Bias-Corrected Subspace Intersection: Minimax-Optimal Shared Subspace Estimation in Multi-View Data

Xianwen Song[1], Yuepeng Yang[2], and Cong Ma[1]

[1]Department of Statistics, University of Chicago
[2]Department of Statistics and Data Science, the Wharton School, University of Pennsylvania

September 4, 2026

Abstract

Estimating a low-dimensional subspace shared across noisy data matrices is a fundamental problem in multi-view matrix estimation. We study this problem under the two-view JIVE model, where each data matrix contains shared and view-specific low-rank components. We demonstrate that standard plug-in subspace intersection, including AJIVE, suffers from a second-order bias caused by direction-dependent leakage of the empirical singular vectors. We propose *bias-corrected subspace intersection* (BCSI), which removes this bias before estimating the shared subspace. We establish finite-sample risk bounds for BCSI that accommodate unequal view dimensions, signal strengths, and view-specific ranks and require no condition-number assumptions on the signal matrices. When the shared and view-specific ranks are comparable, these bounds match our minimax lower bounds up to universal constants. The resulting minimax rate contains a new second-order term, arising from quadratic leakage perturbations relative to the shrinking spectral gap when the view-specific subspaces are nearly aligned. This term is absent from previous JIVE minimax lower bounds. Numerical experiments demonstrate the advantage of BCSI over AJIVE when the leakage bias is pronounced. Along the way, we establish a nonasymptotic concentration result for the bias-corrected leakage Gram matrix of a rectangular spiked matrix, which may be of independent interest.

1 Introduction

Modern data sets often measure the same collection of samples through multiple data modalities, including different genomic assays (Lock et al., 2013), imaging technologies (Zhao et al., 2021), and other heterogeneous sources (Shi et al., 2023). A central goal of multi-view data integration is to separate variation that is shared across views from variation that is specific to an individual view (Lock et al., 2013; Feng et al., 2018; Gaynanova and Li, 2019; Prothero et al., 2024; Gui et al., 2025).

In this paper, we study the statistical problem of estimating the shared subspace in the two-view setting. For $k = 1, 2$, we observe

$$\boldsymbol{Y}_k = \boldsymbol{X}_k + \boldsymbol{E}_k \in \mathbb{R}^{n \times d_k}, \tag{1}$$

The upper- and lower-bound proofs have also been formally verified in Lean 4; see `https://github.com/congma1028/bcsi-lean-verification`.

where $\boldsymbol{X}_k$ is a rank-p_k signal matrix and the entries of $\boldsymbol{E}_k$ are independent $\mathcal{N}(0,\sigma^2)$ random variables. Let $\mathcal{M}_k \coloneqq \mathrm{col}(\boldsymbol{X}_k)$ denote the population signal space of view k. Our goal is to estimate the rank-r shared subspace

$$\mathcal{U} \coloneqq \mathcal{M}_1 \cap \mathcal{M}_2 = \mathrm{col}(\boldsymbol{X}_1) \cap \mathrm{col}(\boldsymbol{X}_2) \tag{2}$$

from the observations $(\boldsymbol{Y}_1, \boldsymbol{Y}_2)$. This is the two-view shared-subspace estimation problem underlying the joint and individual variation explained (JIVE) framework (Lock et al., 2013).

1.1 Bias of plug-in subspace intersection

A natural plug-in strategy is to estimate the two signal spaces separately and then look for directions that are approximately shared by the two estimated spaces. Let $\widehat{\boldsymbol{\Phi}}_k \in \mathbb{R}^{n\times p_k}$ contain the leading p_k left singular vectors of $\boldsymbol{Y}_k$, so that $\mathrm{col}(\widehat{\boldsymbol{\Phi}}_k)$ estimates $\mathcal{M}_k$. The right singular vectors associated with the smallest singular values of $[\widehat{\boldsymbol{\Phi}}_1, \widehat{\boldsymbol{\Phi}}_2]$ identify pairs of directions, one from each estimated signal space, that are nearly the same up to sign. These directions therefore approximate the intersection of the two estimated signal spaces. In the two-view setting, this procedure is equivalent to AJIVE (Feng et al., 2018; Yang and Ma, 2025; Sergazinov et al., 2026); see Section A for a proof.

The problem with this plug-in estimator is that the empirical singular vectors leak outside their population signal spaces. This leakage creates a systematic second-order bias and can make the estimator suboptimal for shared subspace estimation. To see this, for any subspace $\mathcal{S} \subseteq \mathbb{R}^n$, let $\boldsymbol{P}_{\mathcal{S}}$ denote the orthogonal projector onto $\mathcal{S}$, and decompose

$$\widehat{\boldsymbol{\Phi}}_k = \widehat{\boldsymbol{S}}_k + \widehat{\boldsymbol{L}}_k, \qquad \text{where} \qquad \widehat{\boldsymbol{S}}_k \coloneqq \boldsymbol{P}_{\mathcal{M}_k}\widehat{\boldsymbol{\Phi}}_k, \quad \widehat{\boldsymbol{L}}_k \coloneqq \boldsymbol{P}_{\mathcal{M}_k^\perp}\widehat{\boldsymbol{\Phi}}_k.$$

The approximate intersection is determined by the bottom eigenspace of the Gram matrix of the concatenated empirical bases $[\widehat{\boldsymbol{\Phi}}_1, \widehat{\boldsymbol{\Phi}}_2]$. With $\widehat{\boldsymbol{S}} \coloneqq [\widehat{\boldsymbol{S}}_1, \widehat{\boldsymbol{S}}_2]$ and $\widehat{\boldsymbol{L}} \coloneqq [\widehat{\boldsymbol{L}}_1, \widehat{\boldsymbol{L}}_2]$, this Gram matrix decomposes as

$$[\widehat{\boldsymbol{\Phi}}_1, \widehat{\boldsymbol{\Phi}}_2]^\top [\widehat{\boldsymbol{\Phi}}_1, \widehat{\boldsymbol{\Phi}}_2] = \widehat{\boldsymbol{S}}^\top \widehat{\boldsymbol{S}} + \widehat{\boldsymbol{S}}^\top \widehat{\boldsymbol{L}} + \widehat{\boldsymbol{L}}^\top \widehat{\boldsymbol{S}} + \widehat{\boldsymbol{L}}^\top \widehat{\boldsymbol{L}}. \tag{3}$$

If there were no leakage and $\mathrm{col}(\widehat{\boldsymbol{S}}_k) = \mathcal{M}_k$ for $k = 1, 2$, then $\widehat{\boldsymbol{S}}^\top \widehat{\boldsymbol{S}}$ would have an r-dimensional kernel consisting exactly of the linear relations that produce directions in the shared subspace $\mathcal{U}$.

In the noisy case, the two cross terms in (3) are linear in the leakage, while $\widehat{\boldsymbol{L}}^\top \widehat{\boldsymbol{L}}$ is quadratic. Crucially, this quadratic term has a direction-dependent bias and can therefore perturb the nullspace of $\widehat{\boldsymbol{S}}^\top \widehat{\boldsymbol{S}}$. To see why the bias is direction dependent, consider the diagonal entries of $\widehat{\boldsymbol{L}}_k^\top \widehat{\boldsymbol{L}}_k$. Let $(\widehat{s}_{k,i}, \widehat{\boldsymbol{\phi}}_{k,i}, \widehat{\boldsymbol{\psi}}_{k,i})$ be the ith empirical singular triplet and set $a_k \coloneqq n - p_k$ and $b_k \coloneqq d_k - p_k$. The singular-vector relation $\boldsymbol{Y}_k \widehat{\boldsymbol{\psi}}_{k,i} = \widehat{s}_{k,i} \widehat{\boldsymbol{\phi}}_{k,i}$, together with $\boldsymbol{P}_{\mathcal{M}_k^\perp} \boldsymbol{X}_k = \boldsymbol{0}$, implies

$$(\widehat{\boldsymbol{L}}_k^\top \widehat{\boldsymbol{L}}_k)_{ii} = \left\| \boldsymbol{P}_{\mathcal{M}_k^\perp} \widehat{\boldsymbol{\phi}}_{k,i} \right\|^2 = \frac{\left\| \boldsymbol{P}_{\mathcal{M}_k^\perp} \boldsymbol{E}_k \widehat{\boldsymbol{\psi}}_{k,i} \right\|^2}{\widehat{s}_{k,i}^2} \approx \frac{a_k \sigma^2}{\widehat{s}_{k,i}^2}. \tag{4}$$

The last relation is heuristic because $\widehat{\boldsymbol{\psi}}_{k,i}$ depends on the noise. It nevertheless shows that singular directions associated with smaller empirical singular values have larger leakage. Thus the leakage Gram is generally anisotropic, and this anisotropy biases the estimated intersection.

1.2 Bias correction and main results

The heuristic calculation in (4) suggests a natural bias correction: rescale the empirical singular directions within each view so that their leakage Gram is approximately isotropic. Specifically, replace $\widehat{\boldsymbol{\Phi}}_k$ by $\widehat{\boldsymbol{\Phi}}_k \operatorname{diag}(w_{k,1}, \ldots, w_{k,p_k})$. At the heuristic level, this would give weights roughly proportional to $\widehat{s}_{k,i}$. However, the empirical singular vectors themselves depend on the noise, so this heuristic does not give the correct weight. Rectangular spiked-matrix theory gives a more accurate expression for the leakage.

Let $s_{k,i}$ denote the ith population singular value in view k. Classical rectangular spiked-matrix theory (Benaych-Georges and Nadakuditi, 2012; Gavish and Donoho, 2017) predicts

$$\left\|\boldsymbol{P}_{\mathcal{M}_k^\perp} \widehat{\boldsymbol{\phi}}_{k,i}\right\|^2 \approx \frac{a_k \sigma^2 (s_{k,i}^2 + b_k \sigma^2)}{s_{k,i}^2 (s_{k,i}^2 + a_k \sigma^2)}. \tag{5}$$

If $s_{k,i}$ were known, multiplying $\widehat{\boldsymbol{\phi}}_{k,i}$ by $\{s_{k,i}^2 (s_{k,i}^2 + a_k\sigma^2)/(s_{k,i}^2 + b_k\sigma^2)\}^{1/2}$ would make its squared leakage approximately $a_k\sigma^2$ for every i.

This oracle rescaling is not directly implementable because the population singular value $s_{k,i}$ is unknown. The same spiked-matrix theory provides the missing link between the population and empirical singular values:

$$\widehat{s}_{k,i}^2 \approx \frac{(s_{k,i}^2 + a_k\sigma^2)(s_{k,i}^2 + b_k\sigma^2)}{s_{k,i}^2}. \tag{6}$$

We therefore invert (6) to estimate $s_{k,i}^2$ from $\widehat{s}_{k,i}$ and substitute this estimate into the oracle rescaling above.

This leads to our proposed estimator, *bias-corrected subspace intersection* (BCSI). BCSI modifies plug-in subspace intersection in three steps. First, it reweights the empirical singular directions within each view so that their leakage Gram is approximately $a_k\sigma^2 \boldsymbol{I}_{p_k}$. Second, it subtracts this common leakage level from the corresponding diagonal block of the Gram matrix. Finally, it extracts the near-zero eigenspace of the corrected Gram matrix and uses it to reconstruct the shared subspace. We denote the corresponding projector by $\widehat{\boldsymbol{P}}_{\mathrm{BCSI}}$ and give the complete construction in Section 2.2.

Our main results show that BCSI achieves the minimax rate for shared-subspace estimation over a broad range of signal strengths and subspace geometries. For $k = 1, 2$, let s_k be a known lower bound on the smallest nonzero singular value of $\boldsymbol{X}_k$ and define the corresponding effective signal strength by

$$\gamma_k := \frac{s_k^2}{\sqrt{s_k^2 + d_k\sigma^2}}. \tag{7}$$

Also let $\gamma_{\max} := \max\{\gamma_1, \gamma_2\}$, and $\gamma_{\min} := \min\{\gamma_1, \gamma_2\}$. The effective strength γ_k accounts for the column dimension d_k. For column-space estimation, the raw singular value s_k alone does not capture the effect of the d_k noisy columns, so signals with the same s_k can have different estimation accuracy when their aspect ratios differ (Cai and Zhang, 2018; Cai et al., 2021). In particular, γ_k is of order s_k for a strong signal and of order $s_k^2/(\sigma\sqrt{d_k})$ for a weaker one.

Suppose that the shared and view-specific subspaces all have rank of order r. Let $\theta \in (0, 1/2]$ measure the separation between two view-specific subspaces, with smaller θ corresponding to more closely aligned subspaces. Under the signal strength assumption $\gamma_{\min} \gtrsim \sigma\sqrt{n}$, BCSI satisfies

$$\mathbb{E}\left\|\widehat{\boldsymbol{P}}_{\mathrm{BCSI}} - \boldsymbol{P}_{\mathcal{U}}\right\|_{\mathrm{op}} \lesssim 1 \wedge \left\{\frac{\sigma\sqrt{n}}{\gamma_{\max}} + \frac{\sigma\sqrt{r}}{\gamma_{\min}\sqrt{\theta}} + \frac{\sigma^2\sqrt{nr}}{\gamma_{\min}^2\theta}\right\}. \tag{8}$$

A matching minimax lower bound shows that all three terms in (8) are unavoidable up to universal constants.

The first term is the usual cost of estimating the shared subspace when the view-specific directions are known. Since both views contain information about the shared directions, their effective strengths combine, yielding $\sqrt{\gamma_1^2 + \gamma_2^2} \asymp \gamma_{\max}$. The remaining two terms reflect the additional difficulty of identifying which directions are shared when the view-specific subspaces are unknown. This identification becomes harder as the two view-specific subspaces approach one another, leading to the factors $\theta^{-1/2}$ and θ^{-1}. Moreover, these terms are governed by $\gamma_{\min}$: even if the stronger view accurately localizes the relevant signal space, the weaker view is still needed to distinguish the shared directions from the view-specific ones. The final θ^{-1} term is one of our main contributions. It is second order in the noise level and shows that nearly aligned view-specific subspaces create an additional statistical difficulty beyond the usual first-order subspace estimation error. In the balanced-rank regime, the corresponding quadratic term in the existing AJIVE bound is of order $\sigma^2 n/(\gamma_{\min}^2 \theta)$, whereas BCSI improves it to $\sigma^2\sqrt{nr}/(\gamma_{\min}^2\theta)$, a factor $\sqrt{n/r}$ smaller. Our minimax lower bound further shows that the latter θ^{-1} term is unavoidable, and hence is intrinsic to shared-subspace estimation rather than an artifact of the analysis of BCSI. The general results, allowing unequal shared and view-specific ranks, are stated in Sections 3.1 and 3.2.

1.3 Related work

Joint and individual variation. The JIVE framework (Lock et al., 2013; Zhou et al., 2015) models multi-view data through shared and view-specific low-rank components, and has various variants including AJIVE (Feng et al., 2018), CJIVE (Murden et al., 2022), and the probabilistic formulation ProJIVE (Murden et al., 2026). Related multiblock decompositions such as SLIDE (Gaynanova and Li, 2019), hierarchical nuclear-norm penalization (Yi et al., 2023), and DIVAS (Prothero et al., 2024) also accommodate partially shared structure. Among these methods, AJIVE is most closely related to the present work: it first estimates the signal space of each view and then identifies shared directions through the approximate intersection of the estimated spaces. Similarly, Sergazinov et al. (2026) use the product-of-projections method to combine the estimates across different views.

Theory for JIVE. The closest statistical analyses are Yang and Ma (2025) and Li and Lyu (2025). The former establishes finite-sample guarantees and minimax lower bounds for AJIVE in a homogeneous multi-view model, while the latter allows heterogeneous dimensions and signal strengths and studies strength-aware aggregation of the estimated view projectors. Our results allow unequal view dimensions, strengths, and view-specific ranks, impose no condition-number restriction on the loading matrices, and identify an additional θ^{-1} term in the minimax rate arising from the interaction between estimation error and the geometry of the intersection.

Shared-subspace estimation. Related work considers shared-subspace estimation under alternative multi-matrix models. Ma and Ma (2026) analyze complete and partial sharing when the shared and unshared directions are themselves singular vectors of the individual signal matrices, while the partially shared subspace model (Yoon et al., 2026) assumes that the shared directions are eigenvectors of each population covariance. MOP-UP (Tang et al., 2025) considers a decomposition in which one component has a column subspace shared across observations and the other has a shared row subspace. In the fully shared setting, Baharav et al. (2025) derive strength-dependent aggregation rules, while Agterberg (2026) studies estimation and adaptive inference for shared subspaces across symmetric low-rank matrices. Unlike these constructions, our model allows the loading matrices to arbitrarily mix the shared and view-specific coordinates. Consequently, the shared directions need not be singular vectors of any individual view and are instead identified through the intersection of the two population signal spaces.

Spiked random matrix theory. Our bias correction is motivated by classical results for spiked covariance and rectangular low-rank models, including the phase transition and eigenvector behavior studied in Johnstone (2001), Baik et al. (2005), and Paul (2007), and the singular-value and singular-vector asymptotics for rectangular perturbations developed by Benaych-Georges and Nadakuditi (2012). These results underlie later work on data-dependent spectral shrinkage and optimal singular-value transformations (Nadakuditi, 2014; Gavish and Donoho, 2017). BCSI uses the corresponding spike-location and singular-vector formulas to construct direction-dependent weights that equalize the leading leakage of the empirical singular vectors.

Finite-sample subspace estimation and spectral debiasing. Our analysis is also related to finite-sample singular-subspace perturbation theory for rectangular matrices (Cai and Zhang, 2018; Cai et al., 2021) and to spectral debiasing methods such as HeteroPCA (Zhang et al., 2022). HeteroPCA removes a diagonal bias due to heteroskedastic noise. The bias considered here arises instead after estimating the signal space of each view: different empirical singular directions leak outside their population signal spaces by different amounts. BCSI reweights these directions to make the leakage approximately isotropic and then removes the resulting common shift before estimating the intersection.

1.4 Notation

All matrices and vectors are set in bold. For $\boldsymbol{A} \in \mathbb{R}^{n\times d}$, $\|\boldsymbol{A}\|_{\mathrm{op}}$ is the spectral norm, $\boldsymbol{A}^{\dagger}$ the Moore–Penrose pseudoinverse, $\sigma_j(\boldsymbol{A})$ the j-th largest singular value, and $\sigma^{+}_{\min}(\boldsymbol{A})$ the smallest nonzero singular value. For symmetric $\boldsymbol{A}$ we write $\lambda_{\min}(\boldsymbol{A})$ and $\lambda^{+}_{\min}(\boldsymbol{A})$ for the smallest and the smallest positive eigenvalue. We set $\mathrm{St}(n,k) := \{\boldsymbol{W} \in \mathbb{R}^{n\times k} : \boldsymbol{W}^{\top}\boldsymbol{W} = \boldsymbol{I}_k\}$, write $\mathrm{col}(\boldsymbol{A})$ for the column space of $\boldsymbol{A}$, and write $\boldsymbol{P}_{\mathcal{V}}$ for the orthogonal projector onto a subspace $\mathcal{V}$. We write $a \wedge b := \min\{a,b\}$ and $a \vee b := \max\{a,b\}$. For nonnegative quantities x and y, the relation $x \lesssim y$ means that $x \le Cy$ for a universal constant $C > 0$, and $x \gtrsim y$ denotes the reverse inequality. We write $x \asymp y$ when both $x \lesssim y$ and $x \gtrsim y$ hold. For a matrix $\boldsymbol{Z}$, set $\|\boldsymbol{Z}\|_{L^q} := (\mathbb{E}\|\boldsymbol{Z}\|^q_{\mathrm{op}})^{1/q}$. For an event $\mathcal{E}$, let $\mathbf{1}_{\mathcal{E}}$ denote its indicator.

2 Model and bias-corrected subspace intersection

2.1 Model and problem parameters

We now give a concrete parametrization of the two-view JIVE model introduced in Section 1. The parametrization separates the shared and view-specific components and makes explicit the signal strengths and subspace separation that govern the difficulty of estimating the shared subspace.

Recall that $\mathcal{M}_k = \mathrm{col}(\boldsymbol{X}_k)$ and $\mathcal{U} = \mathcal{M}_1 \cap \mathcal{M}_2$. For $k = 1, 2$, we observe

$$\boldsymbol{Y}_k = \boldsymbol{X}_k + \boldsymbol{E}_k, \qquad \boldsymbol{X}_k = [\boldsymbol{U}, \boldsymbol{V}_k]\boldsymbol{B}_k^{\top} \in \mathbb{R}^{n\times d_k}. \tag{9}$$

Here $\boldsymbol{U} \in \mathrm{St}(n,r)$, $\boldsymbol{V}_k \in \mathrm{St}(n,r_k)$, and $[\boldsymbol{U}, \boldsymbol{V}_k] \in \mathrm{St}(n,p_k)$, where $p_k := r + r_k$. The matrix $\boldsymbol{B}_k \in \mathbb{R}^{d_k\times p_k}$ contains the right loadings. The entries of $\boldsymbol{E}_1$ and $\boldsymbol{E}_2$ are mutually independent $\mathcal{N}(0,\sigma^2)$ random variables.

We assume $\sigma_{p_k}(\boldsymbol{B}_k) \geq s_k > 0$. Since $[\boldsymbol{U}, \boldsymbol{V}_k]$ has orthonormal columns, the nonzero singular values of $\boldsymbol{X}_k$ and $\boldsymbol{B}_k$ coincide. In particular, $\boldsymbol{B}_k$ has full column rank and

$$\mathcal{M}_k = \operatorname{col}(\boldsymbol{X}_k) = \operatorname{col}(\boldsymbol{U}) \oplus \operatorname{col}(\boldsymbol{V}_k).$$

Apart from the lower bound on its smallest singular value, we impose no restriction on $\boldsymbol{B}_k$. Thus the right loadings may arbitrarily mix the shared and view-specific coordinates. We use the corresponding effective strength γ_k defined in (7).

It remains to ensure that the common block $\boldsymbol{U}$ spans the entire intersection of the two signal spaces. For a separation parameter $\theta \in (0, 1/2]$, assume

$$\|\boldsymbol{V}_1^\top \boldsymbol{V}_2\|_{\mathrm{op}} \leq 1 - 2\theta. \tag{10}$$

The left-hand side is the cosine of the smallest principal angle between the two view-specific subspaces. Hence any $\theta > 0$ implies $\operatorname{col}(\boldsymbol{V}_1) \cap \operatorname{col}(\boldsymbol{V}_2) = \{\boldsymbol{0}\}$. Consequently, $\mathcal{U} = \mathcal{M}_1 \cap \mathcal{M}_2 = \operatorname{col}(\boldsymbol{U})$. Smaller θ allows the two view-specific subspaces to be more closely aligned.

Our target is the shared projector $\boldsymbol{P}_{\mathcal{U}} = \boldsymbol{U}\boldsymbol{U}^\top$. For an estimator $\widehat{\boldsymbol{P}}$, we measure error by $\mathbb{E}[\|\widehat{\boldsymbol{P}} - \boldsymbol{P}_{\mathcal{U}}\|_{\mathrm{op}}]$.

2.2 The BCSI estimator

We now give the formal construction of the bias-corrected subspace intersection estimator (BCSI). It is also summarized in Algorithm 1. For each view, recall that $a_k = n - p_k$ and $b_k = d_k - p_k$. Let $\widehat{s}_{k,1} \geq \cdots \geq \widehat{s}_{k,p_k}$ be the leading p_k singular values of $\boldsymbol{Y}_k$, with corresponding left singular vectors $\widehat{\boldsymbol{\phi}}_{k,1}, \ldots, \widehat{\boldsymbol{\phi}}_{k,p_k}$, and write $\widehat{\boldsymbol{\Phi}}_k = [\widehat{\boldsymbol{\phi}}_{k,1}, \ldots, \widehat{\boldsymbol{\phi}}_{k,p_k}] \in \mathbb{R}^{n \times p_k}$.

Leakage bias correction. The directionwise correction motivated in Section 1.2 depends on the unknown population singular values. We therefore estimate them from the empirical singular values using the spike-location relation in (6). For an empirical singular value y above the spectral edge, solving this relation for the squared population singular value gives the upper root

$$\tau_k(y) := \begin{cases} \dfrac{y^2 - (a_k + b_k)\sigma^2 + \sqrt{\left\{y^2 - (a_k + b_k)\sigma^2\right\}^2 - 4a_k b_k \sigma^4}}{2}, & y > \sigma\left(\sqrt{a_k} + \sqrt{b_k}\right), \\ \sigma^2\sqrt{a_k b_k}, & \text{otherwise}. \end{cases} \tag{11}$$

At the spectral edge, the two roots coincide at $\sigma^2\sqrt{a_k b_k}$, and we keep $\tau_k(y)$ at this boundary value below the edge.

Substituting $\tau_k(y)$ into the oracle rescaling from Section 1.2 gives the weight

$$w_k(y) := \left\{ \frac{\tau_k(y)\left(\tau_k(y) + a_k\sigma^2\right)}{\tau_k(y) + b_k\sigma^2} \right\}^{1/2}. \tag{12}$$

We then form the bias-corrected empirical factor

$$\boldsymbol{C}_k := \widehat{\boldsymbol{\Phi}}_k \operatorname{diag}\{w_k(\widehat{s}_{k,1}), \ldots, w_k(\widehat{s}_{k,p_k})\} \in \mathbb{R}^{n \times p_k}. \tag{13}$$

The columns of $\boldsymbol{C}_k$ span the same empirical signal space as $\widehat{\boldsymbol{\Phi}}_k$, but the rescaling is chosen so that the leakage Gram of $\boldsymbol{C}_k$ is approximately $a_k\sigma^2\boldsymbol{I}_{p_k}$.

Algorithm 1 Bias-corrected subspace intersection (BCSI)

Input: Data matrices $\boldsymbol{Y}_1, \boldsymbol{Y}_2$; ranks r, r_1, r_2; noise level σ; signal strength lower bounds s_1, s_2.

- **Reweight the empirical signal directions.** For each $k = 1, 2$, let $p_k = r + r_k$, $a_k = n - p_k$, and $b_k = d_k - p_k$. Compute the leading p_k left singular vectors $\widehat{\boldsymbol{\Phi}}_k$ and singular values $\widehat{s}_{k,1}, \ldots, \widehat{s}_{k,p_k}$ of $\boldsymbol{Y}_k$, and set
$$\boldsymbol{C}_k = \widehat{\boldsymbol{\Phi}}_k \operatorname{diag}\{w_k(\widehat{s}_{k,1}), \ldots, w_k(\widehat{s}_{k,p_k})\},$$
where $w_k(\cdot)$ is defined in (12).

- **Form the corrected intersection.** Construct the centered Gram matrix
$$\boldsymbol{G} = \begin{pmatrix} \boldsymbol{C}_1^\top \boldsymbol{C}_1 - a_1\sigma^2 \boldsymbol{I}_{p_1} & \boldsymbol{C}_1^\top \boldsymbol{C}_2 \\ \boldsymbol{C}_2^\top \boldsymbol{C}_1 & \boldsymbol{C}_2^\top \boldsymbol{C}_2 - a_2\sigma^2 \boldsymbol{I}_{p_2} \end{pmatrix}.$$
Let $\boldsymbol{Q}_1$ and $\boldsymbol{Q}_2$ denote the two row blocks of an orthonormal basis for the invariant subspace associated with the r eigenvalues of $\boldsymbol{G}$ nearest zero.

- **Combine the two views.** With $\alpha_k = \gamma_k^2/(\gamma_1^2 + \gamma_2^2)$, form
$$\boldsymbol{M} = \alpha_1 \boldsymbol{C}_1 \boldsymbol{Q}_1 - \alpha_2 \boldsymbol{C}_2 \boldsymbol{Q}_2.$$

- **Output.** Return the projector $\widehat{\boldsymbol{P}}_{\text{BCSI}} = \boldsymbol{P}_{\text{col}(\boldsymbol{M})}$.

We therefore subtract this common leakage contribution from the within-view Gram blocks and define
$$\boldsymbol{G} \coloneqq \begin{pmatrix} \boldsymbol{C}_1^\top \boldsymbol{C}_1 - a_1\sigma^2 \boldsymbol{I}_{p_1} & \boldsymbol{C}_1^\top \boldsymbol{C}_2 \\ \boldsymbol{C}_2^\top \boldsymbol{C}_1 & \boldsymbol{C}_2^\top \boldsymbol{C}_2 - a_2\sigma^2 \boldsymbol{I}_{p_2} \end{pmatrix}. \tag{14}$$
After this correction, $\boldsymbol{G}$ is intended to approximate the Gram matrix obtained by retaining only the components of $\boldsymbol{C}_1$ and $\boldsymbol{C}_2$ inside their population signal spaces. The latter has an r-dimensional nullspace corresponding to the shared subspace.

Let $\boldsymbol{Q} = (\boldsymbol{Q}_1^\top, \boldsymbol{Q}_2^\top)^\top \in \mathbb{R}^{(p_1+p_2)\times r}$ have orthonormal columns spanning the invariant subspace associated with the r eigenvalues of $\boldsymbol{G}$ nearest zero, where $\boldsymbol{Q}_k \in \mathbb{R}^{p_k \times r}$.

Combining views of unequal strength. To account for unequal signal strengths, we follow the weighted aggregation used in prior work on shared-subspace estimation (Baharav et al., 2025; Li and Lyu, 2025). Specifically, we set
$$\boldsymbol{M} \coloneqq \alpha_1 \boldsymbol{C}_1 \boldsymbol{Q}_1 - \alpha_2 \boldsymbol{C}_2 \boldsymbol{Q}_2 \in \mathbb{R}^{n\times r}, \qquad \text{where } \alpha_k \coloneqq \frac{\gamma_k^2}{\gamma_1^2 + \gamma_2^2}, \qquad k = 1, 2. \tag{15}$$
Thus the stronger view contributes more to the reconstruction. Define $\boldsymbol{J} \coloneqq \operatorname{diag}(\alpha_1 \boldsymbol{I}_{p_1}, -\alpha_2 \boldsymbol{I}_{p_2})$, so that the BCSI reconstruction can be written as $\boldsymbol{M} = \boldsymbol{C}\boldsymbol{J}\boldsymbol{Q}$. Finally, define $\widehat{\boldsymbol{P}}_{\text{BCSI}} \coloneqq \boldsymbol{P}_{\text{col}(\boldsymbol{M})}$.

3 Statistical guarantees

3.1 Upper bound

We first state the performance guarantee for BCSI.

Theorem 1. *For $k = 1, 2$, suppose that $p_k \le (n \wedge d_k)/3$. If $\gamma_{\min} \ge \nu\sigma\sqrt{n}$ for some sufficiently large constant ν, then the BCSI estimator obeys*

$$\begin{aligned}
&\mathbb{E}\left\|\widehat{\boldsymbol{P}}_{\mathrm{BCSI}} - \boldsymbol{P}_{\mathcal{U}}\right\|_{\mathrm{op}} \\
&\quad \le C \min\left\{1, \frac{\sigma\sqrt{n}}{\gamma_{\max}} + \frac{\sigma\sqrt{r_1 + r_2}}{\gamma_{\min}} + \frac{\sigma\sqrt{r + (r_1 \wedge r_2)}}{\gamma_{\min}\sqrt{\theta}} + \frac{\sigma^2\sqrt{n\{r + (r_1 \vee r_2)\}}}{\gamma_{\min}^2\theta}\right\} \qquad (16)
\end{aligned}$$

for some universal constant $C > 0$.

Comparison with AJIVE. Consider the homogeneous two-view regime with $\gamma_1 \asymp \gamma_2 \asymp \gamma$ and $r \asymp r_1 \asymp r_2$. Specializing the recent high-probability bound for equal-weight AJIVE in Li and Lyu (2025) to this setting and suppressing logarithmic factors gives

$$\text{AJIVE:} \quad \frac{\sigma\sqrt{nr}}{\gamma} + \frac{\sigma r}{\gamma\sqrt{\theta}} + \frac{\sigma^2 n}{\gamma^2\theta}, \qquad \text{and} \qquad \text{BCSI:} \quad \frac{\sigma\sqrt{n}}{\gamma} + \frac{\sigma\sqrt{r}}{\gamma\sqrt{\theta}} + \frac{\sigma^2\sqrt{nr}}{\gamma^2\theta}.$$

The main difference is the quadratic term: bias correction improves this term by a factor $\sqrt{n/r}$. This gain is most consequential when θ is small, as the quadratic term can then dominate the estimation error; we illustrate this regime in the numerical experiments. The BCSI bound also has sharper rank dependence in the first-order terms.

Signal strength assumption. The condition $\gamma_k \gtrsim \sigma\sqrt{n}$ ensures that the population signal space $\mathcal{M}_k$ in each view can be estimated with nontrivial accuracy. The ratio $\frac{\sigma\sqrt{n}}{\gamma_k} \asymp \frac{\sigma\sqrt{n}}{s_k} + \frac{\sigma^2\sqrt{nd_k}}{s_k^2}$ is also the minimax optimal rate for rectangular column-space estimation (Cai and Zhang, 2018; Cai et al., 2021). This condition is natural for BCSI, which first estimates each signal space before estimating their intersection. It should not be interpreted as a necessary condition for estimating the shared subspace; see Section 6.

No condition-number dependence. An important feature of Theorem 1 is that the bound depends only on the smallest signal scale through γ_1 and γ_2. It imposes no upper bound on the singular values of $\boldsymbol{X}_1$ or $\boldsymbol{X}_2$, and hence no condition-number assumption on either signal matrix. This differs from existing analyses of AJIVE (Yang and Ma, 2025; Li and Lyu, 2025) that require the signal matrices to be well-conditioned.

3.2 Lower bound and minimax optimality

To determine whether the dependence in (16) is unavoidable, we consider the full class of signals satisfying the same rank, strength, and separation conditions. For fixed $s_1, s_2 > 0$, $r, r_1, r_2 \ge 1$, and $\theta \in (0, 1/2]$, define

$$\begin{aligned}
\mathfrak{P} := \Big\{&(\boldsymbol{U}, \boldsymbol{V}_1, \boldsymbol{V}_2, \boldsymbol{B}_1, \boldsymbol{B}_2) : \boldsymbol{U} \in \mathbb{R}^{n \times r}, \quad \boldsymbol{V}_k \in \mathbb{R}^{n \times r_k}, \\
&[\boldsymbol{U}, \boldsymbol{V}_k] \in \mathrm{St}(n, p_k), \quad \boldsymbol{B}_k \in \mathbb{R}^{d_k \times p_k}, \quad \sigma_{p_k}(\boldsymbol{B}_k) \ge s_k, \quad k = 1, 2, \\
&\|\boldsymbol{V}_1^\top \boldsymbol{V}_2\|_{\mathrm{op}} \le 1 - 2\theta\Big\}. \qquad (17)
\end{aligned}$$

Theorem 2. *For $k = 1, 2$, suppose that $p_k \le (n \wedge d_k)/3$. If $\gamma_{\min} \ge \nu\sigma\sqrt{n}$ for some sufficiently large constant ν, then*

$$\inf_{\widehat{\boldsymbol{P}}} \sup_{(\boldsymbol{U},\boldsymbol{V}_1,\boldsymbol{V}_2,\boldsymbol{B}_1,\boldsymbol{B}_2)\in\mathfrak{P}} \mathbb{E}\left\|\widehat{\boldsymbol{P}} - \boldsymbol{U}\boldsymbol{U}^\top\right\|_{\mathrm{op}}$$
$$\ge c\min\left\{1, \frac{\sigma\sqrt{n}}{\gamma_{\max}} + \frac{\sigma\sqrt{r + (r_1 \wedge r_2)}}{\gamma_{\min}\sqrt{\theta}} + \frac{\sigma^2\sqrt{n\{r + (r_1 \wedge r_2)\}}}{\gamma_{\min}^2\theta}\right\} \tag{18}$$

for some universal constant $c > 0$.

When $r \asymp r_1 \asymp r_2$,[1] Theorems 1 and 2 match up to numerical constants and yield the three-term rate in (8). Thus, in the signal-strength regime considered here, the θ-independent term, the $\theta^{-1/2}$ term, and the θ^{-1} second-order term are all intrinsic. For highly unbalanced view-specific ranks, a gap remains in the rank dependence: the upper bound depends on the larger individual rank, whereas the lower bound depends on the smaller one $r_1 \wedge r_2$.

Comparison with existing lower bounds. The closest prior information-theoretic result is the refined minimax lower bound of Yang and Ma (2025, Theorem 5). Their result considers a homogeneous K-view model with a common column dimension and signal strength and with equal shared and view-specific ranks. Specializing their bound to $K = 2$, and writing s for the common signal strength and d for the common column dimension, gives, up to numerical constants,

$$\frac{\sigma\sqrt{n}}{\gamma} + \frac{\sigma\sqrt{r}}{\gamma\sqrt{\theta}}, \qquad \text{where} \qquad \gamma = \frac{s^2}{\sqrt{s^2 + d\sigma^2}}.$$

Thus the prior lower bound captures the ambient and $\theta^{-1/2}$ difficulties, including the rectangular-matrix effect encoded by the effective strength γ. Our result recovers these scales while allowing unequal view dimensions, effective strengths, and view-specific ranks, and additionally identifies the θ^{-1} obstruction $\frac{\sigma^2\sqrt{nr}}{\gamma^2\theta}$ in the balanced-rank regime.

The proof of Theorem 2 is technically involved in general ranks, but the three statistical obstructions are already visible when $r = r_1 = r_2 = 1$. We therefore use the rank-one case to explain where the three terms come from. The general principle behind the lower bound is to choose the shared direction $\boldsymbol{u}$, the individual directions $\boldsymbol{v}_1, \boldsymbol{v}_2$, and the right loadings so that two or more signal instances have well-separated shared directions but nearly indistinguishable observation laws.

Understanding the θ-independent term. For the first term, suppose that the individual directions $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ are fixed and are even revealed to the estimator. We then vary only the shared direction $\boldsymbol{u}$ inside $\mathrm{span}\{\boldsymbol{v}_1, \boldsymbol{v}_2\}^\perp$, a space whose dimension is of order n. There is no difficulty here in deciding which directions are shared and which are individual: the only task is to locate $\boldsymbol{u}$ from two noisy views. Since both views change with $\boldsymbol{u}$, their information adds, giving the combined effective strength $\sqrt{\gamma_1^2 + \gamma_2^2} \asymp \gamma_{\max}$. Packing the possible shared directions over an n-dimensional space therefore gives the lower-bound scale $\frac{\sigma\sqrt{n}}{\sqrt{\gamma_1^2+\gamma_2^2}} \asymp \frac{\sigma\sqrt{n}}{\gamma_{\max}}$. This term is independent of θ because the individual directions are already known. It represents the unavoidable cost of estimating the shared direction itself. The proof follows a similar strategy in the single-matrix subspace estimation literature (Cai et al., 2021).

[1] In fact, the upper and lower bounds match in a broader setting when $r + (r_1 \vee r_2) \asymp r + (r_1 \wedge r_2)$.

Understanding the $1/\sqrt{\theta}$ term. The second construction isolates a different difficulty: even if the signal space of one view is perfectly known, its decomposition into shared and individual directions is not. Suppose without loss of generality that $\gamma_1 \geq \gamma_2$, so that View 2 is the weaker view. Fix orthonormal vectors $\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{z}_0 \in \mathbb{R}^n$. Take $\boldsymbol{e}_1$ to be the shared direction and $\boldsymbol{e}_2$ to be the individual direction in View 1, so that its signal space is the fixed plane $\mathrm{span}\{\boldsymbol{e}_1, \boldsymbol{e}_2\}$.

To vary this decomposition without changing the first-view signal space, rotate the shared and individual directions inside the same plane by an angle t:

$$\boldsymbol{u}_t = \cos(t)\boldsymbol{e}_1 + \sin(t)\boldsymbol{e}_2, \qquad \boldsymbol{v}_{1,t} = -\sin(t)\boldsymbol{e}_1 + \cos(t)\boldsymbol{e}_2. \tag{19}$$

Since $[\boldsymbol{u}_t, \boldsymbol{v}_{1,t}]$ is merely a rotation of $[\boldsymbol{e}_1, \boldsymbol{e}_2]$, the first-view signal matrix can be chosen to be exactly the same for every t and may therefore be revealed. For View 2, define

$$\boldsymbol{v}_{2,t} = (1-2\theta)\boldsymbol{v}_{1,t} + 2\sqrt{\theta(1-\theta)}\,\boldsymbol{z}_0.$$

Then $\boldsymbol{v}_{1,t}^\top \boldsymbol{v}_{2,t} = 1-2\theta$, so every t satisfies the prescribed separation condition. Thus all information about the unknown rotation t must come from View 2.

When θ is small, $\boldsymbol{v}_{2,t}$ is nearly aligned with $\boldsymbol{v}_{1,t}$. As t changes, the shared direction $\boldsymbol{u}_t$ and the individual direction $\boldsymbol{v}_{2,t}$ therefore rotate almost coherently. The component that distinguishes different values of t lies in the fixed direction $\boldsymbol{z}_0$ and has size of order $\sqrt{\theta}$. Consequently, View 2 can distinguish rotations only at the scale $\frac{\sigma}{\gamma_2\sqrt{\theta}} = \frac{\sigma}{\gamma_{\min}\sqrt{\theta}}$. Thus the second term is the cost of identifying which direction inside an otherwise known signal plane is shared.

Understanding the $1/\theta$ term. The third construction starts from exactly the same family and makes only one change: the fixed direction $\boldsymbol{z}_0$ that helps View 2 resolve the rotation is now replaced by an unknown direction $\boldsymbol{z} \in \mathrm{span}\{\boldsymbol{e}_1, \boldsymbol{e}_2\}^\perp$:

$$\boldsymbol{v}_{2,t,\boldsymbol{z}} = (1-2\theta)\boldsymbol{v}_{1,t} + 2\sqrt{\theta(1-\theta)}\,\boldsymbol{z}.$$

For each fixed t, we randomize $\boldsymbol{z}$ over an $(n-2)$-dimensional family. This changes only the individual direction in View 2 and leaves the target $\boldsymbol{u}_t\boldsymbol{u}_t^\top$ fixed. The unknown parameter of interest is still the same scalar rotation t, but the reference direction that helped reveal this rotation in the preceding construction is now itself hidden.

With the fixed $\boldsymbol{z}_0$, View 2 can compare the rotating signal plane against a known direction outside that plane. After $\boldsymbol{z}_0$ is replaced by the hidden direction $\boldsymbol{z}$, this reference must itself be learned from the data. Averaging over the high-dimensional nuisance direction removes the first-order directional information available in the preceding construction. View 2 must now simultaneously infer the hidden direction and determine the rotation t, leaving only a second-order amount of information about the latter. Hiding $\boldsymbol{z}$ in a space whose dimension is of order n leads to indistinguishable rotations of size $\frac{\sigma^2\sqrt{n}}{\gamma_2^2\theta} = \frac{\sigma^2\sqrt{n}}{\gamma_{\min}^2\theta}$, which gives the third lower-bound term.

4 Numerical experiments

In this section, we use synthetic and real data to compare BCSI with AJIVE.

4.1 Synthetic data

Data-generating process. We consider the model with $d_1 = d_2 = n$, $\sigma = 1$, and $r = r_1 = r_2 = 1$. Thus each signal matrix has rank two, with one shared direction and one view-specific direction.

Let $\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3$ be the first three standard basis vectors in $\mathbb{R}^n$. Set the signal strength $s_n = n^{2/3}$, and the separation parameter $\theta_n = \frac{n^{-1/3}}{4}$. We take the shared direction to be $\boldsymbol{u} = \boldsymbol{e}_1$ and define

$$\boldsymbol{v}_1 = \sqrt{1-\theta_n}\,\boldsymbol{e}_2 + \sqrt{\theta_n}\,\boldsymbol{e}_3, \qquad \boldsymbol{v}_2 = \sqrt{1-\theta_n}\,\boldsymbol{e}_2 - \sqrt{\theta_n}\,\boldsymbol{e}_3.$$

Then the two signal matrices and observations are generated according to

$$\boldsymbol{X}_k = [\boldsymbol{u}, \boldsymbol{v}_k]\,\frac{1}{\sqrt{2}}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} s_n & 0 \\ 0 & 2s_n \end{pmatrix}[\boldsymbol{e}_1, \boldsymbol{e}_2]^\top, \qquad \boldsymbol{Y}_k = \boldsymbol{X}_k + \boldsymbol{E}_k, \qquad k = 1, 2,$$

where the entries of $\boldsymbol{E}_1$ and $\boldsymbol{E}_2$ are mutually independent standard Gaussian random variables. The 2×2 matrix in this display rotates $\boldsymbol{u}$ and $\boldsymbol{v}_k$ by $45°$. Consequently, the left singular vectors of $\boldsymbol{X}_k$ are $(\boldsymbol{u}+\boldsymbol{v}_k)/\sqrt{2}$ and $(-\boldsymbol{u}+\boldsymbol{v}_k)/\sqrt{2}$, with singular values s_n and $2s_n$. Moreover, $\boldsymbol{v}_1^\top \boldsymbol{v}_2 = 1 - 2\theta_n$, so the two view-specific subspaces become more closely aligned as n increases. The rotation ensures that the shared and view-specific directions are not themselves singular vectors, allowing unequal leakage from the two empirical singular directions to bias the intersection estimated by AJIVE. Under this scaling, the leading term in the BCSI rate decreases as $n^{-1/6}$, whereas the quadratic bias scale for AJIVE remains constant.

Estimators being compared. We compare four procedures. The first is AJIVE, obtained from the leading rank-one eigenspace of the sum of the two empirical rank-two signal-space projectors. The second is Expected AJIVE: at each value of n, we average the final rank-one AJIVE projectors across Monte Carlo replications and take the leading eigenspace of this average. Expected AJIVE is included as a diagnostic for systematic bias and is not an estimator computed from a single dataset. The third procedure is oracle BCSI, which uses the true noise level and the true effective signal strength $\gamma_n = s_n^2/\sqrt{s_n^2 + n}$. The fourth is estimated-parameter BCSI, which estimates the noise level from the pooled rank-two residual energy and estimates the smallest signal singular value separately in each view. All data-dependent procedures use marginal rank two and joint rank one.

Experimental results. Figure 1 shows that the errors of both AJIVE and Expected AJIVE remain nearly constant as the dimension increases. The persistence of the Expected-AJIVE error indicates that this behavior is caused by systematic bias rather than only by random fluctuations. In contrast, the BCSI error decreases with the dimension and closely follows the predicted $n^{-1/6}$ behavior. The oracle and estimated-parameter BCSI curves are nearly indistinguishable, indicating that estimating the required parameters has little effect in this experiment. The results therefore support the claim that correcting the direction-dependent leakage removes the persistent error suffered by AJIVE in this regime.

4.2 Real data

Data and scientific question. We apply AJIVE and BCSI to the Nutrimouse data (Martin et al., 2007). The data contain gene-expression measurements for 120 genes and concentrations of 21 liver fatty acids from the same 40 mice. The mice belong to two genotypes and five diet groups, with four mice in each genotype-by-diet group. Previous analyses of these data suggest that genotype is represented in the variation shared by the gene-expression and lipid views, whereas diet is primarily represented in the lipid-specific variation (Yuan et al., 2022; Sergazinov et al., 2026). We therefore examine whether the estimated joint subspace retains variation associated with genotype while excluding variation associated with diet.

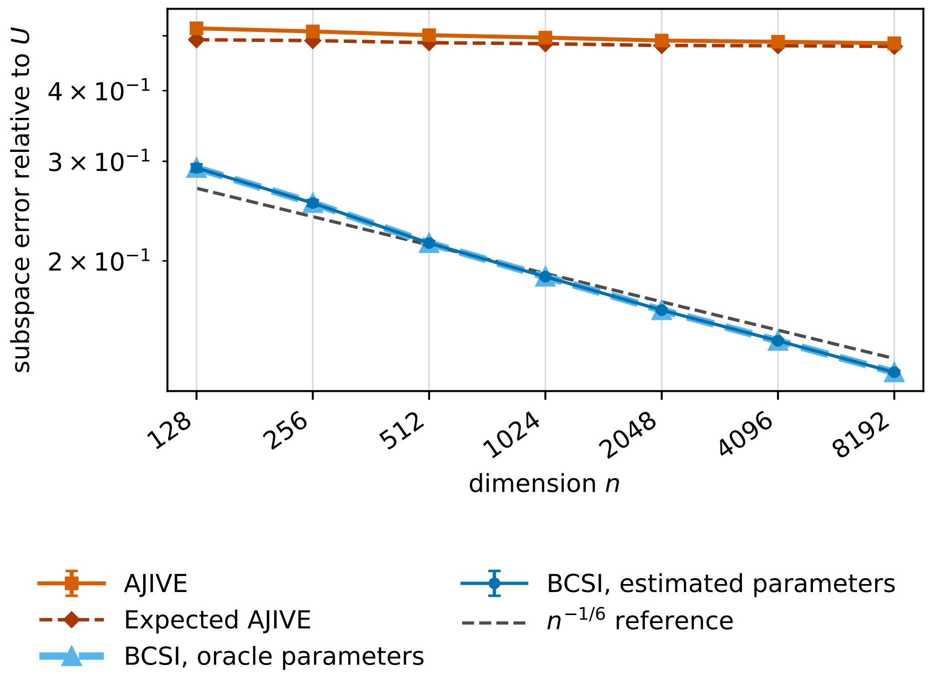

Figure 1: Comparison of AJIVE and BCSI as the dimension increases. The points for AJIVE and BCSI are mean projector errors, and the error bars are ± 2 Monte Carlo standard errors. Expected AJIVE is the leading eigenspace of the Monte Carlo mean of the final AJIVE projectors. Oracle BCSI uses the true noise level and effective signal strength, whereas estimated-parameter BCSI estimates them from the data. The dashed line is an $n^{-1/6}$ reference.

Estimators and rank selection. We center and standardize every feature and then normalize each view by its estimated residual noise level. We use marginal rank $p_1 = 3$ for gene expression and $p_2 = 4$ for lipids. These ranks were used in the earlier Nutrimouse analysis of Yuan et al. (2022). Sergazinov et al. (2026) observed that they agree with the elbows of the two scree plots and give clearer spectral separation than the larger ranks selected by an automatic singular-value threshold. With these marginal ranks, AJIVE selects joint rank $r = 2$. We supply the same marginal and joint ranks to AJIVE and BCSI, so the comparison concerns the estimated joint subspaces rather than differences in rank selection. AJIVE is computed from the two empirical marginal subspaces. BCSI uses the same empirical marginal subspaces but applies the proposed leakage correction, with the noise level and signal strengths estimated from the data.

Evaluation and uncertainty. For each estimated joint subspace, we measure the fraction of its score variation explained by genotype and by diet. This quantity is one minus the SWISS criterion of Cabanski et al. (2010). A larger genotype association indicates that the joint scores retain more of the factor expected to be shared, whereas a smaller diet association indicates that they retain less of the factor expected to be view-specific. The genotype and diet labels are used only for this evaluation and do not enter either estimator. To assess uncertainty, we use 2,000 paired bootstrap refits. Within each bootstrap replication, we sample four mice with replacement from each of the ten genotype-by-diet groups. Each mouse is resampled together with both of its data views, and AJIVE and BCSI are fitted to the same resampled dataset. This preserves the paired multi-view structure and the size of every experimental group. We form 95% percentile intervals from the

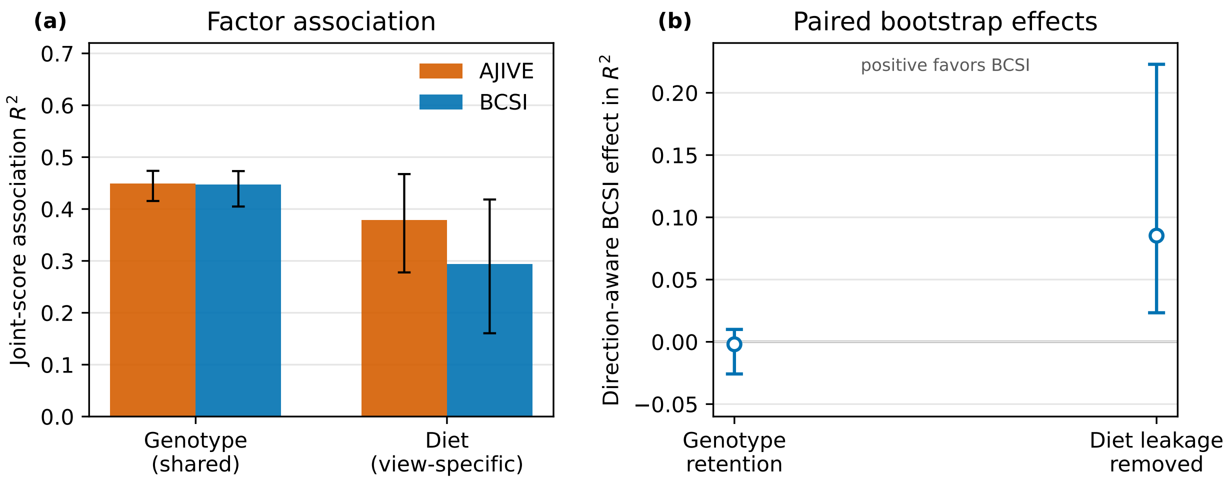

Figure 2: Matched-rank AJIVE and BCSI on the Nutrimouse data. Panel (a) reports the association of the estimated joint scores with genotype and diet. The error bars are 95% percentile intervals from 2,000 paired bootstrap refits. Panel (b) reports the paired differences, with positive values favoring BCSI. For genotype, the difference is BCSI minus AJIVE; for diet, it is AJIVE minus BCSI.

bootstrap distributions of the individual metrics and their paired differences.

Interpretation of the results. Figure 2 shows that the joint scores estimated by AJIVE and BCSI have essentially the same association with genotype, and the paired bootstrap interval does not indicate a meaningful difference between the methods. In contrast, the BCSI joint scores have a weaker association with diet, and the paired bootstrap interval favors BCSI. This suggests that BCSI removes more variation associated with the view-specific factor without a corresponding loss of variation associated with the shared factor.

5 Analysis of BCSI

We now develop the main ingredients for analyzing BCSI. Throughout this section, we view the corrected Gram matrix as a perturbation of a signal-only Gram matrix that encodes the shared subspace.

For each view $k = 1, 2$, decompose the bias-corrected empirical factor into its component inside the corresponding signal space and its orthogonal leakage:

$$\boldsymbol{C}_k = \boldsymbol{S}_k + \boldsymbol{L}_k, \qquad \boldsymbol{S}_k := \boldsymbol{P}_{\mathcal{M}_k}\boldsymbol{C}_k, \qquad \boldsymbol{L}_k := \boldsymbol{P}_{\mathcal{M}_k^\perp}\boldsymbol{C}_k. \tag{20}$$

Here $\boldsymbol{S}_k$ is the part of the bias-corrected empirical factor that remains inside the population signal space, whereas $\boldsymbol{L}_k$ records its leakage into the orthogonal complement.

With $\boldsymbol{C} := [\boldsymbol{C}_1, \boldsymbol{C}_2]$, $\boldsymbol{S} := [\boldsymbol{S}_1, \boldsymbol{S}_2]$, $\boldsymbol{L} := [\boldsymbol{L}_1, \boldsymbol{L}_2]$, and $\boldsymbol{D}_\sigma := \operatorname{diag}(a_1\sigma^2\boldsymbol{I}_{p_1}, a_2\sigma^2\boldsymbol{I}_{p_2})$, the bias-corrected Gram matrix in (14) admits the exact decomposition

$$\boldsymbol{G} = \underbrace{\boldsymbol{S}^\top\boldsymbol{S}}_{=:\,\boldsymbol{G}_0} + \underbrace{\boldsymbol{S}^\top\boldsymbol{L} + \boldsymbol{L}^\top\boldsymbol{S} + \left(\boldsymbol{L}^\top\boldsymbol{L} - \boldsymbol{D}_\sigma\right)}_{=:\,\boldsymbol{\Delta}}. \tag{21}$$

Thus the natural oracle matrix is $\boldsymbol{G}_0 = \boldsymbol{S}^\top \boldsymbol{S}$. The first two terms in $\boldsymbol{\Delta}$ are linear in the leakage, while $\boldsymbol{L}^\top \boldsymbol{L} - \boldsymbol{D}_\sigma$ is the centered quadratic leakage. This decomposition makes precise the role of the bias correction: the weighting makes $\boldsymbol{L}^\top \boldsymbol{L}$ close to the common leakage level $\boldsymbol{D}_\sigma$, which is then subtracted so that $\boldsymbol{G}$ is close to the signal-only Gram matrix $\boldsymbol{G}_0$.

Proof roadmap. The proof proceeds in four steps. First, we show that the oracle matrix $\boldsymbol{G}_0 = \boldsymbol{S}^\top \boldsymbol{S}$ has an r-dimensional nullspace that exactly encodes the shared subspace, and that its positive spectrum is separated from zero by a gap of order $\gamma_{\min}^2 \theta$. Second, we develop a relative perturbation bound showing that the r eigenvectors $\boldsymbol{Q}$ of $\boldsymbol{G} = \boldsymbol{G}_0 + \boldsymbol{\Delta}$ nearest zero approximately satisfy the oracle nullspace relation, i.e., $\|\boldsymbol{S}\boldsymbol{Q}\|_{\mathrm{op}}$ is small. Third, we translate this approximate relation into an error bound for the reconstructed shared subspace and decompose the error into a leading term linear in the leakage and a higher-order remainder. Finally, the stochastic analysis controls these leakage terms, verifies the relative perturbation condition, and yields the rate in Theorem 1. The first three steps are developed in Section 5.1, and the final step in Section 5.2; proofs of the main auxiliary results are deferred to the appendix.

5.1 Deterministic analysis

Fix a sufficiently small numerical constant $c_{\mathrm{emb}} > 0$ and define

$$\mathcal{E}_{\mathrm{emb}} := \bigcap_{k=1}^{2} \left\{ \sigma_{\min}(\boldsymbol{S}_k) \geq c_{\mathrm{emb}} \gamma_k \right\}. \tag{22}$$

On this event, each $\boldsymbol{S}_k$ has full column rank and therefore $\mathrm{col}(\boldsymbol{S}_k) = \mathcal{M}_k$. We work under this deterministic condition throughout the next several arguments; its probability is controlled later.

5.1.1 Oracle geometry

Since BCSI selects the eigenvectors of $\boldsymbol{G}$ nearest zero, we first understand the nullspace of the oracle matrix $\boldsymbol{G}_0 = \boldsymbol{S}^\top \boldsymbol{S}$. This oracle matrix should have an r-dimensional nullspace encoding the coefficient relations that produce the shared subspace. To exhibit these relations, define

$$\boldsymbol{Q}_\star := \begin{pmatrix} \boldsymbol{S}_1^\dagger \boldsymbol{U} \\ -\boldsymbol{S}_2^\dagger \boldsymbol{U} \end{pmatrix} \in \mathbb{R}^{(p_1+p_2)\times r}.$$

Lemma 1. *On $\mathcal{E}_{\mathrm{emb}}$,*

$$\ker(\boldsymbol{G}_0) = \ker(\boldsymbol{S}), \qquad \dim\ker(\boldsymbol{G}_0) = r.$$

Moreover,

$$\boldsymbol{S}\boldsymbol{Q}_\star = \boldsymbol{0}, \qquad \boldsymbol{S}\boldsymbol{J}\boldsymbol{Q}_\star = \boldsymbol{U}, \tag{23}$$

and

$$\lambda_{\min}^{+}(\boldsymbol{G}_0) \geq 2 c_{\mathrm{emb}}^2 \gamma_{\min}^2 \theta. \tag{24}$$

Thus the nullspace of $\boldsymbol{G}_0$ consists exactly of the coefficient relations that reconstruct the shared subspace, while (24) separates this nullspace from the positive spectrum. These are the two properties needed to analyze the perturbed near-zero eigenspace of $\boldsymbol{G}$.

5.1.2 Perturbation of the nullspace

The standard Davis–Kahan argument tells us that if the perturbation size $\|\boldsymbol{\Delta}\|_{\mathrm{op}}$ is small, then the near-zero eigenspace of $\boldsymbol{G} = \boldsymbol{G}_0 + \boldsymbol{\Delta}$ stays close to the oracle nullspace $\ker(\boldsymbol{G}_0)$. However, this does not suffice for two reasons. First, by Lemma 1, the oracle nullspace $\ker(\boldsymbol{G}_0) = \ker(\boldsymbol{S})$ exactly encodes the coefficient relations that produce the shared subspace. Thus what matters for reconstruction is not the Euclidean distance between the two coefficient spaces, but how well the empirical eigenspace continues to satisfy the oracle relation, as measured by $\|\boldsymbol{S}\boldsymbol{Q}\|_{\mathrm{op}}$. Second, a perturbation bound based only on $\|\boldsymbol{\Delta}\|_{\mathrm{op}}$ is too crude. Indeed, the linear part of $\boldsymbol{\Delta}$ contains $\boldsymbol{S}^\top \boldsymbol{L} + \boldsymbol{L}^\top \boldsymbol{S}$, whose operator norm can scale with the largest singular value of $\boldsymbol{S}$. Such a bound would therefore introduce an unnecessary upper signal scale, even though perturbations along stronger signal directions are accompanied by correspondingly larger eigenvalues of $\boldsymbol{G}_0$.

To address these two issues, we use the following self-contained deterministic perturbation result. Let $d \geq 2$, let $\boldsymbol{G}_0 \in \mathbb{R}^{d\times d}$ be symmetric and positive semidefinite, and let $\boldsymbol{\Delta} \in \mathbb{R}^{d\times d}$ be symmetric. Set $\boldsymbol{G} \coloneqq \boldsymbol{G}_0 + \boldsymbol{\Delta}$ and define

$$\begin{aligned} \boldsymbol{\Delta}_{00} &\coloneqq \boldsymbol{P}_{\ker(\boldsymbol{G}_0)} \boldsymbol{\Delta} \boldsymbol{P}_{\ker(\boldsymbol{G}_0)}, \\ \boldsymbol{\Delta}_{+0} &\coloneqq (\boldsymbol{G}_0^\dagger)^{1/2} \boldsymbol{\Delta} \boldsymbol{P}_{\ker(\boldsymbol{G}_0)}, \\ \boldsymbol{\Delta}_{++} &\coloneqq (\boldsymbol{G}_0^\dagger)^{1/2} \boldsymbol{\Delta} (\boldsymbol{G}_0^\dagger)^{1/2}. \end{aligned} \tag{25}$$

These three matrices describe the perturbation relative to the decomposition $\ker(\boldsymbol{G}_0) \oplus \ker(\boldsymbol{G}_0)^\perp$, after scaling the positive-eigenvalue directions by $(\boldsymbol{G}_0^\dagger)^{1/2}$. Specifically, $\boldsymbol{\Delta}_{00}$ acts within the oracle nullspace, $\boldsymbol{\Delta}_{+0}$ maps the oracle nullspace into its orthogonal complement, and $\boldsymbol{\Delta}_{++}$ acts within that orthogonal complement.

Define the perturbation size $\varepsilon_\Delta \coloneqq \|\boldsymbol{\Delta}_{00}\|_{\mathrm{op}} + 2\|\boldsymbol{\Delta}_{+0}\|_{\mathrm{op}}^2$.

Lemma 2. *Assume that* $\dim\ker(\boldsymbol{G}_0) = r$ *for some* $1 \leq r < d$, *and* $\boldsymbol{G} = \boldsymbol{G}_0 + \boldsymbol{\Delta}$ *with*

$$\|\boldsymbol{\Delta}_{++}\|_{\mathrm{op}} \leq \frac{1}{4}, \qquad \varepsilon_\Delta \leq \frac{1}{16}\lambda_{\min}^+(\boldsymbol{G}_0).$$

Then $\boldsymbol{G}$ *has exactly* r *eigenvalues in* $[-\varepsilon_\Delta, \varepsilon_\Delta]$, *while all remaining eigenvalues exceed* $4\varepsilon_\Delta$. *If* $\boldsymbol{Q} \in \mathrm{St}(d, r)$ *spans the invariant subspace associated with these* r *eigenvalues, then*

$$\left\|\boldsymbol{G}_0^{1/2}\boldsymbol{Q}\right\|_{\mathrm{op}} \leq 2\|\boldsymbol{\Delta}_{+0}\|_{\mathrm{op}}. \tag{26}$$

In our application, $d = p_1 + p_2$ and $\boldsymbol{G}_0 = \boldsymbol{S}^\top \boldsymbol{S}$. Hence $\boldsymbol{S}\boldsymbol{Q}$ and $\boldsymbol{G}_0^{1/2}\boldsymbol{Q}$ have the same Gram matrix, so Lemma 2 gives

$$\|\boldsymbol{S}\boldsymbol{Q}\|_{\mathrm{op}} = \left\|\boldsymbol{G}_0^{1/2}\boldsymbol{Q}\right\|_{\mathrm{op}} \leq 2\|\boldsymbol{\Delta}_{+0}\|_{\mathrm{op}}. \tag{27}$$

Moreover, $\boldsymbol{S}\boldsymbol{P}_{\ker(\boldsymbol{G}_0)} = \boldsymbol{0}$. Consequently, the linear leakage terms in $\boldsymbol{\Delta}$ vanish from $\boldsymbol{\Delta}_{00}$, leaving only the centered quadratic leakage.

We therefore define the good event

$$\mathcal{E}_{\mathrm{good}} \coloneqq \mathcal{E}_{\mathrm{emb}} \cap \left\{ \|\boldsymbol{\Delta}_{++}\|_{\mathrm{op}} \leq \frac{1}{4}, \quad \|\boldsymbol{\Delta}_{00}\|_{\mathrm{op}} + 2\|\boldsymbol{\Delta}_{+0}\|_{\mathrm{op}}^2 \leq \frac{1}{8} c_{\mathrm{emb}}^2 \gamma_{\min}^2 \theta \right\}. \tag{28}$$

By Lemma 1 and (24), the conditions of Lemma 2 hold on $\mathcal{E}_{\mathrm{good}}$. Therefore, the invariant subspace of $\boldsymbol{G}$ identified in Lemma 2 is precisely the near-zero eigenspace selected by BCSI and satisfies (27). Since the oracle relation is $\boldsymbol{S}\boldsymbol{Q}_\star = \boldsymbol{0}$, this bound shows that the empirical eigenvectors approximately satisfy the same nullspace relation.

5.1.3 From the near-zero eigenspace perturbation to subspace error

We now translate the bound on $\|\boldsymbol{SQ}\|_{\mathrm{op}}$ into a bound on the error of the estimated shared subspace. The orthonormal normalization of $\boldsymbol{Q}$ is natural for the eigenspace problem, whereas $\boldsymbol{Q}_\star$ is a basis of the oracle nullspace normalized so that $\boldsymbol{SJQ}_\star = \boldsymbol{U}$. Even when $\boldsymbol{Q}$ spans the exact oracle kernel, $\boldsymbol{SJQ}$ need only equal $\boldsymbol{U}$ up to an invertible right factor. We remove this arbitrary factor before measuring subspace error.

Recall that $\boldsymbol{M} = \boldsymbol{CJQ}$. On $\mathcal{E}_{\mathrm{good}}$, $\boldsymbol{U}^\top \boldsymbol{M}$ is invertible (see Lemma 10). Define $\widetilde{\boldsymbol{M}} \coloneqq \boldsymbol{M}(\boldsymbol{U}^\top \boldsymbol{M})^{-1}$. This right transformation does not change the estimated column space, while $\boldsymbol{U}^\top \widetilde{\boldsymbol{M}} = \boldsymbol{I}_r$. Consequently,

$$\left\| \widehat{\boldsymbol{P}}_{\mathrm{BCSI}} - \boldsymbol{P}_{\mathcal{U}} \right\|_{\mathrm{op}} \le \left\| \boldsymbol{P}_{\mathcal{U}^\perp} \widetilde{\boldsymbol{M}} \right\|_{\mathrm{op}}. \tag{29}$$

We next decompose the right-hand side into its leading and higher-order terms. To state the decomposition, define the following linear map. Every $\boldsymbol{x} \in \mathcal{M}_1 + \mathcal{M}_2$ has a unique representation $\boldsymbol{x} = \boldsymbol{Ua} + \boldsymbol{V}_1\boldsymbol{b}_1 + \boldsymbol{V}_2\boldsymbol{b}_2$. Define $\boldsymbol{T} : \mathbb{R}^n \to \mathbb{R}^n$ by

$$\boldsymbol{Tx} \coloneqq \alpha_1 \boldsymbol{V}_1 \boldsymbol{b}_1 - \alpha_2 \boldsymbol{V}_2 \boldsymbol{b}_2 \quad \text{for } \boldsymbol{x} \in \mathcal{M}_1 + \mathcal{M}_2, \qquad \boldsymbol{Tx} \coloneqq \boldsymbol{0} \quad \text{for } \boldsymbol{x} \in (\mathcal{M}_1 + \mathcal{M}_2)^\perp. \tag{30}$$

Lemma 3. *On $\mathcal{E}_{\mathrm{good}}$, the error has the decomposition*

$$\boldsymbol{P}_{\mathcal{U}^\perp} \widetilde{\boldsymbol{M}} = (\boldsymbol{LJ} - \boldsymbol{TL})\boldsymbol{Q}_\star + \textsf{\textbf{Rem}}, \tag{31}$$

where the higher-order remainder satisfies

$$\|\textsf{\textbf{Rem}}\|_{\mathrm{op}} \lesssim \frac{1}{\gamma_{\min}^2 \theta} \left[\left\| \boldsymbol{L}^\top \boldsymbol{L} - \boldsymbol{D}_\sigma \right\|_{\mathrm{op}} + \left\{ \sqrt{\theta}\, \|\boldsymbol{L}\|_{\mathrm{op}} + \left\| \boldsymbol{P}_{\mathrm{col}(\boldsymbol{S})} \boldsymbol{L} \right\|_{\mathrm{op}} + \|\boldsymbol{\Delta}_{+0}\|_{\mathrm{op}} \right\} \|\boldsymbol{\Delta}_{+0}\|_{\mathrm{op}} \right]. \tag{32}$$

The two leading terms in (31) are linear in the leakage $\boldsymbol{L}$, while **Rem** contains only second- and higher-order terms. The stochastic analysis controls these terms separately.

5.2 Stochastic analysis

The deterministic analysis reduces the proof to controlling the first-order term in (31), the remainder in (32), and the probability of $\mathcal{E}_{\mathrm{good}}$.

The main probabilistic input is the following one-view result, which quantifies both the size and the centered quadratic fluctuation of the weighted leakage.[2]

Lemma 4. *Fix $k \in \{1, 2\}$. Suppose that $p_k \le (n \wedge d_k)/3$ and that $\gamma_k \ge \nu\sigma\sqrt{n}$ for a sufficiently large constant ν. Then there exists an event $\mathcal{E}_k$ such that $\sigma_{\min}(\boldsymbol{S}_k) \ge c_{\mathrm{emb}}\gamma_k$ on $\mathcal{E}_k$ and $\mathbb{P}(\mathcal{E}_k^c) \lesssim \sigma^2\sqrt{np_k}/\gamma_k^2$. Moreover,*

$$\left\| \boldsymbol{L}_k^\top \boldsymbol{L}_k - a_k \sigma^2 \boldsymbol{I}_{p_k} \right\|_{L^{32}} \lesssim \sigma^2 \sqrt{a_k p_k}, \qquad \|\boldsymbol{L}_k\|_{L^{64}} \lesssim \sigma\sqrt{a_k}. \tag{33}$$

The lemma can be viewed as a nonasymptotic counterpart of singular-vector overlap results from spiked random matrix theory. Classical rectangular spiked-matrix theory characterizes outlier singular values and singular-vector overlaps in the fixed-rank asymptotic regime (Benaych-Georges and Nadakuditi, 2012; Gavish and Donoho, 2017; Bao et al., 2021). More recent work permits the signal rank to grow with the matrix dimensions, including sublinear- and proportional-rank regimes (Liu et al., 2025; Landau et al., 2023; Shmalo, 2026). These results are mainly asymptotic. The

[2] The results hold for general moments. We use $q = 32$ throughout for concreteness and to simplify notation.

purpose of Lemma 4 is different. It gives finite-sample moment control of the full, data-dependently weighted leakage Gram that enters the analysis of BCSI. This conclusion is obtained under the comparatively strong signal-strength condition $\gamma_k \geq \nu\sigma\sqrt{n}$. Thus the lemma is complementary to the existing asymptotic theory. To our knowledge, this particular finite-sample estimate has not appeared previously.

For BCSI, this result provides the finite-sample justification for subtracting $a_k\sigma^2 \boldsymbol{I}_{p_k}$ from the within-view Gram blocks. Together with the rotational invariance of Gaussian noise, it also supplies the probabilistic input needed to control the linear and higher-order leakage terms. We defer these probabilistic calculations to the appendix and record their consequences below.

Lemma 5. *Under the assumptions of Theorem 1, suppose in addition that the sum of the four rate terms on the right-hand side of (16), before truncation at* 1*, is at most a sufficiently small numerical constant. Then*

$$\mathbb{E}\left[\mathbf{1}_{\mathcal{E}_{\mathrm{good}}} \|(\boldsymbol{L}\boldsymbol{J} - \boldsymbol{T}\boldsymbol{L})\boldsymbol{Q}_\star\|_{\mathrm{op}}\right] \lesssim \frac{\sigma\sqrt{n}}{\gamma_{\max}} + \frac{\sigma\sqrt{r_1 + r_2}}{\gamma_{\min}} + \frac{\sigma\sqrt{r + (r_1 \wedge r_2)}}{\gamma_{\min}\sqrt{\theta}}, \tag{34a}$$

$$\mathbb{E}\left[\mathbf{1}_{\mathcal{E}_{\mathrm{good}}} \|\mathsf{\textbf{Rem}}\|_{\mathrm{op}}\right] \lesssim \frac{\sigma^2\sqrt{n\{r + (r_1 \vee r_2)\}}}{\gamma_{\min}^2\theta}, \tag{34b}$$

$$\mathbb{P}(\mathcal{E}_{\mathrm{good}}^c) \lesssim \frac{\sigma\sqrt{r_1 + r_2}}{\gamma_{\min}} + \frac{\sigma\sqrt{r + (r_1 \wedge r_2)}}{\gamma_{\min}\sqrt{\theta}} + \frac{\sigma^2\sqrt{n\{r + (r_1 \vee r_2)\}}}{\gamma_{\min}^2\theta}. \tag{34c}$$

Completing the proof of Theorem 1. Combining (29) and (31) with Lemma 5, and using that the distance between two orthogonal projectors is at most one on $\mathcal{E}_{\mathrm{good}}^c$, proves (16) when the untruncated rate is sufficiently small. The remaining case follows from the same trivial bound after enlarging the universal constant.

6 Discussion

We propose a bias-corrected subspace intersection estimator (BCSI) for the shared subspace in the two-view JIVE model with Gaussian noise and establish its statistical theory. In the regime $\gamma_{\min} \gtrsim \sigma\sqrt{n}$ and with comparable shared and view-specific ranks, BCSI attains the minimax rate and identifies the distinct first- and second-order difficulties in estimating the shared subspace. The theory also points to several unresolved questions, which we discuss below.

What happens when the ranks are unbalanced? The upper and lower bounds match when $r + (r_1 \vee r_2) \asymp r + (r_1 \wedge r_2)$, but a gap remains when the two individual ranks are substantially different. The upper bound contains terms that depend on the larger individual rank, whereas the corresponding terms in the lower bound depend only on the smaller individual rank. The principal-angle decomposition suggests that the rank dependence in the upper bound may not be sharp: the blocks for the $r_1 \wedge r_2$ paired directions can have smallest eigenvalues of order θ, while each of the remaining $|r_1 - r_2|$ unpaired directions has eigenvalue 1. The dependence on the larger rank may come from controlling the leakage globally rather than separately over these two types of directions. An upper-bound proof that controls the two types of principal-angle blocks separately, or a lower-bound construction that captures the second-order contribution of the unpaired directions, may sharpen the rank dependence.

How strong must the weaker view be? The condition $\gamma_{\min} \gtrsim \sigma\sqrt{n}$ is natural for BCSI because its construction requires the empirical factor in each view to retain all population signal directions. Thus the signal strength must be large enough for both empirical factors to be reliable. The rate itself suggests that the two views play different roles: the stronger view estimates the shared subspace in $\mathbb{R}^n$, as reflected by the term involving $\gamma_{\max}$, whereas the weaker view determines which directions within the stronger signal space are shared, as reflected by the terms involving $\gamma_{\min}$. The weaker view might identify the shared subspace through part of its signal even when it cannot recover its entire signal space. The present parameter class does not cover this possibility, but the rate still distinguishes between estimating the shared subspace in $\mathbb{R}^n$ and separating the shared directions from the individual directions. A full characterization below the current signal threshold would be an interesting next step.

Beyond Gaussian noise. The present finite-sample analysis uses the Gaussian model in two essential ways. First, the one-view leakage isometry relies on the precise behavior of rectangular Gaussian spiked matrices. Second, rotational invariance allows the leakage directions to be symmetrized by an independent Haar rotation, which converts operator-norm bounds that depend on n into bounds that depend on the ranks for the linear leakage terms. The basic bias-correction principle may extend beyond Gaussian noise, but the appropriate correction need not remain a scalar multiple of the identity. Under heteroskedastic, correlated, or anisotropic noise, different directions can have different population leakage even after singular-value rescaling. In such settings, one may need to estimate and subtract a non-scalar leakage operator.

More than two views. With multiple views, the oracle coefficient relations form a higher-dimensional nullspace of a block Gram matrix, and the optimal reconstruction must account for both the signal strengths of the views and the geometry of their individual subspaces. Additional views may improve estimation of the shared subspace in $\mathbb{R}^n$, but it is less clear how they change the error from separating shared and individual directions or the second-order term caused by the individual subspaces. Determining the minimax rate as a function of the number of views, their strengths, and the alignment of their individual subspaces would provide a more complete theory for heterogeneous multi-view integration.

Use of AI

During the preparation of this manuscript, the authors used GPT 5.6 Sol to assist with paper editing, formatting, code writing, and proof writing. All its suggestions were reviewed and revised by the authors, who take full responsibility for the manuscript. The authors have also used GPT 5.6 Sol to formalize the theoretical results in this paper using Lean 4.

Acknowledgement

C.M. was partially supported by the National Science Foundation via the CAREER Award DMS-2443867.

References

Joshua Agterberg. Statistically and computationally optimal estimation and inference of common subspaces. *arXiv preprint arXiv:2606.06483*, 2026. URL https://arxiv.org/abs/2606.06483.

Tavor Z Baharav, Phillip B Nicol, Rafael A Irizarry, and Rong Ma. Stacked SVD or SVD stacked? A random matrix theory perspective on data integration. *arXiv preprint arXiv:2507.22170*, 2025.

Jinho Baik, Gérard Ben Arous, and Sandrine Péché. Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *The Annals of Probability*, 33(5):1643–1697, 2005. ISSN 00911798. URL `http://www.jstor.org/stable/3481698`.

Zhigang Bao, Xiucai Ding, and Ke Wang. Singular vector and singular subspace distribution for the matrix denoising model. *The Annals of Statistics*, 49(1):370–392, 2021. doi: 10.1214/20-AOS1960.

Florent Benaych-Georges and Raj Rao Nadakuditi. The singular values and vectors of low rank perturbations of large rectangular random matrices. *Journal of Multivariate Analysis*, 111:120–135, 2012.

Christopher R. Cabanski, Yuan Qi, Xiaoying Yin, Eric Bair, Michele C. Hayward, Cheng Fan, Jianying Li, Matthew D. Wilkerson, J. S. Marron, Charles M. Perou, and D. Neil Hayes. SWISS MADE: Standardized WithIn class sum of squares to evaluate methodologies and dataset elements. *PLOS ONE*, 5(3):e9905, 2010. doi: 10.1371/journal.pone.0009905.

Changxiao Cai, Gen Li, Yuejie Chi, H. Vincent Poor, and Yuxin Chen. Subspace estimation from unbalanced and incomplete data matrices: $\ell_{2,\infty}$ statistical guarantees. *The Annals of Statistics*, 49(2):944–967, 2021. ISSN 0090-5364. doi: 10.1214/20-aos1986.

T. Tony Cai and Anru Zhang. Rate-optimal perturbation bounds for singular subspaces with applications to high-dimensional statistics. *The Annals of Statistics*, 46(1):60–89, 2018. ISSN 0090-5364. doi: 10.1214/17-AOS1541. URL `https://doi.org/10.1214/17-AOS1541`.

Qing Feng, Meilei Jiang, Jan Hannig, and JS Marron. Angle-based joint and individual variation explained. *Journal of Multivariate Analysis*, 166:241–265, 2018.

Matan Gavish and David L Donoho. Optimal shrinkage of singular values. *IEEE Transactions on Information Theory*, 63(4):2137–2152, 2017.

Irina Gaynanova and Gen Li. Structural learning and integrative decomposition of multi-view data. *Biometrics*, 75(4):1121–1132, 2019.

Yu Gui, Cong Ma, and Zongming Ma. Indiseek learns information-guided disentangled representations. *arXiv preprint arXiv:2509.21584*, 2025.

Iain M Johnstone. On the distribution of the largest eigenvalue in principal components analysis. *The Annals of Statistics*, 29(2):295–327, 2001.

Itamar D. Landau, Gabriel C. Mel, and Surya Ganguli. Singular vectors of sums of rectangular random matrices and optimal estimators of high-rank signals: The extensive spike model. *Physical Review E*, 108:054129, 2023. doi: 10.1103/PhysRevE.108.054129.

Jingyang Li and Zhongyuan Lyu. Spectral joint subspace estimation for heterogeneous multi-view data: Geometry and reweighting. *arXiv preprint arXiv:2512.02866*, 2025.

Xiaoyu Liu, Yiming Liu, Guangming Pan, Lingyue Zhang, and Zhixiang Zhang. Asymptotic limits of spiked eigenvalues and eigenvectors of signal-plus-noise matrices with weak signals and heteroskedastic noise. *Bernoulli*, 31(3):2351–2376, 2025. doi: 10.3150/24-BEJ1808.

Eric F Lock, Katherine A Hoadley, James Stephen Marron, and Andrew B Nobel. Joint and individual variation explained (JIVE) for integrated analysis of multiple data types. *The Annals of Applied Statistics*, 7(1):523–542, 2013.

Zhengchi Ma and Rong Ma. Optimal estimation of shared singular subspaces across multiple noisy matrices. *IEEE Transactions on Information Theory*, 2026.

Olivier Marchal and Julyan Arbel. On the sub-gaussianity of the beta and dirichlet distributions. *Electronic Communications in Probability*, 22(54):1–14, 2017. doi: 10.1214/17-ECP92.

Pascal G. P. Martin, Hervé Guillou, Frédéric Lasserre, Sébastien Déjean, Annaig Lan, Jean-Marc Pascussi, Magali SanCristobal, Philippe Legrand, Philippe Besse, and Thierry Pineau. Novel aspects of PPARα-mediated regulation of lipid and xenobiotic metabolism revealed through a nutrigenomic study. *Hepatology*, 45(3):767–777, 2007. doi: 10.1002/hep.21510.

Raphiel J. Murden, Zhengwu Zhang, Ying Guo, and Benjamin B. Risk. Interpretive JIVE: Connections with CCA and an application to brain connectivity. *Frontiers in Neuroscience*, 16:969510, 2022. doi: 10.3389/fnins.2022.969510.

Raphiel J. Murden, Ganzhong Tian, Deqiang Qiu, Benjamin B. Risk, and Alzheimer's Disease Neuroimaging Initiative. Probabilistic joint and individual variation explained (ProJIVE) for data integration. *Journal of Computational and Graphical Statistics*, 2026. doi: 10.1080/10618600.2026.2639081. Advance online publication.

Raj Rao Nadakuditi. Optshrink: An algorithm for improved low-rank signal matrix denoising by optimal, data-driven singular value shrinkage. *IEEE Transactions on Information Theory*, 60(5):3002–3018, 2014.

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

Jack Prothero, Meilei Jiang, Jan Hannig, Quoc Tran-Dinh, Andrew Ackerman, and JS Marron. Data integration via analysis of subspaces (DIVAS). *TEST*, 33:1–42, 2024. doi: 10.1007/s11749-024-00923-z.

R Sergazinov, A Taeb, and I Gaynanova. A spectral framework for multi-view subspace learning using the product of projections. *Biometrika*, 113(2):asaf088, 2026.

Naichen Shi, Raed Al Kontar, and Salar Fattahi. Heterogeneous matrix factorization: When features differ by datasets. *arXiv preprint arXiv:2305.17744*, 2023.

Yitzchak Shmalo. Mesoscopic rectangular spikes under subspace local laws: Outlier values and singular subspaces. *arXiv preprint arXiv:2608.19594*, 2026.

Runshi Tang, Ming Yuan, and Anru R. Zhang. Mode-wise principal subspace pursuit and matrix spiked covariance model. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 87(1):232–255, 2025. doi: 10.1093/jrsssb/qkae088. URL `https://doi.org/10.1093/jrsssb/qkae088`.

Joel A. Tropp. A comparison principle for functions of a uniformly random subspace. *Probability Theory and Related Fields*, 153(3–4):759–769, 2012. doi: 10.1007/s00440-011-0360-9.

Roman Vershynin. Introduction to the non-asymptotic analysis of random matrices. *arXiv preprint arXiv:1011.3027*, 2010.

Roman Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge University Press, 2018.

Yuepeng Yang and Cong Ma. Estimating shared subspace with AJIVE: the power and limitation of multiple data matrices. *arXiv preprint arXiv:2501.09336*, 2025.

Sangyoon Yi, Raymond Ka Wai Wong, and Irina Gaynanova. Hierarchical nuclear norm penalization for multi-view data integration. *Biometrics*, 79(4):2933–2946, 2023. doi: 10.1111/biom.13893.

Changwon Yoon, Minwoo Kim, Sungkyu Jung, and Jeongyoun Ahn. Joint estimation of high-dimensional spiked covariance matrices via a partially shared subspace. *arXiv preprint arXiv:2607.08123*, 2026. URL https://arxiv.org/abs/2607.08123.

Dongbang Yuan, Yunfeng Zhang, Shuai Guo, Wenyi Wang, and Irina Gaynanova. Exponential canonical correlation analysis with orthogonal variation. *arXiv preprint arXiv:2208.00048*, 2022.

Anru R. Zhang, T. Tony Cai, and Yihong Wu. Heteroskedastic PCA: Algorithm, optimality, and applications. *The Annals of Statistics*, 50(1):53–80, 2022. ISSN 00905364, 21688966. URL https://www.jstor.org/stable/27171694.

Yi Zhao, Lexin Li, and Brian S Caffo. Multimodal neuroimaging data integration and pathway analysis. *Biometrics*, 77(3):879–889, 2021.

Guoxu Zhou, Andrzej Cichocki, Yu Zhang, and Danilo P Mandic. Group component analysis for multiblock data: Common and individual feature extraction. *IEEE Transactions on Neural Networks and Learning Systems*, 27(11):2426–2439, 2015.

A Equivalence of plug-in subspace intersection with AJIVE

Lemma 6. *Let $\widehat{\boldsymbol{\Phi}}_k \in \mathrm{St}(n, p_k)$ for $k = 1, 2$, and set $q \coloneqq p_1 \wedge p_2$. Write*

$$\widehat{\boldsymbol{\Phi}}_1^\top \widehat{\boldsymbol{\Phi}}_2 = \boldsymbol{O}_1 \operatorname{diag}(\zeta_1, \dots, \zeta_q) \boldsymbol{O}_2^\top, \qquad \zeta_1 \ge \cdots \ge \zeta_q \ge 0,$$

where $\boldsymbol{O}_1 \in \mathrm{St}(p_1, q)$ and $\boldsymbol{O}_2 \in \mathrm{St}(p_2, q)$. Suppose $\zeta_r > \zeta_{r+1}$, with $\zeta_{q+1} \coloneqq 0$.

Let $\widehat{\boldsymbol{A}} \coloneqq [\widehat{\boldsymbol{\Phi}}_1, \widehat{\boldsymbol{\Phi}}_2] \in \mathbb{R}^{n \times (p_1 + p_2)}$, and let $\boldsymbol{Q} = (\boldsymbol{Q}_1^\top, \boldsymbol{Q}_2^\top)^\top$ span its r smallest right singular directions. Then

$$\operatorname{col}\big(\widehat{\boldsymbol{\Phi}}_1 \boldsymbol{Q}_1 - \widehat{\boldsymbol{\Phi}}_2 \boldsymbol{Q}_2\big)$$

is exactly the leading r-dimensional eigenspace of $\widehat{\boldsymbol{\Phi}}_1 \widehat{\boldsymbol{\Phi}}_1^\top + \widehat{\boldsymbol{\Phi}}_2 \widehat{\boldsymbol{\Phi}}_2^\top$. Hence the empirical subspace-intersection procedure used in the main text returns the same subspace as two-view AJIVE.

Proof. We first identify the reconstruction from the r smallest right singular directions of $\widehat{\boldsymbol{A}}$, and then show that the resulting directions are exactly the leading AJIVE directions.

Let $\boldsymbol{o}_{1,i}$ and $\boldsymbol{o}_{2,i}$ denote the ith columns of $\boldsymbol{O}_1$ and $\boldsymbol{O}_2$, and set $\widehat{\boldsymbol{a}}_i \coloneqq \widehat{\boldsymbol{\Phi}}_1 o_{1,i}$ and $\widehat{\boldsymbol{b}}_i \coloneqq \widehat{\boldsymbol{\Phi}}_2 \boldsymbol{o}_{2,i}$. Then both vectors have unit norm and $\widehat{\boldsymbol{a}}_i^\top \widehat{\boldsymbol{b}}_i = \zeta_i$.

The Gram matrix of $\widehat{\boldsymbol{A}}$ is

$$\widehat{\boldsymbol{A}}^\top \widehat{\boldsymbol{A}} = \begin{pmatrix} \boldsymbol{I}_{p_1} & \widehat{\boldsymbol{\Phi}}_1^\top \widehat{\boldsymbol{\Phi}}_2 \\ \widehat{\boldsymbol{\Phi}}_2^\top \widehat{\boldsymbol{\Phi}}_1 & \boldsymbol{I}_{p_2} \end{pmatrix},$$

and, for every $i \le q$,

$$\widehat{\boldsymbol{A}}^\top \widehat{\boldsymbol{A}} \frac{1}{\sqrt{2}} \begin{pmatrix} \boldsymbol{o}_{1,i} \\ -\boldsymbol{o}_{2,i} \end{pmatrix} = (1 - \zeta_i) \frac{1}{\sqrt{2}} \begin{pmatrix} \boldsymbol{o}_{1,i} \\ -\boldsymbol{o}_{2,i} \end{pmatrix}.$$

Since $\zeta_1 \ge \cdots \ge \zeta_q$ and $\zeta_r > \zeta_{r+1}$, these vectors span the r smallest right singular directions of $\widehat{\boldsymbol{A}}$. Since only the reconstructed column space matters, we may therefore take $\boldsymbol{Q}_1 = 2^{-1/2}[\boldsymbol{o}_{1,1}, \dots, \boldsymbol{o}_{1,r}]$ and $\boldsymbol{Q}_2 = -2^{-1/2}[\boldsymbol{o}_{2,1}, \dots, \boldsymbol{o}_{2,r}]$, which gives

$$\operatorname{col}\big(\widehat{\boldsymbol{\Phi}}_1 \boldsymbol{Q}_1 - \widehat{\boldsymbol{\Phi}}_2 \boldsymbol{Q}_2\big) = \operatorname{span}\Big\{\widehat{\boldsymbol{a}}_i + \widehat{\boldsymbol{b}}_i : i = 1, \dots, r\Big\}.$$

It remains to identify this span with the leading AJIVE eigenspace. Since $\widehat{\boldsymbol{\Phi}}_1^\top \widehat{\boldsymbol{b}}_i = \zeta_i \boldsymbol{o}_{1,i}$ and $\widehat{\boldsymbol{\Phi}}_2^\top \widehat{\boldsymbol{a}}_i = \zeta_i \boldsymbol{o}_{2,i}$,

$$\Big(\widehat{\boldsymbol{\Phi}}_1 \widehat{\boldsymbol{\Phi}}_1^\top + \widehat{\boldsymbol{\Phi}}_2 \widehat{\boldsymbol{\Phi}}_2^\top\Big)(\widehat{\boldsymbol{a}}_i \pm \widehat{\boldsymbol{b}}_i) = (1 \pm \zeta_i)(\widehat{\boldsymbol{a}}_i \pm \widehat{\boldsymbol{b}}_i).$$

Any unpaired directions have eigenvalue 1. Hence $\zeta_1 \ge \cdots \ge \zeta_q$ and $\zeta_r > \zeta_{r+1}$ imply that the leading r-dimensional eigenspace is precisely $\operatorname{span}\{\widehat{\boldsymbol{a}}_i + \widehat{\boldsymbol{b}}_i : i = 1, \dots, r\}$. Thus the two procedures return the same r-dimensional subspace. □

B Deterministic ingredients for the upper bound

The deterministic argument proceeds in four stages. We first use the population geometry to establish curvature of the oracle Gram matrix and to construct the transfer operator that describes how individual signal directions enter the reconstructed shared subspace. We then apply a relative perturbation argument to isolate the r-dimensional near-zero eigenspace without imposing an upper bound on the signal strengths. Since an orthonormal basis of this eigenspace is determined only up to a right orthogonal factor, the calibration step aligns it with the oracle normalization. Finally, the calibrated eigenvalue equation separates the subspace error into its linear leakage contribution and four higher-order remainder terms.

B.1 Population and oracle geometry

B.1.1 Geometry of the joint signal spaces

The oracle bounds below require a lower bound on the nonzero spectrum of $\boldsymbol{P}_{\mathcal{M}_1} + \boldsymbol{P}_{\mathcal{M}_2}$ and an upper bound on the trace of its pseudoinverse. Both follow from the principal-angle decomposition of the two individual signal spaces. Let $\chi_j = \sigma_j(\boldsymbol{V}_1^\top \boldsymbol{V}_2)$ for $j = 1, \ldots, r_1 \wedge r_2$. These singular values satisfy $1 \geq \chi_j \geq 0$, and (10) gives $\chi_1 \leq 1 - 2\theta$.

Lemma 7. *The operator $\boldsymbol{P}_{\mathcal{M}_1} + \boldsymbol{P}_{\mathcal{M}_2}$ has support $\mathcal{M}_1 + \mathcal{M}_2$ and rank $r + r_1 + r_2$. Its nonzero eigenvalues are*

$$\underbrace{2, \ldots, 2}_{r \text{ times}}, \qquad \underbrace{1 + \chi_j,\ 1 - \chi_j}_{j=1,\ldots,r_1 \wedge r_2}, \qquad \underbrace{1, \ldots, 1}_{|r_1 - r_2| \text{ times}}. \tag{35}$$

Consequently,

$$\lambda_{\min}^{+}\big(\boldsymbol{P}_{\mathcal{M}_1} + \boldsymbol{P}_{\mathcal{M}_2}\big) \geq 2\theta, \tag{36}$$

and

$$\operatorname{tr}\Big\{\big(\boldsymbol{P}_{\mathcal{M}_1} + \boldsymbol{P}_{\mathcal{M}_2}\big)^{\dagger}\Big\} \lesssim r + r_1 + r_2 + \frac{r_1 \wedge r_2}{\theta}. \tag{37}$$

Proof. Define $q := r_1 \wedge r_2$. Choose orthonormal principal-vector pairs $\widetilde{\boldsymbol{v}}_{1,1}, \ldots, \widetilde{\boldsymbol{v}}_{1,q} \in \operatorname{col}(\boldsymbol{V}_1)$ and $\widetilde{\boldsymbol{v}}_{2,1}, \ldots, \widetilde{\boldsymbol{v}}_{2,q} \in \operatorname{col}(\boldsymbol{V}_2)$ such that

$$\widetilde{\boldsymbol{v}}_{1,i}^\top \widetilde{\boldsymbol{v}}_{2,j} = \chi_j \mathbf{1}\{i = j\}.$$

These vectors are obtained from the left and right singular vectors of $\boldsymbol{V}_1^\top \boldsymbol{V}_2$ and may be completed to orthonormal bases of the two individual subspaces.

Since $\mathcal{M}_k = \mathcal{U} \oplus \operatorname{col}(\boldsymbol{V}_k)$,

$$\boldsymbol{P}_{\mathcal{M}_1} + \boldsymbol{P}_{\mathcal{M}_2} = 2\boldsymbol{P}_{\mathcal{U}} + \boldsymbol{V}_1 \boldsymbol{V}_1^\top + \boldsymbol{V}_2 \boldsymbol{V}_2^\top.$$

The operator therefore has eigenvalue 2 on $\mathcal{U}$. For each $j \leq q$, the principal plane $\operatorname{span}\{\widetilde{\boldsymbol{v}}_{1,j}, \widetilde{\boldsymbol{v}}_{2,j}\}$ is invariant, and

$$\big(\boldsymbol{P}_{\mathcal{M}_1} + \boldsymbol{P}_{\mathcal{M}_2}\big)\big(\widetilde{\boldsymbol{v}}_{1,j} \pm \widetilde{\boldsymbol{v}}_{2,j}\big) = (1 \pm \chi_j)\big(\widetilde{\boldsymbol{v}}_{1,j} \pm \widetilde{\boldsymbol{v}}_{2,j}\big).$$

If $r_1 \neq r_2$, the remaining $|r_1 - r_2|$ basis vectors in the larger individual subspace are orthogonal to the other individual subspace, so each has eigenvalue 1. All the displayed eigenvalues are positive because $\chi_j \leq 1 - 2\theta < 1$ for every j, while the operator vanishes on $(\mathcal{M}_1 + \mathcal{M}_2)^\perp$. This proves (35) and shows that the support is $\mathcal{M}_1 + \mathcal{M}_2$. Its rank is therefore $r + r_1 + r_2 = p_1 + p_2 - r$.

By (10), $\chi_j \leq 1 - 2\theta$ for every j. Hence $1 - \chi_j \geq 2\theta$, while $1 \geq 2\theta$ and $2 \geq 2\theta$ because $\theta \leq 1/2$. This proves (36).

Finally, summing the reciprocals of the eigenvalues in (35) gives

$$\begin{aligned}
\operatorname{tr}\Big\{\big(\boldsymbol{P}_{\mathcal{M}_1} + \boldsymbol{P}_{\mathcal{M}_2}\big)^{\dagger}\Big\} &= \frac{r}{2} + |r_1 - r_2| + \sum_{j=1}^{q} \left(\frac{1}{1 + \chi_j} + \frac{1}{1 - \chi_j}\right) \\
&\leq \frac{r}{2} + |r_1 - r_2| + q + \frac{q}{2\theta} \lesssim r + r_1 + r_2 + \frac{r_1 \wedge r_2}{\theta}.
\end{aligned}$$

Here we used $1 + \chi_j \geq 1$, $1 - \chi_j \geq 2\theta$, and $q = r_1 \wedge r_2$. This proves (37). □

B.1.2 Geometry of the oracle Gram matrix

Proof of Lemma 1. The proof has three parts. We first identify the nullspace of the oracle Gram matrix, then verify the normalization identities for $\boldsymbol{Q}_\star$, and finally establish the curvature bound on the orthogonal complement of the nullspace.

Nullspace of the oracle Gram. Since $\boldsymbol{G}_0 = \boldsymbol{S}^\top \boldsymbol{S}$, we have $\boldsymbol{z}^\top \boldsymbol{G}_0 \boldsymbol{z} = \|\boldsymbol{S}\boldsymbol{z}\|_2^2$, and hence $\ker(\boldsymbol{G}_0) = \ker(\boldsymbol{S})$. On $\mathcal{E}_{\mathrm{emb}}$, $\mathrm{col}(\boldsymbol{S}) = \mathcal{M}_1 + \mathcal{M}_2$. Since $\mathcal{M}_1 \cap \mathcal{M}_2 = \mathcal{U}$, we have $\mathrm{rank}(\boldsymbol{S}) = \dim(\mathcal{M}_1 + \mathcal{M}_2) = p_1 + p_2 - r$. As $\boldsymbol{S}$ has $p_1 + p_2$ columns, rank–nullity gives $\dim\ker(\boldsymbol{G}_0) = \dim\ker(\boldsymbol{S}) = r$.

Proof of (23). Next consider $\boldsymbol{Q}_\star$. Since $\mathcal{U} \subseteq \mathrm{col}(\boldsymbol{S}_k)$, we have $\boldsymbol{S}_k \boldsymbol{S}_k^\dagger \boldsymbol{U} = \boldsymbol{U}$ for $k = 1, 2$. Therefore $\boldsymbol{S}\boldsymbol{Q}_\star = \boldsymbol{0}$ and $\boldsymbol{S}\boldsymbol{J}\boldsymbol{Q}_\star = (\alpha_1 + \alpha_2)\boldsymbol{U} = \boldsymbol{U}$, proving (23).

Proof of (24). It remains to lower-bound the positive spectrum of $\boldsymbol{G}_0$. On $\mathcal{E}_{\mathrm{emb}}$, $\mathrm{col}(\boldsymbol{S}_k) = \mathcal{M}_k$ and $\sigma_{\min}(\boldsymbol{S}_k) \geq c_{\mathrm{emb}}\gamma_k \geq c_{\mathrm{emb}}\gamma_{\min}$. Thus, for every $\boldsymbol{x} \in \mathbb{R}^n$,

$$\boldsymbol{x}^\top \boldsymbol{S}_k \boldsymbol{S}_k^\top \boldsymbol{x} = \|\boldsymbol{S}_k^\top \boldsymbol{P}_{\mathcal{M}_k} \boldsymbol{x}\|_2^2 \geq (c_{\mathrm{emb}}\gamma_{\min})^2 \|\boldsymbol{P}_{\mathcal{M}_k}\boldsymbol{x}\|_2^2,$$

and hence

$$\boldsymbol{S}\boldsymbol{S}^\top \succeq (c_{\mathrm{emb}}\gamma_{\min})^2 \big(\boldsymbol{P}_{\mathcal{M}_1} + \boldsymbol{P}_{\mathcal{M}_2}\big). \tag{38}$$

Since $\boldsymbol{G}_0 = \boldsymbol{S}^\top \boldsymbol{S}$ and $\boldsymbol{S}\boldsymbol{S}^\top$ have the same nonzero eigenvalues, (36) and (38) yield

$$\lambda_{\min}^+(\boldsymbol{G}_0) = \lambda_{\min}^+(\boldsymbol{S}\boldsymbol{S}^\top) \geq 2c_{\mathrm{emb}}^2 \gamma_{\min}^2 \theta,$$

which proves (24).

The nullspace calculation, (23), and the curvature bound therefore establish all three claims in Lemma 1. This completes the proof. □

B.1.3 Population norm bounds

We collect several consequences of the oracle geometry that will be used repeatedly in the probabilistic analysis. Write

$$p_{\min} := p_1 \wedge p_2 = r + (r_1 \wedge r_2), \qquad p_{\max} := p_1 \vee p_2 = r + (r_1 \vee r_2).$$

Lemma 8. *For $k = 1, 2$, define $\boldsymbol{R}_k := \boldsymbol{S}_k^\dagger \boldsymbol{U} \in \mathbb{R}^{p_k \times r}$. On $\mathcal{E}_{\mathrm{emb}}$, the following bounds hold.*

$$\|\boldsymbol{R}_k\|_{\mathrm{op}} \lesssim \frac{1}{\gamma_k}, \qquad \|\boldsymbol{R}_k\|_{\mathrm{F}} \lesssim \frac{\sqrt{r}}{\gamma_k}, \qquad k = 1, 2, \tag{39a}$$

$$\left\|\boldsymbol{S}(\boldsymbol{G}_0^\dagger)^{1/2}\right\|_{\mathrm{op}} \leq 1, \qquad \left\|\boldsymbol{S}(\boldsymbol{G}_0^\dagger)^{1/2}\right\|_{\mathrm{F}} = \sqrt{p_1 + p_2 - r} \lesssim \sqrt{p_{\max}}, \tag{39b}$$

$$\left\|(\boldsymbol{G}_0^\dagger)^{1/2}\right\|_{\mathrm{op}} \lesssim \frac{1}{\gamma_{\min}\sqrt{\theta}}, \qquad \left\|(\boldsymbol{G}_0^\dagger)^{1/2}\right\|_{\mathrm{F}} \lesssim \frac{1}{\gamma_{\min}} \left(\sqrt{r_1 + r_2} + \frac{\sqrt{p_{\min}}}{\sqrt{\theta}}\right). \tag{39c}$$

Proof. For (39a), on $\mathcal{E}_{\mathrm{emb}}$ we have $\mathrm{col}(\boldsymbol{S}_k) = \mathcal{M}_k$, $\mathcal{U} \subseteq \mathcal{M}_k$, and $\sigma_{\min}(\boldsymbol{S}_k) \geq c_{\mathrm{emb}}\gamma_k$. Hence $\|\boldsymbol{R}_k\|_{\mathrm{op}} \leq \|\boldsymbol{S}_k^\dagger\|_{\mathrm{op}} \lesssim \gamma_k^{-1}$, and the Frobenius bound follows from $\mathrm{rank}(\boldsymbol{R}_k) \leq r$.

For (39b), observe that

$$\big(\boldsymbol{S}(\boldsymbol{G}_0^\dagger)^{1/2}\big)^\top \boldsymbol{S}(\boldsymbol{G}_0^\dagger)^{1/2} = (\boldsymbol{G}_0^\dagger)^{1/2}\boldsymbol{G}_0(\boldsymbol{G}_0^\dagger)^{1/2} = \boldsymbol{I} - \boldsymbol{P}_{\ker(\boldsymbol{G}_0)}.$$

By Lemma 1, $\mathrm{rank}(\boldsymbol{P}_{\ker(\boldsymbol{G}_0)}) = r$, and therefore $\mathrm{rank}(\boldsymbol{I} - \boldsymbol{P}_{\ker(\boldsymbol{G}_0)}) = p_1 + p_2 - r$. This proves (39b).

For $(\boldsymbol{G}_0^\dagger)^{1/2}$ in (39c), we have $\left\|(\boldsymbol{G}_0^\dagger)^{1/2}\right\|_{\mathrm{op}} = 1/\sqrt{\lambda_{\min}^+(\boldsymbol{G}_0)}$. This together with Lemma 1 yields the operator norm bound. To control the Frobenius norm, observe that, on $\mathcal{E}_{\mathrm{emb}}$, both $\boldsymbol{S}\boldsymbol{S}^\top$ and $\boldsymbol{P}_{\mathcal{M}_1} + \boldsymbol{P}_{\mathcal{M}_2}$ have support $\mathcal{M}_1 + \mathcal{M}_2$. We may therefore restrict (38) to this common support, where both operators are positive definite, and invert the inequality. Extending the inverses by zero on $(\mathcal{M}_1+\mathcal{M}_2)^\perp$ gives

$$(\boldsymbol{S}\boldsymbol{S}^\top)^\dagger \preceq \frac{1}{c_{\mathrm{emb}}^2\gamma_{\min}^2}\big(\boldsymbol{P}_{\mathcal{M}_1}+\boldsymbol{P}_{\mathcal{M}_2}\big)^\dagger.$$

Since $\boldsymbol{G}_0=\boldsymbol{S}^\top\boldsymbol{S}$ and $\boldsymbol{S}\boldsymbol{S}^\top$ have the same nonzero eigenvalues, (37) now yields

$$\begin{aligned}\left\|(\boldsymbol{G}_0^\dagger)^{1/2}\right\|_{\mathrm{F}}^2 &= \operatorname{tr}(\boldsymbol{G}_0^\dagger)=\operatorname{tr}\{(\boldsymbol{S}\boldsymbol{S}^\top)^\dagger\}\\ &\lesssim \frac{1}{\gamma_{\min}^2}\left(r+r_1+r_2+\frac{r_1\wedge r_2}{\theta}\right)\lesssim \frac{1}{\gamma_{\min}^2}\left(r_1+r_2+\frac{p_{\min}}{\theta}\right).\end{aligned}$$

The last inequality uses $p_{\min}=r+(r_1\wedge r_2)$ and $\theta\le 1/2$. Taking square roots and using $\sqrt{x+y}\le\sqrt{x}+\sqrt{y}$ proves the Frobenius bound in (39c). □

B.2 Relative perturbation of the nullspace

Proof of Lemma 2. The proof has three parts. We first obtain a lower bound for the entire spectrum of $\boldsymbol{G}$. We then use the min–max principle to separate the r eigenvalues near zero from the remaining spectrum. Finally, we project the invariant-subspace equation onto $\ker(\boldsymbol{G}_0)^\perp$ to control the relative residual.

Choose orthonormal bases $\boldsymbol{Q}_0\in\mathrm{St}(d,r)$ and $\boldsymbol{Q}_+\in\mathrm{St}(d,d-r)$ for $\ker(\boldsymbol{G}_0)$ and $\ker(\boldsymbol{G}_0)^\perp$, respectively, and define the positive spectral block $\boldsymbol{K}:=\boldsymbol{Q}_+^\top\boldsymbol{G}_0\boldsymbol{Q}_+$. Then $\boldsymbol{K}\succ\boldsymbol{0}$ and $\lambda_{\min}(\boldsymbol{K})=\lambda_{\min}^+(\boldsymbol{G}_0)$. We identify the relative perturbation operators with their nonzero coordinate blocks,

$$\boldsymbol{\Delta}_{00}=\boldsymbol{Q}_0^\top\boldsymbol{\Delta}\boldsymbol{Q}_0,\qquad \boldsymbol{\Delta}_{+0}=\boldsymbol{K}^{-1/2}\boldsymbol{Q}_+^\top\boldsymbol{\Delta}\boldsymbol{Q}_0,\qquad \boldsymbol{\Delta}_{++}=\boldsymbol{K}^{-1/2}\boldsymbol{Q}_+^\top\boldsymbol{\Delta}\boldsymbol{Q}_+\boldsymbol{K}^{-1/2}.$$

This identification preserves their operator norms. The assumptions of Lemma 2 imply

$$\|\boldsymbol{\Delta}_{++}\|_{\mathrm{op}}\le\frac{1}{4},\qquad \varepsilon_\Delta\le\frac{1}{16}\lambda_{\min}^+(\boldsymbol{G}_0). \tag{40}$$

Lower bound for the spectrum. Write any $\boldsymbol{w}\in\mathbb{R}^d$ uniquely as $\boldsymbol{w}=\boldsymbol{Q}_0\boldsymbol{u}+\boldsymbol{Q}_+\boldsymbol{v}$, and set $\boldsymbol{t}:=\boldsymbol{K}^{1/2}\boldsymbol{v}$. In these coordinates,

$$\boldsymbol{w}^\top\boldsymbol{G}\boldsymbol{w}=\|\boldsymbol{t}\|_2^2+\boldsymbol{t}^\top\boldsymbol{\Delta}_{++}\boldsymbol{t}+2\boldsymbol{t}^\top\boldsymbol{\Delta}_{+0}\boldsymbol{u}+\boldsymbol{u}^\top\boldsymbol{\Delta}_{00}\boldsymbol{u}. \tag{41}$$

By (40), the first two terms are at least $\frac{1}{2}\|\boldsymbol{t}\|_2^2$. Young's inequality gives $2|\boldsymbol{t}^\top\boldsymbol{\Delta}_{+0}\boldsymbol{u}|\le\frac{1}{2}\|\boldsymbol{t}\|_2^2+2\|\boldsymbol{\Delta}_{+0}\|_{\mathrm{op}}^2\|\boldsymbol{u}\|_2^2$. Consequently,

$$\boldsymbol{w}^\top\boldsymbol{G}\boldsymbol{w}\ge-\left(\|\boldsymbol{\Delta}_{00}\|_{\mathrm{op}}+2\,\|\boldsymbol{\Delta}_{+0}\|_{\mathrm{op}}^2\right)\|\boldsymbol{u}\|_2^2\ge-\varepsilon_\Delta\|\boldsymbol{w}\|_2^2.$$

It follows that $\lambda_{\min}(\boldsymbol{G})\ge-\varepsilon_\Delta$.

Separation of the near-zero spectrum. Order the eigenvalues as $\lambda_1(\boldsymbol{G})\ge\cdots\ge\lambda_d(\boldsymbol{G})$. On $\ker(\boldsymbol{G}_0)=\mathrm{col}(\boldsymbol{Q}_0)$, the quadratic form of $\boldsymbol{G}$ is $\boldsymbol{u}^\top\boldsymbol{\Delta}_{00}\boldsymbol{u}$. Courant–Fischer therefore gives

$$\lambda_{d-r+1}(\boldsymbol{G})\le\|\boldsymbol{\Delta}_{00}\|_{\mathrm{op}}\le\varepsilon_\Delta.$$

Thus at least r eigenvalues are at most ε_Δ.

To control the remaining spectrum, apply Courant–Fischer on $\ker(\boldsymbol{G}_0)^\perp = \mathrm{col}(\boldsymbol{Q}_+)$. In this case $\boldsymbol{u} = \mathbf{0}$, and (41) gives

$$\boldsymbol{w}^\top \boldsymbol{G} \boldsymbol{w} \geq \lambda^+_{\min}(\boldsymbol{G}_0)\big(1 - \|\boldsymbol{\Delta}_{++}\|_{\mathrm{op}}\big)\|\boldsymbol{w}\|_2^2.$$

By (40), the coefficient on the right is at least $12\varepsilon_\Delta$, which is strictly larger than $4\varepsilon_\Delta$. Hence $\lambda_{d-r}(\boldsymbol{G}) > 4\varepsilon_\Delta$. Combining this separation with the preceding lower bound shows that exactly r eigenvalues lie in $[-\varepsilon_\Delta, \varepsilon_\Delta]$, while all remaining eigenvalues exceed $4\varepsilon_\Delta$.

Bound for $\boldsymbol{G}_0^{1/2}\boldsymbol{Q}$. Let $\boldsymbol{Q} \in \mathrm{St}(d, r)$ span the invariant subspace associated with the r eigenvalues in $[-\varepsilon_\Delta, \varepsilon_\Delta]$, and set $\boldsymbol{\Lambda} \coloneqq \boldsymbol{Q}^\top \boldsymbol{G}\boldsymbol{Q}$. Then $\boldsymbol{G}\boldsymbol{Q} = \boldsymbol{Q}\boldsymbol{\Lambda}$ and $\|\boldsymbol{\Lambda}\|_{\mathrm{op}} \leq \varepsilon_\Delta$.

To express the desired residual in the positive spectral coordinates, define $\boldsymbol{H} \coloneqq \boldsymbol{K}^{1/2}\boldsymbol{Q}_+^\top \boldsymbol{Q}$. Projecting the invariant-subspace equation onto $\mathrm{col}(\boldsymbol{Q}_+)$ and multiplying by $\boldsymbol{K}^{-1/2}$ gives

$$\boldsymbol{H} = -\boldsymbol{\Delta}_{+0}\boldsymbol{Q}_0^\top \boldsymbol{Q} - \boldsymbol{\Delta}_{++}\boldsymbol{H} + \boldsymbol{K}^{-1}\boldsymbol{H}\boldsymbol{\Lambda}. \tag{42}$$

Taking operator norms, using $\|\boldsymbol{Q}_0^\top \boldsymbol{Q}\|_{\mathrm{op}} \leq 1$, and recalling that $\|\boldsymbol{K}^{-1}\|_{\mathrm{op}} = 1/\lambda^+_{\min}(\boldsymbol{G}_0)$, we obtain

$$\|\boldsymbol{H}\|_{\mathrm{op}} \leq \|\boldsymbol{\Delta}_{+0}\|_{\mathrm{op}} + \left(\|\boldsymbol{\Delta}_{++}\|_{\mathrm{op}} + \frac{\|\boldsymbol{\Lambda}\|_{\mathrm{op}}}{\lambda^+_{\min}(\boldsymbol{G}_0)}\right)\|\boldsymbol{H}\|_{\mathrm{op}}.$$

Since $\|\boldsymbol{\Lambda}\|_{\mathrm{op}} \leq \varepsilon_\Delta$, the factor multiplying $\|\boldsymbol{H}\|_{\mathrm{op}}$ is at most $1/2$ by (40). Therefore $\|\boldsymbol{H}\|_{\mathrm{op}} \leq 2\|\boldsymbol{\Delta}_{+0}\|_{\mathrm{op}}$. Finally,

$$\left\|\boldsymbol{G}_0^{1/2}\boldsymbol{Q}\right\|_{\mathrm{op}} = \left\|\boldsymbol{K}^{1/2}\boldsymbol{Q}_+^\top \boldsymbol{Q}\right\|_{\mathrm{op}} = \|\boldsymbol{H}\|_{\mathrm{op}} \leq 2\,\|\boldsymbol{\Delta}_{+0}\|_{\mathrm{op}}.$$

This is exactly the bound in (26), so the result follows. □

B.3 Calibration and deterministic error decomposition

B.3.1 Calibration

The relative perturbation argument controls the residual $\boldsymbol{S}\boldsymbol{Q}$. To calibrate $\boldsymbol{Q}$, we first show that this residual controls the coefficient error once the shared component has been fixed.

Lemma 9. *On $\mathcal{E}_{\mathrm{emb}}$, let $\boldsymbol{Z}$ be any matrix satisfying $\boldsymbol{U}^\top \boldsymbol{S}\boldsymbol{J}\boldsymbol{Z} = \mathbf{0}$. Then*

$$\|\boldsymbol{Z}\|_{\mathrm{op}} \leq \frac{\|\boldsymbol{S}\boldsymbol{Z}\|_{\mathrm{op}}}{c_{\mathrm{emb}}\gamma_{\min}\sqrt{2\theta}}. \tag{43}$$

Proof. It suffices to prove the corresponding vector inequality. Let $\boldsymbol{z} = (\boldsymbol{z}_1^\top, \boldsymbol{z}_2^\top)^\top$ satisfy $\boldsymbol{U}^\top \boldsymbol{S}\boldsymbol{J}\boldsymbol{z} = \mathbf{0}$. Since $\mathrm{col}(\boldsymbol{S}_k) = \mathcal{M}_k$ on $\mathcal{E}_{\mathrm{emb}}$, write $\boldsymbol{S}_k\boldsymbol{z}_k = \boldsymbol{U}\boldsymbol{a}_k + \boldsymbol{V}_k\boldsymbol{b}_k$. The constraint implies $\alpha_1\boldsymbol{a}_1 = \alpha_2\boldsymbol{a}_2$, so there is a vector $\boldsymbol{a}$ such that $\boldsymbol{a}_1 = \alpha_2\boldsymbol{a}$ and $\boldsymbol{a}_2 = \alpha_1\boldsymbol{a}$. Consequently, $\|\boldsymbol{a}_1 + \boldsymbol{a}_2\|_2^2 = \|\boldsymbol{a}\|_2^2$ and, since $\alpha_1^2 + \alpha_2^2 \leq 1$ and $2\theta \leq 1$,

$$\|\boldsymbol{a}_1 + \boldsymbol{a}_2\|_2^2 \geq 2\theta\big(\|\boldsymbol{a}_1\|_2^2 + \|\boldsymbol{a}_2\|_2^2\big).$$

Restricting (36) to $\mathcal{U}^\perp$ gives the same coercive principal-angle bound on the two individual signal spaces. Applied to the coordinate pair $(\boldsymbol{b}_1, \boldsymbol{b}_2)$, it yields $\|\boldsymbol{V}_1\boldsymbol{b}_1 + \boldsymbol{V}_2\boldsymbol{b}_2\|_2^2 \geq 2\theta(\|\boldsymbol{b}_1\|_2^2 + \|\boldsymbol{b}_2\|_2^2)$.

Because $\boldsymbol{U}$ is orthogonal to $\boldsymbol{V}_1$ and $\boldsymbol{V}_2$, it follows that

$$\|\boldsymbol{S}\boldsymbol{z}\|_2^2 = \|\boldsymbol{a}_1 + \boldsymbol{a}_2\|_2^2 + \|\boldsymbol{V}_1\boldsymbol{b}_1 + \boldsymbol{V}_2\boldsymbol{b}_2\|_2^2$$

$$
\begin{aligned}
&\geq 2\theta \sum_{k=1}^{2} \left(\|\boldsymbol{a}_k\|_2^2 + \|\boldsymbol{b}_k\|_2^2 \right) \\
&= 2\theta \sum_{k=1}^{2} \|\boldsymbol{S}_k \boldsymbol{z}_k\|_2^2 \geq 2\theta (c_{\mathrm{emb}} \gamma_{\min})^2 \|\boldsymbol{z}\|_2^2.
\end{aligned}
$$

Thus $\|\boldsymbol{z}\|_2 \leq \|\boldsymbol{S}\boldsymbol{z}\|_2 / (c_{\mathrm{emb}} \gamma_{\min} \sqrt{2\theta})$. Applying this inequality to $\boldsymbol{z} = \boldsymbol{Z}\boldsymbol{x}$ and taking the supremum over $\|\boldsymbol{x}\|_2 = 1$ proves (43). $\square$

We now use this stability estimate to align the empirical near-zero eigenspace with the oracle normalization. Define

$$
\boldsymbol{\Xi} := \boldsymbol{U}^\top \boldsymbol{M} = \boldsymbol{U}^\top \boldsymbol{S}\boldsymbol{J}\boldsymbol{Q}.
$$

This matrix is the change of basis required to impose the normalization $\boldsymbol{S}\boldsymbol{J}\boldsymbol{Q}_\star = \boldsymbol{U}$. Whenever $\boldsymbol{\Xi}$ is invertible, define

$$
\widetilde{\boldsymbol{Q}} := \boldsymbol{Q}\boldsymbol{\Xi}^{-1}, \qquad \boldsymbol{\Delta}_{\boldsymbol{Q}} := \widetilde{\boldsymbol{Q}} - \boldsymbol{Q}_\star.
$$

Lemma 10. *On $\mathcal{E}_{\mathrm{good}}$, the matrix $\boldsymbol{\Xi}$ is invertible, and*

$$
\left\|\boldsymbol{\Xi}^{-1}\right\|_{\mathrm{op}} \lesssim \frac{1}{\gamma_{\min}}, \qquad \|\boldsymbol{S}\boldsymbol{\Delta}_{\boldsymbol{Q}}\|_{\mathrm{op}} \lesssim \frac{\|\boldsymbol{\Delta}_{+0}\|_{\mathrm{op}}}{\gamma_{\min}}, \qquad \|\boldsymbol{\Delta}_{\boldsymbol{Q}}\|_{\mathrm{op}} \lesssim \frac{\|\boldsymbol{\Delta}_{+0}\|_{\mathrm{op}}}{\gamma_{\min}^2 \sqrt{\theta}}.
$$

Proof. Set $\boldsymbol{Z} := \boldsymbol{Q} - \boldsymbol{Q}_\star \boldsymbol{\Xi}$. The identities $\boldsymbol{S}\boldsymbol{Q}_\star = \boldsymbol{0}$ and $\boldsymbol{S}\boldsymbol{J}\boldsymbol{Q}_\star = \boldsymbol{U}$ imply $\boldsymbol{U}^\top \boldsymbol{S}\boldsymbol{J}\boldsymbol{Z} = \boldsymbol{0}$ and $\boldsymbol{S}\boldsymbol{Z} = \boldsymbol{S}\boldsymbol{Q}$. Hence (27) and (43) and the definition of $\mathcal{E}_{\mathrm{good}}$ give

$$
\|\boldsymbol{Z}\|_{\mathrm{op}} \leq \frac{2\,\|\boldsymbol{\Delta}_{+0}\|_{\mathrm{op}}}{c_{\mathrm{emb}} \gamma_{\min} \sqrt{2\theta}} \leq \frac{1}{2}.
$$

This bound implies that $\boldsymbol{\Xi}$ is invertible. Indeed, if $\boldsymbol{\Xi}\boldsymbol{x} = \boldsymbol{0}$, then $\boldsymbol{Q}\boldsymbol{x} = \boldsymbol{Z}\boldsymbol{x}$. Since $\boldsymbol{Q}$ has orthonormal columns, and $\|\boldsymbol{Z}\|_{\mathrm{op}} \leq \frac{1}{2}$, we have $\|\boldsymbol{x}\|_2 \leq \frac{1}{2}\|\boldsymbol{x}\|_2$, and therefore $\boldsymbol{x} = \boldsymbol{0}$.

The identity $\boldsymbol{Q} = \boldsymbol{Q}_\star \boldsymbol{\Xi} + \boldsymbol{Z}$ now gives $\boldsymbol{Q}\boldsymbol{\Xi}^{-1} = \boldsymbol{Q}_\star + \boldsymbol{Z}\boldsymbol{\Xi}^{-1}$. Since $\boldsymbol{Q}$ is an isometry,

$$
\left\|\boldsymbol{\Xi}^{-1}\right\|_{\mathrm{op}} = \left\|\boldsymbol{Q}\boldsymbol{\Xi}^{-1}\right\|_{\mathrm{op}} \leq \|\boldsymbol{Q}_\star\|_{\mathrm{op}} + \frac{1}{2} \left\|\boldsymbol{\Xi}^{-1}\right\|_{\mathrm{op}}.
$$

By the block definition of $\boldsymbol{Q}_\star$ and (39a),

$$
\|\boldsymbol{Q}_\star\|_{\mathrm{op}}^2 = \left\|\boldsymbol{R}_1^\top \boldsymbol{R}_1 + \boldsymbol{R}_2^\top \boldsymbol{R}_2\right\|_{\mathrm{op}} \leq \|\boldsymbol{R}_1\|_{\mathrm{op}}^2 + \|\boldsymbol{R}_2\|_{\mathrm{op}}^2 \lesssim \frac{1}{\gamma_{\min}^2}.
$$

Thus $\left\|\boldsymbol{\Xi}^{-1}\right\|_{\mathrm{op}} \leq 2\,\|\boldsymbol{Q}_\star\|_{\mathrm{op}} \lesssim \gamma_{\min}^{-1}$.

Finally, $\boldsymbol{\Delta}_{\boldsymbol{Q}} = \boldsymbol{Z}\boldsymbol{\Xi}^{-1}$. Since $\boldsymbol{S}\boldsymbol{Z} = \boldsymbol{S}\boldsymbol{Q}$, the relative nullspace bound yields

$$
\|\boldsymbol{S}\boldsymbol{\Delta}_{\boldsymbol{Q}}\|_{\mathrm{op}} \leq \|\boldsymbol{S}\boldsymbol{Q}\|_{\mathrm{op}} \left\|\boldsymbol{\Xi}^{-1}\right\|_{\mathrm{op}} \lesssim \frac{\|\boldsymbol{\Delta}_{+0}\|_{\mathrm{op}}}{\gamma_{\min}}.
$$

The first display in this proof already gives $\|\boldsymbol{Z}\|_{\mathrm{op}} \lesssim \|\boldsymbol{\Delta}_{+0}\|_{\mathrm{op}} / (\gamma_{\min}\sqrt{\theta})$. Combining this estimate with $\boldsymbol{\Delta}_{\boldsymbol{Q}} = \boldsymbol{Z}\boldsymbol{\Xi}^{-1}$ and $\left\|\boldsymbol{\Xi}^{-1}\right\|_{\mathrm{op}} \lesssim \gamma_{\min}^{-1}$ proves the final bound in Lemma 10. This completes the proof. $\square$

B.3.2 The transfer operator

Recall the transfer operator defined in (30). The next lemma records the identity that connects this population operator to the coefficient reconstruction, together with the two norm bounds needed in the probabilistic analysis.

Lemma 11. *For every coefficient vector* $\boldsymbol{z}$*,*

$$\boldsymbol{P}_{\mathcal{U}^\perp}\boldsymbol{S}\boldsymbol{J}\boldsymbol{z} = \boldsymbol{T}(\boldsymbol{S}\boldsymbol{z}). \tag{44}$$

Moreover,

$$\|\boldsymbol{T}\|_{\mathrm{op}} \le \frac{1}{\sqrt{2\theta}}, \qquad \|\boldsymbol{T}\|_{\mathrm{F}} \lesssim \sqrt{r_1 + r_2} + \frac{\sqrt{r_1 \wedge r_2}}{\sqrt{\theta}}. \tag{45}$$

Proof. We first verify the identity from the coordinate definition of $\boldsymbol{T}$. We then factor $\boldsymbol{T}$ through the two individual signal spaces and use their principal-angle spectrum to obtain the norm bounds.

Write $\boldsymbol{z} = (\boldsymbol{z}_1^\top, \boldsymbol{z}_2^\top)^\top$ and decompose $\boldsymbol{S}_k\boldsymbol{z}_k = \boldsymbol{U}\boldsymbol{a}_k + \boldsymbol{V}_k\boldsymbol{b}_k$, where $\boldsymbol{a}_k \in \mathbb{R}^r$ and $\boldsymbol{b}_k \in \mathbb{R}^{r_k}$. Then $\boldsymbol{S}\boldsymbol{z} = \boldsymbol{U}(\boldsymbol{a}_1 + \boldsymbol{a}_2) + \boldsymbol{V}_1\boldsymbol{b}_1 + \boldsymbol{V}_2\boldsymbol{b}_2$, whereas

$$\boldsymbol{P}_{\mathcal{U}^\perp}\boldsymbol{S}\boldsymbol{J}\boldsymbol{z} = \alpha_1\boldsymbol{V}_1\boldsymbol{b}_1 - \alpha_2\boldsymbol{V}_2\boldsymbol{b}_2 = \boldsymbol{T}(\boldsymbol{S}\boldsymbol{z}).$$

The last equality is precisely the definition of $\boldsymbol{T}$, and proves (44).

For the norm bounds, set $\boldsymbol{V} := [\boldsymbol{V}_1, \boldsymbol{V}_2]$ and $\boldsymbol{V}_\alpha := [\alpha_1\boldsymbol{V}_1, -\alpha_2\boldsymbol{V}_2]$. The spectrum in Lemma 7 shows that $\boldsymbol{V}$ has full column rank. Thus $\boldsymbol{V}^\dagger$ recovers the unique individual-space coordinates of every vector in $\mathrm{col}(\boldsymbol{V}_1) + \mathrm{col}(\boldsymbol{V}_2)$ and vanishes on its orthogonal complement, which contains both $\mathcal{U}$ and $(\mathcal{M}_1 + \mathcal{M}_2)^\perp$. Comparing this action with (30) shows that $\boldsymbol{T} = \boldsymbol{V}_\alpha\boldsymbol{V}^\dagger$ on all of $\mathbb{R}^n$.

For $\boldsymbol{b} = (\boldsymbol{b}_1^\top, \boldsymbol{b}_2^\top)^\top$, the triangle and Cauchy–Schwarz inequalities give $\|\boldsymbol{V}_\alpha\boldsymbol{b}\|_2 \le \alpha_1\|\boldsymbol{b}_1\|_2 + \alpha_2\|\boldsymbol{b}_2\|_2 \le \|\boldsymbol{b}\|_2$. Hence $\|\boldsymbol{V}_\alpha\|_{\mathrm{op}} \le 1$.

The same spectrum, together with $\chi_j \le 1 - 2\theta$, shows that $\lambda_{\min}(\boldsymbol{V}^\top\boldsymbol{V}) \ge 2\theta$. Therefore $\|\boldsymbol{V}^\dagger\|_{\mathrm{op}} \le (2\theta)^{-1/2}$, and hence $\|\boldsymbol{T}\|_{\mathrm{op}} \le (2\theta)^{-1/2}$.

For the Frobenius norm, the complete spectrum of $\boldsymbol{V}^\top\boldsymbol{V}$ gives

$$\|\boldsymbol{T}\|_{\mathrm{F}}^2 \le \left\|\boldsymbol{V}^\dagger\right\|_{\mathrm{F}}^2 = |r_1 - r_2| + \sum_{j=1}^{r_1 \wedge r_2}\left(\frac{1}{1+\chi_j} + \frac{1}{1-\chi_j}\right) \lesssim r_1 + r_2 + \frac{r_1 \wedge r_2}{\theta}.$$

Indeed, each summand is $2/(1-\chi_j^2) \le 1/\theta$. Taking square roots and using $\sqrt{x+y} \le \sqrt{x} + \sqrt{y}$ proves (45). □

B.3.3 Deterministic error decomposition

The decomposition isolates the part of the subspace error that is linear in the leakage $\boldsymbol{L}$. The direct contribution is already visible in the reconstruction formula. The remaining linear contribution comes from the displacement of the near-zero eigenspace and is contained in $\boldsymbol{T}\boldsymbol{S}\boldsymbol{\Delta}_{\boldsymbol{Q}}$. We extract this term using the eigenvalue equation for $\boldsymbol{Q}$ and then bound the resulting higher-order remainder.

Proof of Lemma 3. Write $\boldsymbol{\Delta}_{\mathrm{quad}} := \boldsymbol{L}^\top\boldsymbol{L} - \boldsymbol{D}_\sigma$ and $\boldsymbol{\Lambda} := \boldsymbol{Q}^\top\boldsymbol{G}\boldsymbol{Q}$. Since $\boldsymbol{Q}$ has orthonormal columns spanning an invariant subspace of $\boldsymbol{G}$, we have $\boldsymbol{G}\boldsymbol{Q} = \boldsymbol{Q}\boldsymbol{\Lambda}$. On $\mathcal{E}_{\mathrm{good}}$, Lemma 10 gives $\widetilde{\boldsymbol{Q}} = \boldsymbol{Q}\boldsymbol{\Xi}^{-1} = \boldsymbol{Q}_\star + \boldsymbol{\Delta}_{\boldsymbol{Q}}$.

Using the reconstruction identity preceding Lemma 3 and $\boldsymbol{S}\boldsymbol{Q}_\star = \boldsymbol{0}$, we obtain

$$\boldsymbol{P}_{\mathcal{U}^\perp}\widetilde{\boldsymbol{M}} = \boldsymbol{T}\boldsymbol{S}\boldsymbol{\Delta}_{\boldsymbol{Q}} + \boldsymbol{L}\boldsymbol{J}\boldsymbol{Q}_\star + \boldsymbol{L}\boldsymbol{J}\boldsymbol{\Delta}_{\boldsymbol{Q}}. \tag{46}$$

The second term is linear in $\boldsymbol{L}$, whereas the third is higher order. It remains to identify the linear part of $\boldsymbol{T}\boldsymbol{S}\boldsymbol{\Delta}_{\boldsymbol{Q}}$.

The relations $\boldsymbol{G}\boldsymbol{Q} = \boldsymbol{Q}\boldsymbol{\Lambda}$ and $\widetilde{\boldsymbol{Q}} = \boldsymbol{Q}\boldsymbol{\Xi}^{-1}$ imply $\boldsymbol{G}\widetilde{\boldsymbol{Q}} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{\Xi}^{-1}$. Substituting $\boldsymbol{G} = \boldsymbol{G}_0 + \boldsymbol{\Delta}$ and $\widetilde{\boldsymbol{Q}} = \boldsymbol{Q}_\star + \boldsymbol{\Delta}_{\boldsymbol{Q}}$, and using $\boldsymbol{G}_0\boldsymbol{Q}_\star = \boldsymbol{0}$, gives

$$\boldsymbol{G}_0\boldsymbol{\Delta}_{\boldsymbol{Q}} + \boldsymbol{\Delta}\boldsymbol{Q}_\star + \boldsymbol{\Delta}\boldsymbol{\Delta}_{\boldsymbol{Q}} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{\Xi}^{-1}. \tag{47}$$

Since $\boldsymbol{G}_0 = \boldsymbol{S}^\top\boldsymbol{S}$, we have $\boldsymbol{S}\boldsymbol{G}_0^\dagger\boldsymbol{G}_0 = \boldsymbol{S}$ and $\boldsymbol{S}\boldsymbol{G}_0^\dagger\boldsymbol{S}^\top = \boldsymbol{P}_{\mathrm{col}(\boldsymbol{S})}$. Moreover, $\boldsymbol{T}\boldsymbol{P}_{\mathrm{col}(\boldsymbol{S})} = \boldsymbol{T}$ on $\mathcal{E}_{\mathrm{emb}}$. Applying $\boldsymbol{T}\boldsymbol{S}\boldsymbol{G}_0^\dagger$ to (47), expanding $\boldsymbol{\Delta} = \boldsymbol{S}^\top\boldsymbol{L} + \boldsymbol{L}^\top\boldsymbol{S} + \boldsymbol{\Delta}_{\mathrm{quad}}$, and simplifying yields

$$\begin{aligned}\boldsymbol{T}\boldsymbol{S}\boldsymbol{\Delta}_{\boldsymbol{Q}} &= -\boldsymbol{T}\boldsymbol{S}\boldsymbol{G}_0^\dagger\boldsymbol{\Delta}\boldsymbol{Q}_\star - \boldsymbol{T}\boldsymbol{S}\boldsymbol{G}_0^\dagger\boldsymbol{\Delta}\boldsymbol{\Delta}_{\boldsymbol{Q}} + \boldsymbol{T}\boldsymbol{S}\boldsymbol{G}_0^\dagger\boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{\Xi}^{-1}\\ &= -\boldsymbol{T}\boldsymbol{L}\boldsymbol{Q}_\star - \boldsymbol{T}\boldsymbol{S}\boldsymbol{G}_0^\dagger\boldsymbol{\Delta}_{\mathrm{quad}}\boldsymbol{Q}_\star - \boldsymbol{T}\boldsymbol{S}\boldsymbol{G}_0^\dagger\boldsymbol{\Delta}\boldsymbol{\Delta}_{\boldsymbol{Q}} + \boldsymbol{T}\boldsymbol{S}\boldsymbol{G}_0^\dagger\boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{\Xi}^{-1}.\end{aligned} \tag{48}$$

For the second equality, $\boldsymbol{L}^\top\boldsymbol{S}\boldsymbol{Q}_\star = \boldsymbol{0}$, while

$$\boldsymbol{T}\boldsymbol{S}\boldsymbol{G}_0^\dagger\boldsymbol{S}^\top\boldsymbol{L}\boldsymbol{Q}_\star = \boldsymbol{T}\boldsymbol{P}_{\mathrm{col}(\boldsymbol{S})}\boldsymbol{L}\boldsymbol{Q}_\star = \boldsymbol{T}\boldsymbol{L}\boldsymbol{Q}_\star.$$

Thus $-\boldsymbol{T}\boldsymbol{L}\boldsymbol{Q}_\star$ is precisely the linear contribution caused by the displacement of the near-zero eigenspace.

Substituting (48) into (46) proves (31). The higher-order remainder in that decomposition is the sum of the following four terms:

$$\mathsf{Rem} = \underbrace{-\boldsymbol{T}\boldsymbol{S}\boldsymbol{G}_0^\dagger\boldsymbol{\Delta}_{\mathrm{quad}}\boldsymbol{Q}_\star}_{=:\,\boldsymbol{A}_1} + \underbrace{\boldsymbol{L}\boldsymbol{J}\boldsymbol{\Delta}_{\boldsymbol{Q}}}_{=:\,\boldsymbol{A}_2} + \underbrace{\left(-\boldsymbol{T}\boldsymbol{S}\boldsymbol{G}_0^\dagger\boldsymbol{\Delta}\boldsymbol{\Delta}_{\boldsymbol{Q}}\right)}_{=:\,\boldsymbol{A}_3} + \underbrace{\boldsymbol{T}\boldsymbol{S}\boldsymbol{G}_0^\dagger\boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{\Xi}^{-1}}_{=:\,\boldsymbol{A}_4}. \tag{49}$$

It remains to establish the bound in (32). The four estimates repeatedly use

$$\left\|\boldsymbol{T}\boldsymbol{S}\boldsymbol{G}_0^\dagger\right\|_{\mathrm{op}} \le \|\boldsymbol{T}\|_{\mathrm{op}} \left\|\boldsymbol{S}(\boldsymbol{G}_0^\dagger)^{1/2}\right\|_{\mathrm{op}} \left\|(\boldsymbol{G}_0^\dagger)^{1/2}\right\|_{\mathrm{op}} \lesssim \frac{1}{\gamma_{\min}\theta},$$

where the last inequality follows from (39b), (39c), and (45).

Bound $\|\boldsymbol{A}_1\|_{\mathrm{op}}$. By submultiplicativity, the preceding estimate, and the block definition of $\boldsymbol{Q}_\star$ together with (39a), we obtain $\|\boldsymbol{A}_1\|_{\mathrm{op}} \lesssim \|\boldsymbol{\Delta}_{\mathrm{quad}}\|_{\mathrm{op}}/(\gamma_{\min}^2\theta)$.

Bound $\|\boldsymbol{A}_2\|_{\mathrm{op}}$. Using $\|\boldsymbol{J}\|_{\mathrm{op}} \le 1$ and the coefficient bound in Lemma 10, submultiplicativity gives $\|\boldsymbol{A}_2\|_{\mathrm{op}} \lesssim \|\boldsymbol{L}\|_{\mathrm{op}}\|\boldsymbol{\Delta}_{+0}\|_{\mathrm{op}}/(\gamma_{\min}^2\sqrt{\theta})$.

Bound $\|\boldsymbol{A}_3\|_{\mathrm{op}}$. Expanding $\boldsymbol{\Delta} = \boldsymbol{S}^\top\boldsymbol{L} + \boldsymbol{L}^\top\boldsymbol{S} + \boldsymbol{\Delta}_{\mathrm{quad}}$ gives

$$\|\boldsymbol{A}_3\|_{\mathrm{op}} \le \left\|\boldsymbol{T}\boldsymbol{S}\boldsymbol{G}_0^\dagger\boldsymbol{S}^\top\boldsymbol{L}\boldsymbol{\Delta}_{\boldsymbol{Q}}\right\|_{\mathrm{op}} + \left\|\boldsymbol{T}\boldsymbol{S}\boldsymbol{G}_0^\dagger\boldsymbol{L}^\top\boldsymbol{S}\boldsymbol{\Delta}_{\boldsymbol{Q}}\right\|_{\mathrm{op}} + \left\|\boldsymbol{T}\boldsymbol{S}\boldsymbol{G}_0^\dagger\boldsymbol{\Delta}_{\mathrm{quad}}\boldsymbol{\Delta}_{\boldsymbol{Q}}\right\|_{\mathrm{op}}.$$

For the first term, use $\boldsymbol{S}\boldsymbol{G}_0^\dagger\boldsymbol{S}^\top = \boldsymbol{P}_{\mathrm{col}(\boldsymbol{S})}$ and $\boldsymbol{T} = \boldsymbol{T}\boldsymbol{P}_{\mathrm{col}(\boldsymbol{S})}$ to obtain

$$\left\|\boldsymbol{T}\boldsymbol{S}\boldsymbol{G}_0^\dagger\boldsymbol{S}^\top\boldsymbol{L}\boldsymbol{\Delta}_{\boldsymbol{Q}}\right\|_{\mathrm{op}} \le \|\boldsymbol{T}\|_{\mathrm{op}}\left\|\boldsymbol{P}_{\mathrm{col}(\boldsymbol{S})}\boldsymbol{L}\right\|_{\mathrm{op}}\|\boldsymbol{\Delta}_{\boldsymbol{Q}}\|_{\mathrm{op}} \lesssim \frac{\left\|\boldsymbol{P}_{\mathrm{col}(\boldsymbol{S})}\boldsymbol{L}\right\|_{\mathrm{op}}\|\boldsymbol{\Delta}_{+0}\|_{\mathrm{op}}}{\gamma_{\min}^2\theta}.$$

Here the last inequality follows from $\|\boldsymbol{T}\|_{\mathrm{op}} \lesssim \theta^{-1/2}$ and $\|\boldsymbol{\Delta}_{\boldsymbol{Q}}\|_{\mathrm{op}} \lesssim \|\boldsymbol{\Delta}_{+0}\|_{\mathrm{op}}/(\gamma_{\min}^2\sqrt{\theta})$.

For the second term, $\boldsymbol{L}^\top \boldsymbol{S} = (\boldsymbol{P}_{\mathrm{col}(\boldsymbol{S})}\boldsymbol{L})^\top \boldsymbol{S}$, and therefore

$$\left\|\boldsymbol{T}\boldsymbol{S}\boldsymbol{G}_0^\dagger \boldsymbol{L}^\top \boldsymbol{S}\boldsymbol{\Delta}_{\boldsymbol{Q}}\right\|_{\mathrm{op}} \le \left\|\boldsymbol{T}\boldsymbol{S}\boldsymbol{G}_0^\dagger\right\|_{\mathrm{op}} \left\|\boldsymbol{P}_{\mathrm{col}(\boldsymbol{S})}\boldsymbol{L}\right\|_{\mathrm{op}} \left\|\boldsymbol{S}\boldsymbol{\Delta}_{\boldsymbol{Q}}\right\|_{\mathrm{op}} \lesssim \frac{\left\|\boldsymbol{P}_{\mathrm{col}(\boldsymbol{S})}\boldsymbol{L}\right\|_{\mathrm{op}} \left\|\boldsymbol{\Delta}_{+0}\right\|_{\mathrm{op}}}{\gamma_{\min}^2\theta}.$$

The last inequality uses $\|\boldsymbol{T}\boldsymbol{S}\boldsymbol{G}_0^\dagger\|_{\mathrm{op}} \lesssim (\gamma_{\min}\theta)^{-1}$ and $\|\boldsymbol{S}\boldsymbol{\Delta}_{\boldsymbol{Q}}\|_{\mathrm{op}} \lesssim \|\boldsymbol{\Delta}_{+0}\|_{\mathrm{op}}/\gamma_{\min}$.

For the quadratic term,

$$\left\|\boldsymbol{T}\boldsymbol{S}\boldsymbol{G}_0^\dagger \boldsymbol{\Delta}_{\mathrm{quad}}\boldsymbol{\Delta}_{\boldsymbol{Q}}\right\|_{\mathrm{op}} \lesssim \frac{\left\|\boldsymbol{\Delta}_{\mathrm{quad}}\right\|_{\mathrm{op}} \left\|\boldsymbol{\Delta}_{+0}\right\|_{\mathrm{op}}}{\gamma_{\min}^3\theta^{3/2}} \lesssim \frac{\left\|\boldsymbol{\Delta}_{\mathrm{quad}}\right\|_{\mathrm{op}}}{\gamma_{\min}^2\theta}.$$

The final inequality follows from $\|\boldsymbol{\Delta}_{+0}\|_{\mathrm{op}} \lesssim \gamma_{\min}\sqrt{\theta}$ on $\mathcal{E}_{\mathrm{good}}$.

Combining the three bounds gives

$$\left\|\boldsymbol{A}_3\right\|_{\mathrm{op}} \lesssim \frac{\left\|\boldsymbol{P}_{\mathrm{col}(\boldsymbol{S})}\boldsymbol{L}\right\|_{\mathrm{op}} \left\|\boldsymbol{\Delta}_{+0}\right\|_{\mathrm{op}}}{\gamma_{\min}^2\theta} + \frac{\left\|\boldsymbol{\Delta}_{\mathrm{quad}}\right\|_{\mathrm{op}}}{\gamma_{\min}^2\theta}.$$

Bound $\|\boldsymbol{A}_4\|_{\mathrm{op}}$. We first control $\boldsymbol{S}\boldsymbol{G}_0^\dagger\boldsymbol{Q}$. Write the polar decomposition as $\boldsymbol{S} = \boldsymbol{U}_S\boldsymbol{G}_0^{1/2}$, where $\boldsymbol{U}_S$ is a partial isometry from $\ker(\boldsymbol{G}_0)^\perp$ onto $\mathrm{col}(\boldsymbol{S})$. Since the range of $(\boldsymbol{G}_0^\dagger)^{1/2}\boldsymbol{Q}$ lies in $\ker(\boldsymbol{G}_0)^\perp$, the map $\boldsymbol{U}_S$ preserves its norm. Therefore

$$\begin{aligned}\left\|\boldsymbol{S}\boldsymbol{G}_0^\dagger\boldsymbol{Q}\right\|_{\mathrm{op}} &= \left\|\boldsymbol{U}_S(\boldsymbol{G}_0^\dagger)^{1/2}\boldsymbol{Q}\right\|_{\mathrm{op}} = \left\|(\boldsymbol{G}_0^\dagger)^{1/2}\boldsymbol{Q}\right\|_{\mathrm{op}} \\ &\le \left\|\boldsymbol{G}_0^\dagger\right\|_{\mathrm{op}} \left\|\boldsymbol{G}_0^{1/2}\boldsymbol{Q}\right\|_{\mathrm{op}} \lesssim \frac{\left\|\boldsymbol{\Delta}_{+0}\right\|_{\mathrm{op}}}{\gamma_{\min}^2\theta}.\end{aligned}$$

The inequality uses $(\boldsymbol{G}_0^\dagger)^{1/2} = \boldsymbol{G}_0^\dagger\boldsymbol{G}_0^{1/2}$. The final bound follows from (24) and (27).

Next, $\boldsymbol{\Lambda} = \boldsymbol{Q}^\top\boldsymbol{G}\boldsymbol{Q}$ represents the restriction of $\boldsymbol{G}$ to the selected invariant subspace. The eigenvalue localization in Lemma 2 therefore gives

$$\left\|\boldsymbol{\Lambda}\right\|_{\mathrm{op}} \le \left\|\boldsymbol{\Delta}_{00}\right\|_{\mathrm{op}} + 2\left\|\boldsymbol{\Delta}_{+0}\right\|_{\mathrm{op}}^2.$$

Combining these estimates with $\|\boldsymbol{T}\|_{\mathrm{op}} \lesssim \theta^{-1/2}$ and $\|\boldsymbol{\Xi}^{-1}\|_{\mathrm{op}} \lesssim \gamma_{\min}^{-1}$ yields

$$\begin{aligned}\left\|\boldsymbol{A}_4\right\|_{\mathrm{op}} &\le \left\|\boldsymbol{T}\right\|_{\mathrm{op}} \left\|\boldsymbol{S}\boldsymbol{G}_0^\dagger\boldsymbol{Q}\right\|_{\mathrm{op}} \left\|\boldsymbol{\Lambda}\right\|_{\mathrm{op}} \left\|\boldsymbol{\Xi}^{-1}\right\|_{\mathrm{op}} \\ &\lesssim \frac{\left(\left\|\boldsymbol{\Delta}_{00}\right\|_{\mathrm{op}} + \left\|\boldsymbol{\Delta}_{+0}\right\|_{\mathrm{op}}^2\right)\left\|\boldsymbol{\Delta}_{+0}\right\|_{\mathrm{op}}}{\gamma_{\min}^3\theta^{3/2}} \\ &\lesssim \frac{\left\|\boldsymbol{\Delta}_{\mathrm{quad}}\right\|_{\mathrm{op}}}{\gamma_{\min}^2\theta} + \frac{\left\|\boldsymbol{\Delta}_{+0}\right\|_{\mathrm{op}}^2}{\gamma_{\min}^2\theta}.\end{aligned}$$

For the last inequality, observe that $\boldsymbol{\Delta}_{00} = \boldsymbol{P}_{\ker(\boldsymbol{G}_0)}\boldsymbol{\Delta}_{\mathrm{quad}}\boldsymbol{P}_{\ker(\boldsymbol{G}_0)}$, so $\|\boldsymbol{\Delta}_{00}\|_{\mathrm{op}} \le \|\boldsymbol{\Delta}_{\mathrm{quad}}\|_{\mathrm{op}}$. Moreover, $\|\boldsymbol{\Delta}_{+0}\|_{\mathrm{op}} \lesssim \gamma_{\min}\sqrt{\theta}$ on $\mathcal{E}_{\mathrm{good}}$. Consequently,

$$\frac{\left\|\boldsymbol{\Delta}_{00}\right\|_{\mathrm{op}}\left\|\boldsymbol{\Delta}_{+0}\right\|_{\mathrm{op}}}{\gamma_{\min}^3\theta^{3/2}} \lesssim \frac{\left\|\boldsymbol{\Delta}_{\mathrm{quad}}\right\|_{\mathrm{op}}}{\gamma_{\min}^2\theta}, \qquad \frac{\left\|\boldsymbol{\Delta}_{+0}\right\|_{\mathrm{op}}^3}{\gamma_{\min}^3\theta^{3/2}} \lesssim \frac{\left\|\boldsymbol{\Delta}_{+0}\right\|_{\mathrm{op}}^2}{\gamma_{\min}^2\theta}.$$

Combining the bounds for $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_4$ in (49) proves (32) and completes the proof. $\square$

C Probabilistic tools for the upper bound

C.1 Gaussian concentration and matrix bounds

We begin with the standard Gaussian concentration inequality. It will be applied below to a Lipschitz function of the bulk noise matrix.

Lemma 12 (Gaussian Lipschitz concentration)**.** *Let $\boldsymbol{g} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_N)$, and let $f : \mathbb{R}^N \to \mathbb{C}$ be L-Lipschitz with respect to the Euclidean norm. Then, for every $q \geq 2$,*

$$\|f(\boldsymbol{g}) - \mathbb{E} f(\boldsymbol{g})\|_{L^q} \leq C\sqrt{q}\, L. \tag{50}$$

The following bounds are consequences of the standard extreme-singular-value estimates for Gaussian matrices; see (Vershynin, 2010, Theorem 5.32 and Corollary 5.35). The moment bounds follow by integrating the corresponding tail inequalities.

Lemma 13 (Gaussian matrix deviations and moments)**.** *Let $\boldsymbol{G} \in \mathbb{R}^{u \times v}$ have independent $\mathcal{N}(0, b^{-1})$ entries. Then, for every $x \geq 0$,*

$$\mathbb{P}\left\{\|\boldsymbol{G}\|_{\mathrm{op}} > \frac{\sqrt{u} + \sqrt{v}}{\sqrt{b}} + x\right\} \leq 2e^{-cbx^2}. \tag{51}$$

If $v = b$, then, for every $s \geq 0$,

$$\mathbb{P}\left\{\|\boldsymbol{G}\boldsymbol{G}^\top - \boldsymbol{I}_u\|_{\mathrm{op}} > C\left(\sqrt{\frac{u+s}{b}} + \frac{u+s}{b}\right)\right\} \leq 2e^{-cs}. \tag{52}$$

Moreover, for every $q \geq 2$,

$$\|\boldsymbol{G}\|_{L^q} \leq C\left(\sqrt{\frac{u}{b}} + \sqrt{\frac{v}{b}} + \sqrt{\frac{q}{b}}\right), \tag{53}$$

and, when $v = b$,

$$\|\boldsymbol{G}\boldsymbol{G}^\top - \boldsymbol{I}_u\|_{L^q} \leq C\left(\sqrt{\frac{u+q}{b}} + \frac{u+q}{b}\right). \tag{54}$$

We next control the Gaussian coefficient matrices that arise in the one-view expansion. The result covers both independent bilinear forms and centered same-sample quadratic forms.

Lemma 14 (Gaussian coefficient moments)**.** *Let $q \geq 2$, and let $\boldsymbol{G}_1 \in \mathbb{R}^{r_1 \times p}$ and $\boldsymbol{G}_2 \in \mathbb{R}^{r_2 \times p}$ be independent matrices with independent $\mathcal{N}(0, b^{-1})$ entries. For every deterministic $\boldsymbol{A} \in \mathbb{R}^{r_1 \times r_2}$,*

$$\|\boldsymbol{G}_1^\top \boldsymbol{A} \boldsymbol{G}_2\|_{L^q} \leq \frac{C}{b}\left\{(\sqrt{p} + \sqrt{q})\|\boldsymbol{A}\|_{\mathrm{F}} + (p+q)\|\boldsymbol{A}\|_{\mathrm{op}}\right\}. \tag{55}$$

If $r_1 = r_2$ and $\boldsymbol{A} \in \mathbb{R}^{r_1 \times r_1}$, then

$$\left\|\boldsymbol{G}_1^\top \boldsymbol{A} \boldsymbol{G}_1 - b^{-1}\operatorname{tr}(\boldsymbol{A})\boldsymbol{I}_p\right\|_{L^q} \leq \frac{C}{b}\left\{(\sqrt{p} + \sqrt{q})\|\boldsymbol{A}\|_{\mathrm{F}} + (p+q)\|\boldsymbol{A}\|_{\mathrm{op}}\right\}. \tag{56}$$

Both conclusions remain valid conditionally when $\boldsymbol{A}$ is measurable with respect to a sigma-field independent of the displayed Gaussian matrices.

Proof. We first record the Gaussian matrix estimate used for both conclusions. Let $\boldsymbol{H} \in \mathbb{R}^{k\times d}$ have independent standard Gaussian entries, and let $\boldsymbol{B} \in \mathbb{R}^{d\times m}$ be deterministic. Apply Chevet's inequality (Vershynin, 2018, Section 8.7) with $T = \boldsymbol{B}\mathbb{S}^{m-1}$ and $S = \mathbb{S}^{k-1}$. Since

$$\mathrm{rad}(T) = \|\boldsymbol{B}\|_{\mathrm{op}}, \qquad w(T) = \mathbb{E}\|\boldsymbol{B}^\top \boldsymbol{g}_d\|_2 \le \|\boldsymbol{B}\|_{\mathrm{F}},$$

whereas $\mathrm{rad}(S) = 1$ and $w(S) = \mathbb{E}\|\boldsymbol{g}_k\|_2 \le \sqrt{k}$, we obtain

$$\mathbb{E}\|\boldsymbol{H}\boldsymbol{B}\|_{\mathrm{op}} \le \|\boldsymbol{B}\|_{\mathrm{F}} + \sqrt{k}\,\|\boldsymbol{B}\|_{\mathrm{op}}.$$

It remains to control the fluctuation around this mean. The map $\boldsymbol{H} \mapsto \|\boldsymbol{H}\boldsymbol{B}\|_{\mathrm{op}}$ is $\|\boldsymbol{B}\|_{\mathrm{op}}$-Lipschitz with respect to the Frobenius norm, because

$$|\|\boldsymbol{H}_1\boldsymbol{B}\|_{\mathrm{op}} - \|\boldsymbol{H}_0\boldsymbol{B}\|_{\mathrm{op}}| \le \|(\boldsymbol{H}_1 - \boldsymbol{H}_0)\boldsymbol{B}\|_{\mathrm{op}} \le \|\boldsymbol{H}_1 - \boldsymbol{H}_0\|_{\mathrm{F}}\|\boldsymbol{B}\|_{\mathrm{op}}.$$

Therefore, (50) and the preceding expectation bound give

$$\|\boldsymbol{H}\boldsymbol{B}\|_{L^q} \le C\left\{\|\boldsymbol{B}\|_{\mathrm{F}} + (\sqrt{k} + \sqrt{q})\|\boldsymbol{B}\|_{\mathrm{op}}\right\}. \tag{57}$$

We now prove the bilinear estimate. Set $\boldsymbol{Z}_i = \sqrt{b}\,\boldsymbol{G}_i$, so that $\boldsymbol{Z}_1$ and $\boldsymbol{Z}_2$ have independent standard Gaussian entries. Conditional on $\boldsymbol{Z}_2$, apply (57) with $\boldsymbol{H} = \boldsymbol{Z}_1^\top$ and $\boldsymbol{B} = \boldsymbol{A}\boldsymbol{Z}_2$. Taking the L^q norm over $\boldsymbol{Z}_2$ then gives

$$\|\boldsymbol{Z}_1^\top \boldsymbol{A}\boldsymbol{Z}_2\|_{L^q} \le C\left\{\big\|\|\boldsymbol{A}\boldsymbol{Z}_2\|_{\mathrm{F}}\big\|_{L^q} + (\sqrt{p} + \sqrt{q})\big\|\|\boldsymbol{A}\boldsymbol{Z}_2\|_{\mathrm{op}}\big\|_{L^q}\right\}. \tag{58}$$

The two random norms on the right are controlled by the same two ingredients. First, the map $\boldsymbol{Z}_2 \mapsto \|\boldsymbol{A}\boldsymbol{Z}_2\|_{\mathrm{F}}$ is $\|\boldsymbol{A}\|_{\mathrm{op}}$-Lipschitz, and

$$\mathbb{E}\|\boldsymbol{A}\boldsymbol{Z}_2\|_{\mathrm{F}} \le \left(\mathbb{E}\|\boldsymbol{A}\boldsymbol{Z}_2\|_{\mathrm{F}}^2\right)^{1/2} = \sqrt{p}\,\|\boldsymbol{A}\|_{\mathrm{F}}.$$

Gaussian concentration therefore gives

$$\big\|\|\boldsymbol{A}\boldsymbol{Z}_2\|_{\mathrm{F}}\big\|_{L^q} \le \sqrt{p}\,\|\boldsymbol{A}\|_{\mathrm{F}} + C\sqrt{q}\,\|\boldsymbol{A}\|_{\mathrm{op}}.$$

Second, applying (57) to $\boldsymbol{Z}_2^\top \boldsymbol{A}^\top$ yields

$$\big\|\|\boldsymbol{A}\boldsymbol{Z}_2\|_{\mathrm{op}}\big\|_{L^q} \le C\left\{\|\boldsymbol{A}\|_{\mathrm{F}} + (\sqrt{p} + \sqrt{q})\|\boldsymbol{A}\|_{\mathrm{op}}\right\}.$$

Substituting these two bounds into (58) and using $(\sqrt{p} + \sqrt{q})^2 \le 2(p+q)$ gives

$$\|\boldsymbol{Z}_1^\top \boldsymbol{A}\boldsymbol{Z}_2\|_{L^q} \le C\left\{(\sqrt{p} + \sqrt{q})\|\boldsymbol{A}\|_{\mathrm{F}} + (p+q)\|\boldsymbol{A}\|_{\mathrm{op}}\right\}.$$

Since $\boldsymbol{G}_1^\top \boldsymbol{A}\boldsymbol{G}_2 = b^{-1}\boldsymbol{Z}_1^\top \boldsymbol{A}\boldsymbol{Z}_2$, this proves (55).

For the same-sample estimate, let $\boldsymbol{Z}'$ be an independent copy of $\boldsymbol{Z}_1$ and define

$$\boldsymbol{U} := \frac{\boldsymbol{Z}_1 + \boldsymbol{Z}'}{\sqrt{2}}, \qquad \boldsymbol{V} := \frac{\boldsymbol{Z}_1 - \boldsymbol{Z}'}{\sqrt{2}}.$$

The matrices $\boldsymbol{U}$ and $\boldsymbol{V}$ are independent standard Gaussian matrices. By symmetrization and the identity

$$\boldsymbol{Z}_1^\top \boldsymbol{A}\boldsymbol{Z}_1 - \boldsymbol{Z}'^\top \boldsymbol{A}\boldsymbol{Z}' = \boldsymbol{U}^\top \boldsymbol{A}\boldsymbol{V} + \boldsymbol{V}^\top \boldsymbol{A}\boldsymbol{U},$$

the bilinear estimate gives

$$\begin{aligned}\left\|\boldsymbol{Z}_1^\top \boldsymbol{A}\boldsymbol{Z}_1 - \mathrm{tr}(\boldsymbol{A})\boldsymbol{I}_p\right\|_{L^q} &\le \left\|\boldsymbol{U}^\top \boldsymbol{A}\boldsymbol{V} + \boldsymbol{V}^\top \boldsymbol{A}\boldsymbol{U}\right\|_{L^q} \\ &\le C\left\{(\sqrt{p} + \sqrt{q})\|\boldsymbol{A}\|_{\mathrm{F}} + (p+q)\|\boldsymbol{A}\|_{\mathrm{op}}\right\}.\end{aligned}$$

Rescaling by b^{-1} proves (56). The conditional versions follow by applying the deterministic bounds conditionally on the sigma-field with respect to which $\boldsymbol{A}$ is measurable and then taking the outer expectation. $\square$

D One-view weighted leakage isometry

This section proves the one-view result used in Lemma 4. Throughout the section, $C, c > 0$ denote numerical constants whose values may change from line to line.

D.1 Normalized model and theorem

The leakage of an empirical singular vector depends on the corresponding population singular value. The purpose of the weighting is to remove this dependence: after weighting, the leakage Gram matrix should be close to the same scalar matrix for every signal direction. We formulate this statement first in normalized noise units.

Normalized observation. Let $\boldsymbol{X} \in \mathbb{R}^{n\times d}$ have rank p and compact SVD

$$\boldsymbol{X} = \boldsymbol{U}_\star \boldsymbol{\Sigma}_\star \boldsymbol{V}_\star^\top, \qquad \boldsymbol{\Sigma}_\star = \operatorname{diag}(s_1^\star, \dots, s_p^\star),$$

where $\boldsymbol{U}_\star \in \mathrm{St}(n,p)$, $\boldsymbol{V}_\star \in \mathrm{St}(d,p)$, and $s_1^\star \geq \cdots \geq s_p^\star \geq s_\star > 0$. Set $a \coloneqq n-p$, $b \coloneqq d-p$, and $\beta \coloneqq a/b$. We observe

$$\boldsymbol{Y} = \boldsymbol{X} + \boldsymbol{Z}, \qquad Z_{ij} \overset{\text{iid}}{\sim} N(0, b^{-1}). \tag{59}$$

Let $\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top$ be a leading rank-p SVD of $\boldsymbol{Y}$, where

$$\boldsymbol{\Sigma} = \operatorname{diag}(\widehat{s}_1, \dots, \widehat{s}_p), \qquad \widehat{s}_1 \geq \cdots \geq \widehat{s}_p.$$

Choose an orthonormal complement $\boldsymbol{U}_\star^\perp$ of $\boldsymbol{U}_\star$. Then $(\boldsymbol{U}_\star^\perp)^\top \boldsymbol{U}$ contains the components of the empirical left singular vectors outside the population signal space.

Bias-correcting weights. As discussed in Section 1.1, the Gram matrix of these components has a diagonal bias that varies with the signal singular values. To correct it, define

$$\tau_\beta(s) \coloneqq \begin{cases} \dfrac{s^2 - (1+\beta) + \sqrt{\{s^2 - (1+\beta)\}^2 - 4\beta}}{2}, & s > 1 + \sqrt{\beta}, \\ \sqrt{\beta}, & s \leq 1 + \sqrt{\beta}. \end{cases} \tag{60}$$

For each empirical singular value $\widehat{s}_i$, set

$$\tau_i \coloneqq \tau_\beta(\widehat{s}_i), \qquad \ell_i \coloneqq \sqrt{\frac{\tau_i(\tau_i + \beta)}{\tau_i + 1}}, \qquad \boldsymbol{W}_L \coloneqq \operatorname{diag}(\ell_1, \dots, \ell_p).$$

When $\tau_i > \sqrt{\beta}$, the first branch of (60) is equivalent to

$$\widehat{s}_i^2 = 1 + \beta + \tau_i + \frac{\beta}{\tau_i}. \tag{61}$$

Below the threshold $1 + \sqrt{\beta}$, the definition fixes $\tau_i = \sqrt{\beta}$ and hence $\ell_i^2 = \beta$.

The weighted leakage matrix is

$$\boldsymbol{N}_L \coloneqq (\boldsymbol{U}_\star^\perp)^\top \boldsymbol{U} \boldsymbol{W}_L. \tag{62}$$

The choice of $\boldsymbol{W}_L$ is designed so that $\boldsymbol{N}_L^\top \boldsymbol{N}_L$ is centered at $\beta \boldsymbol{I}_p$. The theorem below makes this statement quantitative.

One-view theorem. The relevant signal scale is $s_\star^2 \wedge s_\star^4$. For $s_\star \le 1$, this equals $s_\star^4$, while for $s_\star \ge 1$, it equals $s_\star^2$. Thus the assumption below covers the two signal regimes without imposing an upper bound on $s_1^\star$.

Theorem 3 (One-view weighted leakage isometry)**.** *Suppose that*

$$p \le \frac{1}{3}\min\{n, d\}, \qquad s_\star^2 \wedge s_\star^4 \ge \nu^2 \beta,$$

where $\nu > 0$ is a sufficiently large numerical constant. Then

$$\left\| \boldsymbol{N}_L^\top \boldsymbol{N}_L - \beta \boldsymbol{I}_p \right\|_{L^{32}} \le C\beta \sqrt{\frac{p}{a}}, \tag{63a}$$

$$\|\boldsymbol{N}_L\|_{L^{64}} \le C\sqrt{\beta}. \tag{63b}$$

Moreover, there is an event $\mathcal{E}_{1\mathrm{v}}$ such that

$$\mathbb{P}(\mathcal{E}_{1\mathrm{v}}^c) \le C \exp\left\{ -\frac{b(s_\star^2 \wedge s_\star^4)}{C} \right\}, \tag{64}$$

and, on $\mathcal{E}_{1\mathrm{v}}$,

$$\boldsymbol{W}_L^2 - \beta \boldsymbol{I}_p \succeq c(s_\star^2 \wedge s_\star^4)\boldsymbol{I}_p. \tag{65}$$

Here $C, c > 0$ are numerical constants.

D.2 Proof of Lemma 4

Proof. Fix $k \in \{1, 2\}$. The normalized theorem supplies both the moment bounds for $\boldsymbol{L}_k$ and a lower bound for the weighted empirical Gram matrix. We first return its conclusions to the original noise units and then combine the latter lower bound with the centered leakage estimate to control $\boldsymbol{S}_k^\top \boldsymbol{S}_k$.

Let $\boldsymbol{X}_k = \boldsymbol{U}_{\star,k}\boldsymbol{\Sigma}_{\star,k}\boldsymbol{V}_{\star,k}^\top$ be a compact SVD, and choose an orthonormal complement $\boldsymbol{U}_{\star,k}^\perp$ of $\boldsymbol{U}_{\star,k}$. Apply Theorem 3 to $\widetilde{\boldsymbol{Y}}_k := \boldsymbol{Y}_k/(\sigma\sqrt{b_k})$, with $a = a_k$, $b = b_k$, $\beta = a_k/b_k$, and $s_\star = s_k/(\sigma\sqrt{b_k})$. The noise entries of $\widetilde{\boldsymbol{Y}}_k$ are independent $N(0, b_k^{-1})$, and its empirical left singular vectors are the columns of $\widehat{\boldsymbol{\Phi}}_k$.

The signal assumption follows directly from the definition of γ_k. Indeed, $b_k = d_k - p_k$ and $p_k \le d_k/3$ imply $b_k \le d_k \le 3b_k/2$, so $s_\star^2 \wedge s_\star^4 \asymp \gamma_k^2/(b_k\sigma^2)$. The standing condition $\gamma_k \ge \nu\sigma\sqrt{n}$ therefore gives $s_\star^2 \wedge s_\star^4 \gtrsim \nu^2 a_k/b_k$ because $a_k \le n$. Thus the signal condition in Theorem 3 holds after increasing the numerical constant ν if necessary.

Let $\boldsymbol{W}_L$ and $\boldsymbol{N}_L$ be the matrices in the normalized theorem. Substitution into the definitions of τ_k and w_k gives $w_k(\widehat{s}_{k,i}) = \sigma\sqrt{b_k}\,\ell_i$. Hence $\boldsymbol{C}_k = \sigma\sqrt{b_k}\,\widehat{\boldsymbol{\Phi}}_k\boldsymbol{W}_L$ and $\boldsymbol{L}_k = \sigma\sqrt{b_k}\,\boldsymbol{U}_{\star,k}^\perp\boldsymbol{N}_L$. Rescaling the two conclusions of Theorem 3 gives

$$\left\| \boldsymbol{L}_k^\top \boldsymbol{L}_k - a_k\sigma^2 \boldsymbol{I}_{p_k} \right\|_{L^{32}} \lesssim \sigma^2\sqrt{a_k p_k}, \qquad \|\boldsymbol{L}_k\|_{L^{64}} \lesssim \sigma\sqrt{a_k}. \tag{66}$$

This proves (33).

Let $\widetilde{\mathcal{E}}_k$ denote the event supplied by Theorem 3 for the normalized kth view. Since $\widehat{\boldsymbol{\Phi}}_k$ has orthonormal columns, the lower bound for $\boldsymbol{W}_L^2 - \frac{a_k}{b_k}\boldsymbol{I}_{p_k}$ and the preceding comparison of signal scales imply, on $\widetilde{\mathcal{E}}_k$,

$$\boldsymbol{C}_k^\top \boldsymbol{C}_k - a_k\sigma^2 \boldsymbol{I}_{p_k} \succeq c\gamma_k^2 \boldsymbol{I}_{p_k}.$$

Define

$$\mathcal{E}_k := \widetilde{\mathcal{E}}_k \cap \left\{ \left\| \boldsymbol{L}_k^\top \boldsymbol{L}_k - a_k \sigma^2 \boldsymbol{I}_{p_k} \right\|_{\mathrm{op}} \le \frac{c}{2}\gamma_k^2 \right\}. \tag{67}$$

Because $\boldsymbol{C}_k = \boldsymbol{S}_k + \boldsymbol{L}_k$ and $\boldsymbol{S}_k^\top \boldsymbol{L}_k = \boldsymbol{0}$, on $\mathcal{E}_k$ we have

$$\boldsymbol{S}_k^\top \boldsymbol{S}_k = \left(\boldsymbol{C}_k^\top \boldsymbol{C}_k - a_k\sigma^2 \boldsymbol{I}_{p_k}\right) - \left(\boldsymbol{L}_k^\top \boldsymbol{L}_k - a_k\sigma^2 \boldsymbol{I}_{p_k}\right) \succeq \frac{c}{2}\gamma_k^2 \boldsymbol{I}_{p_k}. \tag{68}$$

Choosing $c_{\mathrm{emb}} \le \sqrt{c/2}$ proves $\sigma_{\min}(\boldsymbol{S}_k) \ge c_{\mathrm{emb}}\gamma_k$ on $\mathcal{E}_k$.

Finally, the probability bound in Theorem 3 and Markov's inequality applied to (66) give

$$\mathbb{P}(\mathcal{E}_k^c) \lesssim \exp\left(-\frac{\gamma_k^2}{C\sigma^2}\right) + \left(\frac{\sigma^2\sqrt{a_k p_k}}{\gamma_k^2}\right)^{32} \lesssim \frac{\sigma^2\sqrt{n p_k}}{\gamma_k^2}.$$

The last inequality uses $a_k \le n$ and $\gamma_k^2 \gtrsim n\sigma^2$. This proves the asserted probability bound and completes the proof. $\square$

D.3 Proof of Theorem 3

We first prove the result in the wide case $a \le b$, so that $\beta \le 1$. Afterwards, we derive a corresponding right-leakage bound under the stronger signal condition $s_\star^2 \ge \nu^2$ and use it to handle the tall case by transposition.

By rotational invariance of the Gaussian noise, we may work in the population singular-vector coordinates, in which

$$\boldsymbol{Y} = \begin{pmatrix} \boldsymbol{\Sigma}_\star + \boldsymbol{Z}_{11} & \boldsymbol{Z}_{12} \\ \boldsymbol{Z}_{21} & \boldsymbol{Z}_{22} \end{pmatrix}, \tag{69}$$

where the four noise blocks have independent $N(0, b^{-1})$ entries and $\boldsymbol{Z}_{22} \in \mathbb{R}^{a\times b}$. Partition the empirical singular vectors conformably as

$$\boldsymbol{U} = \begin{pmatrix} \boldsymbol{U}_\parallel \\ \boldsymbol{U}_\perp \end{pmatrix}, \qquad \boldsymbol{V} = \begin{pmatrix} \boldsymbol{V}_\parallel \\ \boldsymbol{V}_\perp \end{pmatrix}. \tag{70}$$

In these coordinates, $\boldsymbol{N}_L = \boldsymbol{U}_\perp \boldsymbol{W}_L$. Thus (63a) amounts to controlling

$$\boldsymbol{X}_L := \boldsymbol{W}_L \boldsymbol{U}_\perp^\top \boldsymbol{U}_\perp \boldsymbol{W}_L - \beta \boldsymbol{I}_p. \tag{71}$$

We first work on the localization event

$$\mathcal{E}_{\mathrm{loc}} := \left\{ \min_i \tau_i \ge \frac{s_\star^2}{16}, \quad \left\| \boldsymbol{Z}_{22}\boldsymbol{Z}_{22}^\top - \boldsymbol{I}_a \right\|_{\mathrm{op}} \le \frac{s_\star^2}{128} \right\}. \tag{72}$$

The signal condition implies $s_\star^2/16 > \sqrt{\beta}$ for sufficiently large ν. Thus the first condition in $\mathcal{E}_{\mathrm{loc}}$ places every selected singular value on the upper branch, where (61) applies. The second condition will make the singular-vector expansion contractive.

Lemma 15. *Under $s_\star^2 \wedge s_\star^4 \ge \nu^2\beta$,*

$$\mathbb{P}(\mathcal{E}_{\mathrm{loc}}^c) \lesssim \exp\left\{-cb(s_\star^2 \wedge s_\star^4)\right\}. \tag{73}$$

Under the stronger condition $s_\star^2 \ge \nu^2$, $\mathbb{P}(\mathcal{E}_{\mathrm{loc}}^c) \lesssim \exp(-cbs_\star^2)$.

Step 1: control $\boldsymbol{X}_L$ on $\mathcal{E}_{\mathrm{loc}}$. The main difficulty is that the left and right leakage Grams $\boldsymbol{U}_\perp^\top \boldsymbol{U}_\perp$ and $\boldsymbol{V}_\perp^\top \boldsymbol{V}_\perp$ are coupled. The two leakage Grams satisfy a pair of approximate linear relations. To state them, define the following three entrywise operators

$$[\mathcal{A}_U(\boldsymbol{M})]_{ij} \coloneqq \frac{\beta r_i r_j}{\tau_i\tau_j - \beta} M_{ij}, \tag{74a}$$

$$[\mathcal{B}(\boldsymbol{M})]_{ij} \coloneqq \frac{\beta}{\tau_i\tau_j - \beta} M_{ij}, \tag{74b}$$

$$[\mathcal{A}_V(\boldsymbol{M})]_{ij} \coloneqq \frac{\ell_i \ell_j}{\tau_i\tau_j - \beta} M_{ij}, \tag{74c}$$

where $r_i \coloneqq \sqrt{\frac{\tau_i(\tau_i+1)}{\tau_i+\beta}}$ and we define $\boldsymbol{W}_R \coloneqq \mathrm{diag}(r_1, \ldots, r_p)$. These operators are well defined on $\mathcal{E}_{\mathrm{loc}}$ because $\tau_i > \sqrt{\beta}$ for every i. Define the residuals

$$\boldsymbol{\varepsilon}_U \coloneqq \boldsymbol{U}_\perp^\top \boldsymbol{U}_\perp - \mathcal{A}_U(\boldsymbol{V}_\parallel^\top \boldsymbol{V}_\parallel) - \mathcal{B}(\boldsymbol{U}_\parallel^\top \boldsymbol{U}_\parallel), \tag{75a}$$

$$\boldsymbol{\varepsilon}_V \coloneqq \boldsymbol{V}_\perp^\top \boldsymbol{V}_\perp - \mathcal{B}(\boldsymbol{V}_\parallel^\top \boldsymbol{V}_\parallel) - \mathcal{A}_V(\boldsymbol{U}_\parallel^\top \boldsymbol{U}_\parallel). \tag{75b}$$

Since the empirical singular-vector matrices have orthonormal columns, $\boldsymbol{U}_\parallel^\top \boldsymbol{U}_\parallel + \boldsymbol{U}_\perp^\top \boldsymbol{U}_\perp = \boldsymbol{I}_p$ and similarly for $\boldsymbol{V}$. Thus (75) forms a linear system for the two leakage Grams. To solve it, define $\boldsymbol{X}_R \coloneqq \boldsymbol{W}_R \boldsymbol{V}_\perp^\top \boldsymbol{V}_\perp \boldsymbol{W}_R - \boldsymbol{I}_p$. Fix i, j and abbreviate $u = \tau_i\tau_j$, $\ell = \ell_i\ell_j$, $r = r_i r_j$, $x = (\boldsymbol{X}_L)_{ij}$, and $y = (\boldsymbol{X}_R)_{ij}$. Since $\ell_i r_i = \tau_i$, we have $\ell r = u$, and the two equations reduce to

$$\begin{pmatrix} \dfrac{u}{u-\beta} & \dfrac{\beta\ell}{u-\beta} \\ \dfrac{r}{u-\beta} & \dfrac{u}{u-\beta} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \ell(\boldsymbol{\varepsilon}_U)_{ij} \\ r(\boldsymbol{\varepsilon}_V)_{ij} \end{pmatrix}. \tag{76}$$

Direct inversion gives the cancellation identities

$$\boldsymbol{X}_L = \boldsymbol{W}_L \boldsymbol{\varepsilon}_U \boldsymbol{W}_L - \beta \boldsymbol{\varepsilon}_V, \qquad \boldsymbol{X}_R = \boldsymbol{W}_R \boldsymbol{\varepsilon}_V \boldsymbol{W}_R - \boldsymbol{\varepsilon}_U. \tag{77}$$

The following lemma controls the two residuals in precisely the combinations appearing here.

Lemma 16. *Under $s_\star^2 \wedge s_\star^4 \geq \nu^2 \beta$,*

$$\|\mathbf{1}_{\mathcal{E}_{\mathrm{loc}}} \boldsymbol{W}_L \boldsymbol{\varepsilon}_U \boldsymbol{W}_L\|_{L^{32}} + \beta \, \|\mathbf{1}_{\mathcal{E}_{\mathrm{loc}}} \boldsymbol{\varepsilon}_V\|_{L^{32}} \lesssim \frac{\sqrt{ap}}{b}. \tag{78}$$

Under the stronger condition $s_\star^2 \geq \nu^2$,

$$\|\mathbf{1}_{\mathcal{E}_{\mathrm{loc}}} \boldsymbol{W}_R \boldsymbol{\varepsilon}_V \boldsymbol{W}_R\|_{L^{32}} + \|\mathbf{1}_{\mathcal{E}_{\mathrm{loc}}} \boldsymbol{\varepsilon}_U\|_{L^{32}} \lesssim \sqrt{\frac{p}{b}}. \tag{79}$$

By (77) and (78), $\|\mathbf{1}_{\mathcal{E}_{\mathrm{loc}}} \boldsymbol{X}_L\|_{L^{32}} \lesssim \sqrt{ap}/b$.

Step 2: control $\boldsymbol{X}_L$ on $\mathcal{E}_{\mathrm{loc}}^c$. From the definition of $\boldsymbol{X}_L$, $\|\boldsymbol{X}_L\|_{\mathrm{op}} \leq \|\boldsymbol{U}_\perp \boldsymbol{W}_L\|_{\mathrm{op}}^2 + \beta$. We use the following radial estimate.

Lemma 17. *In the wide case $a \leq b$,*

$$\|\boldsymbol{U}_\perp \boldsymbol{W}_L\|_{L^{128}} \lesssim \sqrt{\beta}, \qquad \|\boldsymbol{V}_\perp \boldsymbol{W}_R\|_{L^{128}} \lesssim 1. \tag{80}$$

Hölder's inequality, (73) and (80), and $b(s_\star^2 \wedge s_\star^4) \geq \nu^2 a$ give $\|\mathbf{1}_{\mathcal{E}_{\mathrm{loc}}^c} \boldsymbol{X}_L\|_{L^{32}} \lesssim \beta e^{-ca} \lesssim \sqrt{ap}/b$. Together with the bound on $\mathcal{E}_{\mathrm{loc}}$, this proves (63a) in the wide case.

Completing the wide-case proof. The radial bound (63b) and the localization probability (64) follow directly from Lemmas 15 and 17. Finally, on $\mathcal{E}_{\mathrm{loc}}$, $\tau_i \ge s_\star^2/16$ and

$$\ell_i^2 - \beta = \frac{\tau_i^2 - \beta}{\tau_i + 1} \ge \frac{(s_\star^2/16)^2 - \beta}{s_\star^2/16 + 1} \gtrsim s_\star^2 \wedge s_\star^4,$$

where we used the monotonicity of $t \mapsto (t^2 - \beta)/(t+1)$ on $[\sqrt{\beta}, \infty)$. This proves (65) and completes the wide-case proof.

A right-leakage bound for wide matrices. To pass to the tall case, we need the corresponding control of the *right* leakage for a wide matrix. Choose an orthonormal complement $\boldsymbol{V}_\star^\perp$ of $\boldsymbol{V}_\star$ and define $\boldsymbol{N}_R := (\boldsymbol{V}_\star^\perp)^\top \boldsymbol{V}\boldsymbol{W}_R$. In the population coordinates, $\boldsymbol{N}_R^\top \boldsymbol{N}_R - \boldsymbol{I}_p = \boldsymbol{X}_R$.

Suppose now that $s_\star^2 \ge \nu^2$. On $\mathcal{E}_{\mathrm{loc}}$, the second identity in (77) and (79) give $\|\boldsymbol{1}_{\mathcal{E}_{\mathrm{loc}}} \boldsymbol{X}_R\|_{L^{32}} \lesssim \sqrt{p/b}$. On the complement, $\|\boldsymbol{X}_R\|_{\mathrm{op}} \le \|\boldsymbol{V}_\perp \boldsymbol{W}_R\|_{\mathrm{op}}^2 + 1$, so Lemmas 15 and 17 and Hölder's inequality give the same bound. Hence

$$\begin{aligned} \left\|\boldsymbol{N}_R^\top \boldsymbol{N}_R - \boldsymbol{I}_p\right\|_{L^{32}} &\lesssim \sqrt{\frac{p}{b}}, & \|\boldsymbol{N}_R\|_{L^{64}} &\lesssim 1, \\ \mathbb{P}(\mathcal{E}_{\mathrm{loc}}^c) &\lesssim e^{-cbs_\star^2}, & \min_i (r_i^2 - 1) &\gtrsim s_\star^2 \quad \text{on } \mathcal{E}_{\mathrm{loc}}. \end{aligned} \tag{81}$$

For the last bound, use $r_i^2 - 1 = (\tau_i^2 - \beta)/(\tau_i + \beta)$, $\tau_i \ge s_\star^2/16$, and $s_\star^2 \ge \nu^2$.

The tall case. Suppose now that $a > b$, so $\beta > 1$, and consider the rescaled transpose

$$\widetilde{\boldsymbol{Y}} := \beta^{-1/2} \boldsymbol{Y}^\top.$$

This is a wide matrix with $\widetilde{a} = b$, $\widetilde{b} = a$, $\widetilde{\beta} = \beta^{-1}$, and $\widetilde{s}_\star = \beta^{-1/2} s_\star$. The original signal condition implies $s_\star^2 \ge \nu^2 \beta$, and hence $\widetilde{s}_\star^2 = s_\star^2/\beta \ge \nu^2$. Thus (81) applies to $\widetilde{\boldsymbol{Y}}$.

It remains only to identify its right-leakage quantities with the original left-leakage quantities. The empirical singular values satisfy $\widetilde{\widehat{s}}_i = \widehat{s}_i/\sqrt{\beta}$, and the definition of the bias-correction map gives $\widetilde{\tau}_i = \tau_i/\beta$. Consequently,

$$\widetilde{\boldsymbol{W}}_R = \beta^{-1/2} \boldsymbol{W}_L, \qquad \widetilde{\boldsymbol{N}}_R = \beta^{-1/2} \boldsymbol{N}_L, \qquad \widetilde{\boldsymbol{N}}_R^\top \widetilde{\boldsymbol{N}}_R - \boldsymbol{I}_p = \beta^{-1}\big(\boldsymbol{N}_L^\top \boldsymbol{N}_L - \beta \boldsymbol{I}_p\big). \tag{82}$$

Indeed, the right singular vectors of $\widetilde{\boldsymbol{Y}}$ are the left singular vectors of $\boldsymbol{Y}$, and $\widetilde{r}_i^2 = \ell_i^2/\beta$.

The centered right-leakage bound for $\widetilde{\boldsymbol{Y}}$ therefore gives

$$\left\|\boldsymbol{N}_L^\top \boldsymbol{N}_L - \beta \boldsymbol{I}_p\right\|_{L^{32}} \lesssim \beta \sqrt{\frac{p}{a}} = \frac{\sqrt{ap}}{b},$$

while the radial bound gives $\|\boldsymbol{N}_L\|_{L^{64}} \lesssim \sqrt{\beta}$. Moreover, $\widetilde{b}\,\widetilde{s}_\star^2 = a s_\star^2/\beta = b s_\star^2$. Since the tall-case signal condition forces $s_\star^2 > 1$, we have $s_\star^2 \wedge s_\star^4 = s_\star^2$, so the right-side localization bound gives (64). Finally, $\widetilde{r}_i^2 - 1 = (\ell_i^2 - \beta)/\beta$, and the right-side weight floor gives $\ell_i^2 - \beta \gtrsim s_\star^2 = s_\star^2 \wedge s_\star^4$. This proves all the claims in the tall case and completes the proof.

D.4 Proof of Lemma 15

We first prove the claim under the left signal condition $s_\star^2 \wedge s_\star^4 \ge \nu^2 \beta$. Introduce the auxiliary event

$$\mathcal{E}_0 := \left\{ \|\boldsymbol{Z}_{11}\|_{\mathrm{op}} \le \frac{s_\star}{4}, \quad \left\|\boldsymbol{Z}_{12}\boldsymbol{Z}_{12}^\top - \boldsymbol{I}_p\right\|_{\mathrm{op}} \le \frac{s_\star^2}{8}, \quad \left\|\boldsymbol{Z}_{22}\boldsymbol{Z}_{22}^\top - \boldsymbol{I}_a\right\|_{\mathrm{op}} \le \frac{s_\star^2}{128} \right\}. \tag{83}$$

Step 1: show that $\mathcal{E}_0 \subseteq \mathcal{E}_{\mathrm{loc}}$. Consider the first p rows of $\boldsymbol{Y}$, $\boldsymbol{M} := (\boldsymbol{\Sigma}_\star + \boldsymbol{Z}_{11}, \boldsymbol{Z}_{12})$. By the definition of singular values, $\widehat{s}_p^2 \geq \lambda_{\min}(\boldsymbol{M}\boldsymbol{M}^\top)$. On $\mathcal{E}_0$, Weyl's inequality gives $\sigma_{\min}(\boldsymbol{\Sigma}_\star + \boldsymbol{Z}_{11}) \geq 3s_\star/4$ and $\boldsymbol{Z}_{12}\boldsymbol{Z}_{12}^\top \succeq (1 - s_\star^2/8)\boldsymbol{I}_p$. Consequently, $\widehat{s}_p^2 \geq 1 + 7s_\star^2/16$.

We next translate this bound to the τ-coordinate. The signal condition implies $s_\star^2 \geq \nu\sqrt{\beta}$, so for sufficiently large ν, $\widehat{s}_i \geq \widehat{s}_p > 1 + \sqrt{\beta}$ for every i. Hence every $\widehat{s}_i$ lies on the upper branch of (60), so $\tau_i > \sqrt{\beta}$ and (61) applies. Moreover, the signal condition gives $1 + \beta + s_\star^2/16 + 16\beta/s_\star^2 \leq 1 + 7s_\star^2/16$. Since $t \mapsto 1 + \beta + t + \beta/t$ is increasing on $(\sqrt{\beta}, \infty)$, we obtain $\tau_i \geq s_\star^2/16$ for every i.

The second condition defining $\mathcal{E}_{\mathrm{loc}}$ is already part of $\mathcal{E}_0$, so $\mathcal{E}_0 \subseteq \mathcal{E}_{\mathrm{loc}}$.

Step 2: bound $\mathbb{P}(\mathcal{E}_0^c)$. Since $p/b \leq \beta/2$, $a/b = \beta$, and $s_\star^2 \wedge s_\star^4 \geq \nu^2\beta$, we have $\sqrt{p/b} \vee \sqrt{a/b} \leq \sqrt{\beta} \leq (s_\star^2 \wedge s_\star)/\nu$.

Apply (51) to $\boldsymbol{Z}_{11}$ with $u = v = p$ and $x = s_\star/8$. For sufficiently large ν, $2\sqrt{p/b} \leq s_\star/8$, and hence $\mathbb{P}\{\|\boldsymbol{Z}_{11}\|_{\mathrm{op}} > s_\star/4\} \lesssim \exp\{-cb(s_\star^2 \wedge s_\star^4)\}$.

The block dimensions in (69) are $\boldsymbol{Z}_{12} \in \mathbb{R}^{p\times b}$ and $\boldsymbol{Z}_{22} \in \mathbb{R}^{a\times b}$, and both matrices have independent $N(0, b^{-1})$ entries. Thus (52) applies with $u = p$ and $u = a$, respectively, while $v = b$ in both cases. Choose the tail parameter in that inequality to be $c_0 b(s_\star^2 \wedge s_\star^4)$, where $c_0 > 0$ is a sufficiently small numerical constant. For either $u \in \{p, a\}$, the resulting deviation threshold satisfies

$$C\left[\sqrt{\frac{u}{b} + c_0(s_\star^2 \wedge s_\star^4)} + \frac{u}{b} + c_0(s_\star^2 \wedge s_\star^4)\right] \leq C\left[\sqrt{\frac{u}{b}} + \sqrt{c_0}(s_\star^2 \wedge s_\star) + \frac{u}{b} + c_0(s_\star^2 \wedge s_\star^4)\right] \leq \frac{s_\star^2}{128}.$$

The last inequality follows from the preceding bound $\sqrt{u/b} \leq (s_\star^2 \wedge s_\star)/\nu$, first by choosing c_0 sufficiently small and then by choosing ν sufficiently large. Therefore the failure probabilities for the second and third conditions in (83) are each bounded by $2\exp\{-cb(s_\star^2 \wedge s_\star^4)\}$.

Therefore, a union bound and $\mathcal{E}_0 \subseteq \mathcal{E}_{\mathrm{loc}}$ give

$$\mathbb{P}(\mathcal{E}_{\mathrm{loc}}^c) \lesssim \exp\{-cb(s_\star^2 \wedge s_\star^4)\},$$

which proves the first claim.

The stronger right signal condition. Suppose now that $s_\star^2 \geq \nu^2$. Since $\nu > 1$ and $\beta \leq 1$, the left signal condition also holds, so the inclusion $\mathcal{E}_0 \subseteq \mathcal{E}_{\mathrm{loc}}$ proved above remains valid. In the preceding Gaussian estimates we may now use $s_\star^2 \wedge s_\star = s_\star$, which gives $\mathbb{P}(\mathcal{E}_{\mathrm{loc}}^c) \lesssim e^{-cbs_\star^2}$. This proves the second claim.

D.5 Proof of Lemma 16

On $\mathcal{E}_{\mathrm{loc}}$, the proof proceeds in four stages. We first solve the lower singular-vector equations by an absolutely convergent series and use that series to derive exact expansions for the left and right leakage Gram matrices. We then condition on $\boldsymbol{Z}_{22}$, replace the Gaussian coefficient matrices involving $\boldsymbol{Z}_{12}$ and $\boldsymbol{Z}_{21}$ by their conditional means, and collect the centered fluctuations. The remaining random coefficients are normalized traces of functions of $\boldsymbol{Z}_{22}$; we replace them by the deterministic Marchenko–Pastur transform. Finally, we control the centering and bulk-replacement errors in a common balanced norm and convert that bound into the left and right conclusions of Lemma 16.

The lower block rows of the singular-vector equations are

$$\boldsymbol{Z}_{21}\boldsymbol{V}_\parallel + \boldsymbol{Z}_{22}\boldsymbol{V}_\perp = \boldsymbol{U}_\perp\boldsymbol{\Sigma}, \qquad \boldsymbol{Z}_{12}^\top\boldsymbol{U}_\parallel + \boldsymbol{Z}_{22}^\top\boldsymbol{U}_\perp = \boldsymbol{V}_\perp\boldsymbol{\Sigma}. \tag{84}$$

Eliminating $\boldsymbol{V}_\perp$ gives

$$\boldsymbol{U}_\perp(\boldsymbol{\Sigma}^2 - \boldsymbol{I}_p) = (\boldsymbol{Z}_{22}\boldsymbol{Z}_{22}^\top - \boldsymbol{I}_a)\boldsymbol{U}_\perp + \boldsymbol{Z}_{21}\boldsymbol{V}_\parallel\boldsymbol{\Sigma} + \boldsymbol{Z}_{22}\boldsymbol{Z}_{12}^\top\boldsymbol{U}_\parallel. \tag{85}$$

To simplify notation in this section, we write

$$\boldsymbol{H}_{\text{bulk}} := \boldsymbol{Z}_{22}\boldsymbol{Z}_{22}^\top - \boldsymbol{I}_a, \qquad \boldsymbol{D}_\Sigma := \boldsymbol{\Sigma}^2 - \boldsymbol{I}_p.$$

On $\mathcal{E}_{\text{loc}}$, we have

$$\|\boldsymbol{H}_{\text{bulk}}\|_{\text{op}}\left\|\boldsymbol{D}_\Sigma^{-1}\right\|_{\text{op}} \leq \frac{1}{8}. \tag{86}$$

Indeed, (61) and the definition of $\mathcal{E}_{\text{loc}}$ give $\widehat{s}_i^2 - 1 \geq \tau_i \geq s_\star^2/16$ for every i. Consequently, $\boldsymbol{\Sigma}^2 - \boldsymbol{I}_p$ is positive definite and $\left\|\boldsymbol{D}_\Sigma^{-1}\right\|_{\text{op}} = \frac{1}{\min_i(\widehat{s}_i^2 - 1)} \leq \frac{16}{s_\star^2}$.

Thus, on $\mathcal{E}_{\text{loc}}$, (85) can be solved by a convergent Neumann series:

$$\boldsymbol{U}_\perp = \sum_{j\geq 0}\boldsymbol{H}_{\text{bulk}}^j(\boldsymbol{Z}_{21}\boldsymbol{V}_\parallel\boldsymbol{\Sigma} + \boldsymbol{Z}_{22}\boldsymbol{Z}_{12}^\top\boldsymbol{U}_\parallel)\boldsymbol{D}_\Sigma^{-j-1}. \tag{87}$$

Taking the Gram of this expansion gives the four-term expansion

$$\begin{aligned}\boldsymbol{U}_\perp^\top\boldsymbol{U}_\perp = \sum_{r,s\geq 0}\boldsymbol{D}_\Sigma^{-r-1}\Big(&\boldsymbol{\Sigma}\boldsymbol{V}_\parallel^\top\boldsymbol{Z}_{21}^\top\boldsymbol{H}_{\text{bulk}}^{r+s}\boldsymbol{Z}_{21}\boldsymbol{V}_\parallel\boldsymbol{\Sigma}\\ &+\boldsymbol{\Sigma}\boldsymbol{V}_\parallel^\top\boldsymbol{Z}_{21}^\top\boldsymbol{H}_{\text{bulk}}^{r+s}\boldsymbol{Z}_{22}\boldsymbol{Z}_{12}^\top\boldsymbol{U}_\parallel\\ &+\boldsymbol{U}_\parallel^\top\boldsymbol{Z}_{12}\boldsymbol{Z}_{22}^\top\boldsymbol{H}_{\text{bulk}}^{r+s}\boldsymbol{Z}_{21}\boldsymbol{V}_\parallel\boldsymbol{\Sigma}\\ &+\boldsymbol{U}_\parallel^\top\boldsymbol{Z}_{12}\boldsymbol{Z}_{22}^\top\boldsymbol{H}_{\text{bulk}}^{r+s}\boldsymbol{Z}_{22}\boldsymbol{Z}_{12}^\top\boldsymbol{U}_\parallel\Big)\boldsymbol{D}_\Sigma^{-s-1}.\end{aligned} \tag{88}$$

Note that (88) can be viewed as an equation linking the left leakage Gram matrix $\boldsymbol{U}_\perp^\top\boldsymbol{U}_\perp$ with the two in-signal components $\boldsymbol{U}_\parallel$ and $\boldsymbol{V}_\parallel$. While $\boldsymbol{D}_\Sigma$ and $\boldsymbol{\Sigma}$ are simple diagonal matrices, it is the terms like $\boldsymbol{Z}_{21}^\top\boldsymbol{H}_{\text{bulk}}^{r+s}\boldsymbol{Z}_{21}$ that is problematic as they depend on the noise matrices in a complicated way.

In what follows, we replace the terms like $\boldsymbol{Z}_{21}^\top\boldsymbol{H}_{\text{bulk}}^{r+s}\boldsymbol{Z}_{21}$ by simpler deterministic matrices in two steps: in the first step, we remove the non-bulk noises $\boldsymbol{Z}_{12}, \boldsymbol{Z}_{21}$ by centering them, while in the second step, we remove the bulk noise $\boldsymbol{Z}_{22}$.

D.5.1 Centering the non-bulk noise

The Gram expansion above contains four Gaussian coefficient matrices. The goal of this subsection is to split each coefficient into its conditional mean given $\boldsymbol{Z}_{22}$ and a centered fluctuation. The conditional means will form the empirical Gram operators, while all centered fluctuations will be collected into a remainder.

For $m \geq 0$, define the centered Gaussian coefficient matrices

$$\boldsymbol{C}_{U,11}^{(m)} := \boldsymbol{Z}_{21}^\top\boldsymbol{H}_{\text{bulk}}^m\boldsymbol{Z}_{21} - \frac{\operatorname{tr}(\boldsymbol{H}_{\text{bulk}}^m)}{b}\boldsymbol{I}_p, \tag{89a}$$

$$\boldsymbol{C}_{U,22}^{(m)} := \boldsymbol{Z}_{12}\boldsymbol{Z}_{22}^\top\boldsymbol{H}_{\text{bulk}}^m\boldsymbol{Z}_{22}\boldsymbol{Z}_{12}^\top - \frac{\operatorname{tr}(\boldsymbol{Z}_{22}^\top\boldsymbol{H}_{\text{bulk}}^m\boldsymbol{Z}_{22})}{b}\boldsymbol{I}_p, \tag{89b}$$

$$\boldsymbol{C}_{U,12}^{(m)} := \boldsymbol{Z}_{21}^\top\boldsymbol{H}_{\text{bulk}}^m\boldsymbol{Z}_{22}\boldsymbol{Z}_{12}^\top, \tag{89c}$$

$$\boldsymbol{C}_{U,21}^{(m)} := \boldsymbol{Z}_{12}\boldsymbol{Z}_{22}^\top\boldsymbol{H}_{\text{bulk}}^m\boldsymbol{Z}_{21}. \tag{89d}$$

Conditional on $\boldsymbol{Z}_{22}$, all four matrices are centered. Collecting the corresponding conditional means in (88) motivates the empirical operators

$$\widehat{\mathcal{A}}_U(\boldsymbol{M}) := \sum_{r,s\geq 0} \frac{\operatorname{tr}(\boldsymbol{H}_{\mathrm{bulk}}^{r+s})}{b} \boldsymbol{D}_\Sigma^{-r-1} \boldsymbol{\Sigma} \boldsymbol{M} \boldsymbol{\Sigma} \boldsymbol{D}_\Sigma^{-s-1}, \tag{90a}$$

$$\widehat{\mathcal{B}}(\boldsymbol{M}) := \sum_{r,s\geq 0} \frac{\operatorname{tr}(\boldsymbol{Z}_{22}^\top \boldsymbol{H}_{\mathrm{bulk}}^{r+s} \boldsymbol{Z}_{22})}{b} \boldsymbol{D}_\Sigma^{-r-1} \boldsymbol{M} \boldsymbol{D}_\Sigma^{-s-1}. \tag{90b}$$

With these definitions, the left leakage Gram admits the exact decomposition

$$\boldsymbol{U}_\perp^\top \boldsymbol{U}_\perp = \widehat{\mathcal{A}}_U\big(\boldsymbol{V}_\parallel^\top \boldsymbol{V}_\parallel\big) + \widehat{\mathcal{B}}\big(\boldsymbol{U}_\parallel^\top \boldsymbol{U}_\parallel\big) + \boldsymbol{R}_{U,\mathrm{coef}}, \tag{91}$$

where

$$\begin{aligned} \boldsymbol{R}_{U,\mathrm{coef}} := \sum_{r,s\geq 0} \boldsymbol{D}_\Sigma^{-r-1} \Big\{ & \boldsymbol{\Sigma} \boldsymbol{V}_\parallel^\top \boldsymbol{C}_{U,11}^{(r+s)} \boldsymbol{V}_\parallel \boldsymbol{\Sigma} + \boldsymbol{U}_\parallel^\top \boldsymbol{C}_{U,22}^{(r+s)} \boldsymbol{U}_\parallel \\ & + \boldsymbol{\Sigma} \boldsymbol{V}_\parallel^\top \boldsymbol{C}_{U,12}^{(r+s)} \boldsymbol{U}_\parallel + \boldsymbol{U}_\parallel^\top \boldsymbol{C}_{U,21}^{(r+s)} \boldsymbol{V}_\parallel \boldsymbol{\Sigma} \Big\} \boldsymbol{D}_\Sigma^{-s-1}. \end{aligned} \tag{92}$$

The right subspace leakage. We next carry out the same decomposition for $\boldsymbol{V}_\perp^\top \boldsymbol{V}_\perp$. The only additional bookkeeping comes from the base term $\boldsymbol{Z}_{12}^\top \boldsymbol{U}_\parallel \boldsymbol{\Sigma}^{-1}$ in the expansion of $\boldsymbol{V}_\perp$. We keep this term separate and, for $q \geq 0$, set $\boldsymbol{G}_q := \boldsymbol{Z}_{22}^\top \boldsymbol{H}_{\mathrm{bulk}}^q \boldsymbol{Z}_{22}$ and $\boldsymbol{E}_q := \boldsymbol{D}_\Sigma^{-q-1} \boldsymbol{\Sigma}^{-1}$. Substituting the convergent expansion of $\boldsymbol{U}_\perp$ into the second equation in (84) gives

$$\boldsymbol{V}_\perp = \sum_{r\geq 0} \boldsymbol{Z}_{22}^\top \boldsymbol{H}_{\mathrm{bulk}}^r \boldsymbol{Z}_{21} \boldsymbol{V}_\parallel \boldsymbol{D}_\Sigma^{-r-1} + \boldsymbol{Z}_{12}^\top \boldsymbol{U}_\parallel \boldsymbol{\Sigma}^{-1} + \sum_{q\geq 0} \boldsymbol{G}_q \boldsymbol{Z}_{12}^\top \boldsymbol{U}_\parallel \boldsymbol{E}_q. \tag{93}$$

We now center the four families of quadratic coefficients arising when the Gram of (93) is expanded.

For $r, s, q, u \geq 0$, define four centered terms as

$$\boldsymbol{C}_{V,11}^{(r,s)} := \boldsymbol{Z}_{21}^\top \boldsymbol{H}_{\mathrm{bulk}}^r (\boldsymbol{I}_a + \boldsymbol{H}_{\mathrm{bulk}}) \boldsymbol{H}_{\mathrm{bulk}}^s \boldsymbol{Z}_{21} - \frac{\operatorname{tr}\{\boldsymbol{H}_{\mathrm{bulk}}^r (\boldsymbol{I}_a + \boldsymbol{H}_{\mathrm{bulk}}) \boldsymbol{H}_{\mathrm{bulk}}^s\}}{b} \boldsymbol{I}_p, \tag{94a}$$

$$\boldsymbol{C}_{V,22}^{(q,u)} := \boldsymbol{Z}_{12} \boldsymbol{G}_q \boldsymbol{G}_u \boldsymbol{Z}_{12}^\top - \frac{\operatorname{tr}(\boldsymbol{G}_q \boldsymbol{G}_u)}{b} \boldsymbol{I}_p, \tag{94b}$$

$$\boldsymbol{C}_{V,12}^{(r,u)} := \boldsymbol{Z}_{21}^\top \boldsymbol{H}_{\mathrm{bulk}}^r \boldsymbol{Z}_{22} \boldsymbol{G}_u \boldsymbol{Z}_{12}^\top, \tag{94c}$$

$$\boldsymbol{C}_{V,21}^{(q,s)} := \boldsymbol{Z}_{12} \boldsymbol{G}_q \boldsymbol{Z}_{22}^\top \boldsymbol{H}_{\mathrm{bulk}}^s \boldsymbol{Z}_{21}. \tag{94d}$$

Again, these matrices are centered conditional on $\boldsymbol{Z}_{22}$. Collecting their conditional means leads to the remaining empirical operator

$$\begin{aligned} \widehat{\mathcal{A}}_V(\boldsymbol{M}) := \boldsymbol{\Sigma}^{-1} \boldsymbol{M} \boldsymbol{\Sigma}^{-1} + \sum_{q\geq 0} \frac{\operatorname{tr}(\boldsymbol{G}_q)}{b} \big(\boldsymbol{E}_q \boldsymbol{M} \boldsymbol{\Sigma}^{-1} + \boldsymbol{\Sigma}^{-1} \boldsymbol{M} \boldsymbol{E}_q\big) \\ + \sum_{q,u\geq 0} \frac{\operatorname{tr}(\boldsymbol{G}_q \boldsymbol{G}_u)}{b} \boldsymbol{E}_q \boldsymbol{M} \boldsymbol{E}_u. \end{aligned} \tag{95}$$

Expanding the Gram in (93) and collecting the centered terms therefore gives

$$\boldsymbol{V}_\perp^\top \boldsymbol{V}_\perp = \widehat{\mathcal{B}}\big(\boldsymbol{V}_\parallel^\top \boldsymbol{V}_\parallel\big) + \widehat{\mathcal{A}}_V\big(\boldsymbol{U}_\parallel^\top \boldsymbol{U}_\parallel\big) + \boldsymbol{R}_{V,\mathrm{coef}}, \tag{96}$$

where

$$\begin{aligned}
\boldsymbol{R}_{V,\mathrm{coef}} \coloneqq & \sum_{r,s\ge 0} \boldsymbol{D}_\Sigma^{-r-1}\boldsymbol{V}_\parallel^\top \boldsymbol{C}_{V,11}^{(r,s)}\boldsymbol{V}_\parallel \boldsymbol{D}_\Sigma^{-s-1} + \sum_{q,u\ge 0} \boldsymbol{E}_q \boldsymbol{U}_\parallel^\top \boldsymbol{C}_{V,22}^{(q,u)}\boldsymbol{U}_\parallel \boldsymbol{E}_u \\
& + \sum_{q\ge 0} \boldsymbol{E}_q \boldsymbol{U}_\parallel^\top \left(\boldsymbol{Z}_{12}\boldsymbol{G}_q\boldsymbol{Z}_{12}^\top - \frac{\mathrm{tr}(\boldsymbol{G}_q)}{b}\boldsymbol{I}_p\right)\boldsymbol{U}_\parallel \boldsymbol{\Sigma}^{-1} \\
& + \sum_{u\ge 0} \boldsymbol{\Sigma}^{-1}\boldsymbol{U}_\parallel^\top \left(\boldsymbol{Z}_{12}\boldsymbol{G}_u\boldsymbol{Z}_{12}^\top - \frac{\mathrm{tr}(\boldsymbol{G}_u)}{b}\boldsymbol{I}_p\right)\boldsymbol{U}_\parallel \boldsymbol{E}_u \\
& + \boldsymbol{\Sigma}^{-1}\boldsymbol{U}_\parallel^\top \left(\boldsymbol{Z}_{12}\boldsymbol{Z}_{12}^\top - \boldsymbol{I}_p\right)\boldsymbol{U}_\parallel \boldsymbol{\Sigma}^{-1} \\
& + \sum_{r,u\ge 0} \boldsymbol{D}_\Sigma^{-r-1}\boldsymbol{V}_\parallel^\top \boldsymbol{C}_{V,12}^{(r,u)}\boldsymbol{U}_\parallel \boldsymbol{E}_u + \sum_{r\ge 0} \boldsymbol{D}_\Sigma^{-r-1}\boldsymbol{V}_\parallel^\top \boldsymbol{Z}_{21}^\top \boldsymbol{H}_{\mathrm{bulk}}^r \boldsymbol{Z}_{22}\boldsymbol{Z}_{12}^\top \boldsymbol{U}_\parallel \boldsymbol{\Sigma}^{-1} \\
& + \sum_{q,s\ge 0} \boldsymbol{E}_q \boldsymbol{U}_\parallel^\top \boldsymbol{C}_{V,21}^{(q,s)}\boldsymbol{V}_\parallel \boldsymbol{D}_\Sigma^{-s-1} \\
& + \sum_{s\ge 0} \boldsymbol{\Sigma}^{-1}\boldsymbol{U}_\parallel^\top \boldsymbol{Z}_{12}\boldsymbol{Z}_{22}^\top \boldsymbol{H}_{\mathrm{bulk}}^s \boldsymbol{Z}_{21}\boldsymbol{V}_\parallel \boldsymbol{D}_\Sigma^{-s-1}.
\end{aligned} \tag{97}$$

D.5.2 Replacing the bulk traces

The preceding exact Gram equations have isolated the centered Gaussian coefficient fluctuations in $\boldsymbol{R}_{U,\mathrm{coef}}$ and $\boldsymbol{R}_{V,\mathrm{coef}}$. The remaining randomness in the empirical operators enters only through normalized traces of functions of the bulk matrix $\boldsymbol{Z}_{22}$. The goal of this subsection is to replace these random traces by their deterministic high-dimensional limits and to isolate the resulting replacement error.

Instead of working with the trace of bulk noise such as $\mathrm{tr}(\boldsymbol{H}_{\mathrm{bulk}}^m)$ in (90a) separately for each order m, we rely on the following resolvent

$$\widehat{m}_{\mathrm{bulk}}(z) \coloneqq \frac{1}{a}\,\mathrm{tr}(\boldsymbol{I}_a - z\boldsymbol{H}_{\mathrm{bulk}})^{-1},$$

which is defined on the contour

$$\mathcal{C}_\star \coloneqq \{z \in \mathbb{C} : |z| = 32/s_\star^2\}.$$

Also set $\boldsymbol{K}_0(z) \coloneqq \boldsymbol{D}_\Sigma^{-1}\left(\boldsymbol{I}_p - z^{-1}\boldsymbol{D}_\Sigma^{-1}\right)^{-1}$, and $\boldsymbol{K}_1(z) \coloneqq (1+z^{-1})\boldsymbol{K}_0(z)$. On $\mathcal{E}_{\mathrm{loc}}$, $\|z^{-1}\boldsymbol{D}_\Sigma^{-1}\|_{\mathrm{op}} \le 1/2$, so these matrices admit convergent geometric expansions. These matrices package the trace coefficients appearing in the empirical operators. Cauchy's coefficient formula yields the following contour representations:

$$\widehat{\mathcal{A}}_U(\boldsymbol{M}) = \frac{\beta}{2\pi \mathrm{i}}\oint_{\mathcal{C}_\star} \frac{\widehat{m}_{\mathrm{bulk}}(z)}{z}\,\boldsymbol{\Sigma}\boldsymbol{K}_0(z)\boldsymbol{M}\boldsymbol{K}_0(z)\boldsymbol{\Sigma}\,dz, \tag{98a}$$

$$\widehat{\mathcal{B}}(\boldsymbol{M}) = \frac{\beta}{2\pi \mathrm{i}}\oint_{\mathcal{C}_\star} \frac{\widehat{m}_{\mathrm{bulk}}(z)}{z}\,\boldsymbol{K}_1(z)\boldsymbol{M}\boldsymbol{K}_0(z)\,dz, \tag{98b}$$

$$\widehat{\mathcal{A}}_V(\boldsymbol{M}) = \boldsymbol{\Sigma}^{-1}\boldsymbol{M}\boldsymbol{\Sigma}^{-1} + \frac{\beta}{2\pi \mathrm{i}}\oint_{\mathcal{C}_\star} \frac{\widehat{m}_{\mathrm{bulk}}(z)}{z}\,\boldsymbol{\Sigma}^{-1}\big(\boldsymbol{K}_1(z)\boldsymbol{M} + \boldsymbol{M}\boldsymbol{K}_1(z) + \boldsymbol{K}_1(z)\boldsymbol{M}\boldsymbol{K}_1(z)\big)\boldsymbol{\Sigma}^{-1}\,dz. \tag{98c}$$

We next define the deterministic object that replaces these traces. Let m_β be the unique solution analytic at zero, with $m_\beta(0) = 1$, of

$$m_\beta(z) = 1 + \beta z\, m_\beta(z)\big\{(1+z)m_\beta(z) - 1\big\}. \tag{99}$$

Equivalently, m_β is the moment transform of the centered Marchenko–Pastur law with aspect ratio β.

Replacing $\widehat{m}_{\mathrm{bulk}}$ by m_β in the contour formulas above produces deterministic candidate operators. The next proposition identifies these operators with the population operators $\mathcal{A}_U$, $\mathcal{B}$, and $\mathcal{A}_V$ defined earlier in (74).

Proposition 1. *Recall the population operators defined in (74). On the event $\mathcal{E}_{\mathrm{loc}}$, we have*

$$\mathcal{A}_U(\boldsymbol{M}) = \frac{\beta}{2\pi\mathrm{i}} \oint_{\mathcal{C}_\star} \frac{m_\beta(z)}{z}\, \boldsymbol{\Sigma}\boldsymbol{K}_0(z)\boldsymbol{M}\boldsymbol{K}_0(z)\boldsymbol{\Sigma}\, dz, \tag{100a}$$

$$\mathcal{B}(\boldsymbol{M}) = \frac{\beta}{2\pi\mathrm{i}} \oint_{\mathcal{C}_\star} \frac{m_\beta(z)}{z}\, \boldsymbol{K}_1(z)\boldsymbol{M}\boldsymbol{K}_0(z)\, dz, \tag{100b}$$

$$\mathcal{A}_V(\boldsymbol{M}) = \boldsymbol{\Sigma}^{-1}\boldsymbol{M}\boldsymbol{\Sigma}^{-1} + \frac{\beta}{2\pi\mathrm{i}} \oint_{\mathcal{C}_\star} \frac{m_\beta(z)}{z}\, \boldsymbol{\Sigma}^{-1}\big(\boldsymbol{K}_1(z)\boldsymbol{M} + \boldsymbol{M}\boldsymbol{K}_1(z) + \boldsymbol{K}_1(z)\boldsymbol{M}\boldsymbol{K}_1(z)\big)\boldsymbol{\Sigma}^{-1}\, dz. \tag{100c}$$

Thus the difference between the exact empirical Gram operators and their population counterparts is entirely due to the replacement $\widehat{m}_{\mathrm{bulk}} \mapsto m_\beta$. We collect this difference in the bulk remainders

$$\boldsymbol{R}_{U,\mathrm{bulk}} := (\widehat{\mathcal{A}}_U - \mathcal{A}_U)\big(\boldsymbol{V}_\parallel^\top \boldsymbol{V}_\parallel\big) + (\widehat{\mathcal{B}} - \mathcal{B})\big(\boldsymbol{U}_\parallel^\top \boldsymbol{U}_\parallel\big), \tag{101a}$$

$$\boldsymbol{R}_{V,\mathrm{bulk}} := (\widehat{\mathcal{B}} - \mathcal{B})\big(\boldsymbol{V}_\parallel^\top \boldsymbol{V}_\parallel\big) + (\widehat{\mathcal{A}}_V - \mathcal{A}_V)\big(\boldsymbol{U}_\parallel^\top \boldsymbol{U}_\parallel\big). \tag{101b}$$

Combining these bulk remainders with the exact coefficient-centered Gram equations above, and recalling the definitions of $\boldsymbol{\varepsilon}_U$ and $\boldsymbol{\varepsilon}_V$, we obtain the desired decomposition

$$\boldsymbol{\varepsilon}_U = \boldsymbol{R}_{U,\mathrm{coef}} + \boldsymbol{R}_{U,\mathrm{bulk}}, \qquad \boldsymbol{\varepsilon}_V = \boldsymbol{R}_{V,\mathrm{coef}} + \boldsymbol{R}_{V,\mathrm{bulk}}.$$

D.5.3 Bounding the two errors

It remains to control these two sources of error separately. We first bound the centered Gaussian coefficient fluctuations, and then the error from replacing the empirical bulk traces by their high-dimensional limits.

It is convenient to control both errors first in a common balanced norm. Introduce

$$\boldsymbol{W}_{\mathrm{bal}} := \mathrm{diag}\big(\beta^{1/4}\sqrt{\tau_i}\big).$$

We will obtain bounds for $\boldsymbol{W}_L\boldsymbol{\varepsilon}_U\boldsymbol{W}_L$ and $\boldsymbol{W}_{\mathrm{bal}}\boldsymbol{\varepsilon}_V\boldsymbol{W}_{\mathrm{bal}}$. At the end of the subsection, deterministic comparisons between $\boldsymbol{W}_{\mathrm{bal}}$, $\boldsymbol{W}_L$, and $\boldsymbol{W}_R$ will convert this balanced estimate into the two bounds stated in Lemma 16.

Step 1: Bound the centering error. We first control the remainders $\boldsymbol{R}_{U,\mathrm{coef}}$ and $\boldsymbol{R}_{V,\mathrm{coef}}$, which collect the centered Gaussian coefficient fluctuations from the exact Gram expansions.

Lemma 18. *On $\mathcal{E}_{\mathrm{loc}}$,*

$$\|\mathbf{1}_{\mathcal{E}_{\mathrm{loc}}}\boldsymbol{W}_L\boldsymbol{R}_{U,\mathrm{coef}}\boldsymbol{W}_L\|_{L^{32}} + \|\mathbf{1}_{\mathcal{E}_{\mathrm{loc}}}\boldsymbol{W}_{\mathrm{bal}}\boldsymbol{R}_{V,\mathrm{coef}}\boldsymbol{W}_{\mathrm{bal}}\|_{L^{32}} \le C\beta\sqrt{\frac{p}{a}}. \tag{102}$$

Step 2: Bound the bulk replacement error We next control $\boldsymbol{R}_{U,\mathrm{bulk}}$ and $\boldsymbol{R}_{V,\mathrm{bulk}}$, which arise from replacing the empirical bulk transform $\widehat{m}_{\mathrm{bulk}}$ by its deterministic counterpart m_β.

Lemma 19. *Under the condition that $s_\star^2 \wedge s_\star^4 \geq \nu^2\beta$,*

$$\sup_{z\in\mathcal{C}_\star} \left\| \mathbf{1}_{\mathcal{E}_{\mathrm{loc}}} \left(\widehat{m}_{\mathrm{bulk}}(z) - m_\beta(z)\right) \right\|_{L^{32}} \leq \frac{C}{\sqrt{a}}. \tag{103}$$

We use this estimate to control the two bulk remainders in the same balanced norms as in the preceding subsection.

Lemma 20 (Bulk replacement remainder)**.** *Under the condition that $s_\star^2 \wedge s_\star^4 \geq \nu^2\beta$,*

$$\|\mathbf{1}_{\mathcal{E}_{\mathrm{loc}}} \boldsymbol{W}_L \boldsymbol{R}_{U,\mathrm{bulk}} \boldsymbol{W}_L\|_{L^{32}} + \|\mathbf{1}_{\mathcal{E}_{\mathrm{loc}}} \boldsymbol{W}_{\mathrm{bal}} \boldsymbol{R}_{V,\mathrm{bulk}} \boldsymbol{W}_{\mathrm{bal}}\|_{L^{32}} \leq C\frac{\sqrt{a}}{b}. \tag{104}$$

Step 3: Completing the Proof. We can now return to $\boldsymbol{\varepsilon}_U$ and $\boldsymbol{\varepsilon}_V$. Since $\boldsymbol{\varepsilon}_U = \boldsymbol{R}_{U,\mathrm{coef}} + \boldsymbol{R}_{U,\mathrm{bulk}}$, $\boldsymbol{\varepsilon}_V = \boldsymbol{R}_{V,\mathrm{coef}} + \boldsymbol{R}_{V,\mathrm{bulk}}$, and $\sqrt{a}/b \leq \sqrt{ap}/b$, (102) and (104) yield the balanced estimate

$$\|\mathbf{1}_{\mathcal{E}_{\mathrm{loc}}} \boldsymbol{W}_L \boldsymbol{\varepsilon}_U \boldsymbol{W}_L\|_{L^{32}} + \|\mathbf{1}_{\mathcal{E}_{\mathrm{loc}}} \boldsymbol{W}_{\mathrm{bal}} \boldsymbol{\varepsilon}_V \boldsymbol{W}_{\mathrm{bal}}\|_{L^{32}} \leq C\frac{\sqrt{ap}}{b}. \tag{105}$$

It remains only to translate this common balanced estimate into the two forms stated in Lemma 16.

Under the left signal condition, since $(\boldsymbol{W}_{\mathrm{bal}})_{ii}^2 = \sqrt{\beta}\,\tau_i \geq \beta$, we have $\beta\|\boldsymbol{\varepsilon}_V\|_{\mathrm{op}} \leq \|\boldsymbol{W}_{\mathrm{bal}}\boldsymbol{\varepsilon}_V\boldsymbol{W}_{\mathrm{bal}}\|_{\mathrm{op}}$. Together with (105), this proves (78).

Under the stronger right signal condition, $\tau_i \geq s_\star^2/16 \geq \nu^2/16$. The definitions of the three weights then give $\|\boldsymbol{W}_L^{-1}\|_{\mathrm{op}} \leq C$ and $\|\boldsymbol{W}_R\boldsymbol{W}_{\mathrm{bal}}^{-1}\|_{\mathrm{op}}^2 \leq C\beta^{-1/2}$. Consequently, we have

$$\|\mathbf{1}_{\mathcal{E}_{\mathrm{loc}}} \boldsymbol{W}_R \boldsymbol{\varepsilon}_V \boldsymbol{W}_R\|_{L^{32}} + \|\mathbf{1}_{\mathcal{E}_{\mathrm{loc}}} \boldsymbol{\varepsilon}_U\|_{L^{32}} \leq C\beta^{-1/2}\frac{\sqrt{ap}}{b} \leq C\sqrt{\frac{p}{b}},$$

where the final step uses $a = \beta b$. This proves (79).

D.6 Proof of Lemma 17

We first control $\|\boldsymbol{U}_\perp \boldsymbol{W}_L\|_{\mathrm{op}}$. Partition the coordinates into

$$\mathcal{I}_0 \coloneqq \{i : \tau_i = \sqrt{\beta}\}, \tag{106a}$$

$$\mathcal{I}_{\mathrm{h}} \coloneqq \left\{ i : \tau_i > \sqrt{\beta},\ \widehat{s}_i^2 - 1 \geq 2\left\| \boldsymbol{Z}_{22}\boldsymbol{Z}_{22}^\top - \boldsymbol{I}_a \right\|_{\mathrm{op}} \right\}, \tag{106b}$$

$$\mathcal{I}_\ell \coloneqq \left\{ i : \tau_i > \sqrt{\beta},\ \widehat{s}_i^2 - 1 < 2\left\| \boldsymbol{Z}_{22}\boldsymbol{Z}_{22}^\top - \boldsymbol{I}_a \right\|_{\mathrm{op}} \right\}. \tag{106c}$$

For $\mathcal{I} \subseteq [p]$, use the corresponding subscript to denote the associated column submatrices, and write $\boldsymbol{\Sigma}_{\mathcal{I}} = \mathrm{diag}(\widehat{s}_i : i \in \mathcal{I})$ and $\boldsymbol{W}_{L,\mathcal{I}} = \mathrm{diag}(\ell_i : i \in \mathcal{I})$.

Frozen coordinates. On $\mathcal{I}_0$, $\ell_i = \sqrt{\beta}$, and hence $\|\boldsymbol{U}_{\perp,\mathcal{I}_0}\boldsymbol{W}_{L,\mathcal{I}_0}\|_{\mathrm{op}} \leq \sqrt{\beta}$.

High coordinates. Restrict (85) to $\mathcal{I}_\mathrm{h}$. By definition of this set, $\|\boldsymbol{Z}_{22}\boldsymbol{Z}_{22}^\top - \boldsymbol{I}_a\|_\mathrm{op}\|(\boldsymbol{\Sigma}_{\mathcal{I}_\mathrm{h}}^2 - \boldsymbol{I})^{-1}\|_\mathrm{op} \le 1/2$. Moreover, (61) and the definition of ℓ_i give $\widehat{s}_i\ell_i/(\widehat{s}_i^2-1) = (\tau_i+\beta)/(\tau_i+\beta+\beta/\tau_i) \le 1$, and therefore also $\ell_i/(\widehat{s}_i^2-1) \le 1$. Multiplying the restricted equation by $(\boldsymbol{\Sigma}_{\mathcal{I}_\mathrm{h}}^2 - \boldsymbol{I})^{-1}\boldsymbol{W}_{L,\mathcal{I}_\mathrm{h}}$ and taking operator norms gives

$$\|\boldsymbol{U}_{\perp,\mathcal{I}_\mathrm{h}}\boldsymbol{W}_{L,\mathcal{I}_\mathrm{h}}\|_\mathrm{op} \le 2\left(\|\boldsymbol{Z}_{21}\|_\mathrm{op} + \left\|\boldsymbol{Z}_{22}\boldsymbol{Z}_{12}^\top\right\|_\mathrm{op}\right).$$

Low coordinates. Set $\boldsymbol{Z}_\mathrm{low} := [\boldsymbol{Z}_{21}, \boldsymbol{Z}_{22}]$. Restricting the first equation in (84) to $\mathcal{I}_\ell$ and using the orthonormality of the stacked right singular vectors gives $\|\boldsymbol{U}_{\perp,\mathcal{I}_\ell}\boldsymbol{W}_{L,\mathcal{I}_\ell}\|_\mathrm{op} \le \|\boldsymbol{Z}_\mathrm{low}\|_\mathrm{op}\max_{i\in\mathcal{I}_\ell}\ell_i/\widehat{s}_i$. The same identities give $\ell_i/\widehat{s}_i = \tau_i/(\tau_i+1) \le \tau_i \le \widehat{s}_i^2 - 1$; therefore, the definition of $\mathcal{I}_\ell$ yields

$$\|\boldsymbol{U}_{\perp,\mathcal{I}_\ell}\boldsymbol{W}_{L,\mathcal{I}_\ell}\|_\mathrm{op} \le 2\left\|\boldsymbol{Z}_{22}\boldsymbol{Z}_{22}^\top - \boldsymbol{I}_a\right\|_\mathrm{op}\|\boldsymbol{Z}_\mathrm{low}\|_\mathrm{op}.$$

Combining the three regimes. The preceding bounds give

$$\|\boldsymbol{U}_\perp\boldsymbol{W}_L\|_\mathrm{op} \lesssim \sqrt{\beta} + \|\boldsymbol{Z}_{21}\|_\mathrm{op} + \left\|\boldsymbol{Z}_{22}\boldsymbol{Z}_{12}^\top\right\|_\mathrm{op} + \left\|\boldsymbol{Z}_{22}\boldsymbol{Z}_{22}^\top - \boldsymbol{I}_a\right\|_\mathrm{op}\|\boldsymbol{Z}_\mathrm{low}\|_\mathrm{op}. \tag{107}$$

By (53), $\|\boldsymbol{Z}_{21}\|_{L^{128}} \lesssim \sqrt{\beta}$ and $\|\boldsymbol{Z}_\mathrm{low}\|_{L^{256}} \lesssim 1$. Applying (54) to the stacked Gaussian matrix formed from $\boldsymbol{Z}_{12}$ and $\boldsymbol{Z}_{22}$ also gives $\|\boldsymbol{Z}_{22}\boldsymbol{Z}_{12}^\top\|_{L^{128}} \lesssim \sqrt{\beta}$ and $\|\boldsymbol{Z}_{22}\boldsymbol{Z}_{22}^\top - \boldsymbol{I}_a\|_{L^{256}} \lesssim \sqrt{\beta}$. Thus (107) and Hölder's inequality yield $\|\boldsymbol{U}_\perp\boldsymbol{W}_L\|_{L^{128}} \lesssim \sqrt{\beta}$.

Right leakage. Let $\mathcal{I}_\mathrm{up} := \{i : \tau_i > \sqrt{\beta}\}$. On $\mathcal{I}_0$, $r_i = 1$, so $\|\boldsymbol{V}_{\perp,\mathcal{I}_0}\boldsymbol{W}_{R,\mathcal{I}_0}\|_\mathrm{op} \le 1$. On $\mathcal{I}_\mathrm{up}$, restrict the second equation in (84). By (61) and the definition of r_i, we have $r_i/\widehat{s}_i = \tau_i/(\tau_i+\beta) \le 1$, and hence $\|\boldsymbol{V}_{\perp,\mathcal{I}_\mathrm{up}}\boldsymbol{W}_{R,\mathcal{I}_\mathrm{up}}\|_\mathrm{op} \le \|\boldsymbol{Z}_{12}\|_\mathrm{op} + \|\boldsymbol{Z}_{22}\|_\mathrm{op}$. Consequently, $\|\boldsymbol{V}_\perp\boldsymbol{W}_R\|_\mathrm{op} \le 1 + \|\boldsymbol{Z}_{12}\|_\mathrm{op} + \|\boldsymbol{Z}_{22}\|_\mathrm{op}$, and (53) gives $\|\boldsymbol{V}_\perp\boldsymbol{W}_R\|_{L^{128}} \lesssim 1$.

D.7 Proofs of the Gram-approximation inputs

D.7.1 Proof of Proposition 1

Since $\boldsymbol{\Sigma}$, $\boldsymbol{K}_0(z)$, and $\boldsymbol{K}_1(z)$ are diagonal, it suffices to evaluate the contour formulas entrywise.

Pole locations and scalar identities. On $\mathcal{E}_\mathrm{loc}$, we have $\tau_i > \sqrt{\beta}$, so the upper-branch relation gives $\widehat{s}_i^2 - 1 = \tau_i + \beta + \beta/\tau_i$. Set $\zeta_i := (\widehat{s}_i^2 - 1)^{-1}$. Then $\zeta_i \le 16/s_\star^2$, so ζ_i lies inside $\mathcal{C}_\star$. The standing signal condition also ensures that m_β is analytic on and inside this contour.

Substituting $z = \zeta_i$ into (99) shows that the solution analytic at zero satisfies $m_\beta(\zeta_i) = (\widehat{s}_i^2 - 1)/(\tau_i + \beta)$. Thus, writing $k_{0,i}(z)$ and $k_{1,i}(z)$ for the ith diagonal entries of $\boldsymbol{K}_0(z)$ and $\boldsymbol{K}_1(z)$, respectively,

$$\begin{aligned} \zeta_i m_\beta(\zeta_i) &= \frac{1}{\tau_i+\beta}, \qquad (1+\zeta_i)m_\beta(\zeta_i) = \frac{\tau_i+1}{\tau_i}, \\ k_{0,i}(z) &= \frac{\zeta_i z}{z-\zeta_i}, \qquad k_{1,i}(z) = \frac{\zeta_i(1+z)}{z-\zeta_i}. \end{aligned} \tag{108}$$

The operators $\mathcal{A}_U$ and $\mathcal{B}$. Fix $i, j \in [p]$ and first suppose that $\zeta_i \ne \zeta_j$. The two integrands below have poles only at ζ_i and ζ_j. Using (108), the residue theorem gives

$$\frac{1}{2\pi\mathrm{i}}\oint_{\mathcal{C}_\star}\frac{m_\beta(z)}{z}\,k_{0,i}(z)k_{0,j}(z)\,dz = \zeta_i\zeta_j\frac{\zeta_i m_\beta(\zeta_i) - \zeta_j m_\beta(\zeta_j)}{\zeta_i - \zeta_j}$$

$$
\begin{aligned}
&= \frac{\tau_i\tau_j}{(\tau_i+\beta)(\tau_j+\beta)(\tau_i\tau_j-\beta)},\\
\frac{1}{2\pi\mathrm{i}}\oint_{\mathcal{C}_\star} \frac{m_\beta(z)}{z}\, k_{1,i}(z)k_{0,j}(z)\,dz &= \zeta_i\zeta_j \frac{(1+\zeta_i)m_\beta(\zeta_i)-(1+\zeta_j)m_\beta(\zeta_j)}{\zeta_i-\zeta_j}\\
&= \frac{1}{\tau_i\tau_j-\beta}.
\end{aligned}
$$

The same identities hold when $\zeta_i = \zeta_j$ by continuity, or equivalently by evaluating the resulting double pole.

The (i,j) entry of the right-hand side of (100a) is therefore $\beta\widehat{s}_i\widehat{s}_j\tau_i\tau_j M_{ij}/\{(\tau_i+\beta)(\tau_j+\beta)(\tau_i\tau_j-\beta)\}$. Since the weight definitions give $r_i/\widehat{s}_i = \tau_i/(\tau_i+\beta)$, this equals $\beta r_i r_j M_{ij}/(\tau_i\tau_j-\beta)$, which is $[\mathcal{A}_U(\boldsymbol{M})]_{ij}$. The second residue identity immediately gives $[\mathcal{B}(\boldsymbol{M})]_{ij} = \beta M_{ij}/(\tau_i\tau_j-\beta)$. This proves (100a) and (100b).

The operator $\mathcal{A}_V$. The remaining integrand has poles at $0, \zeta_i, \zeta_j$. At zero, $k_{1,i}(0) = k_{1,j}(0) = -1$, so the residue of

$$
\frac{m_\beta(z)}{z}\bigl\{k_{1,i}(z)+k_{1,j}(z)+k_{1,i}(z)k_{1,j}(z)\bigr\}
$$

is -1. Summing this residue with those at ζ_i and ζ_j, and again using (108), gives

$$
\frac{1}{2\pi\mathrm{i}}\oint_{\mathcal{C}_\star} \frac{m_\beta(z)}{z}\bigl\{k_{1,i}(z)+k_{1,j}(z)+k_{1,i}(z)k_{1,j}(z)\bigr\}\,dz = \frac{\tau_i+\tau_j+\beta+1}{\tau_i\tau_j-\beta}.
$$

As above, the identity extends to $\zeta_i = \zeta_j$ by continuity. Consequently, the coefficient multiplying M_{ij} in the right-hand side of (100c) is

$$
\frac{1}{\widehat{s}_i\widehat{s}_j}\left\{1+\beta\frac{\tau_i+\tau_j+\beta+1}{\tau_i\tau_j-\beta}\right\} = \frac{(\tau_i+\beta)(\tau_j+\beta)}{\widehat{s}_i\widehat{s}_j(\tau_i\tau_j-\beta)} = \frac{\ell_i\ell_j}{\tau_i\tau_j-\beta},
$$

where the last equality uses the direct identity $\widehat{s}_i\ell_i = \tau_i+\beta$. This is precisely $[\mathcal{A}_V(\boldsymbol{M})]_{ij}$ and proves (100c).

D.7.2 Proof of Lemma 18

Proof. We first bound the 11 contribution to $\boldsymbol{R}_{U,\mathrm{coef}}$, namely the term containing $\boldsymbol{C}_{U,11}^{(r+s)}$. The remaining terms can be controlled similarly. Minkowski's inequality and submultiplicativity give

$$
\begin{aligned}
&\left\| \mathbf{1}_{\mathcal{E}_{\mathrm{loc}}} \boldsymbol{W}_L \sum_{r,s\ge 0} \boldsymbol{D}_\Sigma^{-r-1}\boldsymbol{\Sigma}\boldsymbol{V}_\parallel^\top \boldsymbol{C}_{U,11}^{(r+s)}\boldsymbol{V}_\parallel\boldsymbol{\Sigma}\boldsymbol{D}_\Sigma^{-s-1}\boldsymbol{W}_L \right\|_{L^{32}}\\
&\quad\le \sum_{r,s\ge 0} \left\| \mathbf{1}_{\mathcal{E}_{\mathrm{loc}}} \left\|\boldsymbol{W}_L\boldsymbol{\Sigma}\boldsymbol{D}_\Sigma^{-1}\right\|_{\mathrm{op}}^2 \left\|\boldsymbol{D}_\Sigma^{-1}\right\|_{\mathrm{op}}^{r+s} \left\|\boldsymbol{C}_{U,11}^{(r+s)}\right\|_{\mathrm{op}} \right\|_{L^{32}}\\
&\quad\le C\sum_{r,s\ge 0} \left\| \mathbf{1}_{\mathcal{E}_{\mathrm{loc}}} \left\|\boldsymbol{D}_\Sigma^{-1}\right\|_{\mathrm{op}}^{r+s} \left\|\boldsymbol{C}_{U,11}^{(r+s)}\right\|_{\mathrm{op}} \right\|_{L^{32}}.
\end{aligned}
$$

Here we used $\|\boldsymbol{V}_\parallel\|_{\mathrm{op}} \le 1$ in the first inequality. For the second, (61) and the definition of ℓ_i give $\widehat{s}_i\ell_i/(\widehat{s}_i^2-1) \le 1$ on $\mathcal{E}_{\mathrm{loc}}$.

We next bound the centered coefficient itself. Conditional on $\boldsymbol{Z}_{22}$, apply (55) to $\boldsymbol{Z}_{21}$ with $\boldsymbol{A} = \boldsymbol{H}_{\mathrm{bulk}}^m$ and $q = 32$. Since $\|\boldsymbol{H}_{\mathrm{bulk}}^m\|_{\mathrm{F}} \le \sqrt{a}\, \|\boldsymbol{H}_{\mathrm{bulk}}\|_{\mathrm{op}}^m$, we obtain

$$\left(\mathbb{E}\left[\left\|\boldsymbol{C}_{U,11}^{(m)}\right\|_{\mathrm{op}}^{32} \,\middle|\, \boldsymbol{Z}_{22}\right]\right)^{1/32} \le \frac{C}{b}\left(\sqrt{p}\, \|\boldsymbol{H}_{\mathrm{bulk}}^m\|_{\mathrm{F}} + p\, \|\boldsymbol{H}_{\mathrm{bulk}}^m\|_{\mathrm{op}}\right) \le C\beta\sqrt{\frac{p}{a}}\, \|\boldsymbol{H}_{\mathrm{bulk}}\|_{\mathrm{op}}^m.$$

It remains to combine the coefficient estimate with the exterior powers of $\boldsymbol{D}_{\Sigma}^{-1}$. For $m = 0$, the preceding conditional estimate directly gives the required bound. Suppose that $m \ge 1$. If $\|\boldsymbol{H}_{\mathrm{bulk}}\|_{\mathrm{op}} = 0$, then $\boldsymbol{H}_{\mathrm{bulk}}^m = \boldsymbol{0}$ and hence $\boldsymbol{C}_{U,11}^{(m)} = \boldsymbol{0}$, so the contribution vanishes. On the complementary event $\{\|\boldsymbol{H}_{\mathrm{bulk}}\|_{\mathrm{op}} > 0\}$, the contraction bound (86) gives

$$\begin{aligned}\left\|\mathbf{1}_{\mathcal{E}_{\mathrm{loc}}} \|\boldsymbol{D}_{\Sigma}^{-1}\|_{\mathrm{op}}^m \left\|\boldsymbol{C}_{U,11}^{(m)}\right\|_{\mathrm{op}}\right\|_{L^{32}} &\le 8^{-m}\left\|\mathbf{1}\{\|\boldsymbol{H}_{\mathrm{bulk}}\|_{\mathrm{op}} > 0\}\frac{\left\|\boldsymbol{C}_{U,11}^{(m)}\right\|_{\mathrm{op}}}{\|\boldsymbol{H}_{\mathrm{bulk}}\|_{\mathrm{op}}^m}\right\|_{L^{32}} \\ &\le C\beta\sqrt{\frac{p}{a}}\, 8^{-m}.\end{aligned}$$

The second inequality follows by dividing the preceding conditional coefficient estimate by the positive, $\boldsymbol{Z}_{22}$-measurable quantity $\|\boldsymbol{H}_{\mathrm{bulk}}\|_{\mathrm{op}}^m$ and then integrating over $\boldsymbol{Z}_{22}$. Thus the displayed bound holds for every $m \ge 0$ without requiring a convention for division by zero.

All remaining contributions, except the explicitly displayed base–base term involving $\boldsymbol{Z}_{12}\boldsymbol{Z}_{12}^\top - \boldsymbol{I}_p$, follow from the same argument, and we omit the repeated details. Indeed, conditional on $\boldsymbol{Z}_{22}$, each coefficient is a centered Gaussian quadratic or bilinear form whose middle matrix has rank at most a. Its operator norm is bounded by $C(1 + \|\boldsymbol{Z}_{22}\|_{\mathrm{op}})^4$ times the corresponding power of $\|\boldsymbol{H}_{\mathrm{bulk}}\|_{\mathrm{op}}$. The relevant powers are m for the remaining left-Gram coefficients, $r+s$ for $\boldsymbol{C}_{V,11}^{(r,s)}$, $q+u$ for $\boldsymbol{C}_{V,22}^{(q,u)}$, q for either centered base–series coefficient in the $V, 22$ family, $r+u$ for $\boldsymbol{C}_{V,12}^{(r,u)}$, $q+s$ for $\boldsymbol{C}_{V,21}^{(q,s)}$, and r or s for the two base cross terms. Thus (55), the exterior-factor bounds, and (86) give the same geometrically summable bound as for the $U, 11$ contribution. Since $\|(1+\|\boldsymbol{Z}_{22}\|_{\mathrm{op}})^4\|_{L^{32}} \le C$ by (53), their total contribution is at most $C\beta\sqrt{\frac{p}{a}}$.

It remains to treat the base–base term. Using $\|\boldsymbol{U}_{\|}\|_{\mathrm{op}} \le 1$, $\|\boldsymbol{W}_{\mathrm{bal}}\boldsymbol{\Sigma}^{-1}\|_{\mathrm{op}} \le \beta^{1/4}$, and (54), its weighted contribution is at most

$$\|\boldsymbol{W}_{\mathrm{bal}}\boldsymbol{\Sigma}^{-1}\|_{\mathrm{op}}^2 \left\|\boldsymbol{Z}_{12}\boldsymbol{Z}_{12}^\top - \boldsymbol{I}_p\right\|_{L^{32}} \le C\sqrt{\beta}\sqrt{\frac{p}{b}} \le C\beta\sqrt{\frac{p}{a}},$$

where the last inequality uses $a = \beta b$. Combining these bounds proves (102). □

D.7.3 Proof of Lemma 19

Fix $z \in \mathcal{C}_\star$. The resolvent trace is well behaved on $\mathcal{E}_{\mathrm{loc}}$ but need not be globally Lipschitz as a function of the Gaussian matrix. We therefore introduce a cutoff that agrees with the resolvent trace on $\mathcal{E}_{\mathrm{loc}}$. Gaussian concentration will control the resulting statistic around its mean. We then use Gaussian integration by parts to show that this mean satisfies the Marchenko–Pastur equation up to an error of order a^{-1}, after which stability of that equation completes the proof.

A globally Lipschitz statistic. For $\boldsymbol{D} \in \mathbb{R}^{a\times b}$, set $\boldsymbol{H_D} := \boldsymbol{D}\boldsymbol{D}^\top - \boldsymbol{I}_a$. Choose a 4-Lipschitz function $\varphi : [0,\infty) \to [0,1]$ that equals one on $[0,1/2]$ and vanishes on $[3/4,\infty)$, and define

$$\chi(\boldsymbol{D}) := \varphi\left(\frac{64}{s_\star^2}\,\|\boldsymbol{H_D}\|_{\mathrm{op}}\right). \tag{109}$$

The cutoff restricts attention to matrices for which the resolvent is uniformly bounded. Indeed, if $\chi(\boldsymbol{D}) > 0$, then $\|\boldsymbol{H_D}\|_{\mathrm{op}} < 3s_\star^2/256$. Since $|z| = 32/s_\star^2$, the matrix $\boldsymbol{I}_a - z\boldsymbol{H_D}$ is invertible and

$$\left\|(\boldsymbol{I}_a - z\boldsymbol{H_D})^{-1}\right\|_{\mathrm{op}} \le \frac{8}{5}. \tag{110}$$

We may therefore define a function on all of $\mathbb{R}^{a\times b}$ by

$$F(\boldsymbol{D}) := \begin{cases} \dfrac{\chi(\boldsymbol{D})}{a}\,\mathrm{tr}(\boldsymbol{I}_a - z\boldsymbol{H_D})^{-1}, & \chi(\boldsymbol{D}) > 0, \\ 0, & \chi(\boldsymbol{D}) = 0. \end{cases} \tag{111}$$

By (72), $\chi(\boldsymbol{Z}_{22}) = 1$ on $\mathcal{E}_{\mathrm{loc}}$, and hence $F(\boldsymbol{Z}_{22}) = \widehat{m}_{\mathrm{bulk}}(z)$ on this event.

We next verify the Lipschitz estimate needed for Gaussian concentration. Take $\boldsymbol{D}_0, \boldsymbol{D}_1 \in \mathbb{R}^{a\times b}$ and interpolate along $\boldsymbol{D}_t = \boldsymbol{D}_0 + t(\boldsymbol{D}_1 - \boldsymbol{D}_0)$. Where the cutoff changes, $\|\boldsymbol{D}_t\|_{\mathrm{op}} \lesssim \sqrt{1+s_\star^2}$; differentiating $\left\|\boldsymbol{D}_t\boldsymbol{D}_t^\top - \boldsymbol{I}_a\right\|_{\mathrm{op}}$ along this segment therefore gives

$$|\chi(\boldsymbol{D}_1) - \chi(\boldsymbol{D}_0)| \lesssim \frac{\sqrt{1+s_\star^2}}{s_\star^2}\,\|\boldsymbol{D}_1 - \boldsymbol{D}_0\|_{\mathrm{F}}\,.$$

If both cutoffs are positive, the resolvent identity, together with (110), gives the same bound for the normalized traces:

$$\left|\frac{1}{a}\,\mathrm{tr}(\boldsymbol{I}_a - z\boldsymbol{H}_{\boldsymbol{D}_1})^{-1} - \frac{1}{a}\,\mathrm{tr}(\boldsymbol{I}_a - z\boldsymbol{H}_{\boldsymbol{D}_0})^{-1}\right| \lesssim \frac{\sqrt{1+s_\star^2}}{s_\star^2}\,\|\boldsymbol{D}_1 - \boldsymbol{D}_0\|_{\mathrm{F}}\,.$$

If exactly one cutoff vanishes, the first estimate and (110) control the difference of the two values of F; if both vanish, the difference is zero. Thus

$$|F(\boldsymbol{D}_1) - F(\boldsymbol{D}_0)| \lesssim \frac{\sqrt{1+s_\star^2}}{s_\star^2}\,\|\boldsymbol{D}_1 - \boldsymbol{D}_0\|_{\mathrm{F}} \tag{112}$$

for all $\boldsymbol{D}_0, \boldsymbol{D}_1 \in \mathbb{R}^{a\times b}$.

Fluctuation around the mean. Set $\bar{m} := \mathbb{E}F(\boldsymbol{Z}_{22})$. Since $F(\boldsymbol{Z}_{22}) = \widehat{m}_{\mathrm{bulk}}(z)$ on $\mathcal{E}_{\mathrm{loc}}$,

$$\left\|\mathbf{1}_{\mathcal{E}_{\mathrm{loc}}}\big(\widehat{m}_{\mathrm{bulk}}(z) - m_\beta(z)\big)\right\|_{L^{32}} \le \|F(\boldsymbol{Z}_{22}) - \bar{m}\|_{L^{32}} + |\bar{m} - m_\beta(z)|. \tag{113}$$

It remains to bound these two terms.

Let $\boldsymbol{G} = \sqrt{b}\,\boldsymbol{Z}_{22}$, so that the entries of $\boldsymbol{G}$ are independent standard Gaussian variables. By (112), the map $\boldsymbol{G} \mapsto F(\boldsymbol{G}/\sqrt{b})$ has Lipschitz constant at most

$$\frac{C\sqrt{1+s_\star^2}}{\sqrt{b}\,s_\star^2} = \frac{C}{\sqrt{a}}\left(\frac{\beta}{s_\star^4} + \frac{\beta}{s_\star^2}\right)^{1/2} \lesssim \frac{1}{\sqrt{a}},$$

where the last inequality uses $s_\star^2 \wedge s_\star^4 \ge \nu^2\beta$. Applying Gaussian concentration to the real and imaginary parts of F gives

$$\|F(\boldsymbol{Z}_{22}) - \bar{m}\|_{L^{32}} \lesssim \frac{1}{\sqrt{a}}. \tag{114}$$

An approximate Marchenko–Pastur equation. We next show that $\bar{m}$ satisfies the equation in (99) up to an error of order a^{-1}. In the remainder of the proof, $\boldsymbol{D}$ has the same distribution as $\boldsymbol{Z}_{22}$, and we write

$$\chi := \chi(\boldsymbol{D}), \qquad \boldsymbol{R}_z := (\boldsymbol{I}_a - z\boldsymbol{H}_{\boldsymbol{D}})^{-1}, \qquad m := \frac{1}{a}\operatorname{tr}(\boldsymbol{R}_z)$$

on the support of χ. Let $\boldsymbol{d}_\ell$ denote the ℓth column of $\boldsymbol{D}$, and set $h_{\ell i} := \boldsymbol{d}_\ell^\top \boldsymbol{R}_z \boldsymbol{e}_i$. The product $\chi h_{\ell i}$, extended by zero away from the support of χ, is globally Lipschitz by the same argument used for F. We may therefore apply Gaussian integration by parts to it.

The identity $\boldsymbol{R}_z^{-1} = (1+z)\boldsymbol{I}_a - z\boldsymbol{D}\boldsymbol{D}^\top$ gives

$$m = 1 - zm + \frac{z}{a}\sum_{i=1}^{a}\sum_{\ell=1}^{b} D_{i\ell}h_{\ell i}. \tag{115}$$

To evaluate the expectation of the last term, differentiate $h_{\ell i}$. At every point of differentiability of χ,

$$\partial_{i\ell}h_{\ell i} = (\boldsymbol{R}_z)_{ii} + zh_{\ell i}^2 + z(\boldsymbol{d}_\ell^\top \boldsymbol{R}_z \boldsymbol{d}_\ell)(\boldsymbol{R}_z)_{ii}.$$

The three sums produced by this identity reduce to

$$\begin{aligned} \sum_{i,\ell}(\boldsymbol{R}_z)_{ii} &= abm, \\ \sum_{i,\ell} h_{\ell i}^2 &= a\{m + (1+z)m'\}, \\ \sum_{i,\ell}(\boldsymbol{d}_\ell^\top \boldsymbol{R}_z \boldsymbol{d}_\ell)(\boldsymbol{R}_z)_{ii} &= \frac{a^2}{z}m\{(1+z)m - 1\}, \end{aligned} \tag{116}$$

where $m' = \partial_z m$. For the second identity, use $\sum_{i,\ell} h_{\ell i}^2 = \operatorname{tr}(\boldsymbol{D}\boldsymbol{D}^\top \boldsymbol{R}_z^2)$, $\boldsymbol{R}_z^2 - \boldsymbol{R}_z = z\boldsymbol{H}_{\boldsymbol{D}}\boldsymbol{R}_z^2$, and $m' = a^{-1}\operatorname{tr}(\boldsymbol{H}_{\boldsymbol{D}}\boldsymbol{R}_z^2)$. The third follows from $\operatorname{tr}(\boldsymbol{D}\boldsymbol{D}^\top \boldsymbol{R}_z) = a\{(1+z)m - 1\}/z$.

Multiply (115) by χ, take expectations, and apply

$$\mathbb{E}[D_{i\ell}g(\boldsymbol{D})] = \frac{1}{b}\mathbb{E}[\partial_{i\ell}g(\boldsymbol{D})]$$

to $g = \chi h_{\ell i}$. The product rule and (116) give

$$\begin{aligned} \bar{m} = \mathbb{E}\chi &+ \beta z\mathbb{E}\left[\chi m\{(1+z)m - 1\}\right] \\ &+ \frac{z^2}{b}\mathbb{E}\left[\chi\{m + (1+z)m'\}\right] + \frac{z}{ab}\sum_{i,\ell}\mathbb{E}\left[h_{\ell i}\partial_{i\ell}\chi\right]. \end{aligned} \tag{117}$$

The term generated by the first line of (116) cancels the term $-z\bar{m}$ from (115). This cancellation is why the remaining leading term has exactly the form of the Marchenko–Pastur equation.

The error in the fixed-point equation. Define

$$r(z) := \bar{m} - 1 - \beta z\bar{m}\big\{(1+z)\bar{m} - 1\big\}.$$

Comparing (117) with (99) reduces the problem to bounding this residual. Rearranging (117) gives the central decomposition

$$\begin{aligned} r(z) = \mathbb{E}\chi - 1 + \beta z(1+z)\big\{\mathbb{E}[\chi m^2] - \bar{m}^2\big\} \\ + \frac{z^2}{b}\mathbb{E}\left[\chi\{m + (1+z)m'\}\right] + \frac{z}{ab}\sum_{i,\ell}\mathbb{E}\left[h_{\ell i}\partial_{i\ell}\chi\right]. \end{aligned} \tag{118}$$

We bound the four terms in this order.

First, $\chi \neq 1$ implies $\|\boldsymbol{H}_{\boldsymbol{Z}_{22}}\|_{\mathrm{op}} > s_\star^2/128$. The Gaussian Wishart deviation bound and $b(s_\star^2 \wedge s_\star^4) \geq \nu^2 a$ therefore give

$$\mathbb{P}\{\chi(\boldsymbol{Z}_{22}) \neq 1\} \leq C\exp\{-cb(s_\star^2 \wedge s_\star^4)\} \leq Ce^{-ca}. \tag{119}$$

In particular, $|\mathbb{E}\chi - 1| \lesssim a^{-1}$.

For the second term, the exact identity

$$\mathbb{E}[\chi m^2] - \bar{m}^2 = \mathbb{E}[(F-\bar{m})^2] + \mathbb{E}[\chi(1-\chi)m^2]$$

is valid also for complex m. Using (110), (114), and (119), we obtain

$$\left|\mathbb{E}[\chi m^2] - \bar{m}^2\right| \leq \mathbb{E}|F - \bar{m}|^2 + \frac{64}{25}\mathbb{P}\{\chi \neq 1\} \lesssim \frac{1}{a}. \tag{120}$$

Moreover, $\beta|z||1+z| \lesssim \beta/s_\star^2 + \beta/s_\star^4 \lesssim 1$, so the second term in (118) is also of order a^{-1}.

It remains to bound the last two terms. On the support of χ, $m' = a^{-1}\operatorname{tr}(\boldsymbol{H}_{\boldsymbol{D}}\boldsymbol{R}_z^2)$, and hence $|m'| \lesssim s_\star^2$. Where the cutoff changes, $\|\nabla_{\boldsymbol{D}}\chi\|_{\mathrm{F}} \lesssim \sqrt{1+s_\star^2}/s_\star^2$, while $\left\|\boldsymbol{D}^\top\boldsymbol{R}_z\right\|_{\mathrm{F}} \lesssim \sqrt{a(1+s_\star^2)}$. Since $(h_{\ell i})_{\ell,i} = \boldsymbol{D}^\top\boldsymbol{R}_z$, these estimates give

$$\begin{aligned} \frac{|z|^2}{b}\mathbb{E}\left[\chi|m + (1+z)m'|\right] &\lesssim \frac{1+s_\star^2}{bs_\star^4} = \frac{1}{a}\left(\frac{\beta}{s_\star^4} + \frac{\beta}{s_\star^2}\right) \lesssim \frac{1}{a}, \\ \frac{|z|}{ab}\left|\sum_{i,\ell}\mathbb{E}\left[h_{\ell i}\partial_{i\ell}\chi\right]\right| &\leq \frac{|z|}{ab}\mathbb{E}\left[\left\|\boldsymbol{D}^\top\boldsymbol{R}_z\right\|_{\mathrm{F}}\|\nabla_{\boldsymbol{D}}\chi\|_{\mathrm{F}}\right] \\ &\lesssim \frac{1}{a\sqrt{a}}\left(\frac{\beta}{s_\star^4} + \frac{\beta}{s_\star^2}\right) \lesssim \frac{1}{a}. \end{aligned} \tag{121}$$

Combining (119) to (121) in (118) proves

$$|r(z)| \lesssim \frac{1}{a}. \tag{122}$$

Stability of the Marchenko–Pastur equation. It remains to convert the residual bound into a bound for $\bar{m} - m_\beta(z)$. By (110), $|F(\boldsymbol{D})| \leq 8/5$ for every $\boldsymbol{D}$, and therefore $|\bar{m}| \leq 8/5$. The centered Marchenko–Pastur law is supported on $[\beta - 2\sqrt{\beta}, \beta + 2\sqrt{\beta}]$. Since $s_\star^2 \geq \nu\sqrt{\beta}$, the signal condition gives, for sufficiently large ν,

$$|z|(\beta + 2\sqrt{\beta}) \leq C\frac{\sqrt{\beta}}{s_\star^2} \leq \frac{1}{4}.$$

The integral representation of m_β consequently gives $|m_\beta(z)| \leq 4/3$.

Subtracting (99) from the definition of $r(z)$ gives

$$(\bar{m} - m_\beta(z))\left[1 - \beta z\big\{(1+z)(\bar{m} + m_\beta(z)) - 1\big\}\right] = r(z). \tag{123}$$

The bounds on $\bar{m}$ and $m_\beta(z)$ imply

$$\left|\beta z\big\{(1+z)(\bar{m} + m_\beta(z)) - 1\big\}\right| \lesssim \frac{\beta}{s_\star^2} + \frac{\beta}{s_\star^4} \le \frac{C}{\nu^2}.$$

For sufficiently large ν, the bracket in (123) has modulus at least 1/2. Hence (122) and (123) yield

$$|\bar{m} - m_\beta(z)| \lesssim \frac{1}{a}. \tag{124}$$

Finally, (113), (114), and (124) give

$$\left\|\mathbf{1}_{\mathcal{E}_{\mathrm{loc}}}\big(\widehat{m}_{\mathrm{bulk}}(z) - m_\beta(z)\big)\right\|_{L^{32}} \lesssim \frac{1}{\sqrt{a}}.$$

Every bound above is uniform over $z \in \mathcal{C}_\star$, because the argument uses z only through $|z| = 32/s_\star^2$. Taking the supremum over the contour proves (103).

D.7.4 Proof of Lemma 20

Each term in the two bulk remainders has the same structure: the contour integrand contains the scalar difference $\widehat{m}_{\mathrm{bulk}}(z) - m_\beta(z)$ multiplied by a bounded matrix kernel. We give the complete argument for the $\mathcal{A}_U$ term and then record the kernel bounds needed for the other three terms.

By (98a) and (100a),

$$(\widehat{\mathcal{A}}_U - \mathcal{A}_U)(\boldsymbol{V}_\parallel^\top \boldsymbol{V}_\parallel) = \frac{\beta}{2\pi\mathrm{i}} \oint_{\mathcal{C}_\star} \frac{\widehat{m}_{\mathrm{bulk}}(z) - m_\beta(z)}{z}\, \boldsymbol{\Sigma}\boldsymbol{K}_0(z)\boldsymbol{V}_\parallel^\top \boldsymbol{V}_\parallel \boldsymbol{K}_0(z)\boldsymbol{\Sigma}\, dz.$$

On $\mathcal{E}_{\mathrm{loc}}$, the identities $\widehat{s}_i^2 - 1 = \tau_i + \beta + \beta/\tau_i$ and $\widehat{s}_i \ell_i = \tau_i + \beta$ give $\|\boldsymbol{W}_L \boldsymbol{\Sigma}\boldsymbol{K}_0(z)\|_{\mathrm{op}} \le C$ uniformly over $z \in \mathcal{C}_\star$. Since $\|\boldsymbol{V}_\parallel\|_{\mathrm{op}} \le 1$, Minkowski's integral inequality and (103) yield

$$\begin{aligned}
&\left\|\mathbf{1}_{\mathcal{E}_{\mathrm{loc}}} \boldsymbol{W}_L(\widehat{\mathcal{A}}_U - \mathcal{A}_U)(\boldsymbol{V}_\parallel^\top \boldsymbol{V}_\parallel)\boldsymbol{W}_L\right\|_{L^{32}} \\
&\quad \le \frac{\beta}{2\pi} \oint_{\mathcal{C}_\star} \frac{1}{|z|} \left\|\mathbf{1}_{\mathcal{E}_{\mathrm{loc}}} |\widehat{m}_{\mathrm{bulk}}(z) - m_\beta(z)|\, \|\boldsymbol{W}_L \boldsymbol{\Sigma}\boldsymbol{K}_0(z)\|_{\mathrm{op}}^2\right\|_{L^{32}} |dz| \\
&\quad \le C\frac{\beta}{\sqrt{a}} = C\frac{\sqrt{a}}{b},
\end{aligned}$$

where the last identity uses $a = \beta b$.

The remaining three operator differences are controlled identically, so we omit the repeated contour calculations. The only changes are the matrix kernels. On $\mathcal{E}_{\mathrm{loc}}$, uniformly over $z \in \mathcal{C}_\star$, the same scalar identities give

$$\begin{gathered}
|1 + z^{-1}|\, \|\boldsymbol{W}_L \boldsymbol{K}_0(z)\|_{\mathrm{op}}^2 \le C, \\
|1 + z^{-1}|\, \|\boldsymbol{W}_{\mathrm{bal}} \boldsymbol{K}_0(z)\|_{\mathrm{op}}^2 \le C, \\
|1 + z^{-1}| \left\|\boldsymbol{W}_{\mathrm{bal}} \boldsymbol{\Sigma}^{-1} \boldsymbol{K}_0(z)\right\|_{\mathrm{op}} \left\|\boldsymbol{W}_{\mathrm{bal}} \boldsymbol{\Sigma}^{-1}\right\|_{\mathrm{op}} + |1 + z^{-1}|^2 \left\|\boldsymbol{W}_{\mathrm{bal}} \boldsymbol{\Sigma}^{-1} \boldsymbol{K}_0(z)\right\|_{\mathrm{op}}^2 \le C.
\end{gathered}$$

The first bound controls the $\widehat{\mathcal{B}} - \mathcal{B}$ term in $\boldsymbol{R}_{U,\mathrm{bulk}}$, the second controls the corresponding term in $\boldsymbol{R}_{V,\mathrm{bulk}}$, and the third controls $\widehat{\mathcal{A}}_V - \mathcal{A}_V$. For the last difference, the base term $\boldsymbol{\Sigma}^{-1}\boldsymbol{M}\boldsymbol{\Sigma}^{-1}$ cancels between the empirical and population operators. Using $\|\boldsymbol{U}_\|\|_{\mathrm{op}}, \|\boldsymbol{V}_\|\|_{\mathrm{op}} \le 1$, each of these three terms is therefore bounded by the same contour argument by $C\beta/\sqrt{a} = C\sqrt{a}/b$.

Finally, apply the triangle inequality to the decompositions in (101a) and (101b). Combining the four operator bounds gives (104).

E Completion of the upper-bound proof

We now prove Lemma 5. We establish its three conclusions in order.

E.1 Proof of (34a)

Recall that $\boldsymbol{R}_k = \boldsymbol{S}_k^\dagger \boldsymbol{U}$. Using the two blocks of $\boldsymbol{Q}_\star$, we obtain

$$(\boldsymbol{L}\boldsymbol{J} - \boldsymbol{T}\boldsymbol{L})\boldsymbol{Q}_\star = \alpha_1 \boldsymbol{L}_1 \boldsymbol{R}_1 + \alpha_2 \boldsymbol{L}_2 \boldsymbol{R}_2 - \boldsymbol{T}\big(\boldsymbol{L}_1\boldsymbol{R}_1 - \boldsymbol{L}_2\boldsymbol{R}_2\big). \tag{125}$$

Step 1: Bound $\alpha_1 \boldsymbol{L}_1 \boldsymbol{R}_1 + \alpha_2 \boldsymbol{L}_2 \boldsymbol{R}_2$. The first two terms can be bounded directly. On $\mathcal{E}_{\mathrm{emb}}$, (39a) gives $\|\boldsymbol{R}_k\|_{\mathrm{op}} \lesssim \gamma_k^{-1}$, while (33) gives $\|\boldsymbol{L}_k\|_{L^{64}} \lesssim \sigma\sqrt{a_k} \lesssim \sigma\sqrt{n}$. Since $\mathcal{E}_{\mathrm{good}} \subseteq \mathcal{E}_{\mathrm{emb}}$, these estimates imply

$$\left\|\mathbf{1}_{\mathcal{E}_{\mathrm{good}}}(\alpha_1 \boldsymbol{L}_1 \boldsymbol{R}_1 + \alpha_2 \boldsymbol{L}_2 \boldsymbol{R}_2)\right\|_{L^{32}} \lesssim \sigma\sqrt{n}\left(\frac{\alpha_1}{\gamma_1} + \frac{\alpha_2}{\gamma_2}\right) \lesssim \frac{\sigma\sqrt{n}}{\gamma_{\max}}.$$

The last inequality uses $\alpha_1/\gamma_1 + \alpha_2/\gamma_2 = (\gamma_1+\gamma_2)/(\gamma_1^2+\gamma_2^2) \lesssim \gamma_{\max}^{-1}$.

Step 2: Bound $\boldsymbol{T}\big(\boldsymbol{L}_1\boldsymbol{R}_1 - \boldsymbol{L}_2\boldsymbol{R}_2\big)$. The same argument is too crude for the term multiplied by $\boldsymbol{T}$. Submultiplicativity would bound $\|\boldsymbol{T}\boldsymbol{L}_k\boldsymbol{R}_k\|_{\mathrm{op}}$ by $\|\boldsymbol{T}\|_{\mathrm{op}}\|\boldsymbol{L}_k\|_{\mathrm{op}}\|\boldsymbol{R}_k\|_{\mathrm{op}}$ and would therefore retain the factor $\sqrt{a_k}$ from the bound for $\boldsymbol{L}_k$. To obtain the rank dependence in (34a), we need to leverage the rotational invariance of $\boldsymbol{L}_k$.

For $k \in \{1,2\}$, define

$$\mathcal{H}_k := \sigma\left(\boldsymbol{S}_1, \boldsymbol{S}_2, \boldsymbol{L}_{3-k}, \boldsymbol{L}_k^\top \boldsymbol{L}_k\right). \tag{126}$$

The sigma-field $\mathcal{H}_k$ records the quantities that remain unchanged when $\boldsymbol{L}_k$ is left multiplied by an orthogonal transformation on $\mathcal{M}_k^\perp$. The following lemma uses this invariance to bound a product in terms of the operator and Frobenius norms of the matrices on either side of $\boldsymbol{L}_k$.

Lemma 21. *Fix $k \in \{1,2\}$ and $q \ge 2$. Let $\mathcal{E} \in \mathcal{H}_k$, and let $\boldsymbol{\Psi} \in \mathbb{R}^{n\times u}$ and $\boldsymbol{\Theta} \in \mathbb{R}^{p_k \times v}$ be $\mathcal{H}_k$-measurable. Then*

$$\left\|\mathbf{1}_{\mathcal{E}}\boldsymbol{\Psi}^\top \boldsymbol{L}_k \boldsymbol{\Theta}\right\|_{L^q} \lesssim \sqrt{\frac{q}{a_k}}\,\|\boldsymbol{L}_k\|_{L^{2q}}\left\|\mathbf{1}_{\mathcal{E}}\left(\|\boldsymbol{\Psi}\|_{\mathrm{F}}\|\boldsymbol{\Theta}\|_{\mathrm{op}} + \|\boldsymbol{\Psi}\|_{\mathrm{op}}\|\boldsymbol{\Theta}\|_{\mathrm{F}}\right)\right\|_{L^{2q}}. \tag{127}$$

We defer the proof until after the proof of Lemma 5. The lemma does not require $\boldsymbol{\Psi}$ and $\boldsymbol{\Theta}$ to be independent of $\boldsymbol{L}_k$; their $\mathcal{H}_k$-measurability is sufficient. Taking $q = 32$ in (127) and using $\|\boldsymbol{L}_k\|_{L^{64}} \lesssim \sigma\sqrt{a_k}$ gives the form used below:

$$\left\|\mathbf{1}_{\mathcal{E}}\boldsymbol{\Psi}^\top \boldsymbol{L}_k \boldsymbol{\Theta}\right\|_{L^{32}} \lesssim \sigma\left\|\mathbf{1}_{\mathcal{E}}\left(\|\boldsymbol{\Psi}\|_{\mathrm{F}}\|\boldsymbol{\Theta}\|_{\mathrm{op}} + \|\boldsymbol{\Psi}\|_{\mathrm{op}}\|\boldsymbol{\Theta}\|_{\mathrm{F}}\right)\right\|_{L^{64}}. \tag{128}$$

Since $\mathcal{E}_{\mathrm{emb}}$ and $\boldsymbol{R}_k = \boldsymbol{S}_k^{\dagger}\boldsymbol{U}$ are $\mathcal{H}_k$-measurable and $\boldsymbol{T}$ is deterministic, we may apply (128) with $\boldsymbol{\Psi} = \boldsymbol{T}^{\top}$ and $\boldsymbol{\Theta} = \boldsymbol{R}_k$. Thus

$$\begin{aligned}
\left\|\mathbf{1}_{\mathcal{E}_{\mathrm{good}}}\boldsymbol{T}\big(\boldsymbol{L}_1\boldsymbol{R}_1 - \boldsymbol{L}_2\boldsymbol{R}_2\big)\right\|_{L^{32}} &\le \sum_{k=1}^{2} \left\|\mathbf{1}_{\mathcal{E}_{\mathrm{emb}}}\boldsymbol{T}\boldsymbol{L}_k\boldsymbol{R}_k\right\|_{L^{32}} \\
&\overset{\text{(i)}}{\lesssim} \sigma \sum_{k=1}^{2}\Big(\|\boldsymbol{T}\|_{\mathrm{F}}\,\|\boldsymbol{R}_k\|_{\mathrm{op}} + \|\boldsymbol{T}\|_{\mathrm{op}}\,\|\boldsymbol{R}_k\|_{\mathrm{F}}\Big) \\
&\overset{\text{(ii)}}{\lesssim} \sigma \sum_{k=1}^{2}\left\{\frac{\sqrt{r_1+r_2}}{\gamma_k} + \frac{\sqrt{r_1\wedge r_2}+\sqrt{r}}{\gamma_k\sqrt{\theta}}\right\} \\
&\lesssim \frac{\sigma\sqrt{r_1+r_2}}{\gamma_{\min}} + \frac{\sigma\sqrt{p_{\min}}}{\gamma_{\min}\sqrt{\theta}}.
\end{aligned}$$

Here (i) is (128). For (ii), we use (45) in the forms $\|\boldsymbol{T}\|_{\mathrm{F}} \lesssim \sqrt{r_1+r_2} + \sqrt{r_1\wedge r_2}/\sqrt{\theta}$ and $\|\boldsymbol{T}\|_{\mathrm{op}} \lesssim \theta^{-1/2}$, together with (39a) in the forms $\|\boldsymbol{R}_k\|_{\mathrm{op}} \lesssim \gamma_k^{-1}$ and $\|\boldsymbol{R}_k\|_{\mathrm{F}} \lesssim \sqrt{r}/\gamma_k$. The last inequality uses $\gamma_k \ge \gamma_{\min}$ and $\sqrt{r+(r_1\wedge r_2)} \asymp \sqrt{p_{\min}}$.

Combining this estimate with the bound for the first two terms in (125) and using $L^1 \le L^{32}$ proves (34a).

E.2 Proof of (34b)

Write $\boldsymbol{\Delta}_{\mathrm{quad}} \coloneqq \boldsymbol{L}^{\top}\boldsymbol{L} - \boldsymbol{D}_{\sigma}$. By (32), it is enough to bound $\boldsymbol{\Delta}_{\mathrm{quad}}$, $\boldsymbol{P}_{\mathrm{col}(\boldsymbol{S})}\boldsymbol{L}$, and $\boldsymbol{\Delta}_{+0}$.

The matrix $\boldsymbol{\Delta}_{\mathrm{quad}}$. The diagonal blocks of $\boldsymbol{\Delta}_{\mathrm{quad}}$ are $\boldsymbol{L}_k^{\top}\boldsymbol{L}_k - a_k\sigma^2\boldsymbol{I}_{p_k}$ and are controlled by (33). For the off-diagonal block, $\boldsymbol{L}_2$ is $\mathcal{H}_1$-measurable, so (128) applies with $\boldsymbol{\Psi} = \boldsymbol{L}_2$ and $\boldsymbol{\Theta} = \boldsymbol{I}_{p_1}$. Hence

$$\begin{aligned}
\left\|\boldsymbol{L}_2^{\top}\boldsymbol{L}_1\right\|_{L^{32}} &\overset{\text{(i)}}{\lesssim} \sigma\left\|\|\boldsymbol{L}_2\|_{\mathrm{F}} + \sqrt{p_1}\,\|\boldsymbol{L}_2\|_{\mathrm{op}}\right\|_{L^{64}} \\
&\le \sigma(\sqrt{p_1}+\sqrt{p_2})\,\|\boldsymbol{L}_2\|_{L^{64}} \overset{\text{(ii)}}{\lesssim} \sigma^2\sqrt{np_{\max}}.
\end{aligned}$$

Here (i) is (128), and (ii) uses (33) and $a_2 \le n$. The transpose block satisfies the same estimate, so the diagonal and off-diagonal bounds give

$$\|\boldsymbol{\Delta}_{\mathrm{quad}}\|_{L^{32}} \lesssim \sigma^2\sqrt{np_{\max}}. \tag{129}$$

The matrix $\boldsymbol{P}_{\mathrm{col}(\boldsymbol{S})}\boldsymbol{L}$. On $\mathcal{E}_{\mathrm{emb}}$, we have $\mathrm{col}(\boldsymbol{S}) = \mathcal{M}_1 + \mathcal{M}_2$ and therefore $\mathrm{rank}\{\boldsymbol{P}_{\mathrm{col}(\boldsymbol{S})}\} = p_1 + p_2 - r$. The event and the projection are $\mathcal{H}_k$-measurable, so (128) applies with $\boldsymbol{\Psi} = \boldsymbol{P}_{\mathrm{col}(\boldsymbol{S})}$ and $\boldsymbol{\Theta} = \boldsymbol{I}_{p_k}$. Thus

$$\begin{aligned}
\left\|\mathbf{1}_{\mathcal{E}_{\mathrm{emb}}}\boldsymbol{P}_{\mathrm{col}(\boldsymbol{S})}\boldsymbol{L}\right\|_{L^{32}} &\le \sum_{k=1}^{2}\left\|\mathbf{1}_{\mathcal{E}_{\mathrm{emb}}}\boldsymbol{P}_{\mathrm{col}(\boldsymbol{S})}\boldsymbol{L}_k\right\|_{L^{32}} \\
&\lesssim \sigma\sum_{k=1}^{2}\big(\sqrt{p_1+p_2-r} + \sqrt{p_k}\big) \lesssim \sigma\sqrt{p_{\max}}.
\end{aligned} \tag{130}$$

The block $\boldsymbol{\Delta}_{+0}$. Set $\boldsymbol{P}_0 \coloneqq \boldsymbol{P}_{\ker(\boldsymbol{G}_0)}$. Since $\boldsymbol{S}\boldsymbol{P}_0 = \boldsymbol{0}$, the definition of $\boldsymbol{\Delta}$ gives

$$\boldsymbol{\Delta}_{+0} = \big(\boldsymbol{S}(\boldsymbol{G}_0^\dagger)^{1/2}\big)^\top \boldsymbol{L}\boldsymbol{P}_0 + (\boldsymbol{G}_0^\dagger)^{1/2}\boldsymbol{\Delta}_{\mathrm{quad}}\boldsymbol{P}_0.$$

For the first term, apply (128) to each view with $\boldsymbol{\Psi} = \boldsymbol{S}(\boldsymbol{G}_0^\dagger)^{1/2}$ and $\boldsymbol{\Theta} = (\boldsymbol{P}_0)_k$, where $(\boldsymbol{P}_0)_k$ denotes the kth block rows of $\boldsymbol{P}_0$. Combining this estimate with (39b), (39c), and (129) gives

$$\begin{aligned}
\|\mathbf{1}_{\mathcal{E}_{\mathrm{emb}}}\boldsymbol{\Delta}_{+0}\|_{L^{32}} &\le \sum_{k=1}^{2}\left\|\mathbf{1}_{\mathcal{E}_{\mathrm{emb}}}\big(\boldsymbol{S}(\boldsymbol{G}_0^\dagger)^{1/2}\big)^\top \boldsymbol{L}_k(\boldsymbol{P}_0)_k\right\|_{L^{32}} + \left\|(\boldsymbol{G}_0^\dagger)^{1/2}\right\|_{\mathrm{op}}\|\boldsymbol{\Delta}_{\mathrm{quad}}\|_{L^{32}} \\
&\overset{\text{(i)}}{\lesssim} \sigma\sum_{k=1}^{2}\left\{\left\|\boldsymbol{S}(\boldsymbol{G}_0^\dagger)^{1/2}\right\|_{\mathrm{F}}\|(\boldsymbol{P}_0)_k\|_{\mathrm{op}} + \left\|\boldsymbol{S}(\boldsymbol{G}_0^\dagger)^{1/2}\right\|_{\mathrm{op}}\|(\boldsymbol{P}_0)_k\|_{\mathrm{F}}\right\} + \frac{\sigma^2\sqrt{np_{\max}}}{\gamma_{\min}\sqrt{\theta}} \\
&\overset{\text{(ii)}}{\lesssim} \sigma\sqrt{p_{\max}} + \frac{\sigma^2\sqrt{np_{\max}}}{\gamma_{\min}\sqrt{\theta}}.
\end{aligned} \tag{131}$$

Here (i) uses (39c), (128), and (129). For (ii), (39b) gives $\left\|\boldsymbol{S}(\boldsymbol{G}_0^\dagger)^{1/2}\right\|_{\mathrm{op}} \le 1$ and the required Frobenius bound, while $\|(\boldsymbol{P}_0)_k\|_{\mathrm{op}} \le 1$ and $\|(\boldsymbol{P}_0)_k\|_{\mathrm{F}} \le \sqrt{r}$.

We now apply (32). The small-rate assumption gives $\sigma^2\sqrt{np_{\max}}/(\gamma_{\min}^2\theta) \lesssim 1$ and, since $p_{\max} \le n$, $\{\sigma\sqrt{p_{\max}}/(\gamma_{\min}\sqrt{\theta})\}^2 \le \sigma^2\sqrt{np_{\max}}/(\gamma_{\min}^2\theta) \lesssim 1$. We also use the standing condition $\sigma\sqrt{n}/\gamma_{\min} \lesssim 1$. Cauchy–Schwarz and (129) to (131) yield

$$\begin{aligned}
\mathbb{E}\left[\mathbf{1}_{\mathcal{E}_{\mathrm{good}}}\|\mathsf{Rem}\|_{\mathrm{op}}\right] &\lesssim \frac{\sigma^2\sqrt{np_{\max}}}{\gamma_{\min}^2\theta} + \frac{\sigma\sqrt{n}}{\gamma_{\min}^2\sqrt{\theta}}\left(\sigma\sqrt{p_{\max}} + \frac{\sigma^2\sqrt{np_{\max}}}{\gamma_{\min}\sqrt{\theta}}\right) \\
&\quad + \frac{\sigma\sqrt{p_{\max}}}{\gamma_{\min}^2\theta}\left(\sigma\sqrt{p_{\max}} + \frac{\sigma^2\sqrt{np_{\max}}}{\gamma_{\min}\sqrt{\theta}}\right) \\
&\quad + \frac{1}{\gamma_{\min}^2\theta}\left(\sigma\sqrt{p_{\max}} + \frac{\sigma^2\sqrt{np_{\max}}}{\gamma_{\min}\sqrt{\theta}}\right)^2 \lesssim \frac{\sigma^2\sqrt{np_{\max}}}{\gamma_{\min}^2\theta}.
\end{aligned}$$

For the second term, the two summands are absorbed using respectively $\theta \le 1$ and $\sigma\sqrt{n}/\gamma_{\min} \lesssim 1$. For the third, use $p_{\max} \le \sqrt{np_{\max}}$ and $\sigma\sqrt{p_{\max}}/(\gamma_{\min}\sqrt{\theta}) \lesssim 1$. The final term follows from the same inequalities and the small-rate assumption. Since $p_{\max} = r + (r_1 \vee r_2)$, this proves (34b).

E.3 Proof of (34c)

By the definition of $\mathcal{E}_{\mathrm{good}}$,

$$\mathbb{P}(\mathcal{E}_{\mathrm{good}}^c) \le \mathbb{P}(\mathcal{E}_{\mathrm{emb}}^c) + \mathbb{P}\left\{\mathcal{E}_{\mathrm{emb}},\ \|\boldsymbol{\Delta}_{++}\|_{\mathrm{op}} > \frac{1}{4}\right\} + \mathbb{P}\left\{\mathcal{E}_{\mathrm{emb}},\ \|\boldsymbol{\Delta}_{00}\|_{\mathrm{op}} + 2\|\boldsymbol{\Delta}_{+0}\|_{\mathrm{op}}^2 > \frac{1}{8}c_{\mathrm{emb}}^2\gamma_{\min}^2\theta\right\}.$$

Because $\mathcal{E}_1 \cap \mathcal{E}_2 \subseteq \mathcal{E}_{\mathrm{emb}}$, Lemma 4 and a union bound give $\mathbb{P}(\mathcal{E}_{\mathrm{emb}}^c) \lesssim \sigma^2\sqrt{np_{\max}}/\gamma_{\min}^2 \le \sigma^2\sqrt{np_{\max}}/(\gamma_{\min}^2\theta)$.

It remains to bound $\boldsymbol{\Delta}_{00}$ and $\boldsymbol{\Delta}_{++}$; the bound for $\boldsymbol{\Delta}_{+0}$ is already given by (131). Since $\boldsymbol{S}\boldsymbol{P}_0 = \boldsymbol{0}$, the two terms in $\boldsymbol{\Delta}$ containing one factor $\boldsymbol{L}$ vanish after multiplication by $\boldsymbol{P}_0$ on both sides. Therefore

$$\boldsymbol{\Delta}_{00} = \boldsymbol{P}_0\boldsymbol{\Delta}_{\mathrm{quad}}\boldsymbol{P}_0, \qquad \|\mathbf{1}_{\mathcal{E}_{\mathrm{emb}}}\boldsymbol{\Delta}_{00}\|_{L^{32}} \lesssim \sigma^2\sqrt{np_{\max}}, \tag{132}$$

where the bound follows from (129).

For $\boldsymbol{\Delta}_{++}$, expanding (21) gives

$$\begin{aligned}\boldsymbol{\Delta}_{++} = \big(\boldsymbol{S}(\boldsymbol{G}_0^{\dagger})^{1/2}\big)^{\top}\boldsymbol{L}(\boldsymbol{G}_0^{\dagger})^{1/2} + (\boldsymbol{G}_0^{\dagger})^{1/2}\boldsymbol{L}^{\top}\boldsymbol{S}(\boldsymbol{G}_0^{\dagger})^{1/2} \\ + (\boldsymbol{G}_0^{\dagger})^{1/2}\boldsymbol{\Delta}_{\mathrm{quad}}(\boldsymbol{G}_0^{\dagger})^{1/2}.\end{aligned}$$

The first two terms are transposes. Apply (128) to each view with $\boldsymbol{\Psi} = \boldsymbol{S}(\boldsymbol{G}_0^{\dagger})^{1/2}$ and $\boldsymbol{\Theta} = ((\boldsymbol{G}_0^{\dagger})^{1/2})_k$, and use (39b), (39c), and (129). This gives

$$\begin{aligned}\|\mathbf{1}_{\mathcal{E}_{\mathrm{emb}}}\boldsymbol{\Delta}_{++}\|_{L^{32}} &\lesssim \sigma\left\{\left\|\boldsymbol{S}(\boldsymbol{G}_0^{\dagger})^{1/2}\right\|_{\mathrm{F}}\left\|(\boldsymbol{G}_0^{\dagger})^{1/2}\right\|_{\mathrm{op}} + \left\|\boldsymbol{S}(\boldsymbol{G}_0^{\dagger})^{1/2}\right\|_{\mathrm{op}}\left\|(\boldsymbol{G}_0^{\dagger})^{1/2}\right\|_{\mathrm{F}}\right\} + \left\|(\boldsymbol{G}_0^{\dagger})^{1/2}\right\|_{\mathrm{op}}^2\|\boldsymbol{\Delta}_{\mathrm{quad}}\|_{L^{32}} \\ &\lesssim \frac{\sigma\sqrt{r_1+r_2}}{\gamma_{\min}} + \frac{\sigma\sqrt{p_{\min}}}{\gamma_{\min}\sqrt{\theta}} + \frac{\sigma\sqrt{p_{\max}}}{\gamma_{\min}\sqrt{\theta}} + \frac{\sigma^2\sqrt{np_{\max}}}{\gamma_{\min}^2\theta}.\end{aligned} \tag{133}$$

The last event in the initial union bound is contained in the union of $\{\|\boldsymbol{\Delta}_{00}\|_{\mathrm{op}} > c_1\gamma_{\min}^2\theta\}$ and $\{\|\boldsymbol{\Delta}_{+0}\|_{\mathrm{op}} > c_2\gamma_{\min}\sqrt{\theta}\}$ for sufficiently small numerical constants $c_1, c_2 > 0$. Markov's inequality with exponent 32, followed by (131) to (133), therefore gives

$$\begin{aligned}&\mathbb{P}\left\{\mathcal{E}_{\mathrm{emb}},\ \|\boldsymbol{\Delta}_{++}\|_{\mathrm{op}} > \frac{1}{4}\right\} + \mathbb{P}\left\{\mathcal{E}_{\mathrm{emb}},\ \|\boldsymbol{\Delta}_{00}\|_{\mathrm{op}} + 2\|\boldsymbol{\Delta}_{+0}\|_{\mathrm{op}}^2 > \frac{1}{8}c_{\mathrm{emb}}^2\gamma_{\min}^2\theta\right\} \\ &\lesssim \|\mathbf{1}_{\mathcal{E}_{\mathrm{emb}}}\boldsymbol{\Delta}_{++}\|_{L^{32}}^{32} + \left(\frac{\|\mathbf{1}_{\mathcal{E}_{\mathrm{emb}}}\boldsymbol{\Delta}_{00}\|_{L^{32}}}{\gamma_{\min}^2\theta}\right)^{32} + \left(\frac{\|\mathbf{1}_{\mathcal{E}_{\mathrm{emb}}}\boldsymbol{\Delta}_{+0}\|_{L^{32}}}{\gamma_{\min}\sqrt{\theta}}\right)^{32} \\ &\lesssim \frac{\sigma\sqrt{r_1+r_2}}{\gamma_{\min}} + \frac{\sigma\sqrt{p_{\min}}}{\gamma_{\min}\sqrt{\theta}} + \frac{\sigma^2\sqrt{np_{\max}}}{\gamma_{\min}^2\theta}.\end{aligned} \tag{134}$$

For the last inequality, the small-rate assumption bounds the 32nd power of each displayed rate by the rate itself. The only additional term is $\sigma\sqrt{p_{\max}}/(\gamma_{\min}\sqrt{\theta})$, whose 32nd power is bounded by its square and hence by $\sigma^2\sqrt{np_{\max}}/(\gamma_{\min}^2\theta)$ because $p_{\max} \le n$.

Combining this estimate with the bound for $\mathbb{P}(\mathcal{E}_{\mathrm{emb}}^c)$ and using $\sqrt{p_{\min}} \asymp \sqrt{r + (r_1 \wedge r_2)}$ proves (34c). Together with (34a) and (34b), this completes the proof of Lemma 5.

E.4 Proof of Lemma 21

Proof. The proof has two parts. We first use the rotational invariance of the Gaussian noise to randomize the orientation of $\boldsymbol{L}_k$ while preserving every quantity in $\mathcal{H}_k$. Conditional on the original data, the resulting product is then controlled by a fixed-matrix Haar estimate.

Fix an orthonormal matrix $\boldsymbol{N}_k \in \mathrm{St}(n, a_k)$ whose columns span $\mathcal{M}_k^{\perp}$, and let $\boldsymbol{O}$ be an independent Haar matrix in $\mathbb{O}(a_k)$. The distributional identity used below is

$$\big(\boldsymbol{S}_1, \boldsymbol{S}_2, \boldsymbol{L}_{3-k}, \boldsymbol{L}_k^{\top}\boldsymbol{L}_k, \boldsymbol{L}_k\big) \overset{\mathrm{d}}{=} \big(\boldsymbol{S}_1, \boldsymbol{S}_2, \boldsymbol{L}_{3-k}, \boldsymbol{L}_k^{\top}\boldsymbol{L}_k, \boldsymbol{N}_k\boldsymbol{O}\boldsymbol{N}_k^{\top}\boldsymbol{L}_k\big). \tag{135}$$

We first derive the claimed moment bound from this identity and verify the identity at the end of the proof.

Choose measurable representatives of $\mathbf{1}_{\mathcal{E}}$, $\boldsymbol{\Psi}$, and $\boldsymbol{\Theta}$ as functions of the tuple defining $\mathcal{H}_k$. These representatives are unchanged by the replacement of $\boldsymbol{L}_k$ in (135). Define

$$\boldsymbol{H} := \boldsymbol{N}_k^{\top}\boldsymbol{\Psi}, \qquad \boldsymbol{M} := \boldsymbol{N}_k^{\top}\boldsymbol{L}_k\boldsymbol{\Theta}.$$

Since $\boldsymbol{L}_k = \boldsymbol{N}_k\boldsymbol{N}_k^{\top}\boldsymbol{L}_k$, the product obtained after the replacement in (135) is $\boldsymbol{H}^{\top}\boldsymbol{O}\boldsymbol{M}$.

We next establish the fixed-matrix Haar estimate needed after conditioning. Let $\boldsymbol{H}_0 \in \mathbb{R}^{a\times u}$ and $\boldsymbol{M}_0 \in \mathbb{R}^{a\times v}$ be deterministic, and let $\boldsymbol{\Gamma} \in \mathbb{R}^{a\times a}$ have independent standard Gaussian entries. Chevet's inequality (Vershynin, 2018, Section 8.7), applied with $T = \boldsymbol{M}_0\mathbb{S}^{v-1}$ and $S = \boldsymbol{H}_0\mathbb{S}^{u-1}$, gives

$$\mathbb{E}\|\boldsymbol{H}_0^\top \boldsymbol{\Gamma}\boldsymbol{M}_0\|_{\mathrm{op}} \le \|\boldsymbol{H}_0\|_{\mathrm{F}}\|\boldsymbol{M}_0\|_{\mathrm{op}} + \|\boldsymbol{H}_0\|_{\mathrm{op}}\|\boldsymbol{M}_0\|_{\mathrm{F}}.$$

The map $\boldsymbol{\Gamma} \mapsto \|\boldsymbol{H}_0^\top \boldsymbol{\Gamma}\boldsymbol{M}_0\|_{\mathrm{op}}$ is $\|\boldsymbol{H}_0\|_{\mathrm{op}}\|\boldsymbol{M}_0\|_{\mathrm{op}}$-Lipschitz with respect to the Frobenius norm, because

$$\begin{aligned}\Big|\|\boldsymbol{H}_0^\top \boldsymbol{\Gamma}_1\boldsymbol{M}_0\|_{\mathrm{op}} - \|\boldsymbol{H}_0^\top \boldsymbol{\Gamma}_0\boldsymbol{M}_0\|_{\mathrm{op}}\Big| &\le \|\boldsymbol{H}_0^\top(\boldsymbol{\Gamma}_1-\boldsymbol{\Gamma}_0)\boldsymbol{M}_0\|_{\mathrm{op}} \\ &\le \|\boldsymbol{H}_0\|_{\mathrm{op}}\|\boldsymbol{\Gamma}_1-\boldsymbol{\Gamma}_0\|_{\mathrm{F}}\|\boldsymbol{M}_0\|_{\mathrm{op}}.\end{aligned}$$

Combining Gaussian concentration with the expectation bound therefore yields

$$\begin{aligned}\Big(\mathbb{E}\|\boldsymbol{H}_0^\top \boldsymbol{\Gamma}\boldsymbol{M}_0\|_{\mathrm{op}}^q\Big)^{1/q} &\le C\left\{\|\boldsymbol{H}_0\|_{\mathrm{F}}\|\boldsymbol{M}_0\|_{\mathrm{op}} + \|\boldsymbol{H}_0\|_{\mathrm{op}}\|\boldsymbol{M}_0\|_{\mathrm{F}} + \sqrt{q}\,\|\boldsymbol{H}_0\|_{\mathrm{op}}\|\boldsymbol{M}_0\|_{\mathrm{op}}\right\} \\ &\le C\sqrt{q}\left\{\|\boldsymbol{H}_0\|_{\mathrm{F}}\|\boldsymbol{M}_0\|_{\mathrm{op}} + \|\boldsymbol{H}_0\|_{\mathrm{op}}\|\boldsymbol{M}_0\|_{\mathrm{F}}\right\}.\end{aligned}$$

The second inequality uses $q \ge 2$ and bounds each operator norm by the corresponding Frobenius norm.

Apply Tropp's comparison principle (Tropp, 2012, Theorem 1) to the seminorm $\boldsymbol{X} \mapsto \|\boldsymbol{H}_0^\top \boldsymbol{X}\boldsymbol{M}_0\|_{\mathrm{op}}$ and the function $t \mapsto t^q$. Comparing a Haar matrix with $\boldsymbol{\Gamma}/\sqrt{a}$ gives

$$\Big(\mathbb{E}_{\boldsymbol{O}}\|\boldsymbol{H}_0^\top \boldsymbol{O}\boldsymbol{M}_0\|_{\mathrm{op}}^q\Big)^{1/q} \le C\sqrt{\frac{q}{a}}\left\{\|\boldsymbol{H}_0\|_{\mathrm{F}}\|\boldsymbol{M}_0\|_{\mathrm{op}} + \|\boldsymbol{H}_0\|_{\mathrm{op}}\|\boldsymbol{M}_0\|_{\mathrm{F}}\right\}. \tag{136}$$

We apply (136) conditionally with $a = a_k$, $\boldsymbol{H}_0 = \boldsymbol{H}$, and $\boldsymbol{M}_0 = \boldsymbol{M}$. The identity (135) and Tonelli's theorem give

$$\mathbb{E}\left[\mathbf{1}_{\mathcal{E}}\|\boldsymbol{\Psi}^\top \boldsymbol{L}_k\boldsymbol{\Theta}\|_{\mathrm{op}}^q\right] = \mathbb{E}\mathbb{E}_{\boldsymbol{O}}\left[\mathbf{1}_{\mathcal{E}}\|\boldsymbol{H}^\top \boldsymbol{O}\boldsymbol{M}\|_{\mathrm{op}}^q\right].$$

Because $\boldsymbol{N}_k$ has orthonormal columns,

$$\|\boldsymbol{H}\|_{\mathrm{F}}\|\boldsymbol{M}\|_{\mathrm{op}} + \|\boldsymbol{H}\|_{\mathrm{op}}\|\boldsymbol{M}\|_{\mathrm{F}} \le \|\boldsymbol{L}_k\|_{\mathrm{op}}\left(\|\boldsymbol{\Psi}\|_{\mathrm{F}}\|\boldsymbol{\Theta}\|_{\mathrm{op}} + \|\boldsymbol{\Psi}\|_{\mathrm{op}}\|\boldsymbol{\Theta}\|_{\mathrm{F}}\right).$$

Consequently, (136) followed by Hölder's inequality gives

$$\begin{aligned}\left\|\mathbf{1}_{\mathcal{E}}\boldsymbol{\Psi}^\top \boldsymbol{L}_k\boldsymbol{\Theta}\right\|_{L^q} &\lesssim \sqrt{\frac{q}{a_k}}\,\left\|\mathbf{1}_{\mathcal{E}}\|\boldsymbol{L}_k\|_{\mathrm{op}}\left(\|\boldsymbol{\Psi}\|_{\mathrm{F}}\|\boldsymbol{\Theta}\|_{\mathrm{op}} + \|\boldsymbol{\Psi}\|_{\mathrm{op}}\|\boldsymbol{\Theta}\|_{\mathrm{F}}\right)\right\|_{L^q} \\ &\le \sqrt{\frac{q}{a_k}}\,\|\boldsymbol{L}_k\|_{L^{2q}}\left\|\mathbf{1}_{\mathcal{E}}\left(\|\boldsymbol{\Psi}\|_{\mathrm{F}}\|\boldsymbol{\Theta}\|_{\mathrm{op}} + \|\boldsymbol{\Psi}\|_{\mathrm{op}}\|\boldsymbol{\Theta}\|_{\mathrm{F}}\right)\right\|_{L^{2q}}.\end{aligned}$$

This is the conclusion of Lemma 21, subject only to the distributional identity.

It remains to prove (135). Fix $\boldsymbol{o} \in \mathbb{O}(a_k)$ and define

$$\boldsymbol{R} \coloneqq \boldsymbol{P}_{\mathcal{M}_k} + \boldsymbol{N}_k\boldsymbol{o}\boldsymbol{N}_k^\top.$$

The matrix $\boldsymbol{R}$ is orthogonal, acts as the identity on $\mathcal{M}_k$, and acts as $\boldsymbol{o}$ on $\mathcal{M}_k^\perp$. Since $\mathrm{col}(\boldsymbol{X}_k) \subseteq \mathcal{M}_k$, we have $\boldsymbol{R}\boldsymbol{X}_k = \boldsymbol{X}_k$. Left rotational invariance of the Gaussian noise and independence of the two views therefore imply

$$(\boldsymbol{R}\boldsymbol{Y}_k, \boldsymbol{Y}_{3-k}) \overset{\mathrm{d}}{=} (\boldsymbol{Y}_k, \boldsymbol{Y}_{3-k}).$$

The bias-corrected empirical factor is left orthogonally equivariant:

$$\boldsymbol{C}_k(\boldsymbol{R}\boldsymbol{Y}_k) = \boldsymbol{R}\boldsymbol{C}_k(\boldsymbol{Y}_k) \qquad \text{almost surely.}$$

Indeed, choose the right singular vectors measurably as functions of $\boldsymbol{Y}_k^\top \boldsymbol{Y}_k$. Left multiplication by $\boldsymbol{R}$ preserves this Gram matrix and all singular values, while multiplying the corresponding left singular vectors by $\boldsymbol{R}$. The weights depend only on the singular values, so the displayed equivariance follows.

Since $\boldsymbol{R}$ acts as the identity on $\mathcal{M}_k$, this equivariance gives

$$\boldsymbol{S}_k(\boldsymbol{R}\boldsymbol{Y}_k) = \boldsymbol{S}_k(\boldsymbol{Y}_k), \qquad \boldsymbol{L}_k(\boldsymbol{R}\boldsymbol{Y}_k) = \boldsymbol{N}_k\boldsymbol{o}\boldsymbol{N}_k^\top \boldsymbol{L}_k(\boldsymbol{Y}_k).$$

The other view is unchanged, and the transformed leakage has the same Gram matrix:

$$\left(\boldsymbol{N}_k\boldsymbol{o}\boldsymbol{N}_k^\top \boldsymbol{L}_k\right)^\top \left(\boldsymbol{N}_k\boldsymbol{o}\boldsymbol{N}_k^\top \boldsymbol{L}_k\right) = \boldsymbol{L}_k^\top \boldsymbol{L}_k.$$

Thus (135) holds with $\boldsymbol{O}$ replaced by any fixed $\boldsymbol{o}$. Averaging over an independent Haar matrix proves the stated identity and completes the proof. □

F Probabilistic tools for the lower bound

Gaussian conditioning. The following lemma is a simple consequence of the tail bound for the smallest singular value of a Gaussian random matrix.

Lemma 22. *There exist numerical constants $c_B > 0$ and $d_0 \geq 1$ such that, whenever $d \geq d_0$ and $1 \leq p \leq d/3$, a matrix $\boldsymbol{B} \in \mathbb{R}^{d\times p}$ with independent $\mathcal{N}(0, d^{-1})$ entries satisfies*

$$\mathbb{P}\left\{\sigma_p^2(\boldsymbol{B}) < c_B\right\} \leq \frac{1}{16}.$$

Gaussian projector information. For an orthogonal projector $\boldsymbol{P} \in \mathbb{R}^{N\times N}$, $\sigma > 0$, and $\beta \geq 0$, define

$$\mathsf{G}_{\sigma,\beta}(\boldsymbol{P}) \coloneqq \mathcal{N}\left(\boldsymbol{0}, \sigma^2(\boldsymbol{I}_N + \beta\boldsymbol{P})\right).$$

The only covariance identity used later is the following standard specialization: if $\boldsymbol{P}$ and $\boldsymbol{Q}$ are equal-rank orthogonal projectors, then

$$\mathrm{KL}\left(\mathsf{G}_{\sigma,\beta}(\boldsymbol{P}) \,\|\, \mathsf{G}_{\sigma,\beta}(\boldsymbol{Q})\right) = \frac{\beta^2}{4(1+\beta)} \|\boldsymbol{P} - \boldsymbol{Q}\|_{\mathrm{F}}^2. \tag{137}$$

For independent columns, we apply additivity of relative entropy directly at the point of use.

Fano. Below is a simple consequence of classical Fano's inequality and the definition of the TV distance.

Lemma 23 (Fano with proxy laws)**.** *Let $\Pi_1, \ldots, \Pi_N$ be priors supported on a parameter class $\mathfrak{P}$, with induced observation laws $P_1, \ldots, P_N$. Suppose that the target functional is constant under each prior, with respective values $\boldsymbol{\Theta}_1, \ldots, \boldsymbol{\Theta}_N$, and that $d(\boldsymbol{\Theta}_i, \boldsymbol{\Theta}_j) \geq 2\delta$ for all $i \neq j$. Let $Q_1, \ldots, Q_N$ be proxy laws such that* $\max_j \mathrm{TV}(P_j, Q_j) \leq \varepsilon$. *Then, for every reference index $i_0 \in [N]$,*

$$\inf_{\widehat{\boldsymbol{\Theta}}} \sup_{\vartheta\in\mathfrak{P}} \mathbb{E}_\vartheta d\big(\widehat{\boldsymbol{\Theta}}, \Theta(\vartheta)\big) \geq \delta\left\{1 - \varepsilon - \frac{N^{-1}\sum_{j=1}^N \mathrm{KL}(Q_j \| Q_{i_0}) + \log 2}{\log N}\right\}.$$

Assouad. We also need the Assouad lemma.

Lemma 24. *Let $\Omega_m := \{-1,1\}^m$. For each $\boldsymbol{\omega} \in \Omega_m$, let $\Pi_{\boldsymbol{\omega}}$ be a prior supported on $\mathfrak{P}$, with induced observation law $P_{\boldsymbol{\omega}}$. Suppose the target functional is constant under each prior:*

$$\Theta(\vartheta) = \boldsymbol{\Theta}_{\boldsymbol{\omega}} \qquad \textit{for } \Pi_{\boldsymbol{\omega}}\textit{-almost every } \vartheta.$$

Assume that, for some $a > 0$,

$$d(\boldsymbol{\Theta}_{\boldsymbol{\omega}}, \boldsymbol{\Theta}_{\boldsymbol{\omega}'}) \geq a\sqrt{\mathrm{Ham}(\boldsymbol{\omega}, \boldsymbol{\omega}')} \qquad \textit{for all } \boldsymbol{\omega}, \boldsymbol{\omega}' \in \Omega_m,$$

where $\mathrm{Ham}(\boldsymbol{\omega}, \boldsymbol{\omega}')$ denotes the Hamming distance. For $j \in [m]$, let $\boldsymbol{\omega}^{(j)}$ denote the vertex obtained by flipping the jth coordinate of $\boldsymbol{\omega}$. If

$$\max_{\substack{\boldsymbol{\omega} \in \Omega_m \\ j \in [m]}} \mathrm{TV}\big(P_{\boldsymbol{\omega}}, P_{\boldsymbol{\omega}^{(j)}}\big) \leq \eta < 1,$$

then

$$\inf_{\widehat{\boldsymbol{\Theta}}} \sup_{\vartheta \in \mathfrak{P}} \mathbb{E}_\vartheta d\big(\widehat{\boldsymbol{\Theta}}, \Theta(\vartheta)\big) \geq \frac{a\sqrt{m}}{4}(1-\eta).$$

Hellinger path.

Lemma 25 (Hellinger path bound)**.** *Let $\{P_t : t \in [0,T]\}$ be probability measures dominated by μ, with strictly positive densities p_t. Suppose that $t \mapsto \sqrt{p_t}$ is absolutely continuous as an $L^2(\mu)$-valued map and that, for almost every t,*

$$\frac{d}{dt}\sqrt{p_t} = \frac{1}{2} S_t \sqrt{p_t} \quad \textit{in } L^2(\mu), \qquad S_t(x) := \partial_t \log p_t(x).$$

Define $J_t := \mathbb{E}_{P_t}[S_t(X)^2]$. Then, for the unnormalized Hellinger distance $H(P,Q) := \|\sqrt{p} - \sqrt{q}\|_{L^2(\mu)}$,

$$\mathrm{TV}(P_0, P_T) \leq H(P_0, P_T) \leq \frac{1}{2}\int_0^T \sqrt{J_t}\, dt. \tag{138}$$

Consequently,

$$\mathrm{TV}(P_0, P_T) \leq \frac{T}{2} \sup_{0 \leq t \leq T} \sqrt{J_t}. \tag{139}$$

Proof. The standard inequality $\mathrm{TV} \leq H$ and the length bound for the absolutely continuous path $t \mapsto \sqrt{p_t}$ give

$$\mathrm{TV}(P_0, P_T) \leq \|\sqrt{p_T} - \sqrt{p_0}\|_{L^2(\mu)} \leq \int_0^T \left\| \frac{d}{dt}\sqrt{p_t} \right\|_{L^2(\mu)} dt = \frac{1}{2}\int_0^T \sqrt{J_t}\, dt.$$

The final inequality in (138) follows by taking the supremum of $\sqrt{J_t}$ over $[0,T]$. ☐

Bound posterior mean by mutual information.

Lemma 26. *Let $\boldsymbol{z}$ be uniform on $\mathbb{S}^{M-1}$ and let $\boldsymbol{Y}$ be any observation jointly distributed with $\boldsymbol{z}$. Then*

$$\mathbb{E}\big\|\mathbb{E}[\boldsymbol{z} \mid \boldsymbol{Y}]\big\|^2 \leq \frac{C}{M}\,\mathrm{I}(\boldsymbol{z};\boldsymbol{Y}).$$

In the Gaussian mean channel $\boldsymbol{Y} = \lambda\boldsymbol{z} + \sigma\boldsymbol{g}$, where $\boldsymbol{g} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_M)$ is independent of $\boldsymbol{z}$, the sharper bound $\mathbb{E}\big\|\mathbb{E}[\boldsymbol{z} \mid \boldsymbol{Y}]\big\|^2 \leq 9\frac{\lambda^2}{M\sigma^2}$ holds.

Proof. Let ν_M denote the uniform probability measure on $\mathbb{S}^{M-1}$. Its one-dimensional marginals are strictly $1/M$-sub-Gaussian (Marchal and Arbel, 2017, Corollary 1). Thus, by rotational invariance,

$$\log \int \exp(\langle \boldsymbol{t}, \boldsymbol{u}\rangle)\,\nu_M(d\boldsymbol{u}) \leq \frac{\|\boldsymbol{t}\|^2}{2M}, \qquad \boldsymbol{t} \in \mathbb{R}^M.$$

Indeed, for $M \geq 2$ a rescaled spherical coordinate has the symmetric Beta$((M-1)/2, (M-1)/2)$ distribution; for $M = 1$ this is the Rademacher bound.

For an observation value y, let Π_y be the conditional law of $\boldsymbol{z}$ given $\boldsymbol{Y} = y$, set $\boldsymbol{m}_y := \int \boldsymbol{u}\,\Pi_y(d\boldsymbol{u})$, and write $D_y := \mathrm{KL}(\Pi_y \,\|\, \nu_M)$. The Donsker–Varadhan variational inequality states that

$$\int f\,d\Pi_y \leq D_y + \log \int e^f\,d\nu_M.$$

Applying it with $f(\boldsymbol{u}) = M\langle \boldsymbol{m}_y, \boldsymbol{u}\rangle$ gives $M\|\boldsymbol{m}_y\|^2 \leq D_y + M\|\boldsymbol{m}_y\|^2/2$, and therefore $\|\boldsymbol{m}_y\|^2 \leq 2D_y/M$; the inequality is automatic if $D_y = \infty$. Averaging over $\boldsymbol{Y}$ and using $\mathbb{E}D_{\boldsymbol{Y}} = \mathrm{I}(\boldsymbol{z};\boldsymbol{Y})$ proves the first assertion with $C = 2$.

For the Gaussian channel, let $Q_{\boldsymbol{u}} = \mathcal{N}(\lambda\boldsymbol{u}, \sigma^2\boldsymbol{I}_M)$, let Q be the marginal law of $\boldsymbol{Y}$, and set $Q_0 = \mathcal{N}(\mathbf{0}, \sigma^2\boldsymbol{I}_M)$. The information-radius identity and the Gaussian shift formula give

$$\mathrm{I}(\boldsymbol{z};\boldsymbol{Y}) + \mathrm{KL}(Q \,\|\, Q_0) = \int \mathrm{KL}(Q_{\boldsymbol{u}} \,\|\, Q_0)\,\nu_M(d\boldsymbol{u}) = \frac{\lambda^2}{2\sigma^2}.$$

Hence $\mathrm{I}(\boldsymbol{z};\boldsymbol{Y}) \leq \lambda^2/(2\sigma^2)$, so the first part yields the stronger bound

$$\mathbb{E}\big\|\mathbb{E}[\boldsymbol{z} \mid \boldsymbol{Y}]\big\|^2 \leq \frac{\lambda^2}{M\sigma^2}.$$

The stated bound follows. $\square$

G Proof of Theorem 2

To simplify notation, we use $\mathcal{R}_{\text{minimax}}$ to denote the minimax risk. It suffices to prove the following three separate lower bounds:

$$\mathcal{R}_{\text{minimax}} \geq c \min\left\{1, \frac{\sigma\sqrt{n}}{\gamma_{\max}}\right\}, \tag{140a}$$

$$\mathcal{R}_{\text{minimax}} \geq c \min\left\{1, \frac{\sigma(\sqrt{r+(r_1\wedge r_2)})}{\gamma_{\min}\sqrt{\theta}}\right\}, \tag{140b}$$

$$\mathcal{R}_{\text{minimax}} \geq c \min\left\{1, \frac{\sigma^2\sqrt{n}(\sqrt{r+(r_1\wedge r_2)})}{\gamma_{\min}^2\theta}\right\}. \tag{140c}$$

The first term is independent of θ and captures the difficulty of estimating the shared subspace even when the view-specific subspaces are fixed and known. The remaining two terms capture the additional difficulty of distinguishing the shared directions from the view-specific directions when the two view-specific subspaces are nearly aligned. We prove the three bounds in order.

G.1 Proof of (140a)

We restrict the parameter class to a family in which both view-specific subspaces and $r-1$ shared directions are fixed. Only the remaining shared direction varies. We construct exponentially many such directions whose projectors are separated by order δ, and then choose Gaussian loadings for which the corresponding observation laws have relative entropy of order $\delta^2\gamma_{\max}^2/\sigma^2$. Fano's lemma will then give the claimed rate.

A packing of shared subspaces. Choose a fixed matrix with orthonormal columns:

$$\big[\,\underbrace{\boldsymbol{U}_0}_{n\times(r-1)}\ \underbrace{\boldsymbol{V}_1}_{n\times r_1}\ \underbrace{\boldsymbol{V}_2}_{n\times r_2}\ \underbrace{\boldsymbol{u}_0}_{n\times 1}\,\big] \in \mathrm{St}(n, r+r_1+r_2).$$

We use $\boldsymbol{U}_0$ to form the fixed part of the shared subspace, $\boldsymbol{V}_1$ and $\boldsymbol{V}_2$ to form the fixed view-specific subspaces, and $\boldsymbol{u}_0$ as the baseline for the remaining shared direction.

The orthogonal complement of the displayed columns has dimension $n-r-r_1-r_2 \asymp n$. A standard packing argument therefore gives unit vectors $\boldsymbol{z}_1,\ldots,\boldsymbol{z}_N$ in this orthogonal complement such that

$$\log N \geq cn, \qquad \|\boldsymbol{z}_i - \boldsymbol{z}_j\|_2 \geq c \quad \text{for all } i \neq j. \tag{141}$$

For $0<\delta\leq 1/2$, set $\boldsymbol{u}_j := \sqrt{1-\delta^2}\,\boldsymbol{u}_0 + \delta\boldsymbol{z}_j$, $\boldsymbol{U}_j := [\boldsymbol{U}_0, \boldsymbol{u}_j]$, and $\boldsymbol{P}_j := \boldsymbol{U}_j\boldsymbol{U}_j^\top$. The choice of the $\boldsymbol{z}_j$ ensures that every $\boldsymbol{U}_j$ has orthonormal columns. Moreover, $\boldsymbol{P}_i - \boldsymbol{P}_j = \boldsymbol{u}_i\boldsymbol{u}_i^\top - \boldsymbol{u}_j\boldsymbol{u}_j^\top$ and $1-\boldsymbol{z}_i^\top\boldsymbol{z}_j = \frac{1}{2}\|\boldsymbol{z}_i-\boldsymbol{z}_j\|_2^2 \asymp 1$. The formula for the distance between two rank-one projectors therefore gives

$$\begin{aligned}\|\boldsymbol{P}_i - \boldsymbol{P}_j\|_{\text{op}}^2 &= 1-(\boldsymbol{u}_i^\top\boldsymbol{u}_j)^2 \\ &= \delta^2(1-\boldsymbol{z}_i^\top\boldsymbol{z}_j)\left\{2-\delta^2(1-\boldsymbol{z}_i^\top\boldsymbol{z}_j)\right\} \asymp \delta^2, \qquad \|\boldsymbol{P}_i-\boldsymbol{P}_j\|_{\text{F}}^2 = 2\,\|\boldsymbol{P}_i-\boldsymbol{P}_j\|_{\text{op}}^2.\end{aligned} \tag{142}$$

Thus the targets are separated by order δ in operator norm, while their squared Frobenius distance is of order δ^2, as required for the information calculation below.

Admissible priors and unconditioned comparison laws. The loadings are randomized so that the unconditioned observation laws are centered Gaussian, while conditioning on their smallest singular values ensures that every realization belongs to the parameter class. Independently for $k = 1, 2$, let $\boldsymbol{G}_k \in \mathbb{R}^{d_k \times p_k}$ have independent $\mathcal{N}(0, d_k^{-1})$ entries, and define

$$\mathcal{E}_k^{\mathrm{load}} := \left\{ \boldsymbol{G}_k^\top \boldsymbol{G}_k \succeq c_B \boldsymbol{I}_{p_k} \right\}, \qquad \boldsymbol{B}_k := \frac{s_k}{\sqrt{c_B}} \boldsymbol{G}_k, \qquad \boldsymbol{X}_k^{(j)} := [\boldsymbol{U}_j, \boldsymbol{V}_k] \boldsymbol{B}_k^\top.$$

For the prior indexed by j, condition jointly on $\mathcal{E}_1^{\mathrm{load}} \cap \mathcal{E}_2^{\mathrm{load}}$. On this event, $\sigma_{p_k}(\boldsymbol{B}_k) \geq s_k$. In addition, $[\boldsymbol{U}_j, \boldsymbol{V}_k] \in \mathrm{St}(n, p_k)$ and $\boldsymbol{V}_1^\top \boldsymbol{V}_2 = \boldsymbol{0}$, so the resulting prior is supported on $\mathfrak{P}$ and its target is the fixed projector $\boldsymbol{P}_j$. Let P_j denote the induced observation law.

Let Q_j denote the observation law obtained from the same construction without conditioning on $\mathcal{E}_1^{\mathrm{load}} \cap \mathcal{E}_2^{\mathrm{load}}$. The Gaussian smallest-singular-value bound and contraction of total variation under the observation channel give

$$\max_{j \in [N]} \mathrm{TV}(\mathsf{P}_j, \mathsf{Q}_j) \leq \mathbb{P}\left(\left(\mathcal{E}_1^{\mathrm{load}} \right)^c \cup \left(\mathcal{E}_2^{\mathrm{load}} \right)^c \right) \leq \frac{1}{8}. \tag{143}$$

It remains to control the relative entropy between the unconditioned laws.

Applying Fano's bound. Under Q_j, the columns in view k are independent centered Gaussian vectors with covariance $\boldsymbol{\Sigma}_{k,j} = \sigma^2 \{ \boldsymbol{I}_n + \beta_k (\boldsymbol{U}_j \boldsymbol{U}_j^\top + \boldsymbol{V}_k \boldsymbol{V}_k^\top) \}$, where $\beta_k := s_k^2 / (c_B d_k \sigma^2)$. Since all fixed directions cancel when comparing j with 1, (137) and (142) give

$$\begin{aligned} \mathrm{KL}(\mathsf{Q}_j \,\|\, \mathsf{Q}_1) &= \sum_{k=1}^{2} \frac{d_k \beta_k^2}{4(1+\beta_k)} \left\| \boldsymbol{u}_j \boldsymbol{u}_j^\top - \boldsymbol{u}_1 \boldsymbol{u}_1^\top \right\|_{\mathrm{F}}^2 \\ &\lesssim \delta^2 \sum_{k=1}^{2} \frac{d_k \beta_k^2}{1+\beta_k} \lesssim \delta^2 \frac{\gamma_{\max}^2}{\sigma^2}. \end{aligned} \tag{144}$$

We now choose

$$\delta := c_0 \left(1 \wedge \frac{\sigma \sqrt{n}}{\gamma_{\max}} \right),$$

where $c_0 > 0$ is sufficiently small. Then (144) is at most a sufficiently small multiple of n, and hence of $\log N$ by (141). Apply Lemma 23 with the operator-norm distance, reference law Q_1, target separation $c\delta$ from (142), and comparison error at most $1/8$ from (143). Fano's lemma gives a minimax risk lower bound in (140a).

G.2 A common construction for the two θ-dependent terms

We now construct a single family of target subspaces for the two lower bounds that depend on θ. By exchanging the labels of the views, we may assume that $\gamma_2 = \gamma_{\min}$. Put $q := r_1 \wedge r_2$. The construction varies the decomposition of a fixed $(r+q)$-dimensional subspace into r shared directions and q view-specific directions. View 1 will have the same distribution under every hypothesis, so only View 2 can distinguish the resulting target projectors. The two subsequent proofs differ only in whether certain view-specific directions in View 2 are fixed or randomized.

The target subspaces. Set $m := n-r-r_1-r_2+q$. The rank assumptions imply $r+r_1+r_2 \leq 2n/3$, and hence $m \geq q$. We may therefore choose

$$[\underbrace{\boldsymbol{F}}_{n\times(r+q)} \quad \underbrace{\boldsymbol{D}_1}_{n\times(r_1-q)} \quad \underbrace{\boldsymbol{D}_2}_{n\times(r_2-q)} \quad \underbrace{\boldsymbol{H}}_{n\times m}] \in \mathrm{St}(n,n),$$

where a block with zero columns is omitted. The shared subspace will vary inside $\mathrm{col}(\boldsymbol{F})$. The matrices $\boldsymbol{D}_1$ and $\boldsymbol{D}_2$ supply the remaining r_1-q and r_2-q view-specific directions, while $\mathrm{col}(\boldsymbol{H})$ contains the additional View-2 directions used to satisfy the separation condition.

For $\boldsymbol{T} \in \mathbb{R}^{q\times r}$, the columns of $[\boldsymbol{I}_r, \boldsymbol{T}^\top]^\top$ span an r-dimensional subspace of $\mathbb{R}^{r+q}$. We orthonormalize these columns and map them into $\mathbb{R}^n$ by defining

$$\boldsymbol{U}^\circ_{\boldsymbol{T}} := \begin{bmatrix}\boldsymbol{I}_r\\ \boldsymbol{T}\end{bmatrix}(\boldsymbol{I}_r + \boldsymbol{T}^\top\boldsymbol{T})^{-1/2}, \qquad \boldsymbol{U}_{\boldsymbol{T}} := \boldsymbol{F}\boldsymbol{U}^\circ_{\boldsymbol{T}}. \tag{145}$$

The corresponding target is the orthogonal projector

$$\boldsymbol{P}_{\boldsymbol{T}} := \boldsymbol{U}_{\boldsymbol{T}}\boldsymbol{U}_{\boldsymbol{T}}^\top = \boldsymbol{F}\begin{bmatrix}\boldsymbol{I}_r\\ \boldsymbol{T}\end{bmatrix}(\boldsymbol{I}_r + \boldsymbol{T}^\top\boldsymbol{T})^{-1}\begin{bmatrix}\boldsymbol{I}_r & \boldsymbol{T}^\top\end{bmatrix}\boldsymbol{F}^\top. \tag{146}$$

Thus $\boldsymbol{T}$ is only a coordinate used to specify the shared subspace; $\boldsymbol{P}_{\boldsymbol{T}}$ is the quantity being estimated.

To construct the associated view-specific directions, define

$$\boldsymbol{R}^\circ_{\boldsymbol{T}} := \begin{bmatrix}-\boldsymbol{T}^\top\\ \boldsymbol{I}_q\end{bmatrix}(\boldsymbol{I}_q + \boldsymbol{T}\boldsymbol{T}^\top)^{-1/2}. \tag{147}$$

The columns of $\boldsymbol{R}^\circ_{\boldsymbol{T}}$ form an orthonormal basis for the orthogonal complement of $\mathrm{col}(\boldsymbol{U}^\circ_{\boldsymbol{T}})$ in $\mathbb{R}^{r+q}$. Indeed, the inverse-square-root factors orthonormalize the two matrices, and direct multiplication shows that their column spaces are orthogonal. Consequently, $[\boldsymbol{U}^\circ_{\boldsymbol{T}}, \boldsymbol{R}^\circ_{\boldsymbol{T}}] \in \mathbb{O}(r+q)$.

The Assouad family. We choose the matrices $\boldsymbol{T}$ so that the family has $r \vee q$ binary coordinates and every nonzero difference has rank one. For $\boldsymbol{\omega} \in \{-1,1\}^{r\vee q}$, define

$$\boldsymbol{T}_{\boldsymbol{\omega}} := \begin{cases}\delta\boldsymbol{e}_1\boldsymbol{\omega}^\top, & r \geq q,\\ \delta\boldsymbol{\omega}\boldsymbol{e}_1^\top, & q > r,\end{cases}$$

where $\boldsymbol{e}_1 \in \mathbb{R}^q$ in the first case and $\boldsymbol{e}_1 \in \mathbb{R}^r$ in the second. Since every nonzero difference has rank one, its operator and Frobenius norms agree. Therefore

$$\|\boldsymbol{T}_{\boldsymbol{\omega}}\|_{\mathrm{op}} = \delta\sqrt{r\vee q}, \qquad \|\boldsymbol{T}_{\boldsymbol{\omega}} - \boldsymbol{T}_{\boldsymbol{\omega}'}\|_{\mathrm{op}} = \|\boldsymbol{T}_{\boldsymbol{\omega}} - \boldsymbol{T}_{\boldsymbol{\omega}'}\|_{\mathrm{F}} = 2\delta\sqrt{\mathrm{Ham}(\boldsymbol{\omega},\boldsymbol{\omega}')}. \tag{148}$$

The rank-one choice serves two purposes. It makes the operator-norm separation proportional to the square root of the Hamming distance, and it ensures that neighboring hypotheses differ in only one shared direction and one direction in its orthogonal complement.

The following deterministic estimate transfers separation between the coordinate matrices to separation between the target projectors. We state it here because it is used by both θ-dependent lower bounds and defer its proof to the subsection containing the proofs of the auxiliary lemmas.

Lemma 27. *For $\boldsymbol{T}, \boldsymbol{S} \in \mathbb{R}^{q\times r}$,*

$$\|\boldsymbol{P_T} - \boldsymbol{P_S}\|_{\mathrm{op}} \geq \frac{\|\boldsymbol{T}-\boldsymbol{S}\|_{\mathrm{op}}}{\sqrt{1+\|\boldsymbol{T}\|_{\mathrm{op}}^2}\sqrt{1+\|\boldsymbol{S}\|_{\mathrm{op}}^2}}.$$

In particular, if $\|\boldsymbol{T}\|_{\mathrm{op}} \vee \|\boldsymbol{S}\|_{\mathrm{op}} \leq 1/2$, then

$$\|\boldsymbol{P_T} - \boldsymbol{P_S}\|_{\mathrm{op}} \geq \frac{4}{5}\|\boldsymbol{T}-\boldsymbol{S}\|_{\mathrm{op}}.$$

We will choose δ so that $\delta\sqrt{r \vee q} \leq 1/2$. Applying Lemma 27 and (148) then gives

$$\left\|\boldsymbol{P}_{\boldsymbol{T}_{\boldsymbol{\omega}}} - \boldsymbol{P}_{\boldsymbol{T}_{\boldsymbol{\omega}'}}\right\|_{\mathrm{op}} \geq \delta\sqrt{\mathrm{Ham}(\boldsymbol{\omega}, \boldsymbol{\omega}')}. \tag{149}$$

The rank-one property will also be essential when comparing neighboring observation laws. For neighboring vertices, $\boldsymbol{T}_{\boldsymbol{\omega}'} - \boldsymbol{T}_{\boldsymbol{\omega}} = \kappa \boldsymbol{a}\boldsymbol{b}^\top$ for some unit vectors $\boldsymbol{a}$ and $\boldsymbol{b}$. The difference vanishes on $\boldsymbol{b}^\perp$, so the corresponding shared subspaces have $r-1$ directions in common. Its transpose vanishes on $\boldsymbol{a}^\perp$, so their orthogonal complements in $\mathrm{col}(\boldsymbol{F})$ have $q-1$ directions in common. Consequently, only one shared direction and one orthogonal direction change, and the two pairs are related by a rotation in a two-dimensional subspace. The neighboring-law lemmas use this reduction to remove all directions whose distributions are identical under the two hypotheses.

View 1 is identical under all hypotheses. For each $\boldsymbol{T}$, define

$$\boldsymbol{V}_{1,\boldsymbol{T}} := [\boldsymbol{F}\boldsymbol{R}_{\boldsymbol{T}}^\circ, \boldsymbol{D}_1] \in \mathrm{St}(n, r_1). \tag{150}$$

The matrices $[\boldsymbol{U_T}, \boldsymbol{V}_{1,\boldsymbol{T}}]$ and $[\boldsymbol{F}, \boldsymbol{D}_1]$ are orthonormal bases of the same subspace. We can therefore choose a first-view signal matrix that is independent of $\boldsymbol{T}$. Fix $\boldsymbol{W}_1 \in \mathrm{St}(d_1, p_1)$ and set

$$\boldsymbol{X}_1 := s_1[\boldsymbol{F}, \boldsymbol{D}_1]\boldsymbol{W}_1^\top. \tag{151}$$

More precisely, there is an orthogonal matrix $\boldsymbol{O_T} \in \mathbb{O}(p_1)$ such that $[\boldsymbol{U_T}, \boldsymbol{V}_{1,\boldsymbol{T}}] = [\boldsymbol{F}, \boldsymbol{D}_1]\boldsymbol{O_T}$. Taking $\boldsymbol{B}_{1,\boldsymbol{T}} := s_1\boldsymbol{W}_1\boldsymbol{O_T}$ represents (151) in the model, and every singular value of $\boldsymbol{B}_{1,\boldsymbol{T}}$ equals s_1. Hence the distribution of $\boldsymbol{Y}_1$ is identical for every $\boldsymbol{T}$ and contains no information for distinguishing the target projectors.

View 2 satisfies the separation condition. Let $\rho := 1-2\theta$ and $\rho_\perp := \sqrt{1-\rho^2}$. For $\boldsymbol{Z} \in \mathrm{St}(m, q)$, define

$$\boldsymbol{V}_{2,\boldsymbol{T},\boldsymbol{Z}} := [\rho\boldsymbol{F}\boldsymbol{R}_{\boldsymbol{T}}^\circ + \rho_\perp\boldsymbol{H}\boldsymbol{Z},\ \boldsymbol{D}_2] \in \mathrm{St}(n, r_2). \tag{152}$$

The matrices $\boldsymbol{F}\boldsymbol{R}_{\boldsymbol{T}}^\circ$ and $\boldsymbol{H}\boldsymbol{Z}$ have orthonormal columns and are mutually orthogonal. Since $\rho^2 + \rho_\perp^2 = 1$, the first q columns of $\boldsymbol{V}_{2,\boldsymbol{T},\boldsymbol{Z}}$ are orthonormal. They are also orthogonal to $\boldsymbol{U_T}$ and $\boldsymbol{D}_2$, so $[\boldsymbol{U_T}, \boldsymbol{V}_{2,\boldsymbol{T},\boldsymbol{Z}}]$ has orthonormal columns.

The alignment with the first view is

$$\boldsymbol{V}_{1,\boldsymbol{T}}^\top\boldsymbol{V}_{2,\boldsymbol{T},\boldsymbol{Z}} = \begin{bmatrix} \rho\boldsymbol{I}_q & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix}.$$

It follows that

$$\left\|\boldsymbol{V}_{1,\boldsymbol{T}}^\top\boldsymbol{V}_{2,\boldsymbol{T},\boldsymbol{Z}}\right\|_{\mathrm{op}} = \rho = 1-2\theta,$$

so the separation condition holds for every $\boldsymbol{T}$ and every $\boldsymbol{Z}$. The target $\boldsymbol{P_T}$ does not depend on $\boldsymbol{Z}$. Thus $\boldsymbol{Z}$ may be fixed in one lower-bound construction and randomized in the other without changing the target.

Reduction to neighboring laws. For either of the second-view loading constructions below, let $\mathsf{P}_{\boldsymbol{\omega}}$ denote the observation law associated with $\boldsymbol{T}_{\boldsymbol{\omega}}$. Suppose that the priors are supported on $\mathfrak{P}$ and that

$$\max_{\substack{\boldsymbol{\omega}\in\{-1,1\}^{r\vee q}\\ j\in[r\vee q]}} \mathrm{TV}\big(\mathsf{P}_{\boldsymbol{\omega}}, \mathsf{P}_{\boldsymbol{\omega}^{(j)}}\big) \leq \frac{1}{4},$$

where $\boldsymbol{\omega}^{(j)}$ is obtained by changing the jth coordinate of $\boldsymbol{\omega}$. Then (149) and Lemma 24 imply

$$\mathcal{R}_{\mathrm{minimax}} \gtrsim \delta\sqrt{r \vee q}. \tag{153}$$

It therefore remains only to choose the second-view loading and δ so that neighboring observation laws have total variation at most $1/4$. Fixing $\boldsymbol{Z}$ gives the $\theta^{-1/2}$ term, whereas mixing over a Haar-distributed $\boldsymbol{Z}$ gives the θ^{-1} term.

Proof of Lemma 27. To verify this claim, consider two distinct vertices and write $\boldsymbol{T} \coloneqq \boldsymbol{T}_{\boldsymbol{\omega}}$ and $\boldsymbol{S} \coloneqq \boldsymbol{T}_{\boldsymbol{\omega}'}$. Choose a unit vector $\boldsymbol{x} \in \mathbb{R}^r$ satisfying $\|(\boldsymbol{T}-\boldsymbol{S})\boldsymbol{x}\|_2 = \|\boldsymbol{T}-\boldsymbol{S}\|_{\mathrm{op}}$, and consider the unit vector

$$\boldsymbol{v} \coloneqq \frac{1}{\sqrt{1+\|\boldsymbol{T}\boldsymbol{x}\|_2^2}} \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{T}\boldsymbol{x} \end{bmatrix} \in \mathrm{col}(\boldsymbol{U}_{\boldsymbol{T}}^{\circ}).$$

Because $\boldsymbol{F}$ has orthonormal columns, the distance between the target projectors equals the distance between the corresponding projectors in $\mathbb{R}^{r+q}$. Moreover, $\boldsymbol{R}_{\boldsymbol{S}}^{\circ}$ spans the orthogonal complement of $\mathrm{col}(\boldsymbol{U}_{\boldsymbol{S}}^{\circ})$. Applying the difference of the two projectors to $\boldsymbol{v}$ therefore gives

$$\begin{aligned}
\|\boldsymbol{P}_{\boldsymbol{T}} - \boldsymbol{P}_{\boldsymbol{S}}\|_{\mathrm{op}} &\geq \|(\boldsymbol{R}_{\boldsymbol{S}}^{\circ})^{\top}\boldsymbol{v}\|_2 \\
&= \frac{\|(\boldsymbol{I}_q + \boldsymbol{S}\boldsymbol{S}^{\top})^{-1/2}(\boldsymbol{T}-\boldsymbol{S})\boldsymbol{x}\|_2}{\sqrt{1+\|\boldsymbol{T}\boldsymbol{x}\|_2^2}} \\
&\geq \frac{\|\boldsymbol{T}-\boldsymbol{S}\|_{\mathrm{op}}}{\sqrt{1+\|\boldsymbol{S}\|_{\mathrm{op}}^2}\sqrt{1+\|\boldsymbol{T}\|_{\mathrm{op}}^2}} \geq \frac{4}{5}\,\|\boldsymbol{T}-\boldsymbol{S}\|_{\mathrm{op}}\,.
\end{aligned}$$

The penultimate inequality uses the smallest eigenvalue of $(\boldsymbol{I}_q + \boldsymbol{S}\boldsymbol{S}^{\top})^{-1/2}$ and the choice of $\boldsymbol{x}$. The final inequality follows from $\|\boldsymbol{T}\|_{\mathrm{op}} \vee \|\boldsymbol{S}\|_{\mathrm{op}} \leq 1/2$. $\square$

G.3 Proof of (140b)

For this bound, we fix the matrix $\boldsymbol{Z}$ in (152). The additional View-2 directions therefore have a known orientation, and the leading change in the second-view covariance is proportional to $\rho_{\perp}$. We construct an admissible Gaussian-loading prior and state the neighboring-law estimate at the scale $\delta \asymp \sigma/(\gamma_2\rho_{\perp})$. The common Assouad reduction in (153) will then give the result.

An admissible prior for View 2. Fix $\boldsymbol{Z}_0 \in \mathrm{St}(m,q)$. Let $\boldsymbol{G}_2 \in \mathbb{R}^{d_2\times p_2}$ have independent $\mathcal{N}(0, d_2^{-1})$ entries, and define

$$\mathcal{E}_2^{\mathrm{load}} \coloneqq \left\{\boldsymbol{G}_2^{\top}\boldsymbol{G}_2 \succeq c_B \boldsymbol{I}_{p_2}\right\}, \qquad \boldsymbol{B}_2 \coloneqq \frac{s_2}{\sqrt{c_B}}\boldsymbol{G}_2.$$

For each matrix $\boldsymbol{T}$ in the Assouad family, condition on $\mathcal{E}_2^{\mathrm{load}}$ and set

$$\boldsymbol{X}_{2,\boldsymbol{T}} \coloneqq [\boldsymbol{U}_{\boldsymbol{T}}, \boldsymbol{V}_{2,\boldsymbol{T},\boldsymbol{Z}_0}]\boldsymbol{B}_2^{\top}. \tag{154}$$

On $\mathcal{E}_2^{\mathrm{load}}$, the loading satisfies $\sigma_{p_2}(\boldsymbol{B}_2) \geq s_2$. Moreover, $[\boldsymbol{U_T}, \boldsymbol{V}_{2,\boldsymbol{T},\boldsymbol{Z}_0}]$ has orthonormal columns and the separation condition was verified in the preceding subsection. Together with the fixed first-view signal in (151), this defines a prior supported on $\mathfrak{P}$ whose target is $\boldsymbol{P_T}$. Let $\mathsf{P}_{\boldsymbol{T}}$ denote the induced observation law.

It remains to show that the laws corresponding to neighboring matrices in the Assouad family are difficult to distinguish. View 1 has the same distribution under every hypothesis, so only View 2 contributes to their total variation distance. After removing the conditioning on $\mathcal{E}_2^{\mathrm{load}}$, the columns of View 2 are independent centered Gaussian vectors, and changing $\boldsymbol{T}$ changes only the projector onto the complete second-view signal subspace. The following lemma gives the required comparison. Its proof is deferred to the subsection containing the proofs of the neighboring-law lemmas.

Lemma 28. *There is a numerical constant $c > 0$ such that the following holds. Let $\boldsymbol{\omega}$ and $\boldsymbol{\omega}'$ be neighboring vertices of the Assouad family, and suppose that*

$$\delta\sqrt{r \vee q} \leq \frac{1}{2}, \qquad \delta \leq c\frac{\sigma}{\gamma_2\rho_\perp}.$$

Then

$$\mathrm{TV}\big(\mathsf{P}_{\boldsymbol{T}_{\boldsymbol{\omega}}}, \mathsf{P}_{\boldsymbol{T}_{\boldsymbol{\omega}'}}\big) \leq \frac{1}{4}.$$

Choose

$$\delta := c_0\left\{\frac{1}{\sqrt{r \vee q}} \wedge \frac{\sigma}{\gamma_2\rho_\perp}\right\},$$

where $c_0 > 0$ is sufficiently small. The first term ensures that $\delta\sqrt{r \vee q} \leq 1/2$, while the second permits the application of Lemma 28 to every neighboring pair. Consequently, (153) gives a minimax risk lower bound of order

$$\delta\sqrt{r \vee q} \asymp 1 \wedge \frac{\sigma\sqrt{r \vee q}}{\gamma_2\rho_\perp} \asymp 1 \wedge \frac{\sigma(\sqrt{r} + \sqrt{q})}{\gamma_{\min}\sqrt{\theta}}.$$

The final comparison uses $\gamma_2 = \gamma_{\min}$, $\rho_\perp^2 = 4\theta(1-\theta) \asymp \theta$, and $\sqrt{r \vee q} \asymp \sqrt{r} + \sqrt{q}$. Since $q = r_1 \wedge r_2$, this proves (140b).

G.4 Proof of (140c)

For the third term, we randomize the matrix $\boldsymbol{Z}$ in (152). Specifically, let

$$\boldsymbol{Z} \sim \mathrm{Haar}\{\mathrm{St}(m, q)\},$$

independently of the loading and observation noise. This is a prior on the view-specific subspace only: for each $\boldsymbol{T}$, the target remains the fixed projector $\boldsymbol{P_T}$. The purpose of the randomization is to make the orientation of the additional View-2 directions unobserved. A comparison conditional on the entire matrix $\boldsymbol{Z}$ would give only the $\theta^{-1/2}$ rate obtained in the preceding subsection.

Set

$$M := m - q + 1 = n - r - r_1 - r_2 + 1.$$

The rank assumptions imply $M \asymp n$. In the proof of the neighboring-law bounds, we condition on $q - 1$ columns of $\boldsymbol{Z}$. The remaining column is then uniform on a sphere in an M-dimensional subspace. The resulting factor M^{-1} in the information bound is the source of the factor $\sqrt{n}$ in the final rate.

We use different loading distributions according to whether $s_2^2 \le c_B d_2 \sigma^2$. Below this threshold, a conditioned Gaussian loading leads to a covariance model with a bounded covariance ratio. Above the threshold, we instead use a deterministic loading with orthonormal columns. This distinction is technical: both constructions give the same neighboring-law bound at the scale

$$\delta \lesssim \frac{\sigma^2 \sqrt{M}}{\gamma_2^2 \rho_\perp^2}.$$

The low-loading regime. Suppose first that $s_2^2 \le c_B d_2 \sigma^2$. Independently of $\boldsymbol{Z}$, let $\boldsymbol{G}_2 \in \mathbb{R}^{d_2 \times p_2}$ have independent $\mathcal{N}(0, d_2^{-1})$ entries, and define

$$\mathcal{E}_2^{\mathrm{load}} := \left\{ \boldsymbol{G}_2^\top \boldsymbol{G}_2 \succeq c_B \boldsymbol{I}_{p_2} \right\}, \qquad \boldsymbol{B}_2 := \frac{s_2}{\sqrt{c_B}} \boldsymbol{G}_2.$$

For each $\boldsymbol{T}$ in the Assouad family, condition on $\mathcal{E}_2^{\mathrm{load}}$ and set

$$\boldsymbol{X}_{2,\boldsymbol{T}} := [\boldsymbol{U}_{\boldsymbol{T}}, \boldsymbol{V}_{2,\boldsymbol{T},\boldsymbol{Z}}] \boldsymbol{B}_2^\top. \tag{155}$$

Every realization satisfies $\sigma_{p_2}(\boldsymbol{B}_2) \ge s_2$, and the separation condition holds for every realization of $\boldsymbol{Z}$. Together with the fixed first-view signal in (151), this defines a prior supported on $\mathfrak{P}$. Let $\mathsf{P}_{\boldsymbol{T}}$ denote the observation law after integrating over both the conditioned loading and $\boldsymbol{Z}$.

The next lemma controls the total variation distance between neighboring laws. Its proof first removes the conditioning on $\mathcal{E}_2^{\mathrm{load}}$ and then compares the resulting Gaussian covariance mixtures after integrating over the unobserved direction. We defer the proof to the subsection containing the neighboring-law calculations.

Lemma 29. *There is a numerical constant $c > 0$ such that the following holds. Let $\boldsymbol{\omega}$ and $\boldsymbol{\omega}'$ be neighboring vertices of the Assouad family. Suppose that*

$$s_2^2 \le c_B d_2 \sigma^2, \qquad \delta \sqrt{r \vee q} \le \frac{1}{2}, \qquad \delta \le c \frac{\sigma^2 \sqrt{M}}{\gamma_2^2 \rho_\perp^2}.$$

Then

$$\mathrm{TV}\big(\mathsf{P}_{\boldsymbol{T}_{\boldsymbol{\omega}}}, \mathsf{P}_{\boldsymbol{T}_{\boldsymbol{\omega}'}}\big) \le \frac{1}{4}.$$

The high-loading regime. Suppose now that $s_2^2 > c_B d_2 \sigma^2$. In this case, choose $\boldsymbol{W}_{2,0} \in \mathrm{St}(d_2, r+q)$ and $\boldsymbol{W}_{2,1} \in \mathrm{St}(d_2, r_2 - q)$ such that $\boldsymbol{W}_{2,0}^\top \boldsymbol{W}_{2,1} = \boldsymbol{0}$, with the second matrix omitted when $r_2 = q$. For each $\boldsymbol{T}$, define the deterministic loading

$$\boldsymbol{B}_{2,\boldsymbol{T}} := s_2 \, [\boldsymbol{W}_{2,0}[\boldsymbol{U}_{\boldsymbol{T}}^\circ, \boldsymbol{R}_{\boldsymbol{T}}^\circ], \, \boldsymbol{W}_{2,1}] \in \mathbb{R}^{d_2 \times p_2}, \tag{156}$$

and set

$$\boldsymbol{X}_{2,\boldsymbol{T}} := [\boldsymbol{U}_{\boldsymbol{T}}, \boldsymbol{V}_{2,\boldsymbol{T},\boldsymbol{Z}}] \boldsymbol{B}_{2,\boldsymbol{T}}^\top. \tag{157}$$

The columns of $\boldsymbol{W}_{2,0}[\boldsymbol{U}_{\boldsymbol{T}}^\circ, \boldsymbol{R}_{\boldsymbol{T}}^\circ]$ and $\boldsymbol{W}_{2,1}$ are orthonormal and mutually orthogonal. Hence every singular value of $\boldsymbol{B}_{2,\boldsymbol{T}}$ equals s_2, so every realization again belongs to the parameter class. Let $\mathsf{P}_{\boldsymbol{T}}$ now denote the observation law obtained after integrating over $\boldsymbol{Z}$.

The factor $[\boldsymbol{U}_{\boldsymbol{T}}^\circ, \boldsymbol{R}_{\boldsymbol{T}}^\circ]$ in (156) is essential. It rotates the right coordinates together with the shared and view-specific directions on the left. To see the resulting cancellation, consider the limiting case $\rho = 1$ and $\rho_\perp = 0$. Then

$$[\boldsymbol{U}_{\boldsymbol{T}}, \boldsymbol{F} \boldsymbol{R}_{\boldsymbol{T}}^\circ][\boldsymbol{U}_{\boldsymbol{T}}^\circ, \boldsymbol{R}_{\boldsymbol{T}}^\circ]^\top = \boldsymbol{F},$$

so the part of $\boldsymbol{X}_{2,\boldsymbol{T}}$ involving $\boldsymbol{W}_{2,0}$ is independent of $\boldsymbol{T}$. For general θ, only terms proportional to $1-\rho$ or $\rho_\perp$ remain when the signal is differentiated with respect to the rotation between neighboring hypotheses.

The corresponding neighboring-law estimate has the same conclusion as in the low-loading case.

Lemma 30. *There is a numerical constant $c > 0$ such that the following holds. Let $\boldsymbol{\omega}$ and $\boldsymbol{\omega}'$ be neighboring vertices of the Assouad family. Suppose that*

$$s_2^2 > c_B d_2 \sigma^2, \qquad \delta\sqrt{r \vee q} \le \frac{1}{2}, \qquad \delta \le c\frac{\sigma^2\sqrt{M}}{\gamma_2^2\rho_\perp^2}.$$

Then

$$\mathrm{TV}\big(\mathsf{P}_{\boldsymbol{T}_{\boldsymbol{\omega}}}, \mathsf{P}_{\boldsymbol{T}_{\boldsymbol{\omega}'}}\big) \le \frac{1}{4}.$$

Choosing the separation. In either loading regime, choose

$$\delta := c_0 \left\{ \frac{1}{\sqrt{r \vee q}} \wedge \frac{\sigma^2\sqrt{M}}{\gamma_2^2\rho_\perp^2} \right\},$$

where $c_0 > 0$ is sufficiently small. The first term ensures that $\delta\sqrt{r \vee q} \le 1/2$, while the second permits the application of the appropriate neighboring-law lemma. Therefore (153) gives a minimax risk lower bound of order

$$\delta\sqrt{r \vee q} \asymp 1 \wedge \frac{\sigma^2\sqrt{M(r \vee q)}}{\gamma_2^2\rho_\perp^2} \asymp 1 \wedge \frac{\sigma^2\sqrt{n}\,(\sqrt{r} + \sqrt{q})}{\gamma_{\min}^2\theta}.$$

The final comparison uses $M \asymp n$, $\gamma_2 = \gamma_{\min}$, $\rho_\perp^2 = 4\theta(1-\theta) \asymp \theta$, and $\sqrt{r \vee q} \asymp \sqrt{r} + \sqrt{q}$. Since $q = r_1 \wedge r_2$, this proves (140c).

G.5 Proofs of the neighboring-law lemmas

Fix neighboring vertices $\boldsymbol{\omega}, \boldsymbol{\omega}'$ and write $\boldsymbol{T}_0 := \boldsymbol{T}_{\boldsymbol{\omega}}$ and $\boldsymbol{T}_1 := \boldsymbol{T}_{\boldsymbol{\omega}'}$. The construction of the Assouad family gives

$$\begin{aligned} \|\boldsymbol{T}_0\|_{\mathrm{op}} = \|\boldsymbol{T}_1\|_{\mathrm{op}} &= \delta\sqrt{r \vee q} \le 1/2, \\ \|\boldsymbol{T}_0 - \boldsymbol{T}_1\|_{\mathrm{F}} &= 2\delta. \end{aligned}$$

In all three constructions, View 1 has the same distribution under $\boldsymbol{T}_0$ and $\boldsymbol{T}_1$ and is independent of View 2. The total variation distance between the joint observation laws therefore equals the distance between their View-2 marginals, which we consider from now on.

For the regular and hidden low-loading constructions, we next replace the loading conditioned on the loading-band event by its unconditioned Gaussian law and account for the resulting total variation error; the high-loading construction uses a deterministic loading and requires no such reduction. It remains to compare the two View-2 laws. The regular case involves product Gaussian laws with different covariance projectors, the hidden low-loading case involves Haar mixtures of Gaussian covariance laws, and the hidden high-loading case involves Haar mixtures of Gaussian mean laws.

For the two Gaussian-loading constructions, write

$$\beta := \frac{s_2^2}{c_B d_2 \sigma^2}. \tag{158}$$

G.5.1 Proof of Lemma 28

Remove the conditioning. Let Q_0 and Q_1 denote the unconditioned View-2 laws corresponding to $\boldsymbol{T}_0$ and $\boldsymbol{T}_1$. By Lemma 22, we have

$$\mathrm{TV}\big(\mathsf{P}_{\boldsymbol{T}_0}, \mathsf{P}_{\boldsymbol{T}_1}\big) \le \frac{1}{16} + \mathrm{TV}(\mathsf{Q}_0, \mathsf{Q}_1) + \frac{1}{16}.$$

It is therefore enough to prove $\mathrm{TV}(\mathsf{Q}_0, \mathsf{Q}_1) \le 1/8$.

The second-view Gaussian laws. After integrating out the unconditioned Gaussian loading, the d_2 columns of View 2 are independent centered Gaussian vectors. Under Q_i, their common covariance is $\sigma^2(\boldsymbol{I}_n + \beta\boldsymbol{\Pi}_i)$, where

$$\boldsymbol{\Pi}_i := \boldsymbol{P}_{\mathrm{col}[\boldsymbol{U}_{\boldsymbol{T}_i}, \boldsymbol{V}_{2,\boldsymbol{T}_i}, \boldsymbol{Z}_0]}.$$

Pinsker's inequality, additivity of relative entropy over the independent columns, and (137) give

$$\begin{aligned}\mathrm{TV}^2(\mathsf{Q}_0, \mathsf{Q}_1) &\le \frac{1}{2}\,\mathrm{KL}(\mathsf{Q}_0 \,\|\, \mathsf{Q}_1)\\ &= \frac{d_2\beta^2}{8(1+\beta)}\|\boldsymbol{\Pi}_0 - \boldsymbol{\Pi}_1\|_{\mathrm{F}}^2 \lesssim \frac{\gamma_2^2}{\sigma^2}\|\boldsymbol{\Pi}_0 - \boldsymbol{\Pi}_1\|_{\mathrm{F}}^2. \end{aligned} \tag{159}$$

For the last inequality, the definitions of β and γ_2 imply

$$\frac{d_2\beta^2}{1+\beta} = \frac{s_2^4}{c_B\sigma^2(s_2^2 + c_B d_2\sigma^2)} \lesssim \frac{s_2^4}{\sigma^2(s_2^2 + d_2\sigma^2)} = \frac{\gamma_2^2}{\sigma^2}.$$

Consequently, it remains only to bound the Frobenius distance between the two covariance projectors.

Distance between the covariance projectors. Set $\boldsymbol{K}_i := \boldsymbol{R}^{\circ}_{\boldsymbol{T}_i}(\boldsymbol{R}^{\circ}_{\boldsymbol{T}_i})^{\top}$. Because $[\boldsymbol{U}^{\circ}_{\boldsymbol{T}_i}, \boldsymbol{R}^{\circ}_{\boldsymbol{T}_i}]$ is an orthogonal matrix, the projector onto $\mathrm{col}(\boldsymbol{U}^{\circ}_{\boldsymbol{T}_i})$ is $\boldsymbol{I}_{r+q} - \boldsymbol{K}_i$. Expanding $\boldsymbol{\Pi}_i$ using (152) and cancelling the terms that do not depend on i therefore gives

$$\begin{aligned}\boldsymbol{\Pi}_0 - \boldsymbol{\Pi}_1 = &\ -\rho_\perp^2\boldsymbol{F}(\boldsymbol{K}_0 - \boldsymbol{K}_1)\boldsymbol{F}^\top\\ &+ \rho\rho_\perp\Big\{\boldsymbol{F}(\boldsymbol{R}^{\circ}_{\boldsymbol{T}_0} - \boldsymbol{R}^{\circ}_{\boldsymbol{T}_1})\boldsymbol{Z}_0^\top\boldsymbol{H}^\top + \boldsymbol{H}\boldsymbol{Z}_0(\boldsymbol{R}^{\circ}_{\boldsymbol{T}_0} - \boldsymbol{R}^{\circ}_{\boldsymbol{T}_1})^\top\boldsymbol{F}^\top\Big\}. \end{aligned} \tag{160}$$

We next compare the two matrices $\boldsymbol{R}^{\circ}_{\boldsymbol{T}_i}$. Define $\boldsymbol{L}_i := (\boldsymbol{I}_q + \boldsymbol{T}_i\boldsymbol{T}_i^\top)^{-1/2}$, so that $\boldsymbol{R}^{\circ}_{\boldsymbol{T}_i} = [-\boldsymbol{T}_i^\top, \boldsymbol{I}_q]^\top\boldsymbol{L}_i$. For positive definite matrices $\boldsymbol{A}, \boldsymbol{B} \succeq \boldsymbol{I}_q$, the integral representation of the inverse square root and the resolvent identity give $\|\boldsymbol{A}^{-1/2} - \boldsymbol{B}^{-1/2}\|_{\mathrm{F}} \le \frac{1}{2}\|\boldsymbol{A} - \boldsymbol{B}\|_{\mathrm{F}}$. Moreover,

$$\|\boldsymbol{T}_0\boldsymbol{T}_0^\top - \boldsymbol{T}_1\boldsymbol{T}_1^\top\|_{\mathrm{F}} \le \big(\|\boldsymbol{T}_0\|_{\mathrm{op}} + \|\boldsymbol{T}_1\|_{\mathrm{op}}\big)\|\boldsymbol{T}_0 - \boldsymbol{T}_1\|_{\mathrm{F}} \le \|\boldsymbol{T}_0 - \boldsymbol{T}_1\|_{\mathrm{F}}.$$

Using these bounds and $\|\boldsymbol{T}_0\|_{\mathrm{op}} \vee \|\boldsymbol{T}_1\|_{\mathrm{op}} \le 1/2$, we obtain

$$\begin{aligned}\|\boldsymbol{L}_0 - \boldsymbol{L}_1\|_{\mathrm{F}} &\le \frac{1}{2}\|\boldsymbol{T}_0 - \boldsymbol{T}_1\|_{\mathrm{F}},\\ \|\boldsymbol{R}^{\circ}_{\boldsymbol{T}_0} - \boldsymbol{R}^{\circ}_{\boldsymbol{T}_1}\|_{\mathrm{F}} &\le 2\|\boldsymbol{T}_0 - \boldsymbol{T}_1\|_{\mathrm{F}},\\ \|\boldsymbol{K}_0 - \boldsymbol{K}_1\|_{\mathrm{F}} &\le 4\|\boldsymbol{T}_0 - \boldsymbol{T}_1\|_{\mathrm{F}}. \end{aligned} \tag{161}$$

The last inequality follows by expanding $\boldsymbol{K}_0 - \boldsymbol{K}_1$ and using that both $\boldsymbol{R}^\circ_{\boldsymbol{T}_0}$ and $\boldsymbol{R}^\circ_{\boldsymbol{T}_1}$ have orthonormal columns.

The three terms in (160) belong respectively to the col($\boldsymbol{F}$)–col($\boldsymbol{F}$), col($\boldsymbol{F}$)–col($\boldsymbol{H}$), and col($\boldsymbol{H}$)–col($\boldsymbol{F}$) matrix blocks. They are therefore pairwise orthogonal in the Frobenius inner product. Since $\boldsymbol{F}$ and $\boldsymbol{H}\boldsymbol{Z}_0$ have orthonormal columns,

$$\begin{aligned}\|\boldsymbol{\Pi}_0 - \boldsymbol{\Pi}_1\|_{\mathrm{F}}^2 &= \rho_\perp^4 \|\boldsymbol{K}_0 - \boldsymbol{K}_1\|_{\mathrm{F}}^2 + 2\rho^2\rho_\perp^2 \|\boldsymbol{R}^\circ_{\boldsymbol{T}_0} - \boldsymbol{R}^\circ_{\boldsymbol{T}_1}\|_{\mathrm{F}}^2 \\ &\le 24\rho_\perp^2 \|\boldsymbol{T}_0 - \boldsymbol{T}_1\|_{\mathrm{F}}^2. \end{aligned} \tag{162}$$

Completing the proof. Combining this projector bound with (159) and using $\|\boldsymbol{T}_0 - \boldsymbol{T}_1\|_{\mathrm{F}} = 2\delta$ gives

$$\mathrm{TV}^2(\mathsf{Q}_0, \mathsf{Q}_1) \lesssim \frac{\gamma_2^2 \rho_\perp^2 \delta^2}{\sigma^2}.$$

The assumption $\delta \le c\sigma/(\gamma_2\rho_\perp)$ therefore implies $\mathrm{TV}(\mathsf{Q}_0, \mathsf{Q}_1) \le 1/8$ when the numerical constant c is sufficiently small. Returning to the conditioning bound,

$$\mathrm{TV}(\mathsf{P}_{\boldsymbol{T}_0}, \mathsf{P}_{\boldsymbol{T}_1}) \le \frac{1}{16} + \frac{1}{8} + \frac{1}{16} = \frac{1}{4}.$$

This proves Lemma 28.

G.5.2 Proof of Lemma 29

Remove the conditioning. Let Q_i denote the unconditioned View-2 law corresponding to $\boldsymbol{T}_i$, for $i \in \{0, 1\}$. By Lemma 22,

$$\mathrm{TV}(\mathsf{P}_{\boldsymbol{T}_0}, \mathsf{P}_{\boldsymbol{T}_1}) \le \frac{1}{16} + \mathrm{TV}(\mathsf{Q}_0, \mathsf{Q}_1) + \frac{1}{16}.$$

It is therefore enough to prove $\mathrm{TV}(\mathsf{Q}_0, \mathsf{Q}_1) \le 1/8$. We reduce these unconditioned laws to an $(M+2)$-dimensional covariance mixture and compare the resulting laws along a Hellinger path.

The second-view covariance mixtures. For $\boldsymbol{T} \in \{\boldsymbol{T}_0, \boldsymbol{T}_1\}$ and $\boldsymbol{Z} \in \mathrm{St}(m, q)$, define

$$\boldsymbol{\Pi}_{\boldsymbol{T},\boldsymbol{Z}} := \boldsymbol{P}_{\mathrm{col}[\boldsymbol{U}_{\boldsymbol{T}}, \boldsymbol{V}_{2,\boldsymbol{T},\boldsymbol{Z}}]}.$$

After integrating out the Gaussian loading, the d_2 columns of View 2 are conditionally independent centered Gaussian vectors. Thus

$$\mathsf{Q}_i = \int \left[\mathcal{N}\left(\boldsymbol{0}, \sigma^2\{\boldsymbol{I}_n + \beta\boldsymbol{\Pi}_{\boldsymbol{T}_i,\boldsymbol{Z}}\}\right)\right]^{\otimes d_2} d\mu_{\mathrm{Haar}}(\boldsymbol{Z}), \tag{163}$$

where β is defined in (158) and satisfies $\beta \le 1$ in the low-loading regime. We next use Haar right invariance to give the directions that agree under $\boldsymbol{T}_0$ and $\boldsymbol{T}_1$ the same coordinates.

Reduction to an $(M+2)$-dimensional covariance mixture. Because $\boldsymbol{\omega}$ and $\boldsymbol{\omega}'$ are neighboring vertices,

$$\boldsymbol{T}_1 - \boldsymbol{T}_0 = \kappa \boldsymbol{a}\boldsymbol{b}^\top, \qquad \|\boldsymbol{a}\|_2 = \|\boldsymbol{b}\|_2 = 1. \tag{164}$$

Choose $\boldsymbol{A}_\perp \in \mathrm{St}(q, q-1)$ and $\boldsymbol{B}_\perp \in \mathrm{St}(r, r-1)$ whose columns span $\boldsymbol{a}^\perp$ and $\boldsymbol{b}^\perp$, respectively. Then

$$\begin{bmatrix}\boldsymbol{I}_r \\ \boldsymbol{T}_0\end{bmatrix}\boldsymbol{B}_\perp = \begin{bmatrix}\boldsymbol{I}_r \\ \boldsymbol{T}_1\end{bmatrix}\boldsymbol{B}_\perp, \qquad \begin{bmatrix}-\boldsymbol{T}_0^\top \\ \boldsymbol{I}_q\end{bmatrix}\boldsymbol{A}_\perp = \begin{bmatrix}-\boldsymbol{T}_1^\top \\ \boldsymbol{I}_q\end{bmatrix}\boldsymbol{A}_\perp. \tag{165}$$

Choose orthonormal matrices $\boldsymbol{U}_* \in \mathrm{St}(r+q, r-1)$ and $\boldsymbol{R}_* \in \mathrm{St}(r+q, q-1)$ with column spaces equal to the two column spaces in (165). For each $i \in \{0,1\}$, choose unit vectors $\boldsymbol{u}_i, \boldsymbol{r}_i \in \mathbb{R}^{r+q}$ so that

$$\mathrm{col}(\boldsymbol{U}^\circ_{\boldsymbol{T}_i}) = \mathrm{col}[\boldsymbol{U}_*, \boldsymbol{u}_i], \qquad \mathrm{col}(\boldsymbol{R}^\circ_{\boldsymbol{T}_i}) = \mathrm{col}[\boldsymbol{R}_*, \boldsymbol{r}_i]. \tag{166}$$

The matrix $[\boldsymbol{U}_*, \boldsymbol{u}_i, \boldsymbol{R}_*, \boldsymbol{r}_i]$ is orthogonal. Its first $r-1$ and next $q-1$ columns do not depend on i, and hence $\mathrm{col}[\boldsymbol{u}_0, \boldsymbol{r}_0] = \mathrm{col}[\boldsymbol{u}_1, \boldsymbol{r}_1]$.

For each i, define the change-of-basis matrix

$$\boldsymbol{Q}_i := [\boldsymbol{R}_*, \boldsymbol{r}_i]^\top \boldsymbol{R}^\circ_{\boldsymbol{T}_i} \in \mathrm{O}(q), \qquad \boldsymbol{R}^\circ_{\boldsymbol{T}_i} \boldsymbol{Q}_i^\top = [\boldsymbol{R}_*, \boldsymbol{r}_i]. \tag{167}$$

Right multiplication by $\boldsymbol{Q}_i^\top$ does not change a column projector, and Haar right invariance gives $\boldsymbol{Z}\boldsymbol{Q}_i^\top \overset{\mathrm{d}}{=} \boldsymbol{Z}$. Therefore the mixture in (163) is unchanged if $\boldsymbol{\Pi}_{\boldsymbol{T}_i,\boldsymbol{Z}}$ is replaced by

$$\begin{aligned} \widetilde{\boldsymbol{\Pi}}_{i,\boldsymbol{Z}} := {} & \boldsymbol{F}\boldsymbol{U}_*\boldsymbol{U}_*^\top \boldsymbol{F}^\top + (\boldsymbol{F}\boldsymbol{u}_i)(\boldsymbol{F}\boldsymbol{u}_i)^\top \\ & + \big(\rho \boldsymbol{F}[\boldsymbol{R}_*, \boldsymbol{r}_i] + \rho_\perp \boldsymbol{H}\boldsymbol{Z}\big)\big(\rho \boldsymbol{F}[\boldsymbol{R}_*, \boldsymbol{r}_i] + \rho_\perp \boldsymbol{H}\boldsymbol{Z}\big)^\top + \boldsymbol{D}_2\boldsymbol{D}_2^\top. \end{aligned} \tag{168}$$

Write $\boldsymbol{Z} = [\boldsymbol{Z}_-, \boldsymbol{z}]$, where $\boldsymbol{Z}_- \in \mathrm{St}(m, q-1)$. Conditional on $\boldsymbol{Z}_-$, choose $\boldsymbol{O} = \boldsymbol{O}(\boldsymbol{Z}_-) \in \mathrm{St}(m, M)$ by a fixed measurable rule so that $\mathrm{col}(\boldsymbol{O}) = \mathrm{col}(\boldsymbol{Z}_-)^\perp$. Then

$$\boldsymbol{z} = \boldsymbol{O}\boldsymbol{\zeta}, \qquad \boldsymbol{\zeta} \sim \mathrm{Unif}(\mathbb{S}^{M-1}). \tag{169}$$

Define

$$\boldsymbol{\Pi}_{\mathrm{fix}}(\boldsymbol{Z}_-) := \boldsymbol{F}\boldsymbol{U}_*\boldsymbol{U}_*^\top \boldsymbol{F}^\top + (\rho \boldsymbol{F}\boldsymbol{R}_* + \rho_\perp \boldsymbol{H}\boldsymbol{Z}_-)(\rho \boldsymbol{F}\boldsymbol{R}_* + \rho_\perp \boldsymbol{H}\boldsymbol{Z}_-)^\top + \boldsymbol{D}_2\boldsymbol{D}_2^\top. \tag{170}$$

Substitution in (168) gives

$$\begin{aligned} \widetilde{\boldsymbol{\Pi}}_{i,\boldsymbol{Z}} = {} & \boldsymbol{\Pi}_{\mathrm{fix}}(\boldsymbol{Z}_-) + (\boldsymbol{F}\boldsymbol{u}_i)(\boldsymbol{F}\boldsymbol{u}_i)^\top \\ & + (\rho \boldsymbol{F}\boldsymbol{r}_i + \rho_\perp \boldsymbol{H}\boldsymbol{O}\boldsymbol{\zeta})(\rho \boldsymbol{F}\boldsymbol{r}_i + \rho_\perp \boldsymbol{H}\boldsymbol{O}\boldsymbol{\zeta})^\top. \end{aligned} \tag{171}$$

The last two terms in (171) lie in the column space of the following isometry. Define their coordinates by

$$\begin{aligned} & \boldsymbol{J} := [\boldsymbol{F}\boldsymbol{u}_0, \boldsymbol{F}\boldsymbol{r}_0, \boldsymbol{H}\boldsymbol{O}] \in \mathrm{St}(n, M+2), \\ & \boldsymbol{a}_i := \boldsymbol{J}^\top \boldsymbol{F}\boldsymbol{u}_i, \qquad \boldsymbol{b}_i := \boldsymbol{J}^\top \boldsymbol{F}\boldsymbol{r}_i, \quad \widetilde{\boldsymbol{z}} := \begin{bmatrix} 0 & 0 & \boldsymbol{\zeta}^\top \end{bmatrix}^\top. \end{aligned} \tag{172}$$

By construction, $\boldsymbol{\Pi}_{\mathrm{fix}}(\boldsymbol{Z}_-)\boldsymbol{J} = \boldsymbol{0}$. Moreover, $\boldsymbol{J}\boldsymbol{a}_i = \boldsymbol{F}\boldsymbol{u}_i$, $\boldsymbol{J}\boldsymbol{b}_i = \boldsymbol{F}\boldsymbol{r}_i$, and $\boldsymbol{J}\widetilde{\boldsymbol{z}} = \boldsymbol{H}\boldsymbol{O}\boldsymbol{\zeta}$. Hence, with

$$\boldsymbol{\Gamma}_i(\widetilde{\boldsymbol{z}}) := \boldsymbol{a}_i\boldsymbol{a}_i^\top + (\rho \boldsymbol{b}_i + \rho_\perp \widetilde{\boldsymbol{z}})(\rho \boldsymbol{b}_i + \rho_\perp \widetilde{\boldsymbol{z}})^\top, \tag{173}$$

the projector in (171) satisfies

$$\widetilde{\boldsymbol{\Pi}}_{i,\boldsymbol{Z}} = \boldsymbol{\Pi}_{\mathrm{fix}}(\boldsymbol{Z}_-) + \boldsymbol{J}\boldsymbol{\Gamma}_i(\widetilde{\boldsymbol{z}})\boldsymbol{J}^\top. \tag{174}$$

The relation $\boldsymbol{\Pi}_{\mathrm{fix}}(\boldsymbol{Z}_-)\boldsymbol{J} = \boldsymbol{0}$ makes the Gaussian coordinates in $\mathrm{col}(\boldsymbol{J})$ independent of those in its orthogonal complement. Define $\boldsymbol{W}_j := \boldsymbol{J}^\top \boldsymbol{y}_{2,j} \in \mathbb{R}^{M+2}$. Conditional on $\boldsymbol{\zeta}$,

$$\boldsymbol{W}_j \overset{\mathrm{iid}}{\sim} \mathcal{N}\left(\boldsymbol{0}, \sigma^2\{\boldsymbol{I}_{M+2} + \beta \boldsymbol{\Gamma}_i(\widetilde{\boldsymbol{z}})\}\right), \qquad j = 1, \ldots, d_2. \tag{175}$$

Let $\mathcal{Q}_i$ denote the law obtained by first drawing $\boldsymbol{\zeta} \sim \mathrm{Unif}(\mathbb{S}^{M-1})$ and then drawing the vectors in (175). After revealing $(\boldsymbol{Z}_-, \boldsymbol{O})$, the remaining Gaussian coordinates are independent of $\boldsymbol{W}_1, \ldots, \boldsymbol{W}_{d_2}$ and have the same law for $i = 0, 1$. Therefore

$$\mathrm{TV}(\mathsf{Q}_0, \mathsf{Q}_1) \le \mathrm{TV}(\mathcal{Q}_0, \mathcal{Q}_1). \tag{176}$$

Hellinger comparison and conclusion. The two orthonormal bases $(\boldsymbol{a}_0, \boldsymbol{b}_0)$ and $(\boldsymbol{a}_1, \boldsymbol{b}_1)$ span the same two-dimensional subspace. After choosing their signs consistently, there is an angle $\varphi \in [0, \pi/2]$ such that

$$\begin{aligned} \boldsymbol{a}_1 &= \cos(\varphi)\boldsymbol{a}_0 + \sin(\varphi)\boldsymbol{b}_0, \\ \boldsymbol{b}_1 &= -\sin(\varphi)\boldsymbol{a}_0 + \cos(\varphi)\boldsymbol{b}_0. \end{aligned} \tag{177}$$

Because $\boldsymbol{u}_0 \in \operatorname{col}(\boldsymbol{U}^{\circ}_{\boldsymbol{T}_0})$, there is $\boldsymbol{x} \in \mathbb{R}^r$ with $\|\boldsymbol{x}\|_2 \le 1$ such that $\boldsymbol{u}_0 = [\boldsymbol{I}_r; \boldsymbol{T}_0]\boldsymbol{x}$. Since $[\boldsymbol{I}_r; \boldsymbol{T}_1]\boldsymbol{x}$ belongs to $\operatorname{col}(\boldsymbol{U}^{\circ}_{\boldsymbol{T}_1})$,

$$\sin\varphi = \operatorname{dist}\big(\boldsymbol{u}_0, \operatorname{col}(\boldsymbol{U}^{\circ}_{\boldsymbol{T}_1})\big) \le \|(\boldsymbol{T}_0 - \boldsymbol{T}_1)\boldsymbol{x}\|_2 \le 2\delta. \tag{178}$$

It follows that

$$\varphi \le 2\sin\varphi \le 4\delta. \tag{179}$$

For $t \in [0, \varphi]$, set

$$\begin{aligned} \boldsymbol{a}_t &:= \cos(t)\boldsymbol{a}_0 + \sin(t)\boldsymbol{b}_0, \\ \boldsymbol{b}_t &:= -\sin(t)\boldsymbol{a}_0 + \cos(t)\boldsymbol{b}_0, \\ \boldsymbol{\Gamma}_t(\widetilde{\boldsymbol{z}}) &:= \boldsymbol{a}_t\boldsymbol{a}_t^{\top} + (\rho\boldsymbol{b}_t + \rho_{\perp}\widetilde{\boldsymbol{z}})(\rho\boldsymbol{b}_t + \rho_{\perp}\widetilde{\boldsymbol{z}})^{\top}. \end{aligned} \tag{180}$$

Let $\mathcal{Q}_t$ be the law obtained by drawing $\boldsymbol{\zeta} \sim \operatorname{Unif}(\mathbb{S}^{M-1})$, forming $\widetilde{\boldsymbol{z}}$ as in (172), and then drawing

$$\boldsymbol{W}_j \overset{\text{iid}}{\sim} \mathcal{N}\left(\boldsymbol{0}, \, \sigma^2\{\boldsymbol{I}_{M+2} + \beta\boldsymbol{\Gamma}_t(\widetilde{\boldsymbol{z}})\}\right), \qquad j = 1, \ldots, d_2. \tag{181}$$

Then $\mathcal{Q}_{t=0} = \mathcal{Q}_0$ and $\mathcal{Q}_{t=\varphi} = \mathcal{Q}_1$.

Let q_t be the density of $\mathcal{Q}_t$, and define

$$S_t(\boldsymbol{W}) := \partial_t \log q_t(\boldsymbol{W}), \qquad J_t := \mathbb{E}_{\mathcal{Q}_t} S_t(\boldsymbol{W})^2. \tag{182}$$

By Lemma 25,

$$\operatorname{TV}(\mathcal{Q}_0, \mathcal{Q}_1) \le \frac{1}{2}\int_0^{\varphi} \sqrt{J_t}\, dt \le \frac{\varphi}{2} \sup_{0 \le t \le \varphi} \sqrt{J_t}. \tag{183}$$

It remains to bound J_t uniformly over the path. The coefficient that appears in this bound satisfies

$$I := \frac{d_2\beta^2}{1+\beta} = \frac{s_2^4}{c_B\sigma^2(s_2^2 + c_B d_2\sigma^2)} \asymp \frac{\gamma_2^2}{\sigma^2}. \tag{184}$$

Since $\beta \le 1$, we have $I \le d_2$. The signal-strength assumption and $M \asymp n$ also give

$$d_2 \ge I \ge c_1\nu^2 M. \tag{185}$$

The law $\mathcal{Q}_t$ is a mixture over $\widetilde{\boldsymbol{z}}$, so the score cannot be computed as if $\widetilde{\boldsymbol{z}}$ were observed. The following lemma controls the mixture score. Its proof follows the three neighboring-law proofs.

Lemma 31. *Under (185), uniformly for $t \in [0, \varphi]$,*

$$J_t \le C\left(I\rho_{\perp}^4 + \frac{I^2\rho_{\perp}^4}{M} + I\rho_{\perp}^2 e^{-c_2 d_2}\right). \tag{186}$$

Combining (183) and (186) yields

$$\operatorname{TV}^2(\mathcal{Q}_0, \mathcal{Q}_1) \le C\varphi^2\left(I\rho_{\perp}^4 + \frac{I^2\rho_{\perp}^4}{M} + I\rho_{\perp}^2 e^{-c_2 d_2}\right). \tag{187}$$

By the assumed bound on δ, (179) and (184) imply

$$\varphi \le Cc\frac{\sqrt{M}}{I\rho_\perp^2}. \tag{188}$$

Consequently,

$$\varphi^2\frac{I^2\rho_\perp^4}{M} \le Cc^2, \qquad \varphi^2 I\rho_\perp^4 \le Cc^2\frac{M}{I} \le Cc^2\nu^{-2}.$$

Also, $\varphi \le 2$, $I \le d_2$, and $\rho_\perp \le 1$ give $\varphi^2 I\rho_\perp^2 e^{-c_2 d_2} \le 4d_2 e^{-c_2 d_2}$. Since $d_2 \ge c_1\nu^2 M$, the last quantity is sufficiently small for sufficiently large ν. Choosing the numerical constant c in the edge assumption sufficiently small now gives $\mathrm{TV}(\mathcal{Q}_0, \mathcal{Q}_1) \le 1/8$. Hence (176) and the conditioning bound at the start of the proof yield

$$\mathrm{TV}(\mathsf{P}_{\boldsymbol{T}_0}, \mathsf{P}_{\boldsymbol{T}_1}) \le \frac{1}{16} + \frac{1}{8} + \frac{1}{16} = \frac{1}{4}.$$

This proves Lemma 29.

G.5.3 Proof of Lemma 30

Let $\mathsf{P}_i^{(2)}$ denote the View-2 law under $\boldsymbol{T}_i$, for $i \in \{0,1\}$. The loading is deterministic in this construction, so there is no conditioning error. By the common reduction above,

$$\mathrm{TV}\big(\mathsf{P}_{\boldsymbol{T}_0}, \mathsf{P}_{\boldsymbol{T}_1}\big) = \mathrm{TV}\big(\mathsf{P}_0^{(2)}, \mathsf{P}_1^{(2)}\big).$$

We reduce the View-2 laws to an $(M+2)\times 2$ Gaussian mean model and then compare the reduced laws along a Hellinger path.

The second-view mean mixtures. Write

$$\boldsymbol{M}_{\boldsymbol{T},\boldsymbol{Z}} := [\boldsymbol{U}_{\boldsymbol{T}}, \rho\boldsymbol{F}\boldsymbol{R}_{\boldsymbol{T}}^\circ + \rho_\perp\boldsymbol{H}\boldsymbol{Z}]\,[\boldsymbol{U}_{\boldsymbol{T}}^\circ, \boldsymbol{R}_{\boldsymbol{T}}^\circ]^\top.$$

Conditional on $\boldsymbol{Z} \in \mathrm{St}(m,q)$, View 2 has entrywise noise variance σ^2 and the mean appearing below. Consequently,

$$\mathsf{P}_i^{(2)} = \int \mathcal{N}\left(\mathrm{vec}\left\{s_2\boldsymbol{M}_{\boldsymbol{T}_i,\boldsymbol{Z}}\boldsymbol{W}_{2,0}^\top + s_2\boldsymbol{D}_2\boldsymbol{W}_{2,1}^\top\right\}, \sigma^2\boldsymbol{I}_{nd_2}\right) d\mu_{\mathrm{Haar}}(\boldsymbol{Z}). \tag{189}$$

It remains to express the two mixtures in coordinates in which all terms that depend on i appear in one Gaussian block.

Reduction to the Gaussian mean model. Use the matrices and vectors in (166) and (167). The identities in those displays give

$$\begin{aligned}\boldsymbol{F}\boldsymbol{U}_{\boldsymbol{T}_i}^\circ(\boldsymbol{U}_{\boldsymbol{T}_i}^\circ)^\top &= \boldsymbol{F}[\boldsymbol{U}_*, \boldsymbol{u}_i][\boldsymbol{U}_*, \boldsymbol{u}_i]^\top,\\ &(\rho\boldsymbol{F}\boldsymbol{R}_{\boldsymbol{T}_i}^\circ + \rho_\perp\boldsymbol{H}\boldsymbol{Z})(\boldsymbol{R}_{\boldsymbol{T}_i}^\circ)^\top\\ &\qquad= (\rho\boldsymbol{F}[\boldsymbol{R}_*, \boldsymbol{r}_i] + \rho_\perp\boldsymbol{H}\boldsymbol{Z}\boldsymbol{Q}_i^\top)[\boldsymbol{R}_*, \boldsymbol{r}_i]^\top.\end{aligned} \tag{190}$$

Since $\boldsymbol{Z}\boldsymbol{Q}_i^\top \overset{\mathrm{d}}{=} \boldsymbol{Z}$, the mixture in (189) is unchanged if $\boldsymbol{M}_{\boldsymbol{T}_i,\boldsymbol{Z}}$ is replaced by

$$\widetilde{\boldsymbol{M}}_{i,\boldsymbol{Z}} := \boldsymbol{F}[\boldsymbol{U}_*, \boldsymbol{u}_i][\boldsymbol{U}_*, \boldsymbol{u}_i]^\top + (\rho\boldsymbol{F}[\boldsymbol{R}_*, \boldsymbol{r}_i] + \rho_\perp\boldsymbol{H}\boldsymbol{Z})[\boldsymbol{R}_*, \boldsymbol{r}_i]^\top. \tag{191}$$

Write $\boldsymbol{Z} = [\boldsymbol{Z}_-, \boldsymbol{z}]$ and condition on $\boldsymbol{Z}_-$. Choose $\boldsymbol{O} \in \mathrm{St}(m, M)$ with $\mathrm{col}(\boldsymbol{O}) = \mathrm{col}(\boldsymbol{Z}_-)^\perp$ as in (169), so that $\boldsymbol{z} = \boldsymbol{O}\boldsymbol{\zeta}$ for $\boldsymbol{\zeta} \sim \mathrm{Unif}(\mathbb{S}^{M-1})$. Expanding (191) gives

$$\begin{aligned}\widetilde{\boldsymbol{M}}_{i,\boldsymbol{Z}} = \boldsymbol{F}\boldsymbol{U}_*\boldsymbol{U}_*^\top + (\rho\boldsymbol{F}\boldsymbol{R}_* + \rho_\perp\boldsymbol{H}\boldsymbol{Z}_-)\boldsymbol{R}_*^\top \\ + (\boldsymbol{F}\boldsymbol{u}_i)\boldsymbol{u}_i^\top + (\rho\boldsymbol{F}\boldsymbol{r}_i + \rho_\perp\boldsymbol{H}\boldsymbol{O}\boldsymbol{\zeta})\boldsymbol{r}_i^\top.\end{aligned} \tag{192}$$

The first line is independent of i. We now isolate the last two terms by defining

$$\begin{aligned}&\boldsymbol{J}_{\mathrm{L}} := [\boldsymbol{F}\boldsymbol{u}_0, \boldsymbol{F}\boldsymbol{r}_0, \boldsymbol{H}\boldsymbol{O}] \in \mathrm{St}(n, M+2), \quad &&\boldsymbol{J}_{\mathrm{R}} := [\boldsymbol{u}_0, \boldsymbol{r}_0] \in \mathrm{St}(r+q, 2), \\ &\boldsymbol{a}_i := \boldsymbol{J}_{\mathrm{L}}^\top\boldsymbol{F}\boldsymbol{u}_i, &&\boldsymbol{b}_i := \boldsymbol{J}_{\mathrm{L}}^\top\boldsymbol{F}\boldsymbol{r}_i, \\ &\overline{\boldsymbol{a}}_i := \boldsymbol{J}_{\mathrm{R}}^\top\boldsymbol{u}_i, &&\overline{\boldsymbol{b}}_i := \boldsymbol{J}_{\mathrm{R}}^\top\boldsymbol{r}_i, \qquad \widetilde{\boldsymbol{z}} := \begin{bmatrix} 0 & 0 & \boldsymbol{\zeta}^\top \end{bmatrix}^\top.\end{aligned} \tag{193}$$

With these coordinates, the second line of (192) equals $\boldsymbol{J}_{\mathrm{L}}\boldsymbol{\mathcal{M}}_i(\widetilde{\boldsymbol{z}})\boldsymbol{J}_{\mathrm{R}}^\top$, where

$$\boldsymbol{\mathcal{M}}_i(\widetilde{\boldsymbol{z}}) := \boldsymbol{a}_i\overline{\boldsymbol{a}}_i^\top + (\rho\boldsymbol{b}_i + \rho_\perp\widetilde{\boldsymbol{z}})\overline{\boldsymbol{b}}_i^\top. \tag{194}$$

Conditional on $(\boldsymbol{Z}_-, \boldsymbol{O})$, apply orthogonal changes of coordinates whose first left and right blocks are $\boldsymbol{J}_{\mathrm{L}}$ and $\boldsymbol{W}_{2,0}\boldsymbol{J}_{\mathrm{R}}$, respectively. The corresponding block of the mean is exactly $s_2\boldsymbol{\mathcal{M}}_i(\widetilde{\boldsymbol{z}})$; all other mean blocks are independent of i, and the Gaussian noise blocks are independent. Let $\mathcal{Q}_i$ denote the law of

$$\boldsymbol{W} = s_2\boldsymbol{\mathcal{M}}_i(\widetilde{\boldsymbol{z}}) + \sigma\boldsymbol{G}, \qquad \boldsymbol{G} \in \mathbb{R}^{(M+2)\times 2}, \tag{195}$$

where $\boldsymbol{G}$ has independent standard Gaussian entries and $\boldsymbol{\zeta}$ is uniform on $\mathbb{S}^{M-1}$. Revealing $(\boldsymbol{Z}_-, \boldsymbol{O})$ can only increase total variation, and the resulting conditional distance does not depend on the revealed variables. Therefore

$$\mathrm{TV}(\mathsf{P}_0^{(2)}, \mathsf{P}_1^{(2)}) \le \mathrm{TV}(\mathcal{Q}_0, \mathcal{Q}_1). \tag{196}$$

Hellinger comparison and conclusion. The rotation in (177) acts on both the left and right coordinates in (194). For $t \in [0, \varphi]$, define

$$\begin{aligned}\boldsymbol{a}_t &:= \cos(t)\boldsymbol{a}_0 + \sin(t)\boldsymbol{b}_0, \quad \boldsymbol{b}_t := -\sin(t)\boldsymbol{a}_0 + \cos(t)\boldsymbol{b}_0, \\ \overline{\boldsymbol{a}}_t &:= \cos(t)\overline{\boldsymbol{a}}_0 + \sin(t)\overline{\boldsymbol{b}}_0, \quad \overline{\boldsymbol{b}}_t := -\sin(t)\overline{\boldsymbol{a}}_0 + \cos(t)\overline{\boldsymbol{b}}_0.\end{aligned} \tag{197}$$

Set

$$\boldsymbol{\mathcal{M}}_t(\widetilde{\boldsymbol{z}}) := \boldsymbol{a}_t\overline{\boldsymbol{a}}_t^\top + (\rho\boldsymbol{b}_t + \rho_\perp\widetilde{\boldsymbol{z}})\overline{\boldsymbol{b}}_t^\top, \tag{198}$$

and let $\mathcal{Q}_t$ be the law obtained by drawing $\boldsymbol{\zeta}$ uniformly on $\mathbb{S}^{M-1}$ and observing

$$\boldsymbol{W} = s_2\boldsymbol{\mathcal{M}}_t(\widetilde{\boldsymbol{z}}) + \sigma\boldsymbol{G}. \tag{199}$$

Then $\mathcal{Q}_{t=0} = \mathcal{Q}_0$, $\mathcal{Q}_{t=\varphi} = \mathcal{Q}_1$, and (179) gives $\varphi \le 4\delta$.

Let q_t be the density of $\mathcal{Q}_t$, and define

$$S_t(\boldsymbol{W}) := \partial_t \log q_t(\boldsymbol{W}), \qquad J_t := \mathbb{E}_{\mathcal{Q}_t} S_t(\boldsymbol{W})^2. \tag{200}$$

By Lemma 25,

$$\mathrm{TV}(\mathcal{Q}_0, \mathcal{Q}_1) \le \frac{1}{2}\int_0^\varphi \sqrt{J_t}\, dt. \tag{201}$$

The following lemma supplies the uniform bound needed in this inequality; its proof is given below with the proof of the covariance-mixture information bound.

Lemma 32. *Uniformly for $t \in [0, \varphi]$,*

$$J_t \le 8\frac{s_2^2\theta^2}{\sigma^2} + 144\frac{s_2^4\theta^2}{M\sigma^4}. \tag{202}$$

Combining (201) and (202) gives

$$\mathrm{TV}^2(\mathcal{Q}_0, \mathcal{Q}_1) \le C\varphi^2 \left(\frac{s_2^2\theta^2}{\sigma^2} + \frac{s_2^4\theta^2}{M\sigma^4}\right). \tag{203}$$

The definition of γ_2 and the high-loading condition imply

$$\frac{c_B}{1 + c_B} s_2^2 \le \gamma_2^2 \le s_2^2. \tag{204}$$

The signal-strength assumption and $M \asymp n$ then give $M\sigma^2/s_2^2 \le C\nu^{-2}$. The assumed bound on δ, (179) and (204), and $\rho_\perp^2 \ge 2\theta$ yield

$$\varphi \le Cc\frac{\sigma^2\sqrt{M}}{s_2^2\theta}. \tag{205}$$

Substitution in (203) gives

$$\mathrm{TV}^2(\mathcal{Q}_0, \mathcal{Q}_1) \le Cc^2 \left(1 + \frac{M\sigma^2}{s_2^2}\right) \le Cc^2(1 + \nu^{-2}). \tag{206}$$

Choosing the numerical constant c sufficiently small gives $\mathrm{TV}(\mathcal{Q}_0, \mathcal{Q}_1) \le 1/4$. Combining (196) with the equality between the joint and View-2 distances at the start of the proof proves

$$\mathrm{TV}(\mathsf{P}_{\boldsymbol{T}_0}, \mathsf{P}_{\boldsymbol{T}_1}) \le \frac{1}{4}.$$

This proves Lemma 30.

G.6 Information bounds for the hidden constructions

G.6.1 Proof of Lemma 31

Let $q_{t,\widetilde{\boldsymbol{z}}}$ denote the conditional density of the observations in (181) for fixed $\widetilde{\boldsymbol{z}}$, and define its score by

$$S_{t,\widetilde{\boldsymbol{z}}}(\boldsymbol{W}) := \partial_t \log q_{t,\widetilde{\boldsymbol{z}}}(\boldsymbol{W}). \tag{207}$$

Because the conditional law is a product of centered Gaussian laws, differentiating its covariance gives

$$S_{t,\widetilde{\boldsymbol{z}}}(\boldsymbol{W}) = \frac{\beta\rho_\perp}{\sigma^2(1+\beta)} \sum_{j=1}^{d_2} (\boldsymbol{a}_t^\top \boldsymbol{W}_j) \left\langle \rho_\perp \boldsymbol{b}_t - \rho\widetilde{\boldsymbol{z}}, \boldsymbol{W}_j \right\rangle. \tag{208}$$

Define $\widehat{\boldsymbol{z}}_t := \mathbb{E}[\widetilde{\boldsymbol{z}} \mid \boldsymbol{W}]$. Fisher's identity and (208) yield

$$S_t(\boldsymbol{W}) = \frac{\beta\rho_\perp}{\sigma^2(1+\beta)} \sum_{j=1}^{d_2} (\boldsymbol{a}_t^\top \boldsymbol{W}_j) \left\langle \rho_\perp \boldsymbol{b}_t - \rho\widehat{\boldsymbol{z}}_t, \boldsymbol{W}_j \right\rangle. \tag{209}$$

To compute its second moment, set

$$X_j := \boldsymbol{a}_t^\top \boldsymbol{W}_j, \qquad \boldsymbol{R}_j := \boldsymbol{W}_j - X_j \boldsymbol{a}_t.$$

Conditional on $\widetilde{\boldsymbol{z}}$, the variables X_j are independent $\mathcal{N}(0, \sigma^2(1+\beta))$ variables whose joint law does not depend on $\widetilde{\boldsymbol{z}}$, and they are independent of $(\boldsymbol{R}_1, \ldots, \boldsymbol{R}_{d_2})$. These independence statements remain true after integrating over $\widetilde{\boldsymbol{z}}$, and

$$\widehat{\boldsymbol{z}}_t = \mathbb{E}[\widetilde{\boldsymbol{z}} \mid \boldsymbol{R}_1, \ldots, \boldsymbol{R}_{d_2}] \,. \tag{210}$$

Moreover, $\rho_\perp \boldsymbol{b}_t - \rho \widehat{\boldsymbol{z}}_t$ is orthogonal to $\boldsymbol{a}_t$. Thus, if

$$c_j \coloneqq \langle \rho_\perp \boldsymbol{b}_t - \rho \widehat{\boldsymbol{z}}_t, \boldsymbol{W}_j \rangle = \langle \rho_\perp \boldsymbol{b}_t - \rho \widehat{\boldsymbol{z}}_t, \boldsymbol{R}_j \rangle \,, \tag{211}$$

then $c_1, \ldots, c_{d_2}$ are fixed conditional on $(\boldsymbol{R}_1, \ldots, \boldsymbol{R}_{d_2})$. The independence and centering of the X_j therefore give

$$\begin{aligned}
&\mathbb{E}\big[S_t(\boldsymbol{W})^2 \,\big|\, \boldsymbol{R}_1, \ldots, \boldsymbol{R}_{d_2}\big] \\
&\quad = \frac{\beta^2 \rho_\perp^2}{\sigma^4 (1+\beta)^2} \mathbb{E}\left[\left(\sum_{j=1}^{d_2} X_j c_j \right)^2 \,\middle|\, \boldsymbol{R}_1, \ldots, \boldsymbol{R}_{d_2} \right] = \frac{\beta^2 \rho_\perp^2}{\sigma^2 (1+\beta)} \sum_{j=1}^{d_2} c_j^2 .
\end{aligned} \tag{212}$$

After taking expectation and using $(x+y)^2 \le 2x^2 + 2y^2$, we obtain

$$J_t \le \frac{2\beta^2 \rho_\perp^4}{\sigma^2(1+\beta)} \mathbb{E} \sum_{j=1}^{d_2} \langle \boldsymbol{b}_t, \boldsymbol{W}_j \rangle^2 + \frac{2\beta^2 \rho_\perp^2}{\sigma^2(1+\beta)} \mathbb{E} \sum_{j=1}^{d_2} \langle \widehat{\boldsymbol{z}}_t, \boldsymbol{W}_j \rangle^2 . \tag{213}$$

Since $\boldsymbol{b}_t^\top \boldsymbol{W}_j$ has variance $\sigma^2(1+\beta\rho^2)$, the first term in (213) is at most $CI\rho_\perp^4$.

We next bound the second term. The identity

$$\sum_{j=1}^{d_2} \langle \widehat{\boldsymbol{z}}_t, \boldsymbol{W}_j \rangle^2 = \widehat{\boldsymbol{z}}_t^\top \left(\sum_{j=1}^{d_2} \boldsymbol{W}_j \boldsymbol{W}_j^\top \right) \widehat{\boldsymbol{z}}_t \tag{214}$$

requires an operator-norm bound for the sample covariance. Define

$$\mathcal{E}_t \coloneqq \left\{ \left\| \sum_{j=1}^{d_2} \boldsymbol{W}_j \boldsymbol{W}_j^\top \right\|_{\mathrm{op}} \le C d_2 \sigma^2 \right\} . \tag{215}$$

We prove below that, uniformly in t,

$$\mathbb{P}(\mathcal{E}_t^c) \le C e^{-c_2 d_2}, \qquad \mathbb{E}\left[\left\| \sum_{j=1}^{d_2} \boldsymbol{W}_j \boldsymbol{W}_j^\top \right\|_{\mathrm{op}} \mathbf{1}_{\mathcal{E}_t^c} \right] \le C d_2 \sigma^2 e^{-c_2 d_2} . \tag{216}$$

It remains to control $\mathbb{E}\|\widehat{\boldsymbol{z}}_t\|_2^2$. For fixed $\widetilde{\boldsymbol{z}}, \widetilde{\boldsymbol{z}}'$, the corresponding conditional Gaussian laws satisfy

$$\mathrm{KL}\big(\mathcal{Q}_{t,\widetilde{\boldsymbol{z}}} \,\|\, \mathcal{Q}_{t,\widetilde{\boldsymbol{z}}'}\big) \le \frac{I\rho_\perp^2}{2} \|\widetilde{\boldsymbol{z}} - \widetilde{\boldsymbol{z}}'\|_2^2 . \tag{217}$$

Indeed, the covariance projectors share $\boldsymbol{a}_t \boldsymbol{a}_t^\top$ and differ only in the projectors onto $\rho \boldsymbol{b}_t + \rho_\perp \widetilde{\boldsymbol{z}}$ and $\rho \boldsymbol{b}_t + \rho_\perp \widetilde{\boldsymbol{z}}'$, so (137) gives this bound. Averaging over an independent copy $\widetilde{\boldsymbol{z}}'$ and using convexity of relative entropy in its second argument yield $\mathrm{I}(\widetilde{\boldsymbol{z}}; \boldsymbol{W}) \le CI\rho_\perp^2$. Therefore Lemma 26 gives

$$\mathbb{E}\|\widehat{\boldsymbol{z}}_t\|_2^2 \le C \frac{I\rho_\perp^2}{M} . \tag{218}$$

On $\mathcal{E}_t$, combine (214) and (218); on $\mathcal{E}_t^c$, use $\|\widehat{\boldsymbol{z}}_t\|_2 \le 1$ and (216). This gives

$$\mathbb{E}\sum_{j=1}^{d_2}\langle\widehat{\boldsymbol{z}}_t, \boldsymbol{W}_j\rangle^2 \le Cd_2\sigma^2\frac{I\rho_\perp^2}{M} + Cd_2\sigma^2 e^{-c_2 d_2}. \tag{219}$$

Substituting this estimate into (213) and using $I = d_2\beta^2/(1+\beta)$ gives the bound in (186). This proves Lemma 31.

Proof of (216). Conditional on $\widetilde{\boldsymbol{z}}$, the vectors $\boldsymbol{W}_1, \dots, \boldsymbol{W}_{d_2}$ are independent centered Gaussian vectors with covariance

$$\boldsymbol{\Sigma}_{t,\widetilde{\boldsymbol{z}}} \coloneqq \sigma^2\{\boldsymbol{I}_{M+2} + \beta\boldsymbol{\Gamma}_t(\widetilde{\boldsymbol{z}})\}, \qquad \|\boldsymbol{\Sigma}_{t,\widetilde{\boldsymbol{z}}}\|_{\mathrm{op}} \le 2\sigma^2. \tag{220}$$

The norm bound follows because $\boldsymbol{\Gamma}_t(\widetilde{\boldsymbol{z}})$ is an orthogonal projector and $\beta \le 1$. If $\boldsymbol{G} \in \mathbb{R}^{(M+2)\times d_2}$ has independent $\mathcal{N}(0, d_2^{-1})$ entries, then, conditionally on $\widetilde{\boldsymbol{z}}$,

$$[\boldsymbol{W}_1, \dots, \boldsymbol{W}_{d_2}] \overset{\mathrm{d}}{=} \sqrt{d_2}\,\boldsymbol{\Sigma}_{t,\widetilde{\boldsymbol{z}}}^{1/2}\boldsymbol{G}, \qquad \left\|\sum_{j=1}^{d_2}\boldsymbol{W}_j\boldsymbol{W}_j^\top\right\|_{\mathrm{op}} \le 2d_2\sigma^2\left\|\boldsymbol{G}\boldsymbol{G}^\top\right\|_{\mathrm{op}}. \tag{221}$$

By (185), $d_2 \ge I \ge c_1\nu^2 M$. Applying (52) with $u = M+2$ and $b = d_2$ therefore yields, uniformly in $\widetilde{\boldsymbol{z}}$,

$$\mathbb{P}\left(\left\|\sum_{j=1}^{d_2}\boldsymbol{W}_j\boldsymbol{W}_j^\top\right\|_{\mathrm{op}} > Cd_2\sigma^2 \,\middle|\, \widetilde{\boldsymbol{z}}\right) \le Ce^{-c_2 d_2}. \tag{222}$$

Averaging over $\widetilde{\boldsymbol{z}}$ proves the first inequality in (216). For the second, combine (221), Cauchy–Schwarz, and (54) to obtain, uniformly in $\widetilde{\boldsymbol{z}}$,

$$\mathbb{E}\left[\left\|\sum_{j=1}^{d_2}\boldsymbol{W}_j\boldsymbol{W}_j^\top\right\|_{\mathrm{op}}\mathbf{1}_{\mathcal{E}_t^c}\,\middle|\,\widetilde{\boldsymbol{z}}\right] \le Cd_2\sigma^2\|\boldsymbol{G}\boldsymbol{G}^\top\|_{L^2}\mathbb{P}(\mathcal{E}_t^c \mid \widetilde{\boldsymbol{z}})^{1/2} \le Cd_2\sigma^2 e^{-c_2 d_2}.$$

Averaging over $\widetilde{\boldsymbol{z}}$ proves the second inequality in (216).

G.6.2 Proof of Lemma 32

For fixed $\widetilde{\boldsymbol{z}}$, the law in (199) is a Gaussian location model with covariance $\sigma^2\boldsymbol{I}$. Since $\dot{\boldsymbol{a}}_t = \boldsymbol{b}_t$ and $\dot{\boldsymbol{b}}_t = -\boldsymbol{a}_t$, with the same identities for the barred vectors, differentiating (198) gives

$$\dot{\boldsymbol{\mathcal{M}}}_t(\widetilde{\boldsymbol{z}}) = \{(1-\rho)\boldsymbol{b}_t - \rho_\perp\widetilde{\boldsymbol{z}}\}\overline{\boldsymbol{a}}_t^\top + (1-\rho)\boldsymbol{a}_t\overline{\boldsymbol{b}}_t^\top. \tag{223}$$

The terms independent of ρ cancel because the left and right coordinates undergo the same rotation. The remaining terms are proportional to $1-\rho = 2\theta$ or $\rho_\perp$.

Let $S_{t,\widetilde{\boldsymbol{z}}}$ denote the score conditional on $\widetilde{\boldsymbol{z}}$. The Gaussian location score is

$$S_{t,\widetilde{\boldsymbol{z}}}(\boldsymbol{W}) = \frac{s_2}{\sigma^2}\left\langle\boldsymbol{W} - s_2\boldsymbol{\mathcal{M}}_t(\widetilde{\boldsymbol{z}}), \dot{\boldsymbol{\mathcal{M}}}_t(\widetilde{\boldsymbol{z}})\right\rangle_{\mathrm{F}}.$$

Define

$$\boldsymbol{\varepsilon}_t \coloneqq \boldsymbol{W}\overline{\boldsymbol{a}}_t - s_2\boldsymbol{a}_t, \qquad \xi_t \coloneqq \langle\boldsymbol{W}\overline{\boldsymbol{b}}_t, \boldsymbol{a}_t\rangle. \tag{224}$$

Using (223) and the orthogonality of $\boldsymbol{a}_t$ to $\boldsymbol{b}_t$ and $\widetilde{\boldsymbol{z}}$ gives

$$S_{t,\widetilde{\boldsymbol{z}}}(\boldsymbol{W}) = \frac{s_2}{\sigma^2}\left[\langle \boldsymbol{\varepsilon}_t, (1-\rho)\boldsymbol{b}_t - \rho_\perp \widetilde{\boldsymbol{z}}\rangle + (1-\rho)\xi_t\right]. \tag{225}$$

Set $\widehat{\boldsymbol{z}}_t \coloneqq \mathbb{E}[\widetilde{\boldsymbol{z}} \mid \boldsymbol{W}]$. Fisher's identity averages (225) over the conditional distribution of $\widetilde{\boldsymbol{z}}$, so

$$S_t(\boldsymbol{W}) = \frac{s_2}{\sigma^2}\left[\langle \boldsymbol{\varepsilon}_t, (1-\rho)\boldsymbol{b}_t - \rho_\perp \widehat{\boldsymbol{z}}_t\rangle + (1-\rho)\xi_t\right]. \tag{226}$$

Rotating the observed columns by $(\overline{\boldsymbol{a}}_t, \overline{\boldsymbol{b}}_t)$ gives, conditionally on $\widetilde{\boldsymbol{z}}$,

$$\boldsymbol{W}\overline{\boldsymbol{a}}_t = s_2\boldsymbol{a}_t + \sigma\boldsymbol{g}_a, \qquad \boldsymbol{W}\overline{\boldsymbol{b}}_t = s_2(\rho\boldsymbol{b}_t + \rho_\perp\widetilde{\boldsymbol{z}}) + \sigma\boldsymbol{g}_b, \tag{227}$$

where $\boldsymbol{g}_a$ and $\boldsymbol{g}_b$ are independent standard Gaussian vectors. Only the last M coordinates of the second vector in (227) depend on $\widetilde{\boldsymbol{z}}$; they have the Gaussian mean-channel form with signal strength $s_2\rho_\perp$ and noise level σ. Thus Lemma 26 gives

$$\mathbb{E}\|\widehat{\boldsymbol{z}}_t\|_2^2 \le 9\frac{s_2^2\rho_\perp^2}{M\sigma^2}. \tag{228}$$

Conditional on these last M coordinates, $\widehat{\boldsymbol{z}}_t$ is fixed, $\boldsymbol{\varepsilon}_t = \sigma\boldsymbol{g}_a$, and $\xi_t = \sigma\langle \boldsymbol{g}_b, \boldsymbol{a}_t\rangle$. The latter two Gaussian quantities are centered and independent of one another and of the conditioning variables. Since $\widehat{\boldsymbol{z}}_t$ is orthogonal to $\boldsymbol{b}_t$, taking the conditional second moment in (226), then averaging and applying (228), gives

$$J_t = \frac{s_2^2}{\sigma^2}\left\{2(1-\rho)^2 + \rho_\perp^2\mathbb{E}\|\widehat{\boldsymbol{z}}_t\|_2^2\right\} \le 8\frac{s_2^2\theta^2}{\sigma^2} + 144\frac{s_2^4\theta^2}{M\sigma^4},$$

where we used $1-\rho = 2\theta$ and $\rho_\perp^2 \le 4\theta$. This is (202) and completes the proof.