Spectral Thresholds in Correlated Spiked Models and Fundamental Limits of Partial Least Squares

Pierre Mergny
Information, Learning and Physics Laboratory
École Polytechnique Fédérale de Lausanne
1015 Lausanne, Suisse

Lenka Zdeborová
Statistical Physics of Computation Laboratory
École Polytechnique Fédérale de Lausanne
1015 Lausanne, Suisse

Abstract

We provide a rigorous random matrix theory analysis of spiked cross-covariance models where the signals across two high-dimensional data channels are partially aligned. These models are motivated by multi-modal learning and form the standard generative setting underlying Partial Least Squares (PLS), a widely used yet theoretically underdeveloped method. We show that the leading singular values of the sample cross-covariance matrix undergo a Baik–Ben Arous–Péché (BBP)-type phase transition, and we characterize the precise thresholds for the emergence of informative components. Our results yield the first sharp asymptotic description of the signal recovery capabilities of PLS in this setting, revealing a fundamental performance gap between PLS and the Bayes-optimal estimator. In particular, we identify the SNR and correlation regimes where PLS fails to recover any signal, despite detectability being possible in principle. These findings clarify the theoretical limits of PLS and provide guidance for the design of reliable multi-modal inference methods in high dimensions.

1 INTRODUCTION

The challenge of recovering a low-dimensional structure hidden in a high-dimensional noisy output is widespread in statistics, probability, and machine learning. Spiked random matrix models Ben Arous et al. (2005); Johnstone and Lu (2009); Lelarge and Miolane (2017); Zou et al. (2006) have attracted significant attention as a simple yet rich framework for studying this class of problems, especially within the toolbox of random matrix theory (RMT) Anderson et al. (2010); Tao (2012); Potters and Bouchaud (2020), which provides asymptotic characterizations of spectral properties in high dimensions. On the other hand, in more complex data scenarios, one often has access to multiple related outputs. Multi-modal learning Ngiam et al. (2011); Ramachandram and Taylor (2017), a central paradigm in modern data analysis, seeks to leverage the joint information contained in such datasets to improve inference or prediction. This includes, for instance, settings where signals are observed across different modalities or sensors.

Popular classical approaches such as Canonical Correlation Analysis (CCA) Thompson (2000); Guo and Wu (2019); Asendorf and Nadakuditi (2017) and Partial Least Squares (PLS) Wold (1975, 1983); Wold et al. (2001); Wegelin et al. (2000); Pirouz (2006) rely on spectral methods to uncover such cross-dependencies and have been widely applied across various scientific and engineering domains. While CCA has been extensively analyzed in the literature Yang (2022a,b); Guo and Wu (2019); Bykhovskaya and Gorin (2023); Ma and Yang (2023); Bao et al. (2019), notably through the lens of RMT, methods such as PLS, which operate directly on the (empirical) cross-covariance matrix, despite their widespread use, remain less well understood from a theoretical point of view.

To address this issue, we focus on a setting involving two *correlated spiked matrix models* (or *"channels"*), defined as follows:

$$\tilde{\boldsymbol{X}} = \boldsymbol{X} + \sum_{k=1}^{r} \sqrt{\lambda_{x,k}} \cdot \mathbf{u}_{x,k}^{\star} \left(\mathbf{v}_{x,k}^{\star}\right)^T \in \mathbb{R}^{n \times d_x} , \quad (1.1)$$

Proceedings of the 29th International Conference on Artificial Intelligence and Statistics (AISTATS) 2026, Tangier, Morocco. PMLR: Volume 300. Copyright 2026 by the author(s).

$$\tilde{\boldsymbol{Y}} = \boldsymbol{Y} + \sum_{k=1}^{r} \sqrt{\lambda_{y,k}} \cdot \mathbf{u}^{\star}_{y,k} \left(\mathbf{v}^{\star}_{y,k}\right)^T \in \mathbb{R}^{n \times d_y} \,. \tag{1.2}$$

The precise description and assumptions of this model are provided in more detail in the following paragraph. For now, one may consider the matrices $\boldsymbol{X}$ and $\boldsymbol{Y}$ as *noise matrices*. The values $\lambda_{x,k}, \lambda_{y,k} \geq 0$ of the low-rank matrices serve as signal-to-noise ratios (SNR) for the components $\{\mathbf{u}^{\star}_{x,k}, \mathbf{v}^{\star}_{x,k}, \mathbf{u}^{\star}_{y,k}, \mathbf{v}^{\star}_{y,k}\}_{k=1}^{r}$ that one aims to infer. The term *correlated* refers to the assumption that the unit-norm signals in the shared dimension of the two sources exhibit partial alignment in the high-dimensional regime. Specifically, we consider that $\langle \mathbf{u}^{\star}_{x,k}, \mathbf{u}^{\star}_{y,k} \rangle \approx \rho_k \in (-1, 1)$ for some fixed positive constants ρ_k.

Our objective is to provide a quantitative analysis of the Partial Least Squares (PLS) methods and derive its performance analytically to ease and inform the comparison with other estimators. The PLS methods are based on estimating the subspace spanned by the signals based on the singular vectors of the **sample cross-covariance matrices**:

$$\tilde{\boldsymbol{S}} := \tilde{\boldsymbol{X}}^T \tilde{\boldsymbol{Y}} \in \mathbb{R}^{d_x \times d_y} \,, \tag{1.3}$$

and thus a high-dimensional setting where

$$n, d_x, d_y \to \infty \text{ such that } \frac{d_x}{n} \to \alpha_x, \frac{d_y}{n} \to \alpha_y \,,$$

for some positive constants α_x, α_y and without loss of generality, we set $\alpha_x \geq \alpha_y$.

This model has recently been considered as a natural toy model for describing *multi-modal learning* in high-dimensional settings, see Abdelaleem et al. (2023); Keup and Zdeborová (2024). The key intuition is that stronger alignment between the correlated signal vectors may facilitate the recovery of the other components. Authors of Keup and Zdeborová (2024) evaluated the detectability threshold of the Bayes-optimal estimator and compared it empirically with the performance of the PLS and CCA. While for the unimodal spiked matrix model, the detectability threshold of the Bayes-optimal estimator coincides with the BBP threshold for the natural spectral methods, the authors of Keup and Zdeborová (2024) pointed out based on numerical experiments that for the above correlated spiked matrix model the threshold of both the PLS and CCA are suboptimal. This observation is interesting in particular in the view of the contrast with the unimodal case.

Main results – In this work, we provide the high-dimensional limits of the spectral method based on the sample-cross covariance matrix given by Eq. (1.3). Specifically,

- We show that its leading singular values undergo a BBP-like phase transition depending on the value of the SNR, correlation and aspect ratios.
- We obtained a complete characterization of the associated overlap with the hidden signal, and their phase transition, further generalizing the BBP results to this cross-product setting.
- We apply this result to obtain the fundamental limits of Partial Least Square (PLS) methods.
- We discuss the comparison between the PLS, CCA and Bayes-optimal thresholds unveiling the somewhat surprising sub-optimality of these spectral approaches even on a model as simple as the one considered here.

Related works – Many variants of spiked matrix models have been investigated in the high-dimensional regime; see Péché (2014); Ben Arous et al. (2005); Paul (2007); Perry et al. (2018); Benaych-Georges and Nadakuditi (2011); Guionnet et al. (2023) for general references, and in particular Loubaton and Vallet (2011); Capitaine (2018); Benaych-Georges and Nadakuditi (2012) for the case of spectral analysis in a *single-channel* setting (also known as the *information-plus-noise spiked model*) —namely, the study of the asymptotic behavior of the singular values and vectors of $\tilde{\boldsymbol{X}}$ (or equivalently, of $\tilde{\boldsymbol{Y}}$). The properties of the spectrum of sample cross-covariance matrices *without any low-rank perturbation* are also well understood, see Burda et al. (2010); Akemann et al. (2013). To the best knowledge of the authors, this work is the first one to characterize analytically the spectral properties of these cross-covariance matrices *with the inclusion of such spikes*, answering in particular the question left open in Keup and Zdeborová (2024) where such models have been introduced as a multi-modal toy model and studied from a Bayes-Optimal (BO) point of view. As a direct application of our work, we obtain the fundamental limit of the performance of Partial Least Squares (PLS) methods. Despite the lack of previous theoretical guarantees (see Keup and Zdeborová (2024) and Abdelaleem et al. (2023) for an exploratory empirical study), PLS methods are widely used across a variety of fields Hulland (1999); Krishnan et al. (2011). Closely related to PLS, CCA (sometimes refereed to as PLS "mode B") which operates on the MANOVA matrix $(\tilde{\boldsymbol{X}}^T\tilde{\boldsymbol{X}})^{-1/2}\tilde{\boldsymbol{X}}^T\tilde{\boldsymbol{Y}}(\tilde{\boldsymbol{Y}}^T\tilde{\boldsymbol{Y}})^{-1/2}$ rather than the cross covariance matrix $\tilde{\boldsymbol{S}}$ has been recently studied from a theoretical point of view in Guo and Wu (2019); Yang (2022a,b); Bykhovskaya and Gorin (2023) and our work compares the phase diagram of PLS methods with both CCA and BO methods. Similarly, Sergazinov et al. (2025); Lebeau et al.

(2024); Ma and Nandy (2023); Yang et al. (2024); Duranthon and Zdeborová (2023); Tabanelli et al. (2025); Pacco and Ros (2023); Qi and Decurninge (2025) investigated the information theoretical limit for the inference of multiple spiked models with correlated signals, albeit in a different setting. Finally, we mention the related the work Benaych-Georges et al. (2023); Attal and Allez (2025) which analyzes cross-covariance models in a regime where the number of spikes grows linearly with dimension, leading to fundamentally different spectral behavior.

Notations – In the following, generic vectors are written in bold lowercase (e.g. $\mathbf{v}$) while matrices are written in bold uppercase (e.g. $\mathbf{M}$). To further emphasize on the random nature of vectors (or matrices), we will use italics when needed (e.g. $\boldsymbol{v}$ and $\boldsymbol{M}$ respectively). In a similar way, we use the *tilde*-notations (e.g. $\tilde{\boldsymbol{v}}, \tilde{\boldsymbol{M}}$) to highlight the dependency in the low-rank signal components. For $k \in \mathbb{N}$, $[\![k]\!] = \{1, \dots, k\}$. $\mathbb{C}_\pm = \{z \in \mathbb{C} | \pm \mathfrak{Im}(z) > 0\}$ correspond to the upper and lower half complex plane. For $\mathbf{M} \in \mathbb{R}^{n\times m}$ with $n \le m$, we denote by $\sigma_1(\mathbf{M}) \ge \dots \ge \sigma_n(\mathbf{M}) \ge 0$ its singular values and $\mu_{\mathbf{M}} := \frac{1}{n}\sum_{i=1}^n \delta_{\sigma_i(\mathbf{M})^2}$ the associated empirical *squared* singular value distribution. We say that a sequence $(X_n)_n$ of random variables converges exponentially fast to a value x if for all $\epsilon > 0$, there exist $C, c > 0$ such that $\mathbb{P}(|X_n - x| > \epsilon) \le C\mathrm{e}^{-cn}$. $\mathcal{P}_c(\mathbb{R})$ denotes the set of compactly supported measures on $\mathbb{R}$. For a sequence $\mu_n \in \mathcal{P}_c(\mathbb{R})$, we denote $\mu_n \to \mu$ the (almost-sure) weak convergence. We denote by $\eta(\mathbf{A}) := \begin{pmatrix} 0 & \mathbf{A} \\ \mathbf{A}^T & 0 \end{pmatrix}$ the hermitian embedding of $\mathbf{A}$.

2 ASYMPTOTICS OF THE TOP SINGULAR VALUES AND OVERLAPS

2.1 Assumptions

In what follows, we provide a more detailed description of the assumptions underlying our model for the cross-covariance matrix of Eq. (1.3).

Assumption (A1). *The matrices $\boldsymbol{X} = (X_{ij})_{i,j}$ and $\boldsymbol{Y} = (Y_{ij})_{i,j}$ are independent with entries given by $\sqrt{d_x}\, X_{ij} \overset{\text{i.i.d}}{\sim} \mathsf{N}(0,1)$ and $\sqrt{d_y}\, Y_{ij} \overset{\text{i.i.d}}{\sim} \mathsf{N}(0,1)$.*

Although our analysis is carried out under this Gaussian assumption, universality results in RMT suggest that our conclusions should extend beyond this framework, typically requiring only the entries to satisfy a log-Sobolev inequality, see for example Baik and Silverstein (2006) for the within channel covariance matrix. We leave a rigorous investigation of this extension for future work.

Assumption (A2). *For $z \in \{x, y\}$ and any $k \in [\![r]\!]$, as $n \to \infty$ the low-rank signal components satisfy $\|\mathbf{u}^\star_{z,k}\| \xrightarrow[n\to\infty]{\text{a.s.}} 1$, $\|\mathbf{v}^\star_{z,k}\| \xrightarrow[n\to\infty]{\text{a.s.}} 1$ and $\langle \mathbf{u}^\star_{x,k}, \mathbf{u}^\star_{y,k}\rangle \xrightarrow[n\to\infty]{\text{a.s.}} \rho_k$ exponentially fast and any other inner product converges to zero.*

Assumption (A2) covers the case where the signal components are drawn independently from $\sqrt{d_x}\,\mathbf{v}^\star_{x,k} \sim \mathsf{P}_x^{\otimes d_x}$, $\sqrt{d_y}\,\mathbf{v}^\star_{y,k} \sim \mathsf{P}_y^{\otimes d_y}$ and $(\mathbf{u}^\star_{x,k}, \mathbf{u}^\star_{y,k}) \sim \mathsf{Q}_{(\rho_k)}^{\otimes n}$, where P_x, P_y (resp. $\mathsf{Q}_{(\rho_k)}$) are probability measures on $\mathbb{R}$ (resp. on $\mathbb{R}^2$) with mean zero, variance one (resp. covariance $\begin{pmatrix} 1 & \rho_k \\ \rho_k & 1 \end{pmatrix}$) satisfying standard log-Sobolev inequalities, as well as the cases where the signal components are deterministic vectors. We point out that Assumption (A2) is motivated from a technical point of view by wanting the matrix $(\mathbf{u}^\star_{x,1}, \dots, \mathbf{u}^\star_{1,r})^T(\mathbf{v}^\star_{y,1}, \dots, \mathbf{v}^\star_{y,r})$ to be diagonal. In principle, the analysis developed in this paper could be extended beyond this setting, at the expense of greater technical complexity.

2.2 Preliminary Definitions

For $\mu \in \mathcal{P}_c(\mathbb{R}_+)$ and $z \in \mathbb{C} \setminus \mathrm{Supp}(\mu)$, we introduce the *$T$-(moment generating) transform* as:

$$t_\mu(z) := \int \frac{\lambda}{z - \lambda} \mathrm{d}\mu(\lambda)\,, \tag{2.1}$$

which uniquely characterizes $\mu \in \mathcal{P}_c(\mathbb{R}_+)$, see Potters and Bouchaud (2020).

Next, for $z \in \mathbb{C}_-$ we introduce the cubic polynomial $P \in \mathbb{R}_3[X]$ given by

$$\begin{aligned} P_{(\alpha_x,\alpha_y)}(X, z) :=& 1 + (1 + \alpha_x + \alpha_y - z)X \\ &+ (\alpha_x + \alpha_y + \alpha_x\alpha_y)X^2 + \alpha_x\alpha_y X^3\,. \end{aligned} \tag{2.2}$$

We will also consider the following two other cubic polynomials $Q, R \in \mathbb{R}_3[X]$ as

$$Q_{(\alpha_x,\alpha_y)}(X) := 1 - (\alpha_x\alpha_y + \alpha_x + \alpha_y)X^2 - 2\alpha_x\alpha_y X^3\,, \tag{2.3}$$

and

$$\begin{aligned} R_{(\lambda_x,\lambda_y,\rho)}(X) :=& 1 + (1 - \rho^2\lambda_x\lambda_y - \lambda_x - \lambda_y)X \\ &+ (\lambda_x\lambda_y - \lambda_x - \lambda_y)X^2 + \lambda_x\lambda_y X^3\,; \end{aligned} \tag{2.4}$$

and will use the following property.

Lemma 2.1 (Roots of the cubic polynomials)**.** *The polynomial $Q_{(\alpha_x,\alpha_y)}$ has exactly one positive root which*

we denote by $\tau^+ \equiv \tau^+(\alpha_x, \alpha_y)$. *Similarly, the polynomial* $R_{(\lambda_x,\lambda_y,\rho)}$ *has exactly two (counted with multiplicity) positive roots, which we denote by* $\mathrm{r}^+ \equiv \mathrm{r}^+(\lambda_x, \lambda_y, \rho) \geq \mathrm{r}^- \equiv \mathrm{r}^-(\lambda_x, \lambda_y, \rho) > 0$. *Furthermore,* r^- *(resp.* r^+*) is a decreasing (respectively increasing) function of* ρ^2.

Proof. See Appendix B. □

As it will appear clear in the next sections, these roots (and more precisely their inverse) govern the BBP-like phase transition of the model: the terms $1/\mathrm{r}^\pm(\lambda_x, \lambda_y, \rho)$ can be viewed as effective signal-to-noise ratios for the spiked cross-covariance matrix, while $1/\tau^+$ plays the role of a limiting threshold for the appearance of spikes in the spectrum.

For cubic polynomials, the roots can be expressed explicitly using Cardano's formula. However, the resulting expressions are quite cumbersome, and the roots are more conveniently computed numerically using standard root-finding algorithms. In the following, to ease notations we write simply $\mathrm{r}_k^\pm \equiv \mathrm{r}^\pm(\lambda_{x,k}, \lambda_{y,k}, \rho_k)$.

2.3 Bulk Distribution

We recall the result regarding the behavior of the *bulk* of the spectrum. To this end, consider the constant ς_+ defined as

$$\varsigma_+ \equiv \varsigma_+(\alpha_x, \alpha_y) := \sqrt{\frac{\left(1+\tau^+\right)\left(1+\alpha_x\tau^+\right)\left(1+\alpha_y\tau^+\right)}{\tau^+}}\,, \tag{2.5}$$

which corresponds to the solution (in z) of the equation $P_{(\alpha_x,\alpha_y)}(\tau^+/\alpha_x, z^2) = 0$. For notational convenience, we present the proposition under the assumption $\alpha_x > 1$, and refer the reader to Appendix C for the complementary case.

Proposition 2.1 (Bulk Distribution, Swain et al. (2025); Burda et al. (2010))**.** *Under Assumptions (A1)-(A2) with* $\tilde{\boldsymbol{S}}$ *given by Eq.(1.3),*

(i) we have $\mu_{\tilde{\boldsymbol{S}}} \to \mu_{(\alpha_x,\alpha_y)}$ *whose* T*-transform* $t(z) \equiv t_{\mu_{(\alpha_x,\alpha_y)}}(z)$ *defined by Eq. (2.1) is given for any* $z \in \mathbb{C}_-$ *as the unique solution in* $\mathbb{C}_+$ *of*

$$P_{(\alpha_x,\alpha_y)}\left(\frac{t(z)}{\alpha_x}, z\right) = 0\,. \tag{2.6}$$

(ii) Furthermore, in the absence of planted signals ($\lambda_{x,k} = \lambda_{y,k} = 0$ *for all* $k \in [\![r]\!]$)*, the top singular value converges to the rightmost edge of the distribution* $\mu_{(\alpha_x,\alpha_y)}$*, given by the constant* ς_+ *of Eq. (2.5) and* $\lim_{\epsilon\searrow 0} t(\varsigma_+^2 + \epsilon) = \tau^+$ *defined in Lemma 2.1.*

Proof. This follows from free probability results, see Mingo and Speicher (2017) for an introduction to the topic. For completeness, the reader may find the proof in Appendix C. □

Fig. 4 provides an illustration comparing the theoretical bulk of the spectrum with empirical histograms of the eigenvalues for large but finite dimensions.

2.4 Asymptotics of the Top Singular Values

Our first result characterizes how Part-(ii) of this proposition is modified by the presence of low-rank signals leading to a BBP-like phase transition phenomenon. To this end, we introduce the non-increasing function $b : \mathbb{R}^+ \ni r \mapsto b(r) \in \mathbb{R}^+$ defined by

$$b(r) := \begin{cases} \sqrt{\frac{\left(1+r\right)\left(1+\alpha_x r\right)\left(1+\alpha_y r\right)}{r}} & \text{if } 1/r \geq 1/\tau^+\,, \\ \varsigma_+ & \text{otherwise}\,. \end{cases} \tag{2.7}$$

which appears in the following result.

Theorem 2.1 (Phase Transition for the Singular Values)**.** *Under Assumptions (A1)-(A2), the top* $2r$ *singular values of the matrix* $\tilde{\boldsymbol{S}}$ *given by Eq. (1.3) are given asymptotically (up to re-ordering) by*

$$\{\sigma_1(\tilde{\boldsymbol{S}}), \ldots, \sigma_{2r}(\tilde{\boldsymbol{S}})\} \xrightarrow[n\to\infty]{\text{a.s.}} \{b(\mathrm{r}_k^+)\,,\ b(\mathrm{r}_k^-)\}_{k=1}^r\,, \tag{2.8}$$

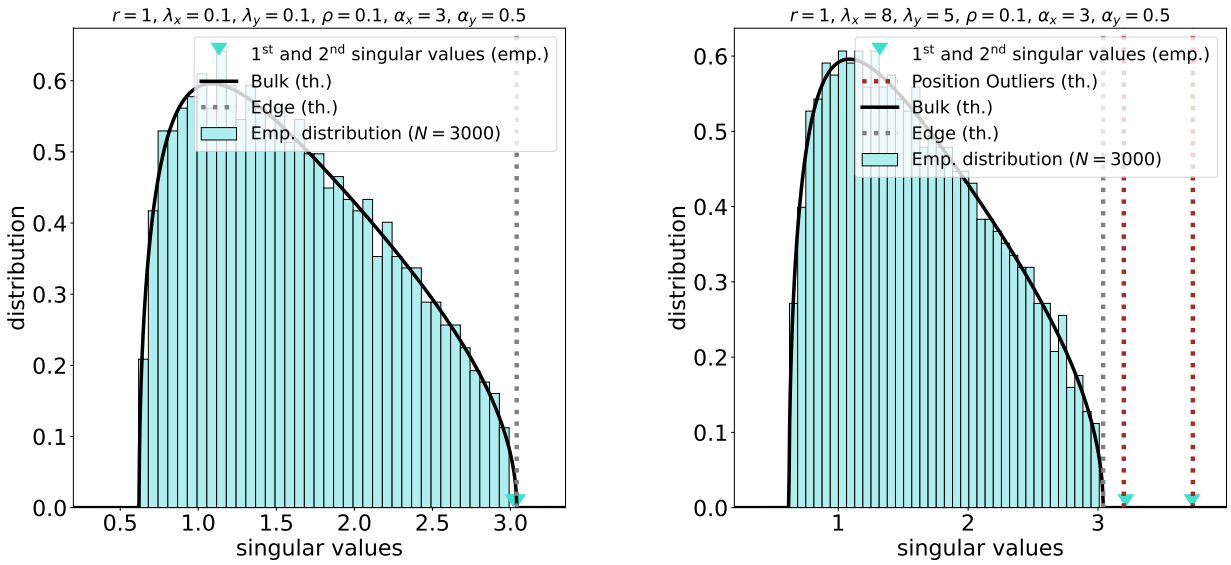
where the $\mathrm{r}_k^\pm \equiv \mathrm{r}^\pm(\lambda_{x,k}, \lambda_{y,k}, \rho_k)$ *are given in Lemma 2.1.*

Proof. The proof of this Theorem is detailed in Sec. 4.1. □

Thm. 2.1 identifies the threshold at which the first outlier separates from the bulk, thus distinguishing the spiked model from the purely noisy one, as the solution to an implicit equation involving the roots of the two cubic polynomials described in Lemma 2.1. Specifically, the model is said to be at criticality, meaning that the planted signal is just strong enough to produce a detectable spectral spike, when

$$\min_{k\in[\![r]\!]}\{\mathrm{r}^-(\lambda_{x,k}, \lambda_{y,k}, \rho_k)\} = \tau^+(\alpha_x, \alpha_y)\,. \tag{2.9}$$

By Lemma 2.1, r^- is a decreasing function of ρ^2, the squared correlation between the latent variables, and $b(\cdot)$ is a non-increasing function of its argument. Therefore if we denote by k_0 the argmin in Eq. (2.9), as the squared correlation increases while keeping other parameters fixed, the leading outlier at $b(\mathrm{r}_{k_0}^-)$ moves further away from the edge of the bulk, thereby



(a) (b)

Figure 1: Singular values of the spiked cross-covariance (1.3) with one spike ($r = 1$) in each channel, for high (b) and low (a) values of the signal components (λ_x, λ_y) while the other parameters $(\alpha_x, \alpha_y, \rho)$ are the same. The figures indicate the presence of outliers outside the bulk for high values while for low values, the top two singular values stick to the edge of the distribution defined in Prop. 2.1. The theoretical positions of the outliers for the right panel follow from our main result Thm. 2.1.

improving detectability. Note that, conversely, by Lemma 2.1 the outlier at $b(r^+_{k_0})$ moves towards the rightmost edge. In Fig. 2, we have illustrated the theoretical prediction given by Thm. 2.1 against empirical simulations at finite but large dimensions, showing good agreement.

The phase diagram in the (λ_x, λ_y) for the spectral methods based on the cross-covariance matrix (and hence of PSL as well) is illustrated in Fig. 3. In the same figures, we have illustrated the threshold for PCA on the single-channel spiked models, the CCA matrix Ma and Yang (2023); Bykhovskaya and Gorin (2023) and the Bayes approach Keup and Zdeborová (2024). In particular, although the parameter region for detectability expands as the squared correlation ρ^2 increases, we note that for small but non-zero values of ρ^2, it is possible to enter a regime where detection is achievable via PCA on the *single-channel*, but *not* on the cross-covariance matrix. This raises questions about the practical advantage of using the cross-covariance matrix and PLS approach in such scenarios, despite their widespread use.

2.5 Asymptotics of the Top Singular Vectors

We now turn to our second main result, which analyzes the overlaps between the singular vectors and the signal components.

Theorem 2.2 (Phase Transition for the Singular Vectors)**.** *For $k \in [\![r]\!]$, let $(\tilde{\boldsymbol{u}}^+_k, \tilde{\boldsymbol{v}}^+_k)$ and $(\tilde{\boldsymbol{u}}^-_k, \tilde{\boldsymbol{v}}^-_k)$ denote the unit-norm singular vectors of $\tilde{\boldsymbol{S}}$ associated with the two singular values whose limiting values are given by $b(\mathrm{r}^+_k)$ and $b(\mathrm{r}^-_k)$ respectively as stated in Thm. 2.1. Then for any $k \in [\![r]\!]$ we have :*

$$\langle \tilde{\boldsymbol{u}}^\pm_k, \mathbf{v}^\star_{x,k} \rangle^2 \xrightarrow[n\to\infty]{\text{a.s.}} \mathrm{m}^\pm_{x,k} \quad , \quad \langle \tilde{\boldsymbol{v}}^\pm_k, \mathbf{v}^\star_{y,k} \rangle^2 \xrightarrow[n\to\infty]{\text{a.s.}} \mathrm{m}^\pm_{y,k},$$

where for $z \in \{x, y\}$, $\mathrm{m}^\pm_{z,k}$ is positive if $1/\mathrm{r}^\pm_k \geq 1/\tau^+$ and null otherwise, with expression given in App. H.

Proof. The proof of this Theorem is detailed in Sec. 4.2. □

Remark (Suboptimal Rotation)**.** This theorem implies that when $1/\mathrm{r}^+_k \geq 1/\tau^+$, the two left (resp. right) singular vectors associated with $b(\mathrm{r}^-_k)$ and $b(\mathrm{r}^+_k)$ **both** correlate with $\mathbf{v}^\star_{x,k}$ (resp. with $\mathbf{v}^\star_{y,k}$). In other words, a more accurate estimation can be achieved by appropriately *rotating* these two vectors. The optimal angle of rotation depends on the parameters of the model and its derivation is given in Appendix I.

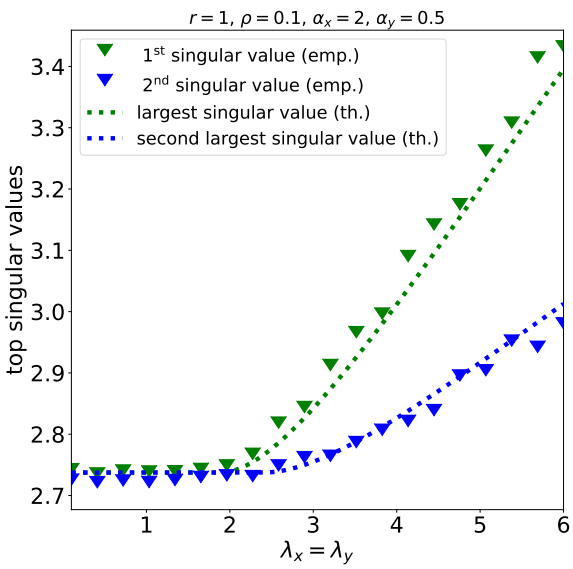


Figure 2: Empirical and theoretical values of the top two singular values of the spiked cross-covariance matrix (1.3) with one spike ($r = 1$), plotted as functions of the signal strengths $\lambda_x = \lambda_y$. All other parameters (α_x, α_y, ρ) are fixed. Each empirical data point is obtained as an average over 10 sample points.

3 APPLICATIONS TO THE FUNDAMENTAL LIMITS OF PLS METHODS

Following the description in Keup and Zdeborová (2024) (see also Wegelin et al. (2000)), the canonical (or "mode-A") PLS algorithm with r_0 steps, constructs rank-r_0 approximations of $\tilde{\boldsymbol{X}}$ and $\tilde{\boldsymbol{Y}}$ in order to estimate the underlying low-rank structures. This is achieved by iterating the following five steps r_0 times:

(1) compute the top left and right singular vectors $\tilde{\boldsymbol{u}}$, $\tilde{\boldsymbol{v}}$ associated with the top singular value of $\tilde{\boldsymbol{S}}$;

(2) estimate the left ('u') components by $\widehat{\boldsymbol{u}}_x^{(\mathrm{PLS})} = \tilde{\boldsymbol{X}}\tilde{\boldsymbol{u}}$ and $\widehat{\boldsymbol{u}}_y^{(\mathrm{PLS})} = \tilde{\boldsymbol{Y}}\tilde{\boldsymbol{v}}$;

(3) refine the estimates of the right ('v') components by $\widehat{\boldsymbol{v}}_x^{(\mathrm{PLS})} = \|\widehat{\boldsymbol{u}}_x^{(\mathrm{PLS})}\|^{-2}\tilde{\boldsymbol{X}}^T\tilde{\boldsymbol{X}}\tilde{\boldsymbol{u}}$ and $\widehat{\boldsymbol{v}}_y^{(\mathrm{PLS})} = \|\widehat{\boldsymbol{u}}_y^{(\mathrm{PLS})}\|^{-2}\tilde{\boldsymbol{Y}}^T\tilde{\boldsymbol{Y}}\tilde{\boldsymbol{v}}$.

(4) subtract the resulting rank-one approximations from each data matrix individually, updating $\tilde{\boldsymbol{X}} \leftarrow \tilde{\boldsymbol{X}} - \widehat{\boldsymbol{u}}_x^{(\mathrm{PLS})}(\widehat{\boldsymbol{v}}_x^{(\mathrm{PLS})})^T$ and $\tilde{\boldsymbol{Y}} \leftarrow \tilde{\boldsymbol{Y}} - \widehat{\boldsymbol{u}}_y^{(\mathrm{PLS})}(\widehat{\boldsymbol{v}}_y^{(\mathrm{PLS})})^T$;

(5) repeat from step (1).

A simplified variant, known as **PLS-SVD**, omits step (3), reducing PLS to a Lanczos-type procedure (a rank-r_0 power method), where the right components v are estimated directly from the top r_0 singular vectors of the cross-covariance matrix.

By construction, both variants rely on the spectral properties of the cross-covariance matrix; their weak recovery thresholds, the point at which the estimates start to correlate with the underlying signals, are identical and corresponds to the point where the top singular vectors acquire non-zero overlap with the planted signals. According to Thm. 2.2, it coincides with the emergence of the first spectral outlier and is explicitly characterized by Eq. (2.9). In other words, in the simple case with a single hidden direction in each source ($r = 1$), this can be written formally as the following result.

Corollary 3.1. *Consider the channels in Eq. (1.1) and Eq. (1.2) with $r = 1$. For $z \in \{x, y\}$, let $\widehat{\boldsymbol{v}}_z^{(\mathrm{PLS})}$ denote the estimator obtained from either of the two PLS variants described above with one ($r_0 = 1$) step. Then,*

$$\langle \widehat{\boldsymbol{v}}_z^{(\mathrm{PLS})}, \mathbf{v}_z^\star \rangle^2 \xrightarrow[n\to\infty]{\text{a.s.}} \mathrm{m}_{v,z}^{(\mathrm{PLS})},$$

where $\mathrm{m}_{v,z}^{(\mathrm{PLS})} > 0$ if and only if $1/\mathrm{r}^{\pm} \geq 1/\tau^{+}$, and $\mathrm{m}_{v,z}^{(\mathrm{PLS})} = 0$ otherwise, and an analogous statement holds for the estimators $\widehat{\boldsymbol{u}}_z^{(\mathrm{PLS})}$.

This behavior is illustrated in Fig. 3, where we compare it with the thresholds achieved by other methods (CCA, single-channel spectral methods, and the Bayes-optimal benchmark) and in particular show that *PLS is most effective in asymmetric regimes* where one data channel carries a strong, clearly detectable signal while the other contains a weak signal. When the two are sufficiently correlated, the strong channel can "lift" the weak one, allowing joint recovery that would not be possible from the weak channel alone. Above this threshold, PLS achieves non-zero overlap with both planted directions, thereby exploiting the cross-channel correlation to transfer information. This leveraging effect is visible in the phase diagram (bottom-right region of Fig. 3), where even if one signal lies below its single-channel detectability threshold, PLS succeeds provided the correlation with the strong channel is large enough. By contrast, *when both signals are weak* (bottom-left region of Fig. 3), *PLS offers no advantage over single-channel spectral methods* (based on the within channel covariance matrices) and may even underperform them. The Bayes-optimal estimator combines the best of both worlds: it constructs a block matrix that integrates the cross-covariance and the within-channel covariance matrices, with weights

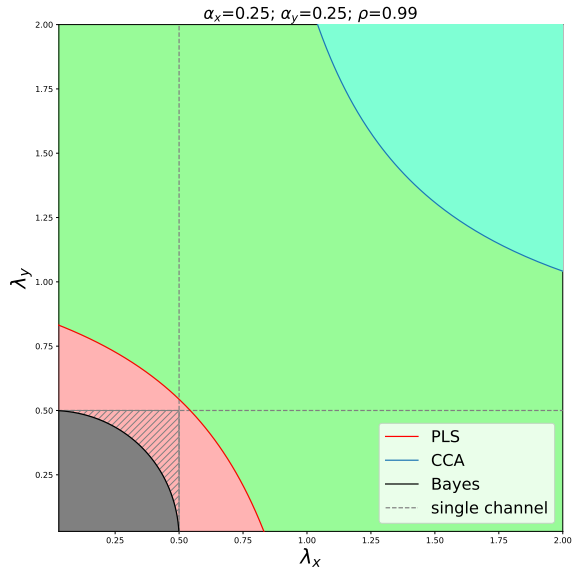


Figure 3: Phase diagram in the (λ_x, λ_y) plane illustrating the detection thresholds for Bayes-optimal method Keup and Zdeborová (2024) (black), single-view channel SVD (gray), PLS (this paper, red), and CCA Bykhovskaya and Gorin (2023); Ma and Yang (2023) (blue) algorithms. The region below each curve corresponds to values of the signal strengths for which spike detection is impossible.

optimally chosen as functions of the model parameters and described explicitly in Keup and Zdeborová (2024).

The performance of PLS when extracting multiple components ($r_0 > 1$) in the presence of several latent directions in each channel ($r > 1$) can, in principle, also be inferred directly from Theorem 2.2 by successively projecting onto the corresponding signal subspaces while this extension is technically straightforward, it yields rather cumbersome expressions. Yet this analysis is more delicate than in the rank-one case. The difficulty arises from the fact that the *limiting positions of the outlier singular values,* $\{b(\mathrm{r}_k^+), b(\mathrm{r}_k^-)\}_{k=1}^r$, are *not necessarily ordered.* For instance, one can encounter configurations where $b(\mathrm{r}_{k_1}^+) > b(\mathrm{r}_{k_2}^-)$ for some $k_1 \neq k_2$. As a result, the leading empirical singular values need not correspond to the same index k, and PLS-SVD therefore align with the "wrong" spike, that is the algorithm produces an estimator correlated with the subspace spanned by $(\mathbf{v}_{x,k_1}, \mathbf{v}_{y,k_1})$ instead of the intended $(\mathbf{v}_{x,k_2}, \mathbf{v}_{y,k_2})$. Although detection thresholds remain the same for PLS and PLS-SVD, the quality of alignment with the true signal subspace may be compromised when multiple spikes interact.

4 OUTLINE OF THE PROOFS

4.1 Proof of the Phase Transition for the Top Singular Values

We introduce the *resolvents* as

$$\boldsymbol{G}(z) = (z - \eta(\boldsymbol{S}))^{-1} \quad \text{and} \quad \tilde{\boldsymbol{G}}(z) = (z - \eta(\tilde{\boldsymbol{S}}))^{-1}\,, \tag{4.1}$$

and to ease notation, we also introduce for $k \in [\![r]\!]$ the matrices

$$\mathbf{Q}_k^\star := \begin{pmatrix} \mathbf{v}_{x,k}^\star & \boldsymbol{X}^T\mathbf{u}_{y,k}^\star & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{v}_{y,k}^\star & \boldsymbol{Y}^T\mathbf{u}_{x,k}^\star \end{pmatrix}, \tag{4.2}$$

and the restrictions of the resolvents into the subspace generated by this matrix:

$$\boldsymbol{H}_k(z) = \mathbf{Q}_k^{\star T}\boldsymbol{G}(z)\mathbf{Q}_k^\star\,,\ \tilde{\boldsymbol{H}}_k(z) = \mathbf{Q}_k^{\star T}\tilde{\boldsymbol{G}}(z)\mathbf{Q}_k^\star\,. \tag{4.3}$$

Both the resolvents and their restrictions into the subspace generated by the $\boldsymbol{Q}_k^\star$ will naturally appear in the proof. We begin with an intermediate result characterizing the possible locations of the outliers.

Lemma 4.1. *The singular values of $\tilde{\boldsymbol{S}}$ that are not the singular values of $\boldsymbol{S} = \boldsymbol{X}^T\boldsymbol{Y}$ are given by the positive solution (in z) of the equation*

$$\prod_{k=1}^r \det\left(\boldsymbol{H}_k(z) - \mathbf{A}_k\right) + \epsilon_n = 0 \tag{4.4}$$

where $\epsilon_n \xrightarrow[n\to\infty]{\text{a.s.}} 0$ *exponentially fast and*

$$\mathbf{A}_k := \begin{pmatrix} 0 & 0 & 0 & \lambda_{x,k}^{-1/2} \\ 0 & 0 & \lambda_{y,k}^{-1/2} & -\rho_k \\ 0 & \lambda_{y,k}^{-1/2} & 0 & 0 \\ \lambda_{x,k}^{-1/2} & -\rho_k & 0 & 0 \end{pmatrix}. \tag{4.5}$$

Proof. The proof follows from the determinant formula and is detailed in App. D. □

We then show that, in the high-dimensional regime $d, n \to \infty$, the high-dimensional random quantities entering the previous lemma admit low-dimensional deterministic equivalents.

Lemma 4.2. *For any* $k \in [\![r]\!]$, *the diagonal entries of the* (4×4) *matrix* $\boldsymbol{H}_k(z) = \big(H_{ij}(z)\big)_{1\le i,j\le 4}$ *of the previous lemma are given asymptotically by*

(i) $H_{11}(z) \xrightarrow[n\to\infty]{\text{a.s.}} h_1(z) := \frac{t(z^2)+1}{z}$,

(ii) $H_{22}(z) \xrightarrow[n\to\infty]{\text{a.s.}} h_2(z) := \frac{z\cdot t(z^2)}{\alpha_x+\alpha_y\cdot t(z^2)}$,

(iii) $H_{33}(z) \xrightarrow[n\to\infty]{\text{a.s.}} h_3(z) := \frac{\alpha_x\alpha_y t(z^2)+1}{z}$,

(iv) $H_{44}(z) \xrightarrow[n\to\infty]{\text{a.s.}} h_4(z) := \frac{1}{\alpha_x}\cdot\frac{z\cdot t(z^2)}{1+t(z^2)}$;

and all other off-diagonal terms go to zero.

Proof. The details of the proofs are given in App. E. They follow from concentration results for Part-(i) and the use of free probability theory for Part-(ii). □

Using the limits of Lemma 4.2 in Lemma 4.1, one ends up with z being an outlier outside the bulk if and only if it is a (positive) zero of one of the functions

$$j_k(z) = \det \begin{pmatrix} h_1(z) & 0 & 0 & \lambda_{x,k}^{-1/2} \\ 0 & h_2(z) & \lambda_{y,k}^{-1/2} & -\rho_k \\ 0 & \lambda_{y,k}^{-1/2} & h_3(z) & 0 \\ \lambda_{x,k}^{-1/2} & -\rho_k & 0 & h_4(z) \end{pmatrix} \tag{4.6}$$

for $k \in [\![r]\!]$, which by expressing everything in terms of the T-transform $t(z)$ thanks to the definition of the $\{h_l\}_{l=1}^4$ writes after simplification:

$$j_k(z) = \left(t(z^2) - \frac{\alpha_x}{\lambda_{x,k}}\right)\left(t(z^2) - \frac{\alpha_x}{\lambda_{y,k}}\right) - \rho_k^2\alpha_x\frac{(1+t(z^2))(\alpha_x+\alpha_y t(z^2))}{z^2}, \tag{4.7}$$

for $k \in [\![r]\!]$. From Prop. 2.1, z and $t(z)$ are related by $P_{(\alpha_x,\alpha_y)}(t/\alpha_x, z) = 0$ which further implies that z is the limiting value of an outlier if it satisfies

$$R_{(\lambda_{x,k},\lambda_{y,k},\rho_k)}(t(z^2)) = 0 \tag{4.8}$$

for some $k \in [\![r]\!]$. Since the T-transform is a continuous decreasing function in the interval (ς_+, ∞) with maximal value obtained at ς_+ and given by τ^+, for the existence of the roots in Eq. (4.8) to hold, one must have that the smallest positive root r^- of the cubic polynomial $R_{(\lambda_{x,k},\lambda_{y,k},\rho_k)}$ is lower than this value τ^+, otherwise Eq. (4.8) is never satisfied by the T-transform. Conversely, if $\tau^+ > \mathrm{r}^-$ (resp. $\tau^+ > \mathrm{r}^+$) one gets the position of an outlier $z > \varsigma_+$ by solving $P_{(\alpha_x,\alpha_y)}(\mathrm{r}^-/\alpha_x, z) = 0$ (respectively $P_{(\alpha_x,\alpha_y)}(\mathrm{r}^+/\alpha_x, z) = 0$), yielding the desired result and concluding the proof.

4.2 Proof of the Phase Transition for the Top Singular Vectors

Lemma 4.3. *Under the same setting as in Thm. 2.2, we have:*

$$\langle \tilde{\boldsymbol{v}}_k^{\pm}, \mathbf{v}_{x,k}^{\star}\rangle^2 = 2\lim_{z\to\tilde{\sigma}_k^{\pm}}(z-\tilde{\sigma}_k^{\pm})\,\tilde{H}_{11}(z)\,,$$
$$\langle \tilde{\boldsymbol{v}}_k^{\pm}, \mathbf{v}_{y,k}^{\star}\rangle^2 = 2\lim_{z\to\tilde{\sigma}_k^{\pm}}(z-\tilde{\sigma}_k^{\pm})\,\tilde{H}_{33}(z)\,,$$

where $\tilde{H}_{ll}(z)$ *denotes the* l*-th diagonal entry of the matrix* $\tilde{\boldsymbol{H}}_k$ *of Eq. (4.3) and* $\{\tilde{\sigma}_k^{\pm}\}_{k=1}^r$ *denotes the set of the top* $2r$ *singular values of* $\tilde{\boldsymbol{S}}$, *ordered such that* $\tilde{\sigma}_k^{\pm} \to b(\mathrm{r}_k^{\pm})$ *as in Thm. 2.1.*

Proof. The proof is detailed in App. F. □

Lemma 4.4. *For any* $k \in [\![r]\!]$, *the diagonal entries of the* (4×4) *matrix* $\tilde{\boldsymbol{H}}_k(z) = \big(\tilde{H}_{ij}(z)\big)_{1\le i,j\le 4}$ *of the previous lemma are given asymptotically by*

$$\tilde{H}_{ll}(z) \xrightarrow[n\to\infty]{\text{a.s.}} h_l(z) + \frac{f_{l,k}(z)}{j_k(z)} \quad \text{for } l \in [\![4]\!], \tag{4.9}$$

where h_l *are given in Lemma 4.2,* j_k *is defined in Eq. (4.7) and* $f_{l,k}$ *is given in App. G.*

Proof. The proof of this result is deferred to App. G. □

Taking the limit $n \to \infty$ in Lemma 4.3 with the asymptotics of Lemma 4.4 one gets that if there is no outlier at z then by Lemma 4.1 $j_k(z) \neq 0$ such that in this case, the associated overlap is asymptotically zero. Conversely, if there is an outlier at $b(\mathrm{r}_k^{\pm})$, then we have $j_k(b(\mathrm{r}_k^{\pm})) = 0$ and one needs to evaluate $\lim_{z\to b(\mathrm{r}_k^{\pm})} 2(z - b(\mathrm{r}_k^{\pm}))f_{l,k}(z)/(j_k(z))$ which by L'Hôpital's rule yields

$$\langle \tilde{\boldsymbol{u}}_k^{\pm}, \mathbf{v}_{x,k}^{\star}\rangle^2 \xrightarrow[n\to\infty]{\text{a.s.}} -2\frac{f_{1,k}\big(b(\mathrm{r}_k^{\pm})\big)}{(j_k)'\big(b(\mathrm{r}_k^{\pm})\big)}\,, \tag{4.10}$$

and similarly for $\langle \tilde{\boldsymbol{v}}_k^{\pm}, \mathbf{v}_{y,k}^{\star}\rangle^2$ with $f_{1,k}$ replaced by $f_{3,k}$. From this point, one differentiates the function $j_k(z)$ using its expression given in Eq. (4.7), which

yields a formula involving z, $t(z^2)$, and its derivative $t'(z^2)$. The latter can be eliminated by differentiating Eq. (2.6), leading to an expression that depends only on z and t. Substituting z with $b(\mathrm{r}_k^{\pm})$ and $t(z^2)$ with $\mathrm{r}_k^{\pm}$, and simplifying, yields the desired result. We refer the reader to Appendix H for details.

5 CONCLUSION

In this work, we provided a rigorous analysis of the spectral properties of spiked cross-covariance matrices in the high-dimensional regime, when the signals across the two channels are partially correlated. Building on tools from random matrix theory, we show the emergence of a Baik Ben-Arous Péché (BBP)-type phase transition in the top singular values and quantified the alignment (overlap) of the corresponding singular vectors with the ground truth signals. As a consequence of these results, we obtain new theoretical insights into the behavior of Partial Least Squares (PLS) methods in high dimensions. In particular, we identified the conditions under which PLS can successfully recover signal structure and compare it with the single-channel setting.

References

Abdelaleem, E., Roman, A., Martini, K. M., and Nemenman, I. (2023). Simultaneous dimensionality reduction: A data efficient approach for multimodal representations learning. *arXiv preprint arXiv:2310.04458*.

Akemann, G., Ipsen, J. R., and Kieburg, M. (2013). Products of rectangular random matrices: singular values and progressive scattering. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, 88(5):052118.

Anderson, G. W., Guionnet, A., and Zeitouni, O. (2010). *An introduction to random matrices*, volume 118 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge.

Asendorf, N. and Nadakuditi, R. R. (2017). Improved detection of correlated signals in low-rank-plus-noise type data sets using informative canonical correlation analysis (icca). *IEEE Transactions on Information Theory*, 63(6):3451–3467.

Attal, E. and Allez, R. (2025). Eigenvector overlaps of random covariance matrices and their submatrices. *Journal of Physics A: Mathematical and Theoretical*, 58(20):205003.

Baik, J. and Silverstein, J. W. (2006). Eigenvalues of large sample covariance matrices of spiked population models. *Journal of multivariate analysis*, 97(6):1382–1408.

Bao, Z., Hu, J., Pan, G., and Zhou, W. (2019). Canonical correlation coefficients of high-dimensional gaussian vectors: Finite rank case. *The Annals of Statistics*, 47(1):612–640.

Belinschi, S. T. and Bercovici, H. (2007). A new approach to subordination results in free probability. *Journal d'Analyse Mathématique*, 101(1):357–365.

Ben Arous, G., Baik, J., and Péché, S. (2005). Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Annals of probability: An official journal of the Institute of Mathematical Statistics*, 33(5):1643–1697.

Benaych-Georges, F., Bouchaud, J.-P., and Potters, M. (2023). Optimal cleaning for singular values of cross-covariance matrices. *The Annals of Applied Probability*, 33(2):1295–1326.

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

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

Burda, Z., Jarosz, A., Livan, G., Nowak, M. A., and Swiech, A. (2010). Eigenvalues and singular values of products of rectangular gaussian random matrices. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, 82(6):061114.

Bykhovskaya, A. and Gorin, V. (2023). High-dimensional canonical correlation analysis. *arXiv preprint arXiv:2306.16393*.

Capitaine, M. (2018). Limiting eigenvectors of outliers for spiked information-plus-noise type matrices. *Séminaire de Probabilités XLIX*, pages 119–164.

Duranthon, O. and Zdeborová, L. (2023). Optimal inference in contextual stochastic block models. *arXiv preprint arXiv:2306.07948*.

Guionnet, A., Ko, J., Krzakala, F., Mergny, P., and Zdeborová, L. (2023). Spectral phase transitions in non-linear wigner spiked models. *arXiv preprint arXiv:2310.14055*.

Guo, C. and Wu, D. (2019). Canonical correlation analysis (cca) based multi-view learning: An overview. *arXiv preprint arXiv:1907.01693*.

Hulland, J. (1999). Use of partial least squares (pls) in strategic management research: A review of four recent studies. *Strategic management journal*, 20(2):195–204.

Johnstone, I. M. and Lu, A. Y. (2009). On consistency and sparsity for principal components anal-

ysis in high dimensions. *Journal of the American Statistical Association*, 104(486):682–693.

Keup, C. and Zdeborová, L. (2024). Optimal thresholds and algorithms for a model of multi-modal learning in high dimensions.

Krishnan, A., Williams, L. J., McIntosh, A. R., and Abdi, H. (2011). Partial least squares (pls) methods for neuroimaging: a tutorial and review. *Neuroimage*, 56(2):455–475.

Lebeau, H., Seddik, M. E. A., and Goulart, J. H. d. M. (2024). Performance gaps in multi-view clustering under the nested matrix-tensor model. *arXiv preprint arXiv:2402.10677*.

Lelarge, M. and Miolane, L. (2017). Fundamental limits of symmetric low-rank matrix estimation. In *Conference on Learning Theory*, pages 1297–1301. PMLR.

Loubaton, P. and Vallet, P. (2011). Almost Sure Localization of the Eigenvalues in a Gaussian Information Plus Noise Model. Application to the Spiked Models. *Electronic Journal of Probability*, 16(none):1934 – 1959.

Ma, Z. and Nandy, S. (2023). Community detection with contextual multilayer networks. *IEEE Transactions on Information Theory*, 69(5):3203–3239.

Ma, Z. and Yang, F. (2023). Sample canonical correlation coefficients of high-dimensional random vectors with finite rank correlations. *Bernoulli*, 29(3):1905–1932.

Mingo, J. A. and Speicher, R. (2017). *Free probability and random matrices*, volume 35. Springer.

Ngiam, J., Khosla, A., Kim, M., Nam, J., Lee, H., Ng, A. Y., et al. (2011). Multimodal deep learning. In *ICML*, volume 11, pages 689–696.

Pacco, A. and Ros, V. (2023). Overlaps between eigenvectors of spiked, correlated random matrices: From matrix principal component analysis to random gaussian landscapes. *Physical Review E*, 108(2):024145.

Paul, D. (2007). Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statistica Sinica*, pages 1617–1642.

Péché, S. (2014). Deformed ensembles of random matrices. In *Proceedings of the International Congress of Mathematicians, Seoul*, volume 3, pages 1059–1174.

Perry, A., Wein, A. S., Bandeira, A. S., and Moitra, A. (2018). Optimality and sub-optimality of pca i: Spiked random matrix models. *The Annals of Statistics*, 46(5):2416–2451.

Pirouz, D. M. (2006). An overview of partial least squares. *Available at SSRN 1631359*.

Potters, M. and Bouchaud, J.-P. (2020). *A first course in random matrix theory: for physicists, engineers and data scientists*. Cambridge University Press.

Qi, Y. and Decurninge, A. (2025). Statistical limits in random tensors with multiple correlated spikes. *arXiv preprint arXiv:2503.03356*.

Ramachandram, D. and Taylor, G. W. (2017). Deep multimodal learning: A survey on recent advances and trends. *IEEE signal processing magazine*, 34(6):96–108.

Sergazinov, R., Taeb, A., and Gaynanova, I. (2025). A spectral method for multi-view subspace learning using the product of projections. *Biometrika*, page asaf088.

Swain, A., Ridout, S. A., and Nemenman, I. (2025). Distribution of singular values in large sample cross-covariance matrices.

Tabanelli, H., Mergny, P., Zdeborova, L., and Krzakala, F. (2025). Computational thresholds in multi-modal learning via the spiked matrix-tensor model.

Tao, T. (2012). *Topics in random matrix theory*, volume 132. American Mathematical Soc.

Thompson, B. (2000). Canonical correlation analysis.

Wegelin, J. A. et al. (2000). A survey of partial least squares (pls) methods, with emphasis on the two-block case. Technical report, Technical report.

Wold, H. (1975). Path models with latent variables: The nipals approach. In *Quantitative sociology*, pages 307–357. Elsevier.

Wold, H. (1983). Systems analysis by partial least squares.

Wold, S., Sjöström, M., and Eriksson, L. (2001). Pls-regression: a basic tool of chemometrics. *Chemometrics and intelligent laboratory systems*, 58(2):109–130.

Yang, F. (2022a). Limiting distribution of the sample canonical correlation coefficients of high-dimensional random vectors. *Electronic Journal of Probability*, 27:1–71.

Yang, F. (2022b). Sample canonical correlation coefficients of high-dimensional random vectors: Local law and tracy–widom limit. *Random Matrices: Theory and Applications*, 11(01):2250007.

Yang, X., Lin, B., and Sen, S. (2024). Fundamental limits of community detection from multi-view data: multi-layer, dynamic and partially labeled block models. *arXiv preprint arXiv:2401.08167*.

Zou, H., Hastie, T., and Tibshirani, R. (2006). Sparse principal component analysis. *Journal of computational and graphical statistics*, 15(2):265–286.

Checklist

1. For all models and algorithms presented, check if you include:
   (a) A clear description of the mathematical setting, assumptions, algorithm, and/or model. [Yes]
   (b) An analysis of the properties and complexity (time, space, sample size) of any algorithm. [Yes]
   (c) (Optional) Anonymized source code, with specification of all dependencies, including external libraries. [Not Applicable]

2. For any theoretical claim, check if you include:
   (a) Statements of the full set of assumptions of all theoretical results. [Yes]
   (b) Complete proofs of all theoretical results. [Yes]
   (c) Clear explanations of any assumptions. [Yes]

3. For all figures and tables that present empirical results, check if you include:
   (a) The code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL). [Yes]
   (b) All the training details (e.g., data splits, hyperparameters, how they were chosen). [Yes]
   (c) A clear definition of the specific measure or statistics and error bars (e.g., with respect to the random seed after running experiments multiple times). [Yes]
   (d) A description of the computing infrastructure used. (e.g., type of GPUs, internal cluster, or cloud provider). [Not Applicable]

4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets, check if you include:
   (a) Citations of the creator If your work uses existing assets. [Not Applicable]
   (b) The license information of the assets, if applicable. [Not Applicable]
   (c) New assets either in the supplemental material or as a URL, if applicable. [Not Applicable]
   (d) Information about consent from data providers/curators. [Not Applicable]
   (e) Discussion of sensible content if applicable, e.g., personally identifiable information or offensive content. [Not Applicable]

5. If you used crowdsourcing or conducted research with human subjects, check if you include:
   (a) The full text of instructions given to participants and screenshots. [Not Applicable]
   (b) Descriptions of potential participant risks, with links to Institutional Review Board (IRB) approvals if applicable. [Not Applicable]
   (c) The estimated hourly wage paid to participants and the total amount spent on participant compensation. [Not Applicable]

Supplementary Materials

A Standard Linear Algebra and Random Matrix Theory Results

In this section, we review some classical results from linear algebra and random matrix theory (RMT); see, e.g., Benaych-Georges and Nadakuditi (2012) for references.

Lemma A.1 (Woodbury Resolvent identity)**.** *Let* $\mathbf{A} \in \mathbb{M}_n^H$, $\mathbf{Q} \in \mathbb{M}_{n,d}$, $\mathbf{C} \in \mathbb{M}_d^H$, $\mathbf{A'} = \mathbf{A} + \mathbf{QCQ}^T$; $\mathbf{G} = (z\mathbf{I} - \mathbf{A})^{-1}$ *and* $\mathbf{G'} = (z\mathbf{I} - \mathbf{A'})^{-1}$*,then the following identity holds:*

$$\mathbf{G'} = \mathbf{G} + \mathbf{GQC}(\mathbf{I} - \mathbf{Q}^T\mathbf{GQC})^{-1}\mathbf{Q}^T\mathbf{G}\,, \tag{A.1}$$

in particular if we define $\mathbf{K} := \mathbf{Q}^T\mathbf{GQ}$ *and* $\mathbf{K'} := \mathbf{Q}^T\mathbf{G'Q}$*, we have:*

$$\mathbf{K'} = \mathbf{K} + \mathbf{KC}(\mathbf{I} - \mathbf{KC})^{-1}\mathbf{K}\,. \tag{A.2}$$

Lemma A.2. *Let* $\mathbf{A} \in \mathbb{M}_n^H$, $\mathbf{G}(z) := (z\mathbf{I} - \mathbf{A})^{-1}$*, we denote by* $\{\mathbf{u}_i\}_{i\in[\![n]\!]}$ *the eigenvectors of* $\mathbf{A}$ *associated to the eigenvalues* $\{\lambda_i\}_{i\in[\![n]\!]}$*, counted with multiplicity. Fix* $\lambda_{i_0} \in \mathrm{Spec}(\mathbf{A})$ *and* $\mathbf{q} \in \mathbb{R}^n$ *then for any analytic function* $f()$ *with* $f(\lambda_{i_0}) \neq 0$*, we have*

$$|\langle \mathbf{q}, \mathrm{Span}\{\mathbf{u}_j | \lambda_j = \lambda_{i_0}\}\rangle|^2 = \lim_{z\to\lambda_{i_0}} \frac{(z - \lambda_{i_0})}{f(\lambda_{i_0})} \cdot |\langle \mathbf{q}\,, f(\mathbf{A})\mathbf{G}(z)\mathbf{q}\rangle|\,. \tag{A.3}$$

Proof. This follows directly from the eigendecomposition of $\mathbf{G}(z) = \frac{1}{z-\lambda_{i_0}}\sum_{i|\lambda_i=\lambda_{i_0}} \mathbf{u}_i\mathbf{u}_i^T + \sum_{i|\lambda_i\neq\lambda_{i_0}} \frac{1}{z-\lambda_i}\mathbf{u}_i\mathbf{u}_i^T$. □

Lemma A.3 (Eigenvalue decomposition)**.** *Let* $\mathbf{B} \in \mathbb{M}_{n,m}$ *with singular value decomposition* $\mathbf{B} = \mathbf{UDV}^T$*, where* $\mathbf{U} \in \mathbb{U}_{n,r}$, $\mathbf{V} \in \mathbb{U}_{m,r}$, $\mathbf{D} \in \mathbb{D}_r$*. Introduce the three matrices* $\mathbf{Q} = \begin{pmatrix} \mathbf{U} & \mathbf{0} \\ \mathbf{0} & \mathbf{V} \end{pmatrix}$, $\mathbf{W}_+ := \frac{1}{\sqrt{2}}[\mathbf{V} \quad \mathbf{U}]^T$, $\mathbf{W}_- := \frac{1}{\sqrt{2}}[\mathbf{V} \quad -\mathbf{U}]^T$*, then we have the following decompositions for the chiral extension of* $\mathbf{B}$*:*

$$(i) \qquad \eta(\mathbf{B}) = \mathbf{Q}\begin{pmatrix} \mathbf{0} & \mathbf{D} \\ \mathbf{D} & \mathbf{0} \end{pmatrix}\mathbf{Q}^T\,; \tag{A.4}$$

$$(ii) \qquad \eta(\mathbf{B}) = [\mathbf{W}_+ \ \mathbf{W}_-]\begin{pmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & -\mathbf{D} \end{pmatrix}[\mathbf{W}_+ \ \mathbf{W}_-]^T\,. \tag{A.5}$$

Lemma A.4 (Resolvent for Chiral Matrices)**.** *For* $\mathbf{B} \in \mathbb{M}_{n,m}$*, we have*

$$\big(z\mathbf{I} - \eta(\mathbf{B})\big)^{-1} = \begin{pmatrix} z\cdot\mathbf{\Gamma} & \mathbf{\Gamma B} \\ \mathbf{B}^T\mathbf{\Gamma} & z\cdot\check{\mathbf{\Gamma}} \end{pmatrix} \tag{A.6}$$

with $\mathbf{\Gamma} := (z^2\mathbf{I} - \mathbf{BB}^T)^{-1}$ *and* $\check{\mathbf{\Gamma}} := (z^2\mathbf{I} - \mathbf{B}^T\mathbf{B})^{-1}$*.*

Lemma A.5 (matrix determinant lemma for chiral matrices)**.** *let* $\mathbf{B} \in \mathbb{M}_{n,m}$, $\mathbf{L} \in \mathbb{M}_{n,d}$, $\mathbf{C} \in \mathbb{M}_{d,d}$ *and invertible,* $\mathbf{R} \in \mathbb{M}_{m,d}$*, then for any* $z \notin \mathrm{Spec}\big(\eta(\mathbf{B})\big)$ *we have:*

$$\det(z\mathbf{I} - \eta(\mathbf{B} + \mathbf{LCR}^T)) = \det(\mathbf{C})^2 \cdot \det(z\mathbf{I} - \eta(\mathbf{B})) \cdot \det(\mathbf{\Upsilon}(z))\,, \tag{A.7}$$

with the $(2d \times 2d)$ *matrix* $\mathbf{\Upsilon}(z)$ *given by*

$$\mathbf{\Upsilon}(z) := \begin{pmatrix} z\cdot\mathbf{L}^T\mathbf{\Gamma L} & \mathbf{L}^T\mathbf{\Gamma BR} \\ (\mathbf{BR})^T\mathbf{\Gamma L} & z\cdot\mathbf{R}^T\check{\mathbf{\Gamma}}\mathbf{R} \end{pmatrix} - \eta\big((\mathbf{C}^T)^{-1}\big)\,. \tag{A.8}$$

and where $\mathbf{\Gamma} := (z^2\mathbf{I} - \mathbf{BB}^T)^{-1}$, $\check{\mathbf{\Gamma}} := (z^2\mathbf{I} - \mathbf{B}^T\mathbf{B})^{-1}$*.*

Lemma A.6 (Concentration Quadratic Form)**.** *For* $\mathbf{A} \in \mathbb{M}_{n,n}^H, \mathbf{B} \in \mathbb{M}_{n,m}$*, and* $\mathbf{u}_i, \mathbf{v}_i$*, we have as* $n \to \infty$

$$\langle \mathbf{u}_i, \mathbf{A}\mathbf{u}_i \rangle - n^{-1}\mathbb{E}\mathrm{Tr}\mathbf{A} \to 0\,, \tag{A.9}$$

$$\langle \mathbf{u}_i, \mathbf{A}\mathbf{u}_j \rangle \to 0\,, \tag{A.10}$$

$$\langle \mathbf{u}_i, \mathbf{B}\mathbf{v}_k \rangle \to 0\,, \tag{A.11}$$

Lemma A.7. *If we denote by* $\mathrm{Spec}_+(\mathbf{A}) = \{z \in \mathrm{Spec}(\mathbf{A}) \setminus 0\}$*, then for* $\mathbf{A} \in \mathbb{R}^{n,m_1}$*,* $\mathbf{B} \in \mathbb{R}^{n,n}$*, we have* $\mathrm{Spec}_+(\mathbf{A}^T\mathbf{B}\mathbf{A}) = \mathrm{Spec}_+((\mathbf{A}\mathbf{A}^T)^{1/2}\mathbf{B}(\mathbf{A}\mathbf{A}^T)^{1/2})$*.*

For $\mu \in \mathcal{P}_c(\mathbb{R}_+)$ with rightmost edge λ_+, recall the definition of the t-transform that uniquely characterizes the distribution μ:

$$t_\mu(z) := \int_{\mathbb{R}} (z-x)^{-1} x \,\mathrm{d}\mu(x)\,, \tag{A.12}$$

which is strictly decreasing on (λ_+, ∞) and thus admits an inverse for the composition that we denote $t^{\langle -1 \rangle}$. The S-transform of a measure μ is defined as

$$S_\mu(\theta) := \frac{\theta + 1}{\theta t_\mu^{\langle -1 \rangle}(\theta)} \tag{A.13}$$

and appears naturally whenever one consider the product of two free elements:

Lemma A.8. *(free product). Let* $\mu, \nu \in \mathcal{P}_c(\mathbb{R}_+)$*, then there exists a unique distribution* $\mu \boxtimes \nu$ *known as the free convolution of* μ *and* ν *such that*

$$S_{\mu \boxtimes \nu}(\theta) = S_\mu(\theta)\, S_\nu(\theta)\,. \tag{A.14}$$

Furthermore for a sequence of $(n \times n)$ *independent symmetric positive definite matrices* $(\mathbf{A}_n, \mathbf{B}_n)_n$ *such that* $\mathbf{A}_n$ *is orthogonally invariant in law (*$\mathbf{A}_n \overset{\mathcal{D}}{=} \mathbf{O}\mathbf{A}_n\mathbf{O}^T$ *for any orthogonal matrix* $\mathbf{O}$*) and* $\mu_{\mathbf{A}_n} \to \mu$ *and* $\mu_{\mathbf{B}_n} \to \nu$*, we have*

$$\mu_{\mathbf{A}_n^{1/2}\mathbf{B}_n\mathbf{A}_n^{1/2}} \to \mu \boxtimes \nu. \tag{A.15}$$

B Properties of the roots of the Cubic Polynomials (Lemma 2.1)

For $Q_{(\alpha_x,\alpha_y)}$ defined by Eq. (2.3), we have $Q(0) = 1$ and $\lim_{x\to\infty} Q_{(\alpha_x,\alpha_y)}(x) = -\infty$, since $Q'_{(\alpha_x,\alpha_y)}(x) = -2(\alpha_x\alpha_y + \alpha_x + \alpha_y)x - 6\alpha_x\alpha_y x^2$ is non-zero for $x > 0$, the function admits exactly one positive root.

Similarly for $R_{(\lambda_x,\lambda_y,\rho)}$ defined by Eq. (2.4), we have $R_{(\lambda_x,\lambda_y,\rho)}(0) = 1$ and $\lim_{x\to\infty} R_{(\lambda_x,\lambda_y,\rho)}(x) = \infty$. Its discriminant is given by

$$\begin{aligned}\mathrm{Dis}(R_{(\lambda_x,\lambda_y,\rho)}) &= (1+\lambda_x)^2(\lambda_x-\lambda_y)^2(1+\lambda_y)^2 \\ &\quad + 2\lambda_x\lambda_y\big((-1+\lambda_y)\lambda_y^2 + \lambda_x^2(-1+\lambda_y)(1+(-1+\lambda_y)\lambda_y) \\ &\qquad + 2\lambda_x\lambda_y(2+\lambda_y+2\lambda_y^2) + \lambda_x^3(1+\lambda_y(4+\lambda_y))\big)\rho^2 \\ &\quad + \lambda_x^2\lambda_y^2\left(10\lambda_x(-1+\lambda_y)\lambda_y + \lambda_y^2 + \lambda_x^2(1+\lambda_y(10+\lambda_y))\right)\rho^4 + 4\lambda_x^4\lambda_y^4\rho^6\end{aligned}$$

which is positive, hence $R_{(\lambda_x,\lambda_y,\rho)}(x)$ has three real roots on the real axis. Its derivative is given by $R'_{(\lambda_x,\lambda_y,\rho)}(x) = 1 - \lambda_x - \lambda_y - \lambda_x\lambda_y\rho^2 + 2(\lambda_x\lambda_y - \lambda_x - \lambda_y)x + 3\lambda_x\lambda_y x^2$ which has one positive root at

$$x_+ = \frac{2\lambda_x + 2\lambda_y - 2\lambda_x\lambda_y + \sqrt{(-2\lambda_x - 2\lambda_y + 2\lambda_x\lambda_y)^2 - 12\lambda_x\lambda_y(1-\lambda_x-\lambda_y-\lambda_x\lambda_y\rho^2)}}{6\lambda_x\lambda_y} \tag{B.1}$$

hence $R_{(\lambda_x,\lambda_y,\rho)}$ itself must have two positive roots.

Next to get the dependency with ρ^2, we differentiate the fixed point equation $R_{(\lambda_x,\lambda_y,\rho)}(\mathrm{r}_\pm(\rho^2)) = 0$ with respect to this parameter:

$$\partial_{\rho^2} R_{(\lambda_x,\lambda_y,\rho)}(\mathrm{r}_\pm(\rho^2)) + (\mathrm{r}_\pm)'(\rho^2) \cdot R'_{(\lambda_x,\lambda_y,\rho)}(\mathrm{r}_\pm(\rho^2)) = 0 \tag{B.2}$$

such that

$$(\mathrm{r}_\pm)'(\rho^2) = -\frac{\partial_{\rho^2} R_{(\lambda_x,\lambda_y,\rho)}(\mathrm{r}_\pm(\rho^2)}{R'_{(\lambda_x,\lambda_y,\rho)}(\mathrm{r}_\pm(\rho^2))} \tag{B.3}$$

We have

$$\partial_{\rho^2} R_{(\lambda_x,\lambda_y,\rho)}(\mathrm{r}_\pm(\rho^2)) = -2\lambda_x\lambda_y \mathrm{r}_\pm(\rho^2) < 0 \tag{B.4}$$

and since $R_{(\lambda_x,\lambda_y,\rho)}$ is decreasing on $(0, x_+)$ and increasing on (x_+, ∞) we have $\mathrm{sign}(R'_{(\lambda_x,\lambda_y,\rho)}(\mathrm{r}_\pm(\rho^2))) = \pm$ from which one gets the desired result.

C Bulk Distribution of Singular Values (Prop. 2.1)

Since we assume $d_x \leq d_y$ without loss of generality, the singular values of $\boldsymbol{S}$ are by definition the square root of the eigenvalues of the matrix $\boldsymbol{X}^T\boldsymbol{Y}\boldsymbol{Y}^T X$ and by Lemma A.7 one has

$$\mathrm{Spec}_+(\boldsymbol{X}^T\boldsymbol{Y}\boldsymbol{Y}^T\boldsymbol{X}) = \mathrm{Spec}_+\big((\boldsymbol{X}\boldsymbol{X}^T)^{1/2}\boldsymbol{Y}\boldsymbol{Y}^T(\boldsymbol{X}\boldsymbol{X}^T)^{1/2}\big)\,, \tag{C.1}$$

from which one can check that the T-transform of the two measures are related by

$$t_{(\boldsymbol{X}\boldsymbol{X}^T)^{1/2}\boldsymbol{Y}\boldsymbol{Y}^T(\boldsymbol{X}\boldsymbol{X}^T)^{1/2}}(z) = \frac{1}{\alpha_x} t_{\boldsymbol{S}\boldsymbol{S}^T}(z) \qquad \text{with } \alpha_x = n/d_x\,. \tag{C.2}$$

From classical random matrix result Potters and Bouchaud (2020), the limiting distribution of $\boldsymbol{X}\boldsymbol{X}^T$ (respectively of $\boldsymbol{Y}\boldsymbol{Y}^T$) is known to be given by the Marcenko-Pastur distribution of aspect ratio α_x (resp. α_y) whose S-transform is

$$S_{\mu_{\mathrm{MP}(\alpha)}}(\theta) = (1+\alpha\theta)^{-1}\,, \tag{C.3}$$

with $\alpha = \alpha_x$ (resp. α_y). Next by Assumption (A1), the two matrices $\boldsymbol{X}, \boldsymbol{Y}$ are independant with Gaussian entries, hence the matrices $\boldsymbol{X}\boldsymbol{X}^T$ is orthogonally invariant and thus using Lemma A.8, the limiting spectral distribution of $(\boldsymbol{X}\boldsymbol{X}^T)^{1/2}\boldsymbol{Y}\boldsymbol{Y}^T(\boldsymbol{X}\boldsymbol{X}^T)^{1/2}$ is given by the free convolution $\mu_{\mathrm{MP}(\alpha_x)} \boxtimes \mu_{\mathrm{MP}(\alpha_y)}$ whose T-transform $t_\boxtimes \equiv t_{\mu_{\mathrm{MP}(\alpha_x)}\boxtimes\mu_{\mathrm{MP}(\alpha_y)}}$ satisfies:

$$t_\boxtimes(z)z = (1+t_\boxtimes(z))(1+\alpha_x t_\boxtimes(z))(1+\alpha_y t_\boxtimes(z))\,, \tag{C.4}$$

and by Eq. (C.2) the one for the distribution of the square of the singular values of $\boldsymbol{S}$ is giving simply by $t(z) = \alpha_x t_\boxtimes(z)$ from which one reads the desired result.

D Determinant Equation for the eigenvalues (Lemma 4.1)

If we denote by $\boldsymbol{S} := \boldsymbol{X}^T\boldsymbol{Y}$, the cross-covariance matrix without any spiked, we have the decomposition

$$\tilde{\boldsymbol{S}} = \boldsymbol{S} + \sum_{k=1}^r \boldsymbol{P}_k\,, \tag{D.1}$$

where the matrix $\boldsymbol{P}_k$ is given by

$$\boldsymbol{P}_k = \sqrt{\lambda_{x,k}\lambda_{y,k}}\,\langle \mathbf{u}^\star_{x,k}, \mathbf{u}^\star_{y,k}\rangle \cdot \mathbf{v}^\star_{x,k}(\mathbf{v}^\star_{y,k})^T + \sqrt{\lambda_{y,k}}\cdot(\boldsymbol{X}^T\mathbf{u}^\star_{y,k})(\mathbf{v}^\star_{y,k})^T \tag{D.2}$$

$$+\sqrt{\lambda_{x,k}}\cdot\mathbf{v}^\star_{x,k}\big(\boldsymbol{Y}^T\mathbf{u}^\star_{x,k}\big)^T\,, \tag{D.3}$$

such that the matrix $\boldsymbol{P} = \sum_{i=1}^r \boldsymbol{P}_k$ can be rewritten as

$$\boldsymbol{P} = (\mathbf{v}^\star_{x,1}, \boldsymbol{X}^T\mathbf{u}^\star_{y,1}, \ldots, \mathbf{v}^\star_{x,r}, \boldsymbol{X}^T\mathbf{u}^\star_{y,r}) \tag{D.4}$$

$$\mathrm{BlockDiag}(\mathbf{A}_1, \ldots, \mathbf{A}_r)(\mathbf{v}^\star_{x,1}, \boldsymbol{Y}^T\mathbf{u}^\star_{y,1}, \ldots, \mathbf{v}^\star_{y,1}, \boldsymbol{Y}^T\mathbf{u}^\star_{y,r})^T\,, \tag{D.5}$$

with the (2×2) matrices $\mathbf{A}_k$ defined by

$$\mathbf{A}_k := \begin{pmatrix} \sqrt{\lambda_{x,k}\lambda_{y,k}}\,\rho_k & \sqrt{\lambda_{x,k}} \\ \sqrt{\lambda_{y,k}} & 0 \end{pmatrix} \qquad \text{with } \rho_k := \langle \mathbf{u}^\star_{x,k}, \mathbf{u}^\star_{y,k} \rangle \,. \tag{D.6}$$

For $\lambda_{x,k}, \lambda_{y,k} > 0$, the matrix $\mathbf{A}_k$ is invertible with inverse

$$\mathbf{A_k}^{-1} = \begin{pmatrix} 0 & \lambda_{x,k}^{-1/2} \\ \lambda_{y,k}^{-1/2} & -\rho_k \end{pmatrix} . \tag{D.7}$$

Next, to ease notations we denote by

$$\tilde{\boldsymbol{Z}}_n = \eta(\tilde{\boldsymbol{S}}) := \begin{pmatrix} \mathbf{0} & \tilde{\boldsymbol{S}} \\ \tilde{\boldsymbol{S}}^T & \mathbf{0} \end{pmatrix} \qquad \text{and} \qquad \boldsymbol{Z}_n = \eta(\boldsymbol{S}) := \begin{pmatrix} \mathbf{0} & \boldsymbol{S} \\ \boldsymbol{S}^T & \mathbf{0} \end{pmatrix} , \tag{D.8}$$

by determinant lemma A.5, the characteristic polynomial of $\tilde{\boldsymbol{Z}}_n$ is given by

$$\det\big(z\mathbf{I} - \tilde{\boldsymbol{Z}}_n\big) = \prod_{i=1}^{k} (\det \mathbf{A}_k)^2 \cdot \det\big(z\mathbf{I} - \boldsymbol{Z}_n\big) \cdot \det \boldsymbol{M}_n(z) \tag{D.9}$$

with the matrix $\boldsymbol{M}_n(z)$

$$\boldsymbol{M}_n(z) := \text{BlockDiag}\left(\begin{pmatrix} \mathbf{0} & \mathbf{A}_k \\ \mathbf{A}_k^T & \mathbf{0} \end{pmatrix}_{k=1,\dots,r} \right)^{-1} - \boldsymbol{\Theta}_n(z) \tag{D.10}$$

where since the components satisfy Assumption **A2**, we have by Lemma A.6 the elements outside the block-diagonal of $\boldsymbol{\Theta}_n(z)$vanishes exponentially fast and $\rho_{n,k} \to \rho_k$, leading to

$$\det(\boldsymbol{\Theta}_n(z) - \text{BlockDiag}(\boldsymbol{H}_k)) \to 0 \,, \tag{D.11}$$

with $\boldsymbol{H}_k$ defined in Eq. (4.3). Since the determinant of block-diagonal matrix is the product of the determinant of its blocks and the inverse of $\mathbf{A}_k$ is given by Eq. (D.7), one gets the desired result.

E Concentration of the diagonal entries (4.2)

We first show that each diagonal entries can be approximated by the trace of certain random matrices. To ease notation, in the following paragraph we write $\mathbf{v}^\star_x \equiv \mathbf{v}^\star_{x,k}$ (and similarly for other components) as the computation is the same for each block.

Lemma E.1. *The diagonal entries of $\boldsymbol{H}_k$ satisfy*

$$\left| h_{1,n}(z) - \frac{z}{d_x} \mathbb{E}\text{Tr}(z^2 - \boldsymbol{S}\boldsymbol{S}^T)^{-1} \right| \to 0 \,, \tag{E.1}$$

$$\left| h_{2,n}(z) - \frac{z}{n} \mathbb{E}\text{Tr}\boldsymbol{X}(z^2 - \boldsymbol{S}\boldsymbol{S}^T)^{-1}\boldsymbol{X} \right| \to 0 \,, \tag{E.2}$$

$$\left| h_{3,n}(z) - \frac{z}{d_y} \mathbb{E}\text{Tr}(z^2 - \boldsymbol{S}^T\boldsymbol{S})^{-1} \right| \to 0 \,, \tag{E.3}$$

$$\left| h_{4,n}(z) - \frac{z}{n} \mathbb{E}\text{Tr}\boldsymbol{Y}^T(z^2 - \boldsymbol{S}^T\boldsymbol{S})^{-1}\boldsymbol{Y}^T \right| \to 0 \,. \tag{E.4}$$

Proof. By definition of the matrix $\boldsymbol{H}$, its diagonal entries $\{h_{i,n}(z)\}_{i \in [\![4]\!]}$ are given by

$$h_{1,n}(z) := \langle (\mathbf{v}^\star_x \quad \mathbf{0}), \boldsymbol{G}(z)(\mathbf{v}^\star_x \quad \mathbf{0}) \rangle \,, \tag{E.5}$$

$$h_{2,n}(z) := \langle (\boldsymbol{X}^T\mathbf{u}^\star_y \quad \mathbf{0}), \boldsymbol{G}(z)(\boldsymbol{X}^T\mathbf{u}^\star_y \quad \mathbf{0}) \rangle \,, \tag{E.6}$$

$$h_{3,n}(z) := \langle (\mathbf{0} \quad \mathbf{v}^\star_y), \boldsymbol{G}(z)(\mathbf{0} \quad \mathbf{v}^\star_y) \rangle \,, \tag{E.7}$$

$$h_{4,n}(z) := \langle (\mathbf{0} \quad \boldsymbol{Y}^T\mathbf{u}^\star_x), \boldsymbol{G}(z)(\mathbf{0} \quad \boldsymbol{Y}^T\mathbf{u}^\star_x) \rangle \,, \tag{E.8}$$

and thus correspond to quadratic forms with either the top-left corner (for $h_{1,n}, h_{2,n}$) or bottom-right corner (for $h_{3,n}, h_{4,n}$) of the resolvent matrix $\boldsymbol{G}(z)$ of the hermitian matrix $\eta(\boldsymbol{S})$. By properties of resolvent of chiral matrices (see Lemma A.4) the latter projections are given as the resolvent of $\boldsymbol{S}\boldsymbol{S}^T$ (resp. of $\boldsymbol{S}^T\boldsymbol{S}$) evaluated at z^2, such that we have:

$$h_{1,n}(z) = \langle \mathbf{v}_x^\star, z(z^2 - \boldsymbol{S}\boldsymbol{S}^T)^{-1}\mathbf{v}_x^\star\rangle\,, \tag{E.9}$$
$$h_{2,n}(z) := \langle \mathbf{u}_y^\star, z\boldsymbol{X}(z^2 - \boldsymbol{S}\boldsymbol{S}^T)^{-1}\boldsymbol{X}\mathbf{u}_y^\star\rangle\,, \tag{E.10}$$
$$h_{3,n}(z) := \langle \mathbf{v}_y^\star, z(z^2 - \boldsymbol{S}^T\boldsymbol{S})^{-1}\mathbf{v}_y^\star\rangle\,, \tag{E.11}$$
$$h_{4,n}(z) := \langle \mathbf{u}_x^\star, \boldsymbol{Y}^T z(z^2 - \boldsymbol{S}^T\boldsymbol{S})^{-1}\boldsymbol{Y}^T\mathbf{u}_x^\star\rangle\,, \tag{E.12}$$

the limits of these quadratic forms are captured by Lemma A.6 □

taking the limit $n \to \infty$, one gets immediately the desired result for Part-(i) of the Lemma. For Part-(ii), we first use the intermediate lemma:

Lemma E.2. *If we denote by* $\boldsymbol{\Gamma}_n = (z^2 - \boldsymbol{S}\boldsymbol{S}^T)^{-1}$ *and* $\check{\boldsymbol{\Gamma}}_n = (z^2 - \boldsymbol{S}^T\boldsymbol{S})^{-1}$*, we have*

$$\frac{z}{n}\mathrm{Tr}\boldsymbol{X}\boldsymbol{\Gamma}_n\boldsymbol{X}^T = \frac{z}{n}\mathrm{Tr}\,\mathbf{G}_{\boldsymbol{S}_x^{1/2}\boldsymbol{S}_y\boldsymbol{S}_x^{1/2}}(z^2)\boldsymbol{S}_x \tag{E.13}$$
$$\frac{z}{n}\mathrm{Tr}\boldsymbol{Y}\check{\boldsymbol{\Gamma}}_n\boldsymbol{Y}^T = \frac{z}{n}\mathrm{Tr}\,\mathbf{G}_{\boldsymbol{S}_y^{1/2}\boldsymbol{S}_x\boldsymbol{S}_y^{1/2}}(z^2)\boldsymbol{S}_y \tag{E.14}$$

with $\mathbf{G}_{\mathbf{A}}(z) := (z - \mathbf{A})^{-1}$ *the resolvent and* $\boldsymbol{S}_x := \boldsymbol{X}\boldsymbol{X}^T$*,* $\boldsymbol{S}_y := \boldsymbol{Y}\boldsymbol{Y}^T$*.*

Proof. Taking z sufficiently away, one has

$$\frac{z}{N}\mathrm{Tr}\boldsymbol{X}\boldsymbol{\Gamma}_n\boldsymbol{X}^T = \frac{z}{N}\mathrm{Tr}\boldsymbol{X}\left(z^2\cdot\sum_{k=0}^{\infty} z^{-2k}\left(\boldsymbol{X}^T\boldsymbol{Y}\boldsymbol{Y}^T\boldsymbol{X}\right)^k\right)\boldsymbol{X}^T\,, \tag{E.15}$$
$$= \frac{z}{N}\mathrm{Tr}\left(z^2\cdot\sum_{k=0}^{\infty} z^{-2k}\left(\boldsymbol{S}_x\boldsymbol{S}_y\right)^k\right)\boldsymbol{S}_x\,, \tag{E.16}$$
$$= \frac{z}{N}\mathrm{Tr}\left(z^2\cdot\sum_{k=0}^{\infty} z^{-2k}\boldsymbol{S}_x^{1/2}\left(\boldsymbol{S}_x^{1/2}\boldsymbol{S}_y\boldsymbol{S}_x^{1/2}\right)^k\right)\boldsymbol{S}_x^{1/2}\,, \tag{E.17}$$
$$= \frac{z}{N}\mathrm{Tr}\left(z^2\cdot\sum_{k=0}^{\infty} z^{-2k}\left(\boldsymbol{S}_x^{1/2}\boldsymbol{S}_y\boldsymbol{S}_x^{1/2}\right)^k\right)\boldsymbol{S}_x\,, \tag{E.18}$$
$$= \frac{z}{N}\mathrm{Tr}\left(z^2 - \boldsymbol{S}_x^{1/2}\boldsymbol{S}_y\boldsymbol{S}_x^{1/2}\right)^{-1}\boldsymbol{S}_x\,, \tag{E.19}$$
$$= \frac{z}{N}\mathrm{Tr}\,\mathbf{G}_{\boldsymbol{S}_x^{1/2}\boldsymbol{S}_y\boldsymbol{S}_x^{1/2}}(z^2)\boldsymbol{S}_x\,, \tag{E.20}$$

and the proof of the other part follows identically. □

Next, we use the following property, which follows from subordination relation in free probability:

Lemma E.3. *Let* $t_{\boxtimes}$ *be the* T*-transform of the measure defined in Lemma A.8, then we have*

$$\frac{1}{n}\mathbb{E}\mathrm{Tr}\,\mathbf{G}_{\boldsymbol{S}_x^{1/2}\boldsymbol{S}_y\boldsymbol{S}_x^{1/2}}(z^2)\boldsymbol{S}_x \to \frac{t_{\boxtimes}(z^2)}{1+\alpha_y t_{\boxtimes}(z^2)} \tag{E.21}$$
$$\frac{1}{n}\mathbb{E}\mathrm{Tr}\,\mathbf{G}_{\boldsymbol{S}_y^{1/2}\boldsymbol{S}_x\boldsymbol{S}_y^{1/2}}(z^2)\boldsymbol{S}_y \to \frac{t_{\boxtimes}(z^2)}{1+\alpha_x t_{\boxtimes}(z^2)} \tag{E.22}$$

Proof. See for example Belinschi and Bercovici (2007). A similar derivation can be found in Chapter 19 of Potters and Bouchaud (2020). □

To conclude, one uses the limit $t = t_{\boxtimes}\alpha_x$ of Eq. (C.2) to express everything in term of the T-transform of the squared singular values and obtains the desired result.

F Proof of the identity between overlaps and resolvent (Lemma 4.3)

This follows simply by applying the result of Lemma A.2 that relates the overlap of the symmetric matrix $\eta(\tilde{\boldsymbol{S}})$ as a proper limit of the quadratic form of the resolvent with the associated vector and then using Lemma A.3 to relate the eigenvectors of $\eta(\tilde{\boldsymbol{S}})$ to the singular vectors of $\tilde{\boldsymbol{S}}$.

G Proof of Lemma 4.4

We can decompose $\tilde{\boldsymbol{H}}_k$ of Eq. (4.3) using Lemma A.1 as

$$\tilde{\boldsymbol{H}}_k = \boldsymbol{H}_k + \boldsymbol{H}_k \eta(\mathbf{A}) \big(\mathbf{I} - \boldsymbol{H}_k \eta(\mathbf{A})\big)^{-1} \boldsymbol{H}_k \,. \tag{G.1}$$

The limit as $n \to \infty$ of the RHS of Eq. (G.1) can thus be written as

$$\begin{aligned}
\boldsymbol{H}_k + \boldsymbol{H}_k \eta(\mathbf{A}) \big(\mathbf{I} - \boldsymbol{H}_k \eta(\mathbf{A})\big)^{-1} \boldsymbol{H}_k = &\begin{pmatrix} h_1 & 0 & 0 & 0 \\ 0 & h_2 & 0 & 0 \\ 0 & 0 & h_3 & 0 \\ 0 & 0 & 0 & h_4 \end{pmatrix} \\
&+ \begin{pmatrix} 0 & 0 & \sqrt{\lambda_{x,k}\lambda_{y,k}}\rho_k h_1 & \sqrt{\lambda_{x,k}} h_1 \\ 0 & 0 & \sqrt{\lambda_{y,k}} h_2 & 0 \\ \sqrt{\lambda_{x,k}\lambda_{y,k}}\rho_k h_3 & \sqrt{\lambda_{y,k}} h_3 & 0 & 0 \\ \sqrt{\lambda_{x,k}} h_4 & 0 & 0 & 0 \end{pmatrix} \\
&\begin{pmatrix} 1 & 0 & -\sqrt{\lambda_{x,k}\lambda_{y,k}}\rho_k \cdot h_1 & -\sqrt{\lambda_{x,k}} \cdot h_1 \\ 0 & 1 & -\sqrt{\lambda_{y,k}} \cdot h_2 & 0 \\ -\sqrt{\lambda_{x,k}\lambda_{y,k}}\rho_k \cdot h_3 & -\sqrt{\lambda_{y,k}} \cdot h_3 & 1 & 0 \\ -\sqrt{\lambda_{x,k}} h_4 & 0 & 0 & 1 \end{pmatrix}^{-1} \\
&\begin{pmatrix} h_1 & 0 & 0 & 0 \\ 0 & h_2 & 0 & 0 \\ 0 & 0 & h_3 & 0 \\ 0 & 0 & 0 & h_4 \end{pmatrix} + \boldsymbol{E}_n \,,
\end{aligned} \tag{G.2}$$

where $\|\boldsymbol{E}_n\|_{\mathrm{op}} \to 0$ exponentially fast. Performing the matrix inversion yields

$$\begin{pmatrix} 1 & 0 & -\sqrt{\lambda_{x,k}\lambda_{y,k}}\rho_k h_1 & -\sqrt{\lambda_{x,k}} h_1 \\ 0 & 1 & -\sqrt{\lambda_{y,k}} h_2 & 0 \\ -\sqrt{\lambda_{x,k}\lambda_{y,k}}\rho_k h_3 & -\sqrt{\lambda_{y,k}} h_3 & 1 & 0 \\ -\sqrt{\lambda_{x,k}} h_4 & 0 & 0 & 1 \end{pmatrix}^{-1} \tag{G.3}$$

$$= \frac{1}{\Delta(z)} \cdot \begin{pmatrix} -1 + \sqrt{\lambda_{x,k}\lambda_{y,k}} h_2 h_3 & -\sqrt{\lambda_{x,k}}\lambda_{y,k}\rho_k h_1 h_3 & -\sqrt{\lambda_{x,k}\lambda_{y,k}}\rho_k h_1 & \\ 0 & 1 & -\sqrt{\lambda_{y,k}} \cdot h_2 & 0 \\ -\sqrt{\lambda_{x,k}\lambda_{y,k}}\rho_k \cdot h_3 & -\sqrt{\lambda_{y,k}} \cdot h_3 & 1 & 0 \\ -\sqrt{\lambda_{x,k}} h_4 & 0 & 0 & 1 \end{pmatrix} , \tag{G.4}$$

with $\Delta(z) = \lambda_{x,k}\lambda_{y,k} j_k(z)$. Doing the matrix multiplication gives after few algebraic manipulations, the equations:

$$\tilde{H}_{11,n}(z) = h_1(z) + \frac{h_1(z)^2 \lambda_{x,k} \left(h_4(z) - h_2(z)h_3(z)h_4(z)\lambda_{y,k} + h_3(z)\lambda_{y,k}\rho^2\right)}{\Delta(z)} + \epsilon_{1,n} \,, \tag{G.5}$$

$$\tilde{H}_{22,n}(z) = h_2(z) + \frac{h_2(z)^2 h_3(z) \left(1 - h_1(z)h_4(z)\lambda_{x,k}\right) \lambda_{y,k}}{\Delta(z)} + \epsilon_{2,n} \tag{G.6}$$

$$\tilde{H}_{33,n}(z) = h_3(z) + \frac{h_3(z)^2 \lambda_{y,k} \left(h_2(z) - h_1(z)h_2(z)h_4\lambda_{x,k} + h_1(z)\lambda_{x,k}\rho^2\right)}{\Delta(z)} + \epsilon_{3,n} \,, \tag{G.7}$$

$$\tilde{H}_{44,n}(z) = h_4(z) + \frac{h_1(z)h_4(z)^2 \lambda_{x,k} \left(1 - h_2(z)h_3(z)\lambda_{y,k}\right)}{\Delta(z)} + \epsilon_{4,n} \,, \tag{G.8}$$

with $\epsilon_{i,n} \to 0$ exponentially fast. The first and third element simplify to

$$\tilde{H}_{11,n}(z) = h_1(z) + \frac{h_1(z)(1-\lambda_{y,k}h_2(z)h_3(z))}{\Delta(z)} + \epsilon_{1,n} \tag{G.9}$$

$$\tilde{H}_{33,n}(z) = h_3(z) + \frac{h_3(z)(1-\lambda_{x,k}h_4(z)h_1(z))}{\Delta(z)} + \epsilon_{3,n}\,. \tag{G.10}$$

Replacing the h_k by their expression given in Lemma 4.2 and simplifying, one may write the result as

$$\tilde{H}_{11,n}(z) \to h_1(z) + \frac{f_{1,k}(z)}{j_k(z)}\,, \tag{G.11}$$

$$\tilde{H}_{33,n}(z) \to h_3(z) + \frac{f_{3,k}(z)}{j_k(z)}\,, \tag{G.12}$$

with

$$f_{1,k}(z) = \frac{\alpha_x\lambda_{x,k}\lambda_{y,k}\rho_k^2(1+t)(\alpha_x+\alpha_y t)}{z} + \lambda_{x,k}t(\alpha_x-\lambda_{y,k}t)z\,, \tag{G.13}$$

$$f_{3,k}(z) = \frac{\alpha_y\lambda_{y,k}\lambda_{x,k}\rho_k^2(1+t)(\alpha_y+\alpha_x t)}{z} + \lambda_{y,k}t(\alpha_y-\lambda_{x,k}t)z\,. \tag{G.14}$$

H Values of the Overlaps

We first differentiate $j_k(z)$ with respect to z which yields after simplification

$$\alpha_x^2 z(j_k)'(z) = 2\left(\alpha_x - \lambda_{x,k}\,t(z^2)\right)\left(\alpha_x - \lambda_{y,k}\,t(z^2)\right) \tag{H.1}$$

$$-\,2\left(\alpha_x(\alpha_x+\alpha_y)\lambda_{x,k}\lambda_{y,k}\rho^2 + \alpha_x(\lambda_{x,k}+\lambda_{y,k})z^2 + 2\lambda_{x,k}\lambda_{y,k}(\alpha_x\alpha_y\rho^2 - z^2)\,t(z^2)\right)t'(z^2) \tag{H.2}$$

Differentiating the relation $P(z, t(z)/\alpha_x) = 0$, one finds that the derivatives of $t'(z)$ writes

$$t'(z) = \frac{\alpha_x^2 t(z)}{\alpha_x^3 + \alpha_x^2\alpha_y - \alpha_x^2 z + 3\alpha_x t(z)^2 + 3\alpha_y t(z)^2 + 6\alpha_x\alpha_y t(z)^2}\,. \tag{H.3}$$

Next we introduce

$$d(z,t) = \Big(2\left(\alpha_x - \lambda_{x,k}\,t\right)\left(\alpha_x - \lambda_{y,k}\,t\right) \tag{H.4}$$

$$+\left(-\alpha_x\lambda_{x,k}\lambda_{y,k}\rho^2\,t\left(\alpha_x+\alpha_y+2\alpha_y t\right) + t\left(-\alpha_x(\lambda_{x,k}+\lambda_{y,k}) + 2\lambda_{x,k}\lambda_{y,k}t\right)z^2\right)\Big) \tag{H.5}$$

$$/\left(\alpha_x+\alpha_y+3(\alpha_x+\alpha_y+2\alpha_x\alpha_y)t^2 - z\right)z) \tag{H.6}$$

from which one deduces that

$$\mathrm{m}_{x,k}^{\pm} = -2\frac{f_{1,k}\left(b(\mathrm{r}_k^{\pm})\right)}{d(b(\mathrm{r}_k^{\pm}),\mathrm{r}_k^{\pm})}\,, \tag{H.7}$$

$$\mathrm{m}_{y,k}^{\pm} = -2\frac{f_{3,k}\left(b(\mathrm{r}_k^{\pm})\right)}{d(b(\mathrm{r}_k^{\pm}),\mathrm{r}_k^{\pm})}\,. \tag{H.8}$$

I Optimal Angle of Rotations

Since both left singular vector $\tilde{\boldsymbol{u}}_+$ and $\tilde{\boldsymbol{u}}_-$ may correlate with the signal components $\mathbf{v}_{x,k}^\star$, one may as well construct a class of unit-norm estimator by performing a rotation in the plane of this two orthogonal vectors, the latter can be parametrized by $\beta \in (0,1)$ in the following way

$$\hat{\boldsymbol{w}}_x(\beta) := \beta\tilde{\boldsymbol{u}}_- + \sqrt{1-\beta^2}\tilde{\boldsymbol{u}}_+\,, \tag{I.1}$$

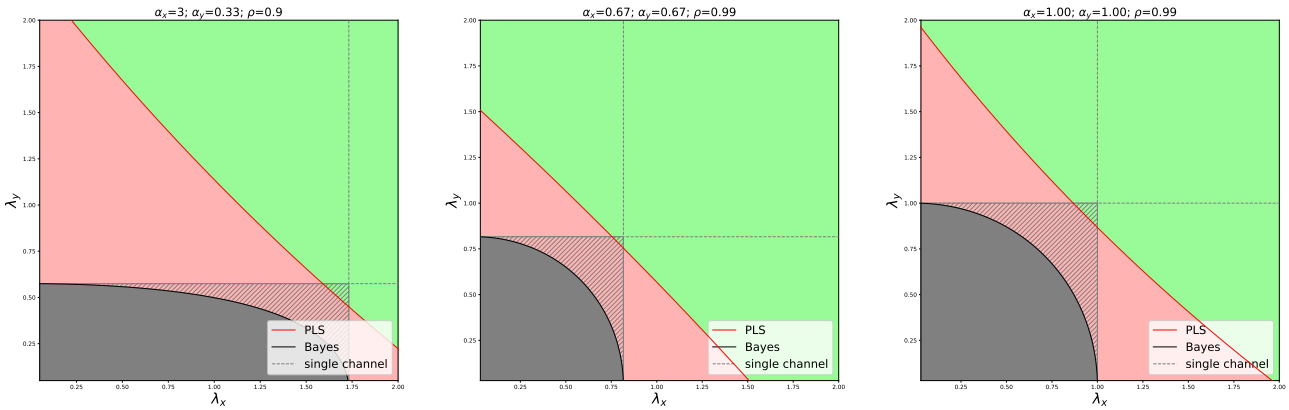
and whose (squared) overlap with $\mathbf{v}_{x,k}^\star$ is given asymptotically by

$$\langle \mathbf{v}_{x,k}^\star, \hat{\boldsymbol{w}}_x(\beta)\rangle^2 \xrightarrow[n\to\infty]{\text{a.s.}} \mathrm{q}_k(\beta) = \beta\,\mathrm{m}_{x,k}^- + \sqrt{1-\beta^2}\,\mathrm{m}_{x,k}^+\,. \tag{I.2}$$

and so to optimize with respect to β one simplify solves $\mathrm{q}_k'(\beta) = 0$ from which one finds that the optimal angle of rotation is given by

$$\beta^{\mathrm{opt}} = \frac{\mathrm{m}_{x,k}^{-}}{\sqrt{(\mathrm{m}_{x,k}^{-})^2 + (\mathrm{m}_{x,k}^{+})^2}} \,. \tag{I.3}$$

J Additional Figures



(a) (b) (c)

Figure 4: Phase diagrams in the (λ_x, λ_y) plane for different values of the aspect ratios α_x, α_y and correlation ρ. The region below each curve corresponds to values of the signal strengths for which spike detection is impossible for the corresponding algorithm.