2} - \alpha\mathbb{E}\frac{\phi^2}{(\alpha + g\phi)^2}. \tag{A.1}$$

By the chain rule,

$$\boxed{\partial_z K_\xi(z) = -\frac{\alpha\mathbb{E}\left[\dfrac{\xi^2\phi^2}{(\alpha+g\phi)^2}\right]}{g^{-2} - \alpha\mathbb{E}\left[\dfrac{\phi^2}{(\alpha+g\phi)^2}\right]}}. \tag{A.2}$$

If $\lambda_\star = K_\xi(\lambda_\star)$ is simple, then near $\lambda_\star$

$$(z - K_\xi(z))^{-1} = \frac{1}{(1 - \partial_z K_\xi(\lambda_\star))(z - \lambda_\star)} + O(1). \tag{A.3}$$

Comparing this pole with the spectral decomposition of the empirical projected resolvent gives

$$\Omega_\xi(\lambda_\star) = \frac{1}{1 - \partial_z K_\xi(\lambda_\star)} + \mathrm{err}_d, \qquad \mathrm{err}_d \to 0. \tag{A.4}$$

B Condition for the first left BBP transition

Let J be a residual set. Put

$$p_J = |J|, \qquad \rho_J = \sum_{j\in J}\mu_j, \qquad S_{2,J} = \sum_{j\in J}\mu_j^2, \qquad a_J = \frac{\rho_J}{p_J + 2}. \tag{B.1}$$

At the isotropic residual plateau the scalar Hessian weight in a fixed index direction is

$$\phi_J = \frac{2}{p_J}\left[a_J\left(3Z_1^2 + \sum_{a=2}^{p_J} Z_a^2\right) - \sum_{r\in J}\mu_r Y_r^2\right], \tag{B.2}$$

with independent standard Gaussians Z_a, Y_r. Wick expansion gives

$$m_1 = 0, \qquad m_2 = \frac{8}{p_J^2}\left[S_{2,J} + (p_J + 8)a_J^2\right]. \tag{B.3}$$

For the residual teacher variable $\xi = Y_i$,

$$\theta_1 = -\frac{4\mu_i}{p_J}, \qquad \theta_2 = \frac{8}{p_J^2}\left[S_{2,J} + (p_J+8)a_J^2 + 4\mu_i^2\right]. \tag{B.4}$$

At the critical scale $\mu_i = O(\alpha^{-1/2})$, the contact equation $K^-(x_-) = x_-$ yields

$$\mu_i^2 = \frac{B_J}{2\alpha} + O(\alpha^{-3/2}), \qquad B_J = S_{2,J} + (p_J+8)a_J^2. \tag{B.5}$$

C Reentry times at the left and the third right BBP transition

For a left residual branch of mode i,

$$\rho_i^-(t) = \mu_i(1 - r_i(t)). \tag{C.1}$$

Applying the threshold expansion gives

$$\mu_i(1 - r_i(t_{i,-}^{\text{in}})) = \frac{c_i^-}{\sqrt{\alpha}} + O(\alpha^{-1}). \tag{C.2}$$

If $\dot{r}_i = 2\omega_i r_i(1-r_i)$ and $r_i(0) \asymp d^{-1}$, then $1 - r_i(t) \asymp de^{-2\omega_i t}$ in the terminal layer and

$$t_{i,-}^{\text{in}} = \frac{\log d}{2\omega_i} + \frac{\log \alpha}{4\omega_i} + O(1). \tag{C.3}$$

For the quadratic model, $2\omega_i = 8\mu_i/p$.

For a right branch, the relevent residual is still $\mu_i(1-r_i)$ but the outlying side is the opposite one:

$$\Omega_i^{+,s}(t) = A_i^{+,s}\big[c_i^{+,s} - \sqrt{\alpha}\mu_i(1 - r_i(t))\big]_+ + \mathrm{err}_\alpha. \tag{C.4}$$

D Overlap at the BBP transition

The margin is computed from the explicit functions $x_\pm$ and $K^\pm$. This gives the expansion below rather than requiring it as a separate assumption. Let ρ be a scalar signal parameter, and let $\sigma \in \{-1, +1\}$ denote the chosen edge. From the large-α expansions,

$$x_\sigma = m_1(\rho) + \sigma\frac{2\sqrt{m_2(\rho)}}{\sqrt{\alpha}} + O(\alpha^{-1}), \qquad K^\sigma(x_\sigma) = \theta_1(\rho) + \sigma\frac{\theta_2(\rho)}{\sqrt{m_2(\rho)}\sqrt{\alpha}} + O(\alpha^{-1}).$$

Thus the signed contact margin has the form

$$\Delta(\rho;\alpha) = a\rho - \frac{b}{\sqrt{\alpha}} + O(\rho^2 + \alpha^{-1}), \qquad a = \partial_\rho(m_1 - \theta_1)(0) > 0. \tag{D.1}$$

Indeed, at zero signal contrast $m_1(0) = \theta_1(0)$. Hence $m_1(\rho) - \theta_1(\rho) = a\rho + O(\rho^2)$, and the edge correction is $b/\sqrt{\alpha} + O(\rho/\sqrt{\alpha} + \alpha^{-1})$.

In the BBP window $\rho = O(\alpha^{-1/2})$, this is the displayed margin expansion. The only non-degeneracy condition is $a > 0$. The threshold is

$$\rho_c(\alpha) = \frac{b}{a\sqrt{\alpha}} + O(\alpha^{-1}). \tag{D.2}$$

At a regular edge g_e,

$$z - z_e = \frac{1}{2}z_{gg}(g_e)(g-g_e)^2 + O((g-g_e)^3), \qquad K - K_e = K_g(g_e)(g-g_e) + O((g-g_e)^2). \tag{D.3}$$

The outlier equation gives

$$\Omega(\rho) = \frac{2m_2^{3/2}a}{\theta_2^2}\,[\sqrt{\alpha}\rho - b/a]_+ + \mathrm{err}_\alpha, \qquad \mathrm{err}_\alpha \to 0. \tag{D.4}$$

Then a newly born outlier has zero residue at the edge. For the left residual branch,

$$\rho_i^-(t) = \mu_i(1 - r_i(t)), \qquad \Omega_i^-(t) = A_i^-\left[\sqrt{\alpha}\rho_i^-(t) - c_i^-\right]_+ + o(1).$$

At the left reentry time,

$$\rho_i^-(t_{i,-}^{\mathrm{in}}) = \frac{c_i^-}{\sqrt{\alpha}} + O(\alpha^{-1}), \qquad \Omega_i^-(t_{i,-}^{\mathrm{in}}) \to 0. \tag{D.5}$$

At the same time $1 - r_i(t_{i,-}^{\mathrm{in}}) = c_i^-/(\mu_i\sqrt{\alpha}) + O(\alpha^{-1})$. Thus the teacher direction is already captured up to the BBP scale, while the residual eigenvector dissolves into the bulk.

For the right branch, with $s_i(t) = 1 - r_i(t)$,

$$\Omega_i^{+,s}(t) = A_i^{+,s}\left[c_i^{+,s} - \sqrt{\alpha}\mu_i s_i(t)\right]_+ + o(1).$$

Hence at the right detachment itself,

$$\mu_i(1 - r_i(t_{i,+}^s)) = \frac{c_i^{+,s}}{\sqrt{\alpha}} + O(\alpha^{-1}), \qquad \Omega_i^{+,s}(t_{i,+}^s) \to 0. \tag{D.6}$$

The right overlap becomes close to one only after the right outlier is macroscopicaly separated from the bulk. Indeed, if $\lambda_\star^+(t) - x_+(t) \geq \delta_0 > 0$, then $g(\lambda_\star^+, t) = O(1)$, and the derivative formula for $\partial_z K$ gives

$$\partial_z K^+(\lambda_\star^+, t) = O(\alpha^{-1}), \qquad \Omega^+(\lambda_\star^+, t) = 1 + O(\alpha^{-1}). \tag{D.7}$$

Therefore the interpretation is left reentry orthogonalizes hte weights, and right reentry aligns them. This is coherent with the algorithm provided in https://arxiv.org/pdf/2508.03688

Table of notation

Notation	Meaning
d	Ambiant dimension, sent to infinity.
p	Number of student directions or neurons.
k	Number of teacher directions or modes.
n	Number of independent samples used to build the empirical Hessian.
$\alpha_d = n_d/d$	Hessian sample ratio. In the main scaling, $\alpha_d = (\log d)^2$, and we write α in scalar expansions.
x, x_ℓ	Gaussian inputs in $\mathbb{R}^d$, with $x_\ell \sim N(0, I_d)$.
X	Data matrix whose rows are $x_\ell^\top$.
$W = (w_1, \dots, w_p)$	Student weight matrix in $\mathbb{R}^{d\times p}$.
$\Theta = (\theta_1, \dots, \theta_k)$	Orthonormal teacher directions in $\mathbb{R}^{d\times k}$.
$\Lambda = \operatorname{diag}(\mu_1, \dots, \mu_k)$	Teacher spectrum. The positive numbers μ_i are the teacher strengths.
$\Sigma_W = p^{-1}WW^\top$	Quadratic form learned by the student.
$A_\star = \Theta\Lambda\Theta^\top$	Teacher quadratic form.
$E_W = \Sigma_W - A_\star$	Error matrix between student and teacher quadratic forms.
$\tau_W = \operatorname{Tr} E_W$	Trace component of the population error.
$R(W)$	Population square loss.
$Q = W^\top W$	Student Gram matrix.
$M = W^\top\Theta$	Student-teacher overlap matrix.
$\mathcal{G}(W)$	Joint Gram matrix of (W, Θ).
$C = M^\top Q^{-1} M$	Captured part of the teacher subspace inside $\operatorname{span}(W)$.
$R_c = I_k - C$	Residual teacher covariance.
$r_i(t)$	One-dimensional captured overlap for mode i.
$h = W^\top x$	Student coordinates of a sample.
$y = \Theta^\top x$	Teacher coordinates of a sample.
$s(h, y)$	Scalar residual $p^{-1}\lVert h\rVert^2 - y^\top\Lambda y$.
ℓ_R	Smooth truncated loss used to apply the effective spectral theorem.
$\Phi^{(R)}_{ab}$	Second derivitive $\partial^2_{h_a h_b}\ell_R$.
$A_R(h, y)$	Matrix Hessian weight with entries $\Phi^{(R)}_{ab}(h, y)$.
$D^{(R)}_{ab}$	Diagonal sample weight matrix associated with Hessian block (a, b).
$H^{(R)}_{ab}(W)$	Empirical Hessian block $n^{-1}X^\top D^{(R)}_{ab}(W)X$.
$S_R(z, t)$	Matrix Stieltjes transform solving the deterministic Dyson equation.
$T_S(A)$	Resolvent weight $A(I_p + \alpha_d^{-1}SA)^{-1}$ from the Ben Arous et al. notation.
$\rho_R(E, t)$	Limiting spectral density of the Hessian.
$x_-(t), x_+(t)$	Left and right spectral edges.
J	Group of teacher modes considered together.
P_J	Coordinate restriction to the modes in J.
ξ_J^-	Residual teacher variable used for left outliers.
ξ_J^+	Parallel teacher variable used for right outliers.
$K^-_{J,R}(z, t)$	Finite matrix governing left outliers.
$K^+_{J,R}(z, t)$	Finite matrix governing right outliers.
Π_{amp}	Projection removing rotational tangent directions of the hidden representation.
$K^+_{J,R,\mathrm{amp}}$	Right outlier matrix after projection onto amplitude directions.
$\Delta^-_J(t), \Delta^+_J(t)$	Contact margins between outliers and spectral edges.
$\Omega^\pm_J(t)$	Squared empirical projection of an outlier eigenvector onto the associated teacher direction.
ϕ	Scalar Hessian weight in the one-dimensional reduction.

Table of notation

Notation	Meaning
g	Scalar Stieltjes coordinate replacing the matrix variable S.
$T_g(\phi)$	Scalar version of $T_S(A)$, equal to $\alpha\phi/(\alpha + g\phi)$.
ξ	Scalar teacher variable whose overlap is measured.
$z(g)$	Scalar Dyson parametrization of the spectral coordinate.
$K_\xi(z)$	Scalar finite outlier function.
$m_\ell = \mathbb{E}[\phi^\ell]$	Moments of the scalar Hessian weight.
$\theta_\ell = \mathbb{E}[\xi^2\phi^\ell]$	Weighted moments controlling teacher overlap.
$p_J, \rho_J, S_{2,J}, a_J$	Plateau quantities: $p_J = \|J\|$, $\rho_J = \sum_{j\in J}\mu_j$, $S_{2,J} = \sum_{j\in J}\mu_j^2$, and $a_J = \rho_J/(p_J+2)$.
B_J	Plateau variance threshold $S_{2,J} + (p_J+8)a_J^2$.

References

[1] G. Ben Arous, R. Gheissari, J. Huang and A. Jagannath, *Local geometry of high-dimensional mixture models: effective spectral theory and dynamical transitions*, arXiv:2502.15655, 2025.

[2] A. Maillard, T. Bonnaire and G. Biroli, *Topological Exploration of High-Dimensional Empirical Risk Landscapes: general approach, and applications to phase retrieval*, arXiv:2602.17779, 2026.

[3] B. L. Annesi, D. Bocchi and C. Cammarota, *Overparametrization bends the landscape: BBP transitions at initialization in simple Neural Networks*, arXiv:2510.18435, 2025.