ADAPTIVE INVARIANCE PCA FROM AUGMENTATION PORTFOLIOS

Anonymous authors
Paper under double-blind review

ABSTRACT

Learning from an augmentation collection requires knowing which background variation it exposes and how strongly to suppress it. We study signal-preserving linear views with correlated backgrounds and shared noise. Complementary channels can identify the background space even when every pair is insufficient; we characterize their required number. We propose Adaptive Invariance PCA (AIPCA), a matrix-power path connecting pooled PCA, normalized alignment, and a limit that suppresses variation across views. Under complete coverage, its population limit eliminates background while retaining all signal coordinates when signal and background spaces are disjoint. Proportional-dimensional risk theory explains the tradeoff between background suppression and estimation error: intermediate powers can outperform the reference methods, whereas sufficiently strong signals favor pooled PCA. Source-wise validation selects the power using covariance profiles and paired comparisons with a reconstruction reference. Synthetic and controlled real-image experiments examine coverage, risk predictions, and adaptation to the available views.

1 INTRODUCTION

Data augmentation creates views of the same original example, or *source*, to preserve semantic signal while varying nuisance background. A reusable augmentation rule is a *channel*; its application produces a *view*. Practitioners assemble a portfolio of channels, such as crops, color changes, and blur, without knowing whether they expose every relevant background direction. Positive views may still share illumination or clutter. The learning problem is therefore to determine which changes identify background and how strongly to use them with finite data.

Pooled principal component analysis (PCA) retains large-variance directions (Jolliffe, 2002); invariance instead favors directions that change little across views. Averaging reduces changing background but retains its shared component. Paired differences cancel signal and reveal changed background. Complementary channels can jointly reveal directions that every fixed pair misses (Figure 1). Our *Adaptive Invariance PCA (AIPCA)* balances pooled variance and within-source stability through one matrix-power path, with a fixed covariance normalization and a power selected from source tuples.

Contributions. We make two contributions. First, we characterize complete background coverage and a necessary, attainable bound on the number b of channels: $b \geq 1 + \lceil m/r \rceil$, when each channel changes at most r of m background directions relative to a reference. Complete coverage identifies the background space; disjoint signal and background spaces then permit background elimination without losing signal coordinates. Orthogonality gives recovery of the original signal space. A Gaussian construction shows that a complementary portfolio can identify this space even when every fixed pair is observationally ambiguous (Theorem B.1).

Second, we develop the power path, its population recovery, and its proportional-dimensional risk. The risk formula separates directional distortion from sampling error and identifies when intermediate powers improve on pooled PCA, normalized alignment, and difference PCA. With other parameters fixed, strong signal favors powers approaching zero, making pooled PCA the leading-order choice. Source-wise validation selects the power through covariance profiles and paired comparisons with a reconstruction reference. Experiments connect coverage, predicted risk, and this adaptive procedure.

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.

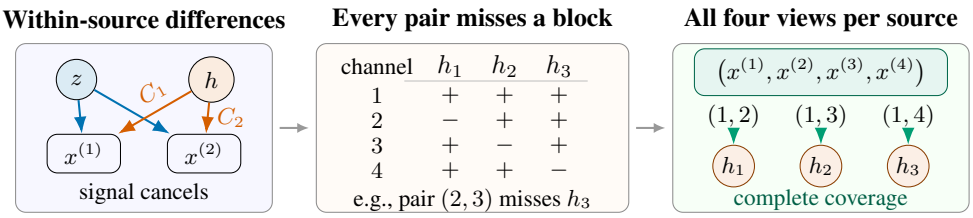


Figure 1: **Complementary channels cover background.** Each source supplies all four views. Differences cancel signal z; actions C_e preserve $(+)$ or flip $(-)$ background blocks h_j. Every pair misses a block, whereas the complete tuple covers all three.

1.1 Related work

PCA++ learns normalized linear representations by cross-view alignment (Wu et al., 2025); slow feature analysis minimizes normalized variation across related observations (Wiskott & Sejnowski, 2002; Escalante-B. & Wiskott, 2013). Our $\alpha = 1$ candidate uses symmetric pooled normalization, connecting to their untruncated generalized eigenproblems; Appendix C states the precise correspondence. Unlike continuum power canonical correlation analysis (CCA) (Swenson, 2015), our power acts on paired differences under fixed pooled normalization.

Invariant risk minimization (IRM) seeks a common optimal predictor across environments (Arjovsky et al., 2019); Principal Invariant Component Analysis (PICA) seeks equal representation distributions (Norman & Meir, 2026). AIPCA instead uses same-source pairing, which can reveal background even when channel marginals agree. Augmentation-based identification (von Kügelgen et al., 2021) and counterfactual matching (Bai et al., 2025; 2026) also exploit nuisance changes. We quantify complementary channel coverage and adapt its statistical use; Appendix A expands these connections.

Notation. For a matrix C, $\mathrm{range}(C) = \mathrm{col}(C)$, $\ker(C)$, $\|C\|_{\mathrm{op}}$, and $\|C\|_{\mathrm{F}}$ denote its column space, kernel, operator norm, and Frobenius norm; $\mathrm{sym}(C) = (C + C^{\mathrm{T}})/2$. For full-column-rank A, let $U_A = A(A^{\mathrm{T}}A)^{-1/2}$ and $P_A = U_A U_A^{\mathrm{T}}$; define U_B, P_B analogously. For rank-k projectors, $\ell(P,Q) = \|P - Q\|_{\mathrm{F}}^2/2$. Matrix powers act on eigenvalues: if $M = U\,\mathrm{diag}(\lambda_j)U^{\mathrm{T}} \succ 0$, then $M^{\alpha} = U\,\mathrm{diag}(\lambda_j^{\alpha})U^{\mathrm{T}}$.

2 Model setup

Each independent source i supplies a complete tuple of $b \geq 2$ views in $\mathbb{R}^d$:

$$x_i^{(e)} = Az_i + BC_e h_i + \varepsilon_i^{(e)}, \qquad z_i \in \mathbb{R}^k, \quad h_i \in \mathbb{R}^p, \quad C_e \in \mathbb{R}^{m\times p}. \tag{1}$$

The full-column-rank loadings $A \in \mathbb{R}^{d\times k}$ and $B \in \mathbb{R}^{d\times m}$ encode signal and background directions and strengths. Channel e preserves Az_i and applies the same unknown background action C_e across sources. The latent h_i can collect both shared background and channel-specific randomness (Remark B.1). Neither factors nor actions are observed.

Assumption 2.1 (Factors and observation noise)**.** The noise is $\varepsilon_i^{(e)} = \sigma\sqrt{1-\chi^2}\,\xi_i + \sigma\chi\,\zeta_i^{(e)}$, with $\sigma \geq 0$, $\chi \in [0,1]$. All generating vectors $z_i, h_i, \xi_i, \zeta_i^{(1)}, \ldots, \zeta_i^{(b)}$ have zero mean and identity covariance and are mutually independent within and across sources; $\xi_i, \zeta_i^{(e)} \in \mathbb{R}^d$.

Thus ξ_i is inherited source noise and $\zeta_i^{(e)}$ is view-specific: $\chi = 0$ gives fully shared noise and $\chi = 1$ independent view noise. Factor scales are absorbed into the loadings and actions; Gaussianity is required only for specific risk results.

With $\overline{x}_i = b^{-1}\sum_e x_i^{(e)}$, define

$$D = \frac{1}{b}\sum_e \mathrm{Cov}(x_i^{(e)}), \qquad S = \frac{1}{b-1}\sum_e \mathbb{E}[(x_i^{(e)} - \overline{x}_i)(x_i^{(e)} - \overline{x}_i)^{\mathrm{T}}]. \tag{2}$$

The pooled covariance D measures total variation; the contrastive covariance S measures within-source variation, canceling signal, shared noise, and unchanged background.

3 From background elimination to signal recovery

We first characterize when channel differences identify the background space. Removing this space need not identify the signal: additional noise directions can remain. Pooled variance then selects a representation that retains all signal coordinates within the background complement.

3.1 Coverage and the role of a reference channel

Let $\overline{C} = b^{-1} \sum_e C_e$. The model gives

$$H = \frac{1}{b-1} \sum_e (C_e - \overline{C})(C_e - \overline{C})^{\mathrm{T}}, \qquad S = \sigma^2 \chi^2 I_d + BHB^{\mathrm{T}}. \tag{3}$$

Complete coverage, $H \succ 0$, means that channel changes jointly span all m background directions. Equivalently, the reference differences $C_e - C_1$, $e > 1$, span this space.

Proposition 3.1 (Coverage and channel budget)**.** Under model (1) and Assumption 2.1, if $H \succ 0$, then

$$\ker(S - \sigma^2 \chi^2 I_d) = \mathrm{col}(B)^{\perp}. \tag{4}$$

Projection onto this space eliminates background. If also $A^{\mathrm{T}} B = 0$ and $d = k + m$, it equals $\mathrm{col}(A)$. If $\mathrm{rank}(C_e - C_1) \le r$ for $e > 1$, where the integer r satisfies $1 \le r \le \min(m, p)$, complete coverage requires

$$b \ge 1 + \lceil m/r \rceil. \tag{5}$$

For every fixed $p \ge r$, suitable actions attain this bound.

For example, $m = 6, r = 2$ requires a reference and three complementary two-dimensional changes; repeating one block cannot suffice. *Additional channels improve coverage when they reveal new directions.* Appendix B proves the result and extends coverage to incomplete pair observations.

The sign-flip construction makes the information gain explicit. With Gaussian factors and matched signal/background variances, all channel marginals coincide. For any fixed pair, an unchanged background coordinate can be exchanged with a signal coordinate without changing the full pair distribution. A complementary portfolio removes this ambiguity (Theorem B.1). Thus pairing and channel diversity provide information beyond marginal covariance differences.

3.2 The population estimator and its interpretation

For $\sigma\chi > 0$, both $D, S \succ 0$. We choose the linear representation $V^{\mathrm{T}} x$, $V \in \mathbb{R}^{d \times k}$, by

$$V_\alpha^\star \in \arg\min_{V^{\mathrm{T}} D V = I_k} \mathrm{tr}(V^{\mathrm{T}} S^\alpha V), \qquad \alpha \ge 0. \tag{6}$$

The fixed *hard uniformity* constraint gives unit-variance, uncorrelated coordinates (Wu et al., 2025; Wang & Isola, 2020; Wang & Liu, 2021). For a common eigenvector of D, S, its cost is s_j^α / d_j: large pooled variance d_j and small contrast variance s_j are preferred. Positive rescaling of S does not change the selected space.

For example, let $k = 1$, $D = \mathrm{diag}(3, 6)$, and $S = \mathrm{diag}(1, 2)$, with the first coordinate signal. Its cost is $1/3$, versus $2^\alpha/6$ for background. Pooled PCA selects background, power one ties, and every $\alpha > 1$ selects signal. A power above one can therefore help when background has both large pooled variance and substantial cross-view variation.

The path gives pooled PCA at $\alpha = 0$ and normalized alignment at $\alpha = 1$, since minimizing $\mathrm{tr}(V^{\mathrm{T}} S V)$ maximizes $\mathrm{tr}\{V^{\mathrm{T}}(D - S)V\}$. With a gap after its kth smallest eigenvalue, the large-power limit is the bottom-k eigenspace of S, or difference PCA. If more than k directions share the smallest contrast variance, pooled variance selects among them (Appendix C).

Proposition 3.2 (Invariant recovery)**.** Under model (1) and Assumption 2.1, suppose $H \succ 0$, $\sigma\chi > 0$, and $\mathrm{col}(A) \cap \mathrm{col}(B) = \{0\}$. Let $Q = I_d - P_B$ and let P_α project onto any minimizing space in (6). Then

$$P_\alpha \longrightarrow P_\infty \quad (\alpha \to \infty), \qquad \mathrm{range}(P_\infty) = \mathrm{col}(QA). \tag{7}$$

This is the leading k-dimensional eigenspace of QDQ. Any basis V_∞ of this space satisfies

$$V_\infty^{\mathrm{T}} B = 0, \qquad \mathrm{rank}(V_\infty^{\mathrm{T}} A) = k. \tag{8}$$

If $A^{\mathrm{T}} B = 0$, then $\mathrm{col}(V_\infty) = \mathrm{col}(A)$, also when $d > k + m$.

Because $QDQ = (QA)(QA)^{\mathrm{T}} + \sigma^2 Q$, pooled variance separates retained signal from additional noise-only directions. Thus the representation preserves every signal coordinate up to an invertible linear transformation and observation noise; nonorthogonality generally changes its loading space. Intersecting signal and background spaces cannot satisfy both conditions in (8). Under orthogonality, sufficiently large finite powers recover the signal; partial coverage additionally requires signal variance to exceed that of unchanged background (Theorem C.1).

With finite data, the power balances background suppression against contrast-covariance estimation error. Section 5 shows why strong signal favors pooled PCA; Section 4 selects the power from data.

4 Adaptive Invariance PCA

AIPCA fits the population power path from complete source tuples and selects its power by source-wise validation. It screens alternatives against a reconstruction reference, then ranks retained powers by their covariance fit.

4.1 Sample moments and fitting

For n fitting sources, write $\overline{x}_i = b^{-1} \sum_e x_i^{(e)}$. Use the known zero means in the model and simulations; for real data, subtract the pooled fitting mean. The sample moments are

$$\widehat{D} = \frac{1}{nb} \sum_{i,e} x_i^{(e)} x_i^{(e)\mathrm{T}}, \qquad \widehat{S} = \frac{1}{n(b-1)} \sum_{i,e} (x_i^{(e)} - \overline{x}_i)(x_i^{(e)} - \overline{x}_i)^{\mathrm{T}}. \tag{9}$$

Assuming $\widehat{D}, \widehat{S} \succ 0$, replace the population moments in (6) to obtain

$$\widehat{V}_\alpha \in \arg \min_{V^{\mathrm{T}} \widehat{D} V = I_k} \mathrm{tr}(V^{\mathrm{T}} \widehat{S}^\alpha V), \qquad \widehat{P}_\alpha = \widehat{V}_\alpha (\widehat{V}_\alpha^{\mathrm{T}} \widehat{V}_\alpha)^{-1} \widehat{V}_\alpha^{\mathrm{T}}. \tag{10}$$

The orthogonal projector $\widehat{P}_\alpha$ makes comparisons independent of basis or scale; $\alpha = 0$ gives pooled PCA. A stable generalized eigensolver is given in Supplement C.1.

4.2 Selecting the power by source-wise validation

Keep all views of each source in the same split. For a finite grid $\mathcal{A}$, fit every $\widehat{P}_\alpha$ on training sources and hold the fitted spaces and centering fixed during validation. Given background rank $1 \le m \le d - k$, let $\widehat{U}_B$ contain the top m training contrast eigenvectors and set $\widehat{Q} = I_d - \widehat{U}_B \widehat{U}_B^{\mathrm{T}}$. Experiments supply m; an unknown rank must be estimated from training data.

Reference and paired screening. For n_{val} held-out sources, choose the reconstruction reference

$$R \in \arg\min_{\alpha \in \mathcal{A}} \widehat{L}_{\mathrm{inv}}(\alpha), \qquad \widehat{L}_{\mathrm{inv}}(\alpha) = \frac{1}{n_{\mathrm{val}} b} \sum_{i,e} \|\widehat{Q} x_i^{(e)} - \widehat{P}_\alpha x_i^{(e)}\|^2. \tag{11}$$

Since an inaccurate $\widehat{Q}$ can mislead reconstruction, we also compare each candidate with R using retained contrast variation $\widehat{C}_\alpha$ and noise-adjusted source-mean variation $\widehat{E}_\alpha$. Put $q = d - k - m$; the latter is defined only when $q > 0$:

$$\widehat{C}_\alpha = \mathrm{tr}(\widehat{P}_\alpha \widehat{S}_{\mathrm{val}}), \qquad \widehat{E}_\alpha = \frac{b}{n_{\mathrm{val}}} \sum_i \left\{ \|\widehat{P}_\alpha \widehat{Q} \overline{x}_i\|^2 - \frac{\mathrm{tr}(\widehat{P}_\alpha \widehat{Q})}{q} \|F_D^{\mathrm{T}} \overline{x}_i\|^2 \right\}.$$

Here $\widehat{S}_{\mathrm{val}}$ uses the held-out sources in (9), and F_D contains the bottom q training pooled-covariance eigenvectors. Thus smaller $\widehat{C}$ favors stability, whereas larger $\widehat{E}$ favors retained source variation above a common noise baseline. The first term in $\widehat{E}$ measures retained energy after estimated background removal. The quantity $\|F_D^{\mathrm{T}}\overline{x}_i\|^2/q$ estimates noise energy per direction; $\mathrm{tr}(\widehat{P}_\alpha\widehat{Q})$ weights that baseline by how much of the candidate survives background removal. The factor b puts the comparison on the scale of $\sqrt{b}\,\overline{x}_i$. When $q = 0$, omit $\widehat{E}$ and its screening condition.

For candidate-minus-reference differences $\widehat{\Delta}^{C,E}_{\alpha R}$, form upper contrast and lower source-mean bounds:

$$U^C_{\alpha R} = \widehat{\Delta}^C_{\alpha R} + \widehat{c}\,\widehat{\mathrm{SE}}^C_{\alpha R}, \qquad L^E_{\alpha R} = \widehat{\Delta}^E_{\alpha R} - \widehat{c}\,\widehat{\mathrm{SE}}^E_{\alpha R}.$$

Standard errors use paired source differences; the critical value $\widehat{c}$ uses a joint Gaussian approximation over all candidate pairs, allowing R to depend on validation data. This is a working calibration for overlapping cross-validation (CV) folds, without guaranteed nominal simultaneous coverage. Retain

$$\widehat{\mathcal{A}}_{\mathrm{adm}} = \{R\} \cup \{\alpha \in \mathcal{A} \setminus \{R\} : U^C_{\alpha R} \le 0, \quad L^E_{\alpha R} \ge 0 \text{ if } q > 0\}. \tag{12}$$

The reference always remains available. An alternative must have an upper contrast bound at most zero and, when measured, a lower source-mean bound at least zero. An inconclusive comparison does not admit that alternative; the covariance ranking chooses among supported alternatives and the reference.

Covariance ranking and refitting. For each candidate, fit a rank-m background space orthogonal to $\widehat{P}_\alpha$ using training data. Partition held-out source-mean energy among the representation, fitted background, and residual spaces, and contrast energy inside and outside the background; omit the residual group when $q = 0$. These groups define measured variances $h^{\mathrm{val}}_{\alpha r}$, with weights w_r counting spatial and contrast coordinates. There are five groups (four when $q = 0$): the mean-representation group has weight k, while contrast groups have an extra factor $b - 1$ for the contrast channels. Rank admitted powers by

$$\widehat{F}_\alpha = \min_{\boldsymbol{\eta} \in \mathcal{V}_q} \sum_r w_r\{\log \eta_r + h^{\mathrm{val}}_{\alpha r}/\eta_r\}, \qquad \widehat{\alpha} \in \arg\min_{\alpha \in \widehat{\mathcal{A}}_{\mathrm{adm}}} \widehat{F}_\alpha.$$

The variance levels $\boldsymbol{\eta}$ satisfy positivity and the signal/background-above-noise and mean-above-contrast orders collected in $\mathcal{V}_q$. Minimizing over them with the training spaces fixed is *profiling*; smaller scores mean a better working Gaussian covariance fit. This background is candidate-specific, whereas $\widehat{E}_\alpha$ uses the common $\widehat{Q}$. Supplement D specifies the groups, variance fitting, calibration, and zero-standard-error convention.

Break exact score ties toward smaller power. Independent holdout returns the selected training fit. Source-wise CV re-estimates all training quantities within each fold, averages foldwise profiled and reconstruction scores using source-count weights, and pools paired source differences for calibration. It then refits the selected power on all sources. Screening compares two validation quantities with the reference; it does not guarantee noninferiority in true subspace risk or downstream accuracy. Its conditional risk interpretation is given in Supplement D.

5 Risk and the Choice of Power

We quantify the tradeoff between suppressing background and estimating invariance in an orthogonal Gaussian submodel.

5.1 Risk along the power path

Consider the special case of model (1)

$$x_i^{(e)} = \sqrt{a}\,U_A z_i + \sqrt{\kappa}\,U_B g_i + \sqrt{\tau}\,U_B h_{ie} + \varepsilon_i^{(e)}, \qquad i = 1, \dots, n, \quad e = 1, \dots, b. \tag{13}$$

Here U_A, U_B have orthonormal columns. Signal z_i and persistent background g_i are shared by all views; background h_{ie} changes across views. Their variance strengths are a, κ, τ, respectively.

Assumption 5.1 (Gaussian model and proportional regime)**.** In addition to Assumption 2.1, assume $k = 1$, $U_A^\top U_B = 0$, and that z_i, g_i, h_{ie} and the generating noise vectors are independent standard Gaussian vectors, with the sharing specified in (13). Let $a > 0$, $\kappa, \tau \geq 0$, $\sigma > 0$, and $0 < \chi \leq 1$. With b fixed, take $m, n \to \infty$, $m/n \to c > 0$, and $(d-1-m)/n \to c_\perp \geq 0$, where $c + c_\perp < b - 1$.

The ratios count background and additional noise-only dimensions per source. Each source supplies $b - 1$ Gaussian contrasts, so their aspect ratio $c_b = (c + c_\perp)/(b - 1) < 1$ ensures asymptotic invertibility. Contrasts have covariance $S = \sigma^2\chi^2 I + \tau P_B$, with $P_B = U_B U_B^\top$; they remove shared components but retain changing background. The source mean instead has covariance

$$\mathrm{Cov}(\overline{x}_i) = aP_A + (\kappa + \tau/b)P_B + \sigma^2(1 - \chi^2 + \chi^2/b)I_d.$$

Thus more views average down changing background and independent noise, whereas persistent background and shared noise remain. Pooled variance can distinguish signal from noise-only directions that contrasts treat alike. Define a direction's score as its pooled variance divided by its powered contrast variance; larger scores are preferred. The population signal-to-background score ratio is

$$\frac{\text{signal score}}{\text{background score}} = \frac{a + \sigma^2}{\kappa + \tau + \sigma^2}\left(1 + \frac{\tau}{\sigma^2\chi^2}\right)^\alpha. \tag{14}$$

Pooled PCA uses the variance advantage in the first factor. Positive powers reward smaller contrast variance, but estimating that variance can rotate the fitted direction. Persistent background lowers the first factor while disappearing from contrasts. Against a noise-only direction, the score ratio is instead $(a + \sigma^2)/\sigma^2$, independent of α. Powers therefore help when contrasts distinguish background from signal; pooled variance is still needed to distinguish signal from noise.

Let $\widehat{u}_\alpha$ be the unit fitted direction in original coordinates, with loss $\ell(\widehat{P}_\alpha, P_A) = 1 - |\widehat{u}_\alpha^\top U_A|^2$. Zero loss means perfect recovery and loss one means orthogonality. We call the signal *detected* when its limiting squared overlap is positive, corresponding to a signal eigenvalue separated from the spectrum generated without the signal. For a detected signal, T_α is the effective inverse covariance response after averaging over source-mean sampling, with the signal removed and the contrast covariance held fixed. Its effect on the signal direction is summarized by

$$\Gamma_\alpha = \operatorname*{plim}_{n\to\infty} \frac{\|T_\alpha U_A\|^2}{(U_A^\top T_\alpha U_A)^2} \geq 1.$$

Here plim denotes convergence in probability. The reciprocal Γ_α^{-1} is the limiting squared cosine between $T_\alpha U_A$ and U_A, so $\Gamma_\alpha = 1$ means no directional distortion. The nonnegative term Ω_α measures additional source-mean sampling error. Appendix E.2 defines the response and both risk factors precisely.

Theorem 5.1 (Risk at every fixed power)**.** Under Assumption 5.1, fix finite $\alpha \geq 0$. If $c_\perp > 0$, assume the regular upper-edge condition (49) in the appendix: the upper edge of the no-signal spectrum is reached before the conditional inverse response becomes singular. No additional edge condition is needed for $c_\perp = 0$. Then

$$\ell(\widehat{P}_\alpha, P_A) \xrightarrow{p} \mathcal{R}_\alpha, \qquad \mathcal{R}_\alpha = 1 - \frac{1}{\Gamma_\alpha + \Omega_\alpha}, \qquad \Gamma_\alpha \geq 1, \quad \Omega_\alpha \geq 0, \tag{15}$$

if the signal eigenvalue separates strictly above the no-signal spectrum, as specified in (45) in the appendix. Otherwise $\mathcal{R}_\alpha = 1$, including at the detection boundary.

Thus useful powers balance suppression against directional distortion and sampling error. The factors are deterministic functions of the model parameters; Appendix E gives their evaluation and proofs, with scalar formulas when $c_\perp = 0$.

5.2 What determines the useful power?

Strong signals favor pooled PCA. First take the proportional limit, then let $a \to \infty$, fixing all other parameters. With contrast-variance ratio $\rho_c = \sigma^2\chi^2/(\sigma^2\chi^2 + \tau)$,

$$\mathcal{R}_0 \longrightarrow 0, \qquad \mathcal{R}_1 \longrightarrow \frac{c\rho_c + c_\perp}{b - 1 - c + c\rho_c} > 0. \tag{16}$$

Pooled variance already identifies a sufficiently strong signal. Indeed, at strong signal the fitted direction is approximately proportional to $\widehat{S}^{-\alpha}U_A$: power zero preserves U_A, whereas positive powers rotate it through an estimated contrast covariance. This sampling cost persists in proportional dimensions. *On any fixed finite candidate grid containing zero, pooled PCA eventually minimizes the limiting risk as signal strength grows.* A continuum of powers allows a small further improvement. On any fixed $[0, \alpha_{\max}]$, $0 < \alpha_{\max} < \infty$, its unique risk minimizer eventually satisfies

$$\alpha_*(a) = \frac{K}{aV_0} + O(a^{-2}), \qquad \mathcal{R}_{\alpha_*(a)} - \mathcal{R}_0 = -\frac{K^2}{a^2 V_0} + O(a^{-3}). \tag{17}$$

Here $K > 0$ combines background suppression with a correction for contrast sampling fluctuations; $V_0 > 0$ is the variance of log contrast eigenvalues weighted by their squared signal projections. Their definitions and proofs are in Appendix E.5. For $\alpha = O(a^{-1})$, the local risk change is $\mathcal{R}_\alpha - \mathcal{R}_0 = V_0\alpha^2 - 2K\alpha/a + O(a^{-3})$: distortion is quadratic in power, while its benefit is linear. The useful power therefore shrinks at rate $1/a$, with only an a^{-2} gain over PCA.

Strong changing background can favor difference PCA. When $m = d - 1$, with no additional noise-only directions, the relative contrast gap $\delta = \tau/(\sigma^2\chi^2 + \tau)$ separates signal from background. Difference PCA detects the signal if $\delta > \sqrt{c/(b-1)}$: the gap must exceed contrast sampling fluctuations. For $b = 2, c = 1/2, a = 2, \kappa = 0, \sigma = \chi = 1, \tau = 9$,

$$(\mathcal{R}_0, \mathcal{R}_1, \mathcal{R}_{\text{diff}}) = (1, 1/3, 5/36).$$

Here the difference-PCA risk $\mathcal{R}_{\text{diff}}$ takes infinite power at each sample size before the proportional limit, separately from the fixed-finite-power theorem. Large background variance defeats pooled PCA, but its strong changes make the signal identifiable through contrasts. Appendix E gives the endpoint formula and the effect of additional noise-only directions.

Intermediate powers use both sources of information. For $b = 2, c = 1/2, c_\perp = 0, a = 2, \kappa = 0, \sigma = \chi = 1, \tau = 4/5$,

$$(\mathcal{R}_0, \mathcal{R}_{1/2}, \mathcal{R}_1, \mathcal{R}_{\text{diff}}) \simeq (0.682, 0.399, 0.735, 1).$$

Contrasts alone cannot detect the signal, but still help suppress background. Power $1/2$ uses this information with less distortion than power one, improving on all three references. Figure 2 compares theoretical curves with empirical losses across three regimes.

More views improve both covariance estimates. At fixed $c, c_\perp$, increasing b lowers the contrast aspect ratio and the changing-background and view-noise variances in source means. Persistent background and shared noise remain. Thus view count alone does not determine the preferred power: signal strength and background persistence also matter. For $\tau > 0$, greater persistence increases K/V_0, favoring more suppression in the strong-signal expansion.

6 Experiments

We test coverage, power selection, and representation quality with supplied background ranks and the prescribed grid

$$\mathcal{A}_{20} = \{0\} \cup \{2^{j/2} : j = -6, \ldots, 12\}. \tag{18}$$

Synthetic studies report normalized subspace loss $\ell(\widehat{P}, P_A)/k$; real-image studies report accuracy after selecting representations without labels. The three-regime study uses five-fold source-wise CV and refitting; coverage and real images use independent-source holdout. Appendix F gives the complete protocols and supporting results.

6.1 Complementary channels and coverage

With two signal and six background coordinates, three channels leave background uncovered; a fourth completes coverage by changing a new two-dimensional block. Across 25 trials with 4,000 fitting and 1,000 validation sources, mean loss drops from 0.6625 to 0.00114. A fifth channel repeats an existing action and adds no new coverage. This illustrates Proposition 3.1: channel complementarity determines what the power path can identify.

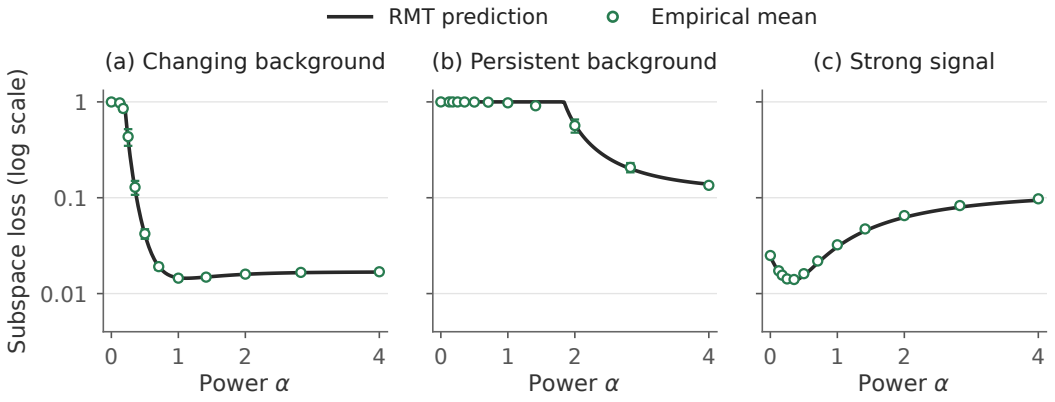


Figure 2: **Power choice in three regimes.** Lines: risk predicted by random matrix theory (RMT; Theorem 5.1); points: empirical mean loss with ± 1.96 Monte Carlo standard errors over 24 trials. Left to right: changing background, persistent background, and stronger signal. Here $b = 8$, with the remaining parameters in Section 6.2; adaptive selection uses the full grid (18).

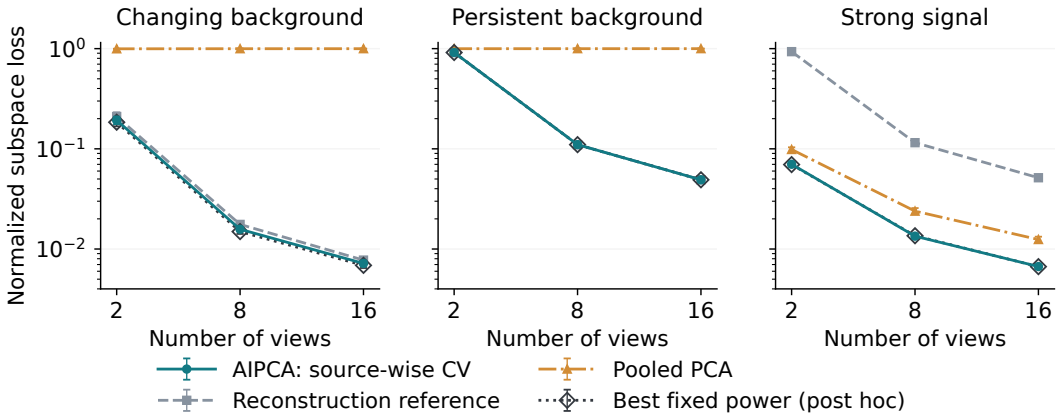


Figure 3: **Source-wise CV from Section 4.2.** Mean subspace losses over 24 fresh trials, with pointwise ± 1.96 Monte Carlo standard errors (log scale). CV refits its selected power on all sources; the best fixed power is chosen retrospectively within each regime–view setting for evaluation only. With $m = 120$ and $q = 0$, screening uses only the contrast comparison.

6.2 Fixed-power risk and adaptive selection

We consider changing background, persistent background, and strong signal in model (13), with $d = 121, n = 340, m = 120$ and $b \in \{2, 8, 16\}$. The strengths (a, κ, τ) are $(3, 0, 4), (2, 4, 1), (6, 0, 1)$, respectively, with $\sigma = \chi = 1$. Figure 2 compares theoretical and empirical fixed-power risks. Figure 3 evaluates the Section 4.2 selector on 24 new independent trials per setting after the rule was fixed. With strong signal and two views, CV attains mean loss 0.0701, versus 0.0985 for pooled PCA and 0.9331 for reconstruction alone; the best fixed-power mean is 0.0696, an evaluation benchmark unavailable to the selector. The finite strength $a = 6$ permits a gain from positive powers, consistent with the eventual PCA limit as $a \to \infty$. The paired improvement over PCA is 0.02836 (Monte Carlo standard error 0.00177). Here reconstruction favors the estimated one-dimensional background complement, while covariance profiles also use the signal information in pooled variation. Selection remains close to the best fixed-power means in the other two spanning regimes as views increase. In the strong-signal, two-view setting, profile-only selection makes the same choices; screening leaves its gain unchanged. Appendix F.2 reports component ablations and additional sample sizes, multiple signals, non-Gaussian sources, and geodesic-loss checks.

Table 1: **Linear representations under controlled illumination.** Accuracy (%), mean ± sample standard deviation over five paired seeds. The columns use different source partitions; the diagnostic setting comes from a broader sharing scan. Deviations quantify randomization variability on this benchmark. Higher is better. Darker/lighter shading marks the best/second-best distinct displayed means within each column; rounding ties share a shade.

Method	Original illumination (official test)	Shared background, $\rho = 0.99$ (exploratory diagnostic)
Pooled PCA	67.55 ± 2.95	71.84 ± 3.10
Fixed $\alpha = 1$	84.42 ± 0.55	72.12 ± 3.12
AIPCA (source-wise holdout)	84.41 ± 0.34	86.04 ± 2.48
Difference PCA	61.66 ± 3.82	63.64 ± 3.87

6.3 Real-image experiments

Four fixed illumination channels augment the 8×8 Optdigits images (Alpaydin & Kaynak, 1998). Source-wise holdout selects an eight-dimensional representation using $m = 3$; a fixed ridge classifier then uses fitting labels. Table 1 compares the original illumination protocol on 1,797 official test images and an exploratory shared-background setting on 1,000 diagnostic images drawn from the official training file. At sharing fraction $\rho = 0.99$, holdout selects $\alpha = 2$ in every split and substantially improves on power one; under original illumination their accuracies are similar. Both evaluations are exploratory, reusing previously examined splits, augmentations, fits, and classifier predictions under a frozen selection rule. The reconstruction reference is the sole admissible candidate in all 30 comparisons across the original protocol and five sharing levels. These results therefore assess power adaptation through the reconstruction reference; Appendix F.4 reports the full scan.

7 Conclusions and Discussions

Complementary channels determine which background directions are identifiable; AIPCA selects how strongly to suppress their empirical variation. Its risk theory explains when intermediate powers help and why sufficiently strong signals favor pooled PCA. Source-wise validation selects the power from observed tuples, adapting the use of invariance to the available views.

The present analysis assumes signal-preserving linear channels, and the exact proportional-risk theorem is established for an orthogonal Gaussian rank-one model. Background unchanged across channels remains invisible to differences, while validation relies on estimated background structure. Risk guarantees for source-wise CV with full-data refitting remain an open problem. Further directions include broader signal and noise models, nonlinear representations, and source versus view allocation under a fixed observation budget.

AI use statement.

GPT-6 Astral assisted with code, language editing, and literature search and summaries. The authors reviewed and verified all AI-assisted work and take full responsibility for the paper.

References

Abubakar Abid, Martin J. Zhang, Vivek K. Bagaria, and James Zou. Exploring patterns enriched in a dataset with contrastive principal component analysis. *Nature Communications*, 9(1):2134, 2018.

Ethem Alpaydin and Cenk Kaynak. Optical recognition of handwritten digits. UCI Machine Learning Repository, 1998.

Massih R. Amini, Nicolas Usunier, and Cyril Goutte. Learning from multiple partially observed views—an application to multilingual text categorization. In *Advances in Neural Information Processing Systems*, volume 22, pp. 28–36. Curran Associates, Inc., 2009.

Martin Arjovsky, Léon Bottou, Ishaan Gulrajani, and David Lopez-Paz. Invariant risk minimization. *arXiv preprint arXiv:1907.02893*, 2019.

Ruqi Bai, Yao Ji, Zeyu Zhou, and David I. Inouye. From invariant representations to invariant data: Provable robustness to spurious correlations via noisy counterfactual matching. *arXiv preprint arXiv:2505.24843*, 2025.

Ruqi Bai, Yao Ji, Mingyu Kim, Easton Currie, Zeyu Zhou, and David I. Inouye. Provable robustness to spurious correlations via invariant data for robust finetuning. AISTATS 2026 Workshop on Causality in the Age of AI Scaling, 2026.

Zhidong Bai and Jian-Feng Yao. On sample eigenvalues in a generalized spiked population model. *Journal of Multivariate Analysis*, 106:167–177, 2012.

R. A. Bailey and Peter J. Cameron. Using graphs to find the best block designs. In Lowell W. Beineke and Robin J. Wilson (eds.), *Topics in Structural Graph Theory*, volume 147 of *Encyclopedia of Mathematics and its Applications*, pp. 282–317. Cambridge University Press, 2013.

Afonso S. Bandeira, March T. Boedihardjo, and Ramon van Handel. Matrix concentration inequalities and free probability. *Inventiones Mathematicae*, 234:419–487, 2023.

Victor Chernozhukov, Denis Chetverikov, and Kengo Kato. Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *The Annals of Statistics*, 41(6): 2786–2819, 2013.

Cian Eastwood, Julius von Kügelgen, Linus Ericsson, Diane Bouchacourt, Pascal Vincent, Bernhard Schölkopf, and Mark Ibrahim. Self-supervised disentanglement by leveraging structure in data augmentations. *arXiv preprint arXiv:2311.08815*, 2023.

Alberto N. Escalante-B. and Laurenz Wiskott. How to solve classification and regression problems on high-dimensional data with a supervised extension of slow feature analysis. *Journal of Machine Learning Research*, 14(116):3683–3719, 2013.

Lea Flumian, Markus Matilainen, Klaus Nordhausen, and Sara Taskinen. Stationary subspace analysis based on second-order statistics. *Journal of Computational and Applied Mathematics*, 436:115379, 2024.

Luigi Gresele, Paul K. Rubenstein, Arash Mehrjou, Francesco Locatello, and Bernhard Schölkopf. The incomplete Rosetta Stone problem: Identifiability results for multi-view nonlinear ICA. In *Proceedings of the Thirty-Fifth Conference on Uncertainty in Artificial Intelligence*, volume 115 of *Proceedings of Machine Learning Research*, pp. 217–227. PMLR, 2020.

Zhiwei Han, Stefan Matthes, and Hao Shen. Provable subspace identification of nonlinear multi-view CCA. In *Proceedings of the 42nd Conference on Uncertainty in Artificial Intelligence*, volume 337 of *Proceedings of Machine Learning Research*, pp. 1947–1982. PMLR, 2026.

Ian T. Jolliffe. *Principal Component Analysis*. Springer, second edition, 2002.

Jon R. Kettenring. Canonical analysis of several sets of variables. *Biometrika*, 58(3):433–451, 1971.

Antti Knowles and Jun Yin. Anisotropic local laws for random matrices. *Probability Theory and Related Fields*, 169:257–352, 2017.

Frank C. Meinecke, Paul von Bünau, Motoaki Kawanabe, and Klaus-Robert Müller. Learning invariances with stationary subspace analysis. In *2009 IEEE 12th International Conference on Computer Vision Workshops*, pp. 87–92. IEEE, 2009.

Yotam Norman and Ron Meir. Unsupervised representation learning—an invariant risk minimization perspective. In *International Conference on Learning Representations*, 2026.

Debashis Paul. Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statistica Sinica*, 17(4):1617–1642, 2007.

Mikael Sørensen, Charilaos I. Kanatsoulis, and Nicholas D. Sidiropoulos. Generalized canonical correlation analysis: A subspace intersection approach. *IEEE Transactions on Signal Processing*, 69:2452–2467, 2021.

Roland Speicher and Carlos Vargas. Free deterministic equivalents, rectangular random matrix models, and operator-valued free probability theory. *Random Matrices: Theory and Applications*, 1(2):1150008, 2012. With an appendix by Tobias Mai.

Erik Swenson. Continuum power CCA: A unified approach for isolating coupled modes. *Journal of Climate*, 28(3):1016–1030, 2015.

Behrooz Tahmasebi, Melanie Weber, and Stefanie Jegelka. Data augmentation: A Fourier analysis perspective. In *Proceedings of the Thirty-Ninth Conference on Learning Theory*, volume 336 of *Proceedings of Machine Learning Research*, pp. 6114–6155. PMLR, 2026.

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

Paul von Bünau, Frank C. Meinecke, Franz C. Király, and Klaus-Robert Müller. Finding stationary subspaces in multivariate time series. *Physical Review Letters*, 103(21):214101, 2009.

Julius von Kügelgen, Yash Sharma, Luigi Gresele, Wieland Brendel, Bernhard Schölkopf, Michel Besserve, and Francesco Locatello. Self-supervised learning with data augmentations provably isolates content from style. In *Advances in Neural Information Processing Systems*, volume 34, pp. 16451–16467. Curran Associates, Inc., 2021.

Feng Wang and Huaping Liu. Understanding the behaviour of contrastive loss. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2495–2504, 2021.

Qinwen Wang and Jianfeng Yao. Extreme eigenvalues of large-dimensional spiked Fisher matrices with application. *The Annals of Statistics*, 45(1):415–460, 2017.

Tongzhou Wang and Phillip Isola. Understanding contrastive representation learning through alignment and uniformity on the hypersphere. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 9929–9939. PMLR, 2020.

Wenjing Wang, Xin Zhang, and Lexin Li. Common reducing subspace model and network alternation analysis. *Biometrics*, 75(4):1109–1120, 2019.

John W. Wilkinson. An analysis of paired comparison designs with incomplete repetitions. *Biometrika*, 44(1–2):97–113, 1957.

Laurenz Wiskott and Terrence J. Sejnowski. Slow feature analysis: Unsupervised learning of invariances. *Neural Computation*, 14(4):715–770, 2002.

Mingqi Wu, Qiang Sun, and Archer Y. Yang. PCA++: How uniformity induces robustness to background noise in contrastive learning. In *Advances in Neural Information Processing Systems*, volume 38, pp. 145293–145340. Curran Associates, Inc., 2025.

Tete Xiao, Xiaolong Wang, Alexei A. Efros, and Trevor Darrell. What should not be contrastive in contrastive learning. In *International Conference on Learning Representations*, 2021.

SUPPLEMENTARY MATERIAL

Overview. Appendix A supplements the related work. Appendix B develops coverage and channel budgets; Appendix C gives spectral computation and population recovery. Appendix D specifies source-wise validation and its conditional risk comparisons. Appendix E develops proportional-dimensional risk and power choice, and Appendix F documents the experiments.

Guide to the proofs of the main results.

- **Coverage and the channel budget.** Appendix B.1 proves Proposition 3.1: channel differences remove shared components, and the contrast covariance above its isotropic noise floor identifies background under complete coverage. A rank bound and complementary-block construction give the channel budget.
- **Population recovery.** Appendix C.2 proves Proposition 3.2 using the spectral construction in Appendix C.1. The power filter converges to the background-complement projector; pooled variance then selects $\mathrm{col}(QA)$. Under orthogonality, Theorem C.1 gives finite-power recovery conditions.
- **Risk at a fixed finite power.** The derivation for Theorem 5.1 starts from the independent source-mean and contrast covariances in Appendix E.1. Appendix E.2 develops the risk calculation under the stated regular-edge condition. Appendix E.3 gives scalar formulas for the spanning submodel.

SUPPLEMENT CONTENTS

A Additional related work **13**

B Model, moments, and coverage **13**

B.1 Coverage and the channel budget 14

B.2 Equal marginals and the information in a portfolio 15

C Population recovery and computation **16**

C.1 Moment identities, computation, and special cases 17

C.2 Population recovery 18

D Validation of invariant signal representations **20**

E Proportional risk: calculation and proofs **23**

E.1 Independent covariance representation 23

E.2 Risk with a noise-only subspace 24

E.3 Explicit scalar formulas for the spanning submodel 29

E.4 Elementary checks with additional noise directions 33

E.5 Strong signals and the contrast-only endpoint 35

F Experimental protocols and supporting evidence **37**

F.1 Common protocols and numerical implementation 37

F.2 Proportional risk and source-wise CV 37

F.3 Independent holdout coverage study 41

F.4 Linear real-image protocols 41

A Additional related work

Augmentation and invariant subspaces. Our paired-source model describes augmentation-specific and common features (Xiao et al., 2021; Tahmasebi et al., 2026; Gresele et al., 2020). Stationary-subspace and invariant-component methods instead identify invariance from changing marginal moments (Meinecke et al., 2009; von Bünau et al., 2009; Norman & Meir, 2026; Flumian et al., 2024; Wang et al., 2019). Here the same-source pairing supplies the difference covariance.

Structured augmentation can also identify content and individual style variables through multiple embedding spaces with different invariances (Eastwood et al., 2023). Our linear noisy model instead quantifies the rank budget needed for complementary channels to expose background directions, then relates their empirical suppression to high-dimensional subspace risk.

Invariant prediction, marginal invariance, and pairing. IRM asks for a representation on which one predictor minimizes risk in every training environment (Arjovsky et al., 2019). PICA instead seeks high-variance representations with equal distributions across environments (Norman & Meir, 2026). For zero-mean Gaussian environments, linear marginal invariance amounts to equality of projected covariances. The equal-marginal construction in (25) satisfies this constraint in every direction and supplies no covariance variation to stationary subspace analysis (Meinecke et al., 2009; von Bünau et al., 2009), whereas paired differences still reveal the changed background. Our identification therefore uses additional paired-source information; it is not a claim of greater identifiability from the same unpaired data. An invariant representation yields a common optimal predictor only with further assumptions on the label mechanism and signal/noise laws.

Identification and projection from augmented pairs. von Kügelgen et al. (2021) identify unchanged latent factors up to an invertible transformation by alignment and entropy maximization, assuming sufficient local augmentation variation in every nuisance coordinate. Noisy counterfactual matching estimates nuisance directions from the leading singular vectors of paired differences and trains predictors orthogonal to them (Bai et al., 2025); GRIT applies this projection to frozen pretrained features (Bai et al., 2026). Our reusable channels may each change only a background subspace. The channel budget characterizes their joint coverage, while the power path addresses finite-sample use of these changes without always projecting fully onto their estimated complement.

Multiview learning and contrastive PCA. The common ambient coordinates and labeled pairs distinguish our sampling design from general multiview correlation and incomplete-view learning (Kettenring, 1971; Sørensen et al., 2021; Han et al., 2026; van de Velden & Takane, 2012; Amini et al., 2009; Wilkinson, 1957; Bailey & Cameron, 2013). Paired records also differ from separate foreground and background samples in contrastive PCA (Abid et al., 2018).

High-dimensional spectral theory. Spiked covariance and Fisher-matrix limits describe the emergence and alignment of sample eigenvectors (Paul, 2007; Bai & Yao, 2012; Wang & Yao, 2017). Anisotropic local laws control resolvents in specified directions (Knowles & Yin, 2017). We use these tools to derive risk along the full fixed-power path, including noninteger powers; normalized alignment and difference PCA supply explicit checks (Appendix E).

B Model, moments, and coverage

We use the model and assumptions of Section 2, with orthonormal bases U_A, U_B for the signal and background spaces. The coverage identities below require only the stated second moments and do not require orthogonal signal and background spaces. The equal-law construction uses a Gaussian subclass with orthogonal signal and background spaces spanning $\mathbb{R}^d$.

Remark B.1 (Examples of augmentation channels). For independent background resampling, take $p = bm$, stack independent standardized m-vectors $h_{i1}, \dots, h_{ib}$ in h_i, and let $C_e = J_e$ select block e. Then $x_i^{(e)} = Az_i + Bh_{ie} + \varepsilon_i^{(e)}$. Taking $b = 2$, $\sigma = \chi = 1$, $A^\top B = 0$, and independent Gaussian factors and noises recovers the contrastive factor model of Wu et al. (2025). Channels may also retain a common background component, so cross-view agreement can reflect background as well as signal. Section 5.1 studies this partially shared case.

B.1 COVERAGE AND THE CHANNEL BUDGET

Proof of Proposition 3.1. The signal and shared observation noise cancel from a channel difference, so

$$\tfrac{1}{2}\operatorname{Cov}(x_i^{(e)} - x_i^{(f)}) = \sigma^2\chi^2 I_d + \tfrac{1}{2}B(C_e - C_f)(C_e - C_f)^{\mathrm{T}}B^{\mathrm{T}}.$$

With $\overline{C} = b^{-1}\sum_e C_e$, the pair-scatter identity gives

$$H = \frac{1}{b(b-1)}\sum_{e<f}(C_e - C_f)(C_e - C_f)^{\mathrm{T}} = \frac{1}{b-1}\sum_e (C_e - \overline{C})(C_e - \overline{C})^{\mathrm{T}}.$$

Averaging the first identity over all pairs gives $S = \sigma^2\chi^2 I_d + BHB^{\mathrm{T}}$, as in (3).

Fix channel 1 as a reference, and write $\Delta_e = C_e - C_1$ for $e > 1$ and $\mathcal{T} = [\Delta_2 \cdots \Delta_b]$. For $u \in \mathbb{R}^m$,

$$u^{\mathrm{T}}Hu = \frac{1}{b(b-1)}\sum_{e<f}\|(C_e - C_f)^{\mathrm{T}}u\|^2.$$

This vanishes exactly when every $\Delta_e^{\mathrm{T}}u = 0$, because $C_e - C_f = \Delta_e - \Delta_f$, with $\Delta_1 = 0$. Taking orthogonal complements gives

$$\operatorname{range}(H) = \operatorname{range}(\mathcal{T}), \qquad \ker(H) = \bigcap_{e=2}^{b}\ker(\Delta_e^{\mathrm{T}}). \tag{19}$$

If $H \succ 0$, then for $v \in \mathbb{R}^d$,

$$v^{\mathrm{T}}(S - \sigma^2\chi^2 I_d)v = \|H^{1/2}B^{\mathrm{T}}v\|^2.$$

Consequently,

$$\begin{aligned}\ker(S - \sigma^2\chi^2 I_d) &= \ker(B^{\mathrm{T}}) = \operatorname{col}(B)^{\perp},\\ \operatorname{range}(S - \sigma^2\chi^2 I_d) &= \operatorname{col}(B).\end{aligned} \tag{20}$$

Full column rank of B makes the excess covariance on its column space positive definite, with smallest eigenvalue $h = \lambda_{\min}(U_B^{\mathrm{T}}BHB^{\mathrm{T}}U_B) > 0$. Let $Q = I_d - U_BU_B^{\mathrm{T}}$. Then $Qx_i^{(e)} = QAz_i + Q\varepsilon_i^{(e)}$, so projection onto this kernel removes all background. If additionally $A^{\mathrm{T}}B = 0$, then $QA = A$ and $QDQ = AA^{\mathrm{T}} + \sigma^2 Q$. Its leading k eigenvectors within $\operatorname{range}(Q)$ therefore span $\operatorname{col}(A)$: their eigenvalues exceed σ^2, whereas every remaining direction has variance σ^2. When $d = k + m$, the kernel itself equals $\operatorname{col}(A)$, recovering the bottom-k characterization.

Let r be an integer with $1 \le r \le \min(m, p)$, and suppose $\operatorname{rank}(\Delta_e) \le r$ for $e > 1$. By (19), $\operatorname{rank}(H) = \operatorname{rank}(\mathcal{T}) \le \sum_{e=2}^{b}\operatorname{rank}(\Delta_e) \le (b-1)r$. Complete coverage thus requires

$$(b-1)r \ge m, \qquad b \ge 1 + \lceil m/r \rceil. \tag{21}$$

For sufficiency, fix any $p \ge r$ and assume (21). Partition $\mathbb{R}^m$ into $b-1$ coordinate blocks of dimensions $s_e \le r$, allowing empty blocks. Let $U_e \in \mathbb{R}^{m\times s_e}$ contain the coordinate vectors of block e, and let $F_e \in \mathbb{R}^{p\times s_e}$ contain the first s_e coordinate vectors of $\mathbb{R}^p$. Set $C_1 = 0$ and $C_e = U_eF_e^{\mathrm{T}}$, taking $C_e = 0$ on empty blocks. Every reference difference $C_e - C_1$ has rank $s_e \le r$, and the reference-pair terms give

$$H \succeq \frac{1}{b(b-1)}\sum_{e=2}^{b}C_eC_e^{\mathrm{T}} = \frac{1}{b(b-1)}\sum_{e=2}^{b}U_eU_e^{\mathrm{T}} = \frac{1}{b(b-1)}I_m \succ 0.$$

The construction works at every allowed fixed p; a common matrix may also be added to all actions without changing coverage. □

The budget constrains $\operatorname{rank}(\Delta_e) = \operatorname{rank}(C_e - C_1) \le r$. Adding the same matrix $C_0 \in \mathbb{R}^{m\times p}$ to every action leaves $(C_e + C_0) - (C_1 + C_0) = \Delta_e$ and hence H unchanged, even though individual action ranks may change. For the example $m = 6, r = 2$ in Section 3.1, take $p = 2, b = 4, C_1 = 0$, with complementary changes

$$[\Delta_2\ \Delta_3\ \Delta_4] = \begin{bmatrix} I_2 & 0 & 0\\ 0 & I_2 & 0\\ 0 & 0 & I_2\end{bmatrix} = I_6.$$

Each change has rank two, while (19) gives $\operatorname{rank}(H) = 6$, so these four channels attain complete coverage at $b = 1 + m/r$. If all nonreference channels instead repeat the same block, $C_1 = 0$ and $C_e = U := [I_2\ 0\ 0]^{\mathrm{T}}$ for $e > 1$, then

$$H = \frac{1}{b(b-1)} \sum_{e<f} (C_e - C_f)(C_e - C_f)^{\mathrm{T}} = \frac{1}{b} UU^{\mathrm{T}}, \qquad \operatorname{rank}(H) = 2 < 6,$$

regardless of the number of repeated channels. For independent resampling, the block selectors in Remark B.1 satisfy, for every $e \neq f$,

$$\frac{1}{2}(J_e - J_f)(J_e - J_f)^{\mathrm{T}} = I_m \succ 0.$$

Thus each pair alone covers all m background directions; $\operatorname{rank}(J_e - J_f) = m$ corresponds to $r = m$, for which the budget requires only $b \geq 2$.

Incomplete pair observations. Suppose each source supplies only a named pair (e, f), assigned independently of its factors and noise. Let $\mathcal{E}_\pi$ be the represented pair types, with positive limiting frequencies π_{ef} summing to one; deterministic schedules are allowed. Then

$$S_\pi = \sigma^2 \chi^2 I_d + B H_\pi B^{\mathrm{T}}, \qquad H_\pi = \tfrac{1}{2} \sum_{(e,f) \in \mathcal{E}_\pi} \pi_{ef} (C_e - C_f)(C_e - C_f)^{\mathrm{T}}. \tag{22}$$

The empirical moments are the averages of $\frac{1}{2}(x_i^{(e)} - x_i^{(f)})(x_i^{(e)} - x_i^{(f)})^{\mathrm{T}}$ and $\frac{1}{2}(x_i^{(e)} x_i^{(e)\mathrm{T}} + x_i^{(f)} x_i^{(f)\mathrm{T}})$ over the recorded sources. Conditional on the pair labels, they target the corresponding realized-frequency mixtures. Uniform frequencies over all pairs recover $S_\pi = S$.

Corollary B.1 (Coverage from represented pairs)**.** The range of H_π is the span of the columns of the represented $C_e - C_f$'s. If $H_\pi \succ 0$, then $\ker(S_\pi - \sigma^2 \chi^2 I_d) = \operatorname{col}(B)^\perp$, and the excess covariance on $\operatorname{col}(B)$ has smallest eigenvalue $\lambda_{\min}(U_B^{\mathrm{T}} B H_\pi B^{\mathrm{T}} U_B) > 0$. Under $A^{\mathrm{T}} B = 0$ and $d = k + m$, this kernel equals $\operatorname{col}(A)$. If every reference pair $(1, e)$, $e > 1$, has positive frequency, then $\operatorname{range}(H_\pi) = \operatorname{range}(H)$.

Proof of Corollary B.1. The identity $u^{\mathrm{T}} H_\pi u = \frac{1}{2} \sum_{(e,f)} \pi_{ef} \|(C_e - C_f)^{\mathrm{T}} u\|^2$ identifies the kernel because all represented weights are positive. Applying the preceding quadratic-form argument with H_π in place of H gives the background kernel and positive excess eigenvalues. Orthogonality and $d = k + m$ identify this kernel with $\operatorname{col}(A)$. Reference differences span all pair differences. □

This extends background identification to two views per source, with signal recovery under the stated geometric conditions, using the same complementary channel actions as in Figure 1. Small frequencies can weaken its gap even when coverage is complete. The implemented path and selection theory use complete tuples, with independent sources as sampling units.

B.2 Equal marginals and the information in a portfolio

We construct a Gaussian subclass in which all view marginals agree, every fixed pair leaves an ambiguity, and complementary channels identify the target. For this subclass, set $d = k + m$. Fix $k \geq 1$, $m > 2r$, $(b-1)r \geq m$, amplitudes $g_0 \geq 0, q_0 > 0$, and $s_0^2 = g_0^2 + q_0^2$. Partition the background coordinates into blocks of size at most r, with orthogonal projectors $P_2, \ldots, P_b$ summing to I_m, and set

$$\begin{aligned} &T_1 = I_m, \quad T_e = I_m - 2P_e \ (e > 1), \\ &C_e = s_0^{-1} [g_0 I_m \ \ q_0 T_e]. \end{aligned} \tag{23}$$

Each action flips at most one block relative to the reference and any pair changes at most two, so $\operatorname{rank}(C_e - C_f) \leq 2r < m$. For any orthogonal matrix $[U_A\ U_B] \in \mathbb{R}^{d \times d}$, take

$$A = s_0 U_A, \qquad B = s_0 U_B. \tag{24}$$

Let Θ_{sign} comprise these models with fixed actions, amplitudes, $\sigma^2 > 0$, and $\chi \in [0,1]$, and independent standard Gaussian generating components. Writing $h_i = (u_i^{\mathrm{T}}, v_i^{\mathrm{T}})^{\mathrm{T}}$, the background is $U_B(g_0 u_i + q_0 T_e v_i)$. Since every T_e is orthogonal,

$$\begin{aligned} C_e C_e^{\mathrm{T}} = I_m, \qquad B C_e C_e^{\mathrm{T}} B^{\mathrm{T}} = s_0^2 U_B U_B^{\mathrm{T}}, \\ \mathrm{Cov}(x_i^{(e)}) = (s_0^2 + \sigma^2) I_d. \end{aligned} \tag{25}$$

Thus neither the channel nor the loading basis changes a view marginal.

Theorem B.1 (Fixed-pair ambiguity and portfolio identification)**.** For every named pair (e, f), there are $\theta_{ef}^{(0)}, \theta_{ef}^{(1)} \in \Theta_{\text{sign}}$ whose full pair laws agree for every sample size, while their target projectors have loss one. Consequently,

$$\inf_{\widehat{P}_{ef}} \sup_{\theta \in \Theta_{\text{sign}}} \mathbb{E}_{\theta,ef}^n \ell\{\widehat{P}_{ef}, P_A(\theta)\} \geq \tfrac{1}{4}, \tag{26}$$

where the infimum is over rank-k orthogonal-projector estimators using the entire raw sample from this pair. At every model in the class, the complete portfolio and every covering reference-pair schedule have positive-definite coverage and identify P_A.

Proof of Theorem B.1. Fix a pair. Since its changed blocks have dimension at most $2r < m$, some background coordinate j satisfies $T_e e_j = T_f e_j = \nu e_j$, with $\nu \in \{-1, 1\}$. Choose orthogonal ambient unit vectors a, c. At the first parameter point use $s_0 a$ as one signal loading and $s_0 c$ as background loading j; at the second interchange these columns. All other loadings, action matrices, and generating laws stay fixed, so both points belong to Θ_{sign}.

In the signal coordinate both views contain the scalar $s_0 z_1$; in background coordinate j both contain $g_0 u_j + \nu q_0 v_j$. These scalars are independent Gaussian variables of variance s_0^2. Including observation noise, either coordinate has the two-view covariance block

$$K = \begin{bmatrix} s_0^2 + \sigma^2 & s_0^2 + \sigma^2(1-\chi^2) \\ s_0^2 + \sigma^2(1-\chi^2) & s_0^2 + \sigma^2 \end{bmatrix}. \tag{27}$$

The joint covariance on $\mathrm{span}(a, c)$ is therefore $\mathrm{diag}(K, K)$, unchanged by the swap. Remaining coordinates are independent of this block and have identical laws at the two points. Hence the full Gaussian pair laws agree for one source and for every n. The target projectors P_0, P_1 replace a by c, giving $\ell(P_0, P_1) = 1$. For any rank-k estimate P, the parallelogram inequality gives

$$\ell(P, P_0) + \ell(P, P_1) \geq \tfrac{1}{2}\ell(P_0, P_1) = \tfrac{1}{2}.$$

Integrating under their common pair law makes at least one risk $1/4$ or larger, proving (26). The alternative points may depend on the pair; the parameter class is fixed.

For the full portfolio the reference-pair contributions satisfy

$$H \succeq \frac{4q_0^2}{b(b-1)s_0^2} \sum_{e=2}^{b} P_e = \frac{4q_0^2}{b(b-1)s_0^2} I_m \succ 0.$$

Because this subclass has orthogonal signal and background spaces with $d = k + m$, Proposition 3.1 identifies P_A. Corollary B.1 gives the same conclusion for positive frequencies on the covering reference pairs. □

C Population recovery and computation

The same spectral construction computes the sample fits and explains the population criterion. We first establish the moment identities and special cases in Section 3.2, then prove the recovery conditions in Section 3.

C.1 Moment identities, computation, and special cases

Population and sample moment identities. At the population level, expanding the difference covariances gives

$$M_+ := \frac{2}{b(b-1)} \sum_{e<f} \operatorname{sym} \operatorname{Cov}(x_i^{(e)}, x_i^{(f)}) = D - S,$$

because each marginal covariance occurs $b-1$ times in the pair sum. The sample moments satisfy the same identities. The all-pairs expression for the sample contrastive covariance is

$$\widehat{S} = \frac{1}{nb(b-1)} \sum_i \sum_{e<f} (x_i^{(e)} - x_i^{(f)})(x_i^{(e)} - x_i^{(f)})^{\mathrm{T}}. \tag{28}$$

The pair-scatter identity makes this equal to the within-source formula in (9), which avoids forming all view pairs. The average cross-view moment is

$$\widehat{M}_+ = \frac{1}{nb(b-1)} \sum_i \sum_{e \neq f} x_i^{(e)} x_i^{(f)\mathrm{T}} = \widehat{D} - \widehat{S}. \tag{29}$$

Writing $\widehat{M}_{\mathrm{av}} = n^{-1} \sum_i \overline{x}_i \overline{x}_i^{\mathrm{T}}$ for the covariance estimate of the source averages using their known zero mean gives $\widehat{D} = \widehat{M}_{\mathrm{av}} + (b-1)\widehat{S}/b$.

A spectral computation. For positive-definite matrices $D, S \in \mathbb{R}^{d \times d}$ and $1 \leq k < d$, write

$$V_\alpha \in \arg\min_{V^{\mathrm{T}} D V = I_k} \operatorname{tr}(V^{\mathrm{T}} S^\alpha V), \qquad P_\alpha = V_\alpha (V_\alpha^{\mathrm{T}} V_\alpha)^{-1} V_\alpha^{\mathrm{T}}, \quad \alpha \geq 0. \tag{30}$$

The columns of $D^{1/2} V_\alpha$ span the bottom-k eigenspace of $D^{-1/2} S^\alpha D^{-1/2}$, with $S^0 = I_d$. To avoid large positive powers, set $s_* = \lambda_{\min}(S)$ and define

$$T_\alpha = (s_* S^{-1})^{\alpha/2}, \qquad C_\alpha = T_\alpha D T_\alpha.$$

If W_α contains orthonormal top-k eigenvectors of C_α with eigenvalue matrix Λ_α, then

$$V_\alpha = T_\alpha W_\alpha \Lambda_\alpha^{-1/2}, \qquad V_\alpha^{\mathrm{T}} D V_\alpha = I_k, \qquad S^\alpha V_\alpha = D V_\alpha s_*^\alpha \Lambda_\alpha^{-1}.$$

Thus the largest eigenvalues of C_α correspond to the smallest generalized eigenvalues of (S^α, D); the largest generalized eigenvalues of (D, S^α) are $\Lambda_\alpha / s_*^\alpha$. This also constructs the fit by filtering the views, applying PCA, and mapping its loading vectors back. A QR decomposition of $T_\alpha W_\alpha$ gives the Euclidean projector in (10). All powers reuse the eigensystem of S. The unregularized procedure requires positive-definite fitting moments; support failures are recorded without adding a ridge. Uniqueness and differentiability of the selected projector require a gap between its selected and unselected eigenvalues.

Derivation of the reference criteria. Let $D, S \succ 0$, and let $s_1 \leq \cdots \leq s_d$ be the eigenvalues of S. At $\alpha = 0$, the whitened matrix is D^{-1}, so the minimizing subspaces in (30) are the leading k-dimensional eigenspaces of D. At $\alpha = 1$, $\operatorname{tr}(V^{\mathrm{T}} D V) = k$ makes the objective equivalent to

$$\max_{V^{\mathrm{T}} D V = I_k} \operatorname{tr}\{V^{\mathrm{T}} (D - S) V\}. \tag{31}$$

If $s_k < s_{k+1}$, let $P_{S,k}$ project onto the bottom-k eigenspace of S, and write $\kappa(D) = \lambda_{\max}(D)/\lambda_{\min}(D)$. Every minimizing subspace satisfies

$$\ell(P_\alpha, P_{S,k}) \leq k \kappa(D) \left(\frac{s_k}{s_{k+1}} \right)^\alpha. \tag{32}$$

In particular, for fixed D, S, $P_\alpha \to P_{S,k}$ as $\alpha \to \infty$.

To prove the bound, let E be an orthonormal basis of the bottom-k space of S. The feasible competitor $E(E^{\mathrm{T}} D E)^{-1/2}$ has objective at most $k s_k^\alpha / \lambda_{\min}(D)$. Thus

$$s_{k+1}^\alpha \| (I_d - E E^{\mathrm{T}}) V_\alpha \|_F^2 \leq \operatorname{tr}(V_\alpha^{\mathrm{T}} S^\alpha V_\alpha) \leq \frac{k s_k^\alpha}{\lambda_{\min}(D)}.$$

Because $V_\alpha^\top V_\alpha \succeq \lambda_{\max}(D)^{-1} I_k$, orthonormalizing V_α gives (32).

If instead the smallest eigenvalue of S has multiplicity $r \geq k$, let Q_0 project onto its eigenspace. Then $T_\alpha \to Q_0$ and $C_\alpha \to Q_0 D Q_0$. When $r = k$, or the restriction of $Q_0 D Q_0$ to $\mathrm{range}(Q_0)$ has a gap after its kth largest eigenvalue, the limiting selected space is its leading k-dimensional eigenspace. The back-transformation $T_\alpha W_\alpha$ has the same limiting space because Q_0 acts as the identity there. Thus pooled variance orders directions tied at the smallest contrast variance; Proposition 3.2 gives the corresponding conclusion under the generating model.

The limit takes $\alpha \to \infty$ at fixed D, S, before any sample-size or dimension limit.

Connection with two-view spectral learning. With two views, AIPCA's normalized-alignment candidate keeps the cross-view objective of PCA++ (Wu et al., 2025), using the average of the two sample covariances in the uniformity constraint. The original PCA++ uses one view's sample covariance; the formulations coincide when those covariances agree. The symmetric candidate has the same generalized eigenproblem as linear slow feature analysis on the view pairs when both use the same fitting moments and centering convention (Wiskott & Sejnowski, 2002). These correspondences concern the untruncated generalized eigenproblem; view-specific normalization and spectral truncation in the original PCA++ remain distinct choices. Covariance powers also interpolate classical multivariate methods through partial whitening, as in continuum power CCA (Swenson, 2015). Here the power acts on within-source differences under a fixed pooled-covariance constraint.

C.2 Population recovery

Proof of Proposition 3.2. Write $q = \sigma^2 \chi^2 > 0$ and $Q = I_d - P_B$. Complete coverage gives $S = qI_d + BHB^\top$, whose minimum-eigenvalue space is $\mathrm{range}(Q)$. The disjointness condition implies $\mathrm{rank}(QA) = k$: if $QAz = 0$, then $Az \in \mathrm{col}(A) \cap \mathrm{col}(B) = \{0\}$ and hence $z = 0$. Expanding model (1) gives

$$QDQ = QAA^\top Q + \sigma^2 Q.$$

Its leading k-dimensional eigenspace is $\mathrm{col}(QA)$, separated from the remaining eigenvalues by a positive gap. In the spectral construction above,

$$T_\alpha = (qS^{-1})^{\alpha/2} \longrightarrow Q, \qquad C_\alpha = T_\alpha D T_\alpha \longrightarrow QDQ.$$

The spectral gap therefore makes the leading k-dimensional eigenspace of C_α converge to $\mathrm{col}(QA)$. Since T_α is the identity on $\mathrm{range}(Q)$, mapping back by $\mathrm{col}(V_\alpha^\star) = \mathrm{col}(T_\alpha W_\alpha)$ preserves this limit, proving $P_\alpha \to P_\infty$.

Any basis V_∞ of this limiting space can be written $V_\infty = QAR$ with R invertible. Consequently,

$$V_\infty^\top B = 0, \qquad V_\infty^\top A = R^\top A^\top Q A,$$

and the second matrix is invertible because $A^\top Q A \succ 0$. Substituting into model (1) proves the stated representation identity. If $A^\top B = 0$, then $QA = A$; Theorem C.1 further gives exact recovery at every sufficiently large finite power. □

The following block analysis assumes $A^\top B = 0$ and allows arbitrary ambient dimension $d \geq k + m$. Under model (1), D and $S = \sigma^2 \chi^2 I_d + G$ are block diagonal between signal and background. When $d > k + m$, their additional block on $(\mathrm{col}(A) \oplus \mathrm{col}(B))^\perp$ contains only observation noise, with $D = \sigma^2 I$ and $S = \sigma^2 \chi^2 I$. Let U_A, U_B be orthonormal bases of the signal and background spaces. Define

$$D_A = U_A^\top D U_A, \quad D_B = U_B^\top D U_B, \quad G_B = U_B^\top G U_B, \quad d_A = \lambda_{\min}(D_A), \quad d_B = \lambda_{\max}(D_B).$$

The matrices D_B and G_B need not commute. Full column rank of A gives $d_A > \sigma^2$.

Theorem C.1 (Population recovery along the power path)**.** Assume model (1), $A^{\mathrm{T}}B = 0$, $d \geq k + m$, full-column-rank A, B, and $\sigma^2\chi^2 > 0$. Under complete coverage, write $h = \lambda_{\min}(G_B) > 0$. Every exponent satisfying

$$\left(1 + \frac{h}{\sigma^2\chi^2}\right)^{\alpha} > \frac{d_B}{d_A} \tag{33}$$

has the unique minimizing subspace $\mathrm{col}(A)$.
For arbitrary coverage, let Π_0 be the orthogonal projector onto $\ker(G_B)$ and set $d_0 = \|\Pi_0 D_B \Pi_0\|_{\mathrm{op}}$, with $d_0 = 0$ when this kernel is trivial. There exists a finite $\alpha \geq 0$ for which the signal space is the unique minimizing subspace if and only if

$$d_A > d_0. \tag{34}$$

When this inequality holds, every sufficiently large exponent recovers the signal. If $G = 0$, every exponent instead gives the same pooled-PCA subspace, which recovers the signal precisely under (34).

Proof of Theorem C.1. The population matrices split into signal, background, and, when $d > k + m$, pure-noise blocks. The signal generalized eigenvalues of (S^α, D) are at most $(\sigma^2\chi^2)^\alpha/d_A$. The pure-noise eigenvalues are $(\sigma^2\chi^2)^\alpha/\sigma^2$, strictly larger because $d_A > \sigma^2$. Under complete coverage the background generalized Rayleigh quotient is at least $(\sigma^2\chi^2 + h)^\alpha/d_B$. The strict inequality (33) therefore places all k signal eigenvalues below the background eigenvalues.

For arbitrary coverage use the reciprocal eigenproblem, multiplied by $(\sigma^2\chi^2)^\alpha$. Its signal block is D_A, and its background block is

$$M_\alpha = Q_\alpha D_B Q_\alpha, \qquad Q_\alpha = (I_m + \frac{G_B}{\sigma^2\chi^2})^{-\alpha/2}.$$

The pure-noise block, when present, is $\sigma^2 I$. Since $d_A > \sigma^2$, unique signal recovery is again equivalent to $\lambda_{\max}(M_\alpha) < d_A$. For every unit vector $u \in \ker(G_B)$, $Q_\alpha u = u$, so $\lambda_{\max}(M_\alpha) \geq d_0$; this also holds when $d_0 = 0$. Thus (34) is necessary. Conversely, $Q_\alpha \to \Pi_0$ and hence $M_\alpha \to \Pi_0 D_B \Pi_0$. Its largest eigenvalue is eventually below d_A whenever (34) holds. If $G = 0$, $S = \sigma^2\chi^2 I_d$ and the objective is a positive multiple of its $\alpha = 0$ version for every exponent, proving the last assertion. □

Complete coverage makes a finite exponent sufficient, although a fixed grid need not contain one. Under partial coverage, d_0 is the largest pooled variance among unchanged background directions, so (34) requires the weakest signal to outrank them. This is the extra variance information used beyond differences.

The large-power population limit has a corresponding variance interpretation that does not require orthogonality. Assume $D \succ 0$, $\sigma^2\chi^2 > 0$, and $\dim\ker G \geq k$, and let Q be the orthogonal projector onto $\ker G$. If $\dim\ker G = k$, or the restriction of QDQ to $\ker G$ has a gap after its kth largest eigenvalue, the limiting loading space solves

$$\max_{V^{\mathrm{T}}V = I_k,\, GV = 0} \mathrm{tr}(V^{\mathrm{T}} D V). \tag{35}$$

Indeed, at population $s_* = \sigma^2\chi^2$, so $T_\alpha = (I_d + \frac{G}{\sigma^2\chi^2})^{-\alpha/2} \to Q$ and $C_\alpha \to QDQ$ in the filtering construction. Thus pooled variance orders the invariant directions when more than k remain.

For the rank-one exchangeable model (13), under its orthogonality assumption, the pooled signal and background variances are $a+\sigma^2$ and $\kappa+\tau+\sigma^2$, while their contrast variances are $\sigma^2\chi^2$ and $\tau+\sigma^2\chi^2$. Their generalized eigenvalue ratio for (D, S^α) gives (14); the exact strict recovery condition is

$$\left(1 + \frac{\tau}{\sigma^2\chi^2}\right)^{\alpha} > \frac{\kappa + \tau + \sigma^2}{a + \sigma^2}. \tag{36}$$

At zero this becomes $a > \kappa + \tau$; at one it becomes $a(\sigma^2\chi^2 + \tau) + \sigma^2(1 - \chi^2)\tau > \sigma^2\chi^2\kappa$. These are population conditions; proportional-dimensional sample thresholds are derived separately in Section 5.

When $\sigma^2\chi^2 = 0$, empirical differences lie in $\mathrm{col}(B)$; if their span has dimension m, their orthogonal complement is $\mathrm{col}(B)^\perp$. Under $A^\top B = 0$ and $d = k + m$, difference PCA therefore recovers P_A exactly. Under the same orthogonal spanning assumption, population complete coverage makes every $\alpha > 0$ have zero objective only on $\mathrm{col}(A)$. More generally, if $D \succ 0$ and $\dim\ker S > k$, every k-subspace of that kernel minimizes it. These singular cases do not use the inverse-power computation above.

D VALIDATION OF INVARIANT SIGNAL REPRESENTATIONS

Reconstruction reference. The reconstruction minimizer R in Section 4.2 anchors the relative comparisons. We first record its covariance form and population target.

For the held-out views in Section 4.2, use the training mean when centering is needed and define

$$\widehat{D}_{\mathrm{val}} = \frac{1}{n_{\mathrm{val}} b} \sum_{i,e} x_i^{(e)} x_i^{(e)\top},$$
$$\widehat{C} = \widehat{D}_{\mathrm{val}}\widehat{Q} + \widehat{Q}\widehat{D}_{\mathrm{val}} - \widehat{D}_{\mathrm{val}}.$$

The score in (11) can then be evaluated as

$$\widehat{L}_{\mathrm{inv}}(\alpha) = \mathrm{tr}(\widehat{Q}\widehat{D}_{\mathrm{val}}\widehat{Q}) - \mathrm{tr}(\widehat{P}_\alpha \widehat{C}).$$

The first term is common to all candidates, so minimizing the score is equivalent to maximizing $\mathrm{tr}(\widehat{P}_\alpha \widehat{C})$.

Let $Q = I_d - P_B$ and let P_∞ be the projector in Proposition 3.2. All projectors below are Euclidean orthogonal projectors. For a rank-k projector P, define

$$L_Q(P) = \frac{1}{b}\sum_{e=1}^{b} \mathbb{E}\|Qx_i^{(e)} - Px_i^{(e)}\|^2, \qquad C_Q = DQ + QD - D.$$

Thus Px reconstructs the background-free target Qx from the candidate's original-input features, using its orthogonal decoder.

Proposition D.1 (Population target of reconstruction validation)**.** Under the conditions of Proposition 3.2, P_∞ uniquely minimizes L_Q among rank-k projectors.

Proof. Using $P^2 = P$, $Q^2 = Q$, and cyclicity of trace gives

$$\begin{aligned} L_Q(P) &= \mathrm{tr}\{(Q-P)D(Q-P)\} = \mathrm{tr}(QDQ) - \mathrm{tr}(PC_Q) \\ &= \mathrm{tr}(QDQ) - \mathrm{tr}(PQDQ) + \mathrm{tr}(PP_BDP_B), \qquad C_Q = QDQ - P_BDP_B. \end{aligned} \tag{37}$$

The last two terms respectively reward retaining variation outside the background space and penalize retaining background variation. Model (1) gives $QDQ = QAA^\top Q + \sigma^2 Q$. The two terms of C_Q act on the orthogonal spaces $\mathrm{range}(Q)$ and $\mathrm{col}(B)$. Since $\mathrm{col}(A) \cap \mathrm{col}(B) = \{0\}$, QA has rank k. On $\mathrm{range}(Q)$, the leading k eigenvectors therefore span $\mathrm{col}(QA)$, with eigenvalues strictly above σ^2; all remaining eigenvalues equal σ^2. On $\mathrm{col}(B)$, the eigenvalues of $-P_BDP_B$ are nonpositive. Thus C_Q has a unique leading k-dimensional eigenspace $\mathrm{col}(QA)$. Maximizing $\mathrm{tr}(PC_Q)$ over rank-k orthogonal projectors uniquely selects P_∞, proving the claim. □

Covariance profile. Write $P_\alpha = U_\alpha U_\alpha^\top$, with orthonormal U_α, for a training candidate, and set $q = d - k - m$. All moments use the same training-fitted centering. The sample identity

$$\widehat{T} = b\widehat{D} - (b-1)\widehat{S} = \frac{b}{n}\sum_i \overline{x}_i \overline{x}_i^\top$$

also defines $\widehat{T}_{\mathrm{val}}$. For each candidate, fit a training background projector $P_{B,\alpha}^{\mathrm{prof}}$ of rank m, orthogonal to P_α, and put $P_{N,\alpha}^{\mathrm{prof}} = I - P_\alpha - P_{B,\alpha}^{\mathrm{prof}}$. These candidate-specific projectors are used only for the profile. Suppressing their subscripts, its working covariance blocks are

$$T_{\boldsymbol{\eta}} = t_A P_\alpha + t_B P_B^{\mathrm{prof}} + t_0 P_N^{\mathrm{prof}}, \qquad S_{\boldsymbol{\eta}} = s_0(P_\alpha + P_N^{\mathrm{prof}}) + s_B P_B^{\mathrm{prof}}.$$

For $q > 0$, the five groups entering the profile are listed below. Each $h^{\mathrm{val}}_{\alpha r}$ is variance per spatial direction; w_r counts the coordinates contributed by a source. The shared contrast variance s_0 outside the fitted background expresses the working assumption that changing background lies in that fitted space.

Group r	$h^{\mathrm{val}}_{\alpha r}$	η_r	w_r
1: means, representation	$\mathrm{tr}(P_\alpha \widehat{T}_{\mathrm{val}})/k$	t_A	k
2: means, fitted background	$\mathrm{tr}(P_B^{\mathrm{prof}} \widehat{T}_{\mathrm{val}})/m$	t_B	m
3: means, residual space	$\mathrm{tr}(P_N^{\mathrm{prof}} \widehat{T}_{\mathrm{val}})/q$	t_0	q
4: contrasts, outside background	$\mathrm{tr}\{(I - P_B^{\mathrm{prof}})\widehat{S}_{\mathrm{val}}\}/(k+q)$	s_0	$(b-1)(k+q)$
5: contrasts, fitted background	$\mathrm{tr}(P_B^{\mathrm{prof}} \widehat{S}_{\mathrm{val}})/m$	s_B	$(b-1)m$

The variance vector is $\boldsymbol{\eta} = (t_A, t_B, t_0, s_0, s_B)$. All variances are positive, with the following necessary variance orders:

$$t_0 \le t_A, \quad t_0 \le t_B, \quad s_0 \le s_B, \quad s_0 \le t_0, \quad s_B \le t_B.$$

For $q = 0$, $P_B^{\mathrm{prof}} = I - P_\alpha$: remove the residual block and t_0, use weights $(k, m, (b-1)k, (b-1)m)$, and impose $s_0 \le s_B$, $s_0 \le t_A$, $s_B \le t_B$. These positivity and order constraints define $\mathcal{V}_q$. This is a relaxation containing the scalar exchangeable model of Section 5. For $q > 0$, that model also implies $t_B - t_0 \ge s_B - s_0$, since the difference equals $b\kappa$. The profile enforces only the displayed orders, leaving this additional coupled inequality unrestricted. The factor $b - 1$ counts orthogonal contrast channels. Indeed, an orthogonal change of coordinates across the views gives one channel $\sqrt{b}\,\overline{x}_i$ and $b - 1$ contrasts. In the specified Gaussian exchangeable model, these channels are independent, with covariances T and S, respectively. Their multiplicities determine the weights without an additional fitted penalty. The ordering constraints allow equality, whereas the variances themselves remain positive.

For a scalar Gaussian coordinate $z \sim N(0, \eta)$, $-2 \log f_\eta(z) = \log(2\pi) + \log \eta + z^2/\eta$. Summing over the w_r coordinates in each group and averaging over sources gives $\sum_r w_r\{\log \eta_r + h^{\mathrm{val}}_{\alpha r}/\eta_r\}$, up to a constant common to all candidates. Minimizing over the allowed variance levels with the training projectors fixed gives the score below; this variance minimization is the profiling step.

$$\widehat{F}_\alpha = \min_{\boldsymbol{\eta} \in \mathcal{V}_q} \sum_r w_r\{\log \eta_r + h^{\mathrm{val}}_{\alpha r}/\eta_r\}, \qquad q = d - k - m. \tag{38}$$

Here $h^{\mathrm{val}}_{\alpha r}$ is the held-out variance of block r, η_r its fitted variance, and w_r its dimension and contrast multiplicity.

When $q > 0$, initialize the background separately with the leading m eigenvectors of training $\widehat{S}$ and $\widehat{T}$, each compressed to $U_\alpha^\perp$. Alternate exact variance minimization with the leading eigenspace of

$$(t_0^{-1} - t_B^{-1})\widehat{T} + (b-1)(s_0^{-1} - s_B^{-1})\widehat{S} \quad \text{compressed to } U_\alpha^\perp.$$

Retain the smaller training objective, preferring the contrast start for numerical ties. Variance minimization enumerates the 52 equality partitions of five parameters (15 for four); each equality block uses its weighted empirical mean. Each block step decreases the objective up to roundoff. A nondegenerate fixed point is stationary, but two starts do not certify a global optimum. We allow 256 background updates and stop when the spectral trace optimality gap is at most 10^{-10} times the larger of total block weight and the trace scale. The latter is the maximum of one, the sum of absolute leading eigenvalues, and the current absolute trace, after removing the operator's scalar identity part. A fit reaching the iteration limit remains eligible and is flagged. Spectral ties use a deterministic ambient-coordinate basis; start-objective ties within $1024\epsilon_{\mathrm{mach}} \max(1, |F_S|, |F_T|)$ prefer the contrast start. Here F_S, F_T are the final objectives from the two starts, and ϵ_{mach} is machine precision.

Finally, hold both training spaces fixed and minimize over variances again using h^{val}_α in (38). This differs from inserting training variance estimates into a predictive loss. A finite positive optimum is required for every candidate: an unavailable profile is reported as a failed run, without a variance floor or candidate deletion. For independent Gaussian source-mean and contrast channels in this

relaxed family, the objective is twice the average negative Gaussian log likelihood up to common constants. For other views it is a working composite criterion; its ordering need not coincide with projector-risk ordering.

Joint paired comparisons. Keep the common $\widehat{Q}$ from reconstruction. For $q > 0$, let F_D contain the bottom q training eigenvectors of $\widehat{D}$, and set $N_D = F_D F_D^{\mathrm{T}}$. For each validation source define

$$S_i = \frac{1}{b-1}\sum_e (x_i^{(e)} - \overline{x}_i)(x_i^{(e)} - \overline{x}_i)^{\mathrm{T}}, \qquad c_i(\alpha) = \mathrm{tr}(P_\alpha S_i),$$

$$e_i(\alpha) = b\left\{\|U_\alpha^{\mathrm{T}}\widehat{Q}\overline{x}_i\|^2 - \frac{\mathrm{tr}(P_\alpha\widehat{Q})}{q}\|F_D^{\mathrm{T}}\overline{x}_i\|^2\right\} \qquad (q > 0).$$

Thus E compares residual mean capture with a common noise baseline; it never uses $P_{B,\alpha}^{\mathrm{prof}}$. E is absent when $q = 0$.

Index all unordered candidate pairs in the applicable C/E families by h, with source differences d_{ih}, means $\bar{d}_h$, and standard errors $\widehat{s}_h/\sqrt{n_{\mathrm{val}}}$, where $\widehat{s}_h^2$ is their sample variance. For positive-variance pairs simulate the joint Gaussian maximum

$$Z^* = \max_h \left|\sum_i \xi_i \frac{d_{ih} - \bar{d}_h}{\{\sum_r (d_{rh} - \bar{d}_h)^2\}^{1/2}}\right|, \qquad \xi_i \overset{\mathrm{iid}}{\sim} N(0,1).$$

Set $Z^* = 0$ if no positive-variance pairs remain. Use its upper empirical 0.95 quantile $\widehat{c}$. The experiments use 9,999 Gaussian draws to approximate this quantile; this prespecified Monte Carlo budget is a numerical choice, and 10,000 draws would also be valid. Numerical SVD compression removes only machine-precision null directions and records covariance error; Monte Carlo error remains. With directed differences defined as candidate minus reference, the bounds entering (12) are

$$U_{\alpha R}^C = \widehat{\Delta}_{\alpha R}^C + \widehat{c}\,\widehat{\mathrm{SE}}_{\alpha R}^C, \qquad L_{\alpha R}^E = \widehat{\Delta}_{\alpha R}^E - \widehat{c}\,\widehat{\mathrm{SE}}_{\alpha R}^E.$$

A nonreference candidate requiring a zero empirical-variance comparison is excluded; empirical constancy does not establish population constancy. For such unavailable comparisons, define $U_{\alpha R}^C = +\infty$ or $L_{\alpha R}^E = -\infty$, respectively, in (12). The reference is always retained. Including all pairs allows the reference itself to depend on validation data.

Risk meaning. The profile ranks candidates within the admissible set. The following risk comparison comes from the paired screening inequalities and is conditional on simultaneous coverage of their population differences. For independent holdout, let $\mathcal{T}$ denote the training data and define the population quantities

$$C_\alpha = \mathbb{E}[c_i(\alpha) \mid \mathcal{T}], \qquad E_\alpha = \mathbb{E}[e_i(\alpha) \mid \mathcal{T}] \quad (q > 0),$$

where i denotes a fresh source. These expectations are used for the risk comparison below; the implemented screening uses the empirical averages $\widehat{C}_\alpha = n_{\mathrm{val}}^{-1}\sum_i c_i(\alpha)$ and $\widehat{E}_\alpha = n_{\mathrm{val}}^{-1}\sum_i e_i(\alpha)$. Define the normalized projector loss $r(P; P_*) = 1 - \mathrm{tr}(PP_*)/k$, equal to $\ell(P, P_*)/k$.

Proposition D.2 (Relative comparison on a simultaneous event)**.** Suppose the conditional population mean of every positive empirical-variance pair difference is covered by its interval $\bar{d}_h \pm \widehat{c}\widehat{s}_h/\sqrt{n_{\mathrm{val}}}$. The selected index j satisfies $C_j \le C_R$, and $E_j \ge E_R$ when $q > 0$. If $q = 0$ and $S = vI + \tau P_B$, $\tau > 0$, then $r(P_j; I - P_B) \le r(P_R; I - P_B)$.
For $q > 0$, suppose the centered source-mean moment is $M = \theta I + F_M$, where $F_M = aP_A + \beta P_B$, $a > 0$, $\beta \ge 0$, and $P_A P_B = 0$. Put

$$\zeta = \mathrm{tr}(N_D F_M)/q, \qquad H = \widehat{Q}F_M\widehat{Q} - \zeta\widehat{Q} - aP_A, \qquad \nu_{jR} = \|P_j - P_R\|_*/k.$$

Then, with $\|\cdot\|_*$ denoting nuclear norm,

$$r(P_j; P_A) \le r(P_R; P_A) + \nu_{jR}\|H\|_{\mathrm{op}}/a. \tag{39}$$

Proof. For an accepted nonreference pair, simultaneous coverage and the two sign constraints give the claimed population comparisons; the reference case is equality. For $q = 0$, $C(P) = vk+\tau kr(P; I-P_B)$. For $q > 0$, the common baseline cancels θ, giving $E(P)/b = \mathrm{tr}\{P(aP_A+H)\}$. Hence $ak\{r(P_j; P_A) - r(P_R; P_A)\} \le \mathrm{tr}\{(P_j - P_R)H\} \le \|P_j - P_R\|_*\|H\|_{\mathrm{op}}$. □

Oracle background and noise projectors give $H = 0$. More generally, if $\epsilon_B = \|\widehat{Q} - (I - P_B)\|_{\mathrm{op}}$, then $\|H\|_{\mathrm{op}} \le 2a\epsilon_B + \beta\epsilon_B^2 + \zeta$. The signal term follows by expanding $\widehat{Q}P_A\widehat{Q} - P_A$; the background term follows from $\|\widehat{Q}P_B\widehat{Q}\|_{\mathrm{op}} \le \epsilon_B^2$. For nonorthogonal or anisotropic signals, E instead measures weighted capture of $QAA^{\mathrm{T}}Q$; the same unweighted loss comparison need not follow. Incomplete contrast coverage can likewise prevent the C comparison from identifying background loss. Training-mean error adds its residual mean outer product to M and must be included in H. These comparisons concern the reference's risk, not its distance to the best candidate. If the paired differences cannot be resolved, the reference can remain the only admissible choice. A favorable fixed-power risk theorem therefore does not ensure that validation will identify that power from the available sources.

In CV, use source-count weights w_f and fresh-source expectations conditional on each fold's own training set. With common population a, b, k across folds and simultaneous coverage of the weighted pair targets, the proposition gives

$$r(P_j^{\mathrm{full}}; P_A) \le r(P_R^{\mathrm{full}}; P_A) + \sum_f w_f \nu_{f,jR}\|H_f\|_{\mathrm{op}}/a + \eta_j^{\mathrm{fit}} + \eta_R^{\mathrm{fit}},$$

$$\eta_\alpha^{\mathrm{fit}} = \sum_f w_f \|P_\alpha^{\mathrm{full}} - P_{f,\alpha}\|_{\mathrm{op}}.$$

Indeed, sum the foldwise identity for E, apply the aggregate comparison, and use $|r(P_\alpha^{\mathrm{full}}; P_A) - \sum_f w_f r(P_{f,\alpha}; P_A)| \le \eta_\alpha^{\mathrm{fit}}$. For $q = 0$, use $I - P_B$ and omit the H term. The nuisance and refitting terms identify sources of error; their numerical size is not established here, so the bound is an error decomposition rather than a demonstrated finite-sample risk certificate.

Independent iid holdout admits conditional asymptotic calibration for a fixed grid under a joint central limit theorem and consistent variance estimation, with positive pair variances; uniform standardized $(2 + \delta)$-moment bounds for some $\delta > 0$ are sufficient conditions. The Gaussian multiplier construction is standard for simultaneous comparisons of source averages (Chernozhukov et al., 2013). Finite Gaussian integration adds approximation error. The implemented CV calibration centers the concatenated source differences globally; overlapping training folds do not inherit the independent-holdout coverage argument. Thus nominal 0.05 does not assert 95% CV coverage, and the proposition supplies neither proportional-dimensional oracle recovery nor classification noninferiority.

E Proportional risk: calculation and proofs

We prove the calculation in Section 5.1 and its consequences. Throughout this appendix, we work under Assumption 5.1. For brevity, write $\mathfrak{a}_\alpha = 1 - \mathcal{R}_\alpha$ for the limiting squared signal overlap in this appendix.

E.1 Independent covariance representation

Model (13) is a special case of model (1), obtained by taking $p = (b + 1)m$ and setting

$$h_i = (g_i^{\mathrm{T}}, h_{i1}^{\mathrm{T}}, \ldots, h_{ib}^{\mathrm{T}})^{\mathrm{T}}, \qquad A = \sqrt{a}\, U_A, \qquad B = U_B, \qquad C_e = [\sqrt{\kappa}\, I_m \;\; \sqrt{\tau}\, J_e],$$

where $J_e \in \mathbb{R}^{m\times bm}$ selects the e-th m-dimensional block of a bm-vector.

Write $P_B = U_B U_B^{\mathrm{T}}$, $P_\perp = I - P_A - P_B$, and $c_b = (c + c_\perp)/(b - 1)$. Let $N = n(b - 1)$, $M_{\mathrm{av}} = \mathbb{E}[\overline{x}_i \overline{x}_i^{\mathrm{T}}]$, and

$$\theta_0 = \sigma^2(1 - \chi^2) + \frac{\sigma^2\chi^2}{b}, \quad \theta_A = \theta_0 + a, \quad \theta_B = \theta_0 + \kappa + \frac{\tau}{b}, \quad s_B = \tau + \sigma^2\chi^2. \tag{40}$$

Source means have signal, background, and noise-only variances $\theta_A, \theta_B, \theta_0$, respectively. The corresponding contrast variances are $\sigma^2\chi^2, s_B, \sigma^2\chi^2$. Means retain persistent background; contrasts

remove it, but treat signal and noise-only directions alike. Pooled variance distinguishes the signal within this latter space. Under equation (13), the two population covariances are

$$M_{\mathrm{av}} = \theta_0 I + (\kappa + \tau/b) P_B + a P_A,$$
$$S = \sigma^2 \chi^2 I + \tau P_B.$$

Let $X \in \mathbb{R}^{d\times n}$ and $Z \in \mathbb{R}^{d\times N}$ have independent standard Gaussian entries and be independent of each other. The fitting moments admit the Gaussian covariance representation

$$\widehat{M}_{\mathrm{av}} = M_{\mathrm{av}}^{1/2} \frac{XX^{\mathrm{T}}}{n} M_{\mathrm{av}}^{1/2}, \qquad \widehat{S} = S^{1/2} \frac{ZZ^{\mathrm{T}}}{N} S^{1/2}, \tag{41}$$

To obtain this representation, apply an orthogonal transformation to the b view coordinates of each source, with first row $b^{-1/2}\mathbf{1}^{\mathrm{T}}$. Its first coordinate is $\sqrt{b}\,\overline{x}_i$; the other $b-1$ coordinates cancel every shared factor. Distinct transformed coordinates have zero covariance and hence are independent Gaussian vectors. Summing squares in these coordinates gives

$$\widehat{D} = \widehat{M}_{\mathrm{av}} + \frac{b-1}{b}\widehat{S}, \qquad \widehat{M}_{\mathrm{av}} = \frac{1}{n}\sum_i \overline{x}_i \overline{x}_i^{\mathrm{T}}. \tag{42}$$

Since $c_b < 1$ and $\sigma^2\chi^2 > 0$, the spectrum of $\widehat{S}$ lies in a fixed compact subset of $(0,\infty)$ with probability tending to one. All fixed powers and their derivatives are therefore bounded on this event.

E.2 Risk with a noise-only subspace

Write $Q_0 = I - P_B$ and $r = d - 1 - m$, so Q_0 contains the signal and the r noise-only directions. Locally abbreviate $v = \sigma^2\chi^2 > 0$ and set

$$\eta = c + c_\perp, \qquad y = \frac{\eta}{b-1} < 1, \qquad q = \frac{b-1}{b}, \qquad \theta_0 = \sigma^2(1-\chi^2) + \frac{v}{b}.$$

Here $q = (b-1)/b$ is a local covariance weight, distinct from the residual dimension q in Section 4.2. The source-mean covariance and contrast covariance are

$$M_{\mathrm{av}} = M_0 + aP_A, \qquad M_0 = \theta_0 Q_0 + \theta_B P_B, \qquad S = vQ_0 + s_B P_B. \tag{43}$$

Matrix definitions of the risk factors. The following matrix equations specify both detection and risk in Theorem 5.1. Write $\beta = (b-1)/b = q$ and use the signal-free source-mean covariance M_0 in (43). The identity (42) separates means from contrasts. For a candidate generalized eigenvalue λ of $(\widehat{D}, \widehat{S}^\alpha)$, define

$$\begin{aligned} T &= (\lambda\widehat{S}^\alpha - \beta\widehat{S} - \varphi M_0)^{-1}, \qquad \varphi^{-1} = 1 - \frac{1}{n}\operatorname{tr}(M_0 T), \\ W &= \frac{\varphi^2}{n}\operatorname{tr}(M_0 T M_0 T). \end{aligned} \tag{44}$$

Here T represents the inverse response after averaging over source-mean sampling while holding the contrast covariance fixed; φ supplies the corresponding scalar correction. Start at large λ, where $\varphi \to 1$, and follow the solution while T is positive definite and $W < 1$. A detected signal corresponds to a root

$$a\varphi U_A^{\mathrm{T}} T U_A = 1 \tag{45}$$

on this branch whose limiting eigenvalue stays strictly above the no-signal spectrum. At that root, define

$$\begin{aligned} \Gamma_\alpha &= \operatorname*{plim}_{n\to\infty} \frac{\|TU_A\|^2}{(U_A^{\mathrm{T}} T U_A)^2}, \\ \Omega_\alpha &= \operatorname*{plim}_{n\to\infty} \frac{\varphi^2 \operatorname{tr}(M_0 T^2)/n}{1-W} \left\{ \frac{U_A^{\mathrm{T}} T M_0 T U_A}{(U_A^{\mathrm{T}} T U_A)^2} + a \right\}. \end{aligned} \tag{46}$$

The notation plim means convergence in probability under Assumption 5.1. These limits are deterministic functions of the model parameters; the equations describe theoretical risk, rather than a rule requiring knowledge of U_A when fitting the method.

The response T at the separated root is denoted by T_α in the main text. The local notation q below is identical to β. The deterministic construction that follows computes the displayed limits without sampling matrices.

Deterministic calculation. First suppose $c_\perp > 0$, and put $\pi_B = c/\eta \in (0,1)$. We describe a deterministic trace calculation; it requires no sampled matrices. Let P be a projection and w a positive operator in a tracial probability space with faithful trace t, such that $\mathsf{t}(\mathsf{P}) = \pi_B$, w has the mean-one Marchenko–Pastur law of aspect y, and P and w are free. Here freeness specifies their mixed moments: the trace of a product of centered factors alternating between the two generated algebras is zero. Together with

$$\mathsf{P}^2 = \mathsf{P}, \qquad \mathsf{t}(\mathsf{w}^j) = \sum_{\pi \in \mathrm{NC}(j)} y^{j-|\pi|}, \qquad j \ge 1,$$

this determines every mixed polynomial moment. The sum is over the noncrossing partitions of $\{1,\dots,j\}$. Define

$$\begin{aligned} \mathsf{s} &= (v+\tau\mathsf{P})^{1/2}\mathsf{w}(v+\tau\mathsf{P})^{1/2}, \\ \mathsf{m} &= \theta_0(1-\mathsf{P}) + \theta_B\mathsf{P}, \\ \mathsf{t}_0(A) &= \frac{\mathsf{t}((1-\mathsf{P})A)}{1-\pi_B}. \end{aligned} \tag{47}$$

The functional t_0 is a normalized positive state; it describes a fixed direction in the noise-only block, which has the same contrast variance as the signal. The spectral support of s is contained in $[v(1-\sqrt{y})^2, s_B(1+\sqrt{y})^2] \subset (0,\infty)$.

For a fixed finite power α, define $K_\lambda = \lambda\mathsf{s}^\alpha - q\mathsf{s}$. Let z be an additive spectral shift; differentiation in z will give the Euclidean overlap. For sufficiently large λ and then by continuation, solve

$$T = (K_\lambda + z - \varphi\mathsf{m})^{-1}, \qquad \varphi^{-1} = 1 - \eta\mathsf{t}(\mathsf{m}T), \qquad j = \mathsf{t}_0(T), \qquad \Phi(\lambda, z) = j^{-1} - a\varphi. \tag{48}$$

The relevant branch has $\varphi \to 1$ as $\lambda \to \infty$ at $z = 0$. Follow it while T is positive and

$$W = \eta\varphi^2\mathsf{t}(\mathsf{m}T\mathsf{m}T) < 1.$$

On this branch the candidate signal eigenvalue is the solution of $\Phi(\lambda, 0) = 0$ above the upper bulk edge. It is unique if it exists. All operators and traces in (48) are deterministic functions of $(\alpha, a, \kappa, \tau, \sigma, \chi, c, c_\perp, b)$.

For deterministic evaluation at any finite α, approximate x^α uniformly by polynomials on the displayed interval, then approximate the inverse in T on its compact positive spectrum. The mixed-moment rules evaluate every resulting trace; increasing the degrees gives the trace quantities and derivatives on compact exterior domains. Start at large λ, where a Neumann series converges, and continue the unique branch. The joint construction retains the noncommutativity of s and P.

The exterior branch and its upper edge. The following deterministic regularity condition is used for the positive-proportion case. As λ decreases from infinity at $z = 0$, the branch in (48) reaches a finite first point $(\lambda_{+,\alpha}, \varphi_+)$ at which

$$W = 1, \qquad \inf \mathrm{spec}(K_{\lambda_{+,\alpha}} - \varphi_+\mathsf{m}) > 0. \tag{49}$$

This deterministic condition requires the sample edge to precede a singularity of the conditional population resolvent. It also covers the signal detection boundary. It is an explicit hypothesis for $c_\perp > 0$; the scalar argument below establishes edge regularity when $c_\perp = 0$.

Proof of the edge and uniqueness claims. Put

$$F(\lambda, z, \varphi) = \varphi^{-1} - 1 + \eta\mathsf{t}\{\mathsf{m}(K_\lambda + z - \varphi\mathsf{m})^{-1}\}.$$

All derivatives below keep the other displayed arguments fixed. At the edge,

$$\begin{aligned} F_\varphi &= -\varphi^{-2} + \eta\mathsf{t}(\mathsf{m}T\mathsf{m}T) = 0, \\ F_z &= -\eta\mathsf{t}(\mathsf{m}T^2) < 0, \\ F_{\varphi\varphi} &= 2\varphi^{-3} + 2\eta\mathsf{t}\{(\mathsf{m}^{1/2}T\mathsf{m}^{1/2})^3\} > 0, \\ F_\lambda &= -\eta\mathsf{t}(\mathsf{m}T\mathsf{s}^\alpha T) < 0. \end{aligned} \tag{50}$$

Thus the edge is a nondegenerate square-root branch point. Above it, $F_\varphi < 0$, so $\varphi_\lambda < 0$. Consequently $j_\lambda < 0$ and $\Phi_\lambda > 0$. The signal equation has at most one root on this upper exterior branch. Equivalently, there is a separated root precisely when the limiting value of $\Phi(\lambda, 0)$ at the edge is negative; equality is the detection boundary. □

At a separated root λ_*, evaluate all quantities at $(\lambda, z) = (\lambda_*, 0)$ and define

$$\begin{aligned}\Gamma_\alpha &= \frac{\mathfrak{t}_0(T^2)}{j^2},\\ \Omega_\alpha &= \frac{\eta\varphi^2\mathfrak{t}(\mathsf{m}T^2)}{1-W}\left\{\frac{\mathfrak{t}_0(T\mathsf{m}T)}{j^2}+a\right\}.\end{aligned} \tag{51}$$

These evaluate the limits in (46): the proof below sends $U_A^{\mathrm{T}}(\cdot)U_A$ to $\mathfrak{t}_0(\cdot)$ and $n^{-1}\operatorname{tr}(\cdot)$ to $\eta\mathfrak{t}(\cdot)$, with M_0 replaced by m. The subsequent compression argument covers $c_\perp = 0$.

Since $\mathfrak{t}_0$ is a probability state, $\Gamma_\alpha = 1 + \operatorname{Var}_{\mathfrak{t}_0}(T)/j^2 \geq 1$; positivity also gives $\Omega_\alpha \geq 0$.

The risk at a fixed power.

Proof of Theorem 5.1 when $c_\perp > 0$. Use the independent Gaussian covariances in (41). Since $d/[n(b-1)] \to y < 1$, the spectrum of $\widehat{S}$ stays in a fixed positive compact interval with probability tending to one.

We first record the needed deterministic limit of the contrast matrix jointly with P_B. Gaussian mixed-moment expansion gives the trace limit (47). The same convergence is strong: polynomial operator norms and spectral supports converge. One way to obtain the latter statement without changing real Gaussian entries to complex entries is to linearize the rectangular Gaussian blocks and use the Gaussian spectral bound of Bandeira et al. (2023, Theorem 2.1). To spell out the reduction, a self-adjoint linearization uses rectangular blocks with coefficients $L \otimes E_{ij} + L^* \otimes E_{ji}$, where the two indices belong to disjoint row and column spaces. Its variance map $B \mapsto \sum_{ij} A_{ij} B A_{ij}$ preserves the finite algebra of matrices constant within the background-row, noise-row, and sample-column blocks. On that algebra it is exactly the rectangular covariance map: the off-diagonal products vanish, and the remaining coefficients are the three block proportions. Thus the free Gaussian comparison in that theorem is the block-valued free model for the same rectangular dimensions. The proportions converge, and the associated finite covariance maps converge; the semicircular Fock-space construction makes their spectra continuous in these coefficients. The Gaussian covariance parameter in the theorem is $O(n^{-1/2})$, whereas the operator variance remains bounded, so its spectral inclusion error tends to zero. Apply this inclusion to the linearization of a self-adjoint polynomial, with a spectral parameter in a compact set outside its limiting spectrum. It excludes zeros there, giving the upper spectral inclusion for the polynomial. The mixed-moment limit and faithfulness of the limiting trace supply the matching lower inclusion. Uniform polynomial approximation then extends it to the powers and exterior resolvents used here.

The distribution of this pair is invariant under orthogonal rotations within Q_0. Since $\operatorname{rank}(Q_0)$ is proportional to n, concentration of a quadratic form of a Haar unit vector yields, for every bounded exterior resolvent expression A_n,

$$U_A^{\mathrm{T}} A_n U_A - \frac{\operatorname{tr}(Q_0 A_n)}{\operatorname{rank}(Q_0)} \xrightarrow{p} 0.$$

Combined with the joint trace limit, this identifies the signal-type state as $\mathfrak{t}_0$. The same argument applies to products used in the derivatives below.

For fixed λ, consider the original-coordinate pencil $L_\lambda = \widehat{M}_{\mathrm{av}} + q\widehat{S} - \lambda\widehat{S}^\alpha$. Conditional on $\widehat{S}$, the covariance resolvent deterministic equivalent, first at large positive z, is

$$(zI - L_\lambda)^{-1} \;\simeq\; (\lambda\widehat{S}^\alpha - q\widehat{S} + zI - \varphi M_{\mathrm{av}})^{-1},$$

where $\simeq$ denotes convergence of the specified bounded bilinear forms, and the scalar equation for φ uses the normalized trace $n^{-1}\operatorname{tr}$. This is the conditional covariance anisotropic resolvent law (Knowles & Yin, 2017). Removing the rank-one term aP_A changes its normalized trace by $o(1)$. The joint contrast limit gives (48). Sherman–Morrison therefore yields the signal resolvent

$$U_A^{\mathrm{T}}(zI - L_\lambda)^{-1} U_A \xrightarrow{p} \frac{j}{1 - a\varphi j} = \Phi(\lambda, z)^{-1}. \tag{52}$$

Initially this holds away from every spectrum. Compact support and uniqueness of the Stieltjes transform then identify the limiting signal spectral measure and its analytic continuation.

To identify actual separated eigenvalues, write $K_n = \lambda \widehat{S}^{\alpha} - q\widehat{S}$. Near the largest generalized eigenvalue, K_n is uniformly positive. Indeed, testing the generalized Rayleigh quotient at a contrast eigenvector maximizing $qx^{1-\alpha}$ adds at least a fixed positive amount from the independent source-mean covariance; its smallest population variance is $\theta_0 > 0$. Conditional on the contrast matrix, the transformed covariance has population matrix

$$\Theta_n = K_n^{-1/2} M_0 K_n^{-1/2} + a K_n^{-1/2} P_A K_n^{-1/2}.$$

The first term has strong limiting spectrum given by $K_\lambda^{-1/2} \mathsf{m} K_\lambda^{-1/2}$, with no extra population eigenvalues. The second term is a positive rank-one perturbation. Its exterior secular equation is $1 - a\varphi j = 0$ at population spectral parameter $1/\varphi$. The usual general-covariance spike map gives

$$\psi(1/\varphi) = \varphi^{-1} + \eta \mathfrak{t}(\mathsf{m}T) = 1, \qquad \psi'(1/\varphi) = 1 - W.$$

Thus exact separation for sample covariances (Bai & Yao, 2012) identifies the separated root above the bulk when it exists, and the upper bulk edge otherwise. A strictly decreasing crossing argument in λ transfers this statement to the largest generalized eigenvalue. In particular, weak spectral convergence alone is not being used to infer the absence of exterior eigenvalues.

It remains to compute the Euclidean overlap. Differentiating (48) in z, at fixed λ, gives

$$\varphi_z = -\frac{\eta \varphi^2 \mathfrak{t}(\mathsf{m}T^2)}{1 - W}, \qquad j_z = -\mathfrak{t}_0(T^2) + \varphi_z \mathfrak{t}_0(T\mathsf{m}T), \qquad \Phi_z = \Gamma_\alpha + \Omega_\alpha.$$

At a separated root the mass of the isolated eigenvalue in (52) is its residue $1/\Phi_z$. Perturbing λ_* to the fitted generalized eigenvalue changes L_λ by $o_p(1)$ in operator norm. Continuity of an isolated spectral projector therefore gives

$$|U_A^{\mathrm{T}} u_\alpha|^2 \xrightarrow{p} (\Gamma_\alpha + \Omega_\alpha)^{-1}.$$

The derivative is taken in $zI - L_\lambda$, so the residue uses the original Euclidean normalization.

Finally, under (49), (50) gives a square-root branch point at the upper edge. At a critical signal root,

$$\partial_\varphi \Phi = -\frac{\mathfrak{t}_0(T\mathsf{m}T)}{j^2} - a < 0.$$

Consequently $\Phi^{-1} = O(z^{-1/2})$, and there is no signal atom at the edge. The same conclusion holds in the strictly subcritical case, where Φ has a nonzero limiting value. Projecting the fitted vector onto a shrinking interval around the limiting edge, and then using weak convergence of the signal spectral measure, gives zero limiting overlap. This proves risk one both below and at the detection boundary. □

A vanishing proportion of noise-only directions. When $c_\perp = 0$, the number r may be zero, fixed, or diverge with $r/n \to 0$. The normalized state in (47) is then undefined and must not be obtained by substituting $\pi_B = 1$. A separate compression argument recovers the scalar formula of Section E.3.

Proof of Theorem 5.1 when $c_\perp = 0$. Let $l = r + 1 = \operatorname{rank}(Q_0) = o(n)$. In bases of Q_0 and P_B, write the independent contrast rows as $Z_0 \in \mathbb{R}^{l \times N}$ and $Z_B \in \mathbb{R}^{m \times N}$, where $N = n(b-1)$. For a spectral parameter away from the background contrast support, the Schur complement is

$$Q_0(zI - \widehat{S})^{-1} Q_0 = \left[zI_l - \frac{v}{N} Z_0 A_B(z) Z_0^{\mathrm{T}} \right]^{-1},$$
$$A_B(z) = I_N + \frac{s_B}{N} Z_B^{\mathrm{T}} (zI_m - s_B Z_B Z_B^{\mathrm{T}}/N)^{-1} Z_B.$$

Conditionally on Z_B, the Gaussian rectangular quadratic form obeys

$$\left\| N^{-1} Z_0 A_B(z) Z_0^{\mathrm{T}} - N^{-1} \operatorname{tr} A_B(z)\, I_l \right\| = O_p(\sqrt{l/N} + N^{-1/2}) = o_p(1)$$

on fixed compact resolvent contours. A finite covering and the resolvent identity make the convergence uniform there. The background MP equation gives the same scalar signal transform $g_\nu(z)$ as in Section E.3. Contour functional calculus, followed by uniform polynomial approximation, yields

$$\left\| Q_0 f(\widehat{S}) Q_0 - \left(\int f \, d\nu \right) Q_0 \right\| \xrightarrow{p} 0. \tag{53}$$

This statement is used for bounded continuous functions on the full contrast spectrum, including its possible isolated small eigenvalue cluster. The Schur equation also confines that cluster to the scalar contrast outlier location; its multiplicity may grow, but is $o(n)$.

Write $h = \theta_0 - \theta_B \le 0$, and let J and φ denote the scalar full-background quantities, with $J = \int R\, d\nu$. The trace equation still has the full-background limit because all removed population terms have rank $o(n)$, with bounded norm in the exterior domain. Using (53) and Woodbury in the whole Q_0 block gives

$$j_0 = \frac{J}{1 - h\varphi J}, \qquad \Phi = j_0^{-1} - a\varphi = J^{-1} - (\theta_A - \theta_B)\varphi.$$

Thus the limiting signal resolvent is exactly the one in the scalar formula, including its pole-safe continuation. The same compression, applied to the bounded resolvent derivatives, gives the residue and hence its stated risk. More explicitly, the two signal-direction ratios in (46) converge to γ_α and $\theta_B\gamma_\alpha + \theta_0 - \theta_B$, respectively, while the trace prefactor converges to $\omega_\alpha/\{\theta_B(1-\omega_\alpha)\}$. Thus, with $\gamma_\alpha, \omega_\alpha$ defined in Section E.3 and $\theta_A = \theta_0 + a$,

$$\Gamma_\alpha = \gamma_\alpha, \qquad \Omega_\alpha = \frac{\omega_\alpha}{1-\omega_\alpha}\left(\gamma_\alpha + \frac{\theta_A}{\theta_B} - 1\right) \ge 0.$$

This verifies the matrix definitions also when $c_\perp = 0$.

One must also check that the extra noise-only directions cannot replace the leading signal root. Their possible exterior cluster has secular equation $\Phi_0 = J^{-1} - h\varphi = 0$. If it lies above the background bulk, j_0 has a pole at its upper location, while $j_0^{-1} - a\varphi$ is negative immediately to its right. It is positive at infinity and strictly increasing on that upper exterior branch, so the fixed positive signal strength a creates a strictly larger separated root. If no such noise cluster separates, the scalar background edge controls the top eigenvalue unless the signal separates. Therefore the additional $o(n)$ noise directions do not change the limiting top signal risk. The scalar edge argument in Section E.3 handles the remaining subcritical and critical cases. □

One signal equation and a positive error decomposition. The following representation separates the background spectrum from signal strength. It also explains the two risk factors without assuming that the contrast covariance and the background projection commute. For $c_\perp > 0$, retain the algebra generated by (s, P) in (47). Adjoin a mean-one Marchenko–Pastur variable w_{av} of aspect η, free from that algebra, and define

$$\mathsf{C} = \mathsf{m}^{1/2}\mathsf{w}_{\mathrm{av}}\mathsf{m}^{1/2}, \qquad \mathsf{L}_\alpha = \mathsf{s}^{-\alpha/2}(q\mathsf{s} + \mathsf{C})\mathsf{s}^{-\alpha/2}, \qquad R_\lambda = (\lambda\mathsf{s}^\alpha - q\mathsf{s} - \mathsf{C})^{-1}. \tag{54}$$

The operator C represents the independent source-mean sampling fluctuations with the signal removed. Thus $\lambda_{+,\alpha} = \sup \mathrm{spec}(\mathsf{L}_\alpha)$ and this bulk model depend on the nuisance parameters, but not on a. The state $\mathfrak{t}_0$ extends by the same projection-weighted trace to these operators.

Proposition E.1 (A monotone detection equation and sampling error)**.** For $\lambda > \lambda_{+,\alpha}$, define

$$g_\alpha(\lambda) = \mathfrak{t}_0(R_\lambda), \quad h_\alpha(\lambda) = \mathfrak{t}(\mathsf{C}R_\lambda), \quad D_\alpha(\lambda) = \{1 + \eta h_\alpha(\lambda)\}g_\alpha(\lambda). \tag{55}$$

Then D_α is positive and strictly decreasing to zero. A separated signal root exists precisely when $aD_\alpha(\lambda^+_{+,\alpha}) > 1$; it is the unique solution of $aD_\alpha(\lambda) = 1$. At that root, $\varphi = 1 + \eta h_\alpha(\lambda)$, and the conditional expectation of R_λ given (s, P) is the response T in (48). With $j = g_\alpha(\lambda)$,

$$\Gamma_\alpha = \frac{\mathfrak{t}_0(T^2)}{j^2}, \qquad \Omega_\alpha = \frac{\mathfrak{t}_0\{(R_\lambda - T)^2\}}{j^2} + a\eta\mathfrak{t}(\mathsf{C}R_\lambda^2). \tag{56}$$

Both terms in Ω_α are nonnegative.

The first sampling term measures how far the full response fluctuates around its conditional mean. The second measures its sensitivity to source-mean sampling in the signal equation. Together with $\Gamma_\alpha - 1 = \mathrm{Var}_{\mathfrak{t}_0}(T)/j^2$, these form a nonnegative decomposition of the excess risk denominator. The equation gives a unique outer scalar inversion; the forward spectral calculation remains necessary. It is not a finite closed form or a data-based tuning rule.

Proof. Let $\mathcal{B}$ be the algebra generated by (s, P), and write $E_{\mathcal{B}}$ for its trace-preserving conditional expectation. Since $1 - \mathsf{P} \in \mathcal{B}$, the weighted state also satisfies $\mathsf{t}_0 \circ E_{\mathcal{B}} = \mathsf{t}_0$. The scalar free cumulants of w_{av} are η^{r-1}. Their conditional cumulant expansion gives $\mathcal{R}_{\mathsf{C}}^{\mathcal{B}}(Z) = \mathsf{m}/\{1-\eta\mathsf{t}(\mathsf{m}Z)\}$, where $\mathcal{R}_{\mathsf{C}}^{\mathcal{B}}$ denotes the conditional cumulant generating map. The corresponding resolvent equation yields $E_{\mathcal{B}}R_\lambda = T$ and the fixed-point equation in (48). This is the free deterministic-equivalent construction (Speicher & Vargas, 2012); it follows first by the convergent expansion at infinity and then by continuation along the exterior positive branch. Taking traces in $(\lambda\mathsf{s}^\alpha - q\mathsf{s})R_\lambda = 1 + \mathsf{C}R_\lambda$ and its conditional expectation gives $h_\alpha = \varphi\mathsf{t}(\mathsf{m}T)$, hence $\varphi = 1+\eta h_\alpha$. The original signal equation is therefore $aD_\alpha = 1$.

The factorization $R_\lambda = \mathsf{s}^{-\alpha/2}(\lambda - \mathsf{L}_\alpha)^{-1}\mathsf{s}^{-\alpha/2}$ shows that g_α and h_α are transforms of positive finite measures. Their derivatives in λ are strictly negative above the spectrum. This proves monotonicity, uniqueness, and the detection criterion. Differentiating the fixed-point equation also gives $(1 - W)\varphi_\lambda = -\eta\varphi^2\mathsf{t}(\mathsf{m}T\mathsf{s}^\alpha T) < 0$; since $\varphi_\lambda = \eta h'_\alpha < 0$, this is the stable branch.

To calculate overlap, add zI inside the inverse and differentiate in z, holding λ fixed. Then $\varphi_z = -\eta\mathsf{t}(\mathsf{C}R_\lambda^2)$ and $\mathsf{t}_0(R_\lambda^2) = \mathsf{t}_0(T^2) - \varphi_z\mathsf{t}_0(T\mathsf{m}T)$. Substitution in (51) proves the formula, because conditional expectation gives $\mathsf{t}_0\{(R_\lambda - T)^2\} = \mathsf{t}_0(R_\lambda^2) - \mathsf{t}_0(T^2) \geq 0$. The scalar-shift derivative is essential: a derivative in λ inserts s^α and does not give Euclidean overlap. □

The proposition is stated for two positive-proportion blocks. At $c_\perp = 0$, the preceding compression argument and the scalar recipe below supply the directional limit, including any isolated contrast atom; one must not substitute $\pi_B = 1$ into the normalized state. Risk one at and below detection uses the regular-edge conclusion of Theorem 5.1.

E.3 Explicit scalar formulas for the spanning submodel

This subsection assumes $m = d - 1$, so $c_b = c/(b - 1)$. It gives the scalar formulas used to evaluate the risk curves. The preceding compression argument extends the fixed-finite-power risk to $d - 1 - m = o(n)$; the contrast-only endpoint below retains the stated $m = d - 1$ assumption.

Contrast spectral distributions. For every continuous function f on the limiting spectral support,

$$\frac{1}{d}\operatorname{tr} f(\widehat{S}) \xrightarrow{p} \int f\, d\mu, \qquad U_A^{\mathrm{T}} f(\widehat{S}) U_A \xrightarrow{p} \int f\, d\nu. \tag{57}$$

The probability measure μ describes the background spectrum. The measure ν weights each eigenvalue by its eigenvector's squared overlap with the true signal. Both are explicit. Set $x_\pm = s_B(1 \pm \sqrt{c_b})^2$. Then

$$\mu(dx) = \frac{\sqrt{(x_+ - x)(x - x_-)}}{2\pi c_b s_B x}\mathbf{1}_{[x_-, x_+]}(x)\, dx. \tag{58}$$

This is the scaled Marchenko–Pastur law. When $\tau = 0$, $\nu = \mu$. Otherwise define

$$\begin{aligned} \delta &= 1 - \sigma^2\chi^2/s_B, \qquad x_0 = \sigma^2\chi^2(1 - c_b/\delta), \\ w_0 &= \begin{cases} (1 - c_b/\delta^2)/(1 - c_b/\delta), & \delta > \sqrt{c_b}, \\ 0, & \delta \leq \sqrt{c_b}. \end{cases} \end{aligned} \tag{59}$$

The signal-weighted distribution is

$$\nu(dx) = \frac{c_b\sigma^2\chi^2}{\delta(x - x_0)}\mu(dx) + w_0\delta_{x_0}(dx), \tag{60}$$

where δ_{x_0} denotes unit mass at x_0. The atom is included only when $w_0 > 0$; then x_0 is an isolated small contrast eigenvalue and w_0 is its squared signal overlap. When the atom is absent, x_0 is only an auxiliary number and no powers of it are evaluated. Write $\mathcal{K}_S = \operatorname{supp}(\mu + \nu)$ for the full limiting contrast spectrum.

Risk calculation. Fix α and a candidate generalized eigenvalue λ of $\widehat{D}w = \lambda\widehat{S}^{\alpha}w$. Solve the scalar fixed-point equation

$$R(x) = \frac{1}{\lambda x^{\alpha} - \frac{b-1}{b}x - \theta_B\varphi}, \qquad \varphi = \left(1 - c\theta_B \int R\, d\mu\right)^{-1}. \tag{61}$$

Here φ is an auxiliary scalar determined by this equation, and R is the resulting spectral response function. At each solution define

$$\omega(\lambda) = c\theta_B^2\varphi^2 \int R^2\, d\mu, \qquad \gamma(\lambda) = \frac{\int R^2\, d\nu}{(\int R\, d\nu)^2}. \tag{62}$$

Start at large λ on the branch with $\varphi \to 1$, and continue downward while the denominator of R is positive on $\operatorname{supp}\mu$ and $\omega(\lambda) < 1$. The upper background eigenvalue edge $\lambda_{+,\alpha}$ is reached when $\omega(\lambda) = 1$. Above that edge, the signal eigenvalue solves

$$\left(\int R\, d\nu\right)^{-1} = (\theta_A - \theta_B)\varphi. \tag{63}$$

Choose its largest root λ_* satisfying $\lambda x^{\alpha} - (b-1)x/b > 0$ for all $x \in \mathcal{K}_S$, and set $\omega_\alpha = \omega(\lambda_*)$ and $\gamma_\alpha = \gamma(\lambda_*)$. If $\theta_A = \theta_B$, a pole of $\int R\, d\nu$ may itself be the root: interpret its reciprocal as zero and $\gamma_\alpha = 1/w_0$ by continuity. If there is no admissible root strictly above the edge, the signal is undetectable by this power.

Here $\gamma_\alpha = (\int R^2\, d\nu)/(\int R\, d\nu)^2 \geq 1$ and $0 \leq \omega_\alpha < 1$. All integrals are one-dimensional; the atomic case uses the displayed continuous extension.

At a separated root the risk specializes to

$$\mathcal{R}_\alpha = 1 - \frac{1 - \omega_\alpha}{\gamma_\alpha + (\theta_A/\theta_B - 1)\omega_\alpha}. \tag{64}$$

If no admissible root separates, the risk is one.

Lemma E.1 (The two contrast spectral measures)**.** The limits in equation (57) hold with the explicit probability measures in equations (58) and (60). In particular, the signal spectral measure $\nu_n(E) = U_A^{\mathrm{T}}\mathbf{1}_E(\widehat{S})U_A$, for Borel sets E, converges weakly in probability to ν. The limiting full contrast spectrum is $\mathcal{K}_S = [x_-, x_+] \cup \{x_0 : w_0 > 0\}$, with the atom and its inclusion condition specified in equation (59).

Proof of Lemma E.1. Use the convention $g_\eta(z) = \int (z-x)^{-1}\eta(dx)$. The Gaussian covariance resolvent equation gives

$$\begin{aligned} z &= \frac{1}{g_\mu(z)} + \frac{s_B}{1 - c_b s_B g_\mu(z)}, \\ g_\nu(z) &= \frac{1}{z - \sigma^2\chi^2/(1 - c_b s_B g_\mu(z))}. \end{aligned} \tag{65}$$

The branch is fixed by $g_\mu(z) \sim 1/z$ at infinity. For completeness, the second equation follows by deleting the first row of Z in equation (41). The resulting background block is $s_B Z_B Z_B^{\mathrm{T}}/N$. Its independent Gaussian first row has a conditional quadratic form concentrating about the resolvent trace. Substitution of the first equation in the Schur complement yields the second. Bounded spectral support and uniqueness of a Stieltjes transform give weak convergence of ν_n.

Writing $\rho_S = \sigma^2\chi^2/s_B$, the second equation is also

$$g_\nu(z) = \frac{g_\mu(z)}{\rho_S + (1 - \rho_S)z g_\mu(z)}.$$

For $x \in (x_-, x_+)$, the quadratic equation for g_μ gives $|g_\mu(x + i0)|^2 = (c_b s_B x)^{-1}$. Taking imaginary parts therefore yields the density ratio in equation (60). The only possible exterior pole solves $z = \sigma^2\chi^2/(1 - c_b s_B g_\mu(z))$, giving $z = x_0$. It lies on the lower exterior branch exactly when $\delta > \sqrt{c_b}$, and its residue is w_0. At equality the density is integrably singular at x_-, with no atom. The one-spike covariance eigenvalue separation result rules out any other exterior spectrum; it also follows by the same Schur equation and the background spectral edge limit. These are the single-spike instances of the covariance results used below (Bai & Yao, 2012; Paul, 2007). □

Conditional covariance reduction and eigenvalue location. For fixed λ, define the original-coordinate pencil

$$L_\lambda = \widehat{M}_{\mathrm{av}} + \tfrac{b-1}{b}\widehat{S} - \lambda \widehat{S}^{\alpha}.$$

For a spectral parameter z above its spectrum, put

$$d(x;\lambda,z) = z + \lambda x^{\alpha} - \tfrac{b-1}{b}x - \theta_B\varphi, \quad I_1 = \int d(x;\lambda,z)^{-1}\mu(dx), \quad \varphi = (1 - c\theta_B I_1)^{-1}. \tag{66}$$

All branches below extend the solution with $\varphi \to 1$ as $z \to +\infty$. Write $J = \int d^{-1}\, d\nu$ and $\Phi = J^{-1} - (\theta_A - \theta_B)\varphi$, extending equation (63) to z. At general z, write $\omega = c\theta_B^2\varphi^2 \int d^{-2}\, d\mu$.

Lemma E.2 (Signal resolvent and exterior roots)**.** The signal spectral measure of L_λ converges weakly in probability. Its transform is

$$\Phi(\lambda, z)^{-1} = \frac{J(\lambda, z)}{1 - (\theta_A - \theta_B)\varphi J(\lambda, z)}. \tag{67}$$

Let $\Lambda_0 = \frac{b-1}{b}\max_{x\in\mathcal{K}_S} x^{1-\alpha}$, and let $\mathcal{Z}_\alpha$ be the exterior zeros of $\Phi(\lambda, 0)$ on the branch $\omega < 1$, with $\lambda > \max(\lambda_{+,\alpha}, \Lambda_0)$. The largest fitted generalized eigenvalue converges to $\max \mathcal{Z}_\alpha$ if this set is nonempty, and to $\lambda_{+,\alpha}$ otherwise. At each separated signal root the corresponding exterior eigenvalue is simple.

Proof of Lemma E.2. Put $K_{\lambda,z} = \lambda\widehat{S}^{\alpha} - (b-1)\widehat{S}/b + zI$. Conditional on $\widehat{S}$, the exact covariance reduction is

$$(zI - L_\lambda)^{-1} = K_{\lambda,z}^{-1/2}\{I - K_{\lambda,z}^{-1/2}\widehat{M}_{\mathrm{av}}K_{\lambda,z}^{-1/2}\}^{-1}K_{\lambda,z}^{-1/2}.$$

The anisotropic covariance law first applies for large z. Here $M_{\mathrm{av}} = \theta_B I + (\theta_A - \theta_B)P_A$; its rank-one term changes the normalized trace equation by $O(n^{-1})$. Sherman–Morrison and Lemma E.1 give (66) and (67). Compact support and uniqueness of the Stieltjes transform identify the limiting signal measure and its continuation.

For eigenvalue location, the independent source-mean quadratic form in any contrast eigenvector has a strictly positive limit. Testing the generalized Rayleigh quotient in a direction maximizing $(b-1)x^{1-\alpha}/b$ therefore puts the largest fitted root strictly above Λ_0. Thus $K_\lambda = \lambda\widehat{S}^{\alpha} - (b-1)\widehat{S}/b$ is uniformly positive near that root. The conditional population covariance is

$$\Theta_\lambda = \theta_B K_\lambda^{-1} + (\theta_A - \theta_B)K_\lambda^{-1/2}P_A K_\lambda^{-1/2}.$$

Its bulk law is the image μ_λ of μ under $x \mapsto \theta_B/\{\lambda x^{\alpha} - (b-1)x/b\}$. The only possible exterior population eigenvalues come from the contrast atom and this rank-one update. Their secular equation is

$$1 - (\theta_A - \theta_B)\int \frac{\nu(dx)}{t\{\lambda x^{\alpha} - (b-1)x/b\} - \theta_B} = 0.$$

For $\theta_A = \theta_B$, the original contrast atom is retained. The population-to-sample covariance map is

$$\psi_\lambda(t) = t\left(1 + c\int \frac{x}{t-x}\mu_\lambda(dx)\right), \qquad \psi_\lambda'(t) = 1 - c\int \frac{x^2}{(t-x)^2}\mu_\lambda(dx).$$

At $t = 1/\varphi$, its value is one and its derivative is $1 - \omega$. Exact separation for general sample covariances (Bai & Yao, 2012) therefore identifies the exterior zeros of Φ and excludes additional sample eigenvalues above the edge. The covariance hypotheses hold conditionally: Gaussian columns are independent, population spectra are bounded, and functional calculus and the single contrast atom specify all exterior population spikes. The largest eigenvalue of $K_\lambda^{-1/2}\widehat{M}_{\mathrm{av}}K_\lambda^{-1/2}$ is strictly decreasing in λ; its crossing of one gives the largest fitted generalized eigenvalue. Bracketing this crossing by fixed exterior points proves the asserted convergence. The secular derivative is nonzero at a separated root; a retained contrast atom has nonzero mass. This also proves simplicity. □

The atomic convention has a useful explicit implementation. If $w_0 > 0$, set $d_{\mathrm{at}} = d(x_0;\lambda,z)$ and $J_c = \int d^{-1}\, d\nu_{\mathrm{ac}}$. Then

$$\Phi(\lambda, z)^{-1} = \frac{d_{\mathrm{at}}J_c + w_0}{d_{\mathrm{at}}\{1 - (\theta_A - \theta_B)\varphi J_c\} - (\theta_A - \theta_B)\varphi w_0}. \tag{68}$$

Solving for zeros of this continuous denominator avoids either losing the $\theta_A = \theta_B$ root or misidentifying a pole of Φ as a root.

Euclidean overlap and the detection boundary.

Proof of the scalar risk formula (64). Write $\upsilon = z - \theta_B\varphi$, and define $I_j = \int d^{-j}\, d\mu$, $J_j = \int d^{-j}\, d\nu$. For fixed λ, the upper exterior branch satisfies

$$\partial_\upsilon I_1 = -I_2, \quad \varphi_\upsilon = -c{\theta_B}^2\varphi^2 I_2, \quad z_\upsilon = 1 - c{\theta_B}^2\varphi^2 I_2 = 1 - \omega.$$

Consequently

$$\varphi_z = -\frac{\omega}{\theta_B(1-\omega)}, \qquad J_z = -\frac{J_2}{1-\omega}, \qquad \Phi_z = \frac{J_2/J^2 + (\theta_A/\theta_B - 1)\omega}{1-\omega}. \tag{69}$$

At a separated root λ_*, lemma E.2 gives one isolated eigenvalue of L_{λ_*} near zero. Its limiting signal mass is the residue of equation (67), namely $1/\Phi_z$. The weak spectral convergence suffices here because the exterior eigenvalue count gives a fixed isolating interval. Since

$$\|L_{\widehat{\lambda}_\alpha} - L_{\lambda_*}\| \le |\widehat{\lambda}_\alpha - \lambda_*|\, \|\widehat{S}^\alpha\| \xrightarrow{p} 0,$$

spectral projection continuity transfers this overlap to the actual fitted direction. This proves equation (64) on the separated branch. It is a Euclidean overlap because the derivative is in $zI - L_\lambda$; a derivative in λ alone would introduce an $\widehat{S}^\alpha$-weighted norm. At an atomic pole, $J_2/J^2 \to 1/w_0$, as stated.

We check the edge and the no-outlier case explicitly. Along the exterior branch,

$$z_{\upsilon\upsilon} = 2c{\theta_B}^2\varphi^2\{I_3 + c\theta_B\varphi I_2^2\} > 0. \tag{70}$$

The branch reaches $\omega = 1$ before its bulk denominator vanishes: as the denominator approaches its minimum on the support, I_2 diverges. At an endpoint this follows from the square-root Marchenko–Pastur density; at an interior minimum it is stronger. The only constant transform, $\alpha = 1, \lambda = \frac{b-1}{b}$, leaves the positive covariance $\widehat{M}_{\mathrm{av}}$ and cannot be a zero upper-edge crossing. Thus the denominator at the relevant edge is strictly positive, and equation (70) is finite and nonzero. The upper edge is a regular square-root edge. Its derivative in λ, evaluated where $\omega = 1$, is

$$-c{\theta_B}^2\varphi^2 \int x^\alpha d(x)^{-2}\mu(dx) < 0.$$

It crosses zero once, from the positive edge at $\lambda = 0$ to a negative edge for large λ. This also proves the edge uniqueness used in equation (61).

Since ν is a probability measure, Cauchy–Schwarz gives $J_2/J^2 \ge 1$. At the edge,

$$\Phi_\upsilon = J_2/J^2 + (\theta_A/\theta_B - 1)\omega \ge \theta_A/\theta_B > 0.$$

If the signal root reaches the edge, Φ has a square-root zero there; hence $\Phi(\lambda, z)^{-1} = O(z^{-1/2})$, rather than a pole of order one. There is no signal atom at zero. The same conclusion holds when Φ is nonzero, when $J = 0$, or when an atomic pole is handled by equation (68). This includes the detection threshold itself.

In the absence of a separated root, $\|L_{\lambda_{+,\alpha}} u_\alpha\| = o_p(1)$. Let E_ϵ be its spectral projector on $[-\epsilon, \epsilon]$. Then

$$|U_A^{\mathrm{T}} u_\alpha| \le \|E_\epsilon U_A\| + \epsilon^{-1}\|L_{\lambda_{+,\alpha}} u_\alpha\|.$$

First take the proportional limit, and then $\epsilon \downarrow 0$. Weak spectral convergence and the absence of an edge atom prove zero overlap. This proves the scalar specialization of Theorem 5.1. The formula indeed lies in $[0, 1]$: its denominator minus numerator is $J_2/J^2 - 1 + (\theta_A/\theta_B)\omega \ge 0$. Bounded loss also gives convergence of its expectation. □

For numerical evaluation, φ itself parameterizes the upper branch. Solve

$$\int \frac{\mu(dx)}{\lambda x^\alpha - \frac{b-1}{b}x - \theta_B\varphi} = \frac{1 - 1/\varphi}{c\theta_B}$$

for λ, starting at $\varphi = 1$, and stop at $\omega = 1$. The stable branch obeys $1 < \varphi < 1 + \sqrt{c}$, because $I_2 \ge (\int R\, d\mu)^2$ implies $\omega \ge (\varphi - 1)^2/c$. The scalar left-hand side is strictly decreasing in λ. Use equation (68) to locate all signal roots on this branch and retain the largest admissible λ.

Risk at special powers. The powers zero and one reduce to a common linear pencil. For a scalar coefficient w, define $L(w) = \widehat{M}_{\mathrm{av}} + w\widehat{S}$, and let λ_{lin} denote its separated top signal eigenvalue. Write Q for the limiting normalized trace of $(zI - U_B^{\mathrm{T}} L(w) U_B)^{-1}$, with z above the background spectrum, and put $\pi_{\mathrm{av}} = 1 - c\theta_B Q$ and $\pi_{\mathrm{ct}} = 1 - c_b w s_B Q$. The corresponding eigenvalue location $z(Q)$ and signal equation $\Phi(Q)$ satisfy

$$z(Q) = Q^{-1} + \frac{\theta_B}{\pi_{\mathrm{av}}} + \frac{w s_B}{\pi_{\mathrm{ct}}}, \qquad \Phi(Q) = z(Q) - \frac{\theta_A}{\pi_{\mathrm{av}}} - \frac{w\sigma^2\chi^2}{\pi_{\mathrm{ct}}}. \tag{71}$$

An upper exterior zero of Φ, on the branch $z_Q < 0$, gives $\lambda_{\mathrm{lin}} \to z(Q)$. Its limiting squared signal overlap is

$$\mathfrak{a} = \frac{z_Q}{\Phi_Q}, \qquad \Phi_Q = z_Q - \frac{c\theta_A\theta_B}{\pi_{\mathrm{av}}^2} - \frac{c_b w^2\sigma^2\chi^2 s_B}{\pi_{\mathrm{ct}}^2}.$$

Here subscripts denote derivatives with respect to Q. The independent Gaussian signal rows in equation (41) give equation (71) by the same Schur-complement calculation as in lemma E.1. Differentiating its signal equation with respect to the eigenvalue location gives the residue $\mathfrak{a} = (\Phi_Q / z_Q)^{-1}$.

At $\alpha = 0$, take $w = (b-1)/b$; the preceding equations give the pooled-PCA overlap $\mathfrak{a}_0$. At $\alpha = 1$, take $w = -\lambda_{\mathrm{av}}$, where λ_{av} is the largest generalized eigenvalue of $(\widehat{M}_{\mathrm{av}}, \widehat{S})$. Then $\lambda_{\mathrm{lin}} = 0$, and the common signal equation simplifies as follows. Define

$$\theta = \frac{\theta_A s_B}{\theta_B \sigma^2\chi^2}, \quad \theta_+ = \frac{1 + \sqrt{c + c_b - cc_b}}{1 - c_b}, \quad \Delta_F = (\theta - 1)^2 - c - c_b\theta^2.$$

Substitution in equation (71) gives

$$\mathfrak{a}_1 = \begin{cases} \dfrac{\Delta_F}{\Delta_F + (\sigma^2\chi^2/s_B)(c\theta + c_b\theta^2)}, & \theta > \theta_+, \\ 0, & \theta \le \theta_+. \end{cases} \tag{72}$$

The eigenvalue threshold agrees with the spiked Fisher calculation (Wang & Yao, 2017). The $\sigma^2\chi^2/s_B$ factor retains the Euclidean normalization. For the difference endpoint, let $\delta = 1 - \sigma^2\chi^2/s_B$. Its overlap is

$$\mathfrak{a}_{\mathrm{diff}} = \begin{cases} (1 - c_b/\delta^2)/(1 - c_b/\delta), & \delta > \sqrt{c_b}, \\ 0, & \delta \le \sqrt{c_b}. \end{cases} \tag{73}$$

Indeed, this is the smallest-eigenvalue direction of $\widehat{S}$, whose signal mass is w_0 in lemma E.1. Here the sample-wise infinite-power endpoint is taken before the proportional limit, separately from the fixed-power theorem.

Strong changing background. Under the spanning condition $m = d - 1$, the difference-PCA risk is $\mathcal{R}_{\mathrm{diff}} = 1 - \mathfrak{a}_{\mathrm{diff}}$ from (73). Its relative contrast gap $\delta = \tau/(\sigma^2\chi^2 + \tau)$ must exceed $\sqrt{c_b}$ for detection. At fixed $c_b < 1$, $\tau/(\sigma^2\chi^2) \to \infty$ yields $\delta \to 1$ and $\mathcal{R}_{\mathrm{diff}} \to 0$. For $b = 2, c = 1/2, a = 2, \kappa = 0, \sigma = \chi = 1, \tau = 9$, $\mathcal{R}_0 = 1$, $\mathcal{R}_1 = 1/3$, and $\mathcal{R}_{\mathrm{diff}} = 5/36$. Large background variance defeats pooled PCA, whereas the much smaller signal contrast variance makes difference PCA more accurate. The spanning condition matters: contrasts cannot distinguish signal from additional noise-only directions, as the bound below shows.

Reducing only τ to $4/5$ gives the intermediate-power comparison in Section 5.2: $\mathcal{R}_0 = 15/22$, $\mathcal{R}_1 = 25/34$, and $\mathcal{R}_{\mathrm{diff}} = 1$. The scalar recipe evaluates $\mathcal{R}_{1/2} \simeq 0.399$. Here $\delta = 4/9 < \sqrt{1/2}$, so contrasts alone do not detect the signal, but their combination with pooled variance does.

E.4 Elementary checks with additional noise directions

The two finite reference powers admit elementary formulas even when $c_\perp > 0$. They extend the preceding spanning checks and provide benchmarks for the general numerical recipe. Retain v, q, η and $\theta_0, \theta_B, \theta_A$ from above. Write $B_v = b - 1$, $d_0 = v$, $d_B = v + \tau$, $c_0 = c_\perp$, and $c_B = c$. All sums below run over $j = 0, B$; a zero-weight block is omitted, while the signal test direction is retained.

Normalized alignment, $\alpha = 1$. Set $e = \theta_A/v$, $\delta_0 = a$, and $\delta_B = ed_B - \theta_B$. First require $\delta_B > 0$, and define

$$H = \sum_j \frac{c_j\theta_j}{\delta_j}, \quad U = \sum_j \frac{c_j\theta_j^2}{\delta_j^2}, \quad K_1 = \sum_j \frac{c_j}{\delta_j}, \quad V_1 = \sum_j \frac{c_j\theta_j}{\delta_j^2}.$$

Put $L = B_v - \eta - 2H - U$ and $Q = B_v - \eta - H$. The top signal is detected exactly when, in addition, $D_1 = (b-\eta)(1-U) - (1+H)^2 > 0$. This also implies $L > 0$. The factors and risk are then

$$\Gamma_1 = 1 + \frac{\theta_A(K_1+V_1)}{L}, \qquad \Omega_1 = \frac{\theta_A Q\{V_1 Q + K_1(H+U)\}}{LD_1}, \qquad \mathcal{R}_1 = \frac{N_1}{D_1+N_1}, \tag{74}$$

where $N_1 = \theta_A\{(1+H)K_1 + (b-\eta)V_1\}$. Thus δ_B measures the population variance-ratio advantage of signal over background; $D_1 > 0$ requires enough advantage to overcome sampling fluctuations. The formula is rational in the model parameters.

Pooled PCA, $\alpha = 0$. Put $s = \sigma^2(1-\chi^2)$, $A_0 = 1 + c_\perp s/a$, $d_* = a - \kappa - \tau/b$, and $E_0 = A_0 d_* + c(s+\kappa) + q\tau$. If $d_* > 0$, form

$$r = \frac{2qd_*}{E_0 + \sqrt{E_0^2 - 4A_0\tau q d_*}}, \qquad \delta_0 = a, \qquad \delta_B = d_* - \tau r. \tag{75}$$

Require $\delta_B > 0$; a zero gap is inadmissible, even when the quadratic produces a real root. Define

$$A_d = \sum_j \frac{c_j d_j}{\delta_j^2}, \quad A_\theta = \sum_j \frac{c_j\theta_j}{\delta_j^2}, \quad A_{dd} = \sum_j \frac{c_j d_j^2}{\delta_j^2}, \quad A_{\theta\theta} = \sum_j \frac{c_j\theta_j^2}{\delta_j^2}, \quad A_{d\theta} = \sum_j \frac{c_j d_j\theta_j}{\delta_j^2}.$$

Let $L_0 = B_v - r^2 A_{dd}$ and $D_0 = (1 - A_{\theta\theta})L_0 - r^2 A_{d\theta}^2$. Detection additionally requires $L_0 > 0, D_0 > 0$, and then

$$\Gamma_0 = 1 + \frac{r^2 v A_d}{L_0}, \qquad \Omega_0 = \frac{(A_\theta L_0 + r^2 A_d A_{d\theta})(\theta_A L_0 + r^2 v A_{d\theta})}{L_0 D_0}. \tag{76}$$

These expressions use rational operations and one square root. At either power the risk is one if any of its strict detection conditions fails. They are finite-power checks, distinct from the sample-wise infinite-power endpoint.

Derivation of the two finite-power formulas. For $\alpha = 1$, put $l = \lambda - q$. The conditional contrast resolvent has directional values $t_j = (l\psi d_j - \varphi\theta_j + z)^{-1}$, where $\psi^{-1} = 1 + l\sum_j c_j d_j t_j/B_v$ and $\varphi^{-1} = 1 - \sum_j c_j\theta_j t_j$. At the signal root $a\varphi t_0 = 1$, $l\psi/\varphi = e$, hence $t_j = (\varphi\delta_j)^{-1}$, $\varphi = 1+H$, and $\psi = Q/B_v$. Writing $\mu = \sum_j c_j\theta_j t_j$, partial differentiation in z, φ, with the other external variables held fixed, gives

$$\begin{aligned} -t_{0,z}/t_0^2 &= 1 + \theta_A(K_1+V_1)/L, & t_{0,\varphi}/t_0^2 + a &= \theta_A Q/L, \\ -\varphi^2\mu_z &= V_1 + (K_1+V_1)(H+U)/L, & \varphi^2\mu_\varphi &= U + (H+U)^2/L. \end{aligned}$$

Since $D_1 = (1-U)L - (H+U)^2$, substitution in (51) yields (74). On $e > \max_j \theta_j/d_j$, H, U decrease and D_1 increases through zero to $B_v - \eta > 0$, identifying the upper stable branch.

For $\alpha = 0$, use $t_j = (\lambda + z - q\psi d_j - \varphi\theta_j)^{-1}$ and $\psi^{-1} = 1 - q\sum_j c_j d_j t_j/B_v$. The signal root gives $\lambda = q\psi v + \varphi\theta_A$. Putting $r = q\psi/\varphi$, the two equations reduce to $r\{A_0 + c(s+\kappa)/(d_* - \tau r)\} = q$. Its admissible smaller root is (75). The same derivatives now give

$$\begin{aligned} -t_{0,z}/t_0^2 &= 1 + r^2 v A_d/L_0, & t_{0,\varphi}/t_0^2 + a &= \theta_A + r^2 v A_{d\theta}/L_0, \\ -\varphi^2\mu_z &= A_\theta + r^2 A_d A_{d\theta}/L_0, & \varphi^2\mu_\varphi &= A_{\theta\theta} + r^2 A_{d\theta}^2/L_0. \end{aligned}$$

Their stability matrix has diagonal entries $1 - A_{\theta\theta}$ and L_0, and off-diagonal entry $-rA_{d\theta}$. It is positive definite precisely when $L_0, D_0 > 0$. Substitution proves (76) and its detection test. □

E.5 Strong signals and the contrast-only endpoint

We return to the general model with $m \le d-1$. Let ν denote the limiting signal spectral measure of $\widehat{S}$: for $c_\perp > 0$, $\int f\, d\nu = \mathfrak{t}_0\{f(\mathsf{s})\}$; for $c_\perp = 0$, use (60), or $\nu = \mu$ when $\tau = 0$.

Proposition E.2 (Strong signals favor pooled PCA)**.** First take the proportional limit, and then let $a \to \infty$ with the other model parameters fixed. Uniformly on any fixed bounded power interval,

$$\mathfrak{a}_\alpha \longrightarrow \frac{\{\int x^{-\alpha}\nu(dx)\}^2}{\int x^{-2\alpha}\nu(dx)}. \tag{77}$$

Consequently, every risk-minimizing power on a fixed closed interval $[0, \alpha_{\max}]$, $\alpha_{\max} < \infty$, tends to zero. On any fixed finite grid $\mathcal{A} \subset [0,\infty)$ containing zero, pooled PCA is the unique limiting-risk minimizer for sufficiently large a.

Proof of Proposition E.2. To prove equation (77), first take the proportional limit and then let $\theta_A \to \infty$ at fixed $\theta_B, \sigma, \chi, s_B, c, c_\perp, b$. The signal block of $\widehat{M}_{\mathrm{av}}/\theta_A$ converges to one; its off-diagonal block is $O_p({\theta_A}^{-1/2})$, and its signal-orthogonal block is $O_p({\theta_A}^{-1})$ in operator norm. Thus $\widehat{D}/\theta_A$ approaches P_A. The leading direction of $\widehat{S}^{-\alpha/2}\widehat{D}\widehat{S}^{-\alpha/2}$ approaches $\widehat{S}^{-\alpha/2}U_A$, so its original-coordinate loading approaches $\widehat{S}^{-\alpha}U_A$. Taking the two quadratic forms in the signal spectral measure just defined proves equation (77). The compact positive contrast spectrum makes the perturbation bound uniform on $\alpha \in [0, \alpha_{\max}]$ for any fixed $\alpha_{\max} < \infty$. The contrast signal measure is nondegenerate: its variance is $c_b\sigma^2\chi^2\{\sigma^2\chi^2 + \tau c/(c+c_\perp)\} > 0$, as follows from the second Gaussian moment. Thus Cauchy–Schwarz is strict for $\alpha > 0$. The limiting risk is uniquely minimized at zero; compactness and uniform convergence give convergence of the minimizing power to zero. On a finite grid, each positive power has a strictly positive limiting risk as $a \to \infty$, whereas $\mathcal{R}_0 \to 0$. The finite minimum of these positive gaps proves the asserted eventual strict preference for zero. □

A quantitative power choice and its computation. Write $\nu_0 = \nu$ for the signal-direction measure above and ν_B for the background-direction measure. For $c_\perp > 0$, its integrals are $\int f\, d\nu_B = \mathfrak{t}(\mathsf{P}f(\mathsf{s}))/\pi_B$; at $c_\perp = 0$, $\nu_B = \mu$ from the scalar recipe. Set $\mathcal{M}_i(t) = \int x^t \nu_i(dx)$, $i = 0, B$. These are moments of positive compactly supported probability measures, so their derivatives in t exist at every real t. In particular, $\mathcal{M}_0(1) = v$, $\mathcal{M}_i'(0) = \int \log x\, \nu_i(dx)$, and $V_0 = \mathrm{Var}_{\nu_0}(\log x)$. The coefficients used in the main text are

$$V_0 = \mathcal{M}_0''(0) - \mathcal{M}_0'(0)^2, \qquad K = q\{\mathcal{M}_0'(1) - v\mathcal{M}_0'(0)\} + c\theta_B\{\mathcal{M}_B'(0) - \mathcal{M}_0'(0)\}. \tag{78}$$

The two terms in K are the contrast-covariance benefit and the background-separation benefit. The first equals $q\,\mathrm{Cov}_{\nu_0}(x, \log x)$.

Proposition E.3 (Rate of the preferred power)**.** Under the assumptions of Proposition E.2, let V_0, K be as in (78). Then $V_0, K > 0$. For any fixed $0 < \alpha_{\max} < \infty$, the limiting risk has a unique minimizer on $[0, \alpha_{\max}]$ for sufficiently large a, and

$$\alpha_*(a) = \frac{K}{aV_0} + O(a^{-2}), \qquad \mathcal{R}_{\alpha_*(a)} - \mathcal{R}_0 = -\frac{K^2}{a^2 V_0} + O(a^{-3}). \tag{79}$$

For $\tau > 0$, $\partial_\kappa(K/V_0) = c\{\mathcal{M}_B'(0) - \mathcal{M}_0'(0)\}/V_0 > 0$. The limits keep all nuisance parameters fixed after taking the proportional limit.

Thus a small positive power can improve on pooled PCA even at strong signal, but its gain is of order a^{-2}. Near equality with PCA in that regime is consistent with the theory. This comparison concerns the limiting risk and a continuum; a fixed finite grid containing zero eventually selects zero as its risk minimizer.

Proof. Let $f = \mathsf{s}^{-\alpha}$, $h = \mathfrak{t}_0(f)$, $h_2 = \mathfrak{t}_0(f^2)$, and $B_0 = q\mathsf{s} + \mathsf{m}$. Put $d_1 = \mathfrak{t}_0(fB_0f)$ and $d_2 = \mathfrak{t}_0(f^2B_0f + fB_0f^2)/2$. Set $\lambda = aL$ in (48). At $a^{-1} = 0$, its solution is $(L, \varphi) = (h, 1)$ with

nonsingular Jacobian. Expanding the inverse yields

$$\Gamma_\alpha = \frac{h_2}{h^2} + \frac{2}{ah^3}\left(d_2 - \frac{h_2 d_1}{h}\right) + O(a^{-2}), \qquad \Omega_\alpha = \frac{\eta \mathsf{t}(\mathsf{m} f^2)}{ah^2} + O(a^{-2}). \tag{80}$$

The compact positive contrast spectrum makes these analytic expansions uniform with their power derivatives on a fixed bounded interval. The directional compression gives the same expansions at $c_\perp = 0$. Hence $\mathcal{R}_\alpha = \mathcal{R}_{\alpha,\infty} + a^{-1}C_1(\alpha) + O(a^{-2})$, with $\mathcal{R}_{\alpha,\infty} = 1 - h^2/h_2$. Differentiating at zero, using that m acts as θ_0 on the signal-type block, gives

$$C_1'(0) = -2q\,\mathrm{Cov}_{\nu_0}(x, \log x) - 2c\theta_B\{\mathcal{M}_B'(0) - \mathcal{M}_0'(0)\} = -2K, \qquad \mathcal{R}''_{\alpha,\infty}\big|_{\alpha=0} = 2V_0.$$

The directional law is nondegenerate, so $V_0 > 0$ and $\mathrm{Cov}_{\nu_0}(x, \log x) > 0$. The positive-resolvent formula below shows $\mathcal{M}_B'(0) - \mathcal{M}_0'(0) \geq 0$, strictly if $\tau > 0$. Thus $K > 0$. Strict convexity of $t \mapsto \log \mathcal{M}_0(t)$ shows that $\mathcal{R}_{\alpha,\infty}$ is strictly increasing for $\alpha > 0$. All minimizers therefore approach zero, where $\mathcal{R}''_\alpha = 2V_0 + O(\alpha) + O(a^{-1}) > 0$ and $\mathcal{R}'_0 = -2K/a + O(a^{-2}) < 0$. The implicit function theorem gives the unique minimizer and its rate. Integrating this derivative from zero to $\alpha_*(a)$ gives the risk difference in (79). Only θ_B in K depends on κ, proving its stated effect. □

Scalar integrals for the coefficients. The coefficients require no random matrices or signal-root inversion. Let $p_B = c/\eta$, $p_0 = 1 - p_B$, $d_0 = v$, and $d_B = v + \tau$; all sums below are over $i = 0, B$. Define the explicit functions

$$\psi(u) = 1 - y\sum_i \frac{p_i d_i}{u + d_i}, \qquad t(u) = u\psi(u), \qquad t'(u) = 1 - y\sum_i \frac{p_i d_i^2}{(u + d_i)^2}, \quad u > 0. \tag{81}$$

Since $1 - y \leq \psi(u), t'(u) \leq 1$, $t(u)$ increases from zero to infinity. The two directional resolvents satisfy $G_i(t) = \int (x + t)^{-1}\nu_i(dx)$ and $G_i(t(u)) = 1/\{\psi(u)(u + d_i)\}$. Consequently, writing $\mu_i = \mathcal{M}_i'(0)$,

$$\mu_i = d_i \int_0^\infty \frac{\log t(u)}{(u + d_i)^2}\, du, \qquad V_0 = v\int_0^\infty \frac{\{\log t(u)\}^2}{(u + v)^2}\, du - \mu_0^2 - \frac{\pi^2}{3}, \tag{82}$$

$$K = q\int_0^\infty \frac{v\{1 - \psi(u)\}t'(u)}{\psi(u)(u + v)}\, du + c\theta_B \int_0^\infty \left\{\frac{1}{u + v} - \frac{1}{u + v + \tau}\right\}\frac{t'(u)}{\psi(u)}\, du. \tag{83}$$

All integrals converge. Their integrands in K are nonnegative, and the second is positive exactly when $\tau > 0$. The formula remains valid at $p_0 = 0$ by the directional resolvent, which retains the signal-type atom.

For completeness, the covariance resolvent equation gives $\psi^{-1} = 1 + y\sum_i p_i d_i/(t + d_i\psi)$; substitution of $u = t/\psi$ proves (81). The fractional-power integral, followed by integration by parts, gives

$$\mathcal{M}_i(z) = d_i \frac{\sin(\pi z)}{\pi z}\int_0^\infty \frac{t(u)^z}{(u + d_i)^2}\, du, \qquad -1 < \Re z < 1,$$

with its continuous value at zero. Differentiating twice proves (82), including the $-\pi^2/3$ term. The identity $\log x = \int_0^\infty \{(1 + t)^{-1} - (x + t)^{-1}\}\, dt$ gives $\mathrm{Cov}_{\nu_0}(x, \log x) = \int_0^\infty \{(v + t)G_0(t) - 1\}\, dt$ and $\mu_B - \mu_0 = \int_0^\infty (G_0 - G_B)\, dt$, proving (83). Finally, the first two inverse moments are

$$\mathcal{M}_0(-1) = \frac{1}{v(1 - y)}, \qquad \mathcal{M}_0(-2) = \frac{1}{v^2(1 - y)^2} + \frac{y(p_0/v + p_B/(v + \tau))}{v(1 - y)^3}.$$

They follow by evaluating $G_0(0)$ and $-G_0'(0)$; substitution in (77) proves (16).

Table 2 checks the coefficient against the full scalar risk. Here $\tau = 0$ isolates the first term of K: small normalization exploits the contrast sampling fluctuations shared with pooled covariance, even without population background separation. The difference from the predicted coefficient decreases as signal strength grows; the theorem keeps y and the view count fixed.

A contrast-only direction has a separate limitation when $m < d - 1$. For the smallest-eigenvalue unit direction u_{diff} of $\widehat{S}$, put $r = d - 1 - m$. Then

$$\mathbb{E}|U_A^{\mathrm{T}} u_{\mathrm{diff}}|^2 \leq \frac{1}{r + 1}.$$

Consequently its signal loss tends to one whenever $r \to \infty$, including every positive $c_\perp$.

Table 2: A deterministic check of the predicted $1/a$ rate. Entries are $a\widetilde{\alpha}(a)$, where $\widetilde{\alpha}(a)$ is a numerical minimizer on $[0, 0.12]$, using 120-point Gaussian quadrature and golden-section search in the scalar risk equations. Here $b = 2, v = 1, \sigma = \chi = 1, \tau = \kappa = c_\perp = 0$. These are limiting-risk evaluations, not finite-sample simulations or CV results.

y	$a = 25$	$a = 100$	$a = 200$	Limit K/V_0
0.2	0.43898	0.45232	0.45461	0.45694
0.5	0.36976	0.38112	0.38307	0.38503
0.8	0.28346	0.29185	0.29328	0.29472

Proof of the contrast-only bound. The contrast distribution is invariant under every orthogonal rotation within $\mathrm{range}(I - P_B)$, a space of dimension $r + 1$. The smallest sample eigenvalue is simple almost surely. Its sign-independent projector consequently has equal expected diagonal entries in any orthonormal basis of that space. Their sum is at most one, proving the bound. Markov's inequality gives convergence of the squared signal overlap to zero if $r \to \infty$. This argument concerns the sample-wise infinite-power endpoint, not a fixed finite power. □

F Experimental protocols and supporting evidence

F.1 Common protocols and numerical implementation

The three-regime, coverage, and real-image studies use the 20-point grid in (18); its spacing and range are prescribed. The exact difference-PCA limit is a separate real-image comparator, excluded from power selection. An additional sample-size check uses smaller fixed grids to examine individual candidates and source-wise CV. After fixing the selection rule, an independent synthetic batch retains the physical designs, candidate grids, and replication counts. It comprises 216 three-regime trials, 100 coverage trials, 240 sample-size trials, and 432 stress trials, for 988 trials across 55 settings. All candidates and selectors share each trial's data and source splits. Ground-truth subspaces enter only evaluation, and every method receives the target rank.

For principal angles $\theta_1, \ldots, \theta_k \in [0, \pi/2]$ between the fitted and true signal spaces, normalized projector loss is $k^{-1} \sum_j \sin^2 \theta_j$. We also report the squared geodesic distance divided by rank, $k^{-1} \sum_j \theta_j^2$, to distinguish large angular errors more strongly. Monte Carlo standard errors (MCSEs) equal the sample standard deviation across independent trials divided by the square root of the trial count; paired MCSEs apply this calculation to within-trial loss differences.

Synthetic fitting and validation use the model's known zero means. Real-image fits instead subtract the pooled fitting-sample mean, with no mean estimated from evaluation images. All folds and splits keep complete source tuples together. For finite-power fits, each fitting covariance must have its smallest eigenvalue exceed 10^{-10} times its largest eigenvalue. No ridge or support change is introduced. The reciprocal calculation uses an SVD of a Cholesky factor times the inverse half-power, with selected-column normalization. A path shares its moment matrices and common spectral decomposition across powers.

F.2 Proportional risk and source-wise CV

Three-regime comparison. Figures 2 and 3 use the same prescribed rank-one design: three regimes, three view counts, and 24 independent trials per configuration. The parameters are $d = 121, n = 340, k = 1, \sigma = \chi = 1$, with $U_A = e_1$ and isotropic background on its orthogonal complement. Changing background, persistent background, and strong signal have $(a, \kappa, \tau) = (3, 0, 4), (2, 4, 1), (6, 0, 1)$, respectively; each uses $b \in \{2, 8, 16\}$. All 20 powers in (18) are fitted to each dataset. Figure 2 retains the original fixed-power experiment. Figure 3 uses 216 new independent trials generated after the selection rule was fixed; the physical parameters, powers, and replication counts are unchanged. Five-fold source-wise CV applies Section 4.2. Within each training fold, we estimate $\widehat{Q}$ using the known background rank $m = 120$ and fit the complete power path. The covariance profile has $q = 0$, so it uses four variance blocks and only the paired contrast comparison. We average profile and reconstruction scores weighted by held-out source counts, calibrate all pairwise

Table 3: Selection components on the same frozen trials. Entries are mean normalized projector loss; parentheses in Joint–Profile give the paired MCSE. Changed counts selected powers differing from the reference; Sole ref. counts singleton admissible sets. Both counts use all replications in the indicated setting. The two stress rows use $b = 2$. Darker/lighter shading marks the smallest/second-smallest distinct displayed means among Reference, Profile, and Joint in each row; rounding ties share a shade.

Setting	Reference	Profile	Joint	Joint–Profile	Changed	Sole ref.
Changing, $b = 2$	0.2101	0.1851	0.1929	$+0.0078$ (0.0037)	13/24	11/24
Persistent, $b = 2$	0.9145	0.9601	0.9154	-0.0447 (0.0138)	1/24	23/24
Strong signal, $b = 2$	0.9331	0.0701	0.0701	$+0.0000$ (0.0000)	24/24	0/24
T1, $n = 240$	0.7619	0.4355	0.6482	$+0.2127$ (0.0348)	17/40	23/40
T1, $n = 480$	0.7373	0.4035	0.5433	$+0.1398$ (0.0315)	26/40	14/40
T1, $n = 960$	0.7587	0.4023	0.5002	$+0.0979$ (0.0325)	32/40	8/40
Strong, $q = 46$, Gaussian	0.1318	0.0933	0.1318	$+0.0385$ (0.0056)	0/12	12/12
Strong, $q = 0$, t_5	0.7294	0.1190	0.2361	$+0.1171$ (0.0793)	10/12	2/12

contrast differences, and refit the selected power on all sources. Reconstruction-only CV is evaluated on the same candidates and folds as a reference. All views of a source remain together; fitting and validation use the known zero means. Reconstruction scores are evaluated as direct squared errors to avoid cancellation when two projectors nearly coincide. All powers use the same data and source-fold partition. Every trial is included in the reported summaries. The best fixed power minimizes the mean evaluated loss within a setting; this post-hoc benchmark never enters selection.

For Figure 2, the full-data losses at $b = 8$ are paired with RMT evaluations at the same generating parameters and $c = 120/340$. The theoretical curves use only the specified population parameters and do not enter empirical power selection. The display shows $0 \le \alpha \le 4$; CV uses the complete grid through 64. All 327 scalar risk evaluations pass successive-quadrature and branch checks. Each empirical mean uses all 24 trials, with Monte Carlo standard error equal to their sample standard deviation divided by $\sqrt{24}$.

Fresh confirmation and calibration diagnostics. The fixed-power RMT figure (Figure 2) retains its original simulations; the CV figure (Figure 3) uses the 216 newly seeded three-regime confirmation trials. These have the same physical designs, grids, sample sizes, and source-fold protocol, but are different data realizations. The reconstruction reference selects powers from 32 to 64; joint validation can also accept intermediate powers. Across the complete 988-trial, 55-setting confirmation, joint validation has lower, numerically equal, or higher setting-mean loss than reconstruction in $20, 32, 3$ settings, respectively, for each of the projector and geodesic losses. The equality tolerance 10^{-12} is numerical, not an equivalence margin.

For the 888 non-block-action trials with scalar population-moment diagnostics, the stored all-pair simultaneous event holds in $706/888$ trials; the selected population C and E conditions fail in 1 and 0 trials. These pooled counts describe the evaluated CV statistic and do not establish exact or worst-setting coverage. Projector fold loss increases in 1 trial and full-data loss in 6, including 5 with no fold increase. The block-action holdout is excluded from those counts, and its $b = 3$ adverse trial is reported separately. Thus the interval diagnostic, fold loss, and full-data refit loss are distinct quantities.

Profile and comparison components. Table 3 separates covariance-profile minimization over the full grid from minimization over the admissible set. The spanning strong-signal gain comes from the profile; paired screening leaves that choice unchanged. At T1 the same screening often retains the reference and sacrifices part of the profile gain. Across all 216 stress trials with noise-only dimensions, and all 30 real-data settings and splits, the reference is the sole admissible candidate. These identical selections demonstrate reference retention, without identifying an incremental benefit from the profile or the particular calibration. The simultaneous-event counts by experiment family are main regimes: 182/216, sample-size comparison: 197/240, stress: 327/432. For changing background with $b = 8, m = 58$, the stress counts are $1/12$ with Rademacher sources and $4/12$ with Gaussian sources.

Table 4: **Joint validation on the additional sample-size designs.** Mean normalized projector loss over 40 fresh trials per row; parentheses are MCSEs. Ref. is reconstruction CV on the same full-data path. Best fixed minimizes the cell mean after evaluation: $\alpha = 0.5$ for T1 and 2 for T2. The per-trial fitted-grid oracle is a different, ground-truth evaluation benchmark; it happens to choose those same powers in all 240 trials. The last column is a paired difference. Darker/lighter shading marks the smallest/second-smallest distinct displayed means among PCA, Ref., and Joint in each row; rounding ties share a shade.

Configuration	PCA	Ref.	Joint (MCSE)	Best fixed	Joint−best (MCSE)
T1, $n = 240$	0.72075	0.76195	0.64820 (0.03876)	0.40971	+0.23849 (0.03339)
T1, $n = 480$	0.69049	0.73732	0.54329 (0.03350)	0.40348	+0.13981 (0.03154)
T1, $n = 960$	0.70431	0.75870	0.50021 (0.03354)	0.40231	+0.09790 (0.03246)
T2, $n = 240$	0.99988	0.01198	0.01198 (0.00023)	0.01198	+0.00000 (0.00000)
T2, $n = 480$	0.99994	0.01163	0.01163 (0.00020)	0.01163	+0.00000 (0.00000)
T2, $n = 960$	0.99997	0.01161	0.01161 (0.00015)	0.01161	+0.00000 (0.00000)

These small-sample frequencies show substantial undercoverage of the all-pair event, including for Gaussian sources; they are descriptive diagnostics, not estimates of a worst-case guarantee.

Additional sample-size comparison. *Design.* Both families use the orthogonal Gaussian model (13), with $d = m + 1$, $U_A = e_1$, $a = 2$, $\sigma = \chi = 1$, and $m/n = 0.5$. T1 has $(\kappa, \tau, b) = (0, 0.8, 2)$ and $\mathcal{A} = \{0, 0.5, 1\}$; T2 has $(\kappa, \tau, b) = (4, 19, 4)$ and $\mathcal{A} = \{0, 0.5, 1, 2\}$. Each uses $(m, n) \in \{(120, 240), (240, 480), (480, 960)\}$, with 40 independent trials per cell. Data and fold streams are independent across trials. Fitting uses the known zero mean.

Source-wise CV. Five folds keep all views of each source together. Within each training fold, we fit the candidate powers and estimate $\widehat{Q}$ as the orthogonal complement of the leading $m = d - 1$ eigenvectors of the training contrast covariance, using the known background rank. We evaluate (11) directly as a mean squared reconstruction error to avoid cancellation between trace terms. This defines the reconstruction reference. The profile and paired contrast rule in Section 4.2 then selects among the same candidates, with fold-size weights and ties favoring the smallest power. Here $q = 0$, so the mean-capture comparison is absent. We refit the selected power on all sources. No signal subspace or downstream label enters selection. These 240 trials use fresh streams generated after the rule was fixed, retaining the original physical design, candidate grids, and 40 replications per cell.

Risk evaluation. Pooled PCA and fixed normalized alignment use powers zero and one on the same full-data fits. The fitted-grid oracle is the smallest actual subspace loss among all prescribed full-data candidates. The best fixed power minimizes the evaluated mean loss within a cell; both are post-hoc benchmarks. Reported paired gaps compare the actual CV refit with that best fixed power on the same trials. All means and Monte Carlo standard errors use the 40 trials in each cell. The fixed-parameter RMT limits use the known design parameters, quadrature orders $96, 192, 384, 768$, root relative tolerance 2×10^{-13}, quadrature tolerance 2×10^{-6}, and relative branch-separation tolerance 2×10^{-7}. Every recorded limit passes its numerical convergence check; these calculations do not select the empirical CV power.

For T1 at $n = 240, 480, 960$, joint validation selects $\alpha = 0.5$ in $14, 26, 32$ of the 40 trials, respectively; across all 120 trials it selects pooled PCA 3 times and $\alpha = 1$ 45 times. The projector-loss gaps from the best fixed $\alpha = 0.5$ are $+0.23849$ (0.03339), $+0.13981$ (0.03154), $+0.09790$ (0.03246), reported as paired mean (MCSE). The corresponding mean-squared-geodesic gaps are $+0.52824$ (0.08262), $+0.30756$ (0.07774), $+0.25995$ (0.09186). These are also the observed per-trial grid-oracle regrets in this sample, since that oracle selects 0.5 in every T1 trial; neither benchmark is used for selection.

For T2, joint validation selects $\alpha = 2$ in all 120 trials and has zero paired grid-oracle regret in each trial. The sample-size designs and grids are unchanged, but these CV results use the independent confirmation seeds; the fixed-power RMT calculations are retained separately.

Here $\widehat{Q}$ has rank one, so a candidate with $\widehat{P} = \widehat{Q}$ would have zero reconstruction loss. Reconstruction therefore favors agreement with the estimated background complement. The profile uses additional

Table 5: **Normalized projector loss: all 36 stress settings.** Joint means have MCSEs in parentheses. Ref. and PCA are the reconstruction-selected and fixed-zero-power means; best fixed minimizes the mean separately within this setting and this loss after evaluation. Δ_{ref} is joint minus reference, with its paired MCSE. Every row uses all 12 trials. Darker/lighter shading marks the smallest/second-smallest distinct displayed means among Joint, Ref., and PCA in each row; rounding ties share a shade.

Case	b	m	Law	Joint (MCSE)	Ref.	PCA	Best fixed	Δ_{ref} (MCSE)
I	2	12	N	0.29827 (0.01367)	0.29827	0.92023	0.28331	+0.00000 (0.00000)
I	2	12	R	0.31468 (0.00959)	0.31468	0.93090	0.30577	+0.00000 (0.00000)
I	2	12	t_5	0.32979 (0.02504)	0.32979	0.88004	0.30902	+0.00000 (0.00000)
I	2	58	N	0.17098 (0.00873)	0.17044	0.97166	0.16492	+0.00055 (0.00124)
I	2	58	R	0.16904 (0.00769)	0.17158	0.96756	0.16531	−0.00254 (0.00147)
I	2	58	t_5	0.18751 (0.01446)	0.18796	0.96732	0.17641	−0.00044 (0.00115)
I	8	12	N	0.11758 (0.00551)	0.11758	0.91448	0.11163	+0.00000 (0.00000)
I	8	12	R	0.12163 (0.00504)	0.12163	0.91623	0.11310	+0.00000 (0.00000)
I	8	12	t_5	0.13631 (0.00449)	0.13631	0.92543	0.13035	+0.00000 (0.00000)
I	8	58	N	0.01493 (0.00050)	0.01522	0.96419	0.01448	−0.00029 (0.00010)
I	8	58	R	0.01580 (0.00047)	0.01586	0.96267	0.01497	−0.00006 (0.00003)
I	8	58	t_5	0.01538 (0.00034)	0.01567	0.94879	0.01518	−0.00029 (0.00015)
P	2	12	N	0.85039 (0.02095)	0.85039	0.98031	0.84717	+0.00000 (0.00000)
P	2	12	R	0.88270 (0.01337)	0.88270	0.98317	0.85717	+0.00000 (0.00000)
P	2	12	t_5	0.85685 (0.01648)	0.85685	0.97670	0.82553	+0.00000 (0.00000)
P	2	58	N	0.73969 (0.01733)	0.73969	0.98984	0.73969	+0.00000 (0.00000)
P	2	58	R	0.74942 (0.02681)	0.74942	0.99016	0.74942	+0.00000 (0.00000)
P	2	58	t_5	0.74007 (0.02881)	0.74007	0.99253	0.74007	+0.00000 (0.00000)
P	8	12	N	0.56018 (0.02002)	0.56018	0.98065	0.51998	+0.00000 (0.00000)
P	8	12	R	0.51572 (0.01916)	0.51572	0.98353	0.48866	+0.00000 (0.00000)
P	8	12	t_5	0.59093 (0.02311)	0.59093	0.98520	0.56782	+0.00000 (0.00000)
P	8	58	N	0.09679 (0.00272)	0.09679	0.99314	0.09679	+0.00000 (0.00000)
P	8	58	R	0.10484 (0.00304)	0.10484	0.99029	0.10484	+0.00000 (0.00000)
P	8	58	t_5	0.10985 (0.00635)	0.10985	0.99155	0.10985	+0.00000 (0.00000)
S	2	12	N	0.13177 (0.00598)	0.13177	0.09924	0.09277	+0.00000 (0.00000)
S	2	12	R	0.13128 (0.00586)	0.13128	0.09926	0.09257	+0.00000 (0.00000)
S	2	12	t_5	0.13065 (0.00754)	0.13065	0.10393	0.09299	+0.00000 (0.00000)
S	2	58	N	0.11149 (0.00499)	0.73828	0.15854	0.11193	−0.62679 (0.02457)
S	2	58	R	0.10745 (0.00296)	0.73156	0.15602	0.10644	−0.62411 (0.01537)
S	2	58	t_5	0.23612 (0.07964)	0.72943	0.16588	0.11695	−0.49331 (0.06903)
S	8	12	N	0.06191 (0.00249)	0.06191	0.06059	0.05438	+0.00000 (0.00000)
S	8	12	R	0.05958 (0.00340)	0.05958	0.06082	0.05388	+0.00000 (0.00000)
S	8	12	t_5	0.05819 (0.00179)	0.05819	0.05679	0.05180	+0.00000 (0.00000)
S	8	58	N	0.03922 (0.00182)	0.10625	0.08527	0.03931	−0.06702 (0.00320)
S	8	58	R	0.03851 (0.00124)	0.09908	0.08650	0.03820	−0.06057 (0.00325)
S	8	58	t_5	0.03960 (0.00107)	0.10673	0.09346	0.03960	−0.06713 (0.00420)

covariance information, but paired comparisons can still retain the reference when the improvement is too small relative to its sampling uncertainty.

Noise-only dimensions and non-Gaussian sources. All 36 settings use $d = 61, n = 180, k = 3$, $m \in \{12, 58\}$, $b \in \{2, 8\}$, the original 20-power grid, and five-fold source-wise joint validation followed by a full-data refit. Each setting has 12 independent trials; the shared and view-specific noise variances are 0.75 and 1. I, P, and S denote $(a, \kappa, \tau) = (3, 0, 4), (2, 4, 1), (6, 0, 1)$, respectively. N, t_5, and R denote normal, variance-standardized Student t_5, and Rademacher draws. Table 5 retains all 36 settings. For all 216 trials with $m = 12$, hence $q = d - k - m > 0$, joint validation retains the reference power; agreement with the reference does not imply agreement with the best fixed power.

In the spanning strong-signal t_5 setting with $b = 2$, joint validation has projector mean loss 0.23612, compared with 0.72943 for the reference, 0.16588 for PCA, and 0.11695 for the best fixed power. Its paired gap from that fixed power is +0.11917 (0.08019). The corresponding geodesic means are 0.34981, 1.19352, 0.17747, and 0.12242, respectively, again improving on the reference but

exceeding PCA. Across all 36 settings, the two losses give the same direction of mean-loss comparison between joint validation and either reconstruction or PCA. All trials, including unfavorable realizations, are retained. Table R1 of the accompanying Reproducibility Addendum reports the complete geodesic results; its data files retain all trial and fixed-power losses.

Interpreting the selection comparison. In the strong-signal, two-view setting of Figure 3, the paired loss difference between joint CV and pooled PCA is -0.02836, with MCSE 0.00177. Here $d - m = k = 1$, so projection onto the estimated background complement has zero reconstruction error, even when that space estimates the signal poorly. Profile minimization alone makes the same choices as the joint rule in this setting (Table 3). The full stress study reports projector loss and geodesic-loss sensitivity; complete geodesic results appear in Table R1 of the accompanying Reproducibility Addendum.

F.3 Independent holdout coverage study

The block-action construction has $k = 2, m = 6, d = 8$. Channels two through four each change two new background coordinates; the fifth repeats the reference action with independent observation noise. Signal, persistent-background, action, and observation-noise variances are 0.5, 4, 1, and 1, respectively. For each $b \in \{2, 3, 4, 5\}$, we generate 5,000 sources, randomly assign 4,000 to fitting and 1,000 to validation, and repeat 25 times. Validation applies Section 4.2 on $\mathcal{A}_{20}$, using $m = 6$ to estimate $\widehat{Q}$ from the fitting sources, and returns the selected fitting-sample model. Here $q = 0$, so the paired rule compares only contrast energy. The profile is a working criterion for these fixed, nonexchangeable channel actions. Data and split streams are independent across channel counts and replications. The reproduction materials retain every channel count, candidate loss, selected power, and random-stream identifier. For $b = 2, 3, 4, 5$, jointly calibrated holdout has mean normalized projector losses (MCSE) 0.73962 (0.03072), 0.66252 (0.03285), 0.00114 (0.00013), 0.00091 (0.00008), respectively, over 25 trials at each channel count. The reference reconstruction means are 0.73962, 0.64187, 0.00114, 0.00091; fixed $\alpha = 64$ gives the same four means. Pooled PCA and fixed $\alpha = 1$ each have mean loss above 0.998 at every channel count. Mean squared geodesic losses (MCSE) for joint holdout are 1.31744 (0.07641), 1.17239 (0.07447), 0.00114 (0.00013), 0.00091 (0.00008), in the same order.

At $b = 3$, one of the 25 trials increases projector loss by 0.51643 relative to the reference; the paired mean increase is $+0.02066$ (0.02066). This incomplete-coverage case is retained along with the sharp improvement at full coverage, $b = 4$. The block-action construction is outside the scalar-model conditional-moment diagnostic for the joint rule.

At $b = 2, 3$, the changing directions do not span the background; knowing its rank does not identify its full space. These cases illustrate incomplete coverage. Full coverage first holds at $b = 4$.

F.4 Linear real-image protocols

Data, channels, and splitting. We retain the official Optdigits training and writer-independent test files (Alpaydin & Kaynak, 1998). Divide their 64 integer image intensities by 16, without estimating a transform from the test set. For each of five seeds, an unstratified permutation of the 3,823 training sources assigns 3,000 to fitting and 823 to validation. Every source's four views remain in the same split. The official test set is held fixed at 1,797 sources. The 20-point evaluation reuses the earlier source splits and augmentation draws; it is a paired rerun on an existing benchmark, rather than a new independent confirmation.

Let $u_1, u_2, u_3 \in \mathbb{R}^{64}$ denote the constant image, a horizontal linear gradient, and a vertical linear gradient. The gradients have eight equally spaced levels from -1 to 1; normalize each pattern to unit root-mean-square pixel intensity. For a source image x_i, draw $c_i \sim N(0, 0.35^2 I_3)$ and use

$$x_i^{(e)} = x_i + \sum_{j=1}^{3} a_{ej} c_{ij} u_j + 0.08\, \xi_i^{(e)}, \qquad (a_{ej}) = \begin{pmatrix} 1 & 1 & 1 \\ -1 & 1 & 1 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \end{pmatrix}, \tag{84}$$

where $\xi_i^{(e)} \sim N(0, I_{64})$ are independent over sources and views and independent of c_i. Thus the four specified channels change complementary illumination components of the same image. The

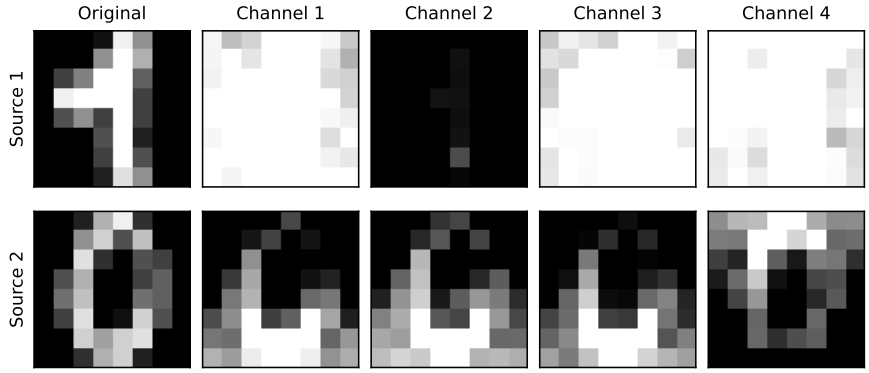


Figure 4: **The fixed illumination portfolio on real images.** The first two fitting sources under the first seed, followed by their four augmented views. Only the display is clipped to $[0, 1]$; fitting and evaluation use the unclipped intensities.

numerical inputs are not clipped. Test augmentations use seed $s + 900000$ when the fitting seed is s; after generating the four views, choose one channel uniformly and independently for each test source. The same test images and channel choices are used for every method within that seed. No model or augmentation parameter is changed in response to test outcomes.

Pixel-space fits. Both real-image protocols use $k = 8$, a single pooled training mean, and (9). Source-wise holdout applies Section 4.2 on the 20-power grid, breaking exact ties toward the smaller power and retaining the selected training fit. The exact difference-PCA comparator uses the eight smallest eigenvectors of $\widehat{S}$ and is not a selection candidate. All 100 finite-power fits in the first protocol pass the same unregularized positive-definite support and normalization checks as the synthetic studies.

Source-wise holdout reuses these finite-power fits and their training means. It takes the known background rank $m = 3$ and sets $\widehat{Q}$ to the complement of the leading three eigenvectors of the training contrast covariance. The reconstruction reference minimizes (11), divided by 64 to express mean squared error per pixel. With $q = 53$, the joint rule also uses the bottom 53 training eigenvectors of $\widehat{D}$ for the common noise probe. Each candidate's covariance profile uses a separately fitted training background space, followed by held-out variance minimization under the same order constraints as in the synthetic studies. These constraints define a working criterion for the fixed illumination channels. Fixed Q-PCA uses the same known background rank and the leading eight eigenvectors of $\widehat{Q}\widehat{D}\widehat{Q}$. It implements PCA within the estimated background complement and is distinct from the infinite-power limit of the sample contrast covariance.

All 600 held-out covariance profiles are finite, and all 1,200 training background initializations converge. Direct source-level replay verifies the block moments, paired statistics, and selected powers. All 30 sets of choices (five original splits and five splits at each of five sharing levels) are frozen before the cached classifier predictions or labels are read. No representation or classifier is refitted for this score comparison. The protocol is exploratory because these images and their earlier outcomes had already been examined.

Common downstream evaluation and uncertainty. Only after all representations and selections have been frozen do we use digit labels. For every candidate, calibrate its eight coordinates to zero mean and identity covariance using fitting views only, then fit a ten-output ridge least-squares classifier to their one-hot digit labels. The objective is mean squared error plus 10^{-3} times the squared Frobenius norm of the coefficient matrix; the intercept is unpenalized. All four fitting views carry their source's label. Classification chooses the largest predicted component. No validation labels or test labels enter fitting or power selection. Original and augmented test images use the same frozen classifier.

Table 6: **All shared-background settings.** Augmented-image diagnostic accuracy (%), mean ± sample standard deviation over five paired splits. Source-wise holdout selects among the 20 finite powers. Source-wise holdout and fixed Q-PCA use the known background rank $m = 3$. All five sharing levels are retained; every comparison reuses the previously examined images, fits, and classifier predictions. Higher is better. Darker/lighter shading marks the best/second-best distinct displayed means within each row; rounding ties share a shade.

ρ	Fixed $\alpha = 1$	Source-wise holdout	Fixed Q-PCA
0	86.72 ± 3.01	87.64 ± 1.64	87.94 ± 1.25
0.9	86.70 ± 2.84	86.70 ± 2.84	87.82 ± 1.44
0.99	72.12 ± 3.12	86.04 ± 2.48	87.64 ± 1.51
0.999	72.16 ± 3.27	77.30 ± 2.95	82.20 ± 1.89
0.9999	72.06 ± 3.44	71.04 ± 2.32	72.14 ± 3.15

The reproduction materials retain all 20 powers. Means and sample standard deviations describe five randomizations of this benchmark. Source-wise holdout is evaluated through the cached classifier of its selected representation. The real images have no known rank-eight signal loading space, so classification measures representation utility rather than subspace recovery.

Shared background. This exploratory extension varies the persistence of the illumination background while retaining the real digit images. Use the same patterns u_j, signs a_{ej}, and pixel scaling as in (84), but replace its illumination by

$$x_i^{(e)} = x_i + 0.35 \sum_{j=1}^{3} \big(\sqrt{\rho}\, g_{ij} + \sqrt{1-\rho}\, a_{ej} h_{ij}\big) u_j + 0.08\, \xi_i^{(e)}, \tag{85}$$

where $g_i, h_i \sim N(0, I_3)$ are independent across sources and of each other and the independent pixel noises $\xi_i^{(e)} \sim N(0, I_{64})$. Both coefficient vectors are shared by a source's views, with only h_i acted on by the channel signs. The parameter ρ is the fraction of the added illumination variance shared unchanged across views. Each view's population illumination covariance stays fixed, while its contribution to the contrast covariance is multiplied by $1 - \rho$. No inputs are clipped. Thus the scan changes background persistence without increasing its marginal strength.

The five sharing levels $\rho \in \{0, 0.9, 0.99, 0.999, 0.9999\}$ and five seeds are fixed before execution. Each seed randomly partitions the 3,823 official training sources into 2,000 for fitting, 823 for validation, and 1,000 for diagnostic evaluation. The official test file is unused. Within a seed, every sharing level uses identical source assignments, Gaussian coefficient draws, pixel noises, and diagnostic channel choices. Each diagnostic image is evaluated in one uniformly drawn view and, as a companion, in its original form. Fitting and downstream classification follow the common protocol above. Source-wise holdout selects among 20 candidates for each sharing level and split. The 25 combinations reuse 500 finite-power fits in total, with all choices frozen before outcome evaluation. All 500 finite-power fits and their evaluations complete without numerical failure.

The reconstruction reference is the sole admissible candidate in all 30 comparisons, giving identical finite-path selections and outcomes. On the prescribed finite grid, source-wise holdout selects $\alpha = 1/\sqrt{2}$ under the original protocol and at $\rho = 0$, $\alpha = 1$ at $\rho = 0.9$, $\alpha = 2$ at $\rho = 0.99$, and $\alpha = 4\sqrt{2}$ at $\rho = 0.999$, consistently across the five splits in each setting. At $\rho = 0.9999$ its selections are $4, 2\sqrt{2}, 2\sqrt{2}, 2\sqrt{2}, 4$.

A reconstruction-only check with Q-PCA added to the 20-power grid selects Q-PCA in all 30 comparisons, reproducing the fixed Q-PCA outcomes. Table 6 reports the shared-background settings. At $\rho = 0.9999$, finite-grid source-wise holdout attains 71.04% and Q-PCA 72.14%, close to fixed $\alpha = 1$ at 72.06%. Here the small changing-background component makes its space difficult to estimate. The validation target and Q-PCA share $\widehat{Q}$; their agreement cannot independently establish that this estimated background space is correct.

Both real-image protocols reuse their respective historical source splits and augmentation draws. The selected powers depend on the prescribed grid; all sharing levels and candidate outcomes are retained in the reproduction materials, together with the random-stream identifiers.

The population threshold in Theorem C.1 motivates this scan: weaker exposure relative to persistent background can require a larger power. The real-image study examines this mechanism and the validation target under controlled additive backgrounds; it is not an independent test of naturally occurring distribution shifts.