SPECTRAL DYNAMICS BEYOND THE LEADING EIGENVECTOR

Anonymous authors
Paper under double-blind review

ABSTRACT

Naive mean-field variational inference on a dense Ising model gives a damped hyperbolic-tangent iteration that reuses one spiked Wigner matrix at every step. We hold that update fixed and ask what state it already requires, tracking not one leading direction but the residual spectrum in the noise eigenbasis. An exact transport identity moves that measure, and a matched-state theorem, certified at the gains we run, shows squared modal energy is insufficient, because the next signal update reads a signed spectral phase. Signed signal–residual atoms then recover exactly the first two input-field moments. For a matrix-independent initialization we prove a Gaussian signal law, a global modal regression whose gain is affine in the noise eigenvalue, and the first-update spectral law. We then cross the first matrix reuse: exact GOE conditioning and a pseudo-Lipschitz empirical law separate the new field into retained Gaussian correlation, response memory, and fresh Gaussian noise, giving the second signal update in closed form, and the response is resolved even on symmetric seeds whose overlap cannot see it. Beyond the first reuse, unsigned spectral transport is a diagnostic, not a closure: its per-step error settles on a floor that stays flat from $n = 200$ to 1200, and an exact error budget attributes that floor mainly to correlation between the nonlinear remainder and the residual, with the signed channel contributing for concentrated seeds. Across five independent matrices at dimension 1200, restoring the signed moments cuts the median free-running signal-update error from 0.112 to 0.012 and, in trajectory-median error, beats a control that drops only the signed atoms on every matrix and seed class. A frozen spectral preconditioner then tests how far the diagnosis transfers to a feasible operator. It does not overtake AMP, and controls attribute its gain to soft preservation of the leading outlier.

1 INTRODUCTION

Iterative estimators that reuse one random matrix are analyzed through a *state*: a functional of the iterate that predicts the next update. AMP, TAP and rotationally invariant variants add Onsager or divergence-free terms that cancel history-dependent conditional means, which leaves a low-dimensional Gaussian channel and often a scalar state evolution (Bayati & Montanari, 2011; Donoho et al., 2009; Javanmard & Montanari, 2013; Bolthausen, 2014). Without that cancellation, Gaussian conditioning is still available, but it retains old directions and response memory, and rigorous dynamical mean-field and trajectory analyses use correlation, response or history descriptions (Liang et al., 2023; Gerbelot et al., 2024; Jones & Pesenti, 2025; Dandi et al., 2026).

The map we study is a common instance of such an uncorrected iteration. For the dense Ising posterior $p(z \mid A) \propto \exp\{(\lambda/2)z^\top A^\circ z\}$ on $z \in \{-1,+1\}^n$ with $A^\circ = A - \operatorname{diag}(A)$, the naive mean-field variational family $q_v(z) = \prod_i (1 + v_i z_i)/2$ has evidence lower bound $(\lambda/2)v^\top A^\circ v + \sum_i h\{(1+v_i)/2\}$, with h the binary entropy, whose stationarity condition is $v = \tanh(\lambda A^\circ v)$ (Wainwright & Jordan, 2008; Fan et al., 2021). Equation 1 below is its damped synchronous solver, run without a memory term; the diagonal convention is immaterial over any fixed horizon since $\|\operatorname{diag}(A)\|_{\mathrm{op}} = O_P(\sqrt{\log n/n})$. The spiked-Wigner Bayes scaling is $\lambda = \beta$, and we keep λ free as an algorithmic gain. Fixed-point layers and message passing that reuse one operator without a memory term raise the same question in learning (Bai et al., 2019; Gu et al., 2020); we make no claim about those models.

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.

Table 1: What each description of the iterate is shown to do. "Every n" results are finite-instance identities or certificates; "$n \to \infty$" results are limit theorems; "measured" results are experiments.

Description of the iterate	Status	Scope	What it gives
$(\nu_t, \kappa_t, \eta_t)$ transport (Lemma 1)	every n	every t	the measure, given the response
(m, ν) is not a state (Thm. 1)	$n \geq 2$, any gain	one step	obstruction
(m, ν, ω, ρ_x) (Prop. 1)	every n	one step	both field moments
scalar law, modal gain (Thm. 2)	$n \to \infty$	$t = 0$	overlap; modal regression
first-update spectra (Prop. 2)	$n \to \infty$	$t = 0$	limits of ν_1, ω_1
conditional law (Thm. 3, Cor. 3)	$n \to \infty$	$t = 1$	coordinate law; m_2; c_1
unsigned transport (Section 6)	measured	$t \leq 10$	diagnostic with a floor
signed moments as predictor (Section 3)	measured	$t \leq 10$	signal update

The question. Instead of designing an update around a scalar state, we hold the update fixed and ask which statistic of the iterate it already requires. We use the spiked Wigner model as an instrument: its noise eigenbasis is explicit, its spectral threshold and optimal overlap are known (Baik et al., 2005; Benaych-Georges & Nadakuditi, 2011; Lelarge & Miolane, 2019), and AMP supplies a calibrated reference. The object we track is the empirical spectral measure of the residual in that eigenbasis.

Contributions. Table 1 places each result by status and scope.

1. **Which information the update reads (Section 3).** Theorem 1 constructs, on almost every instance and at any prescribed gain, two states with the same overlap and the same unsigned spectral atoms but different next overlaps; the missing quantity is a signed signal–residual phase. Proposition 1 recovers both input field moments from signed atoms. Restoring them cuts the free-running signal error from 0.112 to 0.012 and beats, in trajectory-median error on every matrix and seed class, the same predictor with the signed atoms set to zero.
2. **The first step and the first reuse, exactly (Sections 4–5).** For W-independent seeds, Proposition 2 gives explicit weak limits of the unsigned and signed spectra after one update, while Theorem 2 identifies the affine modal gain as a global regression with a strictly positive residual. Theorem 3 resolves the first reuse into retained correlation, response memory and fresh noise; a joint moment detects the response on symmetric blind seeds, whose overlap is uninformative, with a 49-fold separation at $n = 4000$ (Corollary 3).
3. **Where unsigned transport fails, and why (Sections 6–7).** An exact error budget on true states splits the later-time error of the unsigned transport into two neglected cross channels; the reaction observable, supplied as a field mean with no fitted coefficient, removes 77–91% of the squared error of the unsigned signal readout. A frozen spectral filter locates the operator-level effect in the band the modal gain predicts, and an outlier-preserving control explains the rank control's advantage.

Related work. *Low-dimensional states through correction*: AMP for dense matrices (Bayati & Montanari, 2011; Donoho et al., 2009; Javanmard & Montanari, 2013) and its extensions to non-separable and graph-based settings (Berthier et al., 2020; Gerbelot & Berthier, 2023), TAP constructions (Bolthausen, 2014; Opper & Winther, 2001; Opper et al., 2016; Çakmak & Opper, 2019) and OAMP/VAMP or rotationally invariant AMP (Takeuchi, 2020; Rangan et al., 2019; Fan, 2022; Venkataramanan et al., 2022; Zhong et al., 2024) cancel selected matrix–iterate correlations; we keep the uncorrected map and expose what exact conditioning retains. *Uncorrected dynamics*: generating-functional analyses predict two-time response kernels (Sompolinsky & Zippelius, 1981; Mimura, 2011; Bordelon & Pehlevan, 2026), rigorous pathwise results use correlation/response or memory systems (Liang et al., 2023; Gerbelot et al., 2024; Jones & Pesenti, 2025; Dandi et al., 2026), and general first-order methods reduce to AMP (Celentano et al., 2020; Montanari & Venkataramanan, 2021). Our contribution is not the existence of such descriptions but an eigenvalue-resolved observable, a finite-n obstruction, an exact first-reuse specialization, and a measured account of where unsigned transport fails. *Beyond the leading eigenvector*: the outlier transition and optimal overlap (Benaych-Georges & Nadakuditi, 2011; Baik et al., 2005; Lelarge & Miolane, 2019), spectrally initialized AMP (Mondelli & Venkataramanan, 2021; Zhong et al., 2024)

and non-backtracking spectra (Massoulié, 2014; Bordenave et al., 2015) choose a better starting direction; our measure is what remains after the signal direction is removed. Appendix H gives the longer discussion.

2 SETUP

Let $x \in \{-1,+1\}^n$ be uniform and, independently, W a GOE matrix with $W_{ij} \sim \mathcal{N}(0,1/n)$ for $i<j$ and $W_{ii} \sim \mathcal{N}(0,2/n)$, with eigenpairs (μ_k,u_k) and $W = U\Lambda U^\top$. For $\beta > 1$ let $A = (\beta/n)xx^\top + W$. The damped naive mean-field iteration is

$$v_{t+1} = (1-\alpha)v_t + \alpha\tanh(\lambda A v_t), \qquad t \ge 0, \tag{1}$$

with damping $\alpha \in (0,1)$ and gain $\lambda > 0$. Write the overlap $m_t = n^{-1}\langle v_t,\, x\rangle$, the residual $r_t = v_t - m_t x$, and its and the signal's noise-basis coefficients $a_{t,k} = \langle r_t,\, u_k\rangle$ and $b_k = \langle x,\, u_k\rangle$. The three measures we use are

$$\nu_t = \frac{1}{n}\sum_k a_{t,k}^2\delta_{\mu_k}, \qquad \omega_t = \frac{1}{n}\sum_k a_{t,k}b_k\delta_{\mu_k}, \qquad \rho_x = \frac{1}{n}\sum_k b_k^2\delta_{\mu_k}: \tag{2}$$

the unsigned residual measure (mass $q_t = n^{-1}\|r_t\|^2$), the signed signal–residual measure, and the quenched signal measure. Let $S_t = n^{-1}\langle\tanh(\lambda A v_t),\, x\rangle$ be the exact signal projection of the response, $y_t = \tanh(\lambda A v_t) - S_t x$, and $\kappa_t = n^{-1}\sum_k a_{t,k}y_{t,k}\delta_{\mu_k}$, $\eta_t = n^{-1}\sum_k y_{t,k}^2\delta_{\mu_k}$ with $y_{t,k} = \langle y_t,\, u_k\rangle$. For a Gaussian field $H \sim \mathcal{N}(\mu_H,\sigma_H^2)$ write $S^\star(\mu_H,\sigma_H^2) = \mathbb{E}\tanh(\lambda H)$ and $\chi(\mu_H,\sigma_H^2) = \mathbb{E}\,\mathrm{sech}^2(\lambda H)$; χ_t always denotes this Gaussian susceptibility and $\hat\chi_t = n^{-1}\sum_i \mathrm{sech}^2(\lambda(Av_t)_i)$ its empirical counterpart. Every modal gain below is $G_t(\mu) = (1-\alpha) + \alpha\lambda\chi_t\mu$.

Lemma 1 (Exact transport identity)**.** *For every n and t, $\nu_{t+1} = (1-\alpha)^2\nu_t + 2\alpha(1-\alpha)\kappa_t + \alpha^2\eta_t$.*

Proof. $\langle y_t,\, x\rangle = 0$, so projecting equation 1 onto $x^\perp$ gives $a_{t+1,k} = (1-\alpha)a_{t,k} + \alpha y_{t,k}$; square and sum against δ_{μ_k}. □

Protocol. Seeds are *blind* (independent of x, W, A), *aligned* (overlap 0.03, otherwise independent of W), or oracle probes concentrated on noise eigenvectors (*mid-bulk*, *top-bulk*) or built as matched pairs (*phase*); only blind and PCA seeds enter recovery tables. A predictor is *teacher forced* when it is restarted from the true state at every step and *free running* when it is iterated from $t=0$. The signal error is $|m_{t+1}^{\mathrm{pred}} - m_{t+1}|$; the spectral error is $\|g^{\mathrm{pred}} - g\|_2/\|g\|_2$ for Stieltjes transforms $g(z) = \int(\mu - z)^{-1}d\nu(\mu)$ on 41 points of $[-2.5,2.5] + 0.15i$; recovery is the absolute cosine $|v^\top x|/(\|v\|\|x\|)$. The independent unit is one (x,W) draw, and intervals are paired cluster bootstraps over draws. High-gain Gaussian expectations use a composite Gauss–Legendre rule verified against adaptive and 30-digit quadrature to within 10^{-13}, where a fixed 100-node Gauss–Hermite rule errs by up to 0.09 in χ at gain 12; the first-reuse suite at gain 2 keeps a 120-node Gauss–Hermite rule, accurate there to 10^{-9} (Appendix F).

3 WHICH STATISTIC THE UPDATE REQUIRES

Lemma 1 is exact in ν_t, so it is natural to ask whether (m_t,ν_t) is a state. It is not, and the missing information has a closed form.

Theorem 1 (The update depends on signed modal phase)**.** *(a) Let $\bar\mu_x = n^{-1}\sum_k \mu_k b_k^2$, $\theta_k = (\mu_k - \bar\mu_x)b_k$, and suppose $\theta \neq 0$. For $m \in \mathbb{R}$ and $s>0$ put $a = s\sqrt{n}\,\theta/\|\theta\|$ and $v_\pm = mx \pm Ua$. Then $v_\pm$ have the same overlap and atomwise identical ν, their signal-projected fields differ by 2Γ with $\Gamma = s\|\theta\|/\sqrt{n}$, and their next overlaps differ for every small $\lambda \neq 0$ and every λ outside a discrete instance-dependent set.*
(b) Fix $\lambda > 0$ and $|m| < 1$. Let $P = I - xx^\top/n$, $D_\star = \mathrm{diag}(\mathrm{sech}^2(\lambda A m x))$ and $d = PAD_\star x$. If $d \neq 0$, then for every $0 < s < \min\{(1-|m|)/\|d\|_\infty,\ \|d\|^2/(2\lambda\|Ad\|^2)\}$ the states $v_\pm = mx \pm sd$ lie in $(-1,1)^n$, have the same overlap and the same unsigned residual atoms in every orthonormal basis, and

$$m_{1,+} - m_{1,-} \ \ge\ \frac{2\alpha\lambda s}{n}\|d\|^2 - \frac{2\alpha\lambda^2 s^2}{n}\|Ad\|^2 \ >\ 0.$$

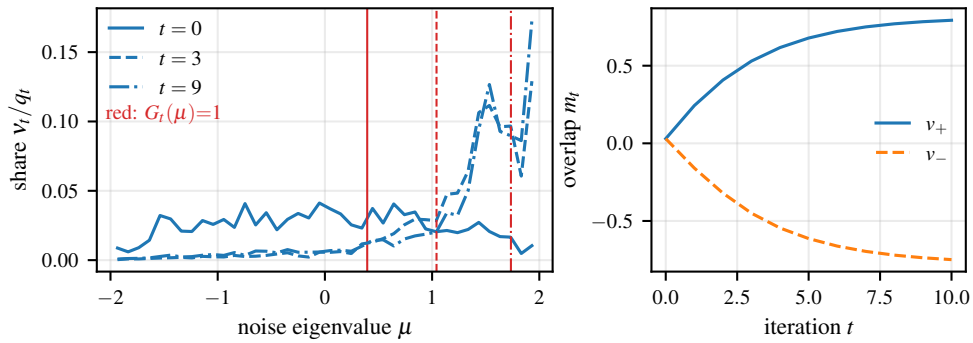


Figure 1: The tracked object at $n = 1200$, $\beta = 1.5$, $\lambda = 8$, $\alpha = 0.3$. *Left*: the residual measure ν_t normalized by its mass q_t at $t = 0, 3, 9$, with the unit-gain crossing $G_t(\mu) = 1$ of each time in red (matching line style). From a blind seed the crossing rises as the iterate saturates, and the residual moves into the upper bulk behind it. *Right*: the matched pair of Theorem 1, identical in overlap and in every ν atom, separating at once.

For $n \geq 2$, GOE W and fixed (m, λ, β), $d \neq 0$ almost surely.

Proof sketch. (a) $\sum_k (\mu_k - \bar{\mu}_x) b_k^2 = 0$ gives $b^\top a = 0$, so negating a keeps the overlap and every atom a_k^2/n, while $n^{-1} \sum_k \mu_k b_k a_k = \Gamma$; a cubic remainder bound and real analyticity in λ give the gain claims. (b) $d \perp x$ fixes the overlap and $\pm sd$ have equal squared coefficients. With $F(s)$ the next overlap of $mx + sd$, $\frac{d}{ds}\{F(s) - F(-s)\}|_{s=0} = 2\alpha\lambda n^{-1} x^\top D_\star A d = 2\alpha\lambda n^{-1} \|d\|^2$ by symmetry of A, and $|\tanh''| \leq 2$ bounds the rest. $\|PAD_\star x\|^2$ is a nonzero real-analytic function of the entries of W. Appendix A gives both proofs in full. □

The missing direction θ is the μ-weighted signal–residual correlation that the next field reads; part (a) holds instance-wise because GOE spectra are simple and signal projections nonzero almost surely, and part (b) certifies the separation at a prescribed gain. On the 48 phase-certificate instances (n from 200 to 1200, $m = 0.03$) the lower bound of part (b) is positive at gain 8 and at gain 12 on 48/48 and 48/48 instances, with minima 0.036 and 0.028 for pair amplitudes near 0.016 and 0.010 (Table 7). The statement is instance-wise and on the state domain: it asserts neither a gap uniform in n nor that the iteration reaches the pair.

Corollary 1 (The required state includes signed phase)**.** *Under Theorem 1(b), no deterministic function of (m, ν), even one allowed to depend on the realized (A, W, x), returns the next overlap for every state in $(-1, 1)^n$.*

Proposition 1 (Exact field moments from signed spectra)**.** *For every n and $h = Av$,*

$$n^{-1} x^\top h = \beta m + m \sum_k \mu_k \rho_{x,k} + \sum_k \mu_k \omega_k,$$
$$n^{-1} \|h\|^2 = \sum_k \mu_k^2 \nu_k + 2m \sum_k \mu_k (\mu_k + \beta) \omega_k + m^2 \sum_k (\mu_k + \beta)^2 \rho_{x,k}.$$

Proof. In the W-basis, $U^\top h = \mu \odot a + m(\mu + \beta) \odot b$. □

The theorem says what ν lacks; the proposition says that the signed atoms supply both moments of the gauged field $x \odot Av$. With the quenched constants of ρ_x, the two moments depend on the iterate only through four scalars, m, $\int \mu^2 d\nu$, $\int \mu \, d\omega$ and $\int \mu^2 d\omega$. They therefore determine the Gaussian readout of the next overlap, not the exact nonlinear update, and propagating them requires the measures: a four-scalar readout, not a four-scalar closure. At finite n, (ν, ω) determines a whenever $b_k \neq 0$, so the proposition is an exact re-encoding; its content is that the encoding is algebraic and local in μ.

Table 2 tests the signed state as a predictor. Keeping ρ_x alone barely helps; keeping ω cuts the pooled signal error 7.1-fold teacher forced and 9.7-fold free running, and the paired matrix-level bootstrap ratio of pooled medians against the quenched control is $[4.9, 9.1]$ and $[7.1, 11.5]$ (five draws describe sampling variability only coarsely). Free running, the signed predictor has the lower

Table 2: Signal-update error at $n = 1200$: median of $|m_{t+1}^{\text{pred}} - m_{t+1}|$ over five (x, W) draws and ten steps ($\beta = 1.5$, $\lambda = 12$, $\alpha = 0.3$). "Quenched" is the signed predictor with ω set to zero: it keeps the finite-n signal measure ρ_x; "wins" counts draws on which the signed predictor's median error over the ten steps is lower. The first two rows pool the five seed classes; the second gives the median spectral error.

	teacher forced				free running			
seed class	unsigned	quenched	signed	wins	unsigned	quenched	signed	wins
all classes	0.020	0.018	0.0028	5/5	0.112	0.109	0.0116	5/5
spectral error	0.203	0.198	0.112		0.491	0.503	0.261	
blind	0.0047	0.0045	0.0030	2/5	0.028	0.030	0.0063	5/5
aligned	0.018	0.018	0.0030	5/5	0.080	0.082	0.0093	5/5
mid-bulk	0.017	0.018	0.0015	5/5	0.050	0.045	0.0095	5/5
phase +	0.022	0.025	0.0027	5/5	0.599	0.596	0.024	5/5
phase −	0.029	0.029	0.0030	5/5	0.713	0.716	0.025	5/5

trajectory-median error on every draw in every seed class, and the lower error at 237/250 individual steps. Teacher forced, it wins on every draw except for blind seeds (2/5), where both errors are already small: from a delocalized blind seed the one-step signed correction is weak, and its value accumulates over the rollout. The constructed phase pairs are where the unsigned rollout fails outright, as Theorem 1 predicts, but they do not carry the result: without them the free-running ratio to the quenched control is 5.3 rather than 9.5. Both predictors start from oracle spectral quantities, and free running neither reads a later true state; this is a diagnostic, not an estimator. The spectral error improves far less, which Section 6 explains.

4 A RIGOROUS FIRST-STEP SPECTRAL LAW

Let v_0 be independent of W given x, with $m_n = n^{-1}x^\top v_0 \to m$, $\varrho_n = n^{-1}\|v_0\|^2 \to \varrho$, $q = \varrho - m^2 > 0$ and $\|v_0\|_\infty = o(\sqrt{n})$ in probability. For $Z \sim \mathcal{N}(0,1)$ let $F_n = \tanh\{\lambda(\beta m_n + \sqrt{\varrho_n}Z)\}$, $S_n^\star = \mathbb{E}F_n$, $\chi_n = \mathbb{E}\,\mathrm{sech}^2\{\lambda(\beta m_n + \sqrt{\varrho_n}Z)\}$ and $V_n = \mathrm{Var}(F_n) - \lambda^2\chi_n^2\varrho_n$, with limits χ and $V > 0$.

Theorem 2 (One-step modal regression)**.** *$S_0 - S_n^\star \to 0$ and $m_1 - \{(1-\alpha)m_n + \alpha S_n^\star\} \to 0$ in probability. With $a_k = \langle u_k, v_0 - m_n x\rangle$, $y_k = \langle u_k, y_0\rangle$, $\epsilon_k = y_k - \lambda\chi_n\mu_k(a_k + m_n b_k)$ and $\zeta_k = \lambda\chi_n m_n\mu_k b_k + \epsilon_k$, the modal update is exactly $a_{1,k} = \{(1-\alpha) + \alpha\lambda\chi_n\mu_k\}a_k + \alpha\zeta_k$, and in probability $n^{-1}\sum_k \epsilon_k^2 \to V$, $n^{-1}\sum_k a_k\epsilon_k$, $n^{-1}\sum_k \mu_k a_k\epsilon_k$, $n^{-1}\sum_k \mu_k b_k\epsilon_k \to 0$, and $n^{-1}\sum_k \zeta_k^2 \to \lambda^2\chi^2 m^2 + V$. Hence the least-squares coefficients of $a_{1,k}$ on $(a_k, \mu_k a_k)$ converge to $(1-\alpha, \alpha\lambda\chi)$: G_0 is an asymptotically exact global regression gain whose normalized modal residual has a strictly positive limit.*

Conditioning on (x, v_0) leaves W Gaussian; the exact GOE covariance, Stein's identity and Gaussian Poincaré give the limits, and strict Cauchy–Schwarz gives $V > 0$ (Appendix B). The gain is a global regression law, not a modewise closure; at the level of spectral measures, however, the first update closes.

Proposition 2 (First-update spectral law)**.** *Under the assumptions of Theorem 2, with $G_0(\mu) = (1-\alpha) + \alpha\lambda\chi\mu$, weakly in probability,*

$$\nu_1 \Rightarrow \{qG_0(\mu)^2 + \alpha^2\lambda^2\chi^2 m^2\mu^2 + \alpha^2 V\}\rho_{\mathrm{sc}}(\mu)\,d\mu, \qquad \omega_1 \Rightarrow \alpha\lambda\chi m\,\mu\,\rho_{\mathrm{sc}}(\mu)\,d\mu.$$

The remainder ϵ is spectrally white and uncorrelated with a and b against every bounded continuous test, by a Schur-complement argument on the part of W that the product Wv_0 leaves unobserved (Appendix B). The limit is the unsigned transport of Section 6 at $t = 0$, so that transport is asymptotically exact at the first update from a W-independent seed; a mid-bulk seed is built from W and lies outside the proposition. On 24 draws per cell, the fitted coefficients at $n = 1200$ are $0.699/0.698$ for a_k (target 0.700) and $0.742/0.744$ for $\mu_k a_k$ (blind/aligned), and all 8/8 log–log error slopes have bootstrap intervals containing $-1/2$. At later times, calibrating $G_t(\mu_k)a_{t,k}$ against the realized coefficients gives correlations 0.944–0.986 but slopes 1.496 and 1.347 for a mid-bulk seed at

Table 3: Response-along-seed moment $n^{-1}\sum_i s_i g_{1,i}$ at $n = 4000$, twelve draws per class: mean absolute error against the full law ($R_{10}\varrho$) and against the response-deleted or fresh-matrix law (theoretical limit 0).

seed class	observed	$R_{10}\varrho$	error, full law	error, no response
blind symmetric	0.2587	0.2586	0.0052	0.2587
aligned Gaussian	0.2318	0.2300	0.0065	0.2318
two-point	0.2314	0.2300	0.0045	0.2314

$t = 3, 9$ (NRMSE 0.455, 0.303), a factor stable over $n \in \{400, 800, 1200\}$: the gain orders the modes correctly but understates the scale for a concentrated seed.

5 THE FIRST MATRIX REUSE

Definition 1 (Reaction observable)**.** $c_t := n^{-1}x^\top W r_t = \int \mu \, d\omega_t$.

Corollary 2 (Reaction at the first reuse)**.** *Under Theorem 2, $c_0 \to 0$ and $c_1 \to \alpha\lambda\chi m$ in probability.*

Theorem 3 (Conditional law at the first matrix reuse)**.** *Let $D_n = \mathrm{diag}(x_n)$ and $s_n = D_n v_{0,n}$, and suppose that, given x_n, $v_{0,n}$ is independent of W_n and $n^{-1}\sum_i \delta_{s_{n,i}}$ converges in 4-Wasserstein distance in probability to a deterministic law π. For $V_0 \sim \pi$ independent of standard Gaussians Z_0, Z_1, let $m = \mathbb{E}V_0$, $\varrho = \mathbb{E}V_0^2 > 0$, $H_0 = \beta m + \sqrt{\varrho}Z_0$, $V_1 = (1-\alpha)V_0 + \alpha\tanh(\lambda H_0)$, $C_{10} = \mathbb{E}V_0V_1$, $C_{11} = \mathbb{E}V_1^2$, $R_{10} = \alpha\lambda\mathbb{E}\,\mathrm{sech}^2(\lambda H_0)$ and $Q_1 = C_{11} - C_{10}^2/\varrho$. With $\widetilde{W}_n = D_n W_n D_n$, $g_0 = \widetilde{W}_n s_n$, $u_1 = D_n v_1$ and $g_1 = \widetilde{W}_n u_1$, for every order-two pseudo-Lipschitz ψ,*

$$\frac{1}{n}\sum_i \psi(s_{n,i}, g_{0,i}, u_{1,i}, g_{1,i}) \xrightarrow{P} \mathbb{E}\psi(V_0, \sqrt{\varrho}Z_0, V_1, \mathcal{G}_1), \qquad \mathcal{G}_1 = \frac{C_{10}}{\varrho}\sqrt{\varrho}Z_0 + R_{10}V_0 + \sqrt{Q_1}Z_1,$$

and $m_{2,n} \to (1-\alpha)m_1 + \alpha\mathbb{E}\tanh\{\lambda(\beta m_1 + R_{10}V_0 + \sqrt{C_{11}}Z)\}$ with $m_1 = \mathbb{E}V_1$.

Proof sketch. The gauged matrix $\widetilde{W}_n$ is GOE and independent of s_n given x_n. Conditional on $g_0 = \widetilde{W}_n s$, exact Gaussian regression of the symmetric matrix on the constraint $\widetilde{W}s = g_0$ gives, with $L = \|s\|^2$,

$$\mathcal{L}(\widetilde{W} \mid g_0) = \mathcal{L}\Big(\frac{g_0 s^\top + s g_0^\top}{L} - \frac{s^\top g_0}{L^2} ss^\top + P_s \widetilde{W}' P_s\Big),$$

with $\widetilde{W}'$ an independent GOE and P_s the projector orthogonal to s. Applied to u_1 this is exactly $\widetilde{W}u_1 = k_n g_0 + e_n s + \xi_n$, where $k_n = s^\top u_1/L$, $e_n = (g_0^\top u_1 - k_n s^\top g_0)/L$ and $\xi_n = P_s\widetilde{W}'w$ for $w = u_1 - k_n s$; conditionally $\mathrm{Cov}(\xi_n) = Q_{1,n}P_s + ww^\top/n$ with normalized trace exactly $Q_{1,n} = \|w\|^2/n$. Stein's identity gives $e_n \to R_{10}$, the seed law gives $k_n \to C_{10}/\varrho$ and $Q_{1,n} \to Q_1$, and transfer lemmas carry the 4-Wasserstein hypothesis into pseudo-Lipschitz limits (Appendix C). □

The response term is independent of the two Gaussian components, so $\mathbb{E}\mathcal{G}_1^2 = C_{11} + R_{10}^2\varrho$: reuse inflates the second moment of the effective field by exactly $R_{10}^2\varrho$ over a fresh matrix, which is what an Onsager correction cancels. On aligned Gaussian and two-point seeds at $n = 4000$, the empirical-law plug-in predicts m_2 with error 0.0033 and 0.0051, against 0.0376 and 0.0400 without the response term; three pseudo-Lipschitz probes fixed in advance have errors 0.0023–0.0101, and an independent fresh GOE reverses the preference, locating the term in reuse.

Corollary 3 (The response is visible on symmetric seeds)**.** *Under Theorem 3, $n^{-1}v_0^\top W v_1 = n^{-1}\sum_i s_{n,i}g_{1,i} \to R_{10}\varrho$ and $n^{-1}\|g_1\|^2 \to C_{11} + R_{10}^2\varrho$ in probability. Deleting the response term, or drawing a fresh matrix for the second product, gives the limits 0 and C_{11}. For symmetric π the two laws predict the same overlaps, $m_1 = m_2 = 0$, but different joint moments.*

Corollary 3 answers what the theorem says about the blind seeds used in recovery experiments: the overlap cannot see the response, the joint law can. For the blind seed the separation grows with n, from 12-fold at $n = 500$ to 49-fold at $n = 4000$ (Table 3). This is evidence about a first-reuse joint statistic, not about recovery; symmetry breaking from an $O(n^{-1/2})$ overlap, later reuses and PCA seeds are outside the theorem.

6 BEYOND FIRST REUSE: LIMITATIONS OF UNSIGNED SPECTRAL TRANSPORT

To continue past the first reuse without evaluating the measures of Lemma 1, one can linearize the response about its first Hermite component. The following identity shows exactly what that drops.

Lemma 2 (Centered transport)**.** *For any state v and any scalar χ, let $L = \lambda\chi$, $\bar{h} = n^{-1}x^\top Av$, $\epsilon = U^\top\{\tanh(\lambda Av) - Sx - L(Av - \bar{h}x)\}$ with S the exact signal projection, $G(\mu) = 1-\alpha+\alpha L\mu$ and $H(\mu) = \alpha L\{m(\mu+\beta) - \bar{h}\}$. Then $a'_k = G_k a_k + H_k b_k + \alpha\epsilon_k$ exactly, and*

$$\nu' = G^2\nu + \underbrace{2GH\,\omega}_{\textit{signed}} + H^2\rho_x + \underbrace{2\alpha G\,\tau_{a\epsilon} + 2\alpha H\,\tau_{b\epsilon}}_{\textit{remainder cross}} + \alpha^2\eta_\epsilon,$$

with atoms $\tau_{a\epsilon,k} = a_k\epsilon_k/n$, $\tau_{b\epsilon,k} = b_k\epsilon_k/n$ and $\eta_{\epsilon,k} = \epsilon_k^2/n$.

The unsigned transport of the diagnostic in Appendix D keeps $G^2\nu$ and the signal injection, and replaces $\alpha^2\eta_\epsilon$ by white mass under cavity assumptions (A1)–(A4): a Gaussian gauged field independent of x_i with the memoryless mean βm_t, delocalized eigenvectors, no atom with a finite share of the mass, and a spectrally white remainder. It must also drop two cross measures that none of these removes: the remainder cross terms, the orthogonality part of (A4), and the signed term, which with $H = \alpha\lambda\chi m\mu$ vanishes under **(A5)**: $\int \varphi\,\mu G\,d\omega_t \to 0$ for the test functions used.

Remark 1 (Moments do not license the recursion). Vanishing of $\int \mu^j d\omega_t$ for $j = 0, 1, 2$ does not make the signed term negligible for the Stieltjes transform. With $P(\mu) = \mu^3 - 2\mu$, the residual $r = P(W)x - n^{-1}x^\top P(W)x\,x$ has $\omega \Rightarrow P\rho_{\mathrm{sc}}$, whose three moments vanish, while $\int(\mu - z)^{-1}\mu G(\mu)P(\mu)\,d\rho_{\mathrm{sc}}(\mu) = ((\sqrt{5}-1)/2)^4(0.24 - 0.70i) \neq 0$ at $z = i$ for $\alpha = 0.3$ and $\lambda\chi = 0.8$. (A5) must therefore be stated for the test functions the transform uses, not for two moments.

By Proposition 1 the memoryless mean assumes $c_t = \int \mu\,d\omega_t \approx 0$, which (A5) also forces (Appendix D) and Corollary 2 contradicts at $t = 1$ for $m \neq 0$. Over updates 5–14 the median $|c_t|$ stays near 0.140–0.152 for delocalized and 0.225–0.234 for mid-bulk seeds from $n = 500$ to 2000, rather than decaying like the $n^{-1/2}$ of Gaussian decoupling; for top-bulk seeds it falls from 0.054 to 0.027. The budget below measures the signed term directly.

The floor, and what it is made of. Under (A1)–(A5) the transport reads $\nu_{t+1}(d\mu) \approx G_t(\mu)^2\nu_t(d\mu) + \alpha^2[\lambda^2\chi_t^2 m_t^2\mu^2 + V_t^{\mathrm{nl}}]\rho_{\mathrm{sc}}(\mu)d\mu$, with V_t^{nl} the variance of the remainder once its first Hermite component is removed. Applying $\int(\mu - z)^{-1}$ and the Laurent identities $\int \mu\,d\nu/(\mu - z) = q + zg$ and $\int \mu^2 d\nu/(\mu - z) = M^{(1)} + zq + z^2 g$ gives, for $g_t(z) = \int(\mu - z)^{-1}d\nu_t$ and $M_t^{(1)} = \int \mu\,d\nu_t$,

$$\begin{aligned} g_{t+1} \approx{}& (1-\alpha)^2 g_t + 2\alpha(1-\alpha)\lambda\chi_t(q_t + zg_t) \\ &+ \alpha^2\lambda^2\chi_t^2\big(M_t^{(1)} + zq_t + z^2 g_t + m_t^2(z + z^2 g_{\mathrm{sc}})\big) + \alpha^2 V_t^{\mathrm{nl}} g_{\mathrm{sc}}, \end{aligned} \tag{3}$$

a closed recursion in the Stieltjes transform (Appendix D). Teacher forced, its first-step error falls from 0.084 to 0.034 between $n = 200$ and $n = 1200$ for blind seeds, as Proposition 2 requires, but the error pooled over the first ten updates stays at 0.09–0.10 for blind and aligned seeds and at 0.32–0.34 for mid-bulk seeds at every n from 200 to 1200: a floor that does not shrink over the dimensions tested. Lemma 2 makes the floor decomposable. On every teacher-forced true state, with the predictor's own χ, the difference between the realized measure and the unsigned prediction is exactly a signed term, the two remainder cross terms, an innovation-energy and an innovation-shape term, and a coefficient term (identity exact to within 10^{-15}). Table 4 reports each term's share of the error after projection onto the total, so cancellation is visible.

Under this linearization, for delocalized seeds the floor is almost entirely the mode-resolved correlation between the nonlinear remainder and the residual: restoring that term alone lowers the error from 0.088 to 0.007 (blind), while the signed term is a few percent and restoring it barely moves the error. The shares depend on the scalar that defines the remainder, but with the least-squares slope the remainder term still carries 88% and 90% (Appendix D). For the concentrated mid-bulk seed the signed term carries 40% of the error, and at its first update 90%; restoring it lowers the error from 0.327 to 0.222. So the signed channel is the decisive repair of the signal update (Table 2) and a partial repair of the spectrum; the spectral floor is mainly a failure of the remainder's orthogonality to the residual, which Proposition 2 proves against every test function at the first update and which fails after reuse.

Table 4: Exact budget of the unsigned transport's teacher-forced spectral error over the first ten updates ($\lambda = 12$, $n \in \{200, \dots, 1200\}$, 240 states per class), under the Gaussian linearization of Lemma 2: median relative error, median shares (%) of the signed and residual–remainder terms, and the error left after restoring only one of them (an oracle ablation).

		share (%)		error with term restored	
seed class	error	signed	remainder	signed	remainder
blind	0.088	1	104	0.086	0.007
aligned	0.092	4	99	0.087	0.011
mid-bulk	0.327	40	69	0.222	0.134
mid-bulk, first update	0.037	90	–	0.012	0.037

Table 5: The reaction as a field mean. Median $|m_{t+1} - m_{t+1}^{\mathrm{pred}}|$ of the Gaussian readout with unsigned moments (U), exact mean (M), exact variance (V) or both (MV); the reduction in squared error relative to U; and the draws on which M beats U.

seed class	U	M	V	MV	reduction by M	by V	draws improved
delocalized	0.0194	0.0060	0.0197	0.0053	86%	-7%	24/24
mid-bulk	0.0170	0.0046	0.0200	0.0019	91%	-13%	24/24
top-bulk	0.0057	0.0027	0.0057	0.0027	77%	-1%	20/24

The reaction as a field mean. $c_t = \int \mu \, d\omega_t$ is the mean of the gauged residual field, so its effect on the signal update can be isolated without any fit. On the true states of the reaction suite ($\lambda = 12$, $n \in \{500, 1000, 2000\}$, 15 steps, 24 draws per class), we compare Gaussian readouts $(1-\alpha)m_t + \alpha S^\star(\mu_H, \sigma_H^2)$ whose mean μ_H is exact, $\beta m_t + m_t d_1 + c_t$, or unsigned, $\beta m_t + m_t d_1$, and whose variance is exact, $e_2 + 2m_t c_t' + m_t^2 d_2 - (m_t d_1 + c_t)^2$, or unsigned, $e_2 + m_t^2(d_2 - d_1^2)$, with $d_j = \int \mu^j d\rho_x$, $c_t' = \int \mu^2 d\omega_t$ and $e_2 = \int \mu^2 d\nu_t$ (Proposition 1). None of the four readouts has a fitted coefficient.

Correcting the mean removes most of the error in every class; correcting the variance alone removes none, though it adds to the mean correction (compare M and MV in Table 5; Figure 2, left). The reaction therefore predicts the signal residual quantitatively, through the field mean, and the linear correlation understates it because the effect passes through a saturated $\tanh$. An earlier baseline that omitted m_t^2 from the field variance made the mid-bulk correlation negative (-0.59); that sign was an artifact of the baseline (Appendix E).

7 An operator-level test of the diagnosis

Modes with $|G_t(\mu)| > 1$ amplify. At the audit's gains $\lambda = \beta \le 1.5$, $G_t \ge 1 - \alpha(1 + 2\lambda) \ge -0.2$ on $[-2, 2]$, so only the branch $G_t > 1$, $\mu > 1/(\lambda\chi_t)$, occurs in the bulk. This suggests attenuating that band, a hypothesis rather than a consequence, since the diagnostic lives in the W-basis while a feasible operator sees only A. The spectral-filtered iteration is $v_{t+1} = v_t + \alpha P(A)(\tanh(\lambda A v_t) - v_t)$ with $P(A) = U_A \operatorname{diag}(P(\mu_k(A)))U_A^\top$, $P = 0.1$ on a band and 1 elsewhere. Under the same scalar-susceptibility surrogate, a filtered unstable mode has gain between 1 and its unfiltered value; this is a local statement, not a convergence guarantee.

What transfers. Table 6 separates location from shape. Of the other three bands, $[1, 2]$ reproduces $[0, 2]$ while $[0, 1]$ and $[-2, 0]$ match naive mean field. At the audit's own parameters the median crossing $1/(\lambda\chi_t)$ starts at 0.78–1.00 and is 1.06–1.09 at $t = 5$ and 1.4–1.6 at $t = 10$, so at these reported times the surrogate's amplifying bulk modes lie in $[1, 2]$ and not in $[0, 1]$. The rank control beats the band at $\beta = 1.1$ and 1.3 for a separate reason: A's leading eigenvalue falls inside $[0, 2]$ on 6/20 and 1/20 draws, and the band then attenuates the signal direction itself. Keeping the leading eigenvector inside the same band recovers the rank control on those draws (0.024 $\to$ 0.235 vs. 0.238 at $\beta = 1.1$; 0.125 $\to$ 0.456 vs. 0.464 at $\beta = 1.3$). What transfers is the location of the amplifying band and the protection of the leading direction; the band's detailed shape is unnecessary.

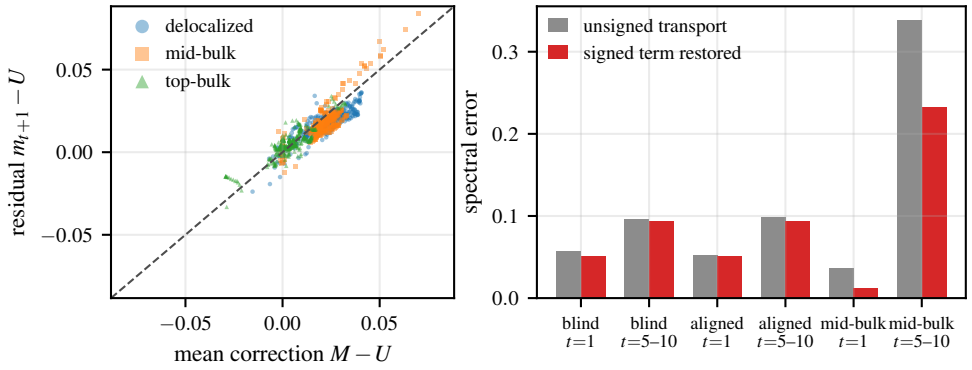


Figure 2: Beyond the first reuse. *Left*: for every true state of the reaction suite, the residual of the unsigned readout against the parameter-free correction from the exact field mean; the diagonal is exact prediction. *Right*: median spectral error of the unsigned transport (grey) and with only the signed term restored (red), by seed class, at the first step and over steps 5 to 10.

Table 6: Mechanism audit: mean final absolute cosine from blind seeds ($n = 800$, $T = 30$, $\lambda = \beta$, 20 draws). Bands are intervals of the A-spectrum attenuated by 0.1; "$[0, 2]$, top kept" is the same band with the leading eigenvector of A left untouched; "all but top" attenuates every mode except the leading one.

β	naive	$[0, 1]$	$[-2, 0]$	$[1, 2]$	$[0, 2]$	$[0, 2]$, top kept	all but top	top only	PCA
1.1	0.151	0.154	0.152	0.297	0.301	0.365	0.391	0.073	0.404
1.3	0.190	0.199	0.188	0.584	0.587	0.604	0.614	0.053	0.615
1.5	0.328	0.341	0.326	0.763	0.764	0.764	0.765	0.051	0.744

Recovery. The filter is a mechanism probe, not a recovery method; Table 13 (Appendix G) gives every cell. At $n = 2000$ and $\beta = 1.1, 1.3, 1.5$ it lifts blind naive MF from 0.118, 0.199, 0.394 to 0.281, 0.600, 0.752, against 0.259, 0.695, 0.829 for AMP; naive MF warm-started from PCA reaches more in every cell (0.375, 0.685, 0.825 here), and PCA-initialized AMP has a higher mean in five cells and agrees to three decimals in the sixth. Paired differences between the filter and AMP are -0.095 $[-0.121, -0.077]$ and -0.077 $[-0.084, -0.071]$ at $\beta = 1.3, 1.5$ for $n = 2000$. State evolution at the gain AMP froze gives fixed points 0.703 and 0.832 at $\beta = 1.3, 1.5$, which AMP reaches. At $\beta = 1.1$ the selected gain at $n = 2000$ is subcritical ($\beta\lambda < 1$), so state evolution has only the uninformative fixed point and the measured 0.259 is a finite-horizon value. At $n = 1200$ the $[0, 2]$ eigenspaces of W and A have normalized overlap $\|U_W^\top U_A\|_F^2 / \min(\text{ranks}) = 0.999$ and their first filtered actions agree to cosine 0.9995, yet the outlier separates the two filters over twenty steps.

8 Discussion

Holding a deployed update fixed and asking which statistic it requires gives checkable answers. At one time, unsigned modal energy provably misses a signed phase at any prescribed gain, and signed atoms give both field moments; used as a free-running predictor they beat the same predictor without them in trajectory-median error on every draw and seed class. At the first reuse the conditional law separates retained correlation, response memory and fresh noise, and the response is resolved even where the overlap is blind to it. Unsigned spectral transport is asymptotically exact at the first update from a W-independent seed, but after reuse it has a floor, flat over the dimensions tested, whose exact budget points mainly at remainder–residual correlation for delocalized seeds and jointly at the signed channel for concentrated ones. Theorem 3 crosses exactly one reuse; general t is the rank-t version of the same conditioning, with a $t \times t$ covariance and a response matrix, and we do not carry it out. It needs a W-independent seed, so PCA seeds are outside it; the later-time results are measurements; and recovery covers one prior, one ensemble and three spikes.

REPRODUCIBILITY STATEMENT

Appendices A–D contain complete proofs of every formal result with all assumptions stated. Appendix F verifies the Gaussian integrals, Appendix G gives the protocol and the selection grids, Supplement Table S8 every frozen hyperparameter, and Supplement S2 the extended tables. The supplementary code runs each experiment from a committed configuration with hash-derived seeds, archives per-trial data with SHA-256 fingerprints of the generating sources, and includes one command that reruns every suite, regenerates every number, table and figure of this paper from the archives, and checks them.

AI USE STATEMENT

In this work, we used generative AI tools for the following tasks with required disclosure: to help develop parts of the theoretical framework, formulate mathematical claims, provide critical ingredients for proving them and assist in writing the proofs, propose or refine hypotheses, design and give feedback on the methodology and experiments, implement the methods, and interpret results. We have not used generative AI tools to generate synthetic data sets (every instance is a random matrix drawn by our simulation code) or to assist with translation, and cleaning or reformatting data sets and qualitative or thematic data analysis are not applicable to this work. Additionally, we used generative AI tools to create and edit software code and figures, suggest experimental parameters, create the reproducibility artifacts, draft and edit parts of the paper and suggest its structure, and search, summarize and verify the literature and format the references. We have reviewed all AI-assisted work: every AI-proposed statement and proof was checked by the authors and by independent AI audits, and the code is covered by unit tests and by an audit that regenerates every reported number from the archived runs. We take responsibility for the final content of this work, including text, claims and artifacts produced with the aid of generative AI.

REFERENCES

Shaojie Bai, J. Zico Kolter, and Vladlen Koltun. Deep equilibrium models. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

Jinho Baik, Gérard Ben Arous, and Sandrine Péché. Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *The Annals of Probability*, 33(5):1643–1697, 2005.

Mohsen Bayati and Andrea Montanari. The dynamics of message passing on dense graphs, with applications to compressed sensing. *IEEE Transactions on Information Theory*, 57(2):764–785, 2011.

Florent Benaych-Georges and Raj Rao Nadakuditi. The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. *Advances in Mathematics*, 227(1):494–521, 2011.

Raphael Berthier, Andrea Montanari, and Phan-Minh Nguyen. State evolution for approximate message passing with non-separable functions. *Information and Inference: A Journal of the IMA*, 9(1):33–79, 2020.

David M. Blei, Alp Kucukelbir, and Jon D. McAuliffe. Variational inference: A review for statisticians. *Journal of the American Statistical Association*, 112(518):859–877, 2017.

Erwin Bolthausen. An iterative construction of solutions of the TAP equations for the Sherrington–Kirkpatrick model. *Communications in Mathematical Physics*, 325(1):333–366, 2014.

Blake Bordelon and Cengiz Pehlevan. Disordered dynamics in high dimensions: Connections to random matrices and machine learning. *arXiv preprint arXiv:2601.01010*, 2026.

Charles Bordenave, Marc Lelarge, and Laurent Massoulié. Non-backtracking spectrum of random graphs: community detection and non-regular Ramanujan graphs. In *2015 IEEE 56th Annual Symposium on Foundations of Computer Science*, pp. 1347–1357, 2015.

Burak Çakmak and Manfred Opper. Memory-free dynamics for the Thouless–Anderson–Palmer equations of Ising models with arbitrary rotation-invariant ensembles of random coupling matrices. *Physical Review E*, 99(6):062140, 2019. doi: 10.1103/PhysRevE.99.062140.

Michael Celentano, Andrea Montanari, and Yuchen Wu. The estimation error of general first order methods. In *Proceedings of the 33rd Conference on Learning Theory*, volume 125 of *Proceedings of Machine Learning Research*, pp. 1078–1141, 2020.

Yatin Dandi, David Gamarnik, Francisco Pernice, and Lenka Zdeborová. Rigorous asymptotics for first-order algorithms through the dynamical cavity method. In *Proceedings of the Thirty-Ninth Conference on Learning Theory*, volume 336 of *Proceedings of Machine Learning Research*, pp. 1611–1646. PMLR, 2026.

Aurelien Decelle, Florent Krzakala, Cristopher Moore, and Lenka Zdeborová. Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Physical Review E*, 84(6):066106, 2011.

Nikita Doikov and Anton Rodomanov. Polynomial preconditioning for gradient methods. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 8162–8187, 2023.

David L Donoho, Arian Maleki, and Andrea Montanari. Message-passing algorithms for compressed sensing. *Proceedings of the National Academy of Sciences*, 106(45):18914–18919, 2009.

Zhou Fan. Approximate message passing algorithms for rotationally invariant matrices. *The Annals of Statistics*, 50(1):197–224, 2022.

Zhou Fan, Song Mei, and Andrea Montanari. TAP free energy, spin glasses and variational inference. *The Annals of Probability*, 49(1):1–45, 2021.

Cédric Gerbelot and Raphaël Berthier. Graph-based approximate message passing iterations. *Information and Inference: A Journal of the IMA*, 12(4):2562–2628, 2023.

Cédric Gerbelot, Emanuele Troiani, Francesca Mignacco, Florent Krzakala, and Lenka Zdeborová. Rigorous dynamical mean-field theory for stochastic gradient descent methods. *SIAM Journal on Mathematics of Data Science*, 6(2):400–427, 2024. doi: 10.1137/23M1594388.

Fangda Gu, Heng Chang, Wenwu Zhu, Somayeh Sojoudi, and Laurent El Ghaoui. Implicit graph neural networks. In *Advances in Neural Information Processing Systems*, volume 33, 2020.

Adel Javanmard and Andrea Montanari. State evolution for general approximate message passing algorithms, with applications to spatial coupling. *Information and Inference: A Journal of the IMA*, 2(2):115–144, 2013.

Chris Jones and Lucas Pesenti. Fourier analysis of iterative algorithms. In *52nd International Colloquium on Automata, Languages, and Programming*, volume 334 of *Leibniz International Proceedings in Informatics*, pp. 102:1–102:21. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2025. doi: 10.4230/LIPIcs.ICALP.2025.102.

Marc Lelarge and Léo Miolane. Fundamental limits of symmetric low-rank matrix estimation. *Probability Theory and Related Fields*, 173:859–929, 2019.

Tengyuan Liang, Subhabrata Sen, and Pragya Sur. High-dimensional asymptotics of Langevin dynamics in spiked matrix models. *Information and Inference: A Journal of the IMA*, 12(4):2720–2752, 2023. doi: 10.1093/imaiai/iaad042.

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, 2020.

Laurent Massoulié. Community detection thresholds and the weak Ramanujan property. *Proceedings of the 46th Annual ACM Symposium on Theory of Computing*, pp. 694–703, 2014.

Kazushi Mimura. Generating functional analysis of iterative algorithms for compressed sensing. *arXiv preprint arXiv:1106.0086*, 2011.

Marco Mondelli and Ramji Venkataramanan. Approximate message passing with spectral initialization for generalized linear models. In *Proceedings of the 24th International Conference on Artificial Intelligence and Statistics*, volume 130 of *Proceedings of Machine Learning Research*, pp. 397–405, 2021.

Andrea Montanari and Ramji Venkataramanan. Estimation of low-rank matrices via approximate message passing. *The Annals of Statistics*, 49(1):321–345, 2021.

Manfred Opper and Ole Winther. Adaptive and self-averaging Thouless–Anderson–Palmer mean-field theory for probabilistic modeling. *Physical Review E*, 64(5):056131, 2001.

Manfred Opper, Burak Çakmak, and Ole Winther. A theory of solving TAP equations for Ising models with general invariant random matrices. *Journal of Physics A: Mathematical and Theoretical*, 49(11):114002, 2016.

Fabian Pedregosa and Damien Scieur. Acceleration through spectral density estimation. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 7553–7562, 2020.

Sundeep Rangan, Philip Schniter, and Alyson K Fletcher. Vector approximate message passing. *IEEE Transactions on Information Theory*, 65(10):6664–6684, 2019.

Haim Sompolinsky and Annette Zippelius. Dynamic theory of the spin-glass phase. *Physical Review Letters*, 47(5):359–362, 1981.

Keigo Takeuchi. Rigorous dynamics of expectation-propagation-based signal recovery from unitarily invariant measurements. *IEEE Transactions on Information Theory*, 66(1):368–386, 2020.

Ramji Venkataramanan, Kevin Kögler, and Marco Mondelli. Estimation in rotationally invariant generalized linear models via approximate message passing. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 22120–22144, 2022.

Martin J. Wainwright and Michael I. Jordan. Graphical models, exponential families, and variational inference. *Foundations and Trends in Machine Learning*, 1(1–2):1–305, 2008.

Xinyi Zhong, Tianhao Wang, and Zhou Fan. Approximate message passing for orthogonally invariant ensembles: Multivariate non-linearities and spectral initialization. *Information and Inference: A Journal of the IMA*, 13(3):iaae024, 2024. doi: 10.1093/imaiai/iaae024.

Yunkai Zhou, Yousef Saad, Murilo L. Tiago, and James R. Chelikowsky. Self-consistent-field calculations using chebyshev-filtered subspace iteration. *Journal of Computational Physics*, 219(1): 172–184, 2006. doi: 10.1016/j.jcp.2006.03.017.

Conventions. Unless a statement says otherwise, every limit is in probability as $n \to \infty$ with (α, β, λ) fixed. Lemma 1, Theorem 1, Proposition 1 and Lemma 2 hold for every realization and every finite n. The one almost-sure argument is the subsequence step that removes the conditioning on a random initialization at the end of Appendix C. We write $m_t = n^{-1}\langle v_t,\, x\rangle$, $r_t = v_t - m_t x$, $a_{t,k} = \langle r_t,\, u_k\rangle$, $b_k = \langle u_k,\, x\rangle$, $q_t = n^{-1}\left\|r_t\right\|^2$ and $\varrho_t = n^{-1}\left\|v_t\right\|^2$, with $W = U\Lambda U^\top$.

A Proofs for Section 3

A.1 Lemma 1

Write $h_t = \tanh(\lambda A v_t)$, so $y_t = h_t - S_t x$ with $S_t = n^{-1}\langle h_t,\, x\rangle$. Since $\left\|x\right\|^2 = n$, $\langle y_t,\, x\rangle = nS_t - nS_t = 0$. Subtracting $m_{t+1}x$ from Eq. (1) and using $m_{t+1} = (1-\alpha)m_t + \alpha S_t$ gives $r_{t+1} = (1-\alpha)r_t + \alpha y_t$, hence $a_{t+1,k} = (1-\alpha)a_{t,k} + \alpha y_{t,k}$ for every k. Squaring, dividing by n and summing against δ_{μ_k} proves the identity; no probabilistic assumption is used.

A.2 THEOREM 1(A)

Step 1: $Ua \perp x$. Since $n^{-1}\sum_k b_k^2 = 1$, $\bar{\mu}_x$ is a weighted average of the μ_k and $\sum_k(\mu_k - \bar{\mu}_x)b_k^2 = 0$, that is $b^\top\theta = 0$ and $b^\top a = 0$. As $x^\top(Ua) = b^\top a$, both states have overlap m, their residuals are exactly $\pm Ua$ with modal coefficients $\pm a_k$, and every atom a_k^2/n of ν coincides.

Step 2: the field gap. With $x^\top v_\pm = nm$, $n^{-1}x^\top A v_\pm = \beta m + n^{-1}b^\top\Lambda(mb \pm a)$. The part $m\bar{\mu}_x$ is common, and

$$\frac{1}{n}\sum_k \mu_k b_k a_k = \frac{s}{\sqrt{n}\,\|\theta\|}\sum_k \mu_k(\mu_k - \bar{\mu}_x)b_k^2 = \frac{s\,\|\theta\|}{\sqrt{n}} = \Gamma,$$

using $\sum_k(\mu_k - \bar{\mu}_x)^2 b_k^2 = \sum_k \mu_k(\mu_k - \bar{\mu}_x)b_k^2$, a consequence of Step 1.

Step 3: through the nonlinearity. With $h_\pm = Av_\pm$, $m_{1,\pm} = (1-\alpha)m + \alpha n^{-1}\sum_i x_i \tanh(\lambda h_{\pm,i})$, and $|\tanh z - z| \le |z|^3/3$ gives $|n^{-1}\sum_i x_i \tanh(\lambda h_{\pm,i}) - \lambda n^{-1}x^\top h_\pm| \le |\lambda|^3\,\|h_\pm\|_3^3/(3n)$. Subtracting the two expansions gives $m_{1,+} - m_{1,-} = 2\alpha\lambda\Gamma + R_\lambda$ with the stated bound, so the gap is nonzero for $0 < |\lambda| < \lambda_0$.

Step 4: the pair lies in the cube. $\|Ua\|_\infty = s\sqrt{n}\,\|U\theta\|_\infty / \|\theta\|$ is linear in s, so $s < (1 - |m|)\,\|\theta\|/(\sqrt{n}\,\|U\theta\|_\infty)$ gives $v_\pm \in (-1,1)^n$.

Genericity. For GOE W and $n \ge 2$ the eigenvalues are almost surely distinct and every $b_k \ne 0$; $\theta = 0$ would force $\mu_k = \bar{\mu}_x$ for all k. Hence $\theta \ne 0$ almost surely.

Step 5: gains outside a discrete set. $D(\lambda) = m_{1,+}(\lambda) - m_{1,-}(\lambda)$ is a finite sum of real-analytic functions of λ with $D'(0) = \alpha n^{-1}x^\top(h_+ - h_-) = 2\alpha\Gamma \ne 0$, so its zero set is discrete. The set need not be finite, and this step certifies no prescribed gain; part (b) does.

A.3 THEOREM 1(B)

Let $P = I - xx^\top/n$, $v_\star = mx$, $D_\star = \mathrm{diag}(\mathrm{sech}^2(\lambda A v_\star))$ and $d = PAD_\star x$. Since $x^\top d = 0$, both $v_\pm = mx \pm sd$ have overlap m and residuals $\pm sd$, whose coefficients in any orthonormal basis differ only in sign, so every unsigned atom coincides. Let $T(v) = (1-\alpha)v + \alpha\tanh(\lambda A v)$ and $F(s) = n^{-1}x^\top T(mx + sd)$. Then $F'(0) = \alpha\lambda n^{-1}x^\top D_\star A d$, and since A is symmetric and $x^\top d = 0$,

$$x^\top D_\star A d = (AD_\star x)^\top d = (PAD_\star x)^\top d = \|d\|^2,$$

so that $\frac{d}{ds}\{F(s) - F(-s)\}|_{s=0} = 2\alpha\lambda n^{-1}\|d\|^2$. Coordinatewise, with $\varphi_i(s) = \tanh(\lambda(Av_\star)_i + s\lambda(Ad)_i)$ and $|\tanh''| \le 2$, Taylor's theorem gives $|\varphi_i(s) - \varphi_i(-s) - 2s\varphi_i'(0)| \le 2s^2\lambda^2(Ad)_i^2$. Summing with the weights $\alpha x_i/n$,

$$\left|F(s) - F(-s) - \frac{2\alpha\lambda s}{n}\|d\|^2\right| \le \frac{2\alpha\lambda^2 s^2}{n}\|Ad\|^2,$$

which is the stated lower bound; it is positive for $s < \|d\|^2/(2\lambda\|Ad\|^2)$ (note $Ad \ne 0$, since $Ad = 0$ would give $\|d\|^2 = (D_\star x)^\top Ad = 0$). The box condition is $s < (1 - |m|)/\|d\|_\infty$.

Nondegeneracy. Let $n \ge 2$; for $n = 1$ the projector P vanishes and so does d. For fixed (m, λ, β, x), $\|PAD_\star x\|^2$ is a real-analytic function of the independent entries of W. It is not identically zero: at $W = 0$ its vector argument vanishes, and its derivative in a symmetric direction B is $\mathrm{sech}^2(\lambda m\beta)PBx$, nonzero for a suitable B because $P \ne 0$. A nonzero real-analytic function has a Lebesgue-null zero set, and the GOE law has a density, so $d \ne 0$ almost surely. The quantifier is each prescribed (m, λ), not a simultaneous statement over all of them.

The certificate on the phase-certificate instances. Table 7 evaluates part (b) at the two gains used in the experiments, with both states of each pair measured independently: their overlaps agree to within 10^{-15} and their unsigned atoms to within 10^{-17}, no coordinate of either state exceeds 0.085 in absolute value, and the realized gap exceeds the certified bound on 96/96 pairs. These are finite-instance checks; the theorem gives no bound uniform in n.

Table 7: Theorem 1(b) at the operating gains on the phase-certificate instances (12 per n, $m = 0.03$, amplitude half the safe radius): the smallest certified lower bound on the next-overlap gap and the median realized gap. The bound is positive on every instance.

	$n = 200$		$n = 400$		$n = 800$		$n = 1200$	
gain	bound	gap	bound	gap	bound	gap	bound	gap
$\lambda = 8$	0.038	0.054	0.036	0.054	0.039	0.056	0.038	0.055
$\lambda = 12$	0.028	0.042	0.029	0.041	0.031	0.043	0.029	0.043

A.4 Corollary 1 and Proposition 1

The pair of Theorem 1(b) gives the same input (m, ν) to any such function and different next overlaps, which proves the corollary. For the proposition, decompose $v = mx + r$ so that $U^\top v = mb + a$. Then $U^\top h = \Lambda(mb + a) + (\beta/n)b\, x^\top v = \mu \odot a + m(\mu + \beta) \odot b$. The inner product with b/n gives $n^{-1}x^\top h = \sum_k \mu_k \omega_k + m \sum_k (\mu_k + \beta)\rho_{x,k}$, which is the first identity since $\sum_k \rho_{x,k} = 1$; the squared norm over n gives the second after expanding the square.

B Proofs of Theorem 2 and Corollary 2

Lemma 3 (GOE action on a fixed vector)**.** *For deterministic u, w, $\mathrm{Cov}((Wu)_i, (Ww)_j) = \delta_{ij} n^{-1} u^\top w + n^{-1} u_j w_i$. Consequently $g := Wv_0$ has covariance $\varrho_n I + v_0 v_0^\top / n$ and the same law as $\sqrt{\varrho_n} z + \gamma v_0 / \sqrt{n}$ with independent $z \sim \mathcal{N}(0, I)$ and $\gamma \sim \mathcal{N}(0, 1)$.*

Proof. $\mathrm{Cov}(W_{ik}, W_{jl}) = n^{-1}(\delta_{ij}\delta_{kl} + \delta_{il}\delta_{kj})$, including the diagonal since W_{ii} has variance $2/n$; sum against $u_k w_l$. □

Proof of Theorem 2. Condition throughout on (x, v_0), which is legitimate because v_0 is independent of W given x.

Step 1: the scalar law. $Av_0 = \beta m_n x + g$ and, by Lemma 3, $g_i = \sqrt{\varrho_n} z_i + \gamma v_{0,i}/\sqrt{n}$. The rank-one part is at most $|\gamma| \, \|v_0\|_\infty / \sqrt{n} = o_P(1)$ uniformly in i, the only use of the delocalization hypothesis. Gauging by $\mathrm{diag}(x)$ and oddness give $x_i \tanh\{\lambda(\beta m_n x_i + g_i)\} = \tanh\{\lambda(\beta m_n + x_i g_i)\}$, and $x \odot z \overset{d}{=} z$. Conditional on γ the reference terms without the rank-one part are i.i.d. and the actual terms differ from them uniformly by $o_P(1)$, so the bounded triangular-array law of large numbers gives $S_0 - S_n^\star \to 0$ and hence the signal claim; the same argument gives $n^{-1} \|y_0\|^2 \to \mathrm{Var}(F_n)$.

Step 2: Stein's identity and concentration. In gauged coordinates let $s = Dv_0$, $\widetilde{W} = DWD$, $G = \widetilde{W}s$, $F_i = \phi(G_i)$ with $\phi(u) = \tanh\{\lambda(\beta m_n + u)\}$, and $Y = F - \bar{F}\mathbf{1} = Dy_0$. For $k \in \{Dr_0, \mathbf{1}\}$ the jointly Gaussian Stein identity applied to $((\widetilde{W}k)_i, G_j)$, whose covariance Lemma 3 gives exactly, yields $\mathbb{E}[n^{-1}(\widetilde{W}k)^\top Y] = (k^\top s/n)\{n^{-1}\sum_i \mathbb{E}\phi'(G_i)\} + o(1)$; since $\mathrm{Var}(G_i) = \varrho_n + s_i^2/n$, delocalization makes the average derivative converge to $\lambda\chi_n$. For the fluctuations, define $\Phi_k = n^{-1}(\widetilde{W}k)^\top (I - \mathbf{1}\mathbf{1}^\top/n)F$; its differential in a symmetric direction B is

$$d\Phi_k[B] = \tfrac{1}{n}(Bk)^\top (I - \mathbf{1}\mathbf{1}^\top/n)F + \tfrac{1}{n}(\widetilde{W}k)^\top (I - \mathbf{1}\mathbf{1}^\top/n)\,\mathrm{diag}(\phi'(G))Bs,$$

whose expected squared Frobenius norm is $O(1)$ because $|F_i| \le 1$, $|\phi'| \le \lambda$ and $\|k\|^2, \|s\|^2 = O(n)$. The GOE Gaussian Poincaré inequality (coordinate variances at most $2/n$) gives $\mathrm{Var}(\Phi_k) = O(1/n)$. Together with $n^{-1} r_0^\top y_0 \to 0$ from the channel representation of Step 1,

$$\tfrac{1}{n}(Wr_0)^\top y_0 \to \lambda\chi q, \qquad \tfrac{1}{n}(Wx)^\top y_0 \to \lambda\chi m, \qquad \tfrac{1}{n} r_0^\top y_0 \to 0.$$

Step 3: isotropic limits. For fixed vectors of squared norm $O(n)$, $n^{-1}u^\top W u \to 0$ and $n^{-1}\|Wu\|^2 - n^{-1}\|u\|^2 \to 0$ with $O(1/n)$ variances, and $n^{-1}x^\top W r_0 \to 0$ because $x^\top r_0 = 0$. Hence $n^{-1}\|Wr_0\|^2 \to q$, $n^{-1}\|Wx\|^2 \to 1$ and $n^{-1} r_0^\top W r_0,\ n^{-1} x^\top W r_0 \to 0$.

Step 4: the residual identities. With $\epsilon = y_0 - \lambda\chi_n W v_0$, whose modal form is $\epsilon_k = y_k - \lambda\chi_n\mu_k(a_k + m_n b_k)$, Parseval turns each inner product above into a modal sum; for instance $n^{-1}\|\epsilon\|^2 \to \mathrm{Var}(H) - 2\lambda^2\chi^2\varrho + \lambda^2\chi^2\varrho = V$. The orthogonality limits and the two ζ identities follow in the same way, using $n^{-1}\|Wx\|^2 \to 1$ for the last.

Step 5: the regression. The Gram matrix of $(a, \mu \odot a)$ over n has entries $n^{-1}\|r_0\|^2$, $n^{-1}r_0^\top W r_0$ and $n^{-1}\|Wr_0\|^2$, which tend to qI_2. With the exact modal update and $n^{-1}\sum_k a_k\zeta_k$, $n^{-1}\sum_k \mu_k a_k\zeta_k \to 0$, the least-squares coefficients converge to $(1-\alpha, \alpha\lambda\chi)$.

Step 6: $V > 0$. Stein's identity for $H(Z) = \tanh\{\lambda(\beta m + \sqrt{\varrho}Z)\}$ gives $\mathbb{E}[ZH(Z)] = \sqrt{\varrho}\lambda\chi$, so Cauchy–Schwarz yields $\mathrm{Var}(H) \geq \lambda^2\chi^2\varrho$, with equality only if H were affine in Z. It is not: it is strictly concave on one side of $-\beta m/\sqrt{\varrho}$ and strictly convex on the other. □

Why the first step is the easy case. A memory correction is a functional of earlier iterates, of which there are none at $t = 0$, so no statement about this step can exhibit what such a correction is designed to remove. Theorem 2 fixes an exact reference point; the first reuse is where the difference appears. Supplement Tables S1 and S2 test its two normal equations and its modal coefficients at every dimension.

Proof of Corollary 2. Since W is symmetric, $c_t = n^{-1}(Wx)^\top r_t$. $c_0 \to 0$ is Step 3 above. By Lemma 1, $r_1 = (1-\alpha)r_0 + \alpha y_0$, so $c_1 = (1-\alpha)n^{-1}(Wx)^\top r_0 + \alpha n^{-1}(Wx)^\top y_0 \to \alpha\lambda\chi m$ by Steps 2–3. Finally $n^{-1}x^\top A v_1 = \beta m_1 + m_1 n^{-1}x^\top Wx + c_1$, and $\mathrm{Var}(n^{-1}x^\top Wx) = 2/n$. □

The first-update spectral law. Proposition 2 upgrades the two normal equations of Theorem 2 to all bounded continuous tests. The extra ingredient is the resolvent of the part of W that the single product Wv_0 leaves unobserved. Let $s(z) = \int(\mu - z)^{-1}\rho_{\mathrm{sc}}(\mu)\,d\mu$, so that $s^2 + zs + 1 = 0$.

Lemma 4 (Resolvent of the unobserved block)**.** *Let W be GOE and independent of a σ-algebra $\mathcal{F}_0$, let e be an $\mathcal{F}_0$-measurable unit vector and $g = We$, and let f, h be $\mathcal{F}_0 \vee \sigma(g)$-measurable with $\|f\|, \|h\| = O_P(1)$, $e^\top f \to 0$ and $g^\top f \to 0$. Then $f^\top(W - z)^{-1}h - s(z)f^\top h \to 0$ for every fixed $z \in \mathbb{C} \setminus \mathbb{R}$.*

Proof. In an orthonormal basis whose first vector is e, write $W = \begin{pmatrix} a & b^\top \\ b & C \end{pmatrix}$, so that $g = (a, b)$. By orthogonal invariance C is a GOE block of dimension $n-1$ independent of (a, b) and $\mathcal{F}_0$, and $a \to 0$, $\|b\|^2 \to 1$. Let $Q = (C - z)^{-1}$. For vectors u, v of bounded norm that are measurable with respect to $\mathcal{F}_0 \vee \sigma(a, b)$, $u^\top Q v - s(z)u^\top v \to 0$: given the spectrum of C its eigenvectors are Haar distributed and independent of u, v, so $u^\top Q v$ has conditional mean $(n-1)^{-1}\,\mathrm{tr}(Q)\,u^\top v$ and conditional variance $O(\|u\|^2\|v\|^2/(n|\operatorname{Im} z|^2))$, and $(n-1)^{-1}\operatorname{tr} Q \to s(z)$ by the semicircle law. With $f = (f_e, f_\perp)$ and $h = (h_e, h_\perp)$ the Schur complement gives

$$f^\top(W - z)^{-1}h = f_\perp^\top Q h_\perp + \frac{(f_e - f_\perp^\top Q b)(h_e - b^\top Q h_\perp)}{a - z - b^\top Q b}.$$

The denominator tends to $-z - s(z) = 1/s(z) \neq 0$. Because $f_e = e^\top f \to 0$ and $b^\top f_\perp = g^\top f - a f_e \to 0$, the first factor of the numerator tends to zero while the second is $O_P(1)$, and $f_\perp^\top Q h_\perp = s(z)f^\top h + o_P(1)$. □

Proof of Proposition 2. Condition on (x, v_0) and let $\epsilon = y_0 - \lambda\chi_n W v_0$ as in Step 4 above, with $\epsilon_k = \langle u_k, \epsilon\rangle$. Lemma 4 is applied with $\mathcal{F}_0 = \sigma(x, v_0)$, and the vectors e, f, h to which it is applied below are measurable with respect to $\sigma(x, v_0, Wv_0)$, because y_0 depends on W only through $Av_0 = \beta m_n x + Wv_0$; the intermediate vectors Wx and Wr_0 are not, and enter only through the limits of Steps 2–4. Steps 2–4 give $n^{-1}\|\epsilon\|^2 \to V$ and $n^{-1}r_0^\top\epsilon$, $n^{-1}(Wr_0)^\top\epsilon$, $n^{-1}(Wx)^\top\epsilon \to 0$, and $x^\top y_0 = 0$ gives $n^{-1}x^\top\epsilon = -\lambda\chi_n n^{-1}x^\top Wv_0 \to 0$. With $e = v_0/\|v_0\|$ and $f = \epsilon/\sqrt{n}$, therefore, $e^\top f \to 0$ and $(We)^\top f = (Wr_0 + m_n Wx)^\top\epsilon/(n\sqrt{\varrho_n}) \to 0$. Lemma 4 with $h \in \{f, r_0/\sqrt{n}, x/\sqrt{n}\}$ gives, for every $z \notin \mathbb{R}$,

$$\frac{1}{n}\sum_k \frac{\epsilon_k^2}{\mu_k - z} \to V s(z), \qquad \frac{1}{n}\sum_k \frac{\epsilon_k a_k}{\mu_k - z} \to 0, \qquad \frac{1}{n}\sum_k \frac{\epsilon_k b_k}{\mu_k - z} \to 0.$$

The three measures $n^{-1}\sum_k \epsilon_k^2\delta_{\mu_k}$, $n^{-1}\sum_k \epsilon_k a_k\delta_{\mu_k}$ and $n^{-1}\sum_k \epsilon_k b_k\delta_{\mu_k}$ have total variation $O_P(1)$ by Cauchy–Schwarz and live on $[-3,3]$ with probability tending to one, and resolvents span a dense subspace of $C_0(\mathbb{R})$ (Appendix D), so they converge weakly to $V\rho_{\mathrm{sc}}$, 0 and 0. Independence of v_0 and W gives $\nu_0 \Rightarrow q\rho_{\mathrm{sc}}$, $\rho_x \Rightarrow \rho_{\mathrm{sc}}$ and $\omega_0 \Rightarrow 0$. Square the exact modal update $a_{1,k} = G_{0,n}(\mu_k)a_k + \alpha\lambda\chi_n m_n\mu_k b_k + \alpha\epsilon_k$, with $G_{0,n}(\mu) = (1-\alpha) + \alpha\lambda\chi_n\mu$, or multiply it by b_k, and test against a bounded continuous φ: every mixed term vanishes by the limits above, which proves both statements. Random initial data are handled by the subsequence argument at the end of Appendix C. □

The limit of ν_1 has mass $(1-\alpha)^2 q + \alpha^2(\lambda^2\chi^2\varrho + V)$, as the scalar first-step law requires, and the limit of ω_1 has first moment $\alpha\lambda\chi m$, Corollary 2. For $m = 0$ the signed limit vanishes while the residual spectrum is already reshaped by G_0^2.

C Proofs of Theorem 3 and Corollary 3

A function $\psi : \mathbb{R}^d \to \mathbb{R}$ is pseudo-Lipschitz of order two ($\psi \in \mathrm{PL}_2$) if $|\psi(a) - \psi(b)| \le L(1 + \|a\| + \|b\|)\,\|a - b\|$; then $|\psi(a)| \le C_\psi(1 + \|a\|^2)$. W_p is the p-Wasserstein distance. The first four lemmas turn the 4-Wasserstein hypothesis on the seed into coordinatewise limits; they are standard, and we prove them because the argument is an iteration of these moves.

Lemma 5 (PL_2 testing is W_2 convergence)**.** *If $W_2(\pi_n, \pi) \to 0$, then for $\psi \in \mathrm{PL}_2$ with constant L, $|\int \psi\, d\pi_n - \int \psi\, d\pi| \le L(1 + M_n + M)W_2(\pi_n, \pi) \to 0$, where $M_n^2 = \int \|a\|^2\, d\pi_n \to M^2 = \int \|a\|^2\, d\pi$.*

Proof. For an optimal coupling (A, B), $\mathbb{E}|\psi(A) - \psi(B)| \le L\mathbb{E}[(1 + \|A\| + \|B\|)\,\|A - B\|] \le L\,\|1 + \|A\| + \|B\|\|_{L^2}\, W_2(\pi_n, \pi)$ by Cauchy–Schwarz, and W_2 convergence implies convergence of second moments. □

Lemma 6 (Tensorization with a Gaussian channel)**.** *If $\pi_n \to \pi$ in $W_2(\mathbb{R}^d)$ and $\sigma_n \to \sigma \ge 0$, then $\pi_n \otimes \mathcal{N}(0, \sigma_n^2) \to \pi \otimes \mathcal{N}(0, \sigma^2)$ in $W_2(\mathbb{R}^{d+1})$. If $\psi \in \mathrm{PL}_2(\mathbb{R}^{d+1})$ has constant L, then $\Psi(a) = \mathbb{E}\psi(a, \sigma Z)$ is in $\mathrm{PL}_2(\mathbb{R}^d)$ with constant $L(1 + 2\sigma\sqrt{2/\pi})$.*

Proof. Independent coupling gives $W_2^2(\mu_1 \otimes \mu_2, \nu_1 \otimes \nu_2) \le W_2^2(\mu_1, \nu_1) + W_2^2(\mu_2, \nu_2)$, and $W_2(\mathcal{N}(0, \sigma_n^2), \mathcal{N}(0, \sigma^2)) = |\sigma_n - \sigma|$. For the last claim, $|\Psi(a) - \Psi(b)| \le L\mathbb{E}[1 + \|a\| + \|b\| + 2\sigma|Z|]\,\|a - b\|$. □

Lemma 7 (Composition with a coordinatewise update)**.** *If $\phi : \mathbb{R}^d \to \mathbb{R}$ is L_ϕ-Lipschitz and $\psi \in \mathrm{PL}_2(\mathbb{R}^{d+1})$ has constant L, then $a \mapsto \psi(a, \phi(a))$ is in $\mathrm{PL}_2(\mathbb{R}^d)$ with constant $L(1 + L_\phi)\max\{1 + 2|\phi(0)|, 1 + L_\phi\}$. Hence an empirical PL_2 limit of an array $\{A_{n,i}\}$ extends to $\{(A_{n,i}, \phi(A_{n,i}))\}$, and bounded fourth empirical moments are preserved.*

Proof. With $c = (a, \phi(a))$ and $d = (b, \phi(b))$: $\|c - d\| \le (1 + L_\phi)\,\|a - b\|$ and $1 + \|c\| + \|d\| \le (1 + 2|\phi(0)|) + (1 + L_\phi)(\|a\| + \|b\|)$; $|\phi(0)|$ enters $\|c\|$ and $\|d\|$ separately. Also $\|c\|^4 \le 8\{(1 + L_\phi)^4\,\|a\|^4 + |\phi(0)|^4\}$. □

Lemma 8 (Normalized-ℓ_2 stability)**.** *If $n^{-1}\sum_i \|A_{n,i} - B_{n,i}\|^2 \to 0$ and $n^{-1}\sum_i(\|A_{n,i}\|^2 + \|B_{n,i}\|^2) = O_P(1)$, then $n^{-1}\sum_i\{\psi(A_{n,i}) - \psi(B_{n,i})\} \to 0$ for every $\psi \in \mathrm{PL}_2$.*

Proof. The pseudo-Lipschitz bound and Cauchy–Schwarz give an $O_P(1)$ factor times an $o_P(1)$ factor. □

The 4-Wasserstein hypothesis is used exactly twice: W_4 convergence implies the W_2 convergence Lemmas 5–6 consume, and it bounds $n^{-1}\sum_i |s_{n,i}|^4$, which the variance bounds below need because the square of a PL_2 function grows quartically. The theorem assumes convergence in probability, which gives only $O_P(1)$ fourth moments; the lemmas below are stated for a deterministic sequence, along which $\sup_n n^{-1}\sum_i |s_{n,i}|^4 < \infty$, and the final subsequence argument connects the two.

Lemma 9 (Empirical GOE channel)**.** *Let s_n be deterministic with $n^{-1}\sum_i \delta_{s_{n,i}} \to \pi$ in W_4 and $\varrho = \mathbb{E}V_0^2 > 0$ for $V_0 \sim \pi$. For GOE W_n and every $\psi \in \mathrm{PL}_2$, $n^{-1}\sum_i \psi(s_{n,i}, (W_n s_n)_i) \to \mathbb{E}\psi(V_0, \sqrt{\varrho}Z_0)$ with Z_0 independent of V_0, and the fourth empirical moments are $O_P(1)$.*

Proof. With $\varrho_n = \|s_n\|^2/n$, $W_n s_n \overset{d}{=} \sqrt{\varrho_n}z_0 + \gamma_0 s_n/\sqrt{n}$; the rank-one term has normalized squared norm $\gamma_0^2\varrho_n/n \to 0$, so Lemma 8 reduces the claim to $(s_{n,i}, \sqrt{\varrho_n}z_{0,i})$. These pairs are independent given s_n, with variances at most $C(1 + |s_{n,i}|^4 + \varrho_n^2\mathbb{E}z^4)$, so the average has variance $O(1/n)$; this is the only step that needs W_4. Its mean is $\int \psi\, d(\widehat{\pi}_n \otimes \mathcal{N}(0, \varrho_n)) \to \mathbb{E}\psi(V_0, \sqrt{\varrho}Z_0)$ by Lemmas 6 and 5. □

Lemma 10 (GOE conditioned on one product)**.** *Fix $s \neq 0$, $L = \|s\|^2$, $P_s = I - ss^\top/L$, $g = Ws$ and $M(s,g) = (gs^\top + sg^\top)/L - (s^\top g/L^2)ss^\top$. Then $\mathcal{L}(W \mid s, Ws = g) = \mathcal{L}\{M(s,g) + P_s W' P_s \mid s, g\}$ with W' an independent GOE.*

Proof. Choose an orthogonal Q with first column $s/\|s\|$. By orthogonal invariance $Q^\top W Q$ has independent blocks $a \sim \mathcal{N}(0, 2/n)$, b with i.i.d. $\mathcal{N}(0, 1/n)$ entries and a symmetric Gaussian C; observing Ws reveals (a, b) and leaves C unchanged. Rotating back gives $M(s,g)$ and $P_s W' P_s$. □

Lemma 11 (Conditional fresh Gaussian channel)**.** *Let $s_n \neq 0$ and w_n be $\mathcal{F}_n$-measurable with $s_n^\top w_n = 0$, let $P_n = I - s_n s_n^\top/\|s_n\|^2$, $Q_n = \|w_n\|^2/n \to Q$, and let W_n' be GOE independent of $\mathcal{F}_n$. If an $\mathcal{F}_n$-measurable array $T_{n,i}$ has an empirical PL_2 limit T with $O_P(1)$ fourth moments, then $n^{-1}\sum_i \psi(T_{n,i}, (P_n W_n' w_n)_i) \to \mathbb{E}\psi(T, \sqrt{Q}Z_1)$ with Z_1 independent of T.*

Proof. Conditionally, $\mathrm{Cov}(P_n W_n' w_n) = Q_n P_n + w_n w_n^\top/n$, so $P_n W_n' w_n \overset{d}{=} \sqrt{Q_n}P_n z_1 + \gamma_1 w_n/\sqrt{n}$; the projection and rank-one corrections are $O_P(1/n)$ in normalized squared norm, and Lemma 8 reduces the claim to $(T_{n,i}, \sqrt{Q}z_{1,i})$. The conditional variance of the average is at most $Cn^{-2}\sum_i(1 + \|T_{n,i}\|^4) = o_P(1)$ and the conditional mean converges because $t \mapsto \mathbb{E}\psi(t, \sqrt{Q}Z)$ is PL_2 (Lemma 6). A conditional variance that is $o_P(1)$ does not give an unconditional bound by Chebyshev alone; instead $\Pr(|X_n - \mathbb{E}[X_n \mid \mathcal{F}_n]| > \varepsilon) \le \mathbb{E}\min\{1, V_n/\varepsilon^2\} \to 0$ by bounded convergence, with V_n the conditional variance. □

Proof of Theorem 3. Write $f(h) = \tanh(\lambda h)$, $m_{0,n} = n^{-1}\sum_i s_{n,i}$ and note that $u_{1,n} = D_n v_{1,n} = (1-\alpha)s_n + \alpha f(\beta m_{0,n}\mathbf{1} + g_{0,n})$ by oddness of f. First fix a deterministic sequence along which the seed law converges in W_4. $\widetilde{W}_n$ is GOE and independent of s_n, so Lemma 9 gives $(s_{n,i}, g_{0,n,i}) \Rightarrow (V_0, \sqrt{\varrho}Z_0)$ against every PL_2 test. Since f is λ-Lipschitz, $n^{-1}\|f(\beta m_{0,n}\mathbf{1} + g_0) - f(\beta m\mathbf{1} + g_0)\|^2 \le \lambda^2\beta^2(m_{0,n} - m)^2 \to 0$ (Lemma 5 with $a \mapsto a$), so Lemmas 8 and 7 with the Lipschitz map $\phi(a,b) = (1-\alpha)a + \alpha f(\beta m + b)$ give $(s_{n,i}, g_{0,n,i}, u_{1,n,i}) \Rightarrow (V_0, \sqrt{\varrho}Z_0, V_1)$ with $O_P(1)$ fourth moments.

With $s = s_n$, $g = g_{0,n}$, $u = u_{1,n}$ and $L = \|s\|^2$, define $k_n = s^\top u/L$, $w_n = u - k_n s$ and $e_n = (g^\top u - k_n s^\top g)/L$. The joint limit and $L/n \to \varrho$ give $k_n \to C_{10}/\varrho$, $n^{-1}s^\top g \to 0$ and $n^{-1}g^\top u \to \alpha\mathbb{E}[\sqrt{\varrho}Z_0 f(H_0)] = \alpha\varrho\mathbb{E}f'(H_0)$ by Stein's identity, hence $e_n \to R_{10}$; and $Q_{1,n} = \|w_n\|^2/n = \|u\|^2/n - (n^{-1}s^\top u)^2/(n^{-1}\|s\|^2) \to C_{11} - C_{10}^2/\varrho = Q_1$, which also shows $Q_1 \ge 0$. Given (s, g) the vector u is measurable, and Lemma 10 gives the exact representation

$$g_{1,n} = \widetilde{W}_n u = k_n g + e_n s + \xi_n, \qquad \xi_n = P_s W_n' w_n,$$

with conditional covariance $Q_{1,n}P_s + w_n w_n^\top/n$, whose normalized trace is exactly $\{Q_{1,n}(n-1) + Q_{1,n}\}/n = Q_{1,n}$. Lemma 11 with $T_{n,i} = (s_i, g_i, u_i)$ gives $(s, g_0, u_1, \xi) \Rightarrow (V_0, \sqrt{\varrho}Z_0, V_1, \sqrt{Q_1}Z_1)$; replacing k_n, e_n by their limits changes the array by $o_P(1)$ in normalized ℓ_2, and Lemma 8 proves the joint law.

For the second overlap, $m_{1,n} \to m_1$ by the joint law with the test $(a,b,c,d) \mapsto c$, and f is λ-Lipschitz, so $m_{1,n}$ may be replaced by m_1 inside f; the Lipschitz map $(c,d) \mapsto (1-\alpha)c + \alpha f(\beta m_1 + d)$ then gives $m_2 = (1-\alpha)m_1 + \alpha\mathbb{E}f(\beta m_1 + \mathcal{G}_1)$. The two centered Gaussian terms of $\mathcal{G}_1$ are independent of V_0 with combined variance $C_{10}^2/\varrho + Q_1 = C_{11}$, which gives the stated formula. For

the reaction, $c_{1,n} = n^{-1}\mathbf{1}^\top g_{1,n} - m_{1,n}n^{-1}\mathbf{1}^\top \widetilde{W}_n\mathbf{1} \to \mathbb{E}\mathcal{G}_1 = R_{10}m$, since $\mathrm{Var}(n^{-1}\mathbf{1}^\top\widetilde{W}_n\mathbf{1}) = 2/n$.

Finally, from every subsequence the convergence of the seed law in probability admits a further subsequence on which it holds almost surely; conditional on almost every such seed sequence the argument above applies, and dominated convergence for the conditional failure probabilities with the subsequence principle gives convergence in probability. □

Proof of Corollary 3. Since $D_n^2 = I$, $n^{-1}\sum_i s_i g_{1,i} = n^{-1}s^\top D_n W_n D_n u_1 = n^{-1}v_0^\top W_n v_1$ and $n^{-1}\|g_1\|^2 = n^{-1}\|W_n v_1\|^2$. Both $(a,b,c,d) \mapsto ad$ and $(a,b,c,d) \mapsto d^2$ are in PL_2, so Theorem 3 gives the limits $\mathbb{E}[V_0\mathcal{G}_1]$ and $\mathbb{E}\mathcal{G}_1^2$. As V_0 is independent of (Z_0, Z_1), $\mathbb{E}[V_0\mathcal{G}_1] = R_{10}\mathbb{E}V_0^2 = R_{10}\varrho$ and $\mathbb{E}\mathcal{G}_1^2 = C_{10}^2/\varrho + R_{10}^2\varrho + Q_1 = C_{11} + R_{10}^2\varrho$. Without the response term, $\mathcal{G}_1^{\mathrm{nr}} = (C_{10}/\varrho)\sqrt{\varrho}Z_0 + \sqrt{Q_1}Z_1$ gives 0 and C_{11}. With a fresh GOE W'' for the second product, Lemma 9 applied to u_1 gives a field $\sqrt{C_{11}}Z$ independent of V_0, hence the same limits 0 and C_{11}. For symmetric π, $m = 0$ makes H_0 symmetric and $m_1 = \mathbb{E}V_1 = 0$; then both laws give $m_2 = \alpha\mathbb{E}\tanh(\lambda X)$ for a symmetric variable X, which is 0. □

Scope. The field of Theorem 3 separates a retained old Gaussian direction, a response-weighted copy of the initial coordinate and fresh Gaussian noise. It is an exact first-reuse limit without (A1)–(A5). It gives no recursion for the signed measure, does not cover matrix-dependent seeds, and does not continue automatically to later reuses. Gaussian seeds are not bounded in the variational cube; the two-point seed is the bounded check. Supplement Tables S11–S15 report the coefficient limits, the reuse and fresh-matrix controls, joint moments and fixed pseudo-Lipschitz tests at every dimension of the first-reuse suite.

D UNSIGNED TRANSPORT, (A5), AND THE ERROR BUDGET

Proof of Lemma 2. By Lemma 1's argument $r' = (1-\alpha)r + \alpha\{\tanh(\lambda Av) - Sx\}$. By the definition of ϵ, $\tanh(\lambda Av) - Sx = L(Av - \bar{h}x) + U\epsilon$, and $U^\top Av = \beta mb + \Lambda(mb + a) = \mu \odot a + m(\mu + \beta) \odot b$ gives $U^\top(Av - \bar{h}x) = \mu \odot a + \{m(\mu+\beta) - \bar{h}\} \odot b$. Hence $a' = (1 - \alpha + \alpha L\mu) \odot a + \alpha L\{m(\mu+\beta) - \bar{h}\} \odot b + \alpha\epsilon$. Both the centered response and the centered field are orthogonal to x, so $b^\top\epsilon = 0$. Squaring a'_k gives the six terms of the lemma. No Gaussian assumption is used.

The unsigned transport. The diagnostic of Section 6 replaces the response by its first Hermite component with the Gaussian χ_t, evaluates the field at its memoryless mean βm_t in place of $\bar{h}_t = \beta m_t + m_t d_1 + c_t$ (Proposition 1), so that $H_t = \alpha\lambda\chi_t m_t \mu$, and keeps, from Lemma 2, only $G_t^2\nu_t$, the signal injection (with the semicircle ρ_{sc} in place of the quenched ρ_x) and a white innovation $\alpha^2 V_t^{\mathrm{nl}}\rho_{\mathrm{sc}}$. This drops (i) the signed term $2G_tH_t\omega_t$, which is (A5), (ii) the remainder cross terms $2\alpha G_t\tau_{a\epsilon} + 2\alpha H_t\tau_{b\epsilon}$, which is the orthogonality part of (A4), and (iii) the spectral shape of η_ϵ, the whiteness part of (A4). The Stieltjes form Eq. (3) follows from $\mu/(\mu - z) = 1 + z/(\mu - z)$ applied once and twice, and from $\int \mu^2\rho_{\mathrm{sc}}(\mu)d\mu/(\mu - z) = z + z^2 g_{\mathrm{sc}}(z)$, which uses $\int \rho_{\mathrm{sc}} = 1$ and $\int \mu\,\rho_{\mathrm{sc}} = 0$.

The memoryless mean, (A5), and the reaction. Since $d_1 = n^{-1}x^\top Wx$ has variance $2/n$, the memoryless mean amounts to $c_t = \int \mu\, d\omega_t \approx 0$. Corollary 2 gives $c_1 \to \alpha\lambda\chi m \neq 0$ for $m \neq 0$; at later times the reaction suite measures $|c_t|$ that does not decay with n for delocalized and mid-bulk seeds (Section 6), a finite-n observation rather than a limit theorem. (A5) is a condition on the whole signed measure, and it implies $c_t \to 0$. The signed measures that occur are not arbitrary: by Cauchy–Schwarz over the modes,

$$\Big|\int f\, d\omega_t\Big| \le \sqrt{q_t}\Big(\int |f|^2\, d\rho_x\Big)^{1/2}$$

for every continuous f, and $q_t \le \varrho_t$ stays bounded along the iteration because $n^{-1/2}\|v_{t+1}\| \le (1-\alpha)n^{-1/2}\|v_t\| + \alpha$.

Lemma 12 (Signed limits have densities)**.** *Suppose all measures live on a fixed compact set K, $q_t \le Q$, $\rho_x \Rightarrow \rho$ and, along a subsequence, $\omega_t \Rightarrow \omega$. Then $d\omega = w\, d\rho$ for some w with $\|w\|_{L^2(\rho)} \le \sqrt{Q}$.*

Proof. In the limit $|\int f\,d\omega| \le \sqrt{Q}\,\|f\|_{L^2(\rho)}$ for $f \in C(K)$. The functional extends continuously to $L^2(\rho)$, in which $C(K)$ is dense, and the Riesz representation gives w; $w\rho$ and ω agree on $C(K)$. □

For GOE, $\rho_x \Rightarrow \rho_{\mathrm{sc}}$ in probability, ρ_{sc} has no atoms, and the spectrum of W lies in $[-3,3]$ with probability tending to one. The resolvents $(\mu - z)^{-1}$, $z \in \mathbb{C} \setminus \mathbb{R}$, span a dense subspace of $C_0(\mathbb{R})$: by partial fractions and difference quotients in z, the closure of their span contains the self-adjoint algebra they generate, which separates points and vanishes nowhere, so Stone–Weierstrass applies. Now let (A5) hold for every resolvent test, with $G_t = (1-\alpha) + \ell_t\mu$ and $\ell_t = \alpha\lambda\chi_t \in [0, \alpha\lambda]$. On events of probability arbitrarily close to one the spectrum lies in $[-3,3]$ and $q_t \le Q$; there the total variation of ω_t is at most $\sqrt{Q}$, so $(\omega_t, \rho_x, \ell_t)$ ranges over a compact metrizable set and the sequence is tight. Along any subsequence a further one converges in distribution and, by Skorokhod's representation, almost surely on another probability space, to some $(\omega, \rho_{\mathrm{sc}}, \ell)$. The bound above and (A5) at a countable dense set of z pass to this limit, so $\omega = w\rho_{\mathrm{sc}}$ by Lemma 12 and the Stieltjes transform of $\mu(1-\alpha+\ell\mu)\,\omega$ vanishes off $\mathbb{R}$; hence $\mu(1-\alpha+\ell\mu)w(\mu) = 0$ for ρ_{sc}-almost every μ. The polynomial is nonzero with at most two roots, which carry no ρ_{sc} mass, so $w = 0$. Every subsequential limit in distribution is thus the deterministic zero measure, so $\omega_t \Rightarrow 0$ and $c_t \to 0$ in probability, even when G_t has a zero inside the bulk. The converse fails. In Remark 1, $P(\mu) = \mu^3 - 2\mu = U_3(\mu/2)$ is orthonormal for ρ_{sc}, so the admissible limit $\omega = P\rho_{\mathrm{sc}}$ has $\int \mu^j\,d\omega = 0$ for $j \le 2$, including $c = 0$, while (A5) fails for it. Two qualifications remain: a fixed grid of resolvents, such as the 41 points of our diagnostic, does not establish (A5); and when $m_t = 0$ the signed term vanishes without it, so (A5) suffices for dropping that term and is necessary only when $\alpha\lambda\chi_t m_t \neq 0$. The budget below measures the signed term directly.

The exact budget. On every teacher-forced true state of the closure suite, with the predictor's own χ_t and innovation variance V_t^{nl}, put $B = G^2\nu + H^2\rho_x + \alpha^2 V^{\mathrm{nl}}\lambda_n$ with λ_n the uniform atoms $1/n$. With ν^U the implemented unsigned prediction,

$$\nu' - \nu^U = \underbrace{2GH\omega}_{\text{signed}} + \underbrace{2\alpha G\tau_{a\epsilon}}_{\text{residual cross}} + \underbrace{2\alpha H\tau_{b\epsilon}}_{\text{signal cross}} + \underbrace{\alpha^2(\|\epsilon\|^2/n - V^{\mathrm{nl}})\lambda_n}_{\text{innovation energy}} + \underbrace{\alpha^2(\eta_\epsilon - (\|\epsilon\|^2/n)\lambda_n)}_{\text{innovation shape}} + \underbrace{B - \nu^U}_{\text{coefficient}}$$

exactly; the identity is linear and holds for the Stieltjes transforms. Each component's share is its projection $\mathrm{Re}\langle g_X, R\rangle / \|R\|^2$ on the total error R, so the shares sum to one and cancellation shows as negative shares. B uses true-state information, so this is an attribution of a teacher-forced error, not a predictor; restoring one term in the error is an oracle ablation, and what remains need not be a positive measure. Table 8 gives the shares by update. The split depends on the scalar L in the definition of ϵ, although the unsigned prediction and the total error do not. Table 9 repeats the attribution with the empirical coordinate susceptibility, which inflates the cross term and offsets it in the coefficient term, and with the least-squares slope of the centered response on the centered field; the residual–remainder term remains the largest for delocalized seeds under all three. Supplement Tables S3, S4 and S10 give the one-step and pooled Stieltjes errors by dimension, the modewise calibration of G_t, and the three signed-closure predictors by dimension.

E THE REACTION READOUT

The reaction suite's historical baseline used the mean βm_t and the variance $e_2 = \int \mu^2 d\nu_t$ of the residual part of the field only, omitting the $m_t^2\, n^{-1}\,\|Wx\|^2$ contribution of the signal part; along the recorded trajectories m_t reaches 0.87. With accurate quadrature this baseline gives correlations between its residual and c_t of opposite signs across seed classes; with the full field variance the mid-bulk sign reverses. The four readouts of Section 6 avoid this ambiguity by using the exact field moments. Table 10 gives them by dimension, and Supplement Table S16 the historical baseline's statistics.

Table 8: Exact budget of the unsigned transport's teacher-forced spectral error by seed class and update ($\lambda = 12$, $n \in \{200, \dots, 1200\}$, six draws per n), under the Gaussian linearization of Lemma 2: median relative error, median projection shares (%) of the signed, residual–remainder, signal–remainder, innovation and coefficient terms, and the error left after restoring only the signed term or only the residual–remainder term. The budget identity holds to $< 10^{-15}$.

class	t	error	signed	res. cross	sig. cross	innov.	coeff.	with signed	with cross
blind	1	0.056	-1	98	-0	4	0	0.051	0.014
blind	2–4	0.058	1	99	-0	-2	0	0.057	0.008
blind	5–10	0.096	1	104	-0	-5	-0	0.094	0.006
aligned	1	0.052	3	88	1	7	0	0.051	0.020
aligned	2–4	0.058	2	97	-0	-2	-0	0.058	0.011
aligned	5–10	0.098	4	100	-0	-5	-0	0.093	0.009
mid-bulk	1	0.037	90	2	-5	3	8	0.012	0.037
mid-bulk	2–4	0.257	52	61	-8	-7	5	0.216	0.152
mid-bulk	5–10	0.338	38	71	-4	-5	1	0.232	0.133

Table 9: Median projection shares (%) of the residual–remainder and signed terms after the first reuse (updates 2–10) when the remainder is defined with the Gaussian $\lambda\chi_t$, the empirical coordinate susceptibility $\lambda\hat{\chi}_t$, or the least-squares slope of the centered response on the centered field. The unsigned prediction and the total error are the same in all three columns; only the split between terms changes.

class	Gaussian $\lambda\chi_t$		coordinate $\lambda\hat{\chi}_t$		least squares	
	cross	signed	cross	signed	cross	signed
blind	104	1	181	1	88	1
aligned	100	4	172	2	90	4
mid-bulk	70	39	98	19	78	33

F Numerical integration

Every Gaussian surrogate needs $S = \mathbb{E}\tanh(\lambda H)$, $\chi = \mathbb{E}\operatorname{sech}^2(\lambda H)$ and $\operatorname{Var}\tanh(\lambda H)$ for $H \sim \mathcal{N}(\mu_H, \sigma_H^2)$. In the standard normal variable the integrand switches sign at $-\mu_H/\sigma_H$ over a width of order $1/(\lambda\sigma_H)$, which a fixed Gauss–Hermite rule under-resolves at large gain. We use a composite 16-node Gauss–Legendre rule on $[-13, 13]$ with panels of width 0.25, refined to $0.25/(\lambda\sigma_H)$ within $40/(\lambda\sigma_H)$ of the switch; the complex singularities of the integrand lie $\pi/(2\lambda\sigma_H)$ off the real axis, about twelve refined half-widths away. The rule refuses inputs outside $|\mu_H| \le 8$, $\sigma_H \le 8$, $|\lambda| \le 16$. Table 11 summarizes its verification. The first-reuse suite at $\lambda = 2$ keeps a fixed 120-node Gauss–Hermite rule; that rule is inaccurate at large $\lambda\sigma_H$, but at the suite's actual parameters its limits R_{10}, C_{10}, C_{11}, m_1 and m_2 agree with the verified rule to within 10^{-9} for every seed class, far below the four decimals reported.

G Experimental protocol

Instances and seeds. Every observation comes from a synthetic spiked-Wigner instance: x uniform on $\{-1, +1\}^n$, $W = (G + G^\top)/\sqrt{2n}$ for a matrix G of independent standard Gaussians, and $A = (\beta/n)xx^\top + W$. There is no external dataset. Each random draw uses a generator seeded by SHA-256 of the configuration's base seed and the labels of the draw (suite, dimension, parameters and draw index), so a rerun of an unchanged configuration on the same platform reproduces every trial and summary; only timestamps and runtimes vary. Across platforms, values can differ in the last digits, and one class of records by more: the gain-12 AMP recursion amplifies rounding (a 10^{-16} change of the seed moves its output by more than 10^{-4}, whereas a 10^{-14} change moves the gain-1.1 output by less than 10^{-12}), so the tuning scores of unselected high-gain AMP candidates can move in the third decimal. In every recovery cell the best such candidate scores at least 0.029 below the shortlist, so no selection depends on them. Table 12 lists the parameters of every suite.

Table 10: The four Gaussian readouts of the signal update by seed class and dimension (eight draws per cell, 15 steps): median absolute error and reduction in squared error relative to U. The replay reproduces the archived trajectories (largest gap $< 10^{-12}$).

class	n	U	M	V	MV	reduction by M	by MV
delocalized	500	0.0149	0.0062	0.0149	0.0052	79%	83%
delocalized	1000	0.0208	0.0061	0.0209	0.0052	89%	92%
delocalized	2000	0.0177	0.0059	0.0179	0.0053	88%	90%
mid-bulk	500	0.0173	0.0056	0.0213	0.0029	91%	96%
mid-bulk	1000	0.0173	0.0037	0.0198	0.0018	93%	97%
mid-bulk	2000	0.0163	0.0041	0.0196	0.0011	91%	97%
top-bulk	500	0.0100	0.0043	0.0100	0.0044	80%	81%
top-bulk	1000	0.0047	0.0026	0.0047	0.0026	70%	70%
top-bulk	2000	0.0043	0.0020	0.0043	0.0020	67%	68%

Table 11: Maximum absolute error in S, χ and $\mathrm{Var}\tanh$. The first three rows verify the rule used in this paper over 6075 states on a dense grid at gains $2, 8, 12$, 4000 random states in the whole accepted region and 7 edge cases (the mpmath row uses the 80 hardest and random cases). The last rows are the fixed rules of an earlier version of the code, on the same grid at the stated gain. Every high-gain number in this paper was recomputed with the verified rule; the first-reuse suite at $\lambda = 2$ is checked separately in the text.

comparison	S	χ	$\mathrm{Var}\tanh$
new rule vs adaptive quadrature	$< 10^{-13}$	$< 10^{-13}$	$< 10^{-13}$
new rule vs halved panels	$< 10^{-14}$	$< 10^{-14}$	$< 10^{-14}$
new rule vs 30-digit mpmath	$< 10^{-14}$	$< 10^{-14}$	–
100-node Gauss–Hermite, $\lambda = 12$	0.078	0.090	0.097
100-node Gauss–Hermite, $\lambda = 8$	0.064	0.084	0.090
160-node Gauss–Hermite, $\lambda = 8$	0.042	0.060	0.064

Seed classes. A *blind* seed is drawn from $\mathcal{N}(0, I)$ and rescaled to RMS 0.3; the function that draws it cannot receive x, W or A. An *aligned* seed has overlap 0.03 with x, RMS 0.3 and an isotropic residual (it is the delocalized class of the reaction suite). A *mid-bulk* or *top-bulk* seed has the same overlap and RMS, with its residual along the middle or the top eigenvector of W, and the *phase* seeds are the matched pair of Theorem 1(a). A *PCA* seed is the leading eigenvector of A at RMS 0.3. In the first-reuse suite the gauged seed coordinates $x_i v_{0,i}$ are i.i.d.: Gaussian with mean 0.25 and RMS 0.65 (aligned), centered Gaussian with RMS 0.65 (blind), or two equally weighted atoms with the aligned mean and second moment (two-point). Only blind and PCA seeds are feasible; the others are oracle probes that isolate mechanisms and never enter a recovery claim.

Recovery and selection. For each $(\beta, \text{seed class})$ cell and each method, the candidate grid is scored by the mean absolute cosine at the final step on a tuning split, the best three candidates are rescored on a validation split, and the winner is frozen before the test split is touched. The three splits never share a realization, and test draws never enter selection. Grids: naive MF gain $\{0.75, 1, 1.5, 2\}\beta \cup \{4, 8, 12\}$ with damping $\{0.3, 0.5, 1\}$; AMP gain $\{0.5, 0.75, 1, 1.25, 1.5, 2\}\beta \cup \{4, 8, 12\}$; filtered MF gain $\{8, 12\}$, depth $\{0.5, 0.9\}$, damping 0.3 and band $[0, 2]$, $[0.6, 2]$ or $[1.2, 2]$. Supplement Table S8 lists the setting every cell froze, so both the grid and the final choice are on record, and Supplement Tables S5–S7 and S9 give the operator-transfer controls, recovery from PCA seeds and the cost of a polynomial filter. Recovery is the absolute cosine $|v^\top x|/(\|v\|\,\|x\|)$ at the final iterate; paired differences resample test draws, with 5000 bootstrap replicates.

Other metrics. Modal NRMSE is $\|a_{t+1}^{\mathrm{pred}} - a_{t+1}\|_2 / \|a_{t+1}\|_2$. The slope reported in the supplement is a one-predictor OLS coefficient with an intercept, whereas Theorem 2 uses two features and no intercept, so the two are reported separately. The spectral error is the relative ℓ_2 error of the Stieltjes transform $g(z) = \int (\mu - z)^{-1} d\nu(\mu)$ on 41 equally spaced points of $[-2.5, 2.5] + i\eta$.

Table 12: Parameters of every suite, read from its committed configuration. "Draws" are independent (x, W) realizations per cell (held-out test draws for recovery); the one-step suite evaluates a single step. Table 7 reuses the phase-certificate instances at the operating gains 8 and 12.

suite	n	β	λ	α	draws	steps	used in
exact identities	$\{32, 64, 128, 256\}$	$\{0, 1.1, 1.5\}$	12	0.3	6	5	Lem. 1
phase certificate	$\{200, 400, 800, 1200\}$	1.5	12	0.3	12	5	Tab. 7
one-step law	$\{200, 400, 800, 1200\}$	1.5	8	0.3	24	1	Thm. 2
first reuse	$\{500, 1000, 2000, 4000\}$	1.5	2	0.6	12	2	Thm. 3
signed closure	$\{400, 800, 1200\}$	1.5	12	0.3	5	10	Tab. 2
Stieltjes closure	$\{200, 400, 800, 1200\}$	1.5	12	0.3	6	10	Tab. 4
reaction	$\{500, 1000, 2000\}$	1.5	12	0.3	8	15	Tab. 5
modal calibration	$\{400, 800, 1200\}$	1.5	8	0.3	10	10	Sec. 4
filter audit	800	$\{1.1, 1.3, 1.5\}$	β, 8	0.3	20	30	Tab. 6
recovery, $n = 800$	800	$\{1.1, 1.3, 1.5\}$	tuned	tuned	20	30	Tab. 13
recovery, $n = 2000$	2000	$\{1.1, 1.3, 1.5\}$	tuned	tuned	40	30	Tab. 13

Table 13: Recovery: mean final absolute cosine at $T = 30$ on held-out draws (20 at $n = 800$, 40 at $n = 2000$), hyperparameters tuned on disjoint splits. Naive MF, filtered MF and AMP start from blind seeds; PCA is the leading eigenvector of A; $\mathrm{MF}_{\mathrm{PCA}}$ and $\mathrm{AMP}_{\mathrm{PCA}}$ start from it.

n	β	naive MF	filtered MF	AMP	PCA	$\mathrm{MF}_{\mathrm{PCA}}$	$\mathrm{AMP}_{\mathrm{PCA}}$
800	1.1	0.159	0.304	0.275	0.404	0.391	0.406
800	1.3	0.260	0.570	0.667	0.628	0.675	0.679
800	1.5	0.530	0.760	0.831	0.745	0.830	0.831
2000	1.1	0.118	0.281	0.259	0.372	0.375	0.375
2000	1.3	0.199	0.600	0.695	0.635	0.685	0.699
2000	1.5	0.394	0.752	0.829	0.744	0.825	0.829

G.1 The AMP baseline

With $f_t(z) = \tanh(\gamma z)$ at the frozen gain γ of the cell, the symmetric-Wigner AMP recursion is

$$z_{t+1} = A f_t(z_t) - b_t f_{t-1}(z_{t-1}), \qquad b_t = \frac{\gamma}{n} \sum_i \{1 - f_t(z_{t,i})^2\}, \tag{4}$$

with $f_{-1} \equiv 0$, z_0 the seed, and estimate $f_t(z_t)$ at time t. The coefficient b_t is the empirical mean derivative of f_t, and subtracting $b_t f_{t-1}$ is the Onsager cancellation of the history-dependent conditional mean that the uncorrected update (equation 1) does not perform; this is why AMP is the calibrating baseline rather than merely a competitor. No damping is mixed into equation 4, since a damped variant would need its own memory term. Writing $z_t \approx \mu_t x + \sqrt{\tau_t}\, g$ with g standard Gaussian gives the scalar state evolution

$$\mu_{t+1} = \beta\, \mathbb{E} f_t(\mu_t + \sqrt{\tau_t} Z), \qquad \tau_{t+1} = \mathbb{E} f_t(\mu_t + \sqrt{\tau_t} Z)^2, \qquad Z \sim \mathcal{N}(0, 1), \tag{5}$$

evaluated with the quadrature of Appendix F at the gain each cell actually froze. Since $\tanh$ is odd and 1-Lipschitz, $|\mu_{t+1}| \le \beta\gamma |\mu_t|$, so for $\beta\gamma < 1$ every fixed point has $\mu = 0$; this is the subcritical case of Section 7.

G.2 The signed-closure experiment

Fix an instance and run the true iteration $v_0, \ldots, v_T$ of equation 1. At each time record the realized state in the eigenbasis of W: $m_t = \langle v_t, x \rangle / n$, $\nu_{t,k} = a_{t,k}^2 / n$, $\omega_{t,k} = a_{t,k} b_k / n$ and $\rho_{x,k} = b_k^2 / n$; the quenched measure ρ_x does not evolve.

The three predictors. Each maps a state at time t to a predicted state at $t + 1$ with no fitted coefficient, through the Gaussian linearization with χ_t evaluated at the field moments the state itself implies.

- *Unsigned* carries (m_t, ν_t) and uses field mean βm_t and variance $m_t^2 + \int \mu^2 d\nu_t$. It is the unsigned transport of Section 6, the map that assumes (A5), with ρ_{sc} replaced by the empirical spectral measure of W (mass $1/n$ per atom).
- *Quenched* carries (m_t, ν_t) and the finite-n measure ρ_x: it is the signed step below with ω_t replaced by 0 at every step, so its field mean is $m_t(\beta + d_1)$ and its variance $\int \mu^2 d\nu_t + m_t^2(d_2 - d_1^2)$ with $d_j = \int \mu^j d\rho_x$.
- *Signed* carries (m_t, ν_t, ω_t) with the exact field moments of Proposition 1 and propagates $a' = G_t(\mu)a + H_t(\mu)b + \alpha\epsilon$ with $H_t(\mu) = \alpha\lambda\chi_t\{m_t(\mu + \beta) - \bar{h}_t\}$ and $\bar{h}_t$ the field mean. Squaring gives $\nu_{t+1} = G_t^2\nu_t + 2G_tH_t\omega_t + H_t^2\rho_x + \alpha^2 V_t^{\mathrm{nl}}/n$ and $\omega_{t+1} = G_t\omega_t + H_t\rho_x$, so the cross term that (A5) discards is retained.

Modes and errors. *Teacher forced* feeds each predictor the realized state at t and scores its one-step output against the realized state at $t + 1$, so errors do not accumulate. *Free running* starts from the realized state at $t = 0$ only and propagates its own state for all T steps. The signal error is $|m_{t+1}^{\mathrm{pred}} - m_{t+1}|$ and the spectral error is computed at $\eta \in \{0.15, 0.3\}$; signal errors use the $\eta = 0.15$ rows only, since rows at another η repeat the same prediction. Rows from one trajectory share an instance, so the paired bootstrap behind Table 2 resamples matrices (all 5^5 ordered resamples), not rows.

Environment and provenance. CPU only: Windows 11 (x86-64, 20 threads, 32 GB RAM), Python 3.13.1, NumPy 2.4.4 with OpenBLAS 0.3.31. The twelve suites run in about 13 minutes, five of them for the $n = 2000$ recovery benchmark. Each run archives its configuration, the SHA-256 of every source file, a machine-readable summary and per-trial records in JSONL and CSV, and all twelve archives share one source fingerprint. Every number in the paper, these appendices and the supplement is written from the archives by scripts in the supplementary code, as LaTeX macros or generated tables, and an audit script regenerates them and compares the result with the submitted files. Supplement S2 collects the extended tables.

H Extended related work

We group prior work by the state it resolves and its relation to our diagnostic.

Low-dimensional states through correction. AMP for dense matrices (Bayati & Montanari, 2011; Donoho et al., 2009; Javanmard & Montanari, 2013) and its non-separable and graph-based extensions (Berthier et al., 2020; Gerbelot & Berthier, 2023), TAP constructions (Bolthausen, 2014; Opper & Winther, 2001; Opper et al., 2016; Çakmak & Opper, 2019), and OAMP/VAMP or rotationally invariant AMP (Takeuchi, 2020; Rangan et al., 2019; Fan, 2022; Venkataramanan et al., 2022; Zhong et al., 2024) all use a memory or divergence-free term to cancel selected matrix–iterate correlations and obtain a low-dimensional state evolution. We instead keep the uncorrected map fixed and expose the state that exact conditioning retains.

Uncorrected and general first-order dynamics. Generating-functional analyses predict two-time response kernels under matrix reuse (Sompolinsky & Zippelius, 1981; Mimura, 2011; Bordelon & Pehlevan, 2026). Rigorous pathwise results for spiked-matrix Langevin dynamics and Gaussian-design gradient methods use correlation/response or memory systems (Liang et al., 2023; Gerbelot et al., 2024); Fourier and dynamical-cavity methods cover broader random-matrix and first-order iterations (Jones & Pesenti, 2025; Dandi et al., 2026); and general first-order methods admit an AMP reduction that pins their achievable error (Celentano et al., 2020; Montanari & Venkataramanan, 2021). Our distinction is the explicit first-update spectral law of the fixed mean-field map (Proposition 2), a finite-n phase obstruction at any prescribed gain, an exact first-reuse specialization that links signed spectrum to reaction, and a matched account of which signed and remainder correlations make the unsigned continuation fail after reuse; it is not the existence of rigorous or general trajectory descriptions of uncorrected dynamics.

Mean-field variational inference and reused operators. The update is the damped synchronous solver of the naive mean-field objective, the simplest member of the variational family (Wainwright

& Jordan, 2008; Blei et al., 2017); on spiked models its landscape is tied to the TAP free energy (Fan et al., 2021). Fixed-point layers and implicit message passing (Bai et al., 2019; Gu et al., 2020) reuse one operator without a memory term; our results concern the Wigner instrument and make no claim about those models.

Working beyond the leading eigenvector. For additive rank-one perturbations of a Wigner matrix the leading eigenvector acquires macroscopic overlap only above $\beta = 1$ (Benaych-Georges & Nadakuditi, 2011), the analogue of the spiked-covariance transition (Baik et al., 2005), and the optimal overlap is known (Lelarge & Miolane, 2019), so PCA is both estimator and warm start. Spectrally initialized AMP is analyzed for generalized linear models (Mondelli & Venkataramanan, 2021) and for orthogonally invariant ensembles (Zhong et al., 2024), and in sparse community detection another operator's spectrum replaces the raw leading eigenvector (Massoulié, 2014; Bordenave et al., 2015). Each chooses a better direction to start from, whereas our residual measure is what remains after the signal direction is removed, tracked while the iteration runs. Polynomial acceleration and preconditioning (Pedregosa & Scieur, 2020; Doikov & Rodomanov, 2023) and Chebyshev filtering (Zhou et al., 2006) supply the machinery for testing whether the diagnosis transfers to a feasible operator.

Mean-field recovery on planted models. Belief propagation and mean-field descriptions organize inference transitions (Decelle et al., 2011), while optimization and Langevin dynamics reveal initialization-dependent landscapes (Mannelli et al., 2020). We fix one elementary damped map and use the solvable GOE conditioning geometry to connect its residual spectrum, its first response coefficient and its recovery behavior.