MISSING-DATA-INDUCED PHASE TRANSITIONS IN SPECTRAL PARTIAL LEAST SQUARES

Anonymous authors
Paper under double-blind review

ABSTRACT

Spectral partial least squares (PLS-SVD) estimates the directions shared by two views of the same samples. We analyze it using a rank-one regression model, with entries of both views missing completely at random and filled with zeros. Missing response entries weaken the signal, whereas missing design entries also tilt the recovered direction. The setting is the proportional regime, with Gaussian response noise and a design that is column-orthogonal before masking. For incoherent designs and planted directions, we prove that as the signal diverges, the squared overlap with the truth approaches 1 on the response side but only a ceiling on the design side. This ceiling is strictly below 1 whenever design entries are missing, unless samples of vanishing leverage carry the signal. For this estimator, a stronger signal therefore cannot remove the design-side error. Under equal leverage and an explicit spectral condition on the masked design, which we prove for designs taken from nested Hadamard matrices, we establish a phase transition and derive the recovery curves. Under the same conditions, the critical signal strength is exactly proportional to $\rho_y^{-1/2}$, with ρ_y the fraction of response entries retained. A heuristic replica derivation reproduces the equations behind these curves, and simulations follow them, including on random orthogonal designs outside the verified class. In semi-synthetic experiments on real biological designs with unequal leverage, the predicted ceiling varies across directions, and high-signal recovery tracks it. These results bring the theory of PLS-SVD closer to what is known for PCA under homogeneous missingness and show that the loss of signal-to-noise ratio observed there does not, by itself, describe missing design entries.

1 INTRODUCTION

Partial least squares (PLS) and canonical correlation analysis (CCA) (Hotelling, 1936) are classical linear methods for extracting the low-dimensional structure shared by two views (X, Y) of the same samples, and PLS remains widely used in high dimensions (Rosipal & Krämer, 2006; Boulesteix & Strimmer, 2007). Its spectral form, PLS-SVD, takes the leading singular vectors of the cross-covariance $X^\top Y$ (Léger & Chatelain, 2025). Two-view data are often incomplete, with entries lost to acquisition failures, measurement dropout, and heterogeneous protocols, and the simplest estimator fills them with zeros. Under homogeneous missing-completely-at-random (MCAR) masking (Little & Rubin, 2002; Schafer & Graham, 2002), inverse-retention weighting only rescales the zero-filled cross-covariance, so it gives the same singular vectors. We ask how missing entries change the recovery threshold and accuracy of zero-filled PLS-SVD, and whether a stronger signal can compensate.[1]

For PCA under homogeneous missingness, the loss of entries acts as a reduced signal-to-noise ratio (Ipsen & Hansen, 2019), which a stronger signal restores. For PLS-SVD, the threshold and the overlaps are known only for complete data (Léger & Chatelain, 2025; Mergny & Zdeborová, 2026). With missing entries, the mask on X sets both the signal direction and the noise covariance of $X^\top Y$. The two are dependent, a case that the standard theory of low-rank perturbations (Benaych-Georges & Nadakuditi, 2012) does not cover, so the noise spectrum alone no longer determines recovery.

We show that the answer depends on which view is incomplete. We study a rank-one regression model, $Y_\star = \theta(X_\star u_0)v_0^\top + Z$, with signal strength θ, Gaussian noise Z, and a complete design $X_\star$

[1] Code is available in an anonymous repository; see Appendix M.

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.

that is column-orthogonal. Entries of $X_\star$ and $Y_\star$ are retained independently with probabilities ρ_x and ρ_y, the dimensions grow proportionally, and R_x^2, R_y^2 denote the squared overlaps of the leading singular pair with u_0, v_0.

An exact decomposition of the masked cross-covariance (Proposition 3.1) separates the roles of the two masks. Response masking attenuates the signal, as in PCA. Design masking also tilts the recovered direction. As the signal diverges, R_y^2 tends to 1 while R_x^2 approaches a ceiling, which under equal leverage is $\kappa = \alpha_x\rho_x/(\alpha_x\rho_x + 1 - \rho_x)$ with α_x the limit of the sample-to-feature ratio N/D_x. At $\alpha_x = 2$ and $\rho_x = 0.3$, $\kappa \approx 0.46$ (Figure 1(a)). For this estimator, a stronger signal cannot remove the design-side error. Our contributions are as follows.

- **High-signal ceiling** (Theorem 3.3). For incoherent designs and planted directions, we prove that the design-side overlap approaches a ceiling as the signal diverges, without any spectral assumptions. The ceiling is independent of ρ_y, depends on the direction through the leverage of the samples carrying the signal, and is below 1 when design entries are missing unless that leverage vanishes.
- **Phase transition** (Theorem 3.4). Under equal leverage and an explicit spectral condition on the masked design, we prove that both overlaps vanish below a critical signal strength θ_c and follow explicit curves above it. These curves depend on θ and ρ_y only through $\theta\sqrt{\rho_y}$, so $\theta_c \propto \rho_y^{-1/2}$ exactly. We prove the condition for designs taken from nested Hadamard matrices (Corollary 3.6).
- **Replica derivation** (Appendix H). A heuristic replica calculation reproduces the equations behind these curves via an independent route that makes no further assumptions about the design.

Experiments test these predictions. Simulations follow the predicted transition, recovery curves, and collapse in $\theta\sqrt{\rho_y}$, also on random orthogonal designs outside the verified class (Figure 1). In semi-synthetic tests on real biological designs with unequal leverage, the direction-specific ceiling predicts an average high-signal recovery of 0.02, compared with 0.05 to 0.08 under equal leverage. We cover one signal component, homogeneous MCAR masking, and a column-orthogonal complete design ($D_x \leq N$). The asymmetry is specific to this model, where only the response is noisy.

2 Related work

Spectral phase transitions. The transition of Theorem 3.4 is of BBP type (Baik et al., 2005). Principal-component asymptotics cover isotropic (Bloemendal et al., 2016) and non-isotropic (Ding & Yang, 2021) covariance spikes, and Benaych-Georges & Nadakuditi (2012) give the singular-vector limits for rectangular low-rank perturbations independent of the noise. For fully observed spiked cross-covariance models, Léger & Chatelain (2025) and Mergny & Zdeborová (2026) characterize the phase transition and the overlaps rigorously in the proportional regime, and Bao et al. (2019) treat finite-rank canonical correlations. We add *entry-wise missingness in both views* to a regression model with a noise-free design. The proofs condition on the design mask, which fixes the signal direction and the noise covariance. With noise in both views, this fails, so that case is open (Appendix J).

Missing data. In the closest single-view result, Ipsen & Hansen (2019) show that homogeneous missingness in PCA acts as a reduced signal-to-noise ratio. Under our spectral condition on the masked design, response masking has this interpretation in the limiting recovery curves, while design masking also changes the signal direction and the noise spectrum. Other spectral approaches correct for missingness, in PCA under heterogeneous missingness, debiased Gram-matrix dimension reduction and covariance estimation from partial observations (Zhu et al., 2022; Ling et al., 2021; Lounici, 2014), whereas we analyze the uncorrected estimator. Signal-plus-noise reductions support prediction from incomplete data (Dobriban et al., 2020). Likelihood methods fit probabilistic PCA and PLS by EM (Dempster et al., 1977; Roweis, 1998; Tipping & Bishop, 1999; el Bouhaddani et al., 2018; 2022), and generalized CCA has been adapted to missing data (van de Velden & Takane, 2012).

Random-matrix tools and replica methods. After masking, the design is a deterministic matrix plus independent noise with a variance profile. This is the setting of information-plus-noise models (Dozier & Silverstein, 2007a; Capitaine, 2014), with deterministic equivalents for variance profiles (Hachem et al., 2007; 2013) and deterministic signals (Vallet et al., 2012). Our proofs need spectral limits along the signal directions, with edge and outlier control. The noise edge is based on classical

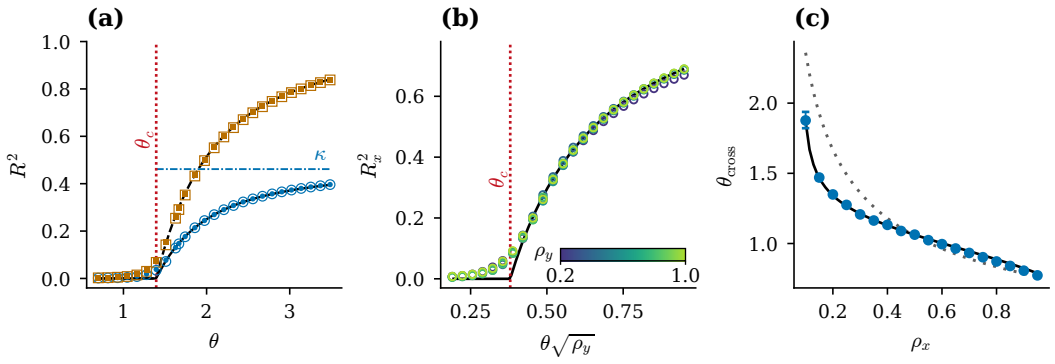


Figure 1: **Recovery under masking in both views.** Black curves are the predictions of Theorem 3.4, and red marks the threshold θ_c in (a) and its rescaled value in (b). **(a)** Squared overlaps R_x^2 (blue) and R_y^2 (orange) against the signal strength θ, on nested Walsh–Hadamard designs (filled markers, the verified class) and on random orthogonal designs (open markers, outside it), with $N = 2048$, $D_x = D_y = 1024$ and retentions $(\rho_x, \rho_y) = (0.3, 0.6)$, for which $\theta_c = 1.39$. The dash-dot line is the ceiling $\kappa = 0.462$ that R_x^2 approaches as θ grows. Markers average 100 trials, and standard errors are below marker size. **(b)** R_x^2 at five response retentions ρ_y falls on one curve against $\theta\sqrt{\rho_y}$ (random orthogonal designs, $N = 2000$, $D_x = 400$, $D_y = 100$, $\rho_x = 0.7$, 60 trials per point). **(c)** The statistic θ_{cross}, the mean of the signal strengths at which the mean R_x^2 curve last crosses three small positive levels, with 95% bootstrap intervals (random orthogonal designs, $N = 2000$, $D_x = D_y = 1000$, $\rho_y = 1$, 100 trials per point). The same statistic is shown for our prediction (black) and for the isotropic reference (gray dotted), which treats masking as a loss of signal-to-noise ratio alone. A crossing of a positive level lies above θ_c (Appendix I.7).

results for sample covariance matrices (Bai & Silverstein, 2010). For the nested Hadamard class, information-plus-noise theory (Dozier & Silverstein, 2007a;b; Bai & Silverstein, 2012) gives the limiting spectrum of the masked design and excludes eigenvalues outside its support. Replica methods derive thresholds and overlaps in spectral inference (Biehl & Mietzner, 1993; Hoyle & Rattray, 2007) and connect to message passing (Bayati & Montanari, 2011; Lesieur et al., 2017). Our heuristic replica derivation follows this line (Section 3.5).

3 Theory of PLS with missing data

3.1 Problem setup

Consider PLS-SVD in a two-view model where both views may have missing entries. Let $X_\star \in \mathbb{R}^{N\times D_x}$ denote the complete X-view, a deterministic design whitened to have orthogonal columns,

$$X_\star^\top X_\star = N I_{D_x}. \tag{1}$$

The normalization forces $D_x \le N$. It also places dimension reduction and whitening before the missingness mechanism, which is a modeling choice. Let $u_0 \in \mathbb{R}^{D_x}$ and $v_0 \in \mathbb{R}^{D_y}$ be fixed unit vectors, and model the complete Y-view as

$$Y_\star = \theta(X_\star u_0)v_0^\top + Z, \quad Z_{ij} \overset{\text{iid}}{\sim} \mathcal{N}(0,1), \quad Z \perp X_\star, \tag{2}$$

where $\theta \ge 0$ is the signal strength. Write $b := X_\star u_0$ for the vector of latent scores, whose norm satisfies $\|b\|^2 = N$ exactly by (1).

Entries are removed by MCAR masks $S_x \in \{0,1\}^{N\times D_x}$ and $S_y \in \{0,1\}^{N\times D_y}$, whose entries are independent Bernoulli variables with fixed retention probabilities $\rho_x, \rho_y \in (0,1]$. The masks are independent of each other and of Z, and we write $\rho := \rho_x\rho_y$ for the joint retention. Missing entries are recorded as zeros, so the observed views are $X := S_x \odot X_\star$ and $Y := S_y \odot Y_\star$. PLS-SVD returns the top singular vectors of the empirical cross-covariance. These vectors are invariant under positive rescaling of the matrix, so we work with the rescaled matrix and its top singular pair,

$$C := \frac{1}{N\sqrt{\rho}} X^\top Y \in \mathbb{R}^{D_x\times D_y}, \qquad (\widehat{u}, \widehat{v}) \in \arg\max_{\|u\|=\|v\|=1} u^\top C v, \tag{3}$$

and measure recovery by the squared overlaps with the planted directions,

$$R_x^2 := (\widehat{u}^\top u_0)^2, \qquad R_y^2 := (\widehat{v}^\top v_0)^2. \tag{4}$$

Asymptotic regime and assumptions. All limits are taken in the proportional regime $N, D_x, D_y \to \infty$, in which the aspect ratios satisfy $N/D_x \to \alpha_x \in [1, \infty)$ and $N/D_y \to \alpha_y \in (0, \infty)$. All convergences hold in probability ($\xrightarrow{P}$). Throughout, we assume the incoherence conditions of Assumption C.1. They require, for a constant C_0, that

$$\|u_0\|_\infty \le C_0 \frac{\log D_x}{\sqrt{D_x}}, \quad \|v_0\|_\infty \le C_0 \frac{\log D_y}{\sqrt{D_y}}, \quad \|X_\star\|_{\max} \le C_0 \log N, \quad \|b\|_\infty \le C_0 \log N,$$

so that no single entry or score dominates the mask fluctuations. The decomposition of Section 3.2 is algebraic, and the high-signal theorem of Section 3.3 needs no assumption beyond incoherence. The finite-signal theorem of Section 3.4 also uses a spectral condition on the masked design (Assumption E.10). We verify this condition for a class of flat orthogonal designs (Corollary 3.6), where **flat** (unimodular) means $|(X_\star)_{ki}| = 1$ for every entry.

3.2 Decomposing the cross-covariance

We first separate the masked signal from the response noise. Averaging the signal over the response mask then writes C as a rank-one spike, product noise, and a remainder. The spike direction is the one PLS can recover at high signal, and the remainder is negligible in operator norm at fixed signal strength.

Proposition 3.1 (Exact decomposition of the masked cross-covariance)**.** *Define the normalized design A, standardized response noise Q, masked Gram matrix G and signal vector a_N by*

$$A := \frac{X}{\sqrt{N\rho_x}}, \quad Q := \frac{S_y \odot Z}{\sqrt{\rho_y}}, \quad G := A^\top A, \quad a_N := \frac{A^\top b}{\sqrt{N}}. \tag{5}$$

For every realization of the data and masks (Appendix B),

$$C = \underbrace{\theta\sqrt{\rho_y}\, a_N v_0^\top}_{\text{Reduced signal}} + \underbrace{\frac{1}{\sqrt{N}} A^\top Q}_{\text{Product noise}} + \mathcal{R}_N, \tag{6}$$

where the reduced signal is the average of the masked signal over the response mask S_y, and the remainder is the fluctuation around this average,

$$\mathcal{R}_N := \frac{\theta}{N\sqrt{\rho}}\, X^\top \mathrm{diag}(b)\big(S_y - \rho_y \mathbf{1}\mathbf{1}^\top\big)\, \mathrm{diag}(v_0). \tag{7}$$

Conditionally on the design mask S_x, the quantities (A, G, a_N) are fixed, and only Q and $\mathcal{R}_N$ remain random. The entries of Q are i.i.d., centered and of unit variance, so the columns of the product noise are independent with conditional covariance G/N given S_x. At $\rho_y = 1$ we have $\mathcal{R}_N = 0$ and $Q = Z$, so C is exactly a rank-one signal plus Gaussian noise with this column covariance.

Write $\bar{C}_N := C - \mathcal{R}_N$ for the reduced matrix. Its signal term is $\theta\sqrt{\rho_y}\, a_N v_0^\top$, which we write as $s_N \widehat{a}_N v_0^\top$. Here $\widehat{a}_N := a_N / \|a_N\|$ is the **corrupted direction** and $s_N := \theta\sqrt{\rho_y}\, \|a_N\|$ is the **effective spike strength**. Response retention scales this whole term, while the design mask changes the vector a_N itself. The response mask still affects the remainder and the law of Q, whose entries are not Gaussian when $\rho_y < 1$.

Under incoherence, $\mathbb{E}\,\|\mathcal{R}_N\|_{\mathrm{op}} \to 0$ at each fixed θ (Lemma B.3). The finite-signal analysis therefore works with $\bar{C}_N$, and Appendix F transfers the conclusions for $\bar{C}_N$ back to C. The bound on $\mathcal{R}_N$ grows in proportion to θ, so the remainder need not vanish when θ grows with N, and the high-signal proof works with C/θ instead.

Two effects of the design mask. Averaging over both masks and the response noise gives $\mathbb{E}[C] = \theta\sqrt{\rho}\, u_0 v_0^\top$, but this mean hides two effects of the realized design mask. The direction $\widehat{a}_N$ generally differs from u_0, and the conditional column covariance of the product noise is G/N rather than I_{D_x}/N. Although $\mathbb{E}[G] = I_{D_x}$ exactly, G need not approach I in operator norm. For a flat design, the expected mean squared distance of the eigenvalues of G from 1 converges to $(1-\rho_x^2)/\alpha_x$ (Lemma C.2 and Appendix E.8), so the eigenvalues stay spread around 1. This spectral spread changes the singular-value bulk of the product noise. We call this the **deformed bulk** and denote its limiting upper edge by z_{edge}. Because $\widehat{a}_N$ and G share S_x, recovery depends on how the signal aligns with the noise eigendirections, as well as on the noise spectrum.

3.3 Geometry and the ceiling

Increasing θ amplifies the part of the reduced signal along u_0 and the part orthogonal to it by the same factor. To quantify their relative energies, write $a_N = \sqrt{\rho_x}\, u_0 + g_N$, where $\sqrt{\rho_x}\, u_0$ is the mean of a_N over the design mask. The energy of the fluctuation g_N is governed by the sample leverages and by how the score energy b_k^2 is spread over samples. The **leverage** of sample k is $\|(X_\star)_{k\cdot}\|^2/N$. For a column-orthogonal design, it is the corresponding diagonal entry of the orthogonal projection onto the design's column space, so it measures the potential influence of that sample in a least-squares fit. The following coupling combines the two.

Definition 3.2 (Score–leverage coupling)**.** Define

$$\tau_N := \sum_{k=1}^{N} \underbrace{\frac{b_k^2}{N}}_{\text{Score weight}} \cdot \underbrace{\frac{\|(X_\star)_{k\cdot}\|^2}{N}}_{\text{Sample leverage}} = \frac{1}{N^2}\sum_{k=1}^{N} b_k^2 \,\|(X_\star)_{k\cdot}\|^2 . \tag{8}$$

The score-energy weights sum to 1, so $\tau_N \in [0,1]$ is the average sample leverage weighted by the signal's energy. Under asymptotically equal leverage, meaning $\|(X_\star)_{k\cdot}\|^2 = D_x(1+o(1))$ uniformly in k, we have $\tau_N \to 1/\alpha_x$.

The geometry calculation (Proposition C.3) gives

$$g_N{}^\top u_0 \xrightarrow{P} 0, \qquad \|g_N\|^2 - (1-\rho_x)\tau_N \xrightarrow{P} 0, \qquad \|a_N\|^2 - \big[\rho_x + (1-\rho_x)\tau_N\big] \xrightarrow{P} 0.$$

Thus ρ_x is the asymptotic energy of a_N along u_0, and $(1-\rho_x)\tau_N$ is the corruption energy. The fraction of the energy that lies along u_0 is therefore $\kappa_{\tau,N}$, with

$$\kappa_{\tau,N} := \frac{\rho_x}{\rho_x + (1-\rho_x)\tau_N}, \qquad \kappa := \frac{\alpha_x\rho_x}{\alpha_x\rho_x + 1 - \rho_x}. \tag{9}$$

Here κ is the value of $\kappa_{\tau,N}$ at $\tau_N = 1/\alpha_x$, which is its limit under asymptotically equal leverage. If $\tau_N \to \tau$, write $\kappa_\tau := \rho_x/[\rho_x + (1-\rho_x)\tau]$ for the limit of $\kappa_{\tau,N}$. Up to $o_P(1)$, $\kappa_{\tau,N}$ is the squared cosine of the angle between the corrupted direction $\widehat{a}_N$ and u_0. For every design, it lies between ρ_x and 1. Concentrating signal energy on high-leverage samples increases τ_N and lowers $\kappa_{\tau,N}$. Orthogonal columns fix only the average leverage, so they do not determine τ_N. Example D.1 gives a column-orthogonal design with unequal block weights in which τ_N converges to a limit above $1/\alpha_x$. Under equal leverage, κ increases from ρ_x at $\alpha_x = 1$ to 1 as $\alpha_x \to \infty$, so more samples per feature reduce the design masking cost.

Theorem 3.3 (High-signal limit)**.** *Suppose Assumption C.1 holds. Then for every deterministic sequence $\theta_N \to \infty$,*

$$R_x^2 - \kappa_{\tau,N} \xrightarrow{P} 0, \qquad R_y^2 \xrightarrow{P} 1,$$

and $R_x^2 \xrightarrow{P} \kappa_\tau$ whenever $\tau_N \to \tau$. The high-signal limit on the X side does not depend on ρ_y.

Origin of the ceiling. The two masks act differently on the signal. The response mask acts independently on each response column. Conditionally on S_x, its fluctuation is the remainder $\mathcal{R}_N$, which is centered with independent columns. Under incoherence, the operator norm of $\mathcal{R}_N$ is $o_P(\theta)$, so at high signal the remainder does not move the leading singular vectors of C. The design mask

enters through the factor $X^\top$ common to all response columns, so its fluctuation enters the signal as the single rank-one term $\theta\sqrt{\rho_y}\, g_N v_0^\top$, which scales with θ exactly as the signal does. This term is nonzero because masking destroys the column orthogonality that gives $X_\star^\top b = N u_0$ in the complete design. The factor $\sqrt{\rho_y}$ scales the useful and corrupted parts of a_N equally, so ρ_y drops out of the high-signal limit of R_x^2, which we call the ceiling.

For example, with equal leverage, $\alpha_x = 2$ and $\rho_x = 0.3$ give $\tau = 0.5$ and $\kappa = 6/13$, so the ceiling is about 0.462 even as $\theta \to \infty$. The theorem is a high-signal limit. It does not by itself assert a bound for every finite θ.

3.4 The finite-signal transition

At finite signal, recovery also depends on whether the spike separates from the deformed bulk. Whether it does depends on how the product noise $E := A^\top Q/\sqrt{N}$ responds along the signal directions at a trial singular value $z > z_{\mathrm{edge}}$, because the corrupted direction and the noise share the design mask. We measure this response using two forms of the noise resolvents, $R_1 := (z^2 I - EE^\top)^{-1}$ and $R_2 := (z^2 I - E^\top E)^{-1}$. The first two are built from R_1, one along the corrupted direction $\widehat{a}_N$ and one between u_0 and $\widehat{a}_N$, and the third is built from R_2 along the response direction v_0. These **spectral responses** have deterministic limits,

$$z\,\widehat{a}_N{}^\top R_1 \widehat{a}_N \xrightarrow{P} m_1(z), \qquad z\,u_0{}^\top R_1 \widehat{a}_N \xrightarrow{P} m_\times(z), \qquad z\,v_0{}^\top R_2 v_0 \xrightarrow{P} m_2(z). \tag{10}$$

Each response is a weighted sum over the singular modes of E. A mode with singular value t enters with weight $z/(z^2 - t^2)$ times its overlaps with the two vectors of the form, so modes with t close to z dominate. The limits depend only on $(\alpha_x, \alpha_y, \rho_x)$, and not on ρ_y or θ. The convergence holds for each fixed z under the condition on the masked Gram matrix G stated next (Proposition F.9).

The spectral condition. Assumption E.10, the spectral condition of the introduction, states that G has a deterministic equivalent, that is, nonrandom limits for its edge and for three forms built from the resolvent $(wI - G)^{-1}$, where the spectral parameter w lies above the spectrum of G. The assumption requires incoherence and, when $\rho_x < 1$, asymptotically equal leverage. The limits solve the **scalar system** of Definition E.2. These explicit scalar equations also give the upper edge w_G of the limiting spectrum of G and the noise edge z_{edge}. The assumption has two parts. **(i)** No eigenvalue of G lies above the edge, in the sense that $\lambda_{\max}(G) \le w_G + o_P(1)$. **(ii)** The trace of $G(wI - G)^{-1}$ divided by N, the quadratic form of $(wI - G)^{-1}$ at $\widehat{a}_N$ and its bilinear form between u_0 and $\widehat{a}_N$ each converge to the value given by the scalar system, uniformly on a complex neighborhood of every compact subset of (w_G, ∞). No limit is assumed for the quadratic form at u_0, because below threshold the proof needs only a finite-N bound on that form, which part (i) provides.

The threshold. The product $\bar{D}_0(z) := m_1(z) m_2(z)$ is positive and decreases to zero above the edge. It is the directional analogue of the D-transform of Benaych-Georges & Nadakuditi (2012). Under equal leverage, s_N^2 converges to the limiting spike energy $s_0^2 := \rho\theta^2/\kappa$, and at $\rho_x = 1$ this holds for any column-orthogonal design. The separation threshold in the model parameter θ is

$$\theta_c = \theta^* \sqrt{\kappa/\rho}, \qquad \theta^* := \bar{D}_0(z_{\mathrm{edge}})^{-1/2}. \tag{11}$$

Here θ^* is the critical value of the effective spike strength s_0. For $\rho_x < 1$ the edge z_{edge} follows from the root of a single scalar equation. Once that root is known, Proposition E.7 evaluates the threshold, the outlier location and the recovery curves explicitly.

Theorem 3.4 (Finite-signal transition under the deterministic equivalent)**.** *In the model of Section 3.1 and under Assumption E.10, let $\theta \neq \theta_c$ be fixed, with θ_c as in* (11)*. Let $\sigma_1(C)$ be the top singular value and $(\widehat{u}, \widehat{v})$ any leading unit singular pair. If $\theta < \theta_c$, then*

$$\sigma_1(C) \xrightarrow{P} z_{\mathrm{edge}}, \qquad R_x^2 \xrightarrow{P} 0, \qquad R_y^2 \xrightarrow{P} 0, \qquad (\widehat{u}^\top \widehat{a}_N)^2 \xrightarrow{P} 0.$$

If $\theta > \theta_c$, there is a unique root $\widehat{z} > z_{\mathrm{edge}}$ of $s_0^2 \bar{D}_0(\widehat{z}) = 1$, and $\sigma_1(C) \xrightarrow{P} \widehat{z}$. The recovery limits are

$$R_x^2 \xrightarrow{P} \frac{-2m_1}{s_0^2 \bar{D}_0'} \Big(\frac{m_\times}{m_1}\Big)^2, \qquad R_y^2 \xrightarrow{P} \frac{-2m_2}{s_0^2 \bar{D}_0'}. \tag{12}$$

All functions are evaluated at $\widehat{z}$, and the prime denotes differentiation with respect to z. The overlap $(\widehat{u}^\top \widehat{a}_N)^2$ converges to $-2m_1/(s_0^2 \bar{D}_0')$.

The X-side formula separates recovery of the corrupted direction from the factor $(m_\times/m_1)^2$ that carries it to u_0. This factor accounts for the dependence between $\widehat{a}_N$ and G, and it generally differs from κ at finite signal. As $\theta \to \infty$ the limits in (12) tend to κ and to 1, which are the equal-leverage values in Theorem 3.3 (Proposition E.7).

At the noise edge, $s_0^2 \bar{D}_0(z_{\mathrm{edge}}) = (\theta/\theta_c)^2$, so the decreasing function $s_0^2 \bar{D}_0$ reaches 1 outside the bulk exactly when $\theta > \theta_c$, and that crossing locates the outlier. Here θ is fixed as $N \to \infty$, so the critical window, in which θ approaches θ_c as N grows, is not covered.

Corollary 3.5 (Unmasked design and response-retention scaling)**.** *At $\rho_x = 1$, Assumption E.10 holds under incoherence, and the limits of Theorem 3.4 have closed forms. With $s_0^2 = \rho_y \theta^2$ and $\alpha_x \alpha_y s_0^4 > 1$,*

$$R_x^2 \xrightarrow{P} \frac{\alpha_x \alpha_y s_0^4 - 1}{\alpha_y s_0^2(\alpha_x s_0^2 + 1)}, \qquad R_y^2 \xrightarrow{P} \frac{\alpha_x \alpha_y s_0^4 - 1}{\alpha_x s_0^2(\alpha_y s_0^2 + 1)}. \tag{13}$$

Both limits are zero below threshold, and $\theta_c = (\alpha_x \alpha_y)^{-1/4}/\sqrt{\rho_y}$. More generally, under Assumption E.10, fixing $(\alpha_x, \alpha_y, \rho_x)$ gives the exact scaling $\theta_c(\rho_y) = \theta_c(1)\rho_y^{-1/2}$.

At fixed design retention, response retention enters the limiting curves only through $\theta\sqrt{\rho_y}$, so reducing ρ_y rescales the signal axis. Design retention changes both the noise spectrum and the corrupted direction, so changing ρ_x has no analogous single rescaling.

Equation (11) separates three effects on the threshold. The factor $\rho^{-1/2}$ compensates for signal attenuation by $\sqrt{\rho}$. Bulk deformation raises θ^* above its isotropic value $(\alpha_x \alpha_y)^{-1/4}$ when $\rho_x < 1$ (Lemma E.8). The factor $\sqrt{\kappa}$ lowers the threshold in the original θ coordinate, since corruption also increases the spike energy $s_0^2 = \rho\theta^2/\kappa$. These effects compete, and the threshold can lie above or below the isotropic reference $(\alpha_x \alpha_y)^{-1/4}\rho^{-1/2}$ of Corollary 3.5 with ρ_y replaced by ρ.

Verified design class. Assumption E.10 is proved for the following class. A sequence of flat column-orthogonal designs is **nested** if, after signed permutations of rows and columns, each $X_\star$ is the $N \times D_x$ corner of one fixed infinite unimodular array. The Sylvester/Walsh–Hadamard array at $N = 2^m$ is the standard example.

Corollary 3.6 (Nested flat orthogonal designs)**.** *Let $(X_\star)$ be a nested sequence of flat orthogonal designs satisfying Assumption C.1. Then the deterministic equivalent of Assumption E.10 holds, and every conclusion of Theorem 3.4 holds unconditionally (Proposition G.1).*

In a flat design every row has squared norm D_x, so every sample has exactly equal leverage. The verification in Appendix G uses flatness for the Gaussian comparison that gives part (ii), and nesting enters through the eigenvalue bound in part (i).

Remark 3.7 (Result status by design class). The finite-signal theorem applies to every design satisfying Assumption E.10. Corollary 3.5 and Corollary 3.6 give the cases in which this assumption is verified. For QR/Haar designs with randomly drawn directions, incoherence holds with probability tending to 1, and Assumption E.10 is supported for them at $\rho_x < 1$ by numerical checks (Appendix E.8) and by the experiments of Section 4. Their finite-signal curves are therefore universality predictions. The curves for biological designs are equal-leverage approximations. We conjecture that the deterministic equivalent also holds for every column-orthogonal design with $\|X_\star\|_{\max} = O(\log N)$ and asymptotically equal leverage under the standing assumptions. Without entrywise incoherence, orthogonal columns and equal leverage are insufficient (Appendix E.8). Appendix A.2 records the full result status.

3.5 Proof outline and replica derivation

The finite-signal proof has four steps. Conditionally on S_x, exact rank-one identities express the outlying singular value and the overlaps of $\bar{C}_N$ through quadratic forms of the noise resolvent (Appendix E). A Gaussian comparison away from the spectrum then replaces these forms by finite-N

transforms of the realized G, and the resulting conditional equations for the outlier and the overlaps do not use Assumption E.10. Next, the edge theorems for sample covariance matrices bound the noise spectrum, and a sign change of the determinant in the rank-one identity locates the outlier. Finally, the deterministic equivalent turns the conditional equations into the scalar prediction (Appendix F). Below threshold, zero overlaps need a bound on the edge spectral subspace, since the absence of a separated root alone does not suffice. Appendix G verifies the deterministic equivalent for the nested flat class by the same comparison on the design side, combined with a result that excludes eigenvalues outside the support (Bai & Silverstein, 2012). The high-signal theorem has a separate proof in Appendix D.

A heuristic replica calculation at a fixed design and a fixed design mask reaches the same conditional equations by an independent route (Appendix H), so Assumption E.10 enters only when the limit in N is taken. The calculation is carried out for the matrix obtained from $\bar{C}_N$ by replacing the product noise with Gaussian noise of the same conditional column covariance G/N. At $\rho_y = 1$ this matrix has the law of $\bar{C}_N$ conditionally on S_x (Corollary B.2). For $\rho_y < 1$ it matches $\bar{C}_N$ only in conditional mean and covariance, so it serves as a comparison model. Replica symmetry, the continuation in the number of replicas, the exchange of limits, and the choice of saddle are not justified rigorously.

4 Experiments

The experiments test the recovery curves and the threshold of Theorem 3.4, on nested Walsh–Hadamard designs in the verified class and on random orthogonal designs outside the verified class. They then test the ceiling of Theorem 3.3 on engineered and biological designs with unequal leverage, and compare zero-filling with imputation.

Setup. In the synthetic experiments we construct a complete column-orthogonal design, generate the response from (2) and apply independent MCAR masks to both views. We take the top singular pair of the matrix C in (3), without re-centering or re-whitening the masked views, and report the squared overlaps of (4). A random orthogonal design, the QR/Haar design of Remark 3.7, is the orthonormal factor in the QR decomposition of a Gaussian matrix (Mezzadri, 2007), rescaled so that $X_\star^\top X_\star = N I_{D_x}$. Appendix I lists all sizes, retentions, signal grids and trial counts.

Transition and recovery curves. In Figure 1(a), both overlaps are small below θ_c and follow the predicted curves above it. The Walsh–Hadamard and random orthogonal designs track the same curves. Over signal strengths above $1.1\,\theta_c$ the mean absolute error of the prediction is below 0.005 for both overlaps on both designs. As θ grows, R_y^2 continues towards 1 while R_x^2 bends towards the ceiling $\kappa = 0.46$ and reaches 0.40 at the largest signal strength shown. In panel (b), on random orthogonal designs, the curves for five response retentions fall on one curve when plotted against $\theta\sqrt{\rho_y}$. This is the scaling of Corollary 3.5, which is a prediction for these designs (Remark 3.7).

Location of the threshold. At finite N the overlaps rise smoothly through θ_c, so the threshold cannot be read off a measured curve. We instead record θ_{cross}, the signal strength at which the mean R_x^2 curve last crosses a small positive level, averaged over the levels 0.05, 0.10 and 0.15, and we apply the same rule to each predicted curve (Appendix I.1). Figure 1(c) shows this statistic on random orthogonal designs, together with the prediction of Theorem 3.4 and the isotropic reference of Section 3.4. The reference treats masking as a loss of signal-to-noise ratio alone, as for PCA. For $\rho_x \le 0.45$ the measured crossings lie up to a quarter below the reference, and every 95% bootstrap interval contains our prediction and excludes the reference. At $\rho_x = 0.5$, where the two predictions nearly meet, the measurements are consistent with both, and Appendix I.7 covers the higher retentions.

Angular recovery. Figure 2 expresses recovery as the angle between $\widehat{u}$ and u_0. Just above θ_c, R_x^2 is positive while this angle can remain large. As θ grows, the angle tends not to zero but to the limiting angle $\arccos\sqrt{\kappa}$, which is the tilt of the recovered direction. For a flat design with $\alpha_x = 2$, the limiting angle is below 30° only if $\rho_x > 0.60$ and below 15° only if $\rho_x > 0.87$, and both retentions decrease as α_x grows.

Direction-specific ceilings. Theorem 3.3 needs incoherence but not equal leverage, and its ceiling $\kappa_{\tau,N}$ then depends on the direction. Whitening equalizes the column covariance but not the sample

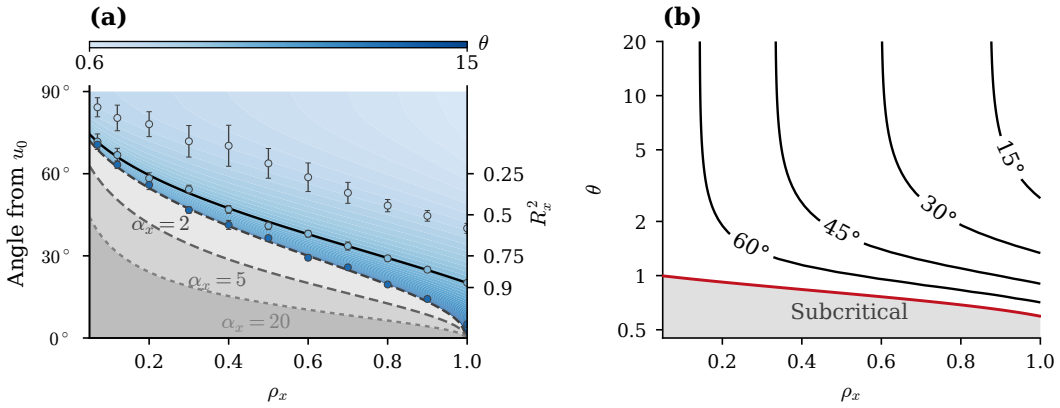


Figure 2: **Angular recovery under design masking. (a)** Angle between $\widehat{u}$ and u_0 against ρ_x for one Walsh–Hadamard design with $N = 1024$, $D_x = 512$, $D_y = 256$ and $\rho_y = 1$, so that $\alpha_x = 2$. The right axis gives the corresponding R_x^2. The colored field is filled between predicted angle curves for θ from 0.6 to 15 and is colored by θ, and the black curve is $\theta = 2$. Points are means over 5 masks with 12 noise draws each at $\theta \in \{1, 2, 8\}$, and bars are one standard deviation. The upper edge of each gray band is the limiting angle $\arccos\sqrt{\kappa}$, with darker bands for larger α_x. **(b)** Predicted angle in the (ρ_x, θ) plane. Contours show equal angles, and red marks θ_c.

leverage, so two directions in the same design can have different couplings and different ceilings. In the engineered designs, rows are scaled by one of two factors before the columns are whitened. The biological designs come from TCGA BRCA, with $N = 873$ samples of RNA-seq and methylation (The Cancer Genome Atlas Network, 2012), and from PBMC Multiome, with $N = 5000$ cells of scRNA-seq and scATAC-seq (10x Genomics, 2021). Both views are reduced to 200 principal components and whitened, and the RNA view is the design. Each of the eight leading singular pairs of the complete-data cross-covariance is planted in turn as (u_0, v_0). The response is then generated from (2), and the whitened design and the response are masked (Appendix I.9). The design is real and the response is synthetic, so the truth is known.

In Figure 8(a) in Appendix I.8, R_x^2 at $30\,\theta_c$ follows $\kappa_{\tau,N}$ on the engineered designs, on the biological directions and on random directions in the biological designs. The equal-leverage value κ is the same for every direction of a design, so it cannot describe this spread. For the biological directions, the mean absolute deviation from $\kappa_{\tau,N}$ is 0.017 on TCGA and 0.018 on PBMC, against 0.078 and 0.046 from κ. Panel (b) tests the finite-signal curves on the engineered designs. These curves assume equal leverage, and their error grows as the leverage becomes more unequal. Agreement with the ceiling at high signal therefore does not justify extending the equal-leverage recovery curve to designs with unequal leverage (Appendix I.8).

Comparison with imputation. The ceiling is proved for the zero-filled estimator, and other estimators may behave differently. In a synthetic comparison at $\rho_x = \rho_y = 0.3$, nuclear-norm imputation tuned on held-out observed entries raises the mean R_x^2 over signal strengths above $1.1\,\theta_c$ from 0.331 to 0.361 (Appendix I.5). At $\rho_x = \rho_y = 0.7$ the two agree within 0.01.

5 Conclusion

We characterized recovery of zero-filled PLS-SVD under MCAR masking in a regression model with a complete column-orthogonal design. The design-side ceiling depends on design retention and score–leverage coupling, while finite-signal recovery also depends on the deformed bulk and its spectral responses. Design-side error can therefore persist as the response signal grows. The recovery curves and ceilings together distinguish informative recovery from accurate direction estimation. Under the deterministic equivalent, verified for nested flat orthogonal designs, response retention enters the limiting curves through $\theta\sqrt{\rho_y}$. The reference $\kappa_{\tau,N}$ can be evaluated for different candidate directions within a complete design (Table 2). Limitations and open problems are discussed in Appendix J.

REFERENCES

10x Genomics. PBMC from a healthy donor - granulocytes removed through cell sorting (10k), 2021. Single Cell Multiome ATAC + Gene Expression dataset, Cell Ranger ARC 2.0.0.

Zhidong Bai and Jack W. Silverstein. No eigenvalues outside the support of the limiting spectral distribution of large dimensional sample covariance matrices. *Annals of Probability*, 26(1):316–345, 1998.

Zhidong Bai and Jack W. Silverstein. *Spectral Analysis of Large Dimensional Random Matrices*. Springer Series in Statistics. Springer, second edition, 2010.

Zhidong Bai and Jack W. Silverstein. No eigenvalues outside the support of the limiting spectral distribution of information-plus-noise type matrices. *Random Matrices: Theory and Applications*, 1(1):1150004, 2012.

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.

Afonso S. Bandeira and Ramon van Handel. Sharp nonasymptotic bounds on the norm of random matrices with independent entries. *Annals of Probability*, 44(4):2479–2506, 2016.

Zhigang Bao, Jiang Hu, Guangming Pan, and Wang Zhou. Canonical correlation coefficients of high-dimensional Gaussian vectors: Finite rank case. *Annals of Statistics*, 47(1):612–640, 2019.

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 singular values and vectors of low rank perturbations of large rectangular random matrices. *Journal of Multivariate Analysis*, 111:120–135, 2012.

Michael Biehl and Andreas Mietzner. Statistical mechanics of unsupervised learning. *Europhysics Letters*, 24(5):421–426, 1993.

Alex Bloemendal, Antti Knowles, Horng-Tzer Yau, and Jun Yin. On the principal components of sample covariance matrices. *Probability Theory and Related Fields*, 164(1/2):459–552, 2016.

Anne-Laure Boulesteix and Korbinian Strimmer. Partial least squares: a versatile tool for the analysis of high-dimensional genomic data. *Briefings in Bioinformatics*, 8(1):32–44, 2007.

Mireille Capitaine. Exact separation phenomenon for the eigenvalues of large information-plus-noise type matrices. Application to spiked models. *Indiana University Mathematics Journal*, 63(6): 1875–1910, 2014.

Arthur P. Dempster, Nan M. Laird, and Donald B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1): 1–38, 1977.

Xiucai Ding and Fan Yang. Spiked separable covariance matrices and principal components. *Annals of Statistics*, 49(2):1113–1138, 2021.

Edgar Dobriban, William Leeb, and Amit Singer. Optimal prediction in the linearly transformed spiked model. *The Annals of Statistics*, 48(1):491–513, 2020. doi: 10.1214/19-AOS1819. URL `https://arxiv.org/abs/1709.03393`.

R. Brent Dozier and Jack W. Silverstein. On the empirical distribution of eigenvalues of large dimensional information-plus-noise-type matrices. *Journal of Multivariate Analysis*, 98(4):678–694, 2007a.

R. Brent Dozier and Jack W. Silverstein. Analysis of the limiting spectral distribution of large dimensional information-plus-noise type matrices. *Journal of Multivariate Analysis*, 98(6):1099–1122, 2007b.

Said el Bouhaddani, Hae-Won Uh, Caroline Hayward, Geurt Jongbloed, and Jeanine Houwing-Duistermaat. Probabilistic partial least squares model: Identifiability, estimation and application. *Journal of Multivariate Analysis*, 167:331–346, 2018.

Said el Bouhaddani, Hae-Won Uh, Geurt Jongbloed, and Jeanine Houwing-Duistermaat. Statistical integration of heterogeneous omics data: Probabilistic two-way partial least squares (PO2PLS). *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 71(5):1451–1470, 2022.

Mary J. Goldman, Brian Craft, Mim Hastie, Kristupas Repečka, Fran McDade, Akhil Kamath, Ayan Banerjee, Yunhai Luo, Dave Rogers, Angela N. Brooks, Jingchun Zhu, and David Haussler. Visualizing and interpreting cancer genomics data via the Xena platform. *Nature Biotechnology*, 38:675–678, 2020.

Walid Hachem, Philippe Loubaton, and Jamal Najim. Deterministic equivalents for certain functionals of large random matrices. *The Annals of Applied Probability*, 17(3):875–930, 2007.

Walid Hachem, Philippe Loubaton, Jamal Najim, and Pascal Vallet. On bilinear forms based on the resolvent of large random matrices. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 49(1):36–63, 2013.

Harold Hotelling. Relations between two sets of variates. *Biometrika*, 28(3/4):321–377, 1936.

David C. Hoyle and Magnus Rattray. Statistical mechanics of learning multiple orthogonal signals: asymptotic theory and fluctuation effects. *Physical Review E*, 75(1):016101, 2007.

Niels Ipsen and Lars Kai Hansen. Phase transition in PCA with missing data: Reduced signal-to-noise ratio, not sample size! In *Proceedings of the 36th International Conference on Machine Learning*, volume 97, pp. 2951–2960. PMLR, 09–15 Jun 2019.

Victor Léger and Florent Chatelain. High-dimensional partial least squares: Spectral analysis and fundamental limitations. *arXiv*, 2025. URL `https://arxiv.org/abs/2512.15684`. Preprint.

Thibault Lesieur, Florent Krzakala, and Lenka Zdeborová. Constrained low-rank matrix estimation: Phase transitions, approximate message passing and applications. *Journal of Statistical Mechanics: Theory and Experiment*, 2017(7):073403, 2017.

Yurong Ling, Zijing Liu, and Jing-Hao Xue. Dimension reduction for data with heterogeneous missingness. In *Proceedings of the Thirty-Seventh Conference on Uncertainty in Artificial Intelligence*, volume 161 of *Proceedings of Machine Learning Research*, pp. 1310–1320, 2021.

Roderick J. A. Little and Donald B. Rubin. *Statistical Analysis with Missing Data*. Wiley Series in Probability and Statistics. Wiley-Interscience, second edition, 2002.

Karim Lounici. High-dimensional covariance matrix estimation with missing observations. *Bernoulli*, 20(3):1029–1058, 2014.

Rahul Mazumder, Trevor Hastie, and Robert Tibshirani. Spectral regularization algorithms for learning large incomplete matrices. *Journal of Machine Learning Research*, 11:2287–2322, 2010.

Pierre Mergny and Lenka Zdeborová. Spectral thresholds in correlated spiked models and fundamental limits of partial least squares. In *Proceedings of the 29th International Conference on Artificial Intelligence and Statistics*, volume 300, pp. 3925–3933. PMLR, 2026. URL `https://proceedings.mlr.press/v300/mergny26a.html`.

Marc Mézard, Giorgio Parisi, and Miguel Angel Virasoro. *Spin Glass Theory and Beyond: An Introduction to the Replica Method and Its Applications*, volume 9 of *World Scientific Lecture Notes in Physics*. World Scientific, 1987.

Francesco Mezzadri. How to generate random matrices from the classical compact groups. *Notices of the American Mathematical Society*, 54(5):592–604, 2007.

Roman Rosipal and Nicole Krämer. Overview and recent advances in partial least squares. In Craig Saunders, Marko Grobelnik, Steve Gunn, and John Shawe-Taylor (eds.), *Subspace, Latent Structure and Feature Selection*, volume 3940 of *Lecture Notes in Computer Science*, pp. 34–51. Springer, Berlin, Heidelberg, 2006.

Sam Roweis. EM algorithms for PCA and SPCA. In *Advances in Neural Information Processing Systems*, volume 10, pp. 626–632, 1998.

Joseph L. Schafer and John W. Graham. Missing data: our view of the state of the art. *Psychological Methods*, 7(2):147–177, 2002.

The Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature*, 490(7418):61–70, 2012.

Michael E. Tipping and Christopher M. Bishop. Probabilistic principal component analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(3):611–622, 1999.

Olga Troyanskaya, Michael Cantor, Gavin Sherlock, Pat Brown, Trevor Hastie, Robert Tibshirani, David Botstein, and Russ B. Altman. Missing value estimation methods for DNA microarrays. *Bioinformatics*, 17(6):520–525, 2001.

Pascal Vallet, Philippe Loubaton, and Xavier Mestre. Improved subspace estimation for multivariate observations of high dimension: The deterministic signals case. *IEEE Transactions on Information Theory*, 58(2):1043–1068, 2012.

Michel van de Velden and Yoshio Takane. Generalized canonical correlation analysis with missing values. *Computational Statistics*, 27(3):551–571, 2012.

Roman Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science*, volume 47 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, 2018.

Per-Åke Wedin. Perturbation bounds in connection with singular value decomposition. *BIT Numerical Mathematics*, 12:99–111, 1972.

Ziwei Zhu, Tengyao Wang, and Richard J. Samworth. High-dimensional principal component analysis with heterogeneous missingness. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 84(5):2000–2031, 2022. doi: 10.1111/rssb.12550.

A NOTATION AND PROOF GUIDE

A.1 NOTATION AND CONVENTIONS

Stars mark complete-data objects ($X_\star$, $Y_\star$). Hats identify the estimated leading singular vectors ($\widehat{u}$, $\widehat{v}$), the corrupted direction ($\widehat{a}_N$), and the supercritical outlier location ($\widehat{z}$). Bars mark conditional or reduced objects ($\bar{C}_N$, $\bar{m}_1$, $\bar{D}$), and a subscript N marks a finite-N quantity. Its deterministic limit, when it exists, drops the subscript ($\tau_N \to \tau$) or replaces it by 0 ($s_N \to s_0$, $\psi_N \to \psi_0$). The bare ψ, Φ, $\Phi^\times$ and T are ψ_0, Φ_0, $\Phi_0^\times$ and T_0 written in the branch coordinate ω (Definition E.2, Lemma E.4). The exact aspect ratios are $\alpha_{x,N} = N/D_x$ and $\alpha_{y,N} = N/D_y$; the unsubscripted α_x, α_y are their limits. Design-side objects are indexed by D_x and response-side objects by D_y. An upright subscript G marks Gaussian objects (W_G, C_G, $g_{\mathrm{G},N}$), and an italic subscript G marks the fold of the branch and the top of the Gram support (ω_G, w_G).

Matrices are generally uppercase italic, while vectors and scalars are lowercase italic and distinguished by context. Calligraphic letters denote structured operators, events, contours, or residuals. Bold uppercase letters are reserved for Hermitized block objects, apart from $\mathbf{1}$ for the all-ones vector.

Model and finite-N random quantities	
$X_\star, Y_\star, u_0, v_0, b, \theta$	complete views, planted directions, latent scores $b = X_\star u_0$, signal strength; (1), (2)
$S_x, S_y, \rho_x, \rho_y, \rho$	masks, retentions, joint retention $\rho = \rho_x \rho_y$
$\alpha_{x,N}, \alpha_{y,N}$	exact finite-N aspect ratios $N/D_x, N/D_y$
$X, Y, C, \widehat{u}, \widehat{v}, R_x^2, R_y^2$	masked views, cross-covariance, leading pair, squared overlaps; (3), (4)
$A, Q, G, a_N, \widehat{a}_N, s_N$	normalized design and noise, masked Gram matrix, signal vector, corrupted direction, effective spike strength; (5), Section 3.2
W_G	standard Gaussian matrix of the Gaussian ensemble $C_\mathrm{G} = s_N \widehat{a}_N v_0^\top + G^{1/2} W_\mathrm{G}/\sqrt{N}$; (B.4), (F.3), Appendix H
$p_N, \Delta_N, \mathcal{R}_N$	projected signal vector, signal residual, remainder; (B.1), (B.3), (7)
$E, g_N, \widehat{p}, D_N$	product noise $A^\top Q/\sqrt{N}$, fluctuation vector $a_N - \sqrt{\rho_x}\, u_0$, unit vector $p_N/\lVert p_N \rVert$, normalized residual Δ_N/θ; Section 3.4, Proposition C.3, Proposition C.5
$\psi_N, \Phi_N, \Phi_N^\times$	conditional resolvent functionals; (E.1), Appendix E.1
$\tau_N, \kappa_{\tau,N}$	score–leverage coupling and the level it implies; (8), (9)
Deterministic limits, design side	
$\alpha_x, \alpha_y, \tau, \kappa_\tau, \kappa, n_a$	aspect ratios, limiting coupling, ceilings, squared spike norm $n_a = \rho_x/\kappa$; Definition 3.2, (9), Definition E.2
g, h, e_1, e_2, ξ	scalar system, $\xi = \sqrt{w}$; (E.4)
ω, W, Λ	branch coordinate $\omega = e_1/d_1$, Gram coordinate map $w = W(\omega)$, Schur determinant $\Lambda = d_1 d_2 - \rho_x$ with $d_i = \xi - e_i$; Appendix E.3
ω_G, w_G	fold of the branch and top of the Gram support; (E.13)
$\omega_R, w_R, z_\mathrm{edge}$	response stationary point and the top edge of the deformed bulk; (E.15)
θ^*, θ_c	detection level and threshold; (11)
$\omega_\theta, \widehat{z}$	supercritical root coordinate and outlier location $\widehat{z} = \sqrt{T(\omega_\theta)}$; Proposition E.7
Response quantities and their limits	
z, δ, K	response coordinate, fixed point, and ratio $K = w/z$; (E.2)
$\bar{m}_1, \bar{m}_2, \bar{m}_1^\times, \bar{D}$	conditional secular transforms; (E.3)
$m_1, m_2, m_\times, \bar{D}_0$	limiting transforms and their product; (E.32), Appendix E.7
$T_N, w_{R,N}, z_{\mathrm{edge},N}, \mathbf{R}_N^\mathrm{det}$	finite-N response map, its minimizer, the response edge, and the deterministic-equivalent resolvent; Appendix E.5
$\widehat{z}_N, \mathfrak{t}_N$	finite-N conditional secular root and design tolerance sequence; Theorem F.10, Lemma F.1
ς	edge-window half-width; Appendix F.7
Replica quantities (Appendix H)	
β, n, u^a, v^a	inverse temperature, replica number, replicated pair
q_v, r_y, χ, ζ	overlap order parameters, response susceptibility, and spherical multiplier

A.2 Proof roadmap and result status

Appendix B gives the exact finite-N reduction, and Appendix C describes the corrupted direction and effective spike strength. Appendix D uses this geometry to prove the high-signal limit under the standing incoherence assumptions, without the deterministic equivalent. For finite signal, Appendix E.1 and Appendix E.6 provide conditional transforms and exact fixed-noise identities under their stated spectral conditions. The scalar comparison in Appendices E.2–E.5 gives the candidate curves of Appendix E.7.

Appendix F uses Gaussian comparison at fixed height, exterior continuation, spike geometry, and residual control to transfer these curves to the estimator under Assumption E.10, with separate arguments above and below threshold. See Appendices F.2–F.7 for the response estimates and the inputs they take from Assumption E.10. Appendix G verifies Assumption E.10 for the stated nested flat class. The replica derivation in Appendix H reproduces the conditional equations for a covariance-matched Gaussian ensemble; it does not establish the deterministic equivalent.

Table 1: Result status by setting.

Setting	**Status**
General column-orthogonal designs under Assumption C.1	**High-signal theorem.** Theorem 3.3; $\tau_N \to \tau$ is needed only for the limiting form.
Every design satisfying Assumption E.10	**Finite-signal theorem.** Theorem 3.4 gives the threshold and curves under the standing hypotheses.
Nested flat orthogonal class of Corollary 3.6	**Unconditional theorem.** Appendix G verifies Assumption E.10 for the stated nested class, including signed row and column permutations of nested Sylvester/Walsh–Hadamard corners under Assumption C.1.
$\rho_x = 1$, any column-orthogonal design under Assumption C.1	**Exact deterministic equivalent and closed forms.** Assumption E.10 holds exactly, and Proposition E.9(i) gives the closed forms.
QR/Haar and other designs without a verified deterministic equivalent	**Universality prediction.** Scope and evidence are recorded in Remark 3.7 and Appendix E.8.
Biological designs in the semi-synthetic experiments	**Equal-leverage approximation.** The theorem is not asserted for these designs.
Covariance-matched Gaussian replica ensemble, conditional on $(X_\star, S_x)$ under (R1)–(R5)	**Replica derivation.** It reproduces the conditional equations, not Assumption E.10; its conditional law equals the reduced one only at $\rho_y = 1$.

B EXACT CONDITIONAL REDUCTION

Proposition 3.1 is proved here by direct computation from (2) and (3). The reduction and its single-view specialization are exact at every finite N and use (1) only where stated. We then bound the residual at fixed signal strength under incoherence.

Proposition B.1 (Dual masking). *Let $q := v_0 \odot v_0$ and define the projected signal vector*

$$p_N := \frac{X^\top (b \odot S_y q)}{N\sqrt{\rho}}. \tag{B.1}$$

For $\rho_y \le 1$ and every realization of the data and masks,

$$C = \underbrace{\theta p_N v_0^\top}_{\text{Projected signal}} + \underbrace{\Delta_N}_{\text{Signal residual}} + \underbrace{\frac{1}{\sqrt{N}} A^\top Q}_{\text{Product noise}}, \tag{B.2}$$

where the signal residual is

$$\Delta_N := \frac{\theta}{N\sqrt{\rho}} X^\top \operatorname{diag}(b) \left[S_y - (S_y q)\mathbf{1}^\top \right] \operatorname{diag}(v_0), \qquad \Delta_N v_0 = 0. \tag{B.3}$$

The noise term obeys the exact pathwise identity

$$E \;=\; \frac{1}{\sqrt{N}} A^\top Q, \qquad A = \frac{X}{\sqrt{N\rho_x}}, \qquad Q = \frac{S_y \odot Z}{\sqrt{\rho_y}},$$

with A and Q as in (5)*. The entries of Q are i.i.d., centered, of unit variance, with $\mathbb{E}\,|Q_{kj}|^m = \rho_y^{1-m/2}\,\mathbb{E}\,|Z_{11}|^m$ for every $m \ge 1$. Conditionally on (S_x, S_y) the columns of E are independent centered Gaussians of covariance $\widetilde{G}_j/N$, where $\widetilde{G}_j := X^\top \operatorname{diag}\big((S_y)_{\cdot j}\big) X/(N\rho)$, and $\mathbb{E}[\widetilde{G}_j \mid S_x] = G$ for every j. The splitting is unique: $\Delta_N v_0 = 0$ pathwise, so $\theta p_N v_0^\top$ is the part of the conditional mean acting along v_0 and Δ_N the part acting on its orthogonal complement.*

Proof. Substituting $Y_\star = \theta b v_0^\top + Z$ into $Y = S_y \odot Y_\star$ and using the entrywise identity $S_y \odot (b v_0^\top) = \operatorname{diag}(b)\, S_y \operatorname{diag}(v_0)$, the cross-covariance (3) splits pathwise as $C = C_{\mathrm{sig}} + E$ with

$$C_{\mathrm{sig}} := \frac{\theta}{N\sqrt{\rho}} X^\top \operatorname{diag}(b)\, S_y \operatorname{diag}(v_0), \qquad E := \frac{X^\top (S_y \odot Z)}{N\sqrt{\rho}}.$$

With A and Q as in (5), $A^\top Q/\sqrt{N} = X^\top (S_y \odot Z)/(N\sqrt{\rho_x \rho_y}) = E$, the claimed identity. The moment identity for Q follows from $|Q_{kj}|^m = (S_y)_{kj}\,|Z_{kj}|^m\,\rho_y^{-m/2}$ for a $0/1$ mask. Conditionally on (S_x, S_y) only Z is random. Writing $D_j := \operatorname{diag}\big((S_y)_{\cdot j}\big)$, the j-th column of E is $X^\top D_j Z_{\cdot j}/(N\sqrt{\rho})$, a fixed linear image of $Z_{\cdot j} \sim \mathcal{N}(0, I_N)$: centered Gaussian with covariance $X^\top D_j^2 X/(N^2\rho) = \widetilde{G}_j/N$, since $D_j^2 = D_j$ for a $0/1$ mask, and independent across j because $E_{\cdot j}$ depends on $Z_{\cdot j}$ alone. Averaging over the response mask at fixed S_x, $\mathbb{E}D_j = \rho_y I_N$ gives $\mathbb{E}[\widetilde{G}_j \mid S_x] = \rho_y\, X^\top X/(N\rho) = G$ for every j.

The signal is the conditional mean $C_{\mathrm{sig}} = \mathbb{E}[C \mid S_x, S_y]$. Its action along the planted response direction is $C_{\mathrm{sig}} v_0 = \theta p_N$, since $\operatorname{diag}(v_0) v_0 = q$ and $\operatorname{diag}(b) S_y q = b \odot S_y q$. Projecting onto this direction and its orthogonal complement therefore gives

$$C_{\mathrm{sig}} = C_{\mathrm{sig}} v_0 v_0^\top + C_{\mathrm{sig}}(I - v_0 v_0^\top) = \theta p_N v_0^\top + \Delta_N.$$

Expanding the second projection recovers (B.3), because $S_y \operatorname{diag}(v_0)(I - v_0 v_0^\top) = [S_y - (S_y q)\mathbf{1}^\top] \operatorname{diag}(v_0)$. It also gives $\Delta_N v_0 = 0$ directly. These two projections uniquely determine the signal matrices, including when $\theta = 0$. □

Averaging the signal over the response mask gives $\mathbb{E}[S_y q] = \rho_y \mathbf{1}$, hence the exact identities

$$\mathbb{E}[p_N \mid S_x] = \sqrt{\rho_y}\, a_N, \qquad \mathbb{E}[C \mid S_x] = \theta\sqrt{\rho_y}\, a_N v_0^\top.$$

Since $\mathbb{E}[A] = \sqrt{\rho_x/N}\, X_\star$, whitening gives $\mathbb{E}[a_N] = \sqrt{\rho_x}\, u_0$ and therefore $\mathbb{E}[C] = \theta\sqrt{\rho}\, u_0 v_0^\top$. These expectations average the masks and noise with the complete design fixed; the decompositions (B.2) and (6) remain pathwise identities.

Corollary B.2 (Single-view specialization)**.** *At $\rho_y = 1$, $S_y \equiv \mathbf{1}$, hence $\Delta_N = 0$, $p_N = a_N$, and $\widetilde{G}_j = G$ for every j. Conditionally on S_x,*

$$C \overset{d}{=} \theta\, a_N v_0^\top + \frac{1}{\sqrt{N}}\, G^{1/2} W_{\mathrm{G}}, \qquad (W_{\mathrm{G}})_{ij} \overset{\mathrm{iid}}{\sim} \mathcal{N}(0,1). \tag{B.4}$$

Proof. At $S_y \equiv \mathbf{1}$, $(S_y q)_k = \|v_0\|^2 = 1$, so $(S_y q)\mathbf{1}^\top = S_y$ and the bracket in (B.3) vanishes, giving $\Delta_N = 0$. With $\rho = \rho_x$, each remaining pathwise claim follows by substitution in (5) and (B.1). For (B.4), $S_y \equiv \mathbf{1}$ is deterministic, so conditioning on S_x coincides with conditioning on (S_x, S_y), and Proposition B.1 makes the columns of $E = A^\top Q/\sqrt{N}$ independent centered Gaussians of covariance $\widetilde{G}_j/N = G/N$. The columns of $G^{1/2}W_{\mathrm{G}}/\sqrt{N}$ have the same joint law, a centered Gaussian vector being determined by its covariance, invertible or not. Adding the S_x-measurable signal term $\theta\, a_N v_0^\top$ gives (B.4). □

The residual $\mathcal{R}_N$ of (6) collects the two mask-fluctuation terms of (B.2),

$$\mathcal{R}_N = \theta\big(p_N - \sqrt{\rho_y}\, a_N\big) v_0^\top + \Delta_N, \tag{B.5}$$

since $\theta p_N v_0^\top + \Delta_N = C_{\mathrm{sig}}$ and $\mathcal{R}_N = C_{\mathrm{sig}} - \theta\sqrt{\rho_y}\, a_N v_0^\top$. The next lemma bounds it in expectation.

Lemma B.3 (L^1 residual bound)**.** *Under Assumption C.1, at each fixed θ, $\mathbb{E}\,\|\mathcal{R}_N\|_{\mathrm{op}} \to 0$.*

Proof of Lemma B.3. The residual $\mathcal{R}_N$ is the fluctuation of the signal about its mean given S_x. Subtracting $\theta\sqrt{\rho_y}\, a_N v_0^\top$ from C_{sig} gives the exact identity

$$\mathcal{R}_N = \frac{\theta}{N\sqrt{\rho}}\, X^\top \operatorname{diag}(b)(S_y - \rho_y \mathbf{1}\mathbf{1}^\top) \operatorname{diag}(v_0).$$

With the complete design and directions fixed, X depends only on S_x and is independent of S_y. Moreover, $\mathbb{E}\,\|X\|_{\mathrm{op}} = O(\sqrt{N})$ by Lemma C.4. The centered response mask has independent entries bounded by 1, so the independent-entry norm bound (Bandeira & van Handel, 2016) gives $\mathbb{E}\,\big\|S_y - \rho_y \mathbf{1}\mathbf{1}^\top\big\|_{\mathrm{op}} = O(\sqrt{N} + \sqrt{D_y}) = O(\sqrt{N})$. Submultiplicativity and independence now give

$$\begin{aligned}\mathbb{E}\,\|\mathcal{R}_N\|_{\mathrm{op}} &\le \frac{\theta\ \|b\|_\infty \|v_0\|_\infty}{N\sqrt{\rho}}\, \mathbb{E}\,\|X\|_{\mathrm{op}}\ \mathbb{E}\,\big\|S_y - \rho_y \mathbf{1}\mathbf{1}^\top\big\|_{\mathrm{op}} \\ &\le C\theta\ \|b\|_\infty \|v_0\|_\infty .\end{aligned}$$

Here C is independent of N and θ, with the retentions and limiting aspect ratios fixed. Thus Assumption C.1(ii) and (iv) imply

$$\mathbb{E}\,\|\mathcal{R}_N\|_{\mathrm{op}} = O\Big(\theta\, \frac{\log N \log D_y}{\sqrt{D_y}}\Big) \longrightarrow 0$$

at each fixed θ. At $\theta = 0$, the residual is identically zero. □

The stated convergence concerns fixed θ. The bound is linear in θ, which also controls the normalized residual used in Appendix D. Weyl's inequality bounds the change in every singular value by $\|\mathcal{R}_N\|_{\mathrm{op}}$, and Wedin's theorem (Wedin, 1972) controls the corresponding singular pair when the outlier is simple and separated from the remaining spectrum. Dependence between $\mathcal{R}_N$ and $A^\top Q$ is irrelevant to this deterministic perturbation step. The free-energy form of the same comparison, used by the replica derivation, is stated at the start of Appendix H.

C GEOMETRY OF THE CORRUPTED DIRECTION

We set out the finite-N geometry of the corrupted direction: the incoherence assumption (Assumption C.1), the exact moments of a_N (Proposition C.3), and the operator-norm bounds (Lemma C.4, Proposition C.5) used throughout.

Assumption C.1 (Polylogarithmic incoherence)**.** There is a constant C_0, independent of N, such that

$$\begin{aligned} &\text{(i)} \quad \|u_0\|_\infty \le C_0 \frac{\log D_x}{\sqrt{D_x}}, && \text{(iii)} \quad \|X_\star\|_{\max} \le C_0 \log N, \\ &\text{(ii)} \quad \|v_0\|_\infty \le C_0 \frac{\log D_y}{\sqrt{D_y}}, && \text{(iv)} \quad \|b\|_\infty \le C_0 \log N. \end{aligned} \tag{C.1}$$

The four conditions prevent any one feature coordinate, response coordinate, design entry, or sample score from dominating the mask fluctuations. For (u_0, v_0) drawn independently of $X_\star$ and of each other, uniformly on their unit spheres or as normalized i.i.d. sign vectors, and $X_\star$ Haar or Walsh–Hadamard, a deterministic C_0 exists for which all four conditions hold jointly with probability tending to 1. Every later bound depends on the data only through C_0, so the conclusions apply whenever these conditions hold. The sample-score condition **(iv)** is separate because **(i)** and **(iii)** do not control $b = X_\star u_0$.

Lemma C.2 (Centering of the masked Gram matrix)**.** $\mathbb{E}[G] = I_{D_x}$ *exactly, for every* N*, and for a flat design* $\mathbb{E}G_{ii'}^2 = (1-\rho_x^2)/N$ *exactly for every* $i \neq i'$*.*

Proof of Lemma C.2. Entrywise $X_{ki} = (S_x)_{ki}(X_\star)_{ki}$, so

$$G_{ii'} = \frac{1}{N\rho_x} \sum_{k=1}^{N} (S_x)_{ki}(S_x)_{ki'}(X_\star)_{ki}(X_\star)_{ki'},$$

and only the mask is random. On the diagonal the mask entries are 0 or 1, so $(S_x)_{ki}^2 = (S_x)_{ki}$ and $\mathbb{E}[G_{ii}] = N^{-1}\sum_k (X_\star)_{ki}^2 = 1$, the i-th diagonal entry of (1). Off the diagonal the two mask entries lie in distinct columns and are independent, so $\mathbb{E}[G_{ii'}] = \rho_x N^{-1}(X_\star^\top X_\star)_{ii'} = 0$, again by (1). For the second moment write $c_k := (X_\star)_{ki}(X_\star)_{ki'}$. Distinct rows are independent and $(S_x)_{ki}^2(S_x)_{ki'}^2 = (S_x)_{ki}(S_x)_{ki'}$, so

$$\mathbb{E}G_{ii'}^2 = \frac{1}{N^2\rho_x^2}\Big(\rho_x^4 \sum_{k\neq l} c_k c_l + \rho_x^2 \sum_k c_k^2\Big).$$

A flat design has $c_k^2 = 1$, hence $\sum_k c_k^2 = N$, while $\sum_k c_k = (X_\star^\top X_\star)_{ii'} = 0$ gives $\sum_{k\neq l} c_k c_l = -N$, and the two terms combine to the stated value. □

Proposition C.3 (Geometry of the corrupted direction)**.** *At every finite* N*,* $\mathbb{E}[a_N] = \sqrt{\rho_x}\, u_0$*, the coordinates of* a_N *are independent, and* $\mathbb{E}\,\|g_N\|^2 = (1-\rho_x)\tau_N$*. Under Assumption C.1,*

$$g_N{}^\top u_0 \xrightarrow{P} 0, \qquad \|g_N\|^2 - (1-\rho_x)\tau_N \xrightarrow{P} 0, \qquad \big\|p_N - \sqrt{\rho_y}\, a_N\big\| \xrightarrow{P} 0,$$

so $p_N{}^\top u_0 \xrightarrow{P} \sqrt{\rho}$ *and* $\|p_N\|^2 - (\rho + \rho_y(1-\rho_x)\tau_N) \xrightarrow{P} 0$*. Let* $\widehat{p} := p_N/\|p_N\|$ *when* $p_N \neq 0$*, with any fixed unit vector used on the event* $p_N = 0$*, whose probability tends to zero. Then* $(\widehat{p}^\top u_0)^2 - \kappa_{\tau,N} \xrightarrow{P} 0$*. If moreover* $\tau_N \to \tau$*, the limits are* $(1-\rho_x)\tau$*,* $\rho + \rho_y(1-\rho_x)\tau$ *and* κ_τ*.*

Proof. The exact moments follow from the Bernoulli mask distribution. The coordinate

$$(a_N)_i = \frac{1}{N\sqrt{\rho_x}} \sum_{k=1}^{N} (S_x)_{ki}(X_\star)_{ki}\, b_k$$

depends on the i-th mask column alone, and distinct columns of S_x are disjoint collections of independent entries, so the coordinates are independent. Taking expectations, $\mathbb{E}[(S_x)_{ki}] = \rho_x$ and $X_\star^\top b = X_\star^\top X_\star u_0 = N u_0$ give $\mathbb{E}[a_N] = \sqrt{\rho_x}\, u_0$, so $g_N = a_N - \sqrt{\rho_x}\, u_0$ is centered with independent coordinates. Summing the mask variances $\rho_x(1-\rho_x)$ down column i yields

$$\sigma_i^2 := \mathrm{Var}\big((a_N)_i\big) = \frac{1-\rho_x}{N^2} \sum_k (X_\star)_{ki}^2 b_k^2,$$

and summing over i, with $\sum_i (X_\star)_{ki}^2 = \|(X_\star)_{k\cdot}\|^2$, gives $\mathbb{E}\,\|g_N\|^2 = \sum_i \sigma_i^2 = (1-\rho_x)\tau_N$ by (8).

For the overlap, coordinate independence, the bound $b_k^2 \le \|b\|_\infty^2$, and the exact column norm $\sum_k (X_\star)_{ki}^2 = N$ from (1) give, uniformly in i,

$$\sigma_i^2 \le \frac{(1-\rho_x)\,\|b\|_\infty^2}{N}, \qquad \mathrm{Var}\big({g_N}^\top u_0\big) = \sum_i u_{0,i}^2 \sigma_i^2 \le \frac{(1-\rho_x)\, C_0^2 \log^2 N}{N} \longrightarrow 0,$$

using $\|u_0\| = 1$ and Assumption C.1(iv). Since g_N is centered, Chebyshev's inequality gives, for each fixed $\varepsilon > 0$,

$$\mathbb{P}\big(\,|{g_N}^\top u_0| > \varepsilon\big) \le \frac{\mathrm{Var}\big({g_N}^\top u_0\big)}{\varepsilon^2} \le \frac{(1-\rho_x)\, C_0^2 \log^2 N}{\varepsilon^2 N} \longrightarrow 0,$$

which is ${g_N}^\top u_0 \xrightarrow{P} 0$.

For the norm, the squares $(g_N)_i^2$ are independent, so

$$\mathrm{Var}\big(\,\|g_N\|^2\,\big) = \sum_i \mathrm{Var}\big((g_N)_i^2\big) \le \sum_i \mathbb{E}[(g_N)_i^4].$$

Each coordinate is a sum over k of independent centered summands

$$\eta_{ki} = \frac{\big((S_x)_{ki} - \rho_x\big)(X_\star)_{ki} b_k}{N\sqrt{\rho_x}}.$$

By Assumption C.1(iii)–(iv), these satisfy $|\eta_{ki}| \le \eta_{\max} := C_0^2 \log^2 N/(N\sqrt{\rho_x})$. In the fourth-moment expansion, independence and centering eliminate every term with an index that appears only once. The surviving terms have either four equal indices or two pairs of equal indices. There are three pairings when the distinct indices are summed as ordered pairs, giving

$$\mathbb{E}[(g_N)_i^4] = \sum_k \mathbb{E}\,\eta_{ki}^4 + 3 \sum_{k \ne l} \mathbb{E}\,\eta_{ki}^2\, \mathbb{E}\,\eta_{li}^2 \le \eta_{\max}^2 \sigma_i^2 + 3\sigma_i^4.$$

For the inequality, $\eta_{ki}^4 \le \eta_{\max}^2 \eta_{ki}^2$ bounds the first sum, and the pair sum is at most $(\sum_k \mathbb{E}\eta_{ki}^2)^2 = \sigma_i^4$. As shown above, $\max_i \sigma_i^2 \le (1-\rho_x) C_0^2 \log^2 N/N$. The projector $X_\star X_\star^\top/N$ has diagonal entries in $[0,1]$, so $\|(X_\star)_{k\cdot}\|^2 \le N$. Together with $\|b\|^2 = N$, this gives $\sum_i \sigma_i^2 = (1-\rho_x)\tau_N \le 1-\rho_x$. Therefore

$$\mathrm{Var}\big(\,\|g_N\|^2\,\big) \le \Big(3 \max_i \sigma_i^2 + \eta_{\max}^2\Big) \sum_i \sigma_i^2 = O\Big(\frac{\log^2 N}{N}\Big),$$

so another application of Chebyshev's inequality gives $\|g_N\|^2 - (1-\rho_x)\tau_N \xrightarrow{P} 0$. The finite-$N$ norm targets are

$$d_{a,N} := \rho_x + (1-\rho_x)\tau_N, \qquad d_{p,N} := \rho_y d_{a,N} = \rho + \rho_y(1-\rho_x)\tau_N.$$

The projector identity used above gives $0 \le \tau_N \le 1$. Consequently, $\rho_x \le d_{a,N} \le 1$ and $d_{p,N} \ge \rho > 0$, uniformly in N. Moreover,

$${a_N}^\top u_0 - \sqrt{\rho_x} = {g_N}^\top u_0 \xrightarrow{P} 0,$$

and

$$\|a_N\|^2 - d_{a,N} = 2\sqrt{\rho_x}\, {g_N}^\top u_0 + \big\{\,\|g_N\|^2 - (1-\rho_x)\tau_N\big\} \xrightarrow{P} 0.$$

Since $d_{a,N} \le 1$, this also gives $\|a_N\| = O_P(1)$ without requiring convergence of τ_N.

For the dual reduction, $\sqrt{\rho_y}\, a_N = X^\top(\rho_y b)/(N\sqrt{\rho})$, so (5) and (B.1) give the pathwise identity

$$p_N - \sqrt{\rho_y}\, a_N = \frac{X^\top\big(b \odot (S_y q - \rho_y \mathbf{1})\big)}{N\sqrt{\rho}}.$$

Only the weighted norm of the response-mask fluctuation is needed here. Each entry $(S_y q)_k - \rho_y$ has mean zero and variance $\rho_y(1-\rho_y)\sum_j q_j^2$. Using $\|b\|^2 = N$ and $\sum_j q_j^2 = \sum_j v_{0,j}^4 \le \|v_0\|_\infty^2$ gives

$$\mathbb{E}\,\|b \odot (S_y q - \rho_y \mathbf{1})\|^2 = N\rho_y(1-\rho_y)\sum_j v_{0,j}^4 \le N\,\|v_0\|_\infty^2 .$$

Markov's inequality therefore gives $\|b \odot (S_y q - \rho_y \mathbf{1})\| = O_P(\sqrt{N}\ \|v_0\|_\infty)$. Combining this with $\|X\|_{\mathrm{op}} = O_P(\sqrt{N})$ from Lemma C.4,

$$\big\|p_N - \sqrt{\rho_y}\, a_N\big\| \le \frac{\|X\|_{\mathrm{op}}}{N\sqrt{\rho}}\ \|b \odot (S_y q - \rho_y \mathbf{1})\| = O_P(\|v_0\|_\infty) \xrightarrow{P} 0$$

by Assumption C.1(ii).

The response-mask remainder $r_N := p_N - \sqrt{\rho_y}\, a_N$ satisfies $\|r_N\| \xrightarrow{P} 0$ by the preceding calculation. Hence

$$\big|{p_N}^\top u_0 - \sqrt{\rho_y}\,{a_N}^\top u_0\big| \le \|r_N\|,$$

so ${p_N}^\top u_0 \xrightarrow{P} \sqrt{\rho}$. Moreover,

$$\|p_N\|^2 - d_{p,N} = \rho_y\big\{\,\|a_N\|^2 - d_{a,N}\big\} + 2\sqrt{\rho_y}\,{a_N}^\top r_N + \|r_N\|^2 \xrightarrow{P} 0,$$

using $\|a_N\| = O_P(1)$. On an event whose probability tends to 1,

$$\|p_N\|^2 \ge d_{p,N}/2 \ge \rho/2.$$

On this event, $p_N \neq 0$ and

$$\big(\widehat{p}^\top u_0\big)^2 = \frac{{p_N}^\top u_0^2}{\|p_N\|^2}.$$

Therefore

$$\left|\frac{{p_N}^\top u_0^2}{\|p_N\|^2} - \frac{\rho}{d_{p,N}}\right| \le \frac{\big|{p_N}^\top u_0^2 - \rho\big|}{\|p_N\|^2} + \frac{\rho\,\Big|\|p_N\|^2 - d_{p,N}\Big|}{\|p_N\|^2\, d_{p,N}} \xrightarrow{P} 0.$$

The comparison term is exactly

$$\frac{\rho}{d_{p,N}} = \frac{\rho_x}{\rho_x + (1-\rho_x)\tau_N} = \kappa_{\tau,N}.$$

If $\tau_N \to \tau$, continuous mapping gives the three limiting conclusions in the proposition. □

Lemma C.4 (Operator norm of the masked design)**.** *Under Assumption C.1(iii),* $\mathbb{E}\,\|X\|_{\mathrm{op}} = O(\sqrt{N})$ *and* $\|X\|_{\mathrm{op}} = O_P(\sqrt{N})$*, hence* $\|G\|_{\mathrm{op}} = O_P(1)$.

Whitening fixes the column variances of X exactly, which removes the logarithmic loss a generic sub-Gaussian bound would incur.

Proof. We write $X = \rho_x X_\star + \Xi$ with $\Xi := (S_x - \rho_x) \odot X_\star$, the fluctuation $\sqrt{N\rho_x}\,\Theta$ of Appendix E. Whitening makes every singular value of $X_\star$ equal to $\sqrt{N}$, so $\|\rho_x X_\star\|_{\mathrm{op}} = \rho_x\sqrt{N}$. The entries of Ξ are independent, centered, bounded by $\|X_\star\|_{\max} \le C_0 \log N$, with $\mathrm{Var}(\Xi_{ki}) = \rho_x(1-\rho_x)(X_\star)_{ki}^2$, so the sharp bound for matrices with independent centered entries (Bandeira & van Handel, 2016) applies:

$$\mathbb{E}\,\|\Xi\|_{\mathrm{op}} \ \lesssim\ \sigma_{\mathrm{col}} + \sigma_{\mathrm{row}} + \sigma_*\sqrt{\log(N \wedge D_x)}.$$

Here $\sigma_{\mathrm{col}}^2 = \max_i \sum_k \mathrm{Var}(\Xi_{ki}) = \rho_x(1-\rho_x)N$ exactly, by the diagonal of (1). Next, $\sigma_{\mathrm{row}}^2 = \rho_x(1-\rho_x)\max_k \|(X_\star)_{k\cdot}\|^2 \le \rho_x(1-\rho_x)N$, since $X_\star X_\star^\top/N$ is an orthogonal projector with

diagonal in $[0, 1]$. Finally $\sigma_* \le C_0 \log N$. Hence $\mathbb{E}\,\|\Xi\|_{\mathrm{op}} = O(\sqrt{N})$, since the logarithmic term is subdominant, and $\mathbb{E}\,\|X\|_{\mathrm{op}} \le \rho_x\sqrt{N} + \mathbb{E}\,\|\Xi\|_{\mathrm{op}} \le C\sqrt{N}$. Markov's inequality now gives, for every $M > 0$,

$$\mathbb{P}(\|X\|_{\mathrm{op}} > M\sqrt{N}) \le \frac{\mathbb{E}\,\|X\|_{\mathrm{op}}}{M\sqrt{N}} \le \frac{C}{M}.$$

This proves $\|X\|_{\mathrm{op}} = O_P(\sqrt{N})$ and consequently $\|G\|_{\mathrm{op}} = \|X\|_{\mathrm{op}}^2/(N\rho_x) = O_P(1)$. □

Proposition C.5 (Residual bound)**.** *Under Assumption C.1, at fixed θ,*

$$\|\Delta_N\|_{\mathrm{op}} \;=\; O_P\big(\theta\;\|b\|_\infty\,\|v_0\|_\infty\big) \;=\; O_P\Big(\theta\,\frac{\log N \log D_y}{\sqrt{D_y}}\Big) \longrightarrow 0.$$

Proof. The residual Δ_N is the part of $\mathcal{R}_N$ acting on the orthogonal complement of v_0. Indeed, (B.5) and $\Delta_N v_0 = 0$ give

$$\mathcal{R}_N(I - v_0 v_0^\top) = \Delta_N(I - v_0 v_0^\top) = \Delta_N.$$

Since $\left\|I - v_0 v_0^\top\right\|_{\mathrm{op}} \le 1$, the estimate in the proof of Lemma B.3 yields

$$\mathbb{E}\,\|\Delta_N\|_{\mathrm{op}} \le \mathbb{E}\,\|\mathcal{R}_N\|_{\mathrm{op}} \le C\theta\;\|b\|_\infty\,\|v_0\|_\infty\,.$$

Markov's inequality gives the stated O_P bound. Inserting Assumption C.1(ii) and (iv) gives the rate, which vanishes at fixed θ since $D_y \asymp N$. □

The normalized residual. For $\theta > 0$, let $D_N := \Delta_N(\theta)/\theta$. The exact expression (B.3) shows that D_N does not depend on θ. Since the constant in the preceding estimate is also independent of θ,

$$\mathbb{E}\,\|D_N\|_{\mathrm{op}} \le C\,\|b\|_\infty\,\|v_0\|_\infty \longrightarrow 0.$$

Thus $\Delta_N(\theta_N)/\theta_N \xrightarrow{P} 0$ for every deterministic $\theta_N \to \infty$, with no upper restriction on its growth. This controls the residual relative to signal strength; the unnormalized residual need not vanish when θ_N grows.

D THE HIGH-SIGNAL LIMIT

Proof of Theorem 3.3. The maximizer in (3) is invariant under positive rescaling. For every deterministic $\theta_N \to \infty$, the pair $(\widehat{u}, \widehat{v})$ is therefore the leading singular pair of

$$\frac{C(\theta_N)}{\theta_N} = p_N v_0^\top + D_N + \frac{E}{\theta_N}$$

for all sufficiently large N. Here D_N is the normalized residual defined after Proposition C.5, which gives $\|D_N\|_{\mathrm{op}} \xrightarrow{P} 0$. The response noise $E = A^\top Q/\sqrt{N}$ from Proposition B.1 does not depend on θ and satisfies

$$\|E\|_{\mathrm{op}} \le \frac{\|A\|_{\mathrm{op}}\,\|Q\|_{\mathrm{op}}}{\sqrt{N}} = O_P(1),$$
$$\|A\|_{\mathrm{op}} = \|G\|_{\mathrm{op}}^{1/2} = O_P(1), \qquad \|Q\|_{\mathrm{op}} = O_P\big(\sqrt{N} + \sqrt{D_y}\big) = O_P(\sqrt{N}).$$

Here Lemma C.4 bounds $\|A\|_{\mathrm{op}}$, and the independent-entry norm bound (Vershynin, 2018, Thm. 4.4.5) controls $\|Q\|_{\mathrm{op}}$. The entries of Q are i.i.d. and uniformly sub-Gaussian by the moment identity of Proposition B.1 with $\rho_y > 0$ fixed. Consequently,

$$\left\| D_N + \frac{E}{\theta_N} \right\|_{\mathrm{op}} \le \|D_N\|_{\mathrm{op}} + \frac{\|E\|_{\mathrm{op}}}{\theta_N} \xrightarrow{P} 0.$$

The leading term is rank one with singular pair $(\widehat{p}, v_0)$ and spectral gap $\|p_N\|$, and $\|p_N\|^2 - (\rho + \rho_y(1-\rho_x)\tau_N) \xrightarrow{P} 0$ with $\rho + \rho_y(1-\rho_x)\tau_N \ge \rho > 0$ by Proposition C.3, so $\|p_N\| \ge \sqrt{\rho}/2$ with probability tending to 1. On the event where this lower bound holds and $\|D_N + E/\theta_N\|_{\mathrm{op}} \le \|p_N\|/2$, whose probability tends to 1, Wedin's theorem (Wedin, 1972) applied to the rank-one matrix $p_N v_0^\top$, whose second singular value is zero, gives

$$\max\big\{ \sin\angle(\widehat{u}, \widehat{p}),\ \sin\angle(\widehat{v}, v_0) \big\} \le \frac{2\,\|D_N + E/\theta_N\|_{\mathrm{op}}}{\|p_N\|} \xrightarrow{P} 0.$$

These are acute angles between the spanned lines, so they are invariant to sign changes, as are R_x^2 and R_y^2. Since v_0 is the leading right singular vector of $p_N v_0^\top$,

$$\big|R_y^2 - 1\big| = \sin^2\angle(\widehat{v}, v_0) \xrightarrow{P} 0.$$

For the left overlap, compare the orthogonal projectors onto the two directions. Since u_0 has unit norm,

$$\big|R_x^2 - (\widehat{p}^\top u_0)^2\big| \le \big\|\widehat{u}\widehat{u}^\top - \widehat{p}\widehat{p}^\top\big\|_{\mathrm{op}} = \sin\angle(\widehat{u}, \widehat{p}) \xrightarrow{P} 0.$$

The projector difference vanishes outside the span of the two directions. If that span is two-dimensional, its eigenvalues within the span are $\pm\sin\angle(\widehat{u}, \widehat{p})$; if it is one-dimensional, the projectors coincide. Finally, $(\widehat{p}^\top u_0)^2 - \kappa_{\tau,N} \xrightarrow{P} 0$ by Proposition C.3, which gives $R_x^2 - \kappa_{\tau,N} \xrightarrow{P} 0$ without requiring convergence of τ_N. □

The high-signal limit is strict, $\kappa_\tau < 1$, exactly when $\rho_x < 1$ and $\tau > 0$:

Example D.1 (Two-block design)**.** Whitening does not force $\tau = 1/\alpha_x$. Let N be even, let $\alpha_{x,N} \ge 2$ with $\alpha_{x,N} \to \alpha_x$, and let $U_1, U_2 \in \mathbb{R}^{(N/2)\times D_x}$ satisfy $U_m^\top U_m = \frac{N}{2} I_{D_x}$ (feasible precisely when $D_x \le N/2$, i.e. $\alpha_{x,N} \ge 2$, since U_m must have D_x orthogonal columns in $\mathbb{R}^{N/2}$). Assume that the row norms within each block converge uniformly to their common target:

$$\max_{k \le N/2} \left| \frac{\|(U_m)_{k\cdot}\|^2}{D_x} - 1 \right| \longrightarrow 0, \qquad m = 1, 2.$$

For $\gamma_1, \gamma_2 > 0$ with $\gamma_1^2 + \gamma_2^2 = 2$, stack $X_\star = (\gamma_1 U_1;\ \gamma_2 U_2)$, so that $X_\star^\top X_\star = NI$ exactly. With Hadamard-derived blocks and random-sign (u_0, v_0), Assumption C.1 holds with probability tending

to 1. In block m, whitening determines the total score energy exactly, while the row-norm assumption determines the leverage uniformly over its rows:

$$\frac{1}{N}\sum_{k \text{ in block } m} b_k^2 = \frac{\gamma_m^2}{N}\left\|U_m u_0\right\|^2 = \frac{\gamma_m^2}{2}, \qquad \frac{\left\|(X_\star)_{k\cdot}\right\|^2}{N} = \frac{\gamma_m^2}{\alpha_{x,N}}(1+o(1)).$$

Their product is the contribution of block m to (8). Summing the two contributions gives

$$\tau_N \;=\; \frac{\gamma_1^4+\gamma_2^4}{2}\cdot\frac{1}{\alpha_{x,N}}\,(1+o(1)),$$

so $\tau \geq 1/\alpha_x$ by the power-mean inequality, with equality only at $\gamma_1 = \gamma_2$. At $\alpha_x = 4$, $\rho_x = \frac{1}{2}$ and $\gamma_1^2 = \frac{3}{2}$, $\gamma_2^2 = \frac{1}{2}$ this gives

$$\tau = \tfrac{5}{16} > \tfrac{1}{4} = \frac{1}{\alpha_x}, \qquad \kappa_\tau = \tfrac{16}{21} < \tfrac{4}{5} = \kappa.$$

The block Gram identity fixes only the average row norm, so equal leverage is a separate hypothesis. Hadamard-derived blocks satisfy it exactly; for those blocks the row leverages have the stated block levels exactly and the formula for τ_N has no error term.

E THE DETERMINISTIC ANALYSIS

E.1 CONDITIONAL SPECTRAL QUANTITIES

This subsection defines the conditional spectral quantities behind the finite-signal formulas in Theorem 3.4. The reduced matrix (6) separates a rank-one spike from product noise. Its noise spectrum and its response along the true and corrupted directions are encoded by three Gram resolvent functionals: one weighted trace and two directional forms.

Given S_x, the quantities G, $\widehat{a}_N$, and s_N are fixed. The exact Gram resolvent $R(w) := (wI - G)^{-1}$ retains spectral and signal-direction information at the Gram coordinate w. We use three scalar functionals:

$$\begin{aligned}\psi_N(w) &:= \tfrac{1}{N}\operatorname{Tr}\left[G\,R(w)\right] = \sum_i \frac{\lambda_i/N}{w-\lambda_i},\\ \Phi_N(w) &:= \widehat{a}_N{}^\top R(w)\widehat{a}_N = \sum_i \frac{(\varphi_i{}^\top \widehat{a}_N)^2}{w-\lambda_i},\\ \Phi_N^\times(w) &:= u_0{}^\top R(w)\widehat{a}_N = \sum_i \frac{\varphi_i{}^\top u_0\,\varphi_i{}^\top \widehat{a}_N}{w-\lambda_i},\end{aligned} \tag{E.1}$$

where ψ_N is the weighted average Gram spectral response, Φ_N the corrupted-direction response, and $\Phi_N^\times$ the true/corrupted cross-response. In the orthonormal eigenbasis $(\varphi_i)_{i\le D_x}$ of G, with eigenvalues (λ_i), the form $x^\top R(w)y$ for $x, y \in \mathbb{R}^{D_x}$ weights the i-th coordinate product by $(w-\lambda_i)^{-1}$. Thus the spectrum alone does not determine the directional responses. All three enter the supercritical calculation. Below threshold only the boundedness of the finite-N true-direction response $u_0{}^\top R(w)u_0$ near the response stationary point w_R of Lemma E.4 is used (Appendix F.7); confinement supplies it, and its limit is not part of the deterministic-equivalent assumption. Each functional is the resolvent transform $\sum_i c_i/(w-\lambda_i)$ of a finite (for $\Phi_N^\times$ signed) measure on $\{\lambda_i\}$. At $\rho_x = 1$, $G = I_{D_x}$ and $\widehat{a}_N = u_0$, so the directional responses coincide while ψ_N retains its N^{-1} trace normalization. Explicitly, $\psi_N(w) = \alpha_{x,N}^{-1}(w-1)^{-1}$, whereas each directional response equals $(w-1)^{-1}$. The response fixed point then relates w to a singular-value coordinate z. Its self-consistency correction δ selects the branch $\delta \in (-1, 0)$ continued from zero at infinity, w is the associated Gram coordinate, and $K = w/z$ is the transport factor. On this branch the fixed point and transforms are

$$\delta = -\alpha_{y,N}(1+\delta)\,\psi_N(w), \qquad w := \alpha_{y,N}(1+\delta)\,z^2, \qquad K := w/z, \tag{E.2}$$

$$\bar{m}_1(z) := K\,\Phi_N(w), \qquad \bar{m}_1^\times(z) := K\,\Phi_N^\times(w), \qquad \bar{m}_2(z) := \frac{1}{z(1+\delta)}, \tag{E.3}$$

and these are conditional deterministic-equivalent transforms, with the exact identity $\bar{D} := \bar{m}_1\bar{m}_2 = \alpha_{y,N}\Phi_N(w)$. An above-edge root of $s_N^2\bar{D}(z) = 1$ remains a candidate until Appendix F identifies the corresponding outlier of the random reduced matrix. The scalar construction in Appendices E.2–E.5 provides deterministic comparison values for these functionals. Appendix E.7 evaluates the corresponding limiting transforms m_1, m_2, $m_\times$, and $\bar{D}_0 := m_1m_2$. Their application to the realized design requires Assumption E.10; Appendix F then transfers the predictions to the estimator.

E.2 THE MASKED DESIGN AND THE SCALAR SYSTEM

The three functionals in (E.1) depend on the realization of the design mask. Whitening gives the exact information-plus-noise split below, with deterministic P and independent centered mask fluctuations Θ. Flat and general profiles correspond to the settings of Dozier & Silverstein (2007b) and Hachem et al. (2007), respectively. For A, a_N, and G as in (5), with $\hat{b} := b/\sqrt{N}$, the split is

$$A = P + \Theta, \qquad P = \sqrt{\rho_x/N}\,X_\star, \qquad \Theta = \frac{(S_x - \rho_x)\odot X_\star}{\sqrt{N\rho_x}},$$

where the deterministic identity $a_N = A^\top\hat{b}$ keeps the signal vector and G coupled through the same mask.

Proposition E.1 (Exact information-plus-noise structure)**.** *At every N, $P^\top P = \rho_x I_{D_x}$ and $P^\top \hat{b} = \sqrt{\rho_x}\, u_0$. The entries of Θ are independent and centered, with $\mathrm{Var}(\Theta_{ki}) = (1-\rho_x)(X_\star)_{ki}^2/N$. Each column sum of these variances equals $1-\rho_x$ exactly, while the k-th row sum equals $(1-\rho_x)\,\|(X_\star)_{k\cdot}\|^2/N$.*

Proof. Rescaling (1) gives $P^\top P = (\rho_x/N)\, X_\star^\top X_\star = \rho_x I_{D_x}$, and $X_\star^\top X_\star u_0 = N u_0$ gives $P^\top \hat{b} = \sqrt{\rho_x}\, u_0$. Each Θ_{ki} depends only on the mask entry $(S_x)_{ki}$, so the entries remain independent and centered. Multiplying the Bernoulli variance $\rho_x(1-\rho_x)$ by $(X_\star)_{ki}^2/(N\rho_x)$ gives the stated variance profile. Summing down column i gives $(1-\rho_x)(X_\star^\top X_\star)_{ii}/N = 1-\rho_x$, again by (1). Summing along row k gives $(1-\rho_x)\,\|(X_\star)_{k\cdot}\|^2/N$. □

The scalar comparison sends the mask self-energies e_1, e_2 to block responses g, h through Schur complements, then returns e_1, e_2 by normalized block traces. Here g records the response on $\mathbb{R}^{D_x}$ and h on $\mathrm{range}\, X_\star$. Exact column regularity determines one trace weight, while the equal leverage of Definition 3.2 determines the other asymptotically. At $\xi := \sqrt{w}$, the system is

$$\begin{aligned} g &= \frac{1}{(\xi - e_2) - \dfrac{\rho_x}{\xi - e_1}}, & h &= \frac{1}{(\xi - e_1) - \dfrac{\rho_x}{\xi - e_2}} \\ e_1 &= \frac{(1-\rho_x)\, g}{\alpha_x}, & e_2 &= (1-\rho_x)\Big(\frac{h}{\alpha_x} + \frac{1 - 1/\alpha_x}{\xi - e_1}\Big). \end{aligned} \tag{E.4}$$

For any real rectangular matrix M, its symmetric dilation is

$$H_M := \begin{pmatrix} 0 & M \\ M^\top & 0 \end{pmatrix}. \tag{E.5}$$

The diagonal zero blocks have the row and column dimensions of M, respectively, so H_M is square. For A, this construction acts on $\mathbb{R}^N \oplus \mathbb{R}^{D_x}$ and places the three spectral quantities in one block resolvent as a normalized trace and two directional forms. For $w = \xi^2$ outside the spectrum of $AA^\top$, its resolvent is

$$\mathbf{R}(\xi) := \big(\xi I - H_A\big)^{-1} = \begin{pmatrix} \xi\,(wI - AA^\top)^{-1} & (wI - AA^\top)^{-1} A \\ A^\top (wI - AA^\top)^{-1} & \xi\,(wI - G)^{-1} \end{pmatrix}. \tag{E.6}$$

Schur complementation and the commutation $A^\top (wI - AA^\top)^{-1} = (wI - G)^{-1} A^\top$ give (E.6). Let $\Pi := X_\star X_\star^\top/N$ be the rank-D_x orthogonal projector onto $\mathrm{range}(X_\star)$, so that $P^\top P = \rho_x I_{D_x}$ (Proposition E.1) and $PP^\top = \rho_x \Pi$. For a scalar g and a complex ξ at which d_1 below is nonzero, the block template at aspect ratio $\alpha_{x,N}$ is

$$\begin{aligned} d_1(g;\xi) &:= \xi - (1-\rho_x)\alpha_{x,N}^{-1} g, & M_{22} &= g\, I_{D_x}, \\ M_{11} &= d_1^{-1}(I_N - \Pi) + \big(d_1^{-1} + \rho_x\, g\, d_1^{-2}\big)\Pi, & M_{21} &= d_1^{-1} g\, P^\top, \end{aligned} \tag{E.7}$$

and $\mathcal{M}_N(g;\xi)$ denotes the matrix on $\mathbb{R}^N \oplus \mathbb{R}^{D_x}$ with these blocks and $M_{12} = M_{21}^\top$, indexed as in (E.6). On a solution (g, h, e_1, e_2) of (E.4) at ratio $\alpha_{x,N}$, $d_1(g;\xi) = \xi - e_1$ and $d_1^{-1} + \rho_x g\, d_1^{-2} = h$, so the template is the two-block inverse $\big(\xi I - H_P - \mathrm{diag}(e_1 I_N, e_2 I_{D_x})\big)^{-1}$. The template takes one of three scalars.

- $g_{\mathrm{G},N}(\xi) := D_x^{-1}\, \mathbb{E}\,\mathrm{Tr}\,\big[(\xi I - H_{A_\mathrm{G}})^{-1}\big]_{22}$, the Gaussian expected trace of the benchmark model $A_\mathrm{G} := P + \Theta_\mathrm{G}$ with iid $\mathcal{N}(0, (1-\rho_x)/N)$ entries of Θ_G; Lemma G.2 shows $\big\|\mathbb{E}(\xi I - H_{A_\mathrm{G}})^{-1} - \mathcal{M}_N(g_{\mathrm{G},N};\xi)\big\|_{\mathrm{op}} = O(N^{-1})$ at fixed positive height.
- $g_{0,N}(\xi)$, the exact finite-aspect scalar: the physical solution of the scalar equations (E.4) at ratio $\alpha_{x,N}$; the exact finite-aspect baseline $M_{0,N} := \mathcal{M}_N(g_{0,N};\xi)$ is displayed in (E.16).
- $g(\xi)$, its limit at ratio α_x, with $|g_{0,N} - g| \to 0$ locally uniformly on the exterior by the aspect-ratio comparison (E.17); $M_0 := \mathcal{M}(g;\xi)$, the template with α_x in place of $\alpha_{x,N}$ in d_1, is the limit of $M_{0,N}$.

Lemma G.2 shows $g_{\mathrm{G},N} - g_{0,N} \to 0$; replacing $g_{\mathrm{G},N}$ by $g_{0,N}$ or g in the template therefore costs $o(1)$, not $O(N^{-1})$. The scalar equations (E.4), the physical branch and the support calculation (Proposition E.3, Lemma E.4) concern $g_{0,N}$ and g; this is the meaning of the scalar comparison. For nested flat orthogonal designs the mask fluctuations have exactly constant variance, so the Gaussian benchmark applies after a comparison step (Lemma F.5); designs with a non-constant variance profile are not covered (Appendix E.8). For the masked design, convergence of the three Gram functionals is Assumption E.10; Appendix G proves it for the nested flat class.

E.3 One variable

The scalar system (E.4) has four unknowns, but in that form it hides which solution is physical and where the spectral edge occurs. Reducing it to one coordinate makes both explicit. Let

$$d_1 = \xi - e_1, \qquad d_2 = \xi - e_2, \qquad \Lambda = d_1 d_2 - \rho_x,$$

and define the coordinate

$$\omega = \frac{e_1}{d_1},$$

which parametrizes the physical branch throughout this appendix and Appendix F. The continued fractions of (E.4) are the Schur-complement formulas $g = d_1/\Lambda$ and $h = d_2/\Lambda$. Since $e_1 = \omega d_1 = (1-\rho_x)g/\alpha_x$ and $d_1 d_2 = \Lambda + \rho_x$, the two trace equations of (E.4) imply

$$\Lambda = \frac{1-\rho_x}{\alpha_x \omega}, \qquad d_1 e_2 = \frac{1-\rho_x}{\alpha_x}\Big(\frac{d_1 d_2}{\Lambda} + \alpha_x - 1\Big) = 1 - \rho_x + \rho_x \omega.$$

Now $\xi = d_1(1+\omega)$ and

$$\frac{1-\rho_x}{\alpha_x \omega} = \Lambda = d_1(\xi - e_2) - \rho_x = d_1^2(1+\omega) - 1 - \rho_x \omega,$$

so

$$d_1^2 = \frac{1 + \rho_x \omega + (1-\rho_x)/(\alpha_x \omega)}{1+\omega}, \qquad \xi^2 = (1+\omega)\Big(1 + \rho_x \omega + \frac{1-\rho_x}{\alpha_x \omega}\Big).$$

The remaining scalars are explicit as well:

$$e_1 = \omega d_1, \qquad e_2 = \frac{1 - \rho_x + \rho_x \omega}{d_1}, \qquad g = \frac{\alpha_x \omega d_1}{1-\rho_x}, \qquad h = \frac{1 - \rho_x + \alpha_x \rho_x \omega}{(1-\rho_x)\, d_1}.$$

The physical solution is the branch continued from infinity, where $g, h \sim \xi^{-1}$, $e_1, e_2 \to 0$, and $\omega \downarrow 0$.

The block resolvent expresses the three functionals of (E.1) through normalized traces and deterministic directions. Block inversion (E.6) gives $D_x^{-1} \operatorname{Tr}(wI - G)^{-1} = (\xi D_x)^{-1} \operatorname{Tr} \mathbf{R}_{22}$, hence $\psi_N(w) = N^{-1} \operatorname{Tr}[w(wI - G)^{-1} - I]$ reads off the lower-right block. The commutation $A(wI - G)^{-1} A^\top = w(wI - AA^\top)^{-1} - I$ moves the form at $a_N = A^\top \hat{b}$ to the N-dimensional side, where $\hat{b}$ is deterministic with $\Pi \hat{b} = \hat{b}$ and $\big\|\hat{b}\big\| = 1$, so the kernel part of M_{11} is annihilated. The identity $(wI - G)^{-1} A^\top = \mathbf{R}_{21}$ with $P^\top \hat{b} = \sqrt{\rho_x}\, u_0$ handles the cross form. Exactly, at every N,

$$a_N{}^\top (wI - G)^{-1} a_N = \xi\, \hat{b}^\top \mathbf{R}_{11} \hat{b} - 1, \qquad u_0{}^\top (wI - G)^{-1} a_N = u_0{}^\top \mathbf{R}_{21} \hat{b}. \tag{E.8}$$

These exact identities express the random direction $a_N = A^\top \hat{b}$ through the deterministic vector $\hat{b}$, preserving its dependence on G. On $\{a_N \neq 0\}$, replacing a_N by the unit direction $\widehat{a}_N$ divides the first form by $\|a_N\|^2$ and the second by $\|a_N\|$. No independence between $\widehat{a}_N$ and G is assumed.

The scalar comparison replaces $\mathbf{R}$ by M_0 in (E.8) and the trace identity above. At $\tau = 1/\alpha_x$, Proposition C.3 gives $\|a_N\|^2 \xrightarrow{P} \rho_x + (1-\rho_x)/\alpha_x = n_a > 0$, so $a_N \neq 0$ with probability tending to 1. Applying the two normalization factors gives the candidate limiting transforms

$$\begin{aligned} \psi_0(w) &= \frac{\xi g - 1}{\alpha_x}, & \Phi_0(w) &= \frac{\xi h - 1}{n_a}, \\ \Phi_0^{\times}(w) &= \frac{\sqrt{\rho_x}\, g}{d_1 \sqrt{n_a}}. \end{aligned} \tag{E.9}$$

Separately, the effective spike strength $s_N = \theta\sqrt{\rho_y}\,\|a_N\|$ satisfies $s_N^2 \xrightarrow{P} s_0^2 := \theta^2\rho_y n_a$. The scalars above turn (E.9) into rational functions of ω alone:

$$\xi g - 1 = \frac{\alpha_x\omega(1+\rho_x\omega)}{1-\rho_x}, \qquad \xi h - 1 = \frac{\alpha_x\omega(n_a+\rho_x\omega)}{1-\rho_x}, \qquad \frac{g}{d_1} = \frac{\alpha_x\omega}{1-\rho_x}.$$

Definition E.2 (The scalar system)**.** With $n_a := \rho_x + (1-\rho_x)/\alpha_x = \rho_x/\kappa$, the limit of $\|a_N\|^2$, the scalar system is (E.4) with the limiting transforms (E.9). On the physical branch it is equivalent to the one-variable form: writing $\psi(\omega) := \psi_0(W(\omega))$, $\Phi(\omega) := \Phi_0(W(\omega))$, and $\Phi^\times(\omega) := \Phi_0^\times(W(\omega))$ for $0 < \omega < \omega_G$,

$$\begin{aligned} W(\omega) &= (1+\omega)\Big(1+\rho_x\omega+\frac{1-\rho_x}{\alpha_x\omega}\Big),\\ \psi(\omega) &= \frac{\omega(1+\rho_x\omega)}{1-\rho_x}, \qquad \Phi(\omega) = \frac{\alpha_x\omega(n_a+\rho_x\omega)}{(1-\rho_x)\,n_a}, \qquad \Phi^\times(\omega) = \frac{\alpha_x\sqrt{\rho_x}\,\omega}{(1-\rho_x)\sqrt{n_a}}. \end{aligned} \tag{E.10}$$

Here ρ_x is the design retention, $1-\rho_x$ the design-missingness fraction, and α_x^{-1} the limiting feature-to-sample ratio, so the two contributions to n_a are the mean design and the centered mask fluctuation (Proposition C.3).

The finite-N and scalar Gram measures can be identified separately. The measure $\nu_N := \frac{1}{N}\sum_{i=1}^{D_x}\delta_{\lambda_i(G)}$ is positive, supported on $[0,\infty)$, has mass $\alpha_{x,N}^{-1}$, and satisfies $\psi_N(w) = \int \lambda(w-\lambda)^{-1}\,\mathrm{d}\nu_N(\lambda)$ by the spectral theorem.

For $0 < \rho_x < 1$, eliminating the self-energies in (E.4) identifies $-g(\sqrt{w})/\sqrt{w}$ with the information-plus-noise Stieltjes solution of Dozier & Silverstein (2007b, (1.1)), at population law δ_{ρ_x}, noise variance $1-\rho_x$, and ratio α_x^{-1}. Its positive probability law and unique physical Stieltjes solution select the analytic continuation from infinity. Let ν_0 be α_x^{-1} times that law. It is supported on $[0,\infty)$, has mass α_x^{-1}, and gives

$$g(\xi) = \alpha_x\int\frac{\xi}{\xi^2-\lambda}\,\mathrm{d}\nu_0(\lambda), \qquad \psi_0(w) = \int\frac{\lambda}{w-\lambda}\,\mathrm{d}\nu_0(\lambda).$$

The first formula is the resolvent of the symmetric lift to $\pm\sqrt{\lambda}$ with equal weights; squaring recovers $\alpha_x\nu_0$. The second follows from $\psi_0 = (\xi g - 1)/\alpha_x$. At $\rho_x = 1$, $g(\xi) = \xi/(\xi^2-1)$ directly, and $\nu_0 = \alpha_x^{-1}\delta_1$. This identifies the scalar comparison measure without assuming convergence of the realized ν_N; that link is supplied by Assumption E.10.

E.4 The branch, the fold, and the response edge

The scalar masked-Gram comparison has top edge w_G. The response map T_0 selects a distinct exterior Gram coordinate $w_R > w_G$, and $z_{\mathrm{edge}} = \sqrt{T_0(w_R)}$ is the resulting candidate product-noise singular-value edge. Assumption E.10 links the realized design functionals to the scalar branch, while Appendix F identifies the realized conditional response edge. Throughout, ν_0 is the measure of Appendix E.3.

Every support statement below concerns the scalar comparison rather than the realized spectrum of G. With aspect ratio $\alpha_{x,N}$, the scalar system (E.4) is the information-plus-noise system with population law δ_{ρ_x}, noise level $1-\rho_x$, and ratio $\alpha_{x,N}^{-1}$, so its boundary is available from the published inverse support map. The notation W_N denotes the map (E.10) with α_x replaced by $\alpha_{x,N}$, and $\omega_{G,N}, w_{G,N}$ denote its turning point and minimum value.

Proposition E.3 (Finite-N support map)**.** *For fixed $0 < \rho_x < 1$, let $\mathfrak{b}$ denote the variable written b in the information-plus-noise support analysis of Bai & Silverstein (2012, (1.6)), following Dozier & Silverstein (2007b), reserving b for the latent scores. At population law δ_{ρ_x}, noise level $1-\rho_x$, and ratio $\alpha_{x,N}^{-1}$, that analysis inverts the support map as*

$$x_N(\mathfrak{b}) = \frac{1}{\mathfrak{b}^2}\,m_{\delta_{\rho_x}}^{-1}\Big(\frac{\alpha_{x,N}}{1-\rho_x}\Big(1-\frac{1}{\mathfrak{b}}\Big)\Big) + \frac{(1-\rho_x)(1-\alpha_{x,N}^{-1})}{\mathfrak{b}}.$$

Under $\mathfrak{b} = (1+\omega)^{-1}$, for every $\omega > 0$,

$$x_N(\mathfrak{b}) = W_N(\omega) = (1+\omega)\Big(1 + \rho_x\omega + \frac{1-\rho_x}{\alpha_{x,N}\omega}\Big), \tag{E.11}$$

and the two coordinates run in opposite directions,

$$\frac{\mathrm{d}\mathfrak{b}}{\mathrm{d}\omega} = -\frac{1}{(1+\omega)^2} < 0, \qquad \textit{so} \qquad x_N'(\mathfrak{b}) > 0 \iff W_N'(\omega) < 0. \tag{E.12}$$

The criterion $x_N'(\mathfrak{b}) > 0$ that marks the exterior of the support therefore holds exactly on $0 < \omega < \omega_{G,N}$, and

$$w_{G,N} = \min_{\omega>0} W_N(\omega) = W_N(\omega_{G,N})$$

is the top boundary of the finite-N support, the minimizer being unique by strict convexity. Since $\alpha_{x,N} \to \alpha_x$, $\omega_{G,N} \to \omega_G$ and $w_{G,N} \to w_G$.

Proof. For a point mass $m_{\delta_{\rho_x}}(q) = (\rho_x - q)^{-1}$, so $m^{-1}_{\delta_{\rho_x}}(y) = \rho_x - 1/y$. At $\mathfrak{b} = (1+\omega)^{-1}$ the argument is $-\alpha_{x,N}\omega/(1-\rho_x)$, giving $m^{-1}_{\delta_{\rho_x}} = \rho_x + (1-\rho_x)/(\alpha_{x,N}\omega)$ and

$$x_N(\mathfrak{b}) = (1+\omega)^2\Big(\rho_x + \frac{1-\rho_x}{\alpha_{x,N}\omega}\Big) + (1-\rho_x)\Big(1 - \frac{1}{\alpha_{x,N}}\Big)(1+\omega) = W_N(\omega),$$

the last step by collecting the common factor $1+\omega$ and canceling the two terms in $(1-\rho_x)\alpha^{-1}_{x,N}$. Equation (E.12) is the chain rule. Finally

$$W_N'(\omega) = 2\rho_x\omega + (1+\rho_x) - \frac{1-\rho_x}{\alpha_{x,N}\omega^2}, \qquad W_N''(\omega) = 2\rho_x + \frac{2(1-\rho_x)}{\alpha_{x,N}\omega^3} > 0,$$

so W_N' has the single positive zero $\omega_{G,N}$, and $W_N \to \infty$ at both ends of $(0,\infty)$. Clearing ω^2 makes $\omega_{G,N}$ the positive root of $2\rho_x\omega^3 + (1+\rho_x)\omega^2 - (1-\rho_x)\alpha^{-1}_{x,N} = 0$, which depends continuously on $\alpha^{-1}_{x,N}$, and W_N depends continuously on both. □

Lemma E.4 (Branch, fold, and response edge)**.** *Let $0 < \rho_x < 1$ be fixed. **(a)** The physical solution of* (E.4) *is the branch with $g, h \sim 1/\xi$ at large ξ, continued downwards. It is parametrized by the coordinate ω: the map $\omega \mapsto W(\omega)$ is strictly convex on $\omega > 0$ with a unique turning point ω_G, the positive root of*

$$2\rho_x\omega_G^3 + (1+\rho_x)\omega_G^2 - \frac{1-\rho_x}{\alpha_x} = 0, \tag{E.13}$$

*and $w = W(\omega)$ maps $(0,\omega_G)$ decreasingly onto (w_G,∞), where $w_G := W(\omega_G)$ is the top boundary of the support of the scalar system. The branch has a square-root fold at w_G and no real continuation below it. **(b)** On the branch, eliminating δ from the response fixed point* (E.2)*, evaluated on the limits* (E.9)*, through $1 + \delta = (1 + \alpha_y\psi_0(w))^{-1}$ sends the Gram coordinate to the squared response coordinate*

$$z(w)^2 \;=\; T_0(w) := w\Big(\alpha_y^{-1} + \psi_0(w)\Big), \qquad T(\omega) := T_0(W(\omega)). \tag{E.14}$$

The derivative $\partial_w T_0$ is strictly increasing on (w_G,∞), tends to $-\infty$ at w_G and to $\alpha_y^{-1} > 0$ at infinity, and therefore has a unique zero $w_R = W(\omega_R)$ with $\omega_R \in (0,\omega_G)$. Equivalently,

$$T'(\omega_R) = 0, \qquad T'(\omega) < 0 \quad (0 < \omega < \omega_R), \qquad z_{\mathrm{edge}} = \sqrt{T(\omega_R)}, \tag{E.15}$$

with primes denoting ω-derivatives, and in explicit form

$$T'(\omega) = W'(\omega)\big(\alpha_y^{-1} + \psi(\omega)\big) + W(\omega)\,\frac{1+2\rho_x\omega}{1-\rho_x}.$$

Thus $0 < \omega < \omega_R$ parametrizes the exterior of the response bulk, with z increasing as ω decreases, and $\omega = \omega_R$ is the bulk edge.

Proof. **(a)** Near infinity the system (E.4) linearizes: $g, h \sim 1/\xi$ and $e_1, e_2 = O(1/\xi)$, which identifies the branch there uniquely, with $\omega \downarrow 0$ as $\xi \to \infty$. The Stieltjes-law identification in Appendix E.3 supplies its analytic continuation and the positive Gram measure ν_0. Along the parametrization of Appendix E.3, every solution with $\xi^2 = w$ real corresponds to a root of $W(\omega) = w$. The first two derivatives are

$$W'(\omega) = 2\rho_x\omega + (1+\rho_x) - \frac{1-\rho_x}{\alpha_x\omega^2}, \qquad W''(\omega) = 2\rho_x + \frac{2(1-\rho_x)}{\alpha_x\omega^3} > 0.$$

Thus W' has exactly one positive zero ω_G, which is (E.13) after clearing ω^2. As $\omega \downarrow 0$, $W(\omega) \to \infty$ and $W'(\omega) < 0$, so the branch continued from infinity is $0 < \omega < \omega_G$, W maps it decreasingly onto (w_G, ∞), and by strict convexity $W(\omega) \geq w_G$ for every $\omega > 0$. At fixed parameters, Taylor expansion at the turning point gives

$$W(\omega) - w_G = \frac{1}{2}W''(\omega_G)(\omega - \omega_G)^2 + O\big(|\omega - \omega_G|^3\big).$$

Since $W''(\omega_G) > 0$, the two nearby real roots meet in a square-root fold. The physical branch has $\omega < \omega_G$, so its inverse takes the minus sign:

$$\omega(w) = \omega_G - \sqrt{\frac{2(w - w_G)}{W''(\omega_G)}} + O(w - w_G), \qquad w \downarrow w_G.$$

There is no nearby real continuation below w_G. Proposition E.3 and its proof use no property of the ratio beyond its positivity, so they hold verbatim with α_x in place of $\alpha_{x,N}$. At ratio α_x^{-1}, population law δ_{ρ_x} and noise level $1 - \rho_x$, the information-plus-noise law is the one defining ν_0 in Appendix E.3, and W is its inverse support map, so $w_G = \min_{\omega>0} W(\omega)$ is the top boundary of $\operatorname{supp}\nu_0$. Under Assumption E.10 the functionals converge to this branch, so ν_0 is the limiting Gram law on the exterior domain; for nested flat orthogonal designs Appendix G proves directly that w_G is the spectral edge. **(b)** The fixed point (E.2), evaluated on the limits (E.9), is $\delta = -\alpha_y(1+\delta)\psi_0(w)$, so $1 + \delta = (1 + \alpha_y\psi_0(w))^{-1}$ and $z^2 = w/(\alpha_y(1+\delta)) = T_0(w)$, which is (E.14). With ν_0 as in part (a),

$$\partial_w T_0(w) = \frac{1}{\alpha_y} - \int \frac{\lambda^2}{(w-\lambda)^2}\,\mathrm{d}\nu_0(\lambda), \qquad \partial_w^2 T_0(w) = 2\int \frac{\lambda^2}{(w-\lambda)^3}\,\mathrm{d}\nu_0(\lambda) > 0.$$

The endpoint behavior comes from the one-variable formulas, with no regularity hypothesis on ν_0. On the exterior branch $\psi_0(W(\omega)) = \psi(\omega)$, so (E.10) gives

$$\psi_0'(w) = \frac{\psi'(\omega)}{W'(\omega)}, \qquad \psi'(\omega) = \frac{1 + 2\rho_x\omega}{1 - \rho_x} > 0.$$

As $w \downarrow w_G$, equivalently $\omega \uparrow \omega_G$, one has $W'(\omega) \uparrow 0^-$, so $\psi_0'(w) \to -\infty$ while w and $\psi_0(w)$ stay bounded, and therefore

$$\partial_w T_0(w) = \alpha_y^{-1} + \psi_0(w) + w\,\psi_0'(w) \longrightarrow -\infty,$$

while $\partial_w T_0(w) \to \alpha_y^{-1} > 0$ at infinity. Strict positivity of $\partial_w^2 T_0$ makes $\partial_w T_0$ increasing, so it vanishes exactly once, at $w_R > w_G$. Transporting through the decreasing parametrization $w = W(\omega)$ gives (E.15), and the chain rule gives the explicit $T'(\omega)$ with ψ' as above. □

E.5 The finite-N baseline and uniform margins

The scalar system below is the exact finite-aspect baseline whose limits appear in Assumption E.10; Appendix G verifies those limits for nested flat orthogonal designs by a Gaussian comparison.

At finite N, the scalar baseline uses $\alpha_{x,N}^{-1}$ in place of α_x^{-1}. Its scalar solution is the physical branch of (E.4) at that ratio; with $e_{1,N} = (1-\rho_x)g_{0,N}/\alpha_{x,N}$ and $h_{0,N} = (\xi - e_{1,N})^{-1} + \rho_x g_{0,N}(\xi - e_{1,N})^{-2}$, the template (E.7) evaluated at $g_{0,N}$ gives

$$M_{0,N;22} = g_{0,N} I_{D_x}, \qquad M_{0,N;11} = h_{0,N}\Pi + \frac{I_N - \Pi}{\xi - e_{1,N}}, \qquad M_{0,N;21} = \frac{g_{0,N}}{\xi - e_{1,N}}P^\top, \quad \text{(E.16)}$$

and taking the two normalized traces returns $e_{1,N}$ and yields $e_{2,N} = \frac{1-\rho_x}{\alpha_{x,N}}\{h_{0,N} + (\alpha_{x,N} - 1)/(\xi - e_{1,N})\}$. This calculation uses only $P^\top P = \rho_x I_{D_x}$, not a probabilistic approximation. Here $g_{0,N}, h_{0,N}, e_{1,N}, e_{2,N}, w_{G,N}$ denote the finite-aspect baseline quantities ($g_{0,N}$ and $h_{0,N}$ are distinct from the response coupling g_B in (E.26) and from the fluctuation vector g_N of Proposition C.3), and the unsubscripted notation is reserved for their $\alpha_{x,N}^{-1} \to \alpha_x^{-1}$ limits, as in Appendix E.2. On every fixed exterior compact and for every fixed $k \geq 0$,

$$\sum_{j=0}^{k} \left\| \partial_\xi^j (M_{0,N} - M_0) \right\|_{\mathrm{op}} = O_k\big(\left| \alpha_{x,N}^{-1} - \alpha_x^{-1} \right| \big). \tag{E.17}$$

The physical branch $\omega(\xi)$ at ratio α_x is analytic on the exterior, and differentiating $W(\omega(\xi)) = \xi^2$ gives $W'(\omega)\,\partial_\xi \omega = 2\xi \neq 0$, so $W'(\omega) \neq 0$ along the branch. The implicit function theorem for $W(\omega) = \xi^2$, with the ratio as an additional variable, then makes ω jointly smooth in ξ near the compact and in the ratio near α_x^{-1}. For real ξ the resulting root at ratio $\alpha_{x,N}^{-1}$ lies in $(0, \omega_{G,N})$ because $\omega_{G,N} \to \omega_G$, so it is the physical root, and analytic continuation in ξ extends this to the whole compact. On the branch $d_1 = \xi/(1+\omega)$, so each block of (E.16) combines $I_N - \Pi$, Π, I_{D_x} and $P^\top$, all of norm at most 1, with coefficients rational in $(\omega, \xi, \alpha_{x,N}^{-1})$ whose denominators stay away from zero. The mean value inequality in the ratio, applied to these coefficients and their ξ-derivatives, then gives (E.17). No convergence rate for $\alpha_{x,N}^{-1} - \alpha_x^{-1}$ is needed; (E.17) is a purely deterministic passage to the displayed limits. The same convergence, without rate, holds locally uniformly on compact subsets of $\{\mathrm{Re}\,\xi > 0, \mathrm{Im}\,\xi > 0\}$: $-g_{0,N}(\xi)/\xi$ is the Stieltjes transform at ξ^2 of a law with support in $[0, w_{G,N}]$, these supports are uniformly bounded, every weakly convergent subsequence satisfies the limiting equation at ratio α_x^{-1}, uniqueness of the solution m with $\mathrm{Im}\, m > 0$ and $\mathrm{Im}(\xi^2 m) \geq 0$ identifies the limit with the physical branch of Appendix E.3, and resolvent bounds with equicontinuity give local uniformity.

These calculations establish the branch structure in Lemma E.4. The fold formulas give a square-root branch point at w_G, strict convexity of W shows that there is no higher fold, and the branch is real analytic for $w > w_G$ with no other obstruction connected to infinity.

At the response edge,

$$z_{\mathrm{edge}}^2 = T_0(w_R), \qquad \frac{d^2}{dw^2}\sqrt{T_0(w)}\Big|_{w_R} = \frac{\partial_w^2 T_0(w_R)}{2 z_{\mathrm{edge}}} > 0,$$

and at the stationary point the integral formulas of Lemma E.4 give the quantitative bound, ν_0 being supported on $[0, \infty)$,

$$\partial_w^2 T_0(w_R) \geq \frac{2}{w_R} \int \frac{\lambda^2}{(w_R - \lambda)^2}\, \mathrm{d}\nu_0(\lambda) = \frac{2}{\alpha_y w_R}. \tag{E.18}$$

All quantities above depend continuously on the parameters. For $(\rho_x, \alpha_x^{-1}, \alpha_y^{-1})$ in a compact subset of $(0,1) \times (0,1] \times (0, \infty)$, ω_G stays bounded above and away from zero, the fold derivative stays bounded below at the fold, and the endpoint signs are uniform. To see the latter without appealing to the limiting measure, use the explicit form of $T'(\omega)$ from Lemma E.4: uniformly on the compact parameter set, $T'(\omega) < 0$ for all sufficiently small $\omega > 0$, while $T'(\omega_G) = W(\omega_G)(1 + 2\rho_x \omega_G)/(1 - \rho_x) > 0$. Strict convexity of T_0 makes the root unique. Compactness then gives

$$w_R - w_G \geq \gamma, \qquad \partial_w^2 T_0(w_R) \geq \gamma, \qquad -W'(\omega_R) \geq \gamma \tag{E.19}$$

for some $\gamma > 0$, where $\omega_R \in (0, \omega_G)$ corresponds to w_R. The derivative formulas of Lemma E.4 together with (E.18) certify these margins analytically on every compact interior parameter set.

When $\rho_x = 1$, $\Theta = 0$ and $G = I_{D_x}$ exactly. The coordinate ω and the inner Gram fold disappear, and the response edge is the ordinary scalar sample-covariance edge, evaluated directly in Proposition E.9(i). When $\alpha_x \to \infty$, the cubic (E.13) for ω_G likewise degenerates. Neither limit belongs to the generic proportional fold statement.

Finite-N response transform and exterior branch. Write $\mathbb{C}^+ := \{\mathrm{Im}\, z > 0\}$ and $\mathbb{C}^{++} := \{\mathrm{Re}\, z > 0,\ \mathrm{Im}\, z > 0\}$. Assume $\|G\|_{\mathrm{op}} \leq K_0$ and $G \neq 0$, with (λ_i, φ_i) the eigenpairs of G as in

(E.1). For real $w > \lambda_{\max}(G)$ define

$$T_N(w) := \frac{w}{\alpha_{y,N}} + w\,\psi_N(w) = \frac{w}{\alpha_{y,N}} + \frac{1}{N}\sum_i \frac{w\lambda_i}{w-\lambda_i}, \qquad z^2 = T_N(w); \tag{E.20}$$

on the exterior branch the response fixed point (E.2) is equivalent to the second relation, with $1+\delta = (1+\alpha_{y,N}\psi_N(w))^{-1}$. Since

$$\partial_w T_N(w) = \frac{1}{\alpha_{y,N}} - \frac{1}{N}\sum_i \frac{\lambda_i^2}{(w-\lambda_i)^2}, \qquad \partial_w^2 T_N(w) = \frac{2}{N}\sum_i \frac{\lambda_i^2}{(w-\lambda_i)^3} > 0, \tag{E.21}$$

$\partial_w T_N$ increases strictly from $-\infty$ at $\lambda_{\max}(G)^+$ to $\alpha_{y,N}^{-1}$ at infinity; it has a unique zero $w_{R,N} > \lambda_{\max}(G)$, the minimizer of T_N on $(\lambda_{\max}(G), \infty)$, and

$$z_{\mathrm{edge},N}^2 := T_N(w_{R,N}) = \min_{w>\lambda_{\max}(G)} T_N(w), \qquad \partial_w^2 T_N(w_{R,N}) \geq \frac{2}{\alpha_{y,N} w_{R,N}}, \tag{E.22}$$

the last since $\frac{1}{N}\sum_i \lambda_i^2 (w_{R,N}-\lambda_i)^{-2} = \alpha_{y,N}^{-1}$ and $(w_{R,N}-\lambda_i)^{-1} \geq w_{R,N}^{-1}$. Two bounds are uniform in the design: $T_N(w) \geq w/\alpha_{y,N}$ gives $w_{R,N} \leq \alpha_{y,N} z_{\mathrm{edge},N}^2$, and $z_{\mathrm{edge},N}^2 \leq T_N(2K_0) \leq 2K_0(\alpha_{y,N}^{-1} + \alpha_{x,N}^{-1})$ because $2K_0 - \lambda_i \geq K_0$.

For real $z > z_{\mathrm{edge},N}$ let $w_N(z)$ be the unique solution of $T_N(w) = z^2$ in $(w_{R,N}, \infty)$, and put $c_N(z) := z/w_N(z)$, the reciprocal of the transport factor $K = w/z$ of (E.2). For $z \geq z_{\mathrm{edge},N} + \varepsilon$, because $\partial_w T_N \leq \alpha_{y,N}^{-1}$,

$$w_N(z) - w_{R,N} \;\geq\; \alpha_{y,N}\big(z^2 - z_{\mathrm{edge},N}^2\big) \;\geq\; \alpha_{y,N}\varepsilon^2,$$

and for $w \geq w_{R,N}$, from $\lambda(w-\lambda)^{-1} \leq (w_{R,N}/w)\,\lambda(w_{R,N}-\lambda)^{-1}$ for $0 \leq \lambda < w_{R,N}$ and the stationarity identity,

$$\partial_w T_N(w) \;\geq\; \frac{1}{\alpha_{y,N}}\Big[1 - \Big(\frac{w_{R,N}}{w}\Big)^2\Big].$$

On any bounded real interval $[z_{\mathrm{edge},N} + \varepsilon, K_1]$ these give $\mathrm{dist}(w_N(z), \mathrm{spec}\, G) \geq \alpha_{y,N}\varepsilon^2$ and $\partial_w T_N(w_N(z)) \geq c(\varepsilon, K_1) > 0$, so by the inverse function theorem there is $\eta(\varepsilon, K_1) > 0$, depending only on $(\varepsilon, K_1, K_0, \alpha_{y,N})$, such that w_N extends analytically to the closed rectangle

$$\mathcal{S}_N(\varepsilon, K_1) := \big\{z :\; z_{\mathrm{edge},N} + \varepsilon \leq \mathrm{Re}\, z \leq K_1,\; |\mathrm{Im}\, z| \leq \eta(\varepsilon, K_1)\big\}, \tag{E.23}$$

with $|w_N'| \leq C(\varepsilon, K_1)$ and $\mathrm{dist}(w_N(z), \mathrm{spec}\, G) \geq \frac{1}{2}\alpha_{y,N}\varepsilon^2$ there. On $\mathcal{S}_N(\varepsilon, K_1)$ the finite-N conditional transforms of (E.3),

$$\bar{m}_1 = c_N^{-1}\Phi_N(w_N), \qquad \bar{m}_2 = \alpha_{y,N} c_N,$$
$$\bar{m}_1^{\times} = c_N^{-1}\Phi_N^{\times}(w_N), \qquad \bar{D} = \bar{m}_1\bar{m}_2 = \alpha_{y,N}\Phi_N(w_N),$$

and the secular function $f_N := 1 - s_N^2 \bar{D}$ are analytic with all derivatives bounded by constants depending only on $(\varepsilon, K_1, K_0, \alpha_{y,N}, s_N)$. On the real branch f_N is real; $f_N \equiv 1$ if $s_N = 0$, and otherwise f_N is strictly increasing, since $f_N' = -s_N^2\alpha_{y,N}\Phi_N'(w_N)\,w_N'$ with $\Phi_N' < 0$ and $w_N' = 2z/\partial_w T_N(w_N) > 0$. At a root $\widehat{z}_N$, Cauchy–Schwarz in the spectral decomposition gives $-\Phi_N' \geq \Phi_N^2$, and $w_N' \geq 2\alpha_{y,N} z$, so

$$f_N'(\widehat{z}_N) \;\geq\; \frac{2\widehat{z}_N}{s_N^2}, \qquad w_N(\widehat{z}_N) \leq \lambda_{\max}(G) + \alpha_{y,N}s_N^2, \qquad \widehat{z}_N^2 - z_{\mathrm{edge},N}^2 \leq s_N^2,$$

the last two from $1 = s_N^2\alpha_{y,N}\Phi_N(w_N(\widehat{z}_N)) \leq s_N^2\alpha_{y,N}/(w_N(\widehat{z}_N) - \lambda_{\max}(G))$ and $\partial_w T_N \leq \alpha_{y,N}^{-1}$.

The same branch at positive height. Let H_N be the empirical spectral distribution of $AA^\top/\alpha_{y,N}$ and, for $\zeta \in \mathbb{C}^+$, let $m_N^{\det}(\zeta)$ be the unique solution in $\mathbb{C}^+$ of Silverstein's equation at ratio $\alpha_{y,N}$ and population law H_N,

$$\zeta = -\frac{1}{m} + \frac{1}{N}\sum_i \frac{\lambda_i}{1 + \lambda_i m/\alpha_{y,N}} \qquad \Longleftrightarrow \qquad \zeta = T_N(w), \quad w = -\frac{\alpha_{y,N}}{m}, \tag{E.24}$$

the equivalence because $\lambda_i/(1+\lambda_i m/\alpha_{y,N}) = w\lambda_i/(w-\lambda_i)$. By Bai & Silverstein (2010, (6.1.5)–(6.1.6)), $m_N^{\det}$ is the Stieltjes transform of the probability measure $F_N := F^{\alpha_{y,N},H_N}$ on $[0,\infty)$, whose support has top $z_{\mathrm{edge},N}^2$ (Lemma F.2). For $z \in \mathbb{C}^{++}$ put $w_N(z) := -\alpha_{y,N}/m_N^{\det}(z^2)$; then $\operatorname{Im} w_N(z) > 0$ and $T_N(w_N(z)) = z^2$. For real $z > z_{\mathrm{edge},N}$ this function is real with $w_N(z) \in (w_{R,N},\infty)$ and $T_N(w_N(z)) = z^2$ (Lemma F.2), so it is the real exterior branch; its analytic continuation coincides with the rectangle extension on $\mathcal{S}_N(\varepsilon,K_1)\cap\mathbb{C}^{++}$, and on $\operatorname{Im} z<0$ by conjugation. The reciprocal transport factor is

$$c_N(z) = \frac{z}{w_N(z)} = -\frac{z\, m_N^{\det}(z^2)}{\alpha_{y,N}} = \frac{1}{2\alpha_{y,N}}\int\Big[\frac{1}{z-\sqrt{\lambda}}+\frac{1}{z+\sqrt{\lambda}}\Big]\mathrm{d}F_N(\lambda),$$

which is analytic on $\mathbb{C}^+$ with $\operatorname{Im} c_N \le 0$ there; hence $\operatorname{Im}(zI-c_NG)\succeq(\operatorname{Im} z)I$, and

$$\big\|(zI-c_NG)^{-1}\big\|_{\mathrm{op}} \le (\operatorname{Im} z)^{-1}, \qquad |\alpha_{y,N}c_N(z)| \le (\operatorname{Im} z)^{-1}.$$

With $(zI-c_NG)^{-1} = c_N^{-1}(w_NI-G)^{-1}$, the matrix

$$\mathbf{R}_N^{\det}(z) := \operatorname{diag}\big((zI-c_NG)^{-1},\ \alpha_{y,N}c_N\, I_{D_y}\big)$$

on $\mathbb{R}^{D_x}\oplus\mathbb{R}^{D_y}$ is the deterministic equivalent, established in Appendix F, of $(zI-H_E)^{-1}$ with $E := A^\top Q/\sqrt{N}$ as in (6); it is distinct from $\mathbf{R}(\xi) = (\xi I - H_A)^{-1}$ of (E.6), and $\big\|\mathbf{R}_N^{\det}(z)\big\|_{\mathrm{op}} \le (\operatorname{Im} z)^{-1}$ on $\mathbb{C}^+$. Its upper block has entries $\bar m_1$, $\bar m_1^{\times}$ and $c_N^{-1}u_0{}^\top(w_NI-G)^{-1}u_0$ in the directions $(\widehat a_N,\widehat a_N)$, $(u_0,\widehat a_N)$ and (u_0,u_0), its lower block is $\bar m_2 = \alpha_{y,N}c_N$ times the identity, and its off-diagonal blocks vanish.

The limiting objects are built from the Gram law ν_0 of Appendix E.3. For $\zeta\in\mathbb{C}^+$ let $m_0(\zeta)$ be the unique solution in $\mathbb{C}^+$ of

$$-\frac{1}{m_0(\zeta)} = \zeta - \int\frac{\lambda\,\mathrm{d}\nu_0(\lambda)}{1+\lambda m_0(\zeta)/\alpha_y}, \tag{E.25}$$

the Stieltjes transform of a probability measure F_0 whose support has top $T_0(w_R) = z_{\mathrm{edge}}^2$ (Lemma F.2); put $w_0(z) := -\alpha_y/m_0(z^2)$, $c_0(z) := z/w_0(z)$, and $f_0 := 1-s_0^2\bar D_0$ with $s_0^2 = \theta^2\rho_y n_a$ and $\bar D_0$ from (E.33). On the real exterior $z > z_{\mathrm{edge}}$, w_0 coincides with the branch $w = W(\omega(z))$ of Lemma E.4: both are analytic there, invert T_0, and tend to infinity with z.

E.6 Exact rank-one identities

The first result locates singular values produced by a rank-one perturbation. The second gives the singular-vector overlaps at a simple singular value above the noise spectrum. The finite-signal proof in Appendix F uses both results, and Appendix E.7 evaluates them on the limiting transforms.

Let $B\in\mathbb{R}^{D_x\times D_y}$ be a deterministic noise matrix and let $C_B := s_N\widehat a_N v_0^\top + B$, with the strength s_N and unit directions $\widehat a_N, u_0, v_0$ held fixed. We use the symmetric dilation H_M defined in (E.5). For real $z > \|B\|_{\mathrm{op}}$, let

$$R_1(z) := (z^2I - BB^\top)^{-1}, \qquad R_2(z) := (z^2I - B^\top B)^{-1}.$$

The five exact scalar forms are

$$\begin{aligned} m_{1,B}(z) &:= z\,\widehat a_N{}^\top R_1(z)\widehat a_N, & m_{2,B}(z) &:= z\,v_0{}^\top R_2(z)v_0,\\ m_{\times,B}(z) &:= z\,u_0{}^\top R_1(z)\widehat a_N, & g_B(z) &:= \widehat a_N{}^\top R_1(z)Bv_0,\\ g_{\times,B}(z) &:= u_0{}^\top R_1(z)Bv_0. \end{aligned} \tag{E.26}$$

Define the secular function

$$F_B(z) := \big(1-s_Ng_B(z)\big)^2 - s_N^2 m_{1,B}(z)m_{2,B}(z). \tag{E.27}$$

Proposition E.5 (Exact rank-one secular identity)**.** *For real $z > \|B\|_{\mathrm{op}}$,*

$$\frac{\det(zI-H_{C_B})}{\det(zI-H_B)} = F_B(z). \tag{E.28}$$

The determinant identity also holds for complex z wherever $R_1(z)$ and $R_2(z)$ exist. For real $z > \|B\|_{\mathrm{op}}$, z is a singular value of C_B if and only if $F_B(z)=0$, or equivalently,

$$\big(1-s_N\,g_B(z)\big)^2 \;=\; s_N^2\,m_{1,B}(z)\,m_{2,B}(z). \tag{E.29}$$

Proof of Proposition E.5. The two columns of U place the spike directions in the left and right blocks, and J couples them:

$$U := \begin{pmatrix} \widehat{a}_N & 0 \\ 0 & v_0 \end{pmatrix}, \qquad J := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Thus $H_{C_B} = H_B + s_N U J U^\top$. The relation $R_1(z)B = BR_2(z)$ follows from $(z^2 I - BB^\top)B = B(z^2 I - B^\top B)$ and gives the block inverse

$$(zI - H_B)^{-1} = \begin{pmatrix} zR_1(z) & R_1(z)B \\ B^\top R_1(z) & zR_2(z) \end{pmatrix}.$$

Compressing this inverse to the two spike directions gives

$$U^\top (zI - H_B)^{-1} U = \begin{pmatrix} m_{1,B} & g_B \\ g_B & m_{2,B} \end{pmatrix}.$$

The matrix determinant lemma therefore yields

$$\frac{\det(zI - H_{C_B})}{\det(zI - H_B)} = \det\big(I_2 - s_N J U^\top (zI - H_B)^{-1} U\big) = \det\begin{pmatrix} 1 - s_N g_B & -s_N m_{2,B} \\ -s_N m_{1,B} & 1 - s_N g_B \end{pmatrix}.$$

The 2×2 determinant is $(1-s_N g_B)^2 - s_N^2 m_{1,B} m_{2,B} = F_B(z)$. For real $z > \|B\|_{\mathrm{op}}$ the denominator is nonzero, so $F_B(z) = 0$ exactly when z is an eigenvalue of H_{C_B}. Its positive eigenvalues are the singular values of C_B, which proves the equivalence. The same block algebra proves the identity for complex z wherever the resolvents exist. It also covers $s_N = 0$, for which $F_B \equiv 1$. □

We next calculate the corresponding singular-vector overlaps.

Proposition E.6 (Exact overlap identities)**.** *Suppose that $\widehat{z} > \|B\|_{\mathrm{op}}$ is a simple singular value of C_B, and let (u, v) be a corresponding unit singular pair. Then $F_B'(\widehat{z}) > 0$ and*

$$(\widehat{a}_N{}^\top u)^2 = \frac{2m_{1,B}}{F_B'}, \qquad (v_0{}^\top v)^2 = \frac{2m_{2,B}}{F_B'}, \tag{E.30}$$

$$u_0{}^\top u = \Big(\big(1 - s_N g_B\big)\frac{m_{\times,B}}{m_{1,B}} + s_N g_{\times,B}\Big)\widehat{a}_N{}^\top u. \tag{E.31}$$

All functions are evaluated at $\widehat{z}$, and primes denote differentiation with respect to z.

Proof of Proposition E.6. Let $\alpha := \widehat{a}_N{}^\top u$ and $\beta := v_0{}^\top v$, and evaluate all functions at $\widehat{z}$ unless an argument is shown. Since $\widehat{z} > \|B\|_{\mathrm{op}}$, we have $s_N \neq 0$. If $\alpha = 0$, the equation $C_B^\top u = \widehat{z} v$ would give $B^\top u = \widehat{z} v$, contradicting $\widehat{z} > \|B\|_{\mathrm{op}}$ on unit vectors. The equation $C_B v = \widehat{z} u$ similarly rules out $\beta = 0$. Also, $R_1(\widehat{z})$ and $R_2(\widehat{z})$ are positive definite, so $m_{1,B}(\widehat{z}), m_{2,B}(\widehat{z}) > 0$.

The two singular-vector equations read

$$Bv + s_N \beta\, \widehat{a}_N = \widehat{z} u, \qquad B^\top u + s_N \alpha\, v_0 = \widehat{z} v.$$

The block inverse in the proof of Proposition E.5 therefore gives

$$u = s_N R_1(\widehat{z})\big(\widehat{z}\beta\, \widehat{a}_N + \alpha\, B v_0\big), \qquad v = s_N R_2(\widehat{z})\big(\widehat{z}\alpha\, v_0 + \beta\, B^\top \widehat{a}_N\big).$$

Projecting onto $\widehat{a}_N$ and v_0, respectively, gives

$$\alpha(1 - s_N g_B) = s_N m_{1,B} \beta, \qquad \beta(1 - s_N g_B) = s_N m_{2,B} \alpha.$$

Multiplying the first by β and the second by α gives $\beta^2 m_{1,B} = \alpha^2 m_{2,B}$.

To impose unit norms, use the same equations in stacked form, with U defined above:

$$\begin{pmatrix} u \\ v \end{pmatrix} = s_N (\widehat{z} I - H_B)^{-1} U \begin{pmatrix} \beta \\ \alpha \end{pmatrix}.$$

For real $z > \|B\|_{\mathrm{op}}$, the resolvent is symmetric and its derivative is $\frac{\mathrm{d}}{\mathrm{d}z}(zI-H_B)^{-1} = -(zI-H_B)^{-2}$. Taking the squared norm of the stacked vector and differentiating the compressed resolvent gives

$$2 = \|u\|^2 + \|v\|^2 = s_N^2(\beta,\alpha)U^\top(\widehat{z}I - H_B)^{-2}U\begin{pmatrix}\beta\\ \alpha\end{pmatrix}$$
$$= -s_N^2\big(\beta^2 m'_{1,B} + 2\alpha\beta\, g'_B + \alpha^2 m'_{2,B}\big).$$

Only the resolvent is differentiated; α and β are the fixed overlaps at $\widehat{z}$. Substituting $\beta^2 = \alpha^2 m_{2,B}/m_{1,B}$ and $\alpha\beta = \alpha^2(1 - s_N g_B)/(s_N\, m_{1,B})$, with the divisions justified above, reduces this normalization identity to

$$\frac{\alpha^2}{m_{1,B}}\Big((m_{1,B}m_{2,B})' + \frac{2}{s_N}\big(1 - s_N g_B\big)g'_B\Big) = -\frac{2}{s_N^2},$$

whose bracket equals $-F'_B/s_N^2$ by (E.27). Hence $\alpha^2 F'_B = 2m_{1,B}$, so $F'_B(\widehat{z}) = 2m_{1,B}/\alpha^2 > 0$, and $\beta^2 F'_B = 2m_{2,B}$, which is (E.30). Finally, projecting the representation of u onto u_0 gives $u_0{}^\top u = s_N\big(m_{\times,B}\,\beta + g_{\times,B}\,\alpha\big)$, and eliminating β through the first projection gives (E.31). □

E.7 The threshold and the recovery curves

This subsection evaluates the exact identities of Appendix E.6 on the deterministic scalar limits from Appendices E.2–E.5. If the secular product exceeds 1 at the edge, the comparison has a candidate separated root, and the same transforms give candidate overlaps. Appendix F combines Assumption E.10 with the spike geometry of Proposition C.3, a Gaussian comparison for the response noise, classical spectral exclusion for the noise edge, and residual control to transfer the root and residues to the random estimator. At an exterior response coordinate z, with $w = W(\omega(z))$ on the branch of Lemma E.4, the notation $m_1, m_2, m_\times$ denotes the transforms $\bar{m}_1, \bar{m}_2, \bar{m}_1^\times$ from (E.3) evaluated on the limits (E.9):

$$m_1(z) = \frac{W(\omega)\,\Phi(\omega)}{z}, \qquad m_2(z) = \frac{1 + \alpha_y\psi(\omega)}{z}, \qquad z^2 = T(\omega), \tag{E.32}$$

using $K = w/z$ and $1 + \delta = (1 + \alpha_y\psi(\omega))^{-1}$. Their product simplifies exactly: since $z^2 = W(\omega)(1 + \alpha_y\psi(\omega))/\alpha_y$ by (E.14),

$$\bar{D}_0(z) := m_1(z)\, m_2(z) = \frac{W(\omega)\,\Phi(\omega)\,(1 + \alpha_y\psi(\omega))}{z^2} = \alpha_y\,\Phi(\omega). \tag{E.33}$$

Proposition E.7 (Threshold, outlier, and recovery curves)**.** *On the scalar system of Definition E.2, on the exterior branch of Lemma E.4, with $s_0^2 = \theta^2\rho_y n_a$:*

(i) *$\bar{D}_0$ is positive and strictly decreasing on $(z_{\mathrm{edge}}, \infty)$, from its edge value $\alpha_y\Phi(\omega_R)$ to zero, and the outlier equation $s_0^2\bar{D}_0(z) = 1$ has a root above the edge if and only if $\theta > \theta_c$, where*

$$\theta_c^2 = \frac{1 - \rho_x}{\alpha_x\rho_y\alpha_y\,\omega_R\,(n_a + \rho_x\omega_R)}. \tag{E.34}$$

Because ω_R does not depend on ρ_y, at fixed $(\alpha_x, \alpha_y, \rho_x)$,

$$\theta_c(\rho_y) = \theta_c(1)\,\rho_y^{-1/2}. \tag{E.35}$$

(ii) *For fixed $\theta > \theta_c$ the root is unique and simple: the secular equation reads*

$$\frac{\alpha_x\theta^2\rho_y\alpha_y}{1 - \rho_x}\,\omega_\theta\,(n_a + \rho_x\omega_\theta) = 1,$$
$$\omega_\theta = \frac{\sqrt{n_a^2 + 4\rho_x(1-\rho_x)/(\alpha_x\theta^2\rho_y\alpha_y)} - n_a}{2\rho_x} \;\in\; (0, \omega_R), \tag{E.36}$$

with outlier location $\widehat{z} = \sqrt{T(\omega_\theta)}$, and

$$\Phi'(\omega) = \frac{\alpha_x(n_a + 2\rho_x\omega)}{(1-\rho_x)\,n_a}, \qquad \frac{d\omega}{dz} = \frac{2z}{T'(\omega)},$$
$$\bar{D}_0'(\widehat{z}) = \frac{2\alpha_x\alpha_y\widehat{z}\,(n_a + 2\rho_x\omega_\theta)}{(1-\rho_x)\,n_a\,T'(\omega_\theta)} < 0. \tag{E.37}$$

***(iii)** At the root, the residue identities of Proposition E.6 evaluated on the limiting transforms assign the squared overlaps, with R_a^2 the limit assigned to $(\widehat{u}^\top \widehat{a}_N)^2$,*

$$R_a^2 = \frac{-2m_1}{s_0^2 \bar{D}_0'}, \qquad R_y^2 = \frac{-2m_2}{s_0^2 \bar{D}_0'}, \qquad R_x^2 = \frac{-2m_1}{s_0^2 \bar{D}_0'} \Big(\frac{\Phi^\times(\omega_\theta)}{\Phi(\omega_\theta)} \Big)^2, \tag{E.38}$$

every unlabeled quantity evaluated at $z = \widehat{z}$. Substitution gives, with every term at $\omega = \omega_\theta$,

$$\begin{aligned} R_a^2 &= -\frac{\alpha_x W(\omega)\, \omega^2 (n_a + \rho_x \omega)^2\, T'(\omega)}{(1-\rho_x)\, n_a\, T(\omega)\, (n_a + 2\rho_x \omega)}, \\ R_x^2 &= -\frac{\alpha_x \rho_x W(\omega)\, \omega^2\, T'(\omega)}{(1-\rho_x)\, T(\omega)\, (n_a + 2\rho_x \omega)}, \qquad R_y^2 = -\frac{\alpha_y \omega\, (n_a + \rho_x \omega)\, T'(\omega)}{W(\omega)\, (n_a + 2\rho_x \omega)}. \end{aligned} \tag{E.39}$$

They are positive, tend to zero as $\theta \downarrow \theta_c$, and satisfy $R_x^2 \to \rho_x / n_a = \kappa$ and $R_y^2 \to 1$ as $\theta \to \infty$.

***(iv)** For $\theta < \theta_c$ the outlier equation has no root on $[z_{\mathrm{edge}}, \infty)$, and the secular function keeps the edge margin*

$$1 - s_0^2 \bar{D}_0(z_{\mathrm{edge}}) = 1 - \frac{\theta^2}{\theta_c^2} > 0. \tag{E.40}$$

Proof. **(i)** By (E.10), $\Phi'(\omega) > 0$, and on the exterior branch $T'(\omega) < 0$ (Lemma E.4), so $d\omega/dz = 2z/T'(\omega) < 0$ and $\bar{D}_0(z) = \alpha_y \Phi(\omega(z))$ decreases strictly, to zero because $\omega \downarrow 0$ as $z \to \infty$. Equality of $s_0^2 \bar{D}_0$ with 1 at the edge defines the threshold: inserting $s_0^2 = \theta^2 \rho_y n_a$ and $\alpha_y \Phi(\omega_R) = \alpha_x \alpha_y \omega_R (n_a + \rho_x \omega_R)/((1-\rho_x) n_a)$ and solving for θ^2 gives (E.34). The coordinate ω_R solves $T'(\omega_R) = 0$, an equation in $(\alpha_x, \alpha_y, \rho_x)$ alone, so $\theta_c^2 \rho_y$ is free of ρ_y, which is (E.35).

(ii) The left side of (E.36) is strictly increasing in ω_θ, so the root is unique, and the quadratic formula gives its closed form. The root satisfies $\omega_\theta < \omega_R$ because the left side evaluated at ω_R exceeds 1 exactly when $\theta > \theta_c$. Differentiating (E.10) gives Φ', the chain rule on $z^2 = T(\omega)$ gives $d\omega/dz$, and $\bar{D}_0' = \alpha_y \Phi'(\omega)\, d\omega/dz$ gives (E.37). The sign comes from $T'(\omega_\theta) < 0$.

(iii) At a simple zero of the secular function, the residue identities (E.30)–(E.31) express the three squared overlaps through $m_{1,B}$, $m_{2,B}$, $m_{\times,B}/m_{1,B}$, and F_B'. On the limiting transforms, g_B and $g_{\times,B}$ are replaced by zero, $m_\times/m_1$ by $\Phi^\times/\Phi$, and F_B' by $-s_0^2 \bar{D}_0'$, which gives (E.38). Appendix F proves, under the deterministic equivalent, that the finite-N overlap forms converge to these limits. The cancellation is clearest before expanding W and Φ. At the root, $s_0^2 \alpha_y \Phi = 1$ and $\bar{D}_0' = 2\alpha_y z \Phi'/T'$. Using $m_1 = W\Phi/z$, $m_2 = \alpha_y z/W$, and $z^2 = T$ in (E.38) gives

$$R_a^2 = -\frac{W \Phi^2 T'}{T \Phi'}, \qquad R_y^2 = -\frac{\alpha_y \Phi T'}{W \Phi'}, \qquad R_x^2 = R_a^2 \Big(\frac{\Phi^\times}{\Phi} \Big)^2.$$

Here all functions are evaluated at ω_θ; T' and Φ' are ω-derivatives, while $\bar{D}_0'$ is a z-derivative. Substituting (E.10), including $(\Phi^\times/\Phi)^2 = \rho_x n_a/(n_a + \rho_x \omega)^2$, gives the expanded formulas (E.39). Their common factor $-T'$ makes the response-edge behavior explicit. Positivity follows from $T'(\omega_\theta) < 0$. As $\theta \downarrow \theta_c$, $\omega_\theta \to \omega_R$ and $T'(\omega_\theta) \to 0$, so all three curves vanish. As $\theta \to \infty$, $\omega_\theta \downarrow 0$ with $W(\omega) \sim (1-\rho_x)/(\alpha_x \omega)$, $T(\omega) \sim (1-\rho_x)/(\alpha_x \alpha_y \omega)$, and $T'(\omega) \sim -(1-\rho_x)/(\alpha_x \alpha_y \omega^2)$. Inserting these asymptotics into (E.39) gives $R_x^2 \to \rho_x/n_a$ and $R_y^2 \to 1$.

(iv) At $\theta < \theta_c$ the edge value satisfies $s_0^2 \bar{D}_0(z_{\mathrm{edge}}) = \theta^2/\theta_c^2 < 1$ by (E.34), and $\bar{D}_0$ decreases beyond the edge, so $1 - s_0^2 \bar{D}_0 > 0$ on the whole exterior branch, which is (E.40). □

The subcritical margin (E.40) rules out a separated root but does not by itself force zero overlaps: that requires control of the edge spectral subspace of the random matrix, carried out in Appendix F.7.

Since $s_0^2 = \theta^2 \rho_y n_a$ and $n_a = \rho_x/\kappa$, the root condition $s_0^2\, \bar{D}_0(z_{\mathrm{edge}}) > 1$ reads $\theta^2 > \theta^{*2} \kappa/\rho$, the factorization (11) of the transition boundary. Both signs of the competition there occur: solving (E.4) by the continuation of Lemma E.4 at $\alpha_x = 4$, $\alpha_y = 2$ places θ_c above the isotropic reference $(\alpha_x \alpha_y)^{-1/4} \rho^{-1/2}$ at $\rho_x = \frac{1}{2}$ and below it at $\rho_x = 0.3$.

Lemma E.8 (The deformation raises the detection level)**.** *For every $\alpha_x \geq 1$, $\alpha_y > 0$ and $0 < \rho_x < 1$, $\theta^* > (\alpha_x\alpha_y)^{-1/4}$, with equality only in the limit $\rho_x \to 1$. The threshold is not ordered this way: $\sqrt{\kappa} < 1$ multiplies θ^* in* (11)*, and θ_c falls on either side of $(\alpha_x\alpha_y)^{-1/4}\rho^{-1/2}$, as above.*

Proof. By (E.33), $\theta^{*2} = 1/(\alpha_y\Phi(\omega_R))$, so the claim is $\Phi(\omega)^2 < \alpha_x/\alpha_y$ at $\omega = \omega_R$. Write $q := (1-\rho_x)/\alpha_x$, so that $n_a = \rho_x + q$ and, by (E.10), $\Phi(\omega) = \omega(n_a + \rho_x\omega)/(q\, n_a)$ and $\psi'(\omega) = (1+2\rho_x\omega)/(1-\rho_x)$; put $p := -W'(\omega) = q/\omega^2 - (1+\rho_x) - 2\rho_x\omega$. Solving $T'(\omega) = 0$ in the explicit form of Lemma E.4 gives $\alpha_y^{-1} = W(\omega)\psi'(\omega)/p - \psi(\omega)$, and substituting it,

$$\frac{\alpha_x}{\alpha_y} - \Phi(\omega)^2 = \frac{Q(q,\rho_x,\omega)}{p\, q^2 n_a^2},$$

where Q has positive integer coefficients, grouped by powers of ω as

$$\begin{aligned} Q = {} & q^2(q+\rho_x)^2(1+\rho_x) + 2q(q+\rho_x)(q^2\rho_x + q\rho_x^2 + 2q\rho_x + q + 2\rho_x^2)\omega \\ & + (3q^3\rho_x^2 + 6q^3\rho_x + 6q^2\rho_x^3 + 12q^2\rho_x^2 + q^2\rho_x + q^2 \\ & \quad + 3q\rho_x^4 + 6q\rho_x^3 + q\rho_x^2 + 2q\rho_x + \rho_x^3 + \rho_x^2)\omega^2 \\ & + 2\rho_x(q+\rho_x)(2q^2\rho_x + 2q\rho_x^2 + q + 2\rho_x + 1)\omega^3 + \rho_x^2(4q + 5\rho_x + 1)\omega^4 + 2\rho_x^3\omega^5. \end{aligned}$$

Every factor of the denominator except p is positive, and $p > 0$ reads $2\rho_x\omega^3 + (1+\rho_x)\omega^2 - (1-\rho_x)/\alpha_x < 0$, which (E.13) identifies with $\omega < \omega_G$. The certificate therefore applies to every exterior stationary pair satisfying the edge equation, and in particular at ω_R. At $\rho_x = 1$ the two sides agree by Proposition E.9(i). □

Proposition E.9 (Limits of the scalar system)**.** ***(i)*** *At $\rho_x = 1$ the deformation is absent: $G = I_{D_x}$ exactly, $\widehat{a}_N = u_0$, $n_a = 1$, both self-energies vanish, and the transforms collapse to*

$$\psi_0(w) = \frac{1}{\alpha_x(w-1)}, \qquad \Phi_0(w) = \Phi_0^{\times}(w) = \frac{1}{w-1}, \qquad T_0(w) = w\Big(\alpha_y^{-1} + \frac{1}{\alpha_x(w-1)}\Big).$$

Here $s_0^2 = \rho_y\theta^2$. The edge, threshold, and outlier have closed forms,

$$\begin{aligned} w_R &= 1 + \sqrt{\alpha_y/\alpha_x}, & z_{\mathrm{edge}} &= \alpha_y^{-1/2} + \alpha_x^{-1/2}, \\ \theta_c^2 &= \frac{1}{\rho_y\sqrt{\alpha_x\alpha_y}}, & w_\theta &= 1 + \alpha_y s_0^2, \end{aligned} \tag{E.41}$$

with $\widehat{z}^2 = (s_0^2 + \alpha_x^{-1})(s_0^2 + \alpha_y^{-1})/s_0^2$. For $\theta > \theta_c$,

$$R_x^2 = \frac{\alpha_x\alpha_y s_0^4 - 1}{\alpha_y s_0^2(\alpha_x s_0^2 + 1)}, \qquad R_y^2 = \frac{\alpha_x\alpha_y s_0^4 - 1}{\alpha_x s_0^2(\alpha_y s_0^2 + 1)}, \tag{E.42}$$

which are the closed forms in (13)*. Since $\widehat{a}_N = u_0$ here, $R_a^2 = R_x^2$. In particular $\theta^* = (\alpha_x\alpha_y)^{-1/4}$ exactly in this case.*

(ii) *Let $0 < \rho_x < 1$, $0 < \rho_y \leq 1$, and $0 < \alpha_y < \infty$ be fixed. As $\alpha_x \to \infty$, $\kappa \to 1$. For every fixed $\theta > 0$, $s_0^2 \to \rho\theta^2$, the root is eventually supercritical, and*

$$\widehat{z}^2 \longrightarrow \rho\theta^2 + \alpha_y^{-1}, \qquad R_a^2, R_x^2 \longrightarrow 1, \qquad R_y^2 \longrightarrow \frac{\alpha_y\rho\theta^2}{1 + \alpha_y\rho\theta^2}.$$

At $\theta = 0$ there is no exterior root and the subcritical recovery curves are zero. The effective threshold retains a relative inflation: $(\alpha_x\alpha_y)^{1/4}\theta^* \to \big(1 + (1-\rho_x^2)/\alpha_y\big)^{1/4} > 1$, with equality only as $\rho_x \to 1$.

Proof. **(i)** At $\rho_x = 1$, $\Theta = 0$, so $G = I_{D_x}$ and $a_N = A^\top\hat{b} = u_0$ exactly; the coordinate ω and the inner Gram fold disappear, and (E.4) collapses to $e_1 = e_2 = 0$ and $g = h = \xi/(w-1)$, giving the

displayed transforms through (E.9). Solving $\partial_w T_0(w_R) = 0$ with

$$\partial_w T_0(w) = \frac{1}{\alpha_y} - \frac{1}{\alpha_x (w-1)^2}$$

gives $w_R = 1 + \sqrt{\alpha_y/\alpha_x}$ and $z^2_{\text{edge}} = T_0(w_R) = (\alpha_x^{-1/2} + \alpha_y^{-1/2})^2$. The outlier equation $s_0^2 \alpha_y \Phi_0(w_\theta) = 1$ gives $w_\theta = 1 + \alpha_y s_0^2$, hence $\widehat{z}^2 = T_0(w_\theta)$ as displayed, and the edge value $\Phi_0(w_R) = \sqrt{\alpha_x/\alpha_y}$ gives θ_c^2. For the curves, $\bar{D}_0(z) = \alpha_y/(w-1)$ and $dw/dz = 2z/\partial_w T_0(w)$, so (E.38) evaluates through

$$\partial_w T_0(w_\theta) = \frac{\alpha_x \alpha_y s_0^4 - 1}{\alpha_x \alpha_y^2 s_0^4}, \qquad R_y^2 = \frac{(w_\theta - 1)^2\, \partial_w T_0(w_\theta)}{s_0^2\, w_\theta}, \qquad R_x^2 = \frac{\alpha_x (w_\theta - 1)^2\, \partial_w T_0(w_\theta)}{s_0^2\, (\alpha_x (w_\theta - 1) + \alpha_y)},$$

where the last two displays substitute $m_2 = \alpha_y z/w$, $m_1 = w\Phi_0(w)/z$, and $\Phi_0^\times/\Phi_0 = 1$ into (E.38). Inserting $w_\theta - 1 = \alpha_y s_0^2$ gives (E.42), which is (13). **(ii)** The limit has two scales: fixed exterior Gram coordinates and coordinates approaching the edge. Evaluating W at $\alpha_x^{-1/2}$ gives $1 \le w_G \le (1 + \alpha_x^{-1/2})^2 \to 1$, so every fixed $w > 1$ is eventually exterior. Its physical solution of $W(\omega) = w$ has $\omega \to 0$, since $\omega < \omega_G$ and (E.13) gives $\omega_G = O(\alpha_x^{-1/2})$. Expanding W gives

$$w - 1 = (1 + \rho_x)\omega + \rho_x \omega^2 + \frac{1 - \rho_x}{\alpha_x} + \frac{1 - \rho_x}{\alpha_x \omega},$$

so $\alpha_x \omega \to (1 - \rho_x)/(w - 1)$. Thus (E.10) and (E.9), with $n_a \to \rho_x$, yield

$$\psi_0(w) \to 0, \qquad \Phi_0(w), \Phi_0^\times(w) \to \frac{1}{w - 1}.$$

These pointwise limits describe the response away from $w = 1$.

To locate the response edge, expand its defining derivative directly:

$$T'(\omega) = -\frac{1 - \rho_x}{\alpha_x \alpha_y \omega^2} + \frac{1 + \rho_x}{\alpha_y} + \frac{1}{1 - \rho_x} + O(\omega + \alpha_x^{-1}).$$

The remainder is uniform for $0 < \omega \le \omega_0$ at any fixed small $\omega_0 > 0$. Indeed, the omitted terms form a polynomial in ω whose constant term is $(1 + \rho_x)/\alpha_x$ and whose positive-power coefficients are bounded for $\alpha_x \ge 1$, with ρ_x and α_y fixed. By Lemma E.4, $0 < \omega_R < \omega_G = O(\alpha_x^{-1/2})$. The uniform expansion therefore applies at ω_R, where $T'(\omega_R) = 0$:

$$\frac{1 - \rho_x}{\alpha_x \alpha_y \omega_R^2} = \frac{1 + \rho_x}{\alpha_y} + \frac{1}{1 - \rho_x} + O(\alpha_x^{-1/2}).$$

The right-hand side tends to a positive constant. Solving for ω_R gives

$$\omega_R = \frac{1 - \rho_x}{\sqrt{\alpha_x(\alpha_y + 1 - \rho_x^2)}}\,(1 + o(1)), \qquad \Phi(\omega_R) = \sqrt{\frac{\alpha_x}{\alpha_y + 1 - \rho_x^2}}\,(1 + o(1)),$$

where the second relation follows from (E.10) and $n_a \to \rho_x$. Therefore $\theta^{*2} = 1/(\alpha_y \Phi(\omega_R))$ gives the stated relative inflation, while $\theta_c^2 = \theta^{*2}/(\rho_y n_a) \to 0$.

For fixed $\theta > 0$, the secular equation (E.36) and $n_a \ge \rho_x > 0$ first give $\omega_\theta = O(\alpha_x^{-1})$, and then

$$\alpha_x \omega_\theta \longrightarrow \frac{1 - \rho_x}{\alpha_y \rho \theta^2}, \qquad \frac{\omega_\theta}{\omega_R} \longrightarrow 0.$$

Thus this root is eventually on the exterior response branch. Substitution into W and T gives

$$W(\omega_\theta) \to 1 + \alpha_y \rho \theta^2, \qquad T(\omega_\theta) \to \rho\theta^2 + \alpha_y^{-1}.$$

The derivative expansion and (E.10) also give, at ω_θ,

$$\omega_\theta T' \to -\rho\theta^2, \qquad \frac{\Phi}{\Phi'} \sim \omega_\theta, \qquad \Phi \to \frac{1}{\alpha_y \rho \theta^2}, \qquad \frac{\Phi^\times}{\Phi} \to 1.$$

Inserting these relations into the compact recovery expressions in the proof of Proposition E.7 yields the three stated overlap limits. At $\theta = 0$, the secular function is identically 1, so there is no exterior root and the subcritical curves remain zero. □

The coordinate formulas require $\rho_x < 1$. The large-aspect-ratio limit also leaves the fixed-parameter fold regime because $\omega_G \to 0$. Part **(ii)** concerns limits of the deterministic curves, with positive signal and the retentions fixed. Where the finite-signal theorem applies at each fixed aspect ratio, the estimator interpretation takes $N \to \infty$ first and then $\alpha_x \to \infty$. The convergence is not uniform as $\theta \downarrow 0$. The effective threshold θ^* retains a ratio strictly greater than 1 to its isotropic counterpart. The variance scale $(1-\rho_x^2)/\alpha_x$ from (E.44) has the same order as $(w_R-1)^2$, so it remains visible at the edge while the fixed-w transforms converge.

E.8 The deterministic equivalent

Appendices E.2–E.5 construct the scalar comparison, Appendix E.6 gives exact fixed-noise identities, and Appendix E.7 evaluates them on scalar limits. Assumption E.10 independently bridges to the realized mask through Gram confinement and convergence of the weighted spectral response and two directional functionals.

> **Assumption E.10** (Deterministic equivalent for the masked design)**.** Under Assumption C.1 and, for $0 < \rho_x < 1$, asymptotically equal leverage as in Definition 3.2, let w_G be the fold of Lemma E.4; at $\rho_x = 1$, where $G = I_{D_x}$ exactly, read $w_G := 1$, the top of the Gram support. Then **(i)** $\lambda_{\max}(G) \le w_G + o_P(1)$, and **(ii)** for every compact $K \subset (w_G, \infty)$ there is a fixed open complex neighborhood $U_K \supset K$, with $\overline{U}_K$ separated from the Gram support, on which the random functionals ψ_N, Φ_N, $\Phi_N^\times$ of (E.1) converge in probability, uniformly, to the analytic continuations of the limits ψ_0, Φ_0, $\Phi_0^\times$ in (E.9). At $\rho_x = 1$, condition **(ii)** is interpreted through $\psi_0(w) = 1/[\alpha_x(w-1)]$ and $\Phi_0(w) = \Phi_0^\times(w) = 1/(w-1)$, as in Proposition E.9(i).

Consequences. On the confinement event in **(i)** the three functionals are analytic and bounded on U_K. Cauchy's formula on a smaller neighborhood of K transfers the uniform convergence in **(ii)** to every fixed w-derivative; in particular $T_N \to T_0$ in C^2 on compacts of (w_G, ∞), with T_N from (E.20) (Appendix E.5). The Gram law is recovered from ψ_N: since ν_N has mass $\alpha_{x,N}^{-1}$, $\int \mathrm{d}\nu_N(\lambda)/(w-\lambda) = (\psi_N(w) + \alpha_{x,N}^{-1})/w$. Fix a countable set $\mathcal{D} \subset U_K \cap \mathbb{C}^+$ with an accumulation point in $U_K \cap \mathbb{C}^+$. Given a subsequence, pass to a further subsequence along which confinement and the convergence $\psi_N(w) \to \psi_0(w)$ hold almost surely for every $w \in \mathcal{D}$. On the almost-sure event the measures ν_N are eventually supported in $[0, w_G + 1]$ with masses $\alpha_{x,N}^{-1} \to \alpha_x^{-1}$, every subsequential weak limit has Stieltjes transform $(\psi_0(w) + \alpha_x^{-1})/w = \int \mathrm{d}\nu_0(\lambda)/(w-\lambda)$ on $\mathcal{D}$, hence on $\mathbb{C}^+$ by analyticity, and a finite measure is determined by its Stieltjes transform. So $\nu_N \Rightarrow \nu_0$ in probability, with ν_0 the Gram law of Appendix E.3, and the empirical spectral distribution of $AA^\top$ converges weakly in probability to $\nu_0 + (1-\alpha_x^{-1})\delta_0$.

Proposition G.1 verifies the assumption for nested flat orthogonal designs. For unequal-leverage designs τ need not equal $1/\alpha_x$, since the coupling of Definition 3.2 depends on the scores as well as the row profile; persistent leverage variation generally makes the deterministic equivalents nonconstant on the row block, so the scalar reduction changes.

Necessity of incoherence. Orthogonal columns and exactly equal leverage do not imply the deterministic equivalent. Let α_x be a fixed integer and consider a constant-ratio sequence $\alpha_{x,N} = N/D_x \equiv \alpha_x$. Partition the N rows into D_x blocks of size α_x. Let column i of $X_\star$ equal $\sqrt{D_x}$ on block i and zero elsewhere. Then

$$X_\star^\top X_\star = N I_{D_x}, \qquad \|(X_\star)_{k\cdot}\|^2 = D_x$$

exactly, but $\|X_\star\|_{\max} = \sqrt{D_x}$, so incoherence fails. The masked Gram matrix is diagonal, with

$$G_{ii} = \frac{\mathrm{Binomial}(\alpha_x, \rho_x)}{\alpha_x \rho_x},$$

and therefore

$$\frac{1}{D_x}\operatorname{Tr} G^2 \;\longrightarrow\; 1 + \frac{1-\rho_x}{\alpha_x \rho_x}. \tag{E.43}$$

For a flat incoherent profile, direct second-moment expansion (or the scalar system (E.4) at infinity) gives instead

$$\frac{1}{D_x}\operatorname{Tr} G^2 \;\longrightarrow\; 1+\frac{1-\rho_x^2}{\alpha_x}. \tag{E.44}$$

The limits differ, for example at $\rho_x = \frac{1}{2}$; hence orthogonal columns and equal leverage do not imply the deterministic equivalent without entrywise incoherence.

In finite-size checks, the mask-averaged functionals for nonnested Paley and Gaussian-QR designs differ from their predicted limits on a scale consistent with $O(N^{-1/2})$ fluctuations of resolvent bilinear forms.

F THE FINITE-SIGNAL PROOF

We prove Theorem 3.4 from Assumption E.10, Assumption C.1, Definition 3.2, Proposition C.3 and Lemma B.3. Classical covariance results locate the noise edge (Bai & Silverstein, 2010, Theorem 4.1, Lemma 6.1, Theorem 6.3(1)). Under the stated moment bounds, matching the first two entry moments gives asymptotic resolvent comparison at fixed positive height (Lemma F.5); a Gaussian calculation identifies the finite-N comparison forms (Lemma F.7). Exterior continuation (Proposition F.9) then gives the outlier by a sign change and interlacing (Theorem F.10); the deterministic equivalent gives the curves (Corollary F.11, Appendix E.7). Below threshold we control the edge subspace, then transfer both regimes through the residual.

F.1 SETTING AND REDUCTION TO DETERMINISTIC DESIGNS

Throughout, $(X_\star, u_0, v_0, b)$ are deterministic and we condition on the design mask S_x. We use the reduction (6), with the notation of (5), the Gram measures of Appendix E.3, and the response objects of Appendix E.5. The remaining randomness is (S_y, Z); the entries of Q are iid, centered and of unit variance, with every fixed moment bounded at the fixed retention $\rho_y \in (0, 1]$, since $\mathbb{E}\,|Q_{kj}|^m = \rho_y^{1-m/2}\,\mathbb{E}\,|Z_{11}|^m$ (Proposition B.1).

For rectangular M, let H_M be the dilation (E.5) and $R_M(z) := (zI - H_M)^{-1}$. We embed first- and second-block vectors as $\mathbf{u} := (u; 0)$ and $\mathbf{v} := (0; v)$; in particular, the three signal directions are $\hat{\mathbf{a}}_N, \mathbf{u}_0, \mathbf{v}_0$. The domains $\mathbb{C}^+$ and $\mathbb{C}^{++}$ are as in Appendix E.5. Write K_0 for a norm bound on G and C_η for constants depending on the height η and the fixed parameters. Constants are uniform for $(\rho_x, \alpha_{x,N}^{-1}, \alpha_{y,N}^{-1})$ in a compact subset of $(0, 1) \times (0, 1] \times (0, \infty)$ and for θ in a compact set separated from θ_c; limits are in probability unless stated.

For $M \in \mathbb{R}^{D_x \times D_y}$,

$$R_M(z) = \begin{pmatrix} z\,(z^2 I - MM^\top)^{-1} & (z^2 I - MM^\top)^{-1} M \\ M^\top (z^2 I - MM^\top)^{-1} & z\,(z^2 I - M^\top M)^{-1} \end{pmatrix}, \qquad \|R_M(z)\|_{\mathrm{op}} \le |\operatorname{Im} z|^{-1}, \qquad R_M(\bar{z}) = \overline{R_M(z)}. \tag{F.1}$$

The five scalar forms and the secular function are those of (E.26) and (E.27); the determinant identity (E.28) applies with $B = E$.

Standing bounds: on $\{\|G\|_{\mathrm{op}} \le K_0\}$, $\|E\|_{\mathrm{op}} \le \|G\|_{\mathrm{op}}^{1/2} \|Q\|_{\mathrm{op}} / \sqrt{N} \le K_E := \sqrt{K_0}\,(1 + \alpha_y^{-1/2}) + 1$ with probability tending to 1 (Bai & Silverstein, 2010, Theorem 5.8); $\|\mathcal{R}_N\|_{\mathrm{op}} = o_P(1)$ (Lemma B.3 and Markov's inequality).

Lemma F.1 (Reduction to deterministic designs)**.** *Choose a deterministic sequence $\mathfrak{t}_N \downarrow 0$ so slowly that each of the following holds with probability tending to 1: $\|G\|_{\mathrm{op}} \le K_0$; $\lambda_{\max}(G) \le w_G + \mathfrak{t}_N$; $\sup_{U_K} \big(|\psi_N - \psi_0| + |\Phi_N - \Phi_0| + \big|\Phi_N^\times - \Phi_0^\times\big| \big) \le \mathfrak{t}_N$ for the finitely many neighborhoods U_K of Assumption E.10 used below; $\big|\|a_N\|^2 - n_a\big| + \big|a_N{}^\top u_0 - \sqrt{\rho_x}\big| \le \mathfrak{t}_N$; $\mathbb{E}[\|\mathcal{R}_N\|_{\mathrm{op}} \mid \text{design}] \le \mathfrak{t}_N$. Such a sequence exists because each quantity tends to zero in probability. It suffices to prove each conclusion of Theorem 3.4, conditionally on the design, along every deterministic sequence of designs satisfying these bounds; the remaining randomness is (S_y, Z).*

Proof. Let $\phi_N(S) \in [0, 1]$ be the conditional probability, given design S, that a fixed conclusion fails by more than ϵ. Let B_N be the event that all displayed bounds hold. If $\mathbb{E}\phi_N \not\to 0$, choose a subsequence and $\delta > 0$ with $\mathbb{E}\phi_N > \delta$. Since $\mathbb{P}(B_N^c) \to 0$, we have $\mathbb{E}[\phi_N \mathbf{1}_{B_N}] \ge \mathbb{E}\phi_N - \mathbb{P}(B_N^c) > \delta/2$ for all large N on that subsequence. Some $S_N \in B_N$ therefore satisfies $\phi_N(S_N) > \delta/2$. The claimed result along every such deterministic design sequence gives $\phi_N(S_N) \to 0$, a contradiction. □

For the assembly of Theorem 3.4, fix a design sequence from Lemma F.1; probabilities now refer to the response randomness. The lemmas and conditional theorem below retain their own stated

assumptions. We record the deterministic consequences of the deterministic equivalent used in the assembly.

By Assumption E.10, its recorded consequences and Proposition C.3,

$$\nu_N \Rightarrow \nu_0, \qquad s_N \to s_0, \qquad T_N \longrightarrow T_0 \quad \text{in } C^2 \text{ on compacts of } (w_G, \infty).$$

On $[w_G + \gamma/2, K_2]$, for large fixed K_2, $\partial_w T_N$ converges uniformly to $\partial_w T_0$, which changes sign only at w_R. Since $\partial_w T_N$ is strictly increasing above $\lambda_{\max}(G) \le w_G + \mathfrak{t}_N$, its unique zero satisfies $w_{R,N} \to w_R$, and $z^2_{\mathrm{edge},N} = T_N(w_{R,N}) \to T_0(w_R)$.

Each of $\psi_N, \Phi_N, \Phi_N^{\times}$ is a transform $\int (w-\lambda)^{-1}\,\mathrm{d}\mu(\lambda)$ with bounded total variation and support in $\operatorname{spec} G$. The measure-identification argument used for ν_N therefore gives convergence of all three, uniformly on every compact of $\mathbb{C} \setminus [0, w_G]$, where they are defined for large N. The identification argument in Lemma F.7 uses only $\nu_N \Rightarrow \nu_0$ and gives $m_N^{\det} \to m_0$ on $\mathbb{C}^+$ and $w_N \to w_0$, $c_N \to c_0$ on $\mathbb{C}^{++}$, locally uniformly. Since $\operatorname{Im} w_0 > 0$ there, composition gives

$$(\bar m_1, \bar m_2, \bar m_1^{\times}, \bar D, f_N) \longrightarrow (m_1, m_2, m_{\times}, \bar D_0, f_0) \quad \text{locally uniformly on } \mathbb{C}^{++}.$$

On real exterior compacts and the rectangles of (E.23), the inverse function theorem applied to $T_N \to T_0$ on U_K gives $w_N \to w_0$. Cauchy's formula then gives convergence of these transforms and all fixed derivatives on smaller neighborhoods.

F.2 The response-noise edge

Lemma F.2 (Support criterion)**.** *Assume $\|G\|_{\mathrm{op}} \le K_0$ and $G \neq 0$, and let F_N be the probability law with Stieltjes transform $m_N^{\det}$ of* (E.24)*, the deterministic equivalent of the ESD of $E^\top E$. If $w > \lambda_{\max}(G)$ and $\partial_w T_N(w) > 0$, then $T_N(w) \notin \operatorname{supp} F_N$; and $\sup \operatorname{supp} F_N = z^2_{\mathrm{edge},N}$. For real $u > z^2_{\mathrm{edge},N}$, $-\alpha_{y,N}/m_N^{\det}(u)$ is the unique solution of $T_N(w) = u$ in $(w_{R,N}, \infty)$. The same statements hold for ν_0, T_0 and the law F_0 with transform m_0 of* (E.25)*; in particular* $\sup \operatorname{supp} F_0 = T_0(w_R) = z^2_{\mathrm{edge}}$.

Proof. For $F_N = F^{\alpha_{y,N}, H_N}$, with H_N the ESD of $T = AA^\top/\alpha_{y,N}$, write $f(m) := -1/m + \alpha_{y,N} \int t/(1+tm)\,\mathrm{d}H_N(t)$. By Bai & Silverstein (2010, Lemma 6.1), the conditions $m \in \mathbb{R} \setminus \{0\}$, $-1/m \notin \operatorname{supp} H_N$ and $f'(m) > 0$ imply $f(m) \notin \operatorname{supp} F_N$. Conversely, an exterior point u has $m = m_N^{\det}(u)$ satisfying these conditions and $u = f(m)$ by their (6.1.6). Substituting (E.24),

$$m = -\frac{\alpha_{y,N}}{w}, \qquad f(m) = T_N(w), \qquad \frac{\mathrm{d}w}{\mathrm{d}m} = \frac{\alpha_{y,N}}{m^2} > 0.$$

For $w > \lambda_{\max}(G)$ the population coordinate $-1/m = w/\alpha_{y,N}$ lies above $\operatorname{spec} T$, and $f'(m) > 0$ exactly when $\partial_w T_N(w) > 0$. Thus (E.22) gives

$$T_N((w_{R,N}, \infty)) = (z^2_{\mathrm{edge},N}, \infty) \subset (\operatorname{supp} F_N)^c, \qquad \sup \operatorname{supp} F_N \le z^2_{\mathrm{edge},N}.$$

For the reverse inequality, take $u > \sup \operatorname{supp} F_N$. Then $m_N^{\det}(u)$ is real, negative and increasing. Hence $w(u) := -\alpha_{y,N}/m_N^{\det}(u)$ is continuous and tends to infinity. By Bai & Silverstein (2010, Lemma 6.1), it avoids $\operatorname{spec} G$; continuity in (E.24), whose denominators are nonzero, gives $u = T_N(w(u))$. Continuity from infinity forces $w(u) > \lambda_{\max}(G)$, so

$$u = T_N(w(u)) \ge \min_{w > \lambda_{\max}(G)} T_N(w) = z^2_{\mathrm{edge},N} \qquad (u > \sup \operatorname{supp} F_N).$$

Letting $u \downarrow \sup \operatorname{supp} F_N$ proves the reverse inequality. For $u > z^2_{\mathrm{edge},N}$, $w(u)$ cannot equal $w_{R,N}$ because their images under T_N differ. Again by continuity from infinity, $w(u) > w_{R,N}$, where T_N is strictly increasing, proving uniqueness of the exterior inverse.

For the limiting law the same argument replaces $\lambda_{\max}(G)$ by $w_G = \sup \operatorname{supp} \nu_0$. The set $\operatorname{supp} \nu_0 \cup \{0\}$ is closed; a continuous positive $w(u)$ avoiding it and tending to infinity stays in (w_G, ∞). Thus $\sup \operatorname{supp} F_0 = T_0(w_R)$ and the same branch identification holds. □

Proposition F.3 (Edge of the response noise). *Along the design sequences of Lemma F.1, for every fixed $\varepsilon > 0$: (1) $\mathbb{P}(\sigma_1(E) > z_{\mathrm{edge}} + \varepsilon) \to 0$; (2) $\sigma_2(E) \xrightarrow{P} z_{\mathrm{edge}}$; (3) $\mathbb{P}(\|E\|_{\mathrm{op}} \le K_E) \to 1$.*

Proof. The standing norm bound gives (3). For (1)–(2), Bai & Silverstein (2010, Theorem 6.3(1)), the exclusion theorem of Bai & Silverstein (1998), applies with

$$T = \frac{AA^\top}{\alpha_{y,N}} \in \mathbb{R}^{N\times N}, \qquad Q \in \mathbb{R}^{N\times D_y}, \qquad \frac{N}{D_y} \to \alpha_y,$$
$$\frac{1}{D_y} Q^\top T Q = \frac{1}{N} Q^\top A A^\top Q = E^\top E.$$

The population is deterministic given the design, $\|T\|_{\mathrm{op}} \le K_0/\alpha_{y,N}$, and its ESD converges to the image of $\nu_0 + (1-\alpha_x^{-1})\delta_0$ under $\lambda \mapsto \lambda/\alpha_y$. The entries of Q are iid with mean zero, unit variance and finite fourth moment. The theorem's hypothesis (f) concerns the finite-N law F_N of this companion.

For any fixed small $\varepsilon' > 0$, choose $\delta' > 0$ small and K_2 large. Closure, (E.19) and $T_N(w) \ge w/\alpha_{y,N}$ give, for large N,

$$w_R + \delta' > w_G + \gamma/2 \ge \lambda_{\max}(G), \qquad \inf_{[w_R+\delta', K_2]} \partial_w T_N \ge c/2 > 0,$$
$$T_N(w_R + \delta') \le z_{\mathrm{edge}}^2 + \varepsilon'/2, \qquad T_N(K_2) \ge K_E^2 + 1.$$

For the final inequality, take $K_2 \ge \sup_N \alpha_{y,N}(K_E^2 + 1)$. By Lemma F.2,

$$[z_{\mathrm{edge}}^2 + \varepsilon', K_E^2] \subset \big(T_N(w_R + \delta'), T_N(K_2)\big) \subset (\operatorname{supp} F_N)^c.$$

Under the nested iid coupling of the Q's, the exclusion theorem places no eigenvalues of $E^\top E$ in this closed interval eventually almost surely. Combining with (3), and choosing ε' smaller than $(z_{\mathrm{edge}} + \varepsilon)^2 - z_{\mathrm{edge}}^2$, proves (1). If the target interval is empty, (3) already gives the bound.

By Bai & Silverstein (2010, Theorem 4.1), $\mathrm{ESD}(E^\top E) \Rightarrow F_0$ almost surely. By Lemma F.2, $\sup \operatorname{supp} F_0 = z_{\mathrm{edge}}^2$, so $F_0((z_{\mathrm{edge}}^2 - \delta'', z_{\mathrm{edge}}^2]) > 0$ for every $\delta'' > 0$. Consequently at least two eigenvalues exceed $z_{\mathrm{edge}}^2 - \delta''$ eventually. This lower bound and (1) give $\sigma_2(E) \xrightarrow{P} z_{\mathrm{edge}}$, proving (2). □

Remark F.4 (No outlier without the deterministic equivalent). Hypothesis (b) of Theorem F.10 below follows from its hypothesis (a) alone, for Q and for the Gaussian Q_{G} alike. If $\mathbb{P}(\sigma_1(E) > z_{\mathrm{edge},N} + \varepsilon') \not\to 0$ along a subsequence, extract a further subsequence along which $\nu_N \Rightarrow \nu_*$ (masses $\alpha_{x,N}^{-1} \to \alpha_x^{-1}$, supports in $[0, K_0]$), the aspect ratios converge, and $z_{\mathrm{edge},N} \to z_*$.

The laws F_N have supports in $[0, z_{\mathrm{edge},N}^2]$ (Lemma F.2), so the fixed interval $[z_*^2 + \varepsilon'', K_E^2]$ lies in an open interval outside $\operatorname{supp} F_N$ for all large N along the subsequence, and Bai & Silverstein (2010, Theorem 6.3(1)) with the norm bound (3) gives $\sigma_1(E) \le z_* + \varepsilon''' \le z_{\mathrm{edge},N} + \varepsilon'$ with probability tending to 1 along it, a contradiction.

F.3 Comparison at fixed positive height

Lemma F.5 (Comparison of resolvent bilinear forms). *Let $P_0 \in \mathbb{R}^{d\times m}$ and $B \in \mathbb{R}^{d\times n}$ be deterministic with $d, m, n \le CN$ and $\|B\|_{\mathrm{op}} \le C$, and let $X, Y \in \mathbb{R}^{n\times m}$ have independent entries with $\mathbb{E}X_{kj} = \mathbb{E}Y_{kj} = 0$, $\mathbb{E}X_{kj}^2 = \mathbb{E}Y_{kj}^2$ and $\sup_{k,j}(\mathbb{E}\,|X_{kj}|^p + \mathbb{E}\,|Y_{kj}|^p) \le M_p$ for every fixed p. Put $Z_X := P_0 + BX/\sqrt{N}$, $Z_Y := P_0 + BY/\sqrt{N}$ ($d \times m$), $R_X(z) := (zI - H_{Z_X})^{-1}$, and R_Y likewise. Then for every compact $\mathcal{K} \subset \mathbb{C}^+$,*

$$\sup_{z\in\mathcal{K}} \sup_{\|r\|=\|s\|=1} \big|\mathbb{E}\, r^\top R_X(z) s - \mathbb{E}\, r^\top R_Y(z) s\big| \longrightarrow 0,$$
$$\sup_{z\in\mathcal{K}} \sup_{\|r\|=\|s\|=1} \big|\mathbb{E}\,\big|r^\top R_X(z) s\big|^2 - \mathbb{E}\,\big|r^\top R_Y(z) s\big|^2\,\big| \longrightarrow 0;$$

Proof. Let $\eta := \min_{\mathcal{K}} \operatorname{Im} z$. We prove the bound $C_{\eta,\epsilon} N^{-1/2+\epsilon}$ for every $\epsilon > 0$. Write

$$H(x) = H_{P_0} + \sum_{k,j} x_{kj} F_{kj}, \qquad F_{kj} = p_k q_j^\top + q_j p_k^\top,$$

$$p_k = \frac{(Be_k; 0)}{\sqrt{N}}, \qquad q_j = (0; e_j), \qquad \|p_k\| \le CN^{-1/2}, \quad \|q_j\| = 1,$$

$$\left\| \sum_k p_k p_k^\top \right\|_{\mathrm{op}} = \frac{\|BB^\top\|_{\mathrm{op}}}{N} \le \frac{C^2}{N}, \qquad \left\| \sum_j q_j q_j^\top \right\|_{\mathrm{op}} = 1.$$

Thus $H_{Z_X} = H(X)$. For fixed $z \in \mathcal{K}$ and unit r, s, put $f(x) := r^\top (zI - H(x))^{-1} s$; every real x satisfies $\|(zI - H(x))^{-1}\|_{\mathrm{op}} \le \eta^{-1}$.

Let X^ϑ, $0 \le \vartheta \le 1$, have independent entries with laws $\vartheta \mathcal{L}(X_{kj}) + (1-\vartheta)\mathcal{L}(Y_{kj})$, independently of X, Y. Differentiating the product measure gives

$$\frac{\mathrm{d}}{\mathrm{d}\vartheta} \mathbb{E} f(X^\vartheta) = \sum_\ell \mathbb{E}[f(X^\vartheta_{(\ell)}, X_\ell) - f(X^\vartheta_{(\ell)}, Y_\ell)].$$

Here $(X^\vartheta_{(\ell)}, t)$ replaces entry $\ell = (k, j)$ by t. Set $R := (zI - H(X^\vartheta))^{-1}$, the same matrix for all ℓ. With $f_i := \partial_\ell^i f(X^\vartheta)$, Taylor expansion and independence give

$$\begin{aligned} f(X^\vartheta_{(\ell)}, X_\ell) - f(X^\vartheta_{(\ell)}, Y_\ell) &= (X_\ell - Y_\ell) f_1 \\ &\quad + \tfrac{1}{2}[(X_\ell - X_\ell^\vartheta)^2 - (Y_\ell - X_\ell^\vartheta)^2] f_2 + \rho_\ell, \\ \mathbb{E}[(X_\ell - Y_\ell) f_1] &= 0, \\ \mathbb{E}\{[(X_\ell - X_\ell^\vartheta)^2 - (Y_\ell - X_\ell^\vartheta)^2] f_2\} &= \mathbb{E}[X_\ell^2 - Y_\ell^2]\mathbb{E} f_2 - 2\mathbb{E}[X_\ell - Y_\ell]\mathbb{E}[X_\ell^\vartheta f_2] = 0. \end{aligned}$$

The last two equalities use the matching moments. It remains to bound

$$|\mathbb{E} f(X) - \mathbb{E} f(Y)| \le \sup_\vartheta \sum_\ell \mathbb{E}|\rho_\ell|,$$

$$|\rho_\ell| \le \tfrac{1}{6}(|X_\ell - X_\ell^\vartheta|^3 + |Y_\ell - X_\ell^\vartheta|^3) \sup_t |\partial_\ell^3 f(X^\vartheta_{(\ell)}, t)|,$$

where t lies between X_ℓ^ϑ and X_ℓ or Y_ℓ.

Differentiation gives $\partial_\ell R = R F_\ell R$ and $\partial_\ell^3 f = 6 r^\top R F_\ell R F_\ell R F_\ell R s$. Expanding the three insertions gives eight products

$$(r^\top R a_1)(b_1^\top R a_2)(b_2^\top R a_3)(b_3^\top R s), \qquad \{a_i, b_i\} = \{p_k, q_j\}.$$

Each contains three p_k's and three q_j's. Bound the middle factors by $\|a\|\,\|b\|/\eta$, but keep the end factors $\alpha_k := |r^\top R p_k|$ and $\beta_j := |r^\top R q_j|$, with primes for s. The bounds $\sum_k \alpha_k^2 \le C^2/(N\eta^2)$ and $\sum_j \beta_j^2 \le \eta^{-2}$ also hold for the primed factors. The three endpoint types give

$$\begin{aligned} (p,p): &\quad \sum_{k,j} \alpha_k \alpha'_k \frac{CN^{-1/2}}{\eta^2} \le \frac{C^3 m N^{-3/2}}{\eta^4}, \\ (p,q): &\quad \sum_{k,j} \alpha_k \beta'_j \frac{C^2 N^{-1}}{\eta^2} \le \frac{C^3 \sqrt{mn}\, N^{-3/2}}{\eta^4}, \\ (q,q): &\quad \sum_{k,j} \beta_j \beta'_j \frac{C^3 N^{-3/2}}{\eta^2} \le \frac{C^3 n N^{-3/2}}{\eta^4}. \end{aligned}$$

These are Cauchy–Schwarz bounds; for the mixed case use $\sum_{k,j} \alpha_k \beta'_j \le \sqrt{n \sum_k \alpha_k^2} \sqrt{m \sum_j (\beta'_j)^2}$. The (q,p) case is identical. Since $m, n \le CN$, their sum is $O_\eta(N^{-1/2})$.

Put $R^{(t)} := (zI - H(X^\vartheta) - (t - X_\ell^\vartheta)F_\ell)^{-1}$. The identity $R^{(t)} = R + (t - X_\ell^\vartheta)RF_\ell R^{(t)}$ gives

$$|r^\top R^{(t)} p_k| \le \alpha_k(1 + |t - X_\ell^\vartheta| CN^{-1/2}\eta^{-1}) + |t - X_\ell^\vartheta| C^2 N^{-1}\eta^{-1}\beta_j,$$
$$|r^\top R^{(t)} q_j| \le \beta_j(1 + |t - X_\ell^\vartheta| CN^{-1/2}\eta^{-1}) + |t - X_\ell^\vartheta|\eta^{-1}\alpha_k.$$

The same bounds hold at the other end, and the middle factors retain their norm bounds. Expanding the two endpoint bounds gives a polynomial of degree at most two in $|t - X_\ell^\vartheta|$; exchanging an endpoint type gives one of the three sums above with the accompanying powers of $N^{-1/2}$. Thus

$$|\partial_\ell^3 f(X_{(\ell)}^\vartheta, t)| \le (1 + |t - X_\ell^\vartheta|)^2 \Xi_\ell, \qquad \sum_\ell \Xi_\ell \le C_\eta N^{-1/2}$$

for every realization.

The cubic Taylor factor and the quadratic perturbation factor give

$$\sum_\ell \mathbb{E}|\rho_\ell| \le C_\eta N^{-1/2}\, \mathbb{E} \max_{\ell \le nm} (1 + |X_\ell| + |Y_\ell| + |X_\ell^\vartheta|)^5$$
$$\le C_{\eta,\epsilon} N^{-1/2} (nm)^\epsilon \le C_{\eta,\epsilon} N^{-1/2+2\epsilon}.$$

The middle inequality bounds the maximum by an $L^{1/\epsilon}$ norm and uses the uniform fixed-moment bounds. Replacing ϵ by $\epsilon/2$ gives the announced bound.

Apply the same interpolation to $f\bar{f}$. The product rule gives

$$\partial_\ell^3(f\bar{f}) = \bar{f}\partial_\ell^3 f + 3\partial_\ell \bar{f}\, \partial_\ell^2 f + 3\partial_\ell^2 \bar{f}\, \partial_\ell f + f\partial_\ell^3 \bar{f}.$$

For the outer terms, $|f| \le \eta^{-1}$ and the third-derivative and remainder estimates above apply. For the middle terms,

$$|\partial_\ell^2 f| \le 8 \|p_k\|^2 \|q_j\|^2 \eta^{-3} \le 8C^2 N^{-1}\eta^{-3},$$
$$\sum_{k,j} (\alpha_k \beta_j' + \beta_j \alpha_k') \le 2\sqrt{mn}\, CN^{-1/2}\eta^{-2}.$$

Their product is $O_\eta(N^{-1/2})$. The perturbed-point and weight estimates are unchanged, up to additional inverse-height factors. All bounds depend on z only through η and are uniform over the unit directions, proving both claims. □

Corollary F.6 (Directional L^2 convergence and traces)**.** *In the setting of Lemma F.5, let $M_N(z)$ be deterministic, complex symmetric, and uniformly bounded on $\mathcal{K}$, and let δ_N be the supremum over $z \in \mathcal{K}$ and unit r, s of $\left\|r^\top R_Y s - r^\top M_N s\right\|_{L^2}$, with $\delta_N \to 0$. Then $\sup_{\mathcal{K}} \sup_{\|r\|=\|s\|=1} \left\|r^\top R_X s - r^\top M_N s\right\|_{L^2} \to 0$, and for every deterministic real Ω with $\|\Omega\|_{\mathrm{op}} \le C$,*

$$\sup_{\mathcal{K}} \left\|N^{-1} \operatorname{Tr}\, \Omega\,(R_X - M_N)\right\|_{L^2} \le \frac{d+m}{N}\, \|\Omega\|_{\mathrm{op}} \sup_{\mathcal{K}} \sup_{\|u\|=1} \left\|u^\top (R_X - M_N) u\right\|_{L^2} \longrightarrow 0.$$

Proof. For the directional conclusion, expand

$$\mathbb{E}|r^\top R_X s - r^\top M_N s|^2 = \mathbb{E}|r^\top R_X s|^2 - 2\operatorname{Re}\left(\overline{r^\top M_N s}\, \mathbb{E} r^\top R_X s\right) + |r^\top M_N s|^2.$$

By Lemma F.5, the first two expectations differ from their Y-versions by $o(1)$ uniformly. Since $|r^\top M_N s| \le \|M_N\|_{\mathrm{op}}$, the whole expression is at most $\delta_N^2 + o(1)$.

For the trace, complex symmetry of $R_X - M_N$ gives $\operatorname{Tr}\Omega(R_X - M_N) = \operatorname{Tr}\Omega_{\mathrm{sym}}(R_X - M_N)$. Diagonalize $\Omega_{\mathrm{sym}} = \sum_i \tau_i u_i u_i^\top$. Minkowski's inequality yields

$$\left\|N^{-1} \operatorname{Tr}\Omega(R_X - M_N)\right\|_{L^2} \le \frac{1}{N} \sum_i |\tau_i| \left\|u_i{}^\top (R_X - M_N) u_i\right\|_{L^2}$$
$$\le \frac{d+m}{N} \|\Omega\|_{\mathrm{op}} \sup_{\|u\|=1} \left\|u^\top (R_X - M_N) u\right\|_{L^2}.$$

The supremum is over deterministic directions outside the expectation. □

F.4 THE GAUSSIAN RESPONSE CALCULATION

Lemma F.7 (Gaussian response model)**.** *Assume* $\|G\|_{\mathrm{op}} \le K_0$ *and* $G \neq 0$*. Let* Q_{G} *have iid* $\mathcal{N}(0,1)$ *entries,* $E_{\mathrm{G}} := A^\top Q_{\mathrm{G}}/\sqrt{N}$*, and* $R := R_{E_{\mathrm{G}}}(z)$*. Then for every compact* $\mathcal{K} \subset \mathbb{C}^{++}$*, with* $\eta := \min_{\mathcal{K}} \operatorname{Im} z$*,*

$$\sup_{\mathcal{K}} \sup_{\|r\|=\|s\|=1} \operatorname{Var}\left(r^\top R s\right) \le \frac{4K_0}{N\eta^4}, \qquad \sup_{\mathcal{K}} \left\|\mathbb{E}R - \mathbf{R}_N^{\det}(z)\right\|_{\mathrm{op}} \longrightarrow 0,$$

with $\mathbf{R}_N^{\det}$ *the deterministic equivalent of Appendix E.5. Consequently, uniformly on* $\mathcal{K}$ *and over unit* r, s*,* $\mathbb{E}\left|r^\top R s - r^\top \mathbf{R}_N^{\det} s\right|^2 \to 0$*.*

Proof. Write $y_k := A^\top e_k/\sqrt{N} \in \mathbb{R}^{D_x}$, $\tilde{y}_k := (y_k; 0)$ and $\tilde{f}_j := (0; e_j) \in \mathbb{R}^{D_x+D_y}$. Then

$$H_{E_{\mathrm{G}}} = \sum_{k,j} Q_{kj} F_{kj}, \qquad F_{kj} = \tilde{y}_k \tilde{f}_j^\top + \tilde{f}_j \tilde{y}_k^\top, \qquad \sum_k \tilde{y}_k \tilde{y}_k^\top = \operatorname{diag}(G/N, 0).$$

Differentiate, use $\|R\|_{\mathrm{op}} \le \eta^{-1}$, and sum with the preceding identity:

$$\partial_{Q_{kj}} r^\top R s = r^\top R \tilde{y}_k \tilde{f}_j{}^\top R s + r^\top R \tilde{f}_j \tilde{y}_k{}^\top R s,$$

$$\sum_k |r^\top R \tilde{y}_k|^2 \le \frac{K_0}{N\eta^2}, \qquad \sum_j |\tilde{f}_j{}^\top R s|^2 \le \eta^{-2},$$

$$\sum_{k,j} |\partial_{Q_{kj}} r^\top R s|^2 \le \frac{4K_0}{N\eta^4}, \qquad \operatorname{Var}(r^\top R s) \le \frac{4K_0}{N\eta^4}.$$

The last inequality is Gaussian Poincaré; linearity of $H_{E_{\mathrm{G}}}$ makes the derivative bound global. Put $\hat{t}_N := N^{-1} \operatorname{Tr} R_{22}$, $t_N := \mathbb{E}\hat{t}_N$ and $m_N^{\mathrm{G}}(\zeta) := D_y^{-1} \operatorname{Tr}(E_{\mathrm{G}}^\top E_{\mathrm{G}} - \zeta I)^{-1}$. Minkowski's inequality and (F.1) give

$$\left\|\hat{t}_N - t_N\right\|_{L^2} \le \frac{D_y}{N} \sup_{\|u\|=1} \operatorname{Var}(u^\top R u)^{1/2} = O(N^{-1/2}), \qquad t_N = -\alpha_{y,N}^{-1} z\, \mathbb{E} m_N^{\mathrm{G}}(z^2).$$

For orthogonal $O \in \mathbb{R}^{D_y \times D_y}$, $Q_{\mathrm{G}} O \overset{d}{=} Q_{\mathrm{G}}$ and $H_{E_{\mathrm{G}}O} = \operatorname{diag}(I, O^\top) H_{E_{\mathrm{G}}} \operatorname{diag}(I, O)$. Thus $\mathbb{E}R_{22}$ commutes with every O, while $\mathbb{E}R_{12} = \mathbb{E}R_{12}O$. Taking traces gives $\mathbb{E}R_{22} = \alpha_{y,N} t_N I_{D_y}$, and $O = -I$ gives $\mathbb{E}R_{12} = 0$.

The block equation is $zR_{11} - E_{\mathrm{G}} R_{21} = I$. Resolvent differentiation and Gaussian integration by parts give

$$(\partial_{Q_{kj}} R)_{21} = (R_{21} y_k)(e_j^\top R_{21}) + (R_{22} e_j)(y_k^\top R_{11}),$$

$$\mathbb{E}[E_{\mathrm{G}} R_{21}] = \sum_{k,j} y_k \mathbb{E}[\partial_{Q_{kj}} (R_{21})_{j\cdot}] = G\mathbb{E}[\hat{t}_N R_{11}] + \frac{G}{N} \mathbb{E}[R_{12} R_{21}].$$

Center both factors in the covariance. For real unit $u, v \in \mathbb{R}^{D_x}$,

$$\begin{aligned} |u^\top \mathbb{E}[(\hat{t}_N - t_N) R_{11}] v| &= \left|\mathbb{E}[(\hat{t}_N - t_N)(u^\top R_{11} v - \mathbb{E} u^\top R_{11} v)]\right| \\ &\le \left\|\hat{t}_N - t_N\right\|_{L^2} \operatorname{Var}(u^\top R_{11} v)^{1/2} = O(N^{-1}). \end{aligned}$$

Real unit tests control the complex operator norm up to a factor 2. Also $\|(G/N) R_{12} R_{21}\|_{\mathrm{op}} \le K_0/(N\eta^2)$. Hence

$$(zI - t_N G)\mathbb{E}R_{11} = I + O(N^{-1}), \qquad \left\|(zI - t_N G)^{-1}\right\|_{\mathrm{op}} \le \eta^{-1},$$

$$\mathbb{E}R_{11} = (zI - t_N G)^{-1} + O(N^{-1}).$$

The inverse bound follows from $\operatorname{Im} t_N \le 0$ and $G \succeq 0$.

Suppose $\sup_{\mathcal{K}} |t_N - c_N| \not\to 0$. Pass to a subsequence with $z_N \to z_* \in \mathcal{K}$ and $|t_N(z_N) - c_N(z_N)| \ge \delta > 0$. The measures ν_N have bounded support and masses $\alpha_{x,N}^{-1} \to \alpha_x^{-1}$, so a further subsequence

satisfies $\nu_N \Rightarrow \nu_*$. Apply Bai & Silverstein (2010, Theorem 4.1) to $E_{\mathrm{G}}^\top E_{\mathrm{G}} = D_y^{-1} Q_{\mathrm{G}}^\top T Q_{\mathrm{G}}$ with $T = AA^\top/\alpha_{y,N}$. It gives $m_N^{\mathrm{G}}(\zeta) \to m_*(\zeta)$ almost surely, where m_* is the unique solution in $\mathbb{C}^+$ of (E.25) with ν_* replacing ν_0 (Bai & Silverstein, 2010, (6.1.5)). Bounded convergence and equicontinuity give $\mathbb{E}m_N^{\mathrm{G}} \to m_*$ locally uniformly on $\mathbb{C}^+$.

The laws with transforms $m_N^{\det}$ have common compact support. Every subsequential limit is therefore a probability Stieltjes transform, with $\operatorname{Im} m_N^{\det}(\zeta) \geq \operatorname{Im}\zeta/(C + |\zeta|)^2 > 0$. This keeps $1 + \lambda m_N^{\det}/\alpha_{y,N}$ away from zero for $\lambda \in [0, K_0]$. Passing to the limit in (E.24) identifies the same unique solution m_*. Thus $m_N^{\det} \to m_*$ locally uniformly and

$$t_N - c_N = -\alpha_{y,N}^{-1} z\big(\mathbb{E}m_N^{\mathrm{G}} - m_N^{\det}\big)(z^2) \longrightarrow 0$$

uniformly near z_*, a contradiction.

By the resolvent identity and the bounds from Appendix E.5,

$$\big\|(zI - t_N G)^{-1} - (zI - c_N G)^{-1}\big\|_{\mathrm{op}} \leq K_0 \eta^{-2} |t_N - c_N| \longrightarrow 0.$$

Together with $\alpha_{y,N}|t_N - c_N| \to 0$ and the symmetry and mean-block identities above, this gives $\big\|\mathbb{E}R - \mathbf{R}_N^{\det}\big\|_{\mathrm{op}} \to 0$. The variance bound then gives the stated directional L^2 convergence. □

Proposition F.8 (Masked model at fixed height)**.** *Assume $\|G\|_{\mathrm{op}} \leq K_0$ and $G \neq 0$. For every compact $\mathcal{K} \subset \mathbb{C}^{++}$,*

$$\sup_{z \in \mathcal{K}} \sup_{\|r\|=\|s\|=1} \mathbb{E}\big|r^\top R_E(z) s - r^\top \mathbf{R}_N^{\det}(z) s\big|^2 \longrightarrow 0;$$

in particular, uniformly on $\mathcal{K}$, $m_{1,E}, m_{2,E}, m_{\times,E}, g_E, g_{\times,E}$ are within $o_P(1)$ of $\bar{m}_1, \bar{m}_2, \bar{m}_1^{\times}, 0, 0$, and $\mathbf{u}_0{}^\top R_E \mathbf{u}_0$ is within $o_P(1)$ of $c_N^{-1} u_0{}^\top (w_N I - G)^{-1} u_0$.

Proof. Lemma F.5 with $P_0 = 0$, $B = A^\top$ ($\|B\|_{\mathrm{op}} \leq \sqrt{K_0}$), $X = Q$, $Y = Q_{\mathrm{G}}$, together with Lemma F.7 and Corollary F.6 ($\mathbf{R}_N^{\det}$ is complex symmetric and bounded on $\mathcal{K}$ by Appendix E.5), gives the display; the entries of $\mathbf{R}_N^{\det}$ in the three directions are those listed in Appendix E.5. A finite net of $\mathcal{K}$ with the Lipschitz bounds $\|R_E(z) - R_E(z')\|_{\mathrm{op}} \leq |z - z'|\,\eta^{-2}$ and $\big\|\mathbf{R}_N^{\det}(z) - \mathbf{R}_N^{\det}(z')\big\|_{\mathrm{op}} \leq C_\eta\,|z - z'|$ turns the pointwise L^2 convergence into uniform convergence in probability. □

F.5 Continuation to the real exterior

Proposition F.9 (Uniform convergence on a rectangle)**.** *Assume $\|G\|_{\mathrm{op}} \leq K_0$, $G \neq 0$ and $s_N \leq C_s$. Fix $\varepsilon > 0$ and $K_1 < \infty$, put $\eta_1 := \eta(\varepsilon/2, K_1 + 1)$ from* (E.23)*, and let*

$$\mathcal{S}_N := \{z :\ z_{\mathrm{edge},N} + \varepsilon/2 \leq \operatorname{Re} z \leq K_1 + 1,\ |\operatorname{Im} z| \leq \eta_1\},$$
$$\mathcal{S}_N^\circ := \{z :\ z_{\mathrm{edge},N} + 3\varepsilon/4 \leq \operatorname{Re} z \leq K_1 + \tfrac{1}{2},\ |\operatorname{Im} z| \leq \eta_1/2\}.$$

Assume $\mathbb{P}(\sigma_1(E) \leq z_{\mathrm{edge},N} + \varepsilon/4) \to 1$. Then

$$\sup_{z \in \mathcal{S}_N} \sup_{r,s \in \{\hat{\mathbf{a}}_N, \mathbf{u}_0, \mathbf{v}_0\}} \big|r^\top R_E(z) s - r^\top \mathbf{R}_N^{\det}(z) s\big| \xrightarrow{P} 0, \qquad \sup_{\mathcal{S}_N} |F_E - f_N| \xrightarrow{P} 0,$$

and the same holds for all z-derivatives on $\mathcal{S}_N^\circ$. Under Assumption E.10 the finite-N objects other than the $(\mathbf{u}_0, \mathbf{u}_0)$ entry may be replaced by their limits, and the rectangles by the fixed ones with left endpoints $z_{\mathrm{edge}} + \varepsilon$ and $z_{\mathrm{edge}} + \frac{5}{4}\varepsilon$; the $(\mathbf{u}_0, \mathbf{u}_0)$ entry stays bounded by $\big|c_N^{-1}\big| \,/\, \operatorname{dist}(w_N, \operatorname{spec} G)$.

Proof. On the confinement event, $\operatorname{spec}(H_E) \subset [-z_{\mathrm{edge},N} - \varepsilon/4, z_{\mathrm{edge},N} + \varepsilon/4]$. By Appendix E.5, with parameters $\varepsilon/2, K_1 + 1$, $\mathbf{R}_N^{\det}$ is analytic on $\mathcal{S}_N$. Thus

$$\sup_{\mathcal{S}_N} \|R_E\|_{\mathrm{op}} \leq 4/\varepsilon, \qquad \sup_{\mathcal{S}_N} \big(\big\|\mathbf{R}_N^{\det}\big\|_{\mathrm{op}} + \big\|(\mathbf{R}_N^{\det})'\big\|_{\mathrm{op}}\big) \leq C.$$

Fix $0 < \eta \le \eta_1$. For $z = x + iy \in \mathcal{S}_N$ with $0 \le y \le \eta$, use $R_E(z) - R_E(x + i\eta) = i(\eta - y)R_E(z)R_E(x + i\eta)$. For the stated unit directions,

$$|r^\top (R_E(z) - R_E(x + i\eta))s| \le 16\eta/\varepsilon^2,$$
$$|r^\top (\mathbf{R}_N^{\det}(z) - \mathbf{R}_N^{\det}(x + i\eta))s| \le C\eta.$$

The part with $y \ge \eta$ is already at positive height; negative heights follow by conjugation. Therefore

$$\sup_{\mathcal{S}_N} |r^\top (R_E - \mathbf{R}_N^{\det})s| \le \sup_{\mathcal{S}_N \cap \{\operatorname{Im} z \ge \eta\}} |r^\top (R_E - \mathbf{R}_N^{\det})s| + C'\eta.$$

The supremum on the right lies in the fixed compact $\{\varepsilon/2 \le \operatorname{Re} z \le K_1 + 1,\ \eta \le \operatorname{Im} z \le \eta_1\} \subset \mathbb{C}^{++}$. Apply Proposition F.8, first letting $N \to \infty$ and then $\eta \downarrow 0$. All stated entries converge uniformly in probability, and their bounded polynomial combination gives $F_E - f_N \xrightarrow{P} 0$ uniformly as well.

Since $\operatorname{dist}(\mathcal{S}_N^\circ, \partial\mathcal{S}_N) \ge \min\{\varepsilon/4, 1/2, \eta_1/2\} > 0$, Cauchy's formula gives convergence of each fixed derivative there. Under the deterministic equivalent, the inputs in Appendix F.1 give convergence of $\bar{m}_1, \bar{m}_2, \bar{m}_1^{\times}, f_N$ and their derivatives on the stated fixed rectangles, which lie inside $\mathcal{S}_N, \mathcal{S}_N^\circ$ for large N. The $(\mathbf{u}_0, \mathbf{u}_0)$ entry has the stated resolvent bound; no limit for it is assumed. □

F.6 The outlier theorem

Theorem F.10 (Conditional outlier and overlaps)**.** *Condition on the design, so that $(A, G, \widehat{a}_N, s_N)$ are deterministic, and assume with constants uniform in N:*

(a) $\|G\|_{\mathrm{op}} \le K_0$, $G \ne 0$, and $0 < s_N \le C_s$;

(b) for every fixed $\varepsilon' > 0$, $\sigma_1(E) \le z_{\mathrm{edge},N} + \varepsilon'$ with conditional probability tending to 1;

(c) the finite-N secular function f_N has a real root $\widehat{z}_N \ge z_{\mathrm{edge},N} + \varepsilon$ for a fixed $\varepsilon > 0$.

Then, with conditional probability tending to 1, $\bar{C}_N$ has exactly one singular value above $z_{\mathrm{edge},N} + \varepsilon/2$; it is simple, equals $\widehat{z}_N + o_P(1)$, and the squared overlaps of its singular pair $(\bar{u}, \bar{v})$ are, up to $o_P(1)$, the residue forms of Proposition E.6 at $\widehat{z}_N$:

$$(\bar{u}^\top \widehat{a}_N)^2 = \frac{-2\bar{m}_1}{s_N^2 \bar{D}'} + o_P(1), \qquad (\bar{v}^\top v_0)^2 = \frac{-2\bar{m}_2}{s_N^2 \bar{D}'} + o_P(1),$$
$$(\bar{u}^\top u_0)^2 = \frac{-2\bar{m}_1}{s_N^2 \bar{D}'} \Big(\frac{\bar{m}_1^{\times}}{\bar{m}_1}\Big)^2 + o_P(1).$$

Proof of Theorem F.10. By Appendix E.5, $f_N'(\widehat{z}_N) \ge 2\widehat{z}_N/s_N^2 \ge c := 2\varepsilon/C_s^2$. Its root bound $\widehat{z}_N^2 \le 2K_0(\alpha_{y,N}^{-1} + \alpha_{x,N}^{-1}) + C_s^2$ gives the fixed cutoff

$$K_1 := 1 + \sup_N \big\{2K_0(\alpha_{y,N}^{-1} + \alpha_{x,N}^{-1}) + C_s^2\big\}^{1/2}, \qquad I := [z_{\mathrm{edge},N} + \varepsilon/2, K_1].$$

It is finite on the compact parameter set, and $\widehat{z}_N \le K_1 - 1$. On I, f_N is real and strictly increasing; the rectangle bounds give $|f_N''| \le M$ for a fixed $M > 0$. With $r := \min\{\varepsilon/4, 1/2, c/(2M)\}$,

$$[\widehat{z}_N - r, \widehat{z}_N + r] \subset [z_{\mathrm{edge},N} + 3\varepsilon/4, K_1 - 1/2] \subset I,$$
$$f_N' \ge c/2 \quad \text{on this interval}, \qquad f_N(\widehat{z}_N - r) \le -cr/2 < 0 < cr/2 \le f_N(\widehat{z}_N + r),$$
$$|f_N(z)| \ge \tfrac{c}{2}|z - \widehat{z}_N| \quad (|z - \widehat{z}_N| \le r,\ z \in \mathbb{R}).$$

By Proposition F.9 and hypothesis (b), with probability tending to 1, $\sup_I |F_E - f_N| < cr/2$ and $\sigma_1(E) \le z_{\mathrm{edge},N} + \varepsilon/4$. On this event F_E is real and continuous on I and has opposite endpoint signs at $\widehat{z}_N \pm r$. It therefore has a real zero $\check{z}$ between them. Since $\check{z} > \sigma_1(E)$, (E.28) makes it a singular value of $\bar{C}_N$. Rank-one interlacing gives $\sigma_2(\bar{C}_N) \le \sigma_1(E) \le z_{\mathrm{edge},N} + \varepsilon/4 < \check{z}$. Thus $\check{z} = \sigma_1(\bar{C}_N)$ is simple and is the only singular value above $z_{\mathrm{edge},N} + \varepsilon/2$. Its displacement obeys

$$\tfrac{c}{2}|\check{z} - \widehat{z}_N| \le |f_N(\check{z})| = |f_N(\check{z}) - F_E(\check{z})| \le \sup_I |F_E - f_N| = o_P(1).$$

By Proposition E.6, the first two squared overlaps at $\check{z}$ are $2m_{1,E}(\check{z})/F'_E(\check{z})$ and $2m_{2,E}(\check{z})/F'_E(\check{z})$; the true-direction overlap is (E.31). On $[\widehat{z}_N - r, \widehat{z}_N + r] \subset \mathcal{S}_N^\circ$, Proposition F.9 compares the five forms and F'_E with $\bar{m}_1, \bar{m}_2, \bar{m}_1^\times, 0, 0$ and f'_N. Uniform derivative bounds and the displacement estimate allow evaluation at $\widehat{z}_N$, at cost $O(|\check{z} - \widehat{z}_N|) = o_P(1)$.

Both denominators stay away from zero. In fact,

$$f'_N(\widehat{z}_N) = -s_N^2 \bar{D}'(\widehat{z}_N) \geq c,$$

$$w_N(\widehat{z}_N) \geq \alpha_{y,N}\varepsilon^2 \quad \Longrightarrow \quad \bar{m}_2(\widehat{z}_N) = \frac{\alpha_{y,N}\widehat{z}_N}{w_N(\widehat{z}_N)} \leq \frac{K_1 - 1}{\varepsilon^2},$$

$$\bar{m}_1(\widehat{z}_N) = \frac{1}{s_N^2 \bar{m}_2(\widehat{z}_N)} \geq \frac{\varepsilon^2}{C_s^2(K_1 - 1)}.$$

Substituting in the residue identities gives all three stated formulas. □

Corollary F.11 (Under the deterministic equivalent)**.** *Let $\theta > \theta_c$ and let $\widehat{z} > z_{\mathrm{edge}}$ be the root of f_0 (Proposition E.7(ii)). Along the design sequences of Lemma F.1, (a)–(c) hold with $\varepsilon := \frac{1}{2}(\widehat{z} - z_{\mathrm{edge}})$ and $\widehat{z}_N \to \widehat{z}$; consequently $\sigma_1(\bar{C}_N) \xrightarrow{P} \widehat{z}$, $\sigma_1(\bar{C}_N) - \sigma_2(\bar{C}_N) \geq \varepsilon/2$ with probability tending to* 1*, and the overlaps converge to* (E.38)*, that is, to the curves* (E.39)*.*

Proof. Hypothesis (a) follows from Lemma F.1 and Proposition C.3: $s_N \to s_0 > 0$ and $\|a_N\|^2 \to n_a > 0$, hence $G \neq 0$ eventually. Hypothesis (b) is Proposition F.3(1) with $z_{\mathrm{edge},N} \to z_{\mathrm{edge}}$.

For (c), the inputs of Appendix F.1 give uniform convergence of f_N and its derivatives near $\widehat{z}$. For every sufficiently small fixed $r > 0$, the signs $f_0(\widehat{z} - r) < 0 < f_0(\widehat{z} + r)$ give $f_N(\widehat{z} - r) < 0 < f_N(\widehat{z} + r)$ for large N. Strict monotonicity gives the unique root $\widehat{z}_N$ between these endpoints. Since r is arbitrary, $\widehat{z}_N \to \widehat{z}$ and $\widehat{z}_N \geq z_{\mathrm{edge},N} + \varepsilon$ eventually. Apply Theorem F.10; its residue forms converge to those of Proposition E.7. □

Remark F.12 (Relation to the replica analysis). The objects of Theorem F.10 are the exact algebraic finite-N quantities of (E.2)–(E.3); the saddle equations of Appendix H are equivalent to $f_N(\widehat{z}_N) = 0$ and the residue forms above, subject to assumptions (R1)–(R5) stated there.

F.7 The subcritical regime

Below threshold there is no separated root, and operator-norm control of $\mathcal{R}_N$ cannot resolve individual eigenvectors at the edge; we control the whole edge spectral subspace of $\bar{C}_N$ and transfer that control through the residual. Throughout, the design is a sequence of Lemma F.1, $\theta < \theta_c$ is fixed, and $C_s \geq \sup_N s_N$.

Fix $0 < \varepsilon \leq 1$, and set $K'_1 := K_1 + C_s + 1$ using the cutoff from Theorem F.10. Since $z_{\mathrm{edge},N} \leq K_1 - 1$, its limit has the same bound. On the event $\sigma_1(E) \leq z_{\mathrm{edge}} + \varepsilon/4$ and $s_N \leq C_s$, we have $\|\bar{C}_N\|_{\mathrm{op}} \leq \sigma_1(E) + s_N \leq z_{\mathrm{edge}} + \varepsilon/4 + C_s \leq K'_1 - 1$. Every singular value above $z_{\mathrm{edge}} + 2\varepsilon$ is therefore a real zero of F_E in the fixed interval $[z_{\mathrm{edge}} + 2\varepsilon, K'_1 - 1]$, by (E.28). On that interval, Proposition F.9 gives $\sup |F_E - f_0| \xrightarrow{P} 0$, while monotonicity and (E.40) give

$$f_0(z) \geq 1 - s_0^2 \bar{D}_0(z_{\mathrm{edge}}) = 1 - \theta^2/\theta_c^2 > 0.$$

Thus there is no such zero with probability tending to 1. For the lower bound, the spike vanishes on $v_0^\perp$, so Courant–Fischer and Proposition F.3(2) give

$$\sigma_1(\bar{C}_N) \geq \sup_{\varphi \perp v_0,\, \|\varphi\| = 1} \|E\varphi\| \geq \sigma_2(E) \xrightarrow{P} z_{\mathrm{edge}}.$$

Since ε was arbitrary, $\sigma_1(\bar{C}_N) \xrightarrow{P} z_{\mathrm{edge}}$.

For $\mathbf{r} \in \{\hat{\mathbf{a}}_N, \mathbf{u}_0, \mathbf{v}_0\}$, let $\mu_{\mathbf{r},N}(\mathcal{I}) := \mathbf{r}^\top \mathbf{1}_{\mathcal{I}}(H_{\bar{C}_N})\mathbf{r}$. We will prove

$$\lim_{\varsigma \downarrow 0} \limsup_{N \to \infty} \mathbb{P}\Big(\mu_{\mathbf{r},N}\big([z_{\mathrm{edge}} - \varsigma, \infty)\big) > \epsilon\Big) = 0 \qquad \text{for every } \epsilon > 0. \tag{F.2}$$

With U, J from the proof of Proposition E.5, Woodbury's identity gives

$$\mathbf{r}^\top R_{\bar{C}_N}(z)\mathbf{r} = \mathbf{r}^\top R_E \mathbf{r} + s_N(\mathbf{r}^\top R_E U)J(I_2 - s_N U^\top R_E U J)^{-1}(U^\top R_E \mathbf{r}),$$
$$\det(I_2 - s_N U^\top R_E U J) = F_E(z).$$

It suffices to bound this expression at $z_\varsigma := z_{\text{edge}} + i\varsigma$, uniformly for small fixed ς after taking $N \to \infty$.

First fix $\varsigma > 0$. By Proposition F.8 and the inputs of Appendix F.1,

$$U^\top R_E(z_\varsigma)U \xrightarrow{P} \begin{pmatrix} m_1(z_\varsigma) & 0 \\ 0 & m_2(z_\varsigma) \end{pmatrix}, \qquad F_E(z_\varsigma) \xrightarrow{P} f_0(z_\varsigma).$$

The true-direction entries obey

$$\mathbf{u}_0{}^\top R_E \mathbf{u}_0 = c_N^{-1} u_0{}^\top (w_N I - G)^{-1} u_0 + o_P(1),$$
$$\mathbf{u}_0^\top R_E U = \left[c_N^{-1} u_0{}^\top (w_N I - G)^{-1}\widehat{a}_N,\ 0\right] + o_P(1),$$
$$w_N = w_N(z_\varsigma) \to w_0(z_\varsigma), \qquad c_N^{-1} \to w_0(z_\varsigma)/z_\varsigma.$$

To control these limits as $\varsigma \downarrow 0$, use the fold at w_R. The map T_0 is real analytic there; (E.18) gives $T_0(w) - z_{\text{edge}}^2 = \chi(w)^2$ locally, with χ conformal, $\chi(w_R) = 0$ and $\chi'(w_R) > 0$. For $\zeta \in \mathbb{C}^+$ close to the edge, the two nearby inverse values have imaginary parts of opposite sign. The upper one gives $m = -\alpha_y/w \in \mathbb{C}^+$ solving (E.25); uniqueness in Bai & Silverstein (2010, (6.1.5)) identifies it with $m_0(\zeta)$. Consequently,

$$w_0(z_\varsigma) \longrightarrow w_R,$$
$$f_0(z_\varsigma) = 1 - s_0^2 \alpha_y \Phi_0(w_0(z_\varsigma)) \longrightarrow 1 - s_0^2 \bar{D}_0(z_{\text{edge}}) = 1 - \theta^2/\theta_c^2 > 0.$$

The last identity is (E.40). Choose $\varsigma_0 > 0$ so that, for $0 < \varsigma \le \varsigma_0$,

$$|w_0(z_\varsigma) - w_R| \le \gamma/4, \quad |w_0(z_\varsigma)/z_\varsigma| \le \frac{3w_R}{2z_{\text{edge}}}, \quad |f_0(z_\varsigma)| \ge \tfrac{1}{2}(1 - \theta^2/\theta_c^2),$$

and $w_0(z_\varsigma) \in U_K$ for $K = [w_R - \gamma/2, w_R + \gamma/2]$ in Assumption E.10.

For each such fixed ς and $N \ge N_0(\varsigma)$, take $|w_N - w_0(z_\varsigma)| \le \min\{\gamma/4, w_R/2\}$. On the confinement event, (E.19) gives

$$\lambda_{\max}(G) \le w_G + \gamma/4, \qquad \operatorname{dist}(w_N, \operatorname{spec} G) \ge \gamma/4, \qquad |c_N^{-1}| \le \frac{2w_R}{z_{\text{edge}}},$$
$$\left|c_N^{-1} u_0{}^\top (w_N I - G)^{-1} u'\right| \le \frac{|c_N^{-1}|}{\operatorname{dist}(w_N, \operatorname{spec} G)} \le \frac{8w_R}{\gamma z_{\text{edge}}} \quad (u' \in \{u_0, \widehat{a}_N\}).$$

The remaining directional limits are also bounded, and $|F_E(z_\varsigma)| \ge \frac{1}{4}(1 - \theta^2/\theta_c^2)$ with probability tending to 1. The 2×2 inverse in Woodbury's formula is bounded by $C/|F_E|$. We have therefore obtained constants $\varsigma_0, C_* > 0$, independent of the chosen ς, such that

$$\mathbb{P}\big(|\mathbf{r}^\top R_{\bar{C}_N}(z_\varsigma)\mathbf{r}| \le C_*\big) \longrightarrow 1 \quad \text{for each fixed } 0 < \varsigma \le \varsigma_0.$$

For $|\lambda - z_{\text{edge}}| \le \varsigma$, the Poisson kernel obeys $\varsigma/((\lambda - z_{\text{edge}})^2 + \varsigma^2) \ge 1/(2\varsigma)$. Integrating gives

$$\mu_{\mathbf{r},N}([z_{\text{edge}} - \varsigma, z_{\text{edge}} + \varsigma]) \le -2\varsigma \operatorname{Im} \mathbf{r}^\top R_{\bar{C}_N}(z_\varsigma)\mathbf{r} \le 2\varsigma |\mathbf{r}^\top R_{\bar{C}_N}(z_\varsigma)\mathbf{r}|.$$

The convergence $\sigma_1(\bar{C}_N) \xrightarrow{P} z_{\text{edge}}$ excludes spectrum above $z_{\text{edge}} + \varsigma$ with probability tending to 1. Hence, for each fixed $\varsigma \le \varsigma_0$, the mass of $[z_{\text{edge}} - \varsigma, \infty)$ is at most $2\varsigma C_*$ with probability tending to 1. First take $N \to \infty$, then $\varsigma \downarrow 0$; choosing $\varsigma < \epsilon/(2C_*)$ proves (F.2).

Put $P_{N,\varsigma} := \mathbf{1}_{[z_{\text{edge}} - \varsigma, \infty)}(H_{\bar{C}_N})$, and let $x_N = 2^{-1/2}(\widehat{u}; \widehat{v})$ be a unit eigenvector of $H_C = H_{\bar{C}_N} + H_{\mathcal{R}_N}$ for $\lambda_N := \sigma_1(C)$. Weyl's inequality gives $|\lambda_N - \sigma_1(\bar{C}_N)| \le \|\mathcal{R}_N\|_{\text{op}} = o_P(1)$, hence

$\lambda_N \xrightarrow{P} z_{\mathrm{edge}}$. On $\{\lambda_N \geq z_{\mathrm{edge}} - \varsigma/2\}$, the gap on $\mathrm{range}(I - P_{N,\varsigma})$ is at least $\varsigma/2$. Projecting the eigenvalue equation therefore gives

$$(\lambda_N - H_{\bar{C}_N})(I - P_{N,\varsigma})x_N = (I - P_{N,\varsigma})H_{\mathcal{R}_N}x_N,$$

$$\|(I - P_{N,\varsigma})x_N\| \leq \frac{2\,\|\mathcal{R}_N\|_{\mathrm{op}}}{\varsigma},$$

$$|\mathbf{r}^\top x_N| \leq \|P_{N,\varsigma}\mathbf{r}\| + \frac{2\,\|\mathcal{R}_N\|_{\mathrm{op}}}{\varsigma}, \qquad \|P_{N,\varsigma}\mathbf{r}\|^2 = \mu_{\mathbf{r},N}([z_{\mathrm{edge}} - \varsigma, \infty)).$$

Let $N \to \infty$ and then $\varsigma \downarrow 0$, using (F.2). Finally,

$$(u_0{}^\top \widehat{u})^2 = 2|\mathbf{u}_0{}^\top x_N|^2, \qquad (\widehat{a}_N{}^\top \widehat{u})^2 = 2|\hat{\mathbf{a}}_N{}^\top x_N|^2,$$
$$(v_0{}^\top \widehat{v})^2 = 2|\mathbf{v}_0{}^\top x_N|^2.$$

Thus all three overlaps vanish for every leading unit singular pair of C.

F.8 Completion of the proof of Theorem 3.4

Proof of Theorem 3.4. Below threshold, Appendix F.7 gives the edge limit and all three vanishing overlaps. Weyl's inequality transfers the singular-value limit, since $|\sigma_1(C) - \sigma_1(\bar{C}_N)| \leq \|\mathcal{R}_N\|_{\mathrm{op}} = o_P(1)$. Above threshold, Corollary F.11 gives the outlier, its overlap limits and a fixed positive gap. Weyl's inequality again transfers the singular value, while Wedin's theorem (Wedin, 1972) gives

$$\max\{\sin\angle(\widehat{u}, \bar{u}), \sin\angle(\widehat{v}, \bar{v})\} \leq \frac{2\,\|\mathcal{R}_N\|_{\mathrm{op}}}{\mathrm{gap}} = o_P(1).$$

Thus the overlap limits transfer to C as well. The reduction in Lemma F.1 removes the conditioning and proves Theorem 3.4 for ρ_x in a compact subset of $(0, 1)$. □

Unmasked design: $\rho_x = 1$. Here $G = I_{D_x}$, $a_N = u_0$, $n_a = 1$, $s_N = \theta\sqrt{\rho_y}$, $\nu_N = \alpha_{x,N}^{-1}\delta_1 \Rightarrow \alpha_x^{-1}\delta_1 = \nu_0$, and Assumption E.10 holds by its $\rho_x = 1$ clause. Everything above applies with $T_0(w) = w/\alpha_y + w/(\alpha_x(w-1))$, $w_R - 1 = \sqrt{\alpha_y/\alpha_x}$, $z_{\mathrm{edge}} = \alpha_y^{-1/2} + \alpha_x^{-1/2}$, the closed forms (E.41)–(E.42) of Proposition E.9(i), and $w_R - w_G = \sqrt{\alpha_y/\alpha_x}$ in place of γ in Appendix F.7.

F.9 The Gaussian ensemble

With $Q_{\mathrm{G}}, W_{\mathrm{G}}$ standard Gaussian of sizes $N \times D_y$ and $D_x \times D_y$, independent of (S_x, S_y, Z), and conditionally on A, $A^\top Q_{\mathrm{G}}/\sqrt{N} \overset{d}{=} G^{1/2}W_{\mathrm{G}}/\sqrt{N}$, so the model of Lemma F.7 is

$$C_{\mathrm{G}} := s_N\,\widehat{a}_N v_0^\top + \frac{1}{\sqrt{N}}\,G^{1/2}W_{\mathrm{G}}. \tag{F.3}$$

At $\rho_y = 1$ it has exactly the conditional law of $\bar{C}_N$ (Corollary B.2); for $\rho_y < 1$ it matches the conditional covariance. Under hypotheses (a) and (c) of Theorem F.10, hypothesis (b) holds for Q_{G} by the remark after Proposition F.3, and the rectangle continuation and the sign-change argument apply to E_{G} without change, so C_{G} has the same conditional outlier and residue-form overlaps up to $o_P(1)$. Fixed-height comparison alone does not suffice; spectral exclusion and the exterior argument are needed. The finite-N laws for $\rho_y < 1$ are not identified; Appendix H analyzes this ensemble.

G NESTED FLAT ORTHOGONAL DESIGNS

This appendix verifies Assumption E.10 for the nested flat orthogonal class of Corollary 3.6: $|(X_\star)_{ki}| = 1$ and $X_\star^\top X_\star = N I_{D_x}$, the designs being the $N \times D_x$ corners of one fixed sign array, with the Sylvester corners at $N = 2^m$ as the standard example, up to row and column signed permutations with u_0 transformed accordingly, and satisfying Assumption C.1. Flatness gives the comparison at fixed positive height (Appendix G.2, Lemma F.5); nesting within one fixed array enters through the spectral confinement of Appendix G.3; Appendix G.4 continues the estimates to the real exterior.

The design is fixed and only the mask S_x is random; the retention $\rho_x \in (0,1)$ and the limit ratio α_x are fixed, and confinement is established at each fixed (α_x, ρ_x) rather than uniformly in the parameters. At $\rho_x = 1$, $G = I_{D_x}$ and there is nothing to prove.

Proposition G.1 (Closure for nested flat orthogonal designs)**.** *Let $(X_\star)$ be a sequence of flat orthogonal designs satisfying Assumption C.1 that is in addition* nested*: after row and column signed permutations, each $X_\star$ is the $N \times D_x$ corner of one fixed infinite unimodular array with orthogonal corners. The standard example is the Sylvester array* (G.1) *at $N = 2^m$. Then the leverage is exactly equal, with $\tau_N = 1/\alpha_{x,N}$, and for every compact $K \subset (w_G, \infty)$ there is a fixed open complex neighborhood $U_K \supset K$, with $\overline{U}_K$ separated from the Gram support, on which*

$$\sup_{w \in U_K} \left\{ |\psi_N(w) - \psi_0(w)| + |\Phi_N(w) - \Phi_0(w)| + \left|\Phi_N^\times(w) - \Phi_0^\times(w)\right| \right\} \xrightarrow{P} 0,$$

and $\lambda_{\max}(G) \le w_G + o_P(1)$, the functionals being those of (E.1) *and the limits those of* (E.9)*. In particular Assumption E.10 holds. At $\rho_x = 1$ the conclusion is immediate from $G = I_{D_x}$.*

G.1 SETTING

We use the design notation of Appendix E.2 and the scalar templates (E.7) and (E.16). By Proposition E.1, flatness gives independent centered entries of Θ, with $\operatorname{Var}\Theta_{ki} = (1-\rho_x)/N$ and $|\Theta_{ki}| \le (N\rho_x)^{-1/2}$.

The trace identity from (E.6) is

$$\psi_N(w) = \frac{1}{N}\big(\xi \operatorname{Tr}\mathbf{R}_{22}(\xi) - D_x\big), \qquad w = \xi^2.$$

It suffices to control $N^{-1}\operatorname{Tr}\mathbf{R}_{22}$, $\hat{b}^\top \mathbf{R}_{11}\hat{b}$ and ${u_0}^\top \mathbf{R}_{21}\hat{b}$ in deterministic directions. The trace and push identities (E.8) recover the three functionals from these forms, with normalization by $\|a_N\|^2$ and $\|a_N\|$ for the directional targets.

G.2 THE GAUSSIAN DESIGN CALCULATION

Lemma G.2 (Gaussian design model)**.** *For the Gaussian benchmark of Appendix E.2, let $0 < \rho_x < 1$ and write $\mathbf{R}_{\mathrm{G}}(\xi) := (\xi I - H_{A_{\mathrm{G}}})^{-1}$.*

1. *For every compact $\mathcal{K} \subset \mathbb{C}^+$, with $\eta := \min_{\mathcal{K}} \operatorname{Im}\xi$: $\operatorname{Im} g_{\mathrm{G},N} \le 0$, $\operatorname{Im} d_1(g_{\mathrm{G},N};\xi) \ge \operatorname{Im}\xi$, $\mathbb{E}\mathbf{R}_{\mathrm{G},22}(\xi) = g_{\mathrm{G},N}(\xi)\, I_{D_x}$ exactly, $\sup_{\mathcal{K}} \|\mathbb{E}\mathbf{R}_{\mathrm{G}} - \mathcal{M}_N(g_{\mathrm{G},N};\xi)\|_{\mathrm{op}} = O(N^{-1})$, and*

$$\sup_{\xi \in \mathcal{K}} \sup_{\|r\|=\|s\|=1} \operatorname{Var}\left(r^\top \mathbf{R}_{\mathrm{G}}(\xi) s\right) \le \frac{4(1-\rho_x)}{N\eta^4}.$$

2. *For every compact $\mathcal{K} \subset \mathbb{C}^{++}$, $g_{\mathrm{G},N} \to g$ uniformly on $\mathcal{K}$, where g is the physical solution of* (E.4) *at ratio α_x; hence $g_{\mathrm{G},N} - g_{0,N} \to 0$, $\mathcal{M}_N(g_{\mathrm{G},N};\xi) - M_{0,N}(\xi) \to 0$ and $M_{0,N}(\xi) - M_0(\xi) \to 0$, uniformly on $\mathcal{K}$ and in operator norm.*

Proof. Write $\Theta_{\mathrm{G}} = \sqrt{1-\rho_x}\, N^{-1/2} Z_{\mathrm{G}}^\top$, where Z_{G} is $D_x \times N$ with iid standard Gaussian entries. Suppress the argument ξ and write $d_1 = d_1(g_{\mathrm{G},N};\xi)$.

For orthogonal $V \in \mathbb{R}^{D_x \times D_x}$, set $U := \rho_x^{-1} P V P^\top + I_N - \Pi$. The identities $P^\top P = \rho_x I_{D_x}$ and $PP^\top = \rho_x \Pi$ give $UU^\top = I_N$ and $UP = PV$. Gaussian bi-invariance therefore gives

$$UA_{\mathrm{G}}V^\top = P + U\Theta_{\mathrm{G}}V^\top \overset{d}{=} A_{\mathrm{G}},$$
$$H_{UA_{\mathrm{G}}V^\top} = \mathrm{diag}(U,V) H_{A_{\mathrm{G}}} \mathrm{diag}(U,V)^\top.$$

Hence $\mathbb{E}\mathbf{R}_{\mathrm{G},22}$ commutes with every V; taking traces gives $\mathbb{E}\mathbf{R}_{\mathrm{G},22} = g_{\mathrm{G},N} I_{D_x}$. The resolvent sign gives $\mathrm{Im}\, g_{\mathrm{G},N} \le 0$ and $\mathrm{Im}\, d_1 \ge \mathrm{Im}\, \xi$.

Using standard basis vectors, differentiation in entry (k,i) and summation over the Gaussian entries gives, for unit r, s,

$$\mathcal{E}_{ki} := \begin{pmatrix} 0 & e_k e_i^\top \\ e_i e_k^\top & 0 \end{pmatrix}, \qquad e_k \in \mathbb{R}^N,\ e_i \in \mathbb{R}^{D_x},$$
$$\partial_{(Z_{\mathrm{G}})_{ik}} r^\top \mathbf{R}_{\mathrm{G}} s = \sqrt{\frac{1-\rho_x}{N}}\, r^\top \mathbf{R}_{\mathrm{G}} \mathcal{E}_{ki} \mathbf{R}_{\mathrm{G}} s,$$
$$\sum_{k,i} \left| r^\top \mathbf{R}_{\mathrm{G}} \mathcal{E}_{ki} \mathbf{R}_{\mathrm{G}} s \right|^2 \le 4 \left\| \mathbf{R}_{\mathrm{G}}^\top r \right\|^2 \|\mathbf{R}_{\mathrm{G}} s\|^2 \le \frac{4}{\eta^4}.$$

Gaussian Poincaré gives the variance bound. Set $\hat{e} := (1-\rho_x) N^{-1} \operatorname{Tr} \mathbf{R}_{\mathrm{G},22}$, so $\mathbb{E}\hat{e} = (1-\rho_x)\alpha_{x,N}^{-1} g_{\mathrm{G},N}$ and $d_1 = \xi - \mathbb{E}\hat{e}$. Minkowski's inequality then gives

$$\|\hat{e} - \mathbb{E}\hat{e}\|_{L^2} \le \frac{(1-\rho_x) D_x}{N} \sup_{\|r\|=1} \mathrm{Var}\left(r^\top \mathbf{R}_{\mathrm{G}} r\right)^{1/2} = O(N^{-1/2}).$$

Differentiate using $\partial_{\Theta_{ki}} \mathbf{R}_{\mathrm{G}} = \mathbf{R}_{\mathrm{G}} \mathcal{E}_{ki} \mathbf{R}_{\mathrm{G}}$. Multiply by the entry variance and sum over i; symmetry $\mathbf{R}_{\mathrm{G},12} = \mathbf{R}_{\mathrm{G},21}^\top$ identifies the product term. Repeating for the lower-right block gives the last identity:

$$\partial_{\Theta_{ki}} (\mathbf{R}_{\mathrm{G},21})_{i\ell} = (\mathbf{R}_{\mathrm{G},21})_{ik} (\mathbf{R}_{\mathrm{G},21})_{i\ell} + (\mathbf{R}_{\mathrm{G},22})_{ii} (\mathbf{R}_{\mathrm{G},11})_{k\ell},$$
$$\left(\mathbb{E}[\Theta_{\mathrm{G}} \mathbf{R}_{\mathrm{G},21}]\right)_{k\ell} = \frac{1-\rho_x}{N} \sum_i \mathbb{E}\left[\partial_{\Theta_{ki}} (\mathbf{R}_{\mathrm{G},21})_{i\ell}\right],$$
$$\mathbb{E}[\Theta_{\mathrm{G}} \mathbf{R}_{\mathrm{G},21}] = \mathbb{E}[\hat{e}\, \mathbf{R}_{\mathrm{G},11}] + \frac{1-\rho_x}{N} \mathbb{E}[\mathbf{R}_{\mathrm{G},12} \mathbf{R}_{\mathrm{G},21}],$$
$$\mathbb{E}[\Theta_{\mathrm{G}} \mathbf{R}_{\mathrm{G},22}] = \mathbb{E}[\hat{e}\, \mathbf{R}_{\mathrm{G},12}] + \frac{1-\rho_x}{N} \mathbb{E}[\mathbf{R}_{\mathrm{G},12} \mathbf{R}_{\mathrm{G},22}].$$

Centering both factors and applying Cauchy–Schwarz, for real unit vectors of the appropriate block dimensions,

$$\left| u^\top \mathbb{E}[(\hat{e} - \mathbb{E}\hat{e}) \mathbf{R}_{\mathrm{G},ab}] v \right| \le \|\hat{e} - \mathbb{E}\hat{e}\|_{L^2} \mathrm{Var}(u^\top \mathbf{R}_{\mathrm{G},ab} v)^{1/2} = O(N^{-1}).$$

Real unit tests control the complex operator norm up to a factor 2. The product remainders in these identities are also $O(N^{-1})$, since $\|\mathbf{R}_{\mathrm{G}}\|_{\mathrm{op}} \le \eta^{-1}$. Insert these estimates into the $(1,1)$ and $(1,2)$ blocks of $(\xi I - H_{A_{\mathrm{G}}}) \mathbf{R}_{\mathrm{G}} = I$. Since $|d_1| \ge \eta$, divide the second relation by d_1, transpose, and substitute into the first, using $PP^\top = \rho_x \Pi$:

$$d_1\, \mathbb{E}\mathbf{R}_{\mathrm{G},11} - P\, \mathbb{E}\mathbf{R}_{\mathrm{G},21} = I + O(N^{-1}),$$
$$d_1\, \mathbb{E}\mathbf{R}_{\mathrm{G},12} - g_{\mathrm{G},N} P = O(N^{-1}),$$
$$\mathbb{E}\mathbf{R}_{\mathrm{G},12} = \frac{g_{\mathrm{G},N}}{d_1} P + O(N^{-1}),$$
$$\mathbb{E}\mathbf{R}_{\mathrm{G},11} = \frac{1}{d_1} I_N + \frac{\rho_x g_{\mathrm{G},N}}{d_1^2} \Pi + O(N^{-1}).$$

Together with the exact lower-right block, these are the blocks of $\mathcal{M}_N(g_{\mathrm{G},N}; \xi)$, uniformly on $\mathcal{K}$, proving (1).

The Gaussian Gram matrix is

$$G_{\mathrm{G}} := A_{\mathrm{G}}^\top A_{\mathrm{G}} = \frac{1}{N} (\sqrt{\rho_x}\, X_\star^\top + \sqrt{1-\rho_x}\, Z_{\mathrm{G}})(\sqrt{\rho_x}\, X_\star^\top + \sqrt{1-\rho_x}\, Z_{\mathrm{G}})^\top.$$

Its deterministic population is $\rho_x X_\star^\top X_\star/N = \rho_x I_{D_x}$, its noise variance is $1-\rho_x$, and $D_x/N \to \alpha_x^{-1}$. Thus Dozier & Silverstein (2007a, Theorem 1.1) gives almost-sure convergence of its empirical Stieltjes transform $s_{G_\mathrm{G}}(\zeta)$ to $s_\rho(\zeta)$ for $\zeta \in \mathbb{C}^+$. Since $|s_{G_\mathrm{G}}(\zeta)| \le (\operatorname{Im}\zeta)^{-1}$, bounded convergence gives

$$g_{\mathrm{G},N}(\xi) = \frac{\xi}{D_x}\mathbb{E}\operatorname{Tr}(\xi^2 I - G_\mathrm{G})^{-1} = -\xi\,\mathbb{E}s_{G_\mathrm{G}}(\xi^2)$$
$$\longrightarrow -\xi s_\rho(\xi^2) = g(\xi), \qquad \xi \in \mathbb{C}^{++}.$$

The last equality uses Appendix E.3: the Dozier–Silverstein equation is (E.4), and its Stieltjes solution selects the physical branch at ratio α_x (branch and support: Dozier & Silverstein (2007b, Section 3), Lemma E.4). The bounds $|g_{\mathrm{G},N}| \le \eta^{-1}$ and $\left|g'_{\mathrm{G},N}\right| \le \eta^{-2}$ upgrade the convergence to uniform convergence on $\mathcal{K} \subset \mathbb{C}^{++}$.

The exact finite-aspect solution $g_{0,N}$ also tends uniformly to g, by (E.17) and the positive-height continuity in Appendix E.5. The triangle inequality therefore gives $\sup_\mathcal{K}|g_{\mathrm{G},N} - g_{0,N}| \to 0$. On $\{\operatorname{Im}\xi \ge \eta\}$, the bounds $\operatorname{Im} g \le 0$, $|g| \le \eta^{-1}$ imply $|d_1(g;\xi)| \ge \eta$. Hence the template is uniformly Lipschitz in g and the aspect ratio on these sets. It follows that

$$\sup_\mathcal{K}\|\mathcal{M}_N(g_{\mathrm{G},N};\xi) - M_{0,N}(\xi)\|_\mathrm{op} \longrightarrow 0,$$
$$\sup_\mathcal{K}\|M_{0,N}(\xi) - M_0(\xi)\|_\mathrm{op} \longrightarrow 0.$$

These substitutions give $o(1)$ errors; the $O(N^{-1})$ estimate in (1) compares the expected Gaussian resolvent with its template before either substitution. This proves (2). □

G.3 Spectral confinement

Proposition G.3 (Confinement)**.** *For the nested flat orthogonal class of Proposition G.1 and fixed $\rho_x \in (0,1)$, there is $K_0 > w_G$ with $\mathbb{P}(\|G\|_\mathrm{op} \le K_0) \to 1$, and $\lambda_{\max}(G) \le w_G + o_P(1)$.*

Proof. Set $(\mathcal{Z}_N)_{ik} := (X_\star)_{ki}((S_x)_{ki} - \rho_x)/\sqrt{\rho_x(1-\rho_x)}$. These entries are independent, centered, of unit variance, and bounded by $\max(\rho_x, 1-\rho_x)/\sqrt{\rho_x(1-\rho_x)}$. The entrywise identity gives the matrix form:

$$\sqrt{\rho_x}\,(X_\star)_{ki} + \sqrt{1-\rho_x}\,(\mathcal{Z}_N)_{ik} = \frac{(X_\star)_{ki}(S_x)_{ki}}{\sqrt{\rho_x}},$$
$$G = \frac{1}{N}\big(\sqrt{\rho_x}\,X_\star^\top + \sqrt{1-\rho_x}\,\mathcal{Z}_N\big)\big(\sqrt{\rho_x}\,X_\star^\top + \sqrt{1-\rho_x}\,\mathcal{Z}_N\big)^\top.$$

For fixed unit $u \in \mathbb{R}^{D_x}$ and $v \in \mathbb{R}^N$, Hoeffding's inequality gives $\mathbb{P}(\left|u^\top \mathcal{Z}_N v\right| > t) \le 2e^{-ct^2}$. Two $1/4$-nets have at most 9^{D_x} and 9^N points, and their largest bilinear form controls $\|\mathcal{Z}_N\|_\mathrm{op}$ up to a factor 2. A union bound therefore gives, for sufficiently large fixed C_Z,

$$\mathbb{P}(\|\mathcal{Z}_N\|_\mathrm{op} > C_Z\sqrt{N}) \le 2\,9^{N+D_x}e^{-cC_Z^2N/4} \le e^{-c'N}.$$

Thus $\|G\|_\mathrm{op}^{1/2} \le \sqrt{\rho_x} + \sqrt{1-\rho_x}\;\|\mathcal{Z}_N\|_\mathrm{op}/\sqrt{N} \le \sqrt{\rho_x} + \sqrt{1-\rho_x}\,C_Z$ on the complementary event. Take $K_0 := (\sqrt{\rho_x} + \sqrt{1-\rho_x}\,C_Z)^2$, enlarged to $K_0 \ge w_G + 1$.

Apply Bai & Silverstein (2012, Theorem 1.1) with row dimension D_x, column dimension N, deterministic part $\sqrt{\rho_x}\,X_\star^\top$ and noise $\sqrt{1-\rho_x}\,\mathcal{Z}_N$. The aspect ratio is $D_x/N = \alpha_{x,N}^{-1} \to \alpha_x^{-1} > 0$. For hypothesis (a), use the fixed unimodular array $\mathcal{A}$ of Proposition G.1 and one infinite array of independent mask variables to define

$$\mathcal{Z}_{ik} := \mathcal{A}[k,i]\,\frac{(S_x)_{ki} - \rho_x}{\sqrt{\rho_x(1-\rho_x)}}.$$

Then $\mathcal{Z}_N$ is a corner of the same array for every selected dimension. Its entries need not be identically distributed. This coupling preserves the law at each N, hence every conclusion in probability. The standard example is the Sylvester array

$$\mathcal{A}[k,i] = (-1)^{\operatorname{popcount}(k\,\&\,i)}, \qquad k,i \ge 0, \tag{G.1}$$

whose top-left $2^m \times 2^m$ corner is the Sylvester matrix of order 2^m; at $N = 2^m$ the design consists of its first D_x columns.

Bounded entries give hypotheses (b)–(c). For (d)–(f), whitening (1) gives

$$\frac{\rho_x}{N} X_\star^\top X_\star = \rho_x I_{D_x}, \qquad \left\| \sqrt{\rho_x/N}\, X_\star^\top \right\|_{\mathrm{op}} = \sqrt{\rho_x}.$$

Thus the population law is exactly δ_{ρ_x}; the norm bound, aspect-ratio limit and matrix form are all as required.

For hypothesis (g), fix $\varepsilon \in (0, 1]$ and choose

$$[a_1, a_2] := [w_G + \varepsilon, K_0],$$
$$(b_1, b_2) := (w_G + \varepsilon/2, K_0 + 1) \supset [a_1, a_2].$$

The limiting law has population δ_{ρ_x}, noise variance $1 - \rho_x$ and ratio α_x^{-1} (Appendix E.3); its support top is w_G (Lemma E.4). For the finite-N laws, Proposition E.3 gives $w_{G,N} = W_N(\omega_{G,N}) \to w_G$, hence $w_{G,N} < b_1$ eventually. Thus (b_1, b_2) avoids both supports as required by (1.7).

Every $x > w_{G,N}$ has the unique exterior representation $x = W_N(\omega)$ with $0 < \omega < \omega_{G,N}$, and $x'_N(\mathfrak{b}) > 0$ at $\mathfrak{b} = (1 + \omega)^{-1}$ by (E.12). The converse stated after (1.6) in Bai & Silverstein (2012, Section 1), at population law δ_{ρ_x}, noise level $1 - \rho_x$ and ratio $\alpha_{x,N}^{-1}$, then identifies $\mathfrak{b}$ with their $b_n^0(x)$, so the population coordinate, their w_n^0 from (1.3), is

$$q_N(x) := \mathfrak{b}^2 x - \mathfrak{b}\,(1 - \rho_x)(1 - \alpha_{x,N}^{-1}).$$

Multiplying the first display of Proposition E.3 by $\mathfrak{b}^2$ shows that $q_N(x)$ is the value of $m_{\delta_{\rho_x}}^{-1}$ computed in its proof, so

$$\begin{aligned} q_N\big(W_N(\omega)\big) &= \rho_x + \frac{1 - \rho_x}{\alpha_{x,N}\omega} \\ &\geq \rho_x + \frac{1 - \rho_x}{\alpha_{x,N}\omega_{G,N}} \geq \rho_x + \frac{1 - \rho_x}{2\alpha_x\omega_G} =: \rho_x + \delta \end{aligned}$$

for all large N, because $\alpha_{x,N}\omega_{G,N} \to \alpha_x\omega_G$. Take $\hat{\varepsilon} := \varepsilon/4$, shrinking it if needed below the margin in the source's (1.8). Then

$$(a_1 - \hat{\varepsilon}, a_2 + \hat{\varepsilon}) = (w_G + \varepsilon - \hat{\varepsilon}, K_0 + \hat{\varepsilon}) \subset (b_1, b_2),$$

so its image under q_N lies in $[\rho_x + \delta, \infty)$.

Deleting column j from $\sqrt{\rho_x}\, X_\star^\top = \sqrt{N}\, P^\top$ gives the population matrix

$$\frac{N}{N-1} P^\top (I_N - e_j e_j^\top) P \preceq \frac{N}{N-1} P^\top P = \frac{\rho_x N}{N-1} I_{D_x}.$$

Since $\rho_x N/(N-1) < \rho_x + \delta$ for all large N, the spectral bound is uniform in j. No deleted-column population matrix has an eigenvalue in the required image; condition (1.10) holds for all N columns.

The theorem applies along the nested dimension sequence: its proof uses the row dimension through summable moment bounds and almost-sure ESD convergence, which remain valid on this subsequence. On the chosen coupling it excludes eigenvalues in $[w_G + \varepsilon, K_0]$ eventually almost surely. Thus, in probability for the original masks,

$$\begin{aligned} \mathbb{P}\big(\lambda_{\max}(G) > w_G + \varepsilon\big) \leq\ & \mathbb{P}(\|G\|_{\mathrm{op}} > K_0) \\ & + \mathbb{P}\big(\operatorname{spec} G \cap [w_G + \varepsilon, K_0] \neq \varnothing\big) \longrightarrow 0. \end{aligned}$$

Since ε is arbitrary, $\lambda_{\max}(G) \leq w_G + o_P(1)$. □

G.4 Verification of the deterministic equivalent

Proof of Proposition G.1. Clause **(i)** of Assumption E.10 is Proposition G.3. For clause **(ii)**, fix a compact $K \subset (w_G, \infty)$, choose K' with $K \subset (w_G, K']$, and put

$$\mathcal{S}_K := \big\{\xi :\ \sqrt{w_G} + \varepsilon \leq \operatorname{Re}\xi \leq \sqrt{K'} + 1,\ |\operatorname{Im}\xi| \leq \eta_1\big\}.$$

Choose $\varepsilon, \eta_1 > 0$ small enough that $\{\xi^2 : \xi \in \mathcal{S}_K\}$ contains an open $U_K \supset K$ whose closure is separated from $[0, w_G]$.

Apply Lemma F.5 with the following substitutions. The entries of X are independent, centered and bounded; those of Y are iid Gaussian, with matching variances:

$$d = n = N, \quad m = D_x, \quad P_0 = P, \quad B = I_N, \quad X = \sqrt{N}\,\Theta, \quad Y = \sqrt{N}\,\Theta_{\mathrm{G}},$$

$$Z_X = A, \quad Z_Y = A_{\mathrm{G}}, \quad X_{ki} = \frac{(X_\star)_{ki}\big((S_x)_{ki} - \rho_x\big)}{\sqrt{\rho_x}}, \quad \mathbb{E}X_{ki}^2 = \mathbb{E}Y_{ki}^2 = 1 - \rho_x.$$

The Gaussian calculation of Lemma G.2 supplies the limiting forms. Since M_0 is complex symmetric and bounded at fixed height, Corollary F.6 applies with the test matrix $\mathrm{diag}(0_N, I_{D_x})$. For every compact $\mathcal{K} \subset \mathbb{C}^{++}$ we obtain

$$\sup_{\xi \in \mathcal{K}} \sup_{\|r\| = \|s\| = 1} \mathbb{E}\left| r^\top \mathbf{R}(\xi) s - r^\top M_0(\xi) s \right|^2 \longrightarrow 0,$$

$$\sup_{\xi \in \mathcal{K}} \left\| N^{-1} \operatorname{Tr}\left(\mathbf{R}_{22}(\xi) - M_{0,22}(\xi)\right) \right\|_{L^2} \longrightarrow 0.$$

The test vectors are deterministic, and their supremum remains outside the expectation.

By Proposition G.3, the event $\{\lambda_{\max}(G) \le w_G + \varepsilon^2/4\}$ has probability tending to 1. On it,

$$\operatorname{spec}(H_A) \subset [-\sqrt{w_G} - \varepsilon/2, \sqrt{w_G} + \varepsilon/2], \qquad \sup_{\xi \in \mathcal{S}_K} \|\mathbf{R}(\xi)\|_{\mathrm{op}} \le \frac{2}{\varepsilon}.$$

The deterministic matrix M_0 is analytic and bounded on a sufficiently thin $\mathcal{S}_K$. Indeed, the integral formula in Appendix E.3 extends g to $\{\xi^2 \notin [0, w_G]\}$, and $d_1 = \xi/(1+\omega) > 0$ on the real exterior branch. Compactness and continuity keep d_1 away from zero throughout the rectangle, so the blocks of M_0 and their derivatives are bounded there.

For $0 < \eta \le \eta_1$ and $\xi = x + iy$ with $0 \le y \le \eta$, compare with $x + i\eta$. The resolvent identity gives

$$\|\mathbf{R}(\xi) - \mathbf{R}(x + i\eta)\|_{\mathrm{op}} \le |\eta - y|\, \|\mathbf{R}(\xi)\|_{\mathrm{op}} \|\mathbf{R}(x + i\eta)\|_{\mathrm{op}} \le \frac{4\eta}{\varepsilon^2},$$

$$\|M_0(\xi) - M_0(x + i\eta)\|_{\mathrm{op}} \le C\eta.$$

Therefore each unit bilinear form satisfies

$$\left| r^\top (\mathbf{R}(\xi) - M_0(\xi)) s \right| \le \left| r^\top (\mathbf{R}(x + i\eta) - M_0(x + i\eta)) s \right| + \frac{4\eta}{\varepsilon^2} + C\eta.$$

The normalized trace obeys the same bound up to the bounded factor D_x/N; negative y follows by conjugation.

For fixed η, $\mathcal{S}_K \cap \{\operatorname{Im}\xi \ge \eta\}$ is a compact subset of $\mathbb{C}^{++}$. A finite net, the positive-height comparison and the resolvent Lipschitz bound $|\xi - \xi'|\,\eta^{-2}$ give uniform convergence in probability there for $r, s \in \{(\hat{b}; 0), (0; u_0)\}$ and for $N^{-1} \operatorname{Tr} \mathbf{R}_{22}$. First let $N \to \infty$ at fixed η; then let $\eta \downarrow 0$ in the displayed bound. The three forms therefore converge to their M_0 counterparts uniformly on all of $\mathcal{S}_K$ in probability.

Use $\Pi\hat{b} = \hat{b}$, $\left\|\hat{b}\right\| = 1$ and $P^\top \hat{b} = \sqrt{\rho_x}\, u_0$ in (E.7). With $d_1 = \xi - (1 - \rho_x)\alpha_x^{-1} g$, the deterministic values are

$$\hat{b}^\top M_{0,11} \hat{b} = d_1^{-1} + \rho_x g d_1^{-2} = h,$$

$$u_0{}^\top M_{0,21} \hat{b} = d_1^{-1} g\, u_0{}^\top P^\top \hat{b} = \frac{\sqrt{\rho_x}\, g}{d_1},$$

$$\frac{1}{N} \operatorname{Tr} M_{0,22} = \frac{D_x}{N} g = \frac{g}{\alpha_{x,N}}.$$

Flatness and $\|b\|^2 = N$ also give the exact row-norm parameter in (8):

$$\tau_N = \frac{1}{N^2} \sum_k b_k^2 \, \|(X_\star)_{k\cdot}\|^2 = \frac{D_x}{N^2} \|b\|^2 = \frac{1}{\alpha_{x,N}}.$$

Thus Proposition C.3 gives $\|a_N\|^2 \xrightarrow{P} n_a > 0$, and $a_N \neq 0$ with probability tending to 1. Substitute the block limits into the trace identity of Appendix G.1 and the push identities (E.8):

$$\psi_N(w) = \frac{\xi}{N} \operatorname{Tr} \mathbf{R}_{22} - \frac{1}{\alpha_{x,N}} \longrightarrow \frac{\xi g - 1}{\alpha_x},$$

$$\Phi_N(w) = \frac{\xi \hat{b}^\top \mathbf{R}_{11} \hat{b} - 1}{\|a_N\|^2} \longrightarrow \frac{\xi h - 1}{n_a},$$

$$\Phi_N^\times(w) = \frac{{u_0}^\top \mathbf{R}_{21} \hat{b}}{\|a_N\|} \longrightarrow \frac{\sqrt{\rho_x}\, g}{d_1 \sqrt{n_a}}.$$

Here $w = \xi^2$, and the convergence is uniform on U_K in probability. These are exactly (E.9): $e_1 = (1 - \rho_x)\alpha_x^{-1} g$, and $h = d_1^{-1} + \rho_x g d_1^{-2}$ is the second continued fraction of (E.4) on its solution. This proves clause **(ii)** and the row-norm assertion. □

Remark G.4 (Signed permutations). The randomized designs of Appendix I.2 transfer from the corners of the fixed array by distributional equivariance. For a row signed permutation U, mask exchangeability and $b \mapsto Ub$ give $A \mapsto UA$ in law together with $\hat{b} \mapsto U\hat{b}$, which preserves the joint law of (G, a_N) and hence of the functionals; for a column signed permutation V, $u_0 \mapsto V^\top u_0$, $G \mapsto V^\top G V$ and $a_N \mapsto V^\top a_N$, so the functionals are invariant when the planted direction is transformed with the design. This includes N-dependent row and column randomizations.

Hence signed-permuted designs are reduced to the nested corners before the fixed-array theorem is invoked in Proposition G.3, and confinement is inherited at the same retention. Assumption C.1 must hold for the pair $(X_\star, u_0)$; selecting columns of a larger flat orthogonal matrix preserves flat orthogonality but not automatically the score bound Assumption C.1(iv).

H THE REPLICA DERIVATION

This appendix studies the covariance-matched Gaussian ensemble (F.3), conditionally on $(X_\star, S_x)$. At $\rho_y = 1$, this ensemble has the conditional law of $\bar{C}_N$. For $\rho_y < 1$, it is only a formal covariance-matched replacement for the actual response-masked noise. Its positive-integer Gaussian average is exact. The continuation to $n \to 0$, replica symmetry, limit exchanges, saddle selection, and source differentiation remain formal under assumptions **(R1)**–**(R5)** below. Subject to those assumptions, the saddle equations are algebraically equivalent to the finite-N conditional deterministic-equivalent secular and overlap equations obtained by evaluating the exact identities of Appendix E.6 on the conditional transforms. Evaluation at the scalar limits from Appendices E.2–E.5 then gives the candidate formulas of Proposition E.7. This calculation neither proves Assumption E.10 nor contributes to the proof of Theorem 3.4.

We work on the mask event $a_N \neq 0$, so $\widehat{a}_N$ is defined and $G \neq 0$. Conditionally on the design and S_x, the functionals of (E.1) are deterministic. We keep the subscript N on them, write the transforms of (E.3) without it, and retain $\alpha_{x,N} = N/D_x$ and $\alpha_{y,N} = N/D_y$ in every finite-N identity. Only after the deterministic limit do these become α_x and α_y. The one-variable forms $\psi(\omega)$ and $\Phi(\omega)$ of Definition E.2 do not appear here. The response overlap is denoted by r_y. The spherical multiplier ζ and susceptibility χ are independent variational quantities; Proposition H.2 relates them to the spectral coordinates after imposing stationarity.

Replica assumptions. The calculation below is conditional on the design and S_x and assumes:
(R1) analytic continuation of the integer-replica expression to $n \to 0$.

(R2) replica symmetry in the right-sector order parameters and the spherical multipliers ζ_a.

(R3) self-averaging of the quenched free energy.

(R4) the scaling $1 - q_v = \chi/\beta$ and the stated zero-temperature/order-of-limits exchange.

(R5) an isolated differentiable saddle when source fields are differentiated.

These assumptions are not used by the exact rank-one identities (Appendix E.6) or the finite-signal proof of Appendix F.

For a matrix $M \in \mathbb{R}^{D_x \times D_y}$ and source fields $h = (h_x, h_p, h_y)$, define the one-copy partition function

$$Z_{\beta,N}(M;h) := \int_{\|u\|=\|v\|=1} \exp\Big(\beta N\big[u^\top M v + h_x u^\top u_0 + h_p u^\top \widehat{a}_N + h_y v^\top v_0\big]\Big)\, \mathrm{d}\sigma(u)\, \mathrm{d}\sigma(v). \tag{H.1}$$

Here $\mathrm{d}\sigma$ is normalized uniform measure on the indicated unit sphere. The fields couple to the planted design direction, the corrupted direction, and the planted response direction, respectively. Let $F_{\beta,N}(M;h) := (\beta N)^{-1} \log Z_{\beta,N}(M;h)$. The residual bound of Lemma B.3 also controls these free energies. The source terms are identical for C and $\bar{C}_N$, and $\big|u^\top C v - u^\top \bar{C}_N v\big| \le \|\mathcal{R}_N\|_{\mathrm{op}}$ on the unit spheres, so

$$e^{-\beta N \|\mathcal{R}_N\|_{\mathrm{op}}}\, Z_{\beta,N}(\bar{C}_N;h) \;\le\; Z_{\beta,N}(C;h) \;\le\; e^{\beta N \|\mathcal{R}_N\|_{\mathrm{op}}}\, Z_{\beta,N}(\bar{C}_N;h),$$

hence, exactly and uniformly in the source fields, $\big|F_{\beta,N}(C;h) - F_{\beta,N}(\bar{C}_N;h)\big| \le \|\mathcal{R}_N\|_{\mathrm{op}}$, and the expected free energies differ by $o(1)$ after averaging over the masks and noise at fixed θ. This bound concerns free-energy values; identifying their limiting source derivatives still requires the formal assumptions. At $\rho_y = 1$, the conditional law of $\bar{C}_N$ is exactly that of C_{G} in (F.3). For $\rho_y < 1$, using C_{G} is the formal Gaussian replacement described above. All Gaussian averages below evaluate (H.1) at $M = C_{\mathrm{G}}$.

H.1 REPLICA-SYMMETRIC CALCULATION

The derivation runs from the replicated partition function to the free energy (H.12) and the stationarity conditions (S1)–(S3), conditionally on the mask, so that in (F.3) the quadruple $(G, \widehat{a}_N, u_0, s_N)$ is deterministic and only W_{G} is random. For a positive integer n, the power $Z_{\beta,N}(C_{\mathrm{G}};h)^n$ introduces n unit-vector pairs (u^a, v^a), indexed by $a = 1, \dots, n$.

Lemma H.1 (Source observables within the RS formalism)**.** *Within the replica-symmetric formalism, let $f(h) := \lim_{\beta\to\infty} \lim_{N\to\infty} \mathbb{E}_{W_{\mathrm{G}}} F_{\beta,N}(C_{\mathrm{G}}; h)$, with the design and mask conditioned on. Define the limiting source observables by*

$$r_x^{\mathrm{RS}} := \lim_{h\to 0} \frac{\partial f}{\partial h_x}, \qquad r_a^{\mathrm{RS}} := \lim_{h\to 0} \frac{\partial f}{\partial h_p}, \qquad r_y^{\mathrm{RS}} := \lim_{h\to 0} \frac{\partial f}{\partial h_y}, \tag{H.2}$$

*where $h_y \downarrow 0$ and $h_x, h_p = o(h_y)$. At fixed N, taking the zero-temperature limit before sending the sources to zero selects the top singular pair of C_{G} with positive response overlap. The derivatives of the expected free energy give its conditional expected overlaps. Identifying these finite-N derivatives with the derivatives of the limiting free energy is the formal step covered by **(R1)–(R5)**.*

Proof. At fixed N and β the integral runs over the compact product of two spheres with an integrand smooth in the sources, so differentiation passes under the integral sign and returns the Gibbs average of the corresponding overlap, $\partial_{h_x} \mathbb{E}_{W_{\mathrm{G}}} F_{\beta,N}(C_{\mathrm{G}}; h) = \mathbb{E}_{W_{\mathrm{G}}} \langle u^\top u_0 \rangle_\beta$, and likewise for h_p and h_y, exactly at every N, β and h. As $\beta \to \infty$, Laplace asymptotics make $F_{\beta,N}(C_{\mathrm{G}}; h)$ converge to the maximum of $u^\top C_{\mathrm{G}} v + h_x u^\top u_0 + h_p u^\top \widehat{a}_N + h_y v^\top v_0$ over the two spheres, with the Gibbs measure concentrating on the maximizing set. At $h = 0$ the maximum is $\sigma_1(C_{\mathrm{G}})$, attained by its top singular pair and the simultaneous sign reversal. This singular value is almost surely simple: the Gaussian noise is absolutely continuous on $\mathrm{range}(G) \otimes \mathbb{R}^{D_y}$, which carries $\widehat{a}_N$ since $a_N \neq 0$ and $a_N \in \mathrm{range}(A^\top) = \mathrm{range}(G)$. But the $h = 0$ integrand is invariant under $(u, v) \mapsto (-u, -v)$ while each of the three overlaps is odd, so every Gibbs average vanishes by exact cancellation, and a source must break the tie. The degeneracy is one global sign, so the three overlaps flip together: $h_y > 0$ selects the representative with positive response overlap, whereas h_x alone would select a positive design overlap, which need not have the same sign as the response overlap. Hence the path $h_x, h_p = o(h_y)$.

At finite N, the formal saddle potential Ψ_h in (H.11) depends on the sources explicitly, through the coefficients c_i of (H.8) and the term $h_y r_y$, and implicitly, through its stationary point $(r_y, \zeta, \chi)(h)$. The envelope theorem removes the latter contribution because the derivatives with respect to these three variables vanish there. The remaining partial derivatives give finite-N saddle predictions. Their limits give (H.2) under the stated formal assumptions. □

The replica overlaps needed for the Gaussian average are

$$(Q_u^G)_{ab} := u^{a\top} G u^b, \qquad (Q_v)_{ab} := v^{a\top} v^b, \qquad (r_y)_a := v^{a\top} v_0. \tag{H.3}$$

The average over W_{G} of the noise factor in $Z_{\beta,N}(C_{\mathrm{G}}; h)^n$ is exact for every positive integer n:

$$\mathbb{E}_{W_{\mathrm{G}}} \exp\Big(\beta\sqrt{N} \sum_{a=1}^{n} u^{a\top} G^{1/2} W_{\mathrm{G}} v^a\Big) = \exp\Big(\frac{\beta^2 N}{2} \operatorname{Tr}\big(Q_u^G Q_v\big)\Big). \tag{H.4}$$

The covariance G therefore enters the coupling between left replicas.

Proof of (H.4). At every positive integer n, this is the Gaussian moment-generating identity for the matrix $\sum_a G^{1/2} u^a (v^a)^\top$, whose squared Frobenius norm is $\sum_{a,b} u^{a\top} G u^b \, v^{a\top} v^b = \mathrm{Tr}(Q_u^G Q_v)$. □

We impose the right-sector overlaps in (H.3) by delta functions and assume replica symmetry: $Q_v = (1 - q_v) I_n + q_v \mathbf{1}\mathbf{1}^\top$ and $(r_y)_a = r_y$ for every a. Thus r_y without a replica index denotes their common scalar value. Every other factor of the replicated integrand, the disorder average (H.4), the signal term and the sources, depends on the right replicas only through Q_v and the response overlaps $(r_y)_a$, so the v-integral at fixed order parameters is a constrained volume.

For the right-sector entropy, split $v^a = (r_y)_a v_0 + \tilde{v}^a$ with $\tilde{v}^a \perp v_0$. The residual Gram matrix has entries $(\tilde{Q}_v)_{ab} = (Q_v)_{ab} - (r_y)_a (r_y)_b$. To leading order in N at fixed integer n, the Gram–Wishart Jacobian contributes $(2\alpha_{y,N})^{-1} \log\det \tilde{Q}_v$ per unit of N (Mézard et al., 1987). Under RS

its eigenvalues are $1-q_v$ on $\mathbf{1}^\perp$ and $1-q_v+n(q_v-r_y^2)$ along $\mathbf{1}$. Thus, on $r_y^2 \le q_v < 1$,

$$\det \tilde{Q}_v = (1-q_v)^{n-1}\big(1-q_v+n(q_v-r_y^2)\big), \qquad \lim_{n\to 0}\frac{\log\det \tilde{Q}_v}{n} = \phi(q_v, r_y), \tag{H.5}$$

where

$$\phi(q,r) := \log(1-q) + \frac{q-r^2}{1-q}.$$

The last identity is the first-order expansion of the displayed determinant. Its continuation in n is part of (R1).

To integrate the left coordinates, represent each constraint $\|u^a\|^2 = 1$ by a delta function and a conjugate multiplier ζ_a. The inner coordinate integral then uses Lebesgue measure; the multiplier contours and normalized-sphere factors remain outside it. Under replica symmetry $\zeta_a = \zeta$, the signal, sources, and Gaussian average (H.4) give the left action

$$\begin{aligned}\mathcal{A}_u \;=\; & \beta N s_N\, r_y \sum_a u^{a\top}\widehat{a}_N \;+\; \frac{\beta^2 N}{2}\sum_{a,b}(Q_v)_{ab}\, u^{a\top} G u^b \\ & + \beta N \sum_a \big(h_x u^{a\top} u_0 + h_p u^{a\top}\widehat{a}_N\big) \;-\; \frac{\beta N\zeta}{2}\sum_a \big(\,\|u^a\|^2 - 1\big).\end{aligned} \tag{H.6}$$

Let $G = \sum_i \lambda_i \varphi_i \varphi_i^\top$ be an orthonormal eigendecomposition, with $\widehat{a}_{N,i} := \varphi_i{}^\top \widehat{a}_N$ and $u_{0,i} := \varphi_i{}^\top u_0$. Every bilinear form in (H.6) is a sum over these eigencoordinates. Let $u_i \in \mathbb{R}^n$ collect the replica coordinates along φ_i. Under RS, the diagonal part of the coupling contributes $\beta^2\lambda_i(1-q_v) = \beta\chi_\beta\lambda_i$, where $\chi_\beta := \beta(1-q_v)$. It combines with the spherical multiplier to give the curvature $\zeta - \chi_\beta\lambda_i$. The common overlap q_v produces a rank-one coupling. Thus

$$\mathcal{A}_u \;=\; \frac{\beta N\zeta n}{2} + \sum_i \Big(-\tfrac{1}{2}\, u_i{}^\top M_i u_i + \beta N c_i\, \mathbf{1}^\top u_i\Big), \tag{H.7}$$

where

$$M_i = \beta N\big[(\zeta-\chi_\beta\lambda_i)I_n - \beta q_v\lambda_i \mathbf{1}\mathbf{1}^\top\big], \qquad c_i := s_N\, r_y \widehat{a}_{N,i} + h_x u_{0,i} + h_p \widehat{a}_{N,i}, \tag{H.8}$$

the three linear terms of (H.6) collecting into the single source c_i. Eigencoordinate i enters only through the triple $(\lambda_i, \widehat{a}_{N,i}, u_{0,i})$.

The matrix $M_i/(\beta N)$ has eigenvalue $\zeta-\chi_\beta\lambda_i$ on $\mathbf{1}^\perp$ and $\zeta-\chi_\beta\lambda_i - n\beta q_v\lambda_i$ along $\mathbf{1}$. Hence, for every positive integer n,

$$\begin{aligned}\det\frac{M_i}{\beta N} \;&=\; (\zeta-\chi_\beta\lambda_i)^{\,n-1}\big(\zeta-\chi_\beta\lambda_i - n\beta q_v\lambda_i\big), \\ \mathbf{1}^\top M_i^{-1}\mathbf{1} \;&=\; \frac{n}{\beta N\big(\zeta-\chi_\beta\lambda_i - n\beta q_v\lambda_i\big)},\end{aligned}$$

valid for the Gaussian integral when $\zeta-\chi_\beta\lambda_i - n\beta q_v\lambda_i > 0$ for every i. Since $q_v \ge r_y^2 \ge 0$ and G is positive semidefinite, this condition makes every M_i positive definite. Completing the square centers the inner integral at $\beta N c_i M_i^{-1}\mathbf{1}$ and gives

$$\int_{\mathbb{R}^n} \exp\Big(-\tfrac{1}{2}u_i{}^\top M_i u_i + \beta N c_i \mathbf{1}^\top u_i\Big)\,\mathrm{d}u_i = \frac{(2\pi)^{n/2}}{\sqrt{\det M_i}}\,\exp\Big(\tfrac{1}{2}(\beta N c_i)^2 \mathbf{1}^\top M_i^{-1}\mathbf{1}\Big).$$

This identity is exact for positive integer n on the stated positivity domain. On the continued branch $\zeta > \chi_\beta\lambda_{\max}(G)$,

$$\begin{aligned}\lim_{n\to 0}\frac{1}{n}\log\det\frac{M_i}{\beta N} &= \log(\zeta-\chi_\beta\lambda_i) - \frac{\beta q_v\lambda_i}{\zeta-\chi_\beta\lambda_i}, \\ \mathbf{1}^\top M_i^{-1}\mathbf{1} &= \frac{n}{\beta N(\zeta-\chi_\beta\lambda_i)} + O(n^2).\end{aligned} \tag{H.9}$$

These are the first-order expansions of the determinant and contraction above. Continuing from the integer-n positivity domain to this branch remains a formal step.

H.2 THE ZERO-TEMPERATURE POTENTIAL

We suppress the N dependence of the saddle potentials below. They use the finite-N spectral data and the leading large-N entropy terms; their identification with the quenched free energy remains formal. Taking the logarithm of the inner Gaussian integral above gives $-\frac{1}{2}\log\det(M_i/2\pi) + \frac{1}{2}(\beta N c_i)^2 \mathbf{1}^\top M_i^{-1}\mathbf{1}$ per eigencoordinate. The source term passes through the v-integration as $\beta N n\, h_y r_y$. Adding it, the multiplier constant $\beta N\zeta n/2$, and the entropy (H.5), dividing by $\beta N n$, dropping additive constants independent of the order parameters and sources, and inserting (H.9) at $n \to 0$ leaves the finite-temperature potential

$$\begin{aligned}\Psi_\beta(r_y, q_v, \zeta) \;=\; & \frac{\zeta}{2} - \frac{1}{2\alpha_{x,N}\beta}\,\langle \log(\zeta - \chi_\beta\lambda)\rangle + \frac{1}{2\alpha_{x,N}}\left\langle \frac{q_v\lambda}{\zeta - \chi_\beta\lambda}\right\rangle \\ & + \frac{1}{2}\sum_i \frac{c_i^2}{\zeta - \chi_\beta\lambda_i} + \frac{\phi(q_v, r_y)}{2\alpha_{y,N}\beta} + h_y r_y,\end{aligned} \tag{H.10}$$

where $\langle g(\lambda)\rangle := D_x^{-1}\sum_{i=1}^{D_x} g(\lambda_i)$. The factors $D_x/N = 1/\alpha_{x,N}$ and $D_y/N = 1/\alpha_{y,N}$ appear outside the spectral averages and the entropy, respectively. Each contribution is Nn times a function of the order parameters up to $o(Nn)$, so the $n \to 0$ limits may be inserted before the $N \to \infty$ Laplace evaluation, whose maximum becomes a stationary point under the continuation. The exchange is formal. The Gaussian prefactors combine with the normalized-sphere factors before the large-N limit; their remaining contribution vanishes at zero temperature. Within the formalism, $f(h) = \lim_{\beta\to\infty}\lim_{N\to\infty}\operatorname{stat}\Psi_\beta$. Now let $\beta \to \infty$ with $q_v \to 1$ and $\chi := \lim\beta(1 - q_v)$ finite, so that $\chi_\beta \to \chi > 0$. At a fixed interior point $\zeta > \chi\lambda_{\max}(G)$, the logarithmic average in (H.10) stays bounded, so its contribution is $O(1/\beta)$. This estimate is uniform on compact sets separated from the spectral boundary. The adjacent average tends to $\frac{1}{2\alpha_{x,N}}\langle\lambda/(\zeta - \chi\lambda)\rangle$, and substituting $1 - q_v = \chi/\beta$ supplies the remaining power of β in the entropy term,

$$\frac{\phi(q_v, r_y)}{2\alpha_{y,N}\beta} \;=\; \frac{\log(\chi/\beta)}{2\alpha_{y,N}\beta} + \frac{1 - \chi/\beta - r_y^2}{2\alpha_{y,N}\chi} \;\xrightarrow[\beta\to\infty]{}\; \frac{1 - r_y^2}{2\alpha_{y,N}\chi}.$$

What remains is the zero-temperature potential

$$\Psi_h(r_y, \zeta, \chi) = \frac{\zeta}{2} + \frac{1}{2\alpha_{x,N}}\left\langle \frac{\lambda}{\zeta - \chi\lambda}\right\rangle + \frac{1}{2}\sum_i \frac{c_i^2}{\zeta - \chi\lambda_i} + \frac{1 - r_y^2}{2\alpha_{y,N}\chi} + h_y r_y, \tag{H.11}$$

on the domain $\{|r_y| \le 1,\ \chi > 0,\ \zeta > \chi\lambda_{\max}(G)\}$, with the overlap bound inherited from $r_y^2 \le q_v < 1$. At $h = 0$ the source is $c_i = s_N r_y \widehat{a}_{N,i}$ and the sum $\sum_i c_i^2/(\zeta - \chi\lambda_i)$ collapses to $s_N^2 r_y^2 F(\zeta, \chi)$ with $F(\zeta, \chi) := \widehat{a}_N{}^\top(\zeta I - \chi G)^{-1}\widehat{a}_N$, leaving the zero-source free energy

$$\Psi(r_y, \zeta, \chi) = \frac{\zeta}{2} + \frac{1}{2\alpha_{x,N}}\left\langle \frac{\lambda}{\zeta - \chi\lambda}\right\rangle + \frac{s_N^2 r_y^2}{2}\,F(\zeta, \chi) + \frac{1 - r_y^2}{2\alpha_{y,N}\chi} \tag{H.12}$$

on the same domain. At $G = I$, every denominator in the spectral average and in F becomes $\zeta - \chi$.

Stationarity of (H.12) in (ζ, χ, r_y) gives

$$\frac{1}{\alpha_{x,N}}\left\langle \frac{\lambda}{(\zeta - \chi\lambda)^2}\right\rangle + s_N^2 r_y^2\,\widehat{a}_N{}^\top(\zeta I - \chi G)^{-2}\widehat{a}_N = 1, \tag{S1}$$

$$\frac{1}{\alpha_{x,N}}\left\langle \frac{\lambda^2}{(\zeta - \chi\lambda)^2}\right\rangle + s_N^2 r_y^2\,\widehat{a}_N{}^\top G(\zeta I - \chi G)^{-2}\widehat{a}_N = \frac{1 - r_y^2}{\alpha_{y,N}\chi^2}, \tag{S2}$$

$$r_y\Big(s_N^2 F(\zeta, \chi) - \frac{1}{\alpha_{y,N}\chi}\Big) = 0. \tag{S3}$$

These equations follow by direct differentiation of (H.12). At $r_y = 0$ and $h = 0$, every c_i vanishes, so all three source derivatives of Ψ_h are zero. This is the zero-overlap RS branch. To locate it, let $w = \zeta/\chi$ and use T_N, $w_{R,N}$, and $z_{\mathrm{edge},N}$ from (E.20) and (E.22). Multiplying (S2) by χ^2 gives

$$\frac{1}{N}\sum_i \frac{\lambda_i^2}{(w - \lambda_i)^2} = -\partial_w\big[w\psi_N(w)\big] = \frac{1}{\alpha_{y,N}}, \qquad \partial_w T_N(w) = 0.$$

Since $G \neq 0$, the convexity formula (E.21) and the endpoint signs give a unique root above $\lambda_{\max}(G)$, namely $w_{R,N}$. Equation (S1) separately determines the susceptibility:

$$\chi^2 = \frac{1}{N}\sum_i \frac{\lambda_i}{(w-\lambda_i)^2} = -\psi_N'(w) > 0.$$

Using $\partial_w T_N(w) = \alpha_{y,N}^{-1} + \psi_N(w) + w\psi_N'(w) = 0$ then gives

$$\zeta^2 = w^2\chi^2 = -w^2\psi_N'(w) = T_N(w_{R,N}) = z_{\mathrm{edge},N}^2.$$

Since $\zeta > 0$, this saddle sits at the noise edge. The combination (H.19) gives $\Psi = \zeta$ on both zero-source branches, as shown in the saddle-value calculation below. The replica prescription selects the branch with the larger ζ, the supercritical root whenever it exists. This selection remains part of the formal assumptions (R1)–(R5).

The derivation produces the transforms (E.3) as eigenvalue sums of the deformed model. Writing $d_i(z) := z^2 - \lambda_i/(\alpha_{y,N}(1+\delta))$, the change of variables $w = \alpha_{y,N}(1+\delta)z^2$ gives

$$\bar{m}_1(z) = z\sum_i \frac{\widehat{a}_{N,i}^2}{d_i(z)}, \qquad \bar{m}_1^{\times}(z) = z\sum_i \frac{\widehat{a}_{N,i}u_{0,i}}{d_i(z)}.$$

The third transform follows directly from the fixed point. Since $w - \lambda_i = \alpha_{y,N}(1+\delta)d_i(z)$, that equation reads $N^{-1}\sum_i \lambda_i/d_i(z) = -\delta$. Since $\alpha_{y,N}D_y = N$,

$$\sum_i \frac{1}{d_i(z)} + \frac{D_y - D_x}{z^2} = \frac{1}{z^2}\left[D_y + \frac{\sum_i \lambda_i/d_i(z)}{\alpha_{y,N}(1+\delta)}\right] = \frac{D_y}{z^2(1+\delta)}.$$

Multiplying by z/D_y gives, identically on the branch,

$$\frac{z}{D_y}\left[\sum_{i=1}^{D_x}\frac{1}{d_i(z)} + \frac{D_y - D_x}{z^2}\right] \;=\; \frac{1}{z(1+\delta)} \;=\; \bar{m}_2(z). \tag{H.13}$$

The left side is the deterministic-equivalent right-sector co-resolvent trace of the deformed model.

One derivative identity will be used below. Differentiating $z^2 = T_N(w)$ from (E.20) gives $w'(z)\,\partial_w T_N(w) = 2z$. Since $K = w/z$ and $\partial_w T_N(w) = w(1 + K^2\psi_N'(w))/K^2$, this becomes

$$w'(z)\left(1 + K^2\psi_N'(w)\right) \;=\; 2K. \tag{H.14}$$

H.3 Equivalence with the secular system

At a stationary point, we identify the spectral coordinates of (E.2) through

$$\widehat{z} := \zeta, \qquad \delta := \frac{1}{\alpha_{y,N}\chi\zeta} - 1, \qquad \text{equivalently} \quad \chi = 1/K(\widehat{z}). \tag{H.15}$$

Proposition H.2 (Algebraic equivalence of the positive RS saddle and the increasing-branch secular solution)**.** *Let (r_y, ζ, χ) with $r_y > 0$, $\chi > 0$, and $\zeta > \chi\lambda_{\max}(G)$ solve the stationarity system* (S1)–(S3)*, and define $(\widehat{z}, \delta)$ by* (H.15)*. Then $\delta \in (-1, 0)$ solves* (E.2) *at $z = \widehat{z}$, the value $\widehat{z}$ solves $\bar{D}(\widehat{z}) = 1/s_N^2$, and at this saddle,*

$$\begin{aligned} r_y^2 &= \frac{-2\bar{m}_2(\widehat{z})}{s_N^2\,\bar{D}'(\widehat{z})}, \\ \partial_{h_p}\Psi_h\big|_{h=0} &= s_N\, r_y\, \bar{m}_1(\widehat{z}), \qquad \partial_{h_x}\Psi_h\big|_{h=0} = s_N\, r_y\, \bar{m}_1^{\times}(\widehat{z}). \end{aligned} \tag{H.16}$$

*The derivative with respect to h_y is r_y. Under **(R1)–(R5)**, the limits of these finite-N saddle derivatives give the source observables in* (H.2)*. Conversely, every real solution of* (E.2) *and $\bar{D}(\widehat{z}) = 1/s_N^2$ with $\widehat{z} > 0$, $\delta \in (-1, 0)$, $w > \lambda_{\max}(G)$, and $1 + K^2\psi_N'(w) > 0$ arises in this way.*

Proof of Proposition H.2. Equation (H.15) gives $1/(\alpha_{y,N}\chi) = \zeta(1+\delta)$, hence, with $w = \alpha_{y,N}(1+\delta)\widehat{z}^2$ and $K = \widehat{z}\,\alpha_{y,N}(1+\delta)$,

$$\zeta - \chi\lambda_i = \frac{\widehat{z}^2\alpha_{y,N}(1+\delta) - \lambda_i}{\widehat{z}\,\alpha_{y,N}(1+\delta)} = \frac{w-\lambda_i}{K}, \tag{H.17}$$

so $(\zeta I - \chi G)^{-1} = K(wI - G)^{-1}$ and every spectral form of (S1)–(S3) becomes a resolvent form of G at w:

$$\begin{aligned} &F(\zeta,\chi) = K\,\Phi_N(w) = \bar{m}_1(\widehat{z}), \qquad \widehat{a}_N{}^\top(\zeta I - \chi G)^{-2}\widehat{a}_N = -K^2\Phi_N'(w), \\ &\frac{1}{\alpha_{x,N}}\left\langle \frac{\lambda}{(\zeta-\chi\lambda)^2}\right\rangle = -K^2\psi_N'(w), \qquad \frac{1}{\alpha_{x,N}}\left\langle \frac{\lambda}{\zeta-\chi\lambda}\right\rangle = K\,\psi_N(w). \end{aligned} \tag{H.18}$$

(a) The secular equation. On the branch $r_y > 0$, (S3) reads $s_N^2 F(\zeta,\chi) = 1/(\alpha_{y,N}\chi)$. By (H.18) and $1/(\alpha_{y,N}\chi) = \widehat{z}(1+\delta) = K/\alpha_{y,N}$ this is $s_N^2 K\Phi_N(w) = K/\alpha_{y,N}$, that is $\alpha_{y,N}s_N^2\Phi_N(w) = 1$. Since $\bar{D} = \bar{m}_1\bar{m}_2 = \alpha_{y,N}\Phi_N(w)$ by (E.3), and step (b) identifies δ with the branch solution of (E.2) at $\widehat{z}$, this is $\bar{D}(\widehat{z}) = 1/s_N^2$.

(b) The fixed point. The identity

$$\zeta(\zeta I - \chi G)^{-2} - \chi G(\zeta I - \chi G)^{-2} = (\zeta I - \chi G)^{-1}$$

contracts the two signal terms in $\zeta\cdot$(S1)$-\chi\cdot$(S2). It also applies to the spectral averages, giving

$$\frac{1}{\alpha_{x,N}}\left\langle \frac{\lambda}{\zeta-\chi\lambda}\right\rangle + s_N^2 r_y^2\,F(\zeta,\chi) = \zeta - \frac{1-r_y^2}{\alpha_{y,N}\chi}. \tag{H.19}$$

On the branch $r_y > 0$, (S3) gives $s_N^2 F = 1/(\alpha_{y,N}\chi)$ exactly, so the left r_y^2 term is $r_y^2/(\alpha_{y,N}\chi)$ and cancels its counterpart on the right, leaving a relation between ζ and χ alone,

$$\frac{1}{\alpha_{x,N}}\left\langle \frac{\lambda}{\zeta-\chi\lambda}\right\rangle = \zeta - \frac{1}{\alpha_{y,N}\chi}. \tag{H.20}$$

Translating through (H.18), the left side is $K\psi_N(w)$ and the right side is $\widehat{z} - \widehat{z}(1+\delta) = -\widehat{z}\,\delta$. Dividing $K\psi_N(w) = -\widehat{z}\delta$ by $\widehat{z}$ and using $K/\widehat{z} = \alpha_{y,N}(1+\delta)$ gives $\delta = -\alpha_{y,N}(1+\delta)\,\psi_N(w)$, which is (E.2) at $z = \widehat{z}$. The branch is the required one: $\chi, \zeta > 0$ give $1+\delta = 1/(\alpha_{y,N}\chi\zeta) > 0$, while $\zeta > \chi\lambda_{\max}(G)$ puts w above the spectrum by (H.17), so $\psi_N(w) > 0$ and $\delta \in (-1, 0)$.

(c) The right overlap. Translating (S1) through (H.18), $-K^2\psi_N'(w) - s_N^2 r_y^2 K^2\Phi_N'(w) = 1$, that is

$$r_y^2 = \frac{1 + K^2\psi_N'(w)}{-s_N^2 K^2\Phi_N'(w)}. \tag{H.21}$$

Along the branch $\bar{D}(z) = \alpha_{y,N}\Phi_N(w(z))$, so (H.14) gives $\bar{D}'(\widehat{z}) = 2K\alpha_{y,N}\Phi_N'(w)/(1 + K^2\psi_N'(w))$, while $\bar{m}_2(\widehat{z}) = 1/(\widehat{z}(1+\delta)) = \alpha_{y,N}/K$ by (E.3). Hence

$$\frac{-2\bar{m}_2(\widehat{z})}{s_N^2\,\bar{D}'(\widehat{z})} = \frac{-2\alpha_{y,N}/K}{2K\alpha_{y,N}s_N^2\Phi_N'(w)/(1+K^2\psi_N'(w))} = \frac{1+K^2\psi_N'(w)}{-s_N^2K^2\Phi_N'(w)} = r_y^2,$$

by (H.21), the first claim of (H.16).

(d) The sources. The only h_x-, h_p-dependence of (H.11) enters through $c_i = s_N r_y\widehat{a}_{N,i} + h_x u_{0,i} + h_p\widehat{a}_{N,i}$ of (H.8), and by Lemma H.1 together with the envelope theorem only this explicit dependence survives at the saddle. Differentiating $\frac{1}{2}\sum_i c_i^2/(\zeta - \chi\lambda_i)$ at $h = 0$,

$$\partial_{h_p}\Psi_h\big|_{h=0} = \sum_i \frac{s_N r_y\widehat{a}_{N,i}\cdot\widehat{a}_{N,i}}{\zeta-\chi\lambda_i} = s_N\,r_y\,F(\zeta,\chi) = s_N\,r_y\,\bar{m}_1(\widehat{z}),$$

and with $\widehat{a}_{N,i}$ replaced by $u_{0,i}$ in the differentiated factor, $\partial_{h_x}\Psi_h|_{h=0} = s_N\,r_y\,K\Phi_N^{\times}(w) = s_N\,r_y\,\bar{m}_1^{\times}(\widehat{z})$. Also, $\partial_{h_y}\Psi_h = r_y$. These are the remaining claims of (H.16). Squaring them and using $\bar{D} = \bar{m}_1\bar{m}_2 = 1/s_N^2$ at the root gives the formal squared overlap formulas, with the factor $(\bar{m}_1^{\times}/\bar{m}_1)^2$ as the squared ratio of the two source derivatives.

Converse. Let $(\widehat{z}, \delta)$ solve (E.2) and $\bar{D}(\widehat{z}) = 1/s_N^2$, with $\widehat{z} > 0$, $\delta \in (-1, 0)$, and $w = \alpha_{y,N}(1 + \delta)\widehat{z}^2 > \lambda_{\max}(G)$ on the increasing branch $1 + K^2\psi_N'(w) > 0$. Let $\zeta := \widehat{z}$ and $\chi := 1/K$, and define $r_y > 0$ by (H.21), positive there because $\Phi_N'(w) < 0$. Since $w > \lambda_{\max}(G)$, (H.17) gives $\zeta > \chi\lambda_{\max}(G)$, the domain of (H.12). Reading (a) backwards recovers (S3). The fixed point translated back through (H.18) is (H.20), which with (S3) restores the combination (H.19), and (H.21) is (S1) itself, so (S2) follows from the combination on dividing by $\chi > 0$. Its left side is strictly positive since $G \neq 0$, so $r_y^2 < 1$, as required for a spherical overlap. By (H.14) the branch condition says $w'(\widehat{z}) > 0$, so $\widehat{z}$ lies on the increasing arc of the branch and the correspondence between saddle and root is a bijection. □

H.4 The saddle value

The same calculation gives the stationary value on both zero-source branches. Multiplying (S1) by ζ and subtracting χ times (S2) gives (H.19). This identity uses only the first two stationarity equations, so it holds at both zero and positive response overlap. Substituting it into (H.12),

$$\Psi = \frac{\zeta}{2} + \frac{1}{2}\Big(\zeta - \frac{1 - r_y^2}{\alpha_{y,N}\chi}\Big) + \frac{1 - r_y^2}{2\alpha_{y,N}\chi} = \zeta.$$

On the positive branch $\zeta = \widehat{z}$, the candidate outlier location; at zero overlap $\zeta = z_{\mathrm{edge},N}$ by the calculation above. Under the formal limit exchanges, Lemma H.1 identifies $f(0)$ with $\lim_{N\to\infty} \mathbb{E}_{W_{\mathrm{G}}}\sigma_1(C_{\mathrm{G}})$. The positive saddle therefore predicts $f(0) = \lim_{N\to\infty}\widehat{z}$, where $\widehat{z}$ here is the finite-N conditional secular root. The zero-overlap branch gives the corresponding limit of $z_{\mathrm{edge},N}$. The squared overlaps do not depend on the sign choice of Lemma H.1. The signed identities in (H.16) do. At $G = I$ and $\widehat{a}_N = u_0$ the RS system reduces to the standard rectangular spiked-matrix formulas (13), evaluated in Proposition E.9(i).

I ADDITIONAL EXPERIMENTS

I.1 SIMULATION CONFIGURATIONS

We distinguish the asymptotic spectral threshold θ_c of (11), a detection half-crossing, and a recovery-level crossing. The detection half-crossing is the signal strength at which the probability that the top singular value exceeds a fixed null quantile reaches one half. A recovery-level crossing is the interpolated signal strength at which a trial-mean recovery curve last crosses a specified level. We apply the same crossing statistic to each available candidate curve; otherwise, we report its offset from the predicted transition.

A *supercritical* mean absolute error averages over grid points with $\theta > 1.1\,\theta_c$. We exclude $(\theta_c, 1.1\theta_c]$ because finite-size rounding dominates this band. Unless stated otherwise, bands and error bars show ± 1 SD across trials. Confidence intervals use percentile bootstraps over whole trial curves, with 2000 resamples.

For the response-retention collapse in Figure 1(b), we use Gaussian-QR designs with $N = 2000$, $(\alpha_x, \alpha_y) = (5, 20)$, $\rho_x = 0.7$, and $\rho_y \in \{1, 0.8, 0.6, 0.4, 0.2\}$, with 60 trials per point on a grid expressed in $\theta\sqrt{\rho_y}$. The angular study in Figure 2 uses one fixed flat orthogonal design with $N = 1024$, $D_x = 512$, $D_y = 256$, $(\alpha_x, \alpha_y) = (2, 4)$, $\rho_x \in [0.07, 1]$, and $\rho_y = 1$. It uses 60 mask and noise draws per (ρ_x, θ) grid point, with $\theta \in \{1, 2, 8\}$.

I.2 THEOREM-CLASS VALIDATION ON NESTED FLAT ORTHOGONAL DESIGNS

Corollary 3.6 covers nested Sylvester/Walsh–Hadamard designs, so a simulation inside the theorem class is available at every sample size of the form $N = 2^m$, the orders at which the nested construction exists. We build the design from a Walsh–Hadamard matrix of that order, take D_x of its columns, and randomize by independent row and column permutations and row and column sign changes, all of which the corollary allows. Every such design satisfies $|(X_\star)_{ki}| = 1$ and $X_\star^\top X_\star = N I_{D_x}$ to machine precision, so its leverage is exactly equal and $\tau_N = 1/\alpha_{x,N}$ exactly, and it meets the entry bound of Assumption C.1 with $\|X_\star\|_{\max} = 1$.

Figure 1(a) runs the transition protocol on such a design and on a Gaussian-QR design of the same size, against the same curves, at $N = 2048$, $D_x = D_y = 1024$, $(\alpha_x, \alpha_y) = (2, 2)$, $\rho_x = 0.3$ and $\rho_y = 0.6$. The grid has 25 equally spaced signal strengths in $[0.5, 2.5]\theta_c$ plus the isotropic reference threshold, with 100 trials per point. Within each trial, the design, both masks, and the noise are shared across the signal grid.

Supercritical mean absolute error is 0.001 for R_x^2 and 0.002 for R_y^2 on the flat orthogonal design, and 0.001 and 0.004 on Gaussian-QR. The recovery-level crossing, averaged over the levels $\{0.05, 0.10, 0.15, 0.20\}$ on the trial-mean R_x^2 curve with a percentile bootstrap over whole trial curves, is 1.619 and 1.626 against 1.629 for the limiting curve of Theorem 3.4 under the same rule, and both intervals cover it. At this resolution both design classes track the same limiting curves. This supports the universality prediction in Remark 3.7, suggesting that the deterministic equivalent may extend beyond the class in which we verify it.

I.3 THE PHASE DIAGRAM

Figure 3(a) maps recovery over the whole (θ, ρ) plane. The grid is 90×90 over $\theta \in [0.2, 2.0]$ and $\rho \in [0.05, 0.95]$ at $N = 2000$, $D_x = 300$, $D_y = 240$, with 30 trials per cell for $\theta \in [0.6, 1.8]\,\theta_c(\rho)$ and 6 trials elsewhere. Every cell reuses the same trial seeds, so the Monte Carlo errors of different cells are correlated. Each joint retention is realized symmetrically, $\rho_x = \rho_y = \sqrt{\rho}$, so this is one slice of the two-retention plane, and Figure 1(b) isolates the response direction. Over the grid the supercritical mean absolute error against the prediction is 0.007.

A separate run resolves the boundary itself, on a band of $\pm 35\%$ around $\theta_c(\rho)$ at every fourth retention with 60 trials per point. Over its 460 points the mean absolute deviation from the predicted curves is 0.029, the larger part of it below the threshold, where the prediction is identically zero and the deviation is the finite-N floor.

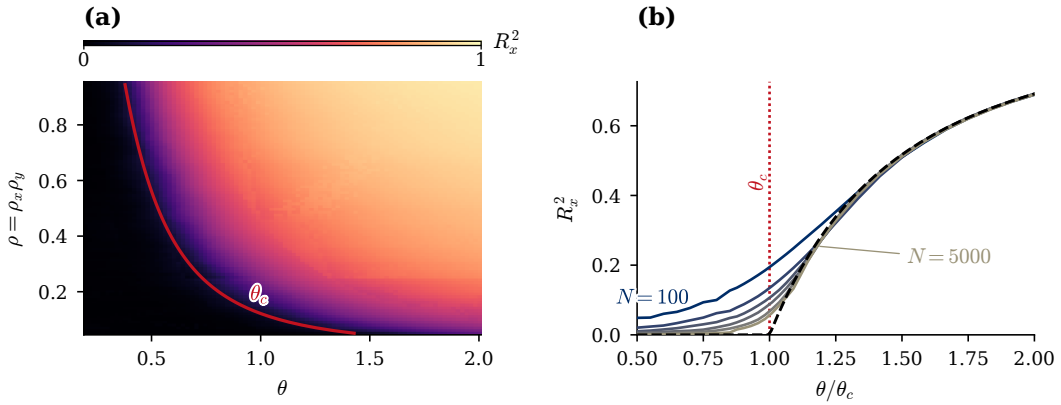


Figure 3: **Recovery across the (θ, ρ) plane. (a)** Measured R_x^2 on a 90×90 grid at $N = 2000$, $D_x = 300$, $D_y = 240$, with 30 trials per cell for $\theta \in [0.6, 1.8]\,\theta_c(\rho)$ and 6 elsewhere, and the predicted boundary $\theta_c(\rho)$ of (11) in red. Each joint retention is realized symmetrically, $\rho_x = \rho_y = \sqrt{\rho}$. The change in trial count at $1.8\,\theta_c(\rho)$ and the shared trial seeds leave faint seams in the field, along that curve and near $\rho = 0.25$. **(b)** The transition sharpens with sample size at $\alpha_x = \alpha_y = 2.5$ and $\rho_x = \rho_y = 0.8$, against the limiting curve in black. The unlabeled curves are $N = 250$, 500, 1000 and 2000, in that order between the two named. Repeats fall from 800 at $N = 100$ to 25 at $N = 5000$.

Panel (b) shows the transition sharpening as N grows. This finite-size sweep uses $(\alpha_x, \alpha_y) = (2.5, 2.5)$, $\rho_x = \rho_y = 0.8$, and 32 signal strengths in $[0.5, 2]\theta_c$. The sample sizes are $N \in \{100, 250, 500, 1000, 2000, 5000\}$, with trial counts decreasing from 800 to 25.

I.4 Robustness to non-Gaussian noise

Our analysis assumes Gaussian response noise. We test the predicted transition and recovery under changes in the noise tails and column variances.

Experimental setup. We use $N = 1000$, $D_x = 200$, $D_y = 150$, $(\alpha_x, \alpha_y) = (5, 6.67)$, and $\rho_x = \rho_y = 0.7$ ($\rho = 0.49$), with 100 trials at each of 25 signal strengths in $[0.6, 2.0]\theta_c$.

The five i.i.d. noise families have unit variance: standard Gaussian noise (excess kurtosis 0), standardized Student-t noise with $\nu = 5$, 4.5, and 3 (excess kurtosis 6, 12, and infinite, respectively), and standardized Laplace noise (excess kurtosis 3).

For the sixth, column-heteroskedastic setting, we draw independent $V_j \sim \mathrm{Uniform}[0.5, 1.5]$ and then $Z_{ij} \mid V_j \sim \mathcal{N}(0, V_j)$. It has unit average variance and marginal excess kurtosis 0.25. The shared column scales also change the covariance structure, so this setting tests anisotropy as well as the marginal noise distribution.

Results. In Figure 4 the i.i.d. noise families with finite fourth moments stay close to the Gaussian-model prediction curves, while Student-$t(\nu = 3)$ degrades in the supercritical regime. Supercritical mean absolute errors are 0.008 (Gaussian), 0.007 (t_5), 0.006 ($t_{4.5}$), 0.239 (t_3), 0.005 (Laplace) and 0.046 (column-heteroskedastic). The column-heteroskedastic setting also deviates and changes the response covariance structure. These results do not establish a monotone effect of kurtosis.

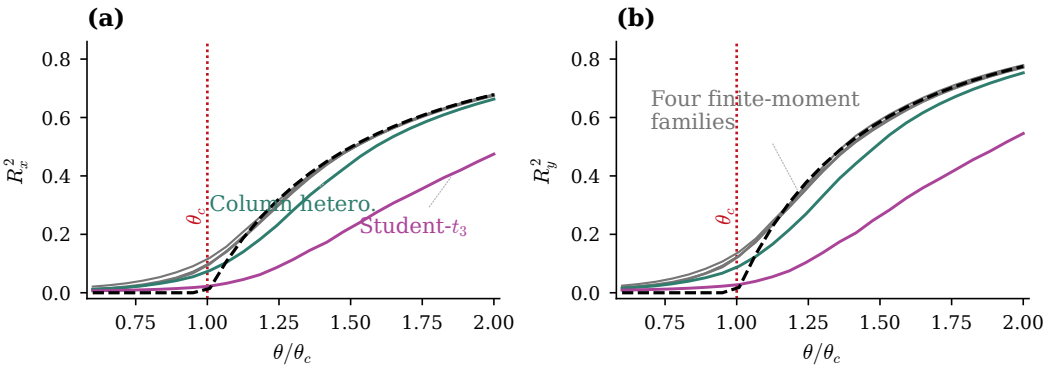


Figure 4: **Robustness of the phase transition to non-Gaussian noise.** Black is the Gaussian-model prediction and red is θ_c. Gaussian, Student-t at $\nu = 5$ and $\nu = 4.5$, and Laplace lie within 0.025 of one another, no more than the trial standard deviation, so their curves share a gray color and a label. **(a)** R_x^2, with separate labels for the two families that deviate, and **(b)** R_y^2, with the same ordering of the noise families. Parameters: $N = 1000$, $D_x = 200$, $D_y = 150$, $\rho_x = \rho_y = 0.7$, 100 trials.

I.5 Comparison with baseline imputation methods under MCAR

This experiment asks whether imputation improves on the zero-filled estimator under MCAR. We select all baseline hyperparameters on held-out observed entries to avoid arbitrary choices that could affect the comparison.

Methods. We compare seven estimators.

(1) **PLS-SVD:** zero-fill both views and form $C = (N\sqrt{\rho})^{-1} X^\top Y$.

(2) **Mean imputation:** replace missing entries by column means, then center both views.

(3) **Rank-one response imputation:** initialize both views by column means, then alternate PLS-SVD estimation with rank-one prediction of missing response entries. The design remains mean-imputed and observed response entries are retained. This alternating method has no tuning parameters.

(4) **Iterative SVD:** alternate rank-truncated reconstruction and restoration of observed entries (Troyanskaya et al., 2001). The rank is selected separately for each view from $\{1, 2, 5, 10, 20\}$ by squared error on a held-out tenth of its observed entries.

(5) **Nuclear-norm imputation:** use soft-impute (Mazumder et al., 2010), which alternates soft-thresholded SVD reconstruction with restoration of observed entries, selecting the threshold by squared error on a held-out tenth of each view's observed entries. We denote by λ_0 the largest singular value of that view's zero-filled training matrix after the held-out entries have been removed. The threshold grid includes λ_0, which gives the zero solution in the selection fit, and smaller thresholds. Appendix I.11 gives the full grid and the iteration limits.

(6) **Complete-data reference:** apply PLS-SVD to unmasked $(X_\star, Y_\star)$.

(7) **Prewhitened PLS-SVD:** re-whiten the masked design before forming the cross-covariance, then map the left direction back before scoring. This estimator is outside the scope of Theorem 3.3.

Selection protocol. Selection uses a realization independent of the evaluation trials, so no trial both chooses and scores its own hyperparameters. Selecting at each signal strength gives the baselines information a practitioner would not have, which makes the comparison conservative for the zero-filled estimator. The iterative-SVD rank selected on each view is 1 at every signal strength in both configurations.

Experimental setup. Figure 5 shows both configurations, sweeping θ from $0.5\theta_c$ to $2.0\theta_c$ (i) and to $3.5\theta_c$ (ii), with 20 and 24 grid points, respectively, and 30 trials per point: (i) $N = 1000$, $D_x = 200$, $D_y = 150$, $\rho_x = \rho_y = 0.7$; and (ii) $N = 800$, $D_x = 400$, $D_y = 200$ ($\alpha_x = 2$, $\alpha_y = 4$), $\rho_x = \rho_y = 0.3$.

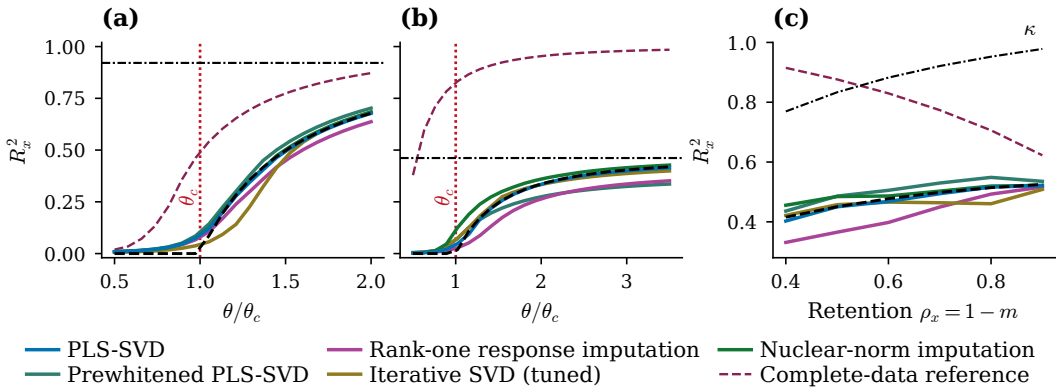


Figure 5: **Baseline imputation methods under MCAR.** The black dashed curve is the prediction for the zero-filled estimator, and the black dash-dot line is its ceiling κ, shown for reference and not as a bound on alternatives. Red marks θ_c. Centered mean imputation agrees with the zero-filled estimator to within 0.004, so we omit its curve. **(a,b)** Recovery curves in a moderate- and a low-retention configuration. The prewhitened direction is mapped back to the original design coordinates before scoring. Every imputation baseline has its hyperparameters selected on held-out observed entries. Complete-data PLS-SVD is a reference rather than an upper bound. Bands over trials are omitted for legibility. **(c)** Recovery at $\theta = 1.5\theta_c(\rho_x)$ versus the retention $\rho_x = \rho_y = 1 - m$ over $\{0.4, 0.5, \ldots, 0.9\}$, with m the missing fraction of each view, the sizes of configuration (i) and hyperparameters selected at each retention. Parameters: configuration (i), $N = 1000$, $D_x = 200$, $D_y = 150$, $\rho_x = \rho_y = 0.7$; configuration (ii), $N = 800$, $D_x = 400$, $D_y = 200$, $\rho_x = \rho_y = 0.3$; 30 trials.

Results. The zero-filled estimator tracks its predicted curves in both configurations. Under light masking (i) nuclear-norm and mean imputation give recovery similar to zero-fill, with mean supercritical R_x^2 of 0.499 and 0.491 against 0.492. The tuned iterative SVD gives 0.450, and re-whitening the design does better at 0.523.

Under heavy masking (ii), where the ceiling falls to $\kappa = 0.462$, the methods differ more: nuclear-norm imputation reaches 0.361 against 0.331 for zero-fill and 0.329 for the tuned iterative SVD, and re-whitening falls to 0.276.

Mean imputation matches zero-fill in both configurations once the imputed views are centered; without centering, the cross-product retains the rank-one mean term $N\hat{\mu}_x\hat{\mu}_y^\top$, with $\hat{\mu}_x, \hat{\mu}_y$ the column means of the imputed views. The tuned nuclear-norm baseline improves on the zero-filled estimator by 0.030 in the strongly masked regime. This paper makes no optimality claim for the zero-filled estimator; it locates that estimator's transition and the overlap it can recover, at both retentions.

On the whitened design view, the hold-out selector chooses the zero solution at every sampled signal strength in both configurations. On the response view, which carries the planted spike, the selector returns the zero solution at a fraction 0.17 of the supercritical grid in configuration (i), with interior thresholds in $[0.85, 0.97]\,\lambda_0$ otherwise, while in configuration (ii) interior thresholds in $[0.41, 0.90]\,\lambda_0$ win throughout its supercritical grid. Up to the scalar $\sqrt{\rho_x\rho_y}$, the zero-filled cross-covariance is unbiased for $X_\star^\top Y_\star/N$ under MCAR (Lounici, 2014); whether imputation improves on it is the question the comparison answers. Rank-one response imputation gives 0.450 and 0.263, so centering the initial fill does not resolve its poor recovery under heavy masking.

I.6 Robustness to correlated noise

The model of Section 3.1 assumes i.i.d. Gaussian noise. We examine structured noise covariances using a Toeplitz model $(\Sigma_Z)_{jk} = r^{|j-k|}$ with correlation parameter r and a factor model $\Sigma_Z = I + \sigma^2 FF^\top$. The factor matrix $F \in \mathbb{R}^{D_y \times 5}$ has independent standard Gaussian entries and is drawn once and held fixed. All runs use $N = 1000$, $D_x = 200$, $D_y = 150$, $\rho_x = \rho_y = 0.7$, 30 signal strengths in $[0.5, 4]\theta_c$, and 400 trials per point. Figure 6 shows the Toeplitz model at twenty

correlations from 0 to 0.95. Separate diagnostic runs use $r \in \{0, 0.3, 0.6, 0.9\}$ and factor variances $\sigma^2 \in \{0, 1, 3, 10\}$; their errors are reported below.

Naive PLS-SVD. Without response whitening, recovery degrades sharply with correlation. At Toeplitz $r = 0.9$ ($\mathrm{cond}(\Sigma_Z) \approx 350$), the naive curve differs substantially from the white-noise prediction (MAE 0.547). The factor model also inflates the noise variance. At $\sigma^2 = 1$, 3 and 10 the mean diagonal entry of Σ_Z is 6.2, 16.5 and 52.8, with $\mathrm{cond}(\Sigma_Z) \approx 205$, 612 and 2039. The naive MAE is 0.689, 0.690 and 0.690, and naive R_x^2 stays below 0.01 at every signal strength.

Oracle whitening. As an oracle diagnostic, we apply the known transformation $\Sigma_Z^{-1/2}$ to the complete response $Y_\star$ before masking and compute the adjusted spike $\theta \, \|\Sigma_Z^{-1/2} v_0\|$ per trial. This checks whether the theory, evaluated at the per-trial adjusted spike, describes the post-whitening effective model. It is not an in-sample estimator of Σ_Z. Across the four Toeplitz diagnostic levels and the four factor variances the whitened MAE is at most 0.004 (Toeplitz $r = 0.9$: 0.002).

I.7 Discriminating the finite-signal threshold

Figure 7 gives the detailed threshold check underlying Figure 1(c). At $(\alpha_x, \alpha_y) = (2, 2)$ and $\rho_y = 1$ we sweep ρ_x on an eighteen-point grid from 0.10 to 0.95, spanning regimes where the masked and isotropic predictions are well separated, nearly coincident, and reversed. Each configuration sweeps 44 signal strengths over $N \in \{500, 1000, 2000\}$ at $D_x = D_y = N/2$, with 100 trials per point and 500 null draws per configuration. Each null draw uses a new design. The signal grid is densified around both predictions so that neither is favored by its resolution.

Both crossings of Appendix I.1 are used. The detection half-crossing uses the 0.95 quantile of the simulated $\theta = 0$ top-singular-value distribution as its fixed null quantile. The recovery-level crossing averages the last crossings of the trial-mean R_x^2 curve through $L \in \{0.05, 0.10, 0.15\}$. Because a positive recovery-level crossing is offset from the asymptotic threshold, each candidate is scored by applying the same crossing statistic to its own predicted curve. At $N = 2000$ the measurements lie $+0.13$, -0.44 and -1.20 interval half-widths from the masked curve at $\rho_x = 0.15$, 0.30 and 0.80; Figure 7(b) shows the comparison with the reference curve across sample sizes and over the whole retention grid. For $\rho_x \geq 0.70$ the measurement at $N = 2000$ lies 1.3 to 2.1% below the crossing of the masked curve, and every interval excludes that crossing, narrowly at $\rho_x = 0.70$. The shortfall is 2.1 to 3.6% at $N = 1000$ and 5.4 to 6.9% at $N = 500$. At $\rho_x = 0.95$ the $N = 2000$ interval contains the comparator's crossing. The detection half-crossing has a positive finite-N offset that is smaller at $N = 2000$ than at $N = 500$ at every retention and agrees with neither candidate: at $\rho_x = 0.15$ it lies $+5.3$ interval half-widths from θ_c and -24.1 from the comparator. Comparing recovery-level crossings accounts for the offset between a positive-level crossing and the asymptotic threshold. The near-coincident $\rho_x = 0.50$ case serves as a calibration rather than a discriminating test.

The null runs also give the first direct check of the deformed edge. The reported singular values are normalized to the scale of $C = X^\top Y/(N\sqrt{\rho})$ in (3). On that scale, the median top singular value under the null approaches z_{edge} from below as N grows, at $\rho_x = 0.15$ by 0.93%, 0.64% and 0.46% at $N = 500$, 1000 and 2000, with the same monotone pattern at every retention.

I.8 Direction-specific ceilings on engineered and real designs

Theorem 3.3 holds for column-orthogonal designs that are incoherent; it does not require equal leverage, and the design enters only through the coupling τ_N. The finite-signal analysis of Appendix E, by contrast, is reduced to one scalar variable under equal leverage. Figure 8 separates the two statements on designs with deliberately unequal leverage, whitened to machine precision, and on the two biological designs.

Rows are drawn with a two-point row-scale pattern and the columns are then whitened, which makes $X_\star^\top X_\star = N I_{D_x}$ exact to machine precision. The whitening map compresses the nominal contrast, since the squared row norms are N times the diagonal of a hat matrix, so each resulting design is reported by its realized coefficient of variation rather than by its nominal one; no squared row norm can exceed $\alpha_{x,N}$ times the mean D_x. The planted direction is drawn toward the leading or

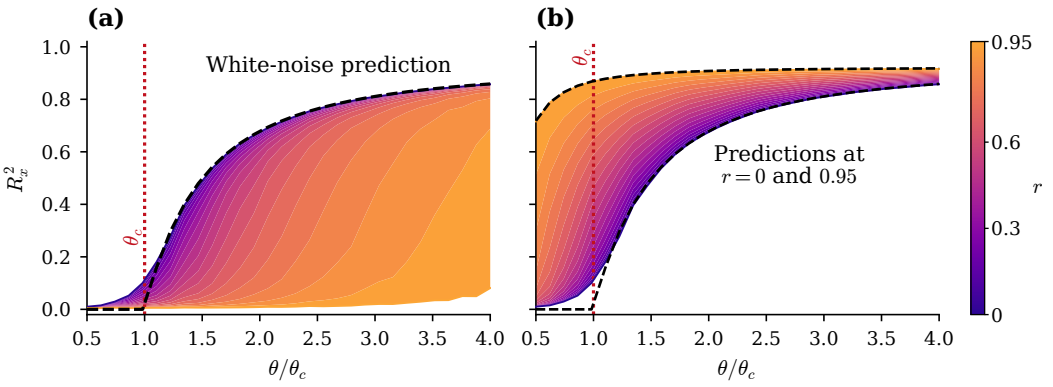


Figure 6: **Phase transition under correlated noise.** Color encodes the Toeplitz correlation r across twenty simulated levels from 0 to 0.95, and red marks θ_c. **(a)** Without whitening, recovery falls steadily as r rises, with the strongest correlations suppressing recovery across most of the displayed range. The black dashed curve is the white-noise prediction. **(b)** After oracle whitening by $\Sigma_Z^{-1/2}$ every level tracks its own prediction, the same curve evaluated at that trial's adjusted spike, with supercritical MAE at most 0.004. The two black dashed curves are the predictions at $r = 0$ and $r = 0.95$, and they bracket the field. Both panels use the same unadjusted horizontal axis, which is why the whitened field shifts left rather than collapsing. Parameters: $N = 1000$, $D_x = 200$, $D_y = 150$, $\rho_x = \rho_y = 0.7$, with 400 trials per signal strength.

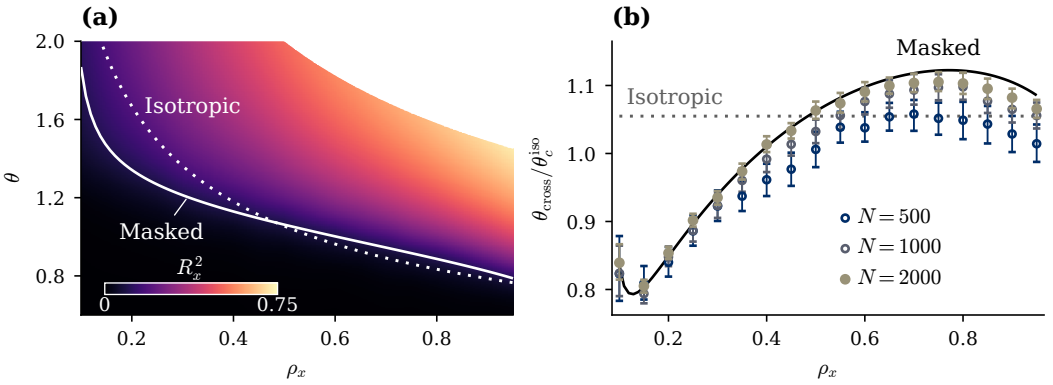


Figure 7: **Detailed finite-signal threshold validation.** **(a)** Mean R_x^2 over the (ρ_x, θ) plane at $N = 2000$ and eighteen retentions, with the level crossing θ_{cross} predicted by each candidate: solid for the masked theory, dotted for the isotropic comparator. The masked curve is the field's 0.1 contour to within 0.02, near the edge of the recovered region. At $\rho_x = 0.25$ the comparator lies at twice that recovery level, inside the recovered region. For $\rho_x > 0.5$ the field is truncated at the end of the signal grid. **(b)** Measured θ_{cross} in units of θ_c^{iso} for three sample sizes, with 95% bootstrap intervals. Figure 1(c) shows the same comparison on the original scale at $N = 2000$. The comparator's own crossing is 1.055, rather than 1, because a positive-level crossing lies above the asymptotic threshold of its curve. At $N = 2000$ the measurement stays within 1.6 interval half-widths of the masked curve. It lies below that curve at the highest retentions, with larger shortfalls at smaller sample sizes, and up to a quarter below the comparator where the two predictions separate. All runs use $D_x = D_y = N/2$, $\rho_y = 1$, and 100 trials per grid point.

the trailing eigenvector of $X_\star^\top \operatorname{diag}(\|(X_\star)_{k\cdot}\|^2) X_\star$, which in these realizations reaches couplings on both sides of the equal-leverage value $1/\alpha_x$; the construction does not guarantee couplings below the equal-leverage value. Parameters: $N = 2000$, $D_x = D_y = 500$, $(\alpha_x, \alpha_y) = (4, 4)$, $\rho_x = 0.5$, $\rho_y = 1$, 40 trials per cell.

At thirty times the equal-leverage threshold used for this experiment, the realized coupling spans 0.133 to 0.558 around the equal-leverage value 0.25, and the implied level $\kappa_{\tau,N}$ spans 0.642 to 0.883

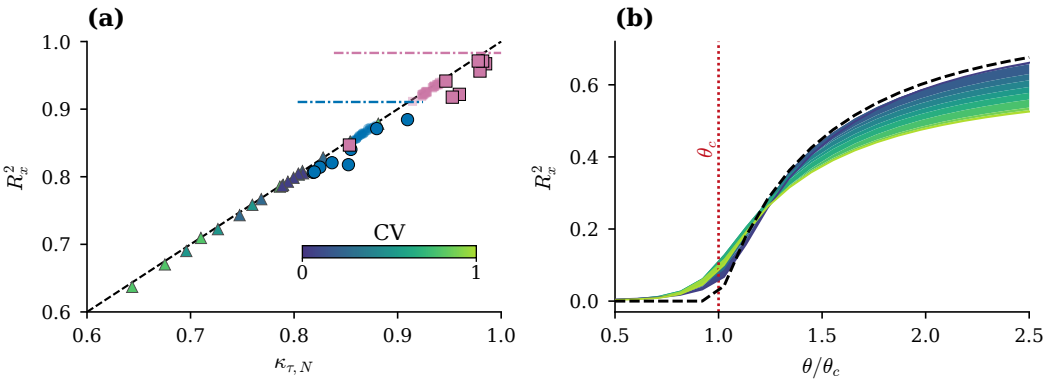


Figure 8: **Direction-specific ceilings on engineered and biological designs.** **(a)** Measured R_x^2 at $\theta = 30\,\theta_c$ against $\kappa_{\tau,N}$, with θ_c computed under equal leverage. On the dashed identity line, measured recovery equals the ceiling of Theorem 3.3. Triangles are engineered designs ($N = 2000$, $D_x = D_y = 500$, $\rho_x = 0.5$, $\rho_y = 1$, 40 trials), colored by the coefficient of variation (CV) of their sample leverages. Large markers are the eight leading directions of the biological designs TCGA (blue circles) and PBMC (pink squares), and small markers are 30 random directions in each ($\rho_x = \rho_y = 0.7$, 100 trials). Dash-dot lines show the equal-leverage value κ of each biological design. Standard errors are below marker size. **(b)** Measured R_x^2 on engineered designs at ten CV levels from 0.05 to 1.0, against the equal-leverage curve of Theorem 3.4 (black dashed). The region between adjacent CV levels is shaded.

around the equal-leverage $\kappa = 0.800$. At that signal strength the equal-leverage finite-signal shortfall is 0.001. The equal-leverage configurations provide a control confirming that it is negligible: their mean lies within 0.003 of κ, inside the shortfall plus twice the standard error. Raw recovery, with no finite-signal adjustment, follows the coupling each trial realizes: the mean absolute deviation from $\kappa_{\tau,N}$ is 0.011, against 0.049 from the equal-leverage ceiling, and a regression on the diagonal has slope 1.021 and intercept -0.019; the shortfall-adjusted deviations, reported as a secondary diagnostic, are 0.011 and 0.048. The high-signal limit therefore tracks the measured coupling on designs well outside the equal-leverage case, as Theorem 3.3 asserts.

The finite-signal equal-leverage approximation degrades in a measurable way. A separate finite-signal sweep uses 20 signal strengths in $[0.5, 2.5]\theta_c$ across 10 row-scale ratios, with 40 trials per cell. At the strongest coupling, the supercritical error of the equal-leverage curves grows from 0.018 at equal leverage (realized coefficient of variation 0.055) to 0.099 at a realized coefficient of variation of 0.998. This quantifies where the equal-leverage reduction remains usable, and connects the verified class to the biological designs of Section 4, which are neither flat nor equal in leverage.

A mask-only oracle. We remove response noise and take a leading left singular vector of $M = X^\top \operatorname{diag}(b) S_y \operatorname{diag}(v_0)$, where $X = S_x \odot X_\star$ and $b = X_\star u_0$. Its squared overlap with u_0 is the mask-only oracle recovery. The exact identity $M/(N\sqrt{\rho}) = p_N v_0^\top + D_N$ retains the finite-N residual defined after Proposition C.5. Under Assumption C.1, Appendix D gives oracle recovery $\kappa_{\tau,N} + o_P(1)$ as $N \to \infty$. For each biological dataset, we use $D_x = D_y = 200$, $\rho_x = \rho_y = 0.7$, 100 trials, and signal strength $30\theta_c$, with the equal-leverage reference threshold. The two datasets use independent simulations. We pair the oracle and measured recovery on each trial's masks as a finite-sample diagnostic.

The couplings vary widely across directions. Raw recovery follows them: the mean absolute deviation from $\kappa_{\tau,N}$ is 0.017 on TCGA and 0.018 on PBMC for the leading complete-data directions, against 0.078 and 0.046 from the equal-leverage value κ, and 0.003 and 0.002 for random directions in the same designs.

The random comparison directions are drawn once and held fixed across trials. The shortfall below $\kappa_{\tau,N}$ is -0.017 ± 0.0007 and -0.018 ± 0.0011, formed per trial since the eight directions of a dataset share their masks. The paired mask-only oracle accounts for most of the shortfall below $\kappa_{\tau,N}$. After pairing, both datasets retain small residuals that are distinguishable from zero at the reported Monte Carlo precision.

Their magnitudes are comparable to the equal-leverage finite-signal gaps. These gaps are direction-independent, so we report them only for scale and do not subtract them from the per-direction residuals. No direction-specific finite-signal prediction is available, so the comparison does not establish zero error or an exact deterministic equivalent.

Detection. We compare the detection half-crossings with matched Haar controls of the same size and aspect ratio. On TCGA the detection half-crossing is $1.130\,\theta_c$ at $\rho_x = \rho_y = 0.7$ against $1.129\,\theta_c$ for its control, and $1.071\,\theta_c$ against 1.143 at $\rho_x = \rho_y = 0.5$. On PBMC it lies above its control by 0.097 at $\rho_x = \rho_y = 0.7$ and by 0.238 at 0.5, in units of θ_c. These finite-sample detection crossings therefore show a larger departure from the matched control as retention falls; they do not directly estimate the asymptotic threshold.

Table 2 reports $\kappa_{\tau,N}$ at $\rho_x \in \{0.9, 0.7, 0.5\}$ for the eight leading complete-data directions of each dataset. These values are calculated directly from the designs, without Monte Carlo trials.

Table 2: Direction-specific ceilings and finite-N incoherence readings on the two biological designs. For each leading complete-data direction the coupling τ_N fixes $\kappa_{\tau,N}$ at any retention, with no Monte Carlo. Columns (i)–(iv) are the finite-N readings of Assumption C.1 at this one size, each divided by its logarithmic bound; they describe the designs rather than test the asymptotic condition.

			$\kappa_{\tau,N}$ at ρ_x			Incoherence reading			
Design	Direction	τ_N	0.9	0.7	0.5	(i)	(ii)	(iii)	(iv)
PBMC ($N = 5000$)	1	0.036	0.996	0.985	0.966	2.57	2.52	8.08	0.25
	2	0.098	0.989	0.960	0.911	2.34	2.47	8.08	0.93
	3	0.116	0.987	0.953	0.896	2.42	2.38	8.08	1.35
	4	0.049	0.995	0.980	0.953	2.60	2.57	8.08	0.56
	5	0.401	0.957	0.853	0.714	1.68	1.41	8.08	5.23
	6	0.043	0.995	0.982	0.959	2.27	2.14	8.08	0.70
	7	0.053	0.994	0.978	0.950	1.67	1.72	8.08	0.70
	8	0.132	0.986	0.947	0.884	1.97	1.03	8.08	1.66
	equal leverage	0.040	0.996	0.983	0.962	n/a	n/a	n/a	n/a
TCGA ($N = 873$)	1	0.404	0.957	0.852	0.712	2.15	1.50	2.70	0.49
	2	0.232	0.975	0.910	0.812	1.66	1.65	2.70	0.45
	3	0.396	0.958	0.855	0.716	1.48	1.19	2.70	0.97
	4	0.319	0.966	0.880	0.758	1.16	0.87	2.70	0.78
	5	0.456	0.952	0.837	0.687	1.03	1.21	2.70	1.42
	6	0.518	0.946	0.818	0.659	0.78	0.89	2.70	1.71
	7	0.495	0.948	0.825	0.669	0.78	0.94	2.70	1.87
	8	0.514	0.946	0.820	0.661	0.70	1.06	2.70	1.24
	equal leverage	0.229	0.975	0.911	0.814	n/a	n/a	n/a	n/a

I.9 Semi-synthetic preprocessing

Table 3: Datasets used in semi-synthetic experiments, with the measured score–leverage couplings τ_N of Definition 3.2 for biological and random signal directions, the implied ceilings $\kappa_{\tau,N}$, and the equal-leverage value κ, all at $\rho_x = 0.7$. The biological columns give the leading complete-data direction, direction 1 of Table 2. The random columns give the mean $\pm$ SD of τ_N and the mean of $\kappa_{\tau,N}$ over the 30 random directions of Figure 8(a).

Dataset	Views	N	$D_x = D_y$	τ_N^{bio}	$\kappa_{\tau,N}^{\mathrm{bio}}$	τ_N^{rand}	$\kappa_{\tau,N}^{\mathrm{rand}}$	κ
TCGA BRCA	RNA-seq / methylation	873	200	0.404	0.852	0.361 ± 0.021	0.866	0.911
PBMC Multiome	scRNA-seq / scATAC-seq	5000	200	0.036	0.985	0.171 ± 0.020	0.932	0.983

For TCGA BRCA we use the `HiSeqV2` gene-level expression matrix ($\log_2$-transformed normalized counts, as distributed) and the `HumanMethylation450` beta-value matrix, both from the TCGA Breast Cancer (BRCA) cohort on the UCSC Xena public hub (Goldman et al., 2020), transposed to samples by features and paired by sample identifier ($N = 873$). Features containing missing values are dropped, each remaining feature is standardized, each view is reduced to 200 principal

components, and the component scores are eigen-whitened: centered, rotated to the eigenbasis of the empirical covariance, and scaled so that $X_w^\top X_w = NI$.

For PBMC Multiome we use the filtered feature matrix of the 10k PBMC Multiome dataset from which granulocytes were removed by cell sorting (10x Genomics, 2021). The RNA view is filtered to cells with at least 200 detected genes and genes detected in at least 3 cells, depth-normalized to 10^4 counts per cell, log-transformed ($\log(1+x)$), and restricted to the 2000 most highly variable genes. The ATAC view is filtered with the same thresholds and restricted to its first 5000 peak features in file order, with no further transformation before standardization. The file orders peaks by chromosome and then by position, so these 5000 peaks span positions 9790 to 68,214,518 of `chr1` in GRCh38. Cells present in both views are paired by barcode and subsampled to $N = 5000$. Both views are then standardized, reduced to 200 principal components, and eigen-whitened as above.

All preprocessing, including whitening, uses complete data, and the MCAR masks of the protocol in Section 4 are applied afterwards, to the whitened design and the synthetic response.

I.10 Aspect-ratio sensitivity

Figure 9 tests the predicted curves of Theorem 3.4 on Gaussian-QR designs across extreme full-column-rank aspect ratios.

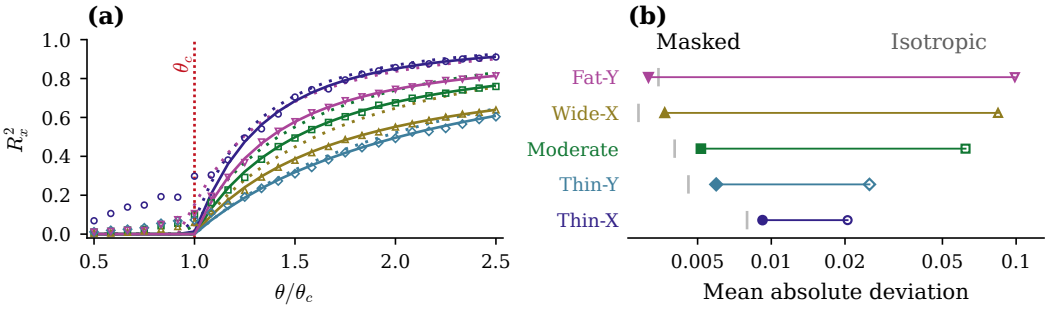


Figure 9: **Aspect-ratio sensitivity.** **(a)** Five aspect-ratio regimes ($\alpha_x, \alpha_y \geq 2$), one color each: markers show measurements over 100 trials, solid curves show the masked prediction, and dotted curves show the isotropic comparator. Below the threshold both predictions are zero while measured recovery has begun to increase. **(b)** Mean absolute deviation from the measurement over $\theta > 1.1\theta_c$: filled for the masked prediction, open for the comparator, gray for the trial-noise level, which the masked prediction stays near in every regime. All runs use $N = 1000$ and $\rho_x = \rho_y = 0.7$.

Extreme aspect ratios. We use $N = 1000$, $\rho_x = \rho_y = 0.7$ ($\rho = 0.49$), and 100 trials at each of 25 signal strengths in $[0.5, 2.5]\theta_c$. The five full-column-rank configurations have the following supercritical MAEs: Thin-X ($\alpha_x = 50$, $\alpha_y = 6.7$, MAE 0.009), Moderate ($\alpha_x = 5$, $\alpha_y = 6.7$, MAE 0.005), Wide-X ($\alpha_x = 2$, $\alpha_y = 6.7$, MAE 0.004), Thin-Y ($\alpha_x = 5$, $\alpha_y = 50$, MAE 0.006), and Fat-Y ($\alpha_x = 5$, $\alpha_y = 2$, MAE 0.003). All achieve $r \geq 0.996$. The isotropic comparator deviates most in the Wide-X and Fat-Y regimes, up to MAE 0.099: Wide-X has the largest ceiling gap and Fat-Y the largest edge gap.

I.11 Implementation of the baseline estimators

Rank-one response imputation centers both mean-imputed views, recenters the response after each fill, and stops after 50 updates or once the absolute cosine between successive estimates of each direction exceeds $1 - 10^{-6}$. Iterative SVD starts from column means and stops after 50 iterations or once the relative change of the filled matrix in Frobenius norm falls below 10^{-4}. Nuclear-norm imputation starts from zero and stops after 200 iterations or once the relative change of its low-rank estimate falls below 10^{-4}. Its threshold grid is $\{1, 0.99, 0.97, 0.95, 0.92, 0.90, 0.85, 0.80\}\,\lambda_0$ followed by 14 geometrically spaced values from $0.70\lambda_0$ to $0.02\lambda_0$, and the selection fits them in decreasing order, each starting from the previous fit. Iterative SVD and nuclear-norm imputation use these limits in selection and in evaluation, and both apply PLS-SVD to the filled views without centering.

J LIMITATIONS

The model assumes a rank-one signal, fixed MCAR retentions, i.i.d. Gaussian response noise, and whitening before masking, with $D_x \leq N$. The finite-signal theorem treats fixed $\theta \neq \theta_c$, leaving critical-window behavior open. For design masking with $\rho_x < 1$, the finite-signal curves for QR/Haar designs remain universality predictions; proving the deterministic equivalent there requires control of a random variance profile. The finite-signal curves for biological designs are equal-leverage approximations, and persistent leverage variation requires a different deterministic equivalent. The high-signal theorem applies more generally under incoherence. The biological experiments combine measured designs with synthetic responses and known directions, so they test recovery within the specified response model. Structured missingness and multiple signal components remain open. Further priorities are whitening estimated from incomplete observations and characterizing recovery thresholds and ceilings in latent-factor models with additive noise in both views.

In the latent-factor model with additive noise in both views, a shared score generates the design and the response. The shared score couples the noisy design, corrupted direction, and product noise, so this specification does not supply the conditional structure used here. The present guarantees should therefore not be transferred to that setting without new analysis.

K BROADER IMPACTS

Multimodal datasets with missing entries are the norm in computational biology and healthcare, and shared components extracted from them enter downstream analyses that are expensive to validate. For specified model parameters and a specified candidate direction, the model threshold and the direction-specific ceiling $\kappa_{\tau,N}$ are computable from the design, and Table 2 shows that on real designs the ceiling at moderate retention varies materially across the leading complete-data directions. Both are model-based reference quantities rather than guarantees: passing the threshold says recovery is possible within the model, not that the model matches the data.

The theory assumes MCAR masking and a design whitened before masking. Missingness in biomedical data is often structured, so applying the MCAR threshold there requires additional justification. The design-side ceiling κ_τ, strictly below 1 when $\rho_x < 1$ and the coupling τ is positive, means design-side loadings are not consistent estimates even at arbitrarily strong signal.

L AI USE STATEMENT

We used generative AI tools for several tasks with required disclosure. They assisted in writing code, including estimators, simulations, and data preprocessing, and provided feedback on the experimental design. They did not write the proofs. They assisted in revising proof text, gave feedback and suggested clarity improvements, most substantially in Appendix F and Appendix G, and helped restate theorem hypotheses so that each matches the class its proof covers. We did not use them to develop the model, to propose hypotheses, to supply the key ideas of the proofs, to generate data sets, to translate or to interpret results, and qualitative data analysis does not apply to this work. We also used them for tasks with recommended disclosure, namely reviewing and finding literature, drafting and editing text for readability, producing the figures, and checking and formatting references.

We reviewed all AI-assisted work. The authors checked every proof in the appendix line by line. Separate AI sessions verified the proofs and the references as an additional check, not as a substitute for our own. We checked each cited result against its source and each bibliography entry against the

publisher record. We take responsibility for the final content of this work, including text, claims, or artifacts produced with the aid of generative AI.

M REPRODUCIBILITY STATEMENT

Code is available at:

`https://anonymous.4open.science/r/Missing-Data-Induced-Phase-Transitions-in-Spectral-PLS/`

The assumptions behind every result are stated in Appendix A, and Table 1 records what is proved for which design class. Complete proofs are in Appendices B–G, and Appendix H is a replica derivation rather than a proof. Appendix I.1 gives the experimental configurations and Appendix I.9 the preprocessing of the biological designs.