The Annals of Statistics
2022, Vol. 50, No. 1, 53–80
https://doi.org/10.1214/21-AOS2074

HETEROSKEDASTIC PCA: ALGORITHM, OPTIMALITY, AND APPLICATIONS

BY ANRU R. ZHANG[1,2], T. TONY CAI[3] AND YIHONG WU[4]

[1]*Department of Statistics, University of Wisconsin-Madison*
[2]*Department of Biostatistics & Bioinformatics, Duke University, anru.zhang@duke.edu*
[3]*Department of Statistics, The Wharton School, University of Pennsylvania, tcai@wharton.upenn.edu*
[4]*Department of Statistics and Data Science, Yale University, yihong.wu@yale.edu*

A general framework for principal component analysis (PCA) in the presence of heteroskedastic noise is introduced. We propose an algorithm called HeteroPCA, which involves iteratively imputing the diagonal entries of the sample covariance matrix to remove estimation bias due to heteroskedasticity. This procedure is computationally efficient and provably optimal under the generalized spiked covariance model. A key technical step is a deterministic robust perturbation analysis on singular subspaces, which can be of independent interest. The effectiveness of the proposed algorithm is demonstrated in a suite of problems in high-dimensional statistics, including singular value decomposition (SVD) under heteroskedastic noise, Poisson PCA, and SVD for heteroskedastic and incomplete data.

1. Introduction. Principal component analysis (PCA) is a ubiquitous tool in statistics, econometrics, machine learning, and applied mathematics. The central aim of PCA is to extract hidden low-rank structures from noisy observations. The spiked covariance model has been well studied and used as a baseline for both methodological and theoretical developments for PCA [3, 4, 20, 30, 43, 45]. Under this model, one observes $Y_1, \ldots, Y_n \overset{\text{iid}}{\sim} N(\mu, \Sigma_0 + \sigma^2 I_p)$, where $\Sigma_0 = U\Lambda U^\top$ is a symmetric low-rank matrix and I_p is a p-dimensional identity matrix. The spiked covariance model can be equivalently written as

$$Y_k = X_k + \varepsilon_k, \qquad X_k \overset{\text{iid}}{\sim} N(\mu, \Sigma_0), \qquad \varepsilon_k \overset{\text{iid}}{\sim} N\big(0, \sigma^2 I_p\big), \quad k = 1, \ldots, n. \tag{1}$$

The goal is either to recover Σ_0, Λ, or U, and it is often done through the sample covariance matrix of $Y_1, \ldots, Y_n$, that is,

$$\widehat{\Sigma} = \frac{1}{n-1}\big(\mathbf{Y} - \bar{Y}1_n^\top\big)\big(\mathbf{Y} - \bar{Y}1_n^\top\big)^\top = \frac{1}{n-1}\sum_{k=1}^{n}(Y_k - \bar{Y})(Y_k - \bar{Y})^\top, \tag{2}$$

where $\mathbf{Y} = [Y_1, \ldots, Y_n]$ and $\bar{Y} = \frac{1}{n}\sum_{k=1}^{n} Y_k$. The asymptotic properties of eigenvalues and eigenvectors of $\widehat{\Sigma}$ have been well established in literature and their estimation based on the eigendecomposition of $\widehat{\Sigma}$ has been introduced and studied. A key assumption in these analyses is homoskedasticity, in the sense that each ε_k is assumed to be spherically symmetric Gaussian.

1.1. *Heteroskedastic PCA.* In many applications, the noise term ε_k can be highly heteroskedastic in the sense that the magnitude of noise entries varies significantly in the data

Received October 2018; revised April 2021.
MSC2020 subject classifications. Primary 62H12, 62H25; secondary 62C20.
Key words and phrases. Factor analysis model, heteroskedasticity, perturbation bound, principal component analysis, singular value decomposition.

matrix. Heteroskedastic noise is especially common in datasets with different types of variables. For example, in various biological sequencing and photon imaging data, the observations are discrete counts that are commonly modeled by Poisson, multinomial, or negative binomial distributions [14, 48] and are naturally heteroskedastic. In network analysis and recommender systems, the observations are usually binary or ordinal, which are heteroskedastic as well.

Motivated by these applications, it is natural to relax the homoskedasticity assumption in (1) and consider the following generalized spiked covariance model [2, 58]:

$$
\begin{aligned}
&Y = X + \varepsilon, \qquad \mathbb{E}X = \mu, \qquad \operatorname{Cov}(X) = \Sigma_0, \\
&\mathbb{E}\varepsilon = 0, \qquad \operatorname{Cov}(\varepsilon_j) = \sigma_j^2, \\
&\varepsilon = (\varepsilon_1, \ldots, \varepsilon_p)^\top; \quad X, \varepsilon_1, \ldots, \varepsilon_p \text{ are independent.}
\end{aligned} \tag{3}
$$

Here, Σ_0 is rank-r and admits eigendecomposition $\Sigma_0 = U\Lambda U^\top$ with $U \in \mathbb{R}^{p\times r}$ and $\Lambda \in \mathbb{R}^{r\times r}$. $\sigma_1^2, \ldots, \sigma_p^2$ are unknown and not necessarily identical. This model is also widely used as the standard model in the literature of *factor analysis* (see, e.g., [24, 50] and the references therein). Given i.i.d. copies $Y_1, \ldots, Y_n$ drawn from (3) and the rank r, the goal is to estimate U.

Performing the classical PCA on data with heteroskedastic noise can often lead to inconsistent estimates. The estimation of U using the classical PCA is equivalent to the estimation of eigenvectors of the sample covariance matrix $\widehat{\Sigma}$. Since $\mathbb{E}\widehat{\Sigma} = \Sigma_0 + \operatorname{diag}(\sigma_1^2, \ldots, \sigma_p^2)$, the top eigenvectors of $\mathbb{E}\widehat{\Sigma}$ and Σ_0 will coincide when $\sigma_1^2, \ldots, \sigma_p^2$ are the same. But in the case of heteroskedastic noise, the differences in the bias terms $\sigma_1^2, \ldots, \sigma_p^2$ can lead to significant difference between the principal components of $\mathbb{E}\widehat{\Sigma}$ and those of Σ_0. Similar phenomena appear in other problems with heteroskedastic noise (see Section 3 for details).

To cope with the bias on the diagonal elements of covariance matrix, [21] introduced the *diagonal-deletion SVD* in the context of bipartite stochastic block model. The idea is to set the diagonal of the sample covariance matrix to zero before performing singular value decomposition. However, it is a priori unclear whether zeroing out the diagonals is always the best choice, because it may change the singular subspace entirely.

In this paper, we introduce *HeteroPCA*, a novel method for heteroskedastic principal component analysis. Instead of zeroing out the diagonal entries of the sample covariance/Gram matrix, we propose to iteratively update the diagonal entries based on the off-diagonals, so that the bias incurred on the diagonal is significantly reduced and more accurate estimation can be achieved. The performance of the proposed procedure is studied both theoretically and numerically. By establishing matching minimax upper and lower bounds, we show that HeteroPCA achieves the optimal rate of convergence for a range of settings under the generalized spiked covariance model.

Classic perturbation bounds, such as Davis–Kahan and Wedin's theorems [17, 57], play key roles in the theoretical analysis of various PCA methods. These tools may not be suitable for the analysis of heteroskedastic PCA due to the aforementioned bias on the diagonal entries of the sample covariance matrix. To tackle this difficulty, we develop a new deterministic subspace perturbation bound (Theorem 3), which provides the key technical tool for analyzing HeteroPCA and may be of independent interest.

In addition to heteroskedastic PCA for the generalized spiked covariance model, the proposed HeteroPCA algorithm is applicable to a collection of high-dimensional problems with heteroskedastic data. Several applications are discussed in detail in Section 3, including *SVD under heteroskedastic noise*, *Poisson PCA*, and *SVD for heteroskedastic and incomplete data*. Our results can also be useful in heteroskedastic canonical correlation analysis, heteroskedastic tensor SVD, exponential family PCA, and community detection in bipartite stochastic network.

1.2. *Related literature.* [2, 58] extended the theory for regular spiked covariance model (1) to the generalized spiked covariance model and studied the limiting distribution of eigenvalues of the sample covariance matrix. [25–27] introduced an alternative model for heteroskedastic data, where the noise is nonuniform *across* different samples but uniform *within* each sample. Under this model, [25, 26] studied the asymptotic performance of PCA and [27] developed the optimal weights for weighted PCA with theoretical guarantees. [51–54] provide a comprehensive study on the methodology and theory of PCA where data-dependent, nonisotropic, or correlated noise and missing values may appear. The detailed comparison of our results and [53, 54] are given in later Remarks 2 and 9.

Our work is also closely related to a substantial body of literature on factor model analysis [1, 24, 34, 44, 49, 50, 56]. There have been various approaches developed to estimate the principal components in factor models, such as the regression method [49], weighted least squares [5], EM [50], and Bayesian MCMC [24]. The asymptotic theory for factor model analysis was also extensively studied (e.g., [1, 56] and the references therein). Different from the previous results, this paper mainly concerns a nonasymptotic framework, providing algorithms with provable guarantees and allowing heteroskedastic noise within each sample in the high-dimensional regime that n, p, r can all grow.

Matrix denoising [19, 20, 23, 42], where the central goal is to estimate low-rank matrices from noisy observations, is closely related to this work. In order to get an accurate estimation of overall low-rank matrices from observations perturbed by random noise, the singular value thresholding [15, 19] and the singular value shrinkage [20, 23, 42] were proposed and widely studied recently. Departing from these previous results, this paper focuses on estimating the singular subspace instead of the overall matrix, which achieves better performance in singular subspace estimation than denoising the whole matrix by previous methods and then performing a rank-r SVD.

In Section 3.3, we discuss the application of HeteroPCA to SVD for heteroskedastic and incomplete data. This problem is related to a body of literature on matrix completion that we will give a review in Section 3.3.

1.3. *Organization of the paper.* After a brief introduction of notation and definitions (Section 2.1), we focus on the generalized spiked covariance model, present the HeteroPCA algorithm (Section 2.2), and develop matching minimax upper and lower bounds of the estimation error (Section 2.3). Then, we introduce a deterministic robust perturbation analysis that serves as a key technical step in our analysis (Section 2.4). We also illustrate main proof ideas in Section 2.5. In Section 3, we discuss the applications of established results. Numerical results are given in Section 4. The proofs of main results are given in Section 6. The additional proofs and technical lemmas are provided in the Supplementary Material [60].

2. Optimal heteroskedastic principal component analysis.

2.1. *Notation and preliminaries.* We use lowercase letters, for example, x, y, to denote scalars or vectors and use uppercase letters, for example, U, M to denote matrices. For sequences of positive numbers $\{a_k\}$ and $\{b_k\}$, we write $a_k \lesssim b_k$ or $b_k \gtrsim a_k$ if there exists a uniform constant $C > 0$ such that $a_k \le Cb_k$ for all k. We also write $a_k \asymp b_k$ if $a_k \lesssim b_k$ and $a_k \gtrsim b_k$ both hold. For any matrix $M \in \mathbb{R}^{p_1 \times p_2}$, let $\lambda_k(M)$ be the k-th largest singular value. Then, the SVD of M can be written as $M = \sum_{k=1}^{p_1 \wedge p_2} \lambda_k(M) u_k v_k^\top$. Let $\mathrm{SVD}_r(M) = [u_1 \cdots u_r]$ be the collection of leading r left singular vectors and $\mathrm{QR}(M)$ be the Q part of QR orthogonalization of M. The matrix spectral norm and Frobenius norm are defined as $\|M\| = \sup_{\|u\|_2=1} \|Mu\|_2 = \lambda_1(M)$ and $\|M\|_F = (\sum_{i,j} M_{ij}^2)^{1/2} = (\sum_k \lambda_k^2(M))^{1/2}$. Let I_r, $0_{m\times n}$, and $1_{m\times n}$ be the r-by-r identity, $m \times n$ zero, and $m \times n$ all-one matrices, respectively. Also let 0_m and 1_m denote the m-dimensional zero and all-one column vectors.

Denote $\mathbb{O}_{p,r} = \{U \in \mathbb{R}^{p\times r} : U^\top U = I_r\}$ as the set of all p-by-r matrices with orthonormal columns. For $U \in \mathbb{O}_{p,r}$, we note $U_\perp \in \mathbb{O}_{p,p-r}$ as the orthogonal complement so that $[UU_\perp] \in \mathbb{R}^{p\times p}$ is a complete orthogonal matrix.

Motivated by incoherence condition, a widely used assumption in the matrix completion literature [11], we define *incoherence constant* of $U \in \mathbb{O}_{p,r}$ as

$$I(U) = (p/r)\max_{i\in[p]} \|e_i^\top U\|_2^2. \tag{4}$$

For any $U_1, U_2 \in \mathbb{O}_{p,r}$, we define the $\sin\Theta$ distance $\|\sin\Theta(U_1, U_2)\| \triangleq \|U_{1\perp}^\top U_2\| = \|U_{2\perp}^\top U_1\|$. For any square matrix A, let $\Delta(A)$ be A with all diagonal entries set to zero and $D(A)$ be A with all off-diagonal entries set to zero. Then $A = \Delta(A) + D(A)$. We define the Orlicz-ϕ_2 norm of any random variable Y as $\|Y\|_{\psi_2} = \sup_{q\geq 1} q^{-1/2}(\mathbb{E}|Y|^q)^{1/q}$. A random variable Y is called σ^2-sub-Gaussian if $\|Y/\sigma\|_{\psi_2} \leq C$ for some constant $C > 0$; a random vector X is called Σ-sub-Gaussian if $\max_{q\geq 1, v\in\mathbb{R}^p} \|v^\top \Lambda^{-1/2} U^\top Y\|_{\psi_2} \leq C$ (here, $\Sigma = U\Lambda U^\top$ is the eigendecomposition of Σ) for some constant $C > 0$. We use $C, C_1, \ldots, c, c_1, \ldots$ to respectively represent generic large and small constants, whose values may differ in different lines.

2.2. *Methods for heteroskedastic PCA.* Suppose one observes i.i.d. copies $Y_1, \ldots, Y_n$ of Y from the generalized spiked covariance model (3). Let $\widehat{\Sigma}$ be the sample covariance matrix defined as (2). The regular SVD estimator $\widetilde{U} = \mathrm{SVD}_r(\widehat{\Sigma})$, that is, the leading r left singular vectors of $\widehat{\Sigma}$, is the natural estimator of U, the leading singular vectors of Σ_0. An important variant of Davis–Kahan's theorem [17] given by Yu, Wang, and Samworth [59] yields

$$\|\sin\Theta(\widetilde{U}, U)\| \lesssim \frac{\|\widehat{\Sigma} - (\Sigma_0 + \beta I_p)\|}{\lambda_r(\Lambda)} \wedge 1, \tag{5}$$

which holds for any scalar $\beta \geq 0$ and cannot be improved in general. As briefly discussed earlier, since $\mathbb{E}\widehat{\Sigma} = \Sigma_0 + \mathrm{diag}(\sigma_1^2, \ldots, \sigma_p^2)$, when $\sigma_1^2, \ldots, \sigma_p^2$ have different values, the diagonal entries of the perturbation matrix $\widehat{\Sigma} - (\Sigma_0 + \beta I_p)$ may be significantly larger than the rest. As a result, $\widetilde{U}$ can be a suboptimal estimator for U.

To achieve a more accurate estimate of U, we propose the following Algorithm 1 named HeteroPCA. The central idea is to iteratively impute the diagonal entries of the sample covariance matrix $\widehat{\Sigma}$ by the diagonals of its low-rank approximation. In Algorithm 1, since $\widehat{\Sigma}$, $N^{(t)}$ are symmetric, we have $u_i^{(t)} = v_i^{(t)}$ or $-v_i^{(t)}$. In contrast to most previous work on matrix completion and robust PCA, where the entries to be imputed are missing at random, here our goal is to impute the diagonal entries. Moreover, HeteroPCA can be interpreted as the projection gradient descent (PGD) for the following rank-constrained (nonconvex) optimization problem:

$$\min_{\mathrm{rank}(\tilde{N})\leq r} \|\Delta(\hat{\Sigma} - \tilde{N})\|_F^2. \tag{7}$$

To see this connection, we first note that $\tilde{N}^{(t)}$ is the best rank-r approximation of $N^{(t)}$, which correspond to the projection step in PGD; next, $\nabla_{\tilde{N}}\|\Delta(\hat{\Sigma} - \tilde{N})\|_F^2 = 2\Delta(\tilde{N} - \hat{\Sigma})$ and the operator norm of $\nabla_{\tilde{N}}^2\|\Delta(\hat{\Sigma} - \tilde{N})\|_F^2$ is 2, where $\nabla_{\tilde{N}}$ and $\nabla_{\tilde{N}}^2$ are the gradient and Hessian with respect to $\tilde{N}$, respectively. Based on the theory of PGD (see, e.g., Section 3.3 in [6]), the smoothness parameter of the loss function $\beta = 2$ and the update,

$$\begin{aligned} N^{(t+1)} &= \tilde{N}^{(t)} - (1/\beta)\nabla_{\tilde{N}^{(t)}}\|\Delta(\hat{\Sigma} - \tilde{N}^{(t)})\|_F^2 = \tilde{N}^{(t)} - \Delta(\tilde{N}^{(t)} - \hat{\Sigma}) \\ &= D(\tilde{N}^{(t)}) + \Delta(\hat{\Sigma}) = D(\tilde{N}^{(t)}) + \Delta(N^{(t)}), \end{aligned}$$

Algorithm 1 HeteroPCA

Input: matrix $\widehat{\Sigma}$, rank r, maximum number of iterations T.
Initialize by setting the diagonal of $\widehat{\Sigma}$ to zero: $N^{(0)} = \Delta(\widehat{\Sigma})$, $t = 0$.
repeat
    Perform SVD on $N^{(t)}$ and let $\widetilde{N}^{(t)}$ be its best rank-r approximation:

$$N^{(t)} = U^{(t)}\Sigma^{(t)}\big(V^{(t)}\big)^\top = \sum_i \lambda_i^{(t)} u_i^{(t)} \big(v_i^{(t)}\big)^\top, \lambda_1^{(t)} \geq \cdots \geq \lambda_m^{(t)} \geq 0,$$

$$\widetilde{N}^{(t)} = \sum_{i=1}^{r} \lambda_i^{(t)} u_i^{(t)} \big(v_i^{(t)}\big)^\top.$$

    Update $N^{(t+1)} = D(\widetilde{N}^{(t)}) + \Delta(N^{(t)})$, that is, replace the diagonal entries of $N^{(t)}$ by those in $\widetilde{N}^{(t)}$:

$$N_{ij}^{(t+1)} = \begin{cases} N_{ij}^{(t)} = \widetilde{N}_{ij}^{(t)}, & i = j; \\ \widehat{\Sigma}_{ij}, & i \neq j. \end{cases} \tag{6}$$

    $t = t + 1$.
until convergence or maximum number of iterations reached.
Output: $\widehat{U} = U^{(T)} = [u_1^{(T)} \cdots u_r^{(T)}]$.

corresponds to the gradient descent step in PGD. Due to the nonconvexity of (7), existing convergence results for PGD do not apply to Algorithm 1, for which we provide a direct analysis next.

2.3. *Theoretical analysis*. Denote

$$\sigma_{\max}^2 \triangleq \max_i \sigma_i^2, \qquad \sigma_{\text{sum}}^2 \triangleq \sum_i \sigma_i^2.$$

Recall that $\Sigma_0 = U\Lambda U^\top$ is rank-r, so $\lambda_r(\Lambda)$ is the smallest nonzero eigenvalue of Σ_0. We have the following theoretical guarantee for Algorithm 1.

THEOREM 1 (Heteroskedastic PCA: upper bound). *Consider the generalized spiked covariance model* (3), *where* X_i *and* ε_i *are* Σ*-sub-Gaussian and* σ_i*-Gaussian, respectively. Let* $Y_1, \ldots, Y_n$ *be i.i.d. samples from* (3). *Assume* $n \geq (C_0 r) \wedge C_0 \log(\sigma_r(\Lambda)/\sigma_{\text{sum}}^2)$ *and* $\|\Lambda\|/\lambda_r(\Lambda) \leq C_0$ *for constant* $C_0 > 0$. *There exists constant* $c_I > 0$ *such that if the incoherence constant* $I(U)$ (*defined in* (4)) *satisfies* $I(U) \leq c_I p/r$, *then the output* $\widehat{U}$ *of Algorithm 1 applied to the sample covariance matrix* $\widehat{\Sigma}$ *with the number of iterations* $T = \Omega(\log(n\lambda_r(\Lambda)/\sigma_{\text{sum}}^2) \vee 1)$ *satisfies*

$$\mathbb{E}\|\sin\Theta(\widehat{U}, U)\| \leq \frac{C}{\sqrt{n}}\left(\frac{\sigma_{\text{sum}} + r^{1/2}\sigma_{\max}}{\lambda_r^{1/2}(\Lambda)} + \frac{\sigma_{\text{sum}}\sigma_{\max}}{\lambda_r(\Lambda)}\right) \wedge 1. \tag{8}$$

Here, the constant C *relies on* c_I, C_0, *but not* X, σ_i, p, r, n.

REMARK 1. Let $\widetilde{p} = \sigma_{\text{sum}}^2/\sigma_{\max}^2$. Then (8) can be rewritten as

$$\mathbb{E}\|\sin\Theta(\widehat{U}, U)\| \lesssim \left(\sqrt{\frac{\widetilde{p} \vee r}{n}}\frac{\sigma_{\max}}{\lambda_r^{1/2}(\Lambda)} + \sqrt{\frac{\widetilde{p}}{n}}\frac{\sigma_{\max}^2}{\lambda_r(\Lambda)}\right) \wedge 1. \tag{9}$$

Consider the homoskedastic PCA setting where $\sigma_1^2 = \cdots = \sigma_p^2 = \sigma_{\max}^2$. This special case of Theorem 1 yields

$$\mathbb{E}\|\sin\Theta(\widehat{U}, U)\| \lesssim \sqrt{\frac{p}{n}}\left(\frac{\sigma_{\max}}{\lambda_r^{1/2}(\Lambda)} + \frac{\sigma_{\max}^2}{\lambda_r(\Lambda)}\right) \wedge 1. \tag{10}$$

Comparing (9) with (10), we see that a weighted average between $\widetilde{p} \vee r$ and $\widetilde{p}$ can be viewed as the "effective dimension" for heteroskedastic PCA.

REMARK 2. Recently, [52, 54] studied the PCA for matrix data with nonisotropic and data-dependent noise. In our notation and with some mild regularity conditions, [54], Part 2, Corollary 2.7, shows that the regular SVD estimator $\widetilde{U} = \mathrm{SVD}_r(Y)$ satisfies

$$\|\sin\Theta(\widetilde{U}, U)\| \lesssim \left(\sqrt{\frac{p}{n}}\frac{\sigma_{\max}}{\lambda_r^{1/2}(\Lambda)} + \sqrt{\frac{p}{n}}\frac{\sigma_{\max}^2}{\lambda_r(\Lambda)} + \frac{\|U_\perp^\top \Sigma_\varepsilon U\|}{\lambda_{\min}(\Lambda)}\right) \wedge 1 \tag{11}$$

with high probability. Here, Σ_ε is the covariance matrix of the noise vector ε_k. Since $\sigma_{\mathrm{sum}} \le \sqrt{p}\sigma_{\max}$ and $\|U_\perp^\top \Sigma_\varepsilon U\| \ge 0$, our Theorem 1 yields a better estimation error rate.

Next, we establish the optimality of Theorem 1. Consider the following class of generalized spiked covariance matrices:

$$\begin{aligned}
&\mathcal{F}_{p,n,r}(\check{\sigma}_{\mathrm{sum}}, \check{\sigma}_{\max}, \nu, \kappa)\\
&\quad= \Big\{\Sigma = U\Lambda U^\top + D:\\
&\qquad D \text{ is nonnegative diagonal}, \textstyle\sum_i D_{ii} \le \check{\sigma}_{\mathrm{sum}}^2, \max_i D_{ii} \le \check{\sigma}_{\max}^2,\\
&\qquad U \in \mathbb{O}_{p,r}, I(U) \le c_I p/r, \|\Lambda\|/\lambda_r(\Lambda) \le \kappa, \lambda_r(\Lambda) \ge \nu \Big\}.
\end{aligned} \tag{12}$$

THEOREM 2 (Heteroskedastic PCA: lower bound). *Suppose* $\sqrt{p}\check{\sigma}_{\max} \ge \check{\sigma}_{\mathrm{sum}} \ge \check{\sigma}_{\max} > 0$, $\kappa \ge 1$. *There exists constant* $C > 0$, *such that if* $p \ge Cr$, *we have*

$$\begin{aligned}
&\inf_{\widehat{U}} \sup_{\Sigma \in \mathcal{F}_{p,n,r}(\check{\sigma}_{\mathrm{sum}}, \check{\sigma}_{\max}, \nu, \kappa)} \mathbb{E}\|\sin\Theta(\widehat{U}, U)\|\\
&\quad\gtrsim \frac{1}{\sqrt{n}}\left(\frac{\check{\sigma}_{\mathrm{sum}} + r^{1/2}\check{\sigma}_{\max}}{\nu^{1/2}} + \frac{\check{\sigma}_{\mathrm{sum}}\check{\sigma}_{\max}}{\nu}\right) \wedge 1.
\end{aligned} \tag{13}$$

REMARK 3. By combining Theorems 1 and 2, the proposed Algorithm 1 achieves the following optimal estimation error rate in $\mathcal{F}_{p,n,r}(\check{\sigma}_{\mathrm{sum}}, \check{\sigma}_{\max}, \nu, \kappa)$ when the condition number κ is a constant:

$$\inf_{\widehat{U}} \sup_{\Sigma \in \mathcal{F}_{p,n,r}(\check{\sigma}_{\mathrm{sum}}, \check{\sigma}_{\max}, \nu, C)} \mathbb{E}\|\sin\Theta(\widehat{U}, U)\| \asymp \frac{1}{\sqrt{n}}\left(\frac{\check{\sigma}_{\mathrm{sum}} + r^{1/2}\check{\sigma}_{\max}}{\nu^{1/2}} + \frac{\check{\sigma}_{\mathrm{sum}}\check{\sigma}_{\max}}{\nu}\right) \wedge 1.$$

Next, we consider the performance of HeteroPCA if the covariance matrix Σ_0 is approximately low-rank.

PROPOSITION 1 (HeteroPCA for approximately low-rank covariance). *Consider the generalized spiked covariance model* (1). *Suppose* $\Sigma_0 = \widetilde{U}\Lambda\widetilde{U}^\top$ *is the eigenvalue decomposition, where* $\widetilde{U} = [UU_\perp]$ *and* U *is the collection of leading* r *singular vectors. Assume* X *and* ε_i *are* Σ*-sub-Gaussian and* σ_i^2*-sub-Gaussian, respectively. Also assume that* $n \ge Cr$, $n \wedge p \ge C(\sigma_{\mathrm{sum}}^2/\sigma_r(\Lambda))$, *and* $\|\Lambda\|/\lambda_r(\Lambda) \le C$ *for some constant* $C > 0$. *Then there exists*

some constant $c_I > 0$ such that if the incoherence constant $I(U)$ (defined in (4)) satisfies $I(U) \le c_I p/r$, then the output $\widehat{U}$ of Algorithm 1 with the input matrix $\widehat{\Sigma}$ and number of iterations $T = \Omega(\log(n\lambda_r(\Lambda)/\sigma_{\mathrm{sum}}^2) \vee 1)$ satisfies

$$\mathbb{E}\|\sin\Theta(\widehat{U}, U)\| \lesssim \left(\frac{\sigma_{\mathrm{sum}} + \sqrt{r}\sigma_{\max}}{n^{1/2}\lambda_r^{1/2}(\Lambda)} + \frac{\sigma_{\mathrm{sum}}\sigma_{\max}}{n^{1/2}\lambda_r(\Lambda)} + \frac{((np)^{1/2} + p)\lambda_{r+1}^{1/2}(\Lambda)}{n\lambda_r^{1/2}(\Lambda)} + \frac{\lambda_{r+1}(\Lambda)}{\lambda_r(\Lambda)}\right) \wedge 1.$$

Proposition 1 shows that HeteroPCA can estimate the loading matrix U accurately if there exists a significant gap between $\lambda_r(\Sigma_0)$ and $\lambda_{r+1}(\Sigma_0)$.

2.4. *A deterministic robust perturbation analysis.* In this section, we temporarily ignore the randomness of X_ks and ε_ks and focus on a more general prototypical model of the heteroskedastic PCA problem in Section 1.1. Let N, M, Z be deterministic symmetric matrices (not necessarily positive definite) that satisfy

$$N = M + Z \in \mathbb{R}^{p\times p}. \tag{14}$$

Here N is the observation, M is the rank-r matrix of interest, and $Z \in \mathbb{R}^{p\times p}$ is the perturbation that possibly has significantly large amplitude in its diagonal entries. In the heteroskedastic PCA model, N, M, Z may represent the sample covariance matrix $\widehat{\Sigma}$, population covariance matrix Σ_0, and their difference, respectively. Let $U \in \mathbb{O}_{p,r}$ be the first r singular vectors of M. As discussed earlier, applying the proposed HeteroPCA (Algorithm 1) to matrix N provides an adaptive estimate of U. In the following theorem, we demonstrate the theoretical property for the proposed Algorithm 1 under the general robust perturbation model (14).

THEOREM 3 (Robust $\sin\Theta$ theorem). *Suppose $M \in \mathbb{R}^{p\times p}$ is a rank-r symmetric matrix and $U \in \mathbb{O}_{p,r}$ consists of the eigenvectors of M. Let $\widehat{U}^{(t)} = [u_1^{(t)} \cdots u_r^{(t)}]$ be the intermediate result of Algorithm 1 with input matrix N after t iterations. There exists a universal constant $c_I > 0$ such that if*

$$I(U)\|M\|/\lambda_r(M) \le c_I p/r, \tag{15}$$

where $I(U)$ is the incoherence constant defined in (4), then

$$\|\sin\Theta(\widehat{U}^{(t)}, U)\| \le \frac{4\|\Delta(Z)\|}{\lambda_r(M)} + 2^{-(t+3)}.$$

In particular if $T = \Omega(\log\frac{\lambda_r(M)}{\eta\|\Delta(Z)\|} \vee 1)$, the final outcome $\widehat{U}$ satisfies

$$\|\sin\Theta(\widehat{U}, U)\| \lesssim \frac{\|\Delta(Z)\|}{\lambda_r(M)} \wedge 1. \tag{16}$$

A matching lower bound and several discussions to the robust $\sin\Theta$ theorem is given in Section 1 in the Supplementary Material.

REMARK 4. Distinct from the matrix completion, where most entries are missing from the target matrix, the substantial corrupted entries only lie in the diagonal of the target Gram matrix/sample covariance matrix in our problem. Thus, a much looser condition on incoherence, $I(U) < cp/r$, is sufficient compared to the one required by matrix completion, $I(U) < \mu$ with μ being a constant.

2.5. *Proof sketches of main technical results.* We briefly discuss the proofs of Theorems 1, 2, and 3 in this section.

The proof of Theorem 1 consists of three main steps. First, we define $\widehat{\Sigma}_X$ as the sample covariance matrix of signal vectors $X_1, \ldots, X_n$ and $E = [\varepsilon_1 \cdots \varepsilon_n]$ as the noise matrix. We aim to develop a concentration inequality for $\Delta((n-1)(\widehat{\Sigma} - \widehat{\Sigma}_X))$, that is, the off-diagonal part of the perturbation. To this end, we decompose $(n-1)(\widehat{\Sigma} - \widehat{\Sigma}_X)$ into $(XE^\top + EX^\top)$, $(EE^\top)$, $n(\bar{X}\bar{E}^\top + \bar{E}\bar{X}^\top + \bar{E}\bar{E}^\top)$, then bound them separately by heteroskedastic Wishart concentration inequality [8] and Lemma 1 in the Supplementary Material. Second, we develop a lower bound for $\lambda_r(\widehat{\Sigma}_X)$, that is, the least nontrivial singular value of the signal covariance matrix. Finally, we apply the robust $\sin\Theta$ theorem (Theorem 3), to complete the proof.

To show the lower bound in Theorem 2, it suffices to show the two terms in (13) separately; c.f., (6) and (7) in the detailed proof. To show each individual lower bound, we construct a series of "candidate matrices" $\{U^{(k)}, \Sigma^{(k)}\}_{k=1}^N$ in $\mathcal{F}_{p,n,r}(\sigma_{\text{sum}}, \sigma_{\max}, \nu)$ so that $\{U^{(k)}\}_{k=1}^N$ are well separated while distinguishing them apart based on random sample $Y_1, \ldots, Y_n \sim N(0, \Sigma^{(k)})$ is impossible. This implies the desired lower bound by applying Fano's method.

The proof of Theorem 3 is the main technical contribution of this paper. Specifically, we analyze how the estimation error $K_t = \|N^{(t)} - M\|$ decays at each iteration. We first obtain an initialization error bound. Then for each t, we decompose K_t into four terms, bound them separately, and obtain an inequality that relates K_t to K_{t-1} (see (43)). By induction, this recursive inequality leads to the exponential decay of K_t and implies the desired upper bound. Note that Algorithm 1 can be viewed as successive compositions involving the projection operator $P_U(\cdot)$ and the diagonal-deletion operator $D(\cdot)$. We thus introduce Lemma 1 to give sharp operator norm upper bounds for compositions of $P_U(\cdot)$ and $D(\cdot)$. At the heart of the proof of Theorem 3, this lemma is useful for bounding the error at both the initialization and the subsequent iterations.

3. Further applications in high-dimensional statistics.

3.1. *SVD under heteroskedastic noise.* Suppose one observes

$$Y = X + E, \tag{17}$$

where X is the low-rank matrix of interest and the entries of noise E are independent, zero-mean, but not necessarily homoskedastic. The goal is to recover the left singular subspace of X based on noisy observation Y. The problem arises naturally in a range of applications, such as magnetic resonance imaging (MRI) and relaxometry [12]. This model can also be viewed as a prototype of various problems in high-dimensional statistics and machine learning, including Poisson PCA [48], bipartite stochastic block model [21], and exponential family PCA [35]. Let the sample and population Gram matrices be $N = YY^\top$ and $M = XX^\top$, respectively. Then,

$$(\mathbb{E}N)_{ij} = \begin{cases} M_{ij}, & i \neq j; \\ M_{ij} + \sum_{k=1}^{p_2} \operatorname{Var}(E_{ik}), & i = j. \end{cases}$$

Thus, only the off-diagonal entries of N are unbiased estimators of the corresponding entries of M. When $\operatorname{Var}(E_{ij})$ are unequal, there can be significant differences between the spectrum of $\mathbb{E}N$, $\mathbb{E}\Delta(N)$, and M. Since left singular vectors of Y and X are respectively, identical to those of N and M, the regular SVD or diagonal-deletion SVD on Y can result in inconsistent estimates of the left singular subspace of X.

Compared to the regular or diagonal-deletion SVD, the next theorem shows the proposed HeteroPCA can be a better approach.

THEOREM 4. *Consider the model* (17). *Suppose* $X \in \mathbb{R}^{p_1 \times p_2}$ *is a fixed rank-*r *matrix, the noise matrix* E *has independent entries,* $\mathbb{E}E_{ij} = 0$, $\mathrm{Var}(E_{ij}) = \sigma_{ij}^2$, *and* E_{ji} *is* σ_{ij}^2*-sub-Gaussian. Suppose the left singular subspace of* X *is* $U \in \mathbb{O}_{p_1,r}$. *Assume that the condition number of* X *is at most some absolute constant* C, *that is,* $\|X\| \le C\lambda_r(X)$. *Denote*

$$\sigma_R^2 = \max_i \sum_{j=1}^{p_2} \sigma_{ij}^2, \qquad \sigma_C^2 = \max_j \sum_{i=1}^{p_1} \sigma_{ij}^2, \qquad \sigma_{\max}^2 = \max_{ij} \sigma_{ij}^2 \tag{18}$$

as the rowwise, columnwise, and entrywise noise variances. Then there exists a constant $c_I > 0$ *such that if* U *satisfies* $I(U) = \max_{1 \le i \le p_1} \frac{p_1}{r} \|e_i^\top U\|_2^2 \le c_I p_1/r$, *Algorithm* 1 *applied to* $YY^\top$ *with rank* r *and number of iterations* $T = \Omega(\log(\lambda_r(X)/\sigma_C) \vee 1)$ *outputs* $\widehat{U}$ *that satisfies*

$$\begin{aligned} &\mathbb{E}\|\sin\Theta(\widehat{U}, U)\| \\ &\lesssim \left(\frac{\sigma_C + \sqrt{r}\sigma_{\max}}{\lambda_r(X)} + \frac{\sigma_R\sigma_C + \sigma_R\sigma_{\max}\sqrt{\log(p_1 \wedge p_2)} + \sigma_{\max}^2 \log(p_1 \wedge p_2)}{\lambda_r^2(X)}\right) \wedge 1. \end{aligned} \tag{19}$$

If $\sigma_{\max} \lesssim \sigma_C / \max\{\sqrt{r}, \sqrt{\log(p_1 \wedge p_2)}\}$ *additionally holds, that is, the variance array* $\{\sigma_{ij}^2\}$ *is not too "spiky," we further have*

$$\mathbb{E}\|\sin\Theta(\widehat{U}, U)\| \lesssim \left(\frac{\sigma_C}{\lambda_r(X)} + \frac{\sigma_R\sigma_C}{\lambda_r^2(X)}\right) \wedge 1. \tag{20}$$

REMARK 5. When $\sigma_{ij} = \sigma_{\max}$ for all i, j, (19) reduces to

$$\mathbb{E}\|\sin\Theta(\widehat{U}, U)\| \lesssim \left(\frac{\sqrt{p_1}\sigma_{\max}}{\lambda_r(X)} + \frac{\sqrt{p_1 p_2}\sigma_{\max}}{\lambda_r^2(X)}\right),$$

which matches the optimal rate for SVD under homoskedastic noise in the literature [9], Theorems 3 and 4.

REMARK 6. In contrast to the scaling of $\lambda_r(\Lambda)$ in Theorems 1 and 2, the scale of $\lambda_r(X)$ in Theorem 4 implicitly grows with both p_1 and p_2 as X is a p_1-by-p_2 matrix.

3.2. *Poisson PCA.* Poisson PCA [35] is an important problem in statistics and engineering with a range of applications, including photon-limited imaging [48] and biological sequencing data analysis [14]. Suppose we observe $Y \in \mathbb{R}^{p_1 \times p_2}$, where $Y_{ij} \overset{\text{ind}}{\sim} \text{Poisson}(X_{ij})$ and $X \in \mathbb{R}^{p_1 \times p_2}$ is rank-r. Let $X = U\Lambda V^\top$ be the singular value decomposition, where $U \in \mathbb{O}_{p_1,r}$, $V \in \mathbb{O}_{p_2,r}$. The goal is to estimate the leading singular vectors of X, that is, U or V, based on Y. HeteroPCA is an appropriate method for Poisson PCA since it can well handle the heteroskedasticity of Poisson distribution. Although the aforementioned heteroskedastic low-rank matrix denoising can be seen as a prototype problem of Poisson PCA, Theorem 4 is not directly applicable and more careful analysis is needed since the Poisson distribution has heavier tail than sub-Gaussian.

THEOREM 5 (Poisson PCA). *Suppose* X *is a nonnegative* p_1*-by-*p_2 *matrix,* $\mathrm{rank}(X) = r$, $\lambda_1(X)/\lambda_r(X) \le C$, $X_{ij} \ge c$ *for constant* $c > 0$, $U \in \mathbb{O}_{p_1,r}$ *is the left singular subspace of* X. *Denote*

$$\sigma_R^2 = \max_i \sum_{j=1}^{p_2} X_{ij}, \qquad \sigma_C^2 = \max_j \sum_{i=1}^{p_1} X_{ij}, \qquad \sigma_*^2 = \max_{i,j} X_{ij}. \tag{21}$$

Suppose one observes $Y \in \mathbb{R}^{p_1 \times p_2}$, $Y_{ij} \overset{\text{ind}}{\sim} \text{Poisson}(X_{ij})$. *Then there exists constant* $c_I > 0$ *such that if* U *satisfies* $I(U) = \max_i \frac{p_1}{r} \|e_i^\top U\|_2^2 \le c_I p_1/r$, *the proposed HeteroPCA procedure* (*Algorithm* 1) *on matrix* $YY^\top$ *with rank* r *and number of iterations* $T = \Omega(\log(\lambda_r(X)/\sigma_C) \vee 1)$ *yields*

$$\mathbb{E}\|\sin\Theta(\widehat{U}, U)\| \lesssim \left(\frac{\sigma_C + r\sigma_{\max}}{\lambda_r(X)} + \frac{\{\sigma_R + \sigma_C + \sigma_{\max}\sqrt{\log(p_2)\log(p_1)}\}^2 - \sigma_R^2}{\lambda_r^2(X)}\right) \wedge 1. \tag{22}$$

In addition, if $\sigma_{\max} \le \sigma_C / \max\{r, \sqrt{\log(p_1)\log(p_2)}\}$, *then*

$$\mathbb{E}\|\sin\Theta(\widehat{U}, U)\| \lesssim \left(\frac{\sigma_C}{\lambda_r(X)} + \frac{\sigma_R\sigma_C}{\lambda_r^2(X)}\right) \wedge 1.$$

3.3. *SVD for heteroskedastic and incomplete data.* Missing data problems arise frequently in high-dimensional statistics. Let $X \in \mathbb{R}^{p_1 \times p_2}$ be a rank-r unknown matrix. Suppose only a small fraction of entries of X, denoted by $\Omega \subseteq [p_1] \times [p_2]$, are observable with random noise,

$$Y_{ij} = X_{ij} + Z_{ij}, \quad (i, j) \in \Omega.$$

Here, each entry Y_{ij} is observed or missing with probability θ or $1 - \theta$ for some $0 < \theta < 1$ and Z_{ij}'s are independent, zero-mean, and possibly heteroskedastic. Let $R \in \mathbb{R}^{p_1 \times p_2}$ be the indicator of observable entries:

$$R_{ij} = \begin{cases} 1, & (i, j) \in \Omega; \\ 0, & (i, j) \notin \Omega, \end{cases}$$

and R and Y are independent. Assume $X = U\Lambda V^\top$ is the singular value decomposition, where $U \in \mathbb{O}_{p_1,r}$ and $V \in \mathbb{O}_{p_2,r}$. Denote $\widetilde{Y}$ as the entry-wise product of Y and R, that is, $\widetilde{Y}_{ij} = Y_{ij}R_{ij}$, $\forall (i, j) \in [p_1] \times [p_2]$. We aim to estimate U based on $\{Y_{ij}, (i, j) \in \Omega\}$ or equivalently $\widetilde{Y}_{ij}$s. This problem is heteroskedastic since $\mathbb{E}\widetilde{Y}_{ij} = \theta X_{ij}$ and $\text{Var}(\widetilde{Y}_{ij})$ may vary for different (i, j) pairs. We can apply HeteroPCA to $\widetilde{Y}\widetilde{Y}^\top$ to estimate U. The following theoretical guarantee holds.

THEOREM 6. *Let* X *be a* p_1*-by-*p_2 *rank-*r *matrix, whose left singular subspace is denoted by* $U \in \mathbb{O}_{p_1,r}$. *Assume that* $\mathbb{E}Y = X$. *Suppose* Y *satisfies* $\max_{ij} \|Y_{ij}\|_{\psi_2} \le C$ *and all entries* Y_{ij} *are independent. Suppose* $0 < \theta \le 1 - c$ *for constant* $c > 0$. *There exists constant* $c_I > 0$ *such that if* $U \in \mathbb{O}_{p_1 \times r}$ *satisfies* $I(U)\|X\|/\lambda_r(X) \le c_I p_1/r$, *HeteroPCA applied to* $\widetilde{Y}\widetilde{Y}^\top$ *with* $T = \Omega(\log(\theta\lambda_r^2(X)/p_1) \vee 1)$ *outputs an estimator* $\widehat{U}$ *satisfying*

$$\|\sin\Theta(\widehat{U}, U)\| \lesssim \frac{\max\{\sqrt{p_2(\theta + \theta^3 p_1^2)\log(p_1)}, \theta p_1 \log^2(p_1)\}}{\theta^2\lambda_r^2(X)} \wedge 1 \tag{23}$$

with probability at least $1 - p_1^{-C}$.

REMARK 7 (Comparison with matrix completion). Our result is related to a substantial body of literature on low-rank matrix completion. For example, [11, 13, 46] analyzed the performance of nuclear norm minimization; [39] introduced the spectral regularization algorithm for incomplete matrix learning and developed the software package *SoftImpute*;[1] [29,

[1] https://cran.r-project.org/web/packages/softImpute/index.html

31–33] analyzed the alternating gradient descent and spectral algorithm for matrix completion with/without noise; [42] developed *OptShrink*, an algorithm for matrix estimation based on the optimal shrinkage of singular values and truncated SVD guided by random matrix theory; [47] studied the low-rank model for count data with missing values; also see [7] for a recent survey of matrix completion. Different from the literature on matrix completion, our goal here is to estimate the singular subspace $U \in \mathbb{O}_{p_1,r}$ rather than the whole matrix $X \in \mathbb{R}^{p_1 \times p_2}$. We apply HeteroPCA to impute the diagonal entries of $XX^\top$, not the missing entries in X itself as in most of the aforementioned matrix completion literature.

In addition, when the average amplitude of all entries in X is a constant (i.e., $\|X\|_F^2 \asymp p_1 p_2$) and X is well conditioned (i.e., $\lambda_1(X) \asymp \lambda_r(X)$), Theorem 6 implies that the HeteroPCA estimator is consistent as long as the expected sample size satisfies

$$\mathbb{E}|\Omega| \gg \max\bigl\{p_1^{1/3} p_2^{2/3} r^{2/3} \log^{1/3}(p_1),\, p_1 r^2 \log(p_1),\, p_1 r \log(p_1)\log(p_1 p_2)\bigr\}. \tag{24}$$

In the classic literature on matrix completion [32, 46], the sample size requirement is $|\Omega| \gtrsim (p_1 + p_2) r \cdot \mathrm{polylog}(p)$. When $p_1 \gtrsim p_2$, these sample size requirements nearly match and coincide with existing lower bounds in the literature [13], Theorem 1.7. When $p_1 \ll p_2$, (24) requires much fewer samples than what is needed for matrix completion; in other words, HeteroPCA can consistently estimate the p_1-by-r subspace U_1, even if most columns of X are completely missing and estimating the whole p_1-by-p_2 matrix accurately is impossible. To our best knowledge, we are among the first to show such a result.

REMARK 8 (Time complexity). If the target matrix X is p_1-by-p_2 and rank-r, the time complexity of HeteroPCA, regular SVD, diagonal-deletion SVD, OptShrink [42], and SoftImpute [39] are $O(|\Omega|^2/p_2 + Tp_1^2 r)$, $O(T(|\Omega| r + r^3))$, $O(|\Omega|^2/p_2 + Tp_1^2 r)$, $O(T(|\Omega| r + r^3))$, and $O(T(|\Omega| + p_1 p_2 r))$, respectively. Here, T denotes the number of iterations in each method.

REMARK 9. Recently, [53, 54] studied PCA with sparse data-dependent noise and incomplete data. They proved that if the signal-to-noise ratio is strong enough, the uncorrelated noise is small enough, and the proportion of missing values is small enough, one can estimate the subspace accurately. Under the model setting of Theorem 6 and some regularity conditions, [53], Corollary 3.7, can imply $\widetilde{U}_1 = \mathrm{SVD}_r(Y)$ satisfies

$$\bigl\|\sin\Theta(\widetilde{U}_1, U_1)\bigr\| \lesssim \left(\sqrt{\frac{sbr}{p_1 p_2}} + \frac{p_2}{\lambda_r^2(X)} + \sqrt{\frac{r^2 s \log p_1}{p_2}} + \sqrt{\frac{r \log p_1}{\lambda_r^2(X)}}\right) \wedge 1. \tag{25}$$

Here, s, b are the maximum number of missing values in each row and in each column, respectively. To ensure (25) gives an nontrivial upper bound, one must have $(s/p_1)(b/p_2) \lesssim 1/r$. In contrast, Theorem 6 implies that HeteroPCA can consistently recover U_1 even if $s \approx p_1$ and $b \approx p_2$, that is, only a smaller fracture of entries are observable, if the observable entries are uniform randomly selected from the target matrix.

REMARK 10. PCA for heteroskedastic and incomplete data is another closely related problem. Suppose one observes incomplete i.i.d. samples $Y_1, \ldots, Y_n \in \mathbb{R}^p$ from the generalized spiked covariance model (3) with missingness:

$$\forall 1 \le i \le p, 1 \le k \le n, \quad R_{ik} = \begin{cases} 1, & Y_{ik} \text{ is observable;} \\ 0, & Y_{ik} \text{ is missing,} \end{cases}$$

where $\{R_{ik}\}_{1 \le i \le p, 1 \le k \le n}$ are independent of $Y_1, \ldots, Y_n$. The goal is to estimate U. Many existing literature on PCA with incomplete data focused on regular SVD methods under

the homoskedastic noisy setting (see, e.g., [9, 36]), which are not directly suitable here. To estimate U using HeteroPCA, we can evaluate the generalized sample covariance matrix,

$$\widehat{\Sigma}^* = (\widehat{\sigma}^*_{ij})_{1\le i,j\le p}, \quad \text{with } \widehat{\sigma}^*_{ij} = \frac{\sum_{k=1}^n (Y_{ik} - \bar{Y}^*_i)(Y_{ik} - \bar{Y}^*_j) R_{ik} R_{jk}}{\sum_{k=1}^n R_{ik} R_{jk}} \quad \text{and}$$

$$\bar{Y}^*_i = \frac{\sum_{k=1}^n Y_{ik} R_{ik}}{\sum_{k=1}^n R_{ik}}.$$

Then U can be estimated by applying Algorithm 1 on $\widehat{\Sigma}^*$. A similar consistent upper bound result to Theorem 6 can be developed for this procedure.

In a more general scenario that noise ε_k has nondiagonal covariance or depends linearly on the signal X_k, the readers are referred to [53, 54] for a theory of the SVD estimator.

4. Numerical results. In this section, we investigate the numerical performance of the proposed procedure. All simulation results are based on 1000 repeated independent experiments. The average and the standard deviation of estimation errors are respectively indicated by markers and error bars in each plot.

4.1. *PCA under the generalized spiked covariance model.* We first consider PCA under the generalized spiked covariance model (3). Let $p = 30$, $n \in [60, 600]$, and $r \in \{3, 5\}$. We generate a p-by-r random matrix U_0 with i.i.d. standard Gaussian entries, $w_1, \dots, w_p \overset{\text{iid}}{\sim} \text{Unif}[0, 1]$, and $\sigma_1, \dots, \sigma_p \overset{\text{iid}}{\sim} \text{Unif}[0, 1]$. The purpose of generating uniform random vectors w, σ is to introduce heteroskedasticity into observations. Then, we let $U = \text{QR}(\text{diag}(w) \cdot U_0) \in \mathbb{O}_{p,r}$ and $\Sigma_0 = U\,\text{diag}(1, \dots, r)U^\top \in \mathbb{R}^{p\times p}$. We aim to recover U based on i.i.d. observations $\{Y_k = X_k + \varepsilon_k\}_{k=1}^n$, where $X_1, \dots, X_n \overset{\text{iid}}{\sim} N(0, \Sigma_0)$, $\varepsilon_1, \dots, \varepsilon_n \overset{\text{iid}}{\sim} N(0, \text{diag}(\sigma_1^2, \dots, \sigma_n^2))$. We implement the proposed HeteroPCA, diagonal-deletion, and regular SVD approaches and plot the average estimation errors and standard deviation in $\sin\Theta$ distance. We also implement the classic factor analysis method [34, 49], `factanal` function in R `stats` package, and the Bayesian factor analysis method, `MCMCfactanal` function from R `MCMCpack` package [38]. The simulation results are summarized in Figure 1. It can be seen that the proposed HeteroPCA estimator significantly outperforms other methods; the regular SVD yields larger estimation error; and the diagonal-deletion estimator performs unstably across different settings. This matches the theoretical findings in Section 2.

Next, we study how the degree of heteroskedasticity affects the performance. Let

$$v_1, \dots, v_p \overset{\text{iid}}{\sim} \text{Unif}[0, 1], \qquad \sigma_k^2 = \frac{0.1 \cdot p \cdot v_k^\alpha}{\sum_{i=1}^p v_i^\alpha}, \quad k = 1, \dots, p.$$

In such case, $\sigma^2_{\text{sum}} = \sigma_1^2 + \cdots + \sigma_p^2$ always equals $0.1p$ and α characterizes the degree of heteroskedasticity: the larger α results in a more imbalanced distribution of $(\sigma_1, \dots, \sigma_p)$; if $\alpha = 0$, $\sigma_1 = \cdots = \sigma_p$ and the setting becomes homoskedastic. Now, we generate U, Σ_0 and $\{Y_k, X_k, \varepsilon_k\}_{k=1}^n$ in the same way as the previous setting. We only compare HeteroPCA with regular SVD and diagonal-deletion estimator since it takes a too long time to run factor analysis methods in this setting. The average estimation errors for U are plotted in Figure 2. The results again suggest that the performance of diagonal-deletion estimator is unstable across different settings. When $\alpha = 0$, that is, the noise is homoskedastic, the performance of HeteroPCA and regular SVD are comparable; but as α increases, the estimation error of HeteroPCA grows significantly slower than that of the regular SVD, which is consistent with the theoretical results in Theorem 1.

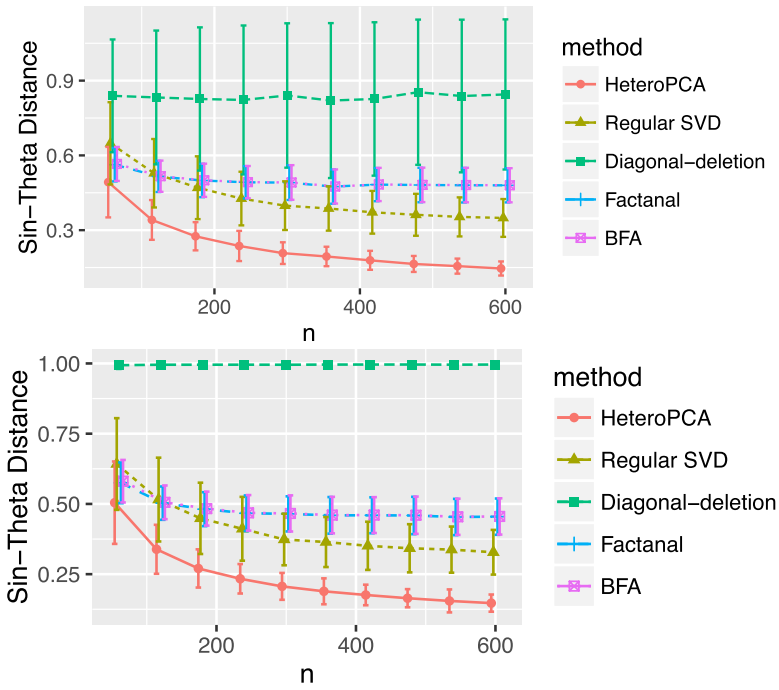
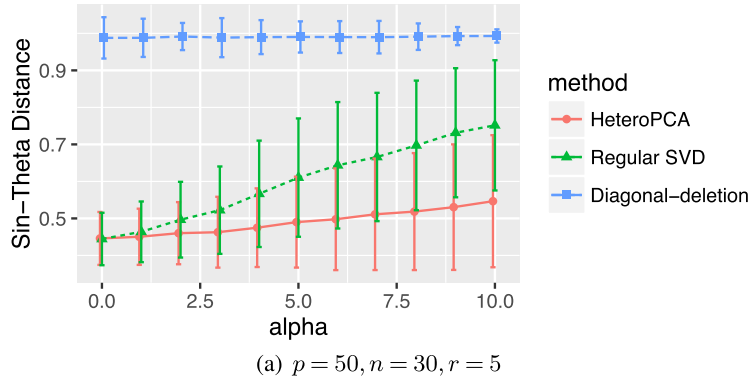
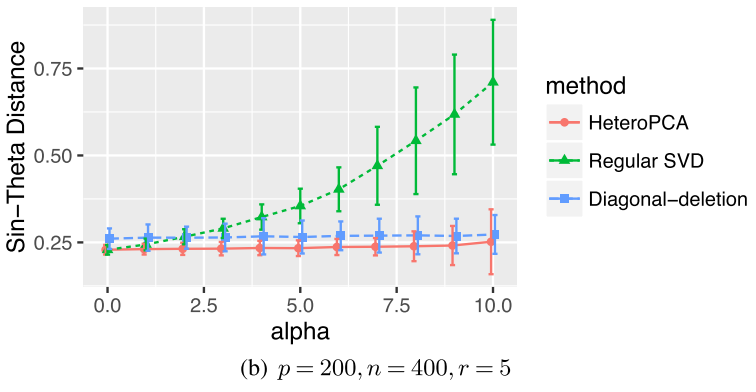


FIG. 1. *Average* $\sin\Theta$ *loss versus sample size* n *under the generalized spiked covariance model* (*Section* 4.1). *Upper panel*: $r = 3$; *lower panel*: $r = 5$.



(a) $p = 50, n = 30, r = 5$



(b) $p = 200, n = 400, r = 5$

FIG. 2. *Average* $\sin\Theta$ *loss versus heteroskedastic level* α *under the generalized spiked covariance model* (*Section* 4.1).

4.2. *SVD under heteroskedastic noise.* Next, we consider the problem of SVD under heteroskedastic noise discussed in Section 3.1. Let $U_0 \in \mathbb{R}^{p_1 \times r}$ and $V_0 \in \mathbb{R}^{p_2 \times r}$ be i.i.d. Gaussian ensembles for $(p_1, p_2) = (50, 200), (200, 1000)$ and $r = 3$. To introduce heteroskedasticity, we also randomly draw $w, v_1 \in \mathbb{R}^{p_1}$, and $v_2 \in \mathbb{R}^{p_2}$ with i.i.d. Unif[0, 1] entries. Then we evaluate $U = \mathrm{QR}(U_0 \cdot \mathrm{diag}(w)^4)$, $V = \mathrm{QR}(V_0)$, and construct the signal matrix $X = (p_1 p_2)^{1/4} \cdot U \,\mathrm{diag}(1, \ldots, r) V^\top$. The noise matrix is drawn as $E_{ij} \overset{\text{ind}}{\sim} N(0, \sigma_0^2 \cdot \sigma_{ij}^2)$, where $\sigma_{ij} = (v_1)_i^4 \cdot (v_2)_j^4$, σ_0 varies from 0 to 2, $1 \le i \le p_1$, and $1 \le j \le p_2$. Based on the p_1-by-p_2 observation $Y = X + E$, we implement HeteroPCA with input of $YY^\top$, regular-SVD, diagonal-deletion, and *OptShrink*[2] ([42], an algorithm for matrix estimation based on the optimal shrinkage of singular values and truncated SVD guided by random matrix theory) to evaluate $\widehat{U}$, $\widehat{V}$. For each of the estimators $\widehat{U}$ and $\widehat{V}$, we also estimate X by $\widehat{X} = \widehat{U}\widehat{U}^\top Y \widehat{V}\widehat{V}^\top$. The average $\sin\Theta$ norm errors of $\widehat{U}$, $\widehat{V}$ and the average Frobenius norm error of $\widehat{X}$ are presented in Figure 3. We can see the proposed HeteroPCA outperforms other methods in all estimations for U, V, and X, and the advantage of HeteroPCA is more significant when the noise level increases.

4.3. *Poisson PCA.* We generate $U_0 \in \mathbb{R}^{p_1 \times r}$ and $V_0 \in \mathbb{R}^{p_2 \times r}$ with i.i.d. standard normal entries for $(p_1, p_2, r) = (50, 500, 3)$ or $(200, 1000, 3)$. Similarly to previous settings, we introduce heteroskedasticity by generating a vector $w \in \mathbb{R}^{p_1}$ with i.i.d. Unif[0, 1] entries. Let $U = |U_0 \cdot \mathrm{diag}(w)^4| \in \mathbb{R}^{p_1 \times r}$, $V = |V_0| \in \mathbb{R}^{p_2 \times r}$, $X = \lambda U \,\mathrm{diag}(1, \ldots, r) V^\top \in \mathbb{R}^{p_1 \times p_2}$, and $Y_{ij} \sim \mathrm{Poisson}(X_{ij})$ independently. Here, $\lambda > 0$ measures the signal strength. The performance of HeteroPCA, regular SVD, diagonal-deletion, and OptShrink on estimation of left singular subspaces are provided in Figure 4. These plots again illustrate the merit of the proposed HeteroPCA method.

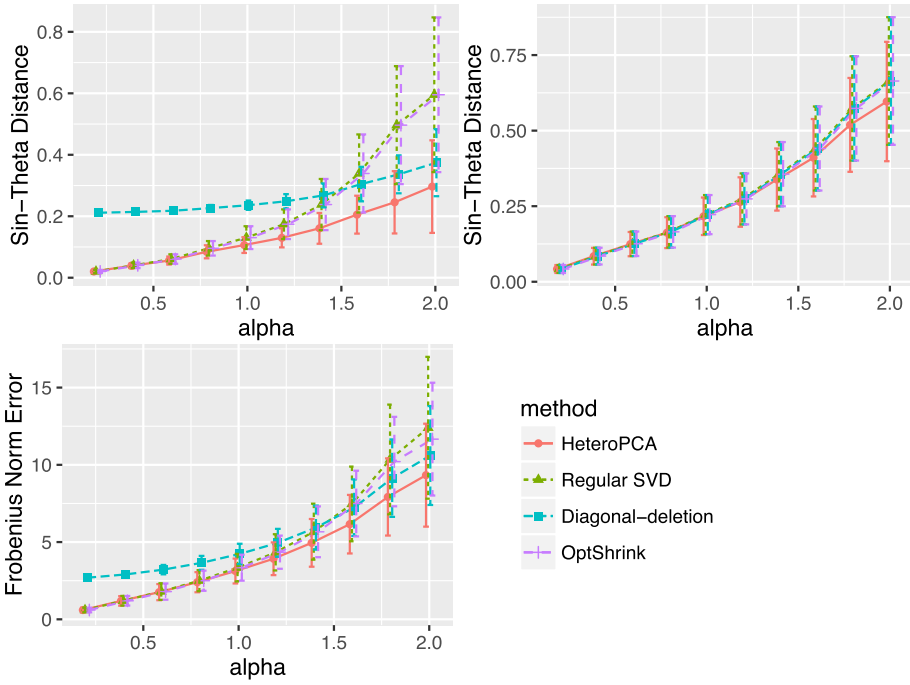
4.4. *SVD based on heteroskedastic and incomplete data.* Finally, in the following experiment, we study SVD based on heteroskedastic and incomplete data in the setting of Section 3.3. Generate $Y, X, Z \in \mathbb{R}^{p_1 \times p_2}$ in the same way as the previous heteroskedastic SVD setting with $p_1 = 50, 100$, $r = 3, 5$, $\sigma_0 = 0.2$, and p_2 ranging from 800 to 3200. Each entry of Y is observed independently with probability $\theta = 0.1$. We aim to estimate U based on $\{Y_{ij} : (i, j) \in \Omega\}$. In addition to HeteroPCA, regular SVD, diagonal-deletion SVD, and OptShrink, we also apply the nuclear norm minimization via *Soft-Impute* package ([39], also see Remark 10)

$$\widehat{X}_* = \underset{\widehat{X} \in \mathbb{R}^{p_1 \times p_2}}{\arg\min} \sum_{(i,j) \in \Omega} (\widetilde{Y}_{ij} - \widehat{X}_{ij})^2 + \nu \|\widehat{X}\|_*, \qquad \widehat{U} = \mathrm{SVD}_r(\widehat{X}).$$

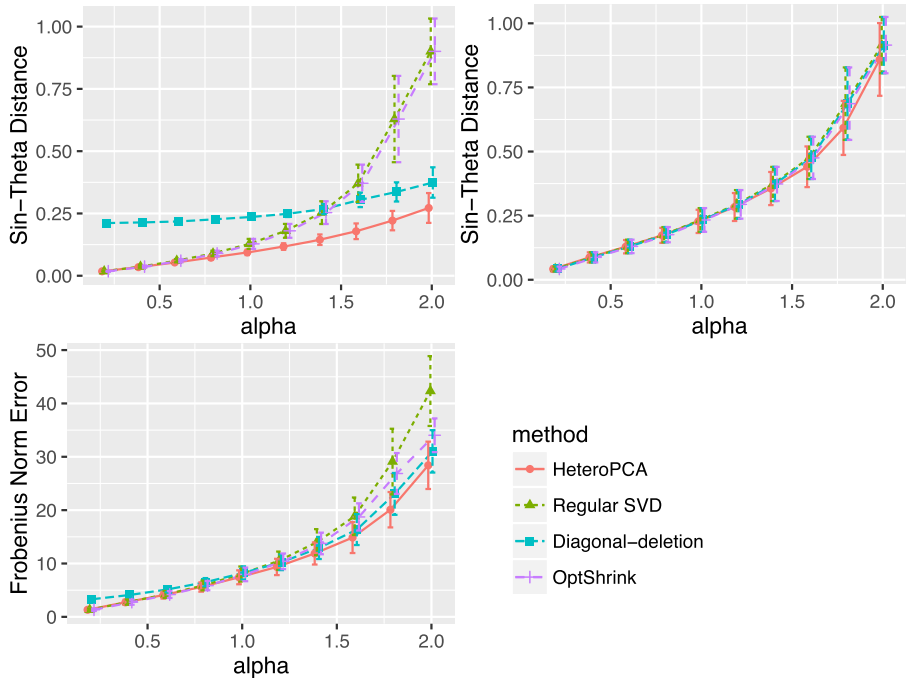
To avoid the cumbersome issue of parameter ν selection, we evaluate the above nuclear norm minimization estimator for a grid of values of ν, then record the outcome with the minimum $\sin\Theta$ distance error $\|\sin\Theta(\widehat{U}, U)\|$. From the results plotted in Figure 5, we can see that HeteroPCA significantly outperforms all other methods when $p_1 \ll p_2$, which matches the discussion in Remark 7.

5. Discussion. We consider PCA in the presence of heteroskedastic noise in this paper. To alleviate the significant bias incurred on diagonal entries of the Gram matrix due to heteroskedastic noise, we introduced a new procedure named HeteroPCA that adaptively imputes diagonal entries to remove the bias. The proposed procedure achieves optimal rate of convergence in a range of settings. In addition, we discuss the applications of the proposed

[2]Software package available at https://web.eecs.umich.edu/~rajnrao/optshrink/

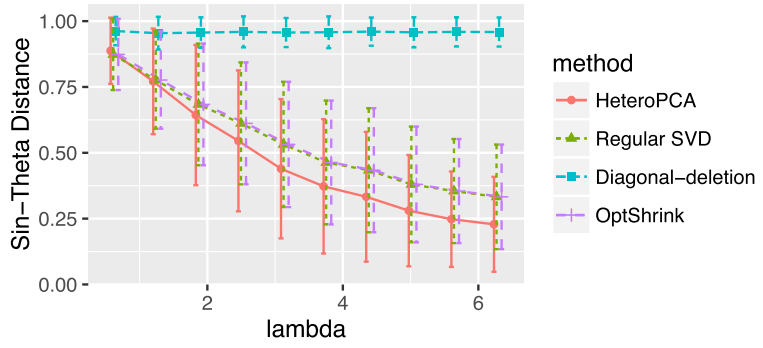


(a) $p_1 = 50, p_2 = 200$

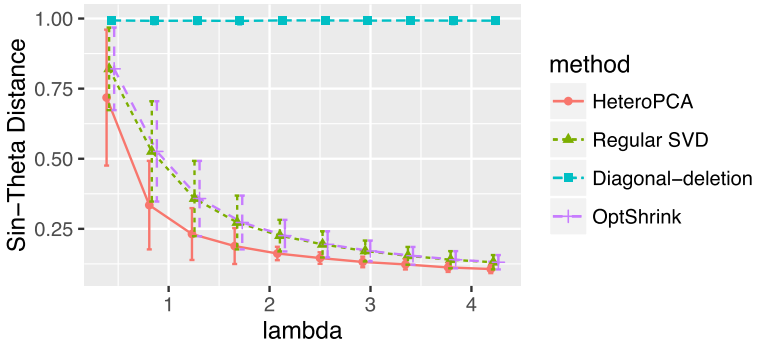


(b) $p_1 = 200, p_2 = 1000$

FIG. 3. *Estimation errors of* $\widehat{U}$ *(top left),* $\widehat{V}$ *(top right), and* $\widehat{X}$ *(bottom left) in SVD under heteroskedastic noise (Section 4.2).*

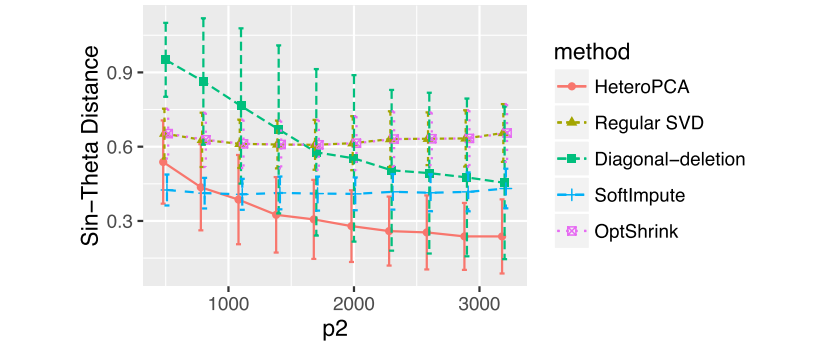


(a) $p_1 = 50, p_2 = 500$



(b) $p_1 = 200, p_2 = 1000$

FIG. 4. *Estimation errors for a ranging value of signal strength* λ *under the Poisson PCA model* (*Section* 4.3).



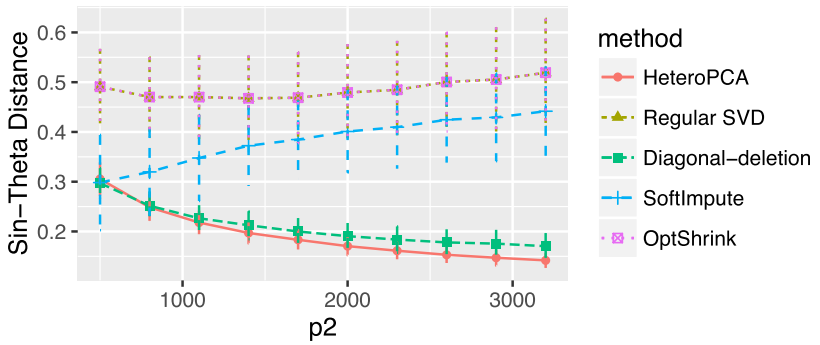


FIG. 5. *Average* $\sin\Theta$ *distance error for SVD based on heteroskedastic and incomplete data* (*Section* 4.4). *Here,* $p_1 = 50$, $r = 5$, $\theta = 0.2$ (*Upper Panel*) *and* $p_1 = 100$, $r = 3$, $\theta = 0.2$ (*Lower Panel*); p_2 *varies from* 800 *to* 3200.

algorithm to heteroskedastic low-rank matrix denoising, Poisson PCA, and SVD based on heteroskedastic and incomplete data.

The proposed HeteroPCA procedure can also be applied to many other problems where the noise is heteroskedastic. First, *exponential family PCA* is a commonly used technique for dimension reduction on nonreal-valued datasets [16, 41]. As discussed in the Introduction, the exponential family distributions, for example, exponential, binomial, and negative binomial, may be highly heteroskedastic. As in the case of Poisson PCA considered in Section 3.2, the proposed HeteroPCA algorithm can be applied to exponential family PCA.

In addition, community detection in social networks has attracted significant attention in the recent literature [22]. Although most of existing results focused on unipartite graphs, bipartite graphs, that is, all edges are between two groups of nodes, often appear in practice [21, 40, 62]. The proposed HeteroPCA can also be applied to *community detection for bipartite stochastic block model*. Similarly to the analysis for heteroskedastic low-rank matrix denoising in Section 3.1, HeteroPCA can be shown to have advantages over other baseline methods.

The proposed framework is also applicable to solve the *heteroskedastic tensor SVD* problem, which aims to recover the low-rank structure from the tensorial observation corrupted by heteroskedastic noise. Suppose one observes $\mathbf{Y} = \mathbf{X} + \mathbf{Z} \in \mathbb{R}^{p_1 \times p_2 \times p_3}$, where $\mathbf{X}$ is a Tucker low-rank signal tensor and $\mathbf{Z}$ is the noise tensor with independent and zero-mean entries. If $\mathbf{Z}$ is homoskedastic, the higher-order orthogonal iteration (HOOI) [18] was shown to achieve the optimal performance for recovering $\mathbf{X}$ [61]. If $\mathbf{Z}$ is heteroskedastic, we can apply HeteroPCA instead of the regular SVD to obtain a better initialization for HOOI. Similarly to the argument in this article, we are able to show that this modified HOOI yields more stable and accurate estimates than the regular HOOI.

Canonical correlation analysis (CCA) is one of the most important tools in multivariate analysis for exploring the relationship between two sets of vector samples [28]. In the standard procedure of CCA, the core step is a regular SVD on the adjusted cross-covariance matrix between samples. When the observations contain heteroskedastic noise, one can replace the regular SVD procedure by HeteroPCA to achieve better performance.

6. Proofs. In this section, we prove the main results, namely, Theorems 1 and 3. For reasons of space, the other proofs are given in the Supplementary Material [60].

6.1. *Proofs for heteroskedastic PCA.* PROOF OF THEOREM 1. First, we introduce

$$E = [\varepsilon_1, \ldots, \varepsilon_n] \in \mathbb{R}^{p \times n}, \qquad \gamma_k = \Lambda^{1/2} U^\top (X_k - \mu) \in \mathbb{R}^r, \qquad \Gamma = [\gamma_1, \ldots, \gamma_n] \in \mathbb{R}^{r \times n}.$$

Then the observations can be written as

$$Y_k = X_k + \varepsilon_k = \mu + U\Lambda^{1/2}\gamma_k + \varepsilon_k, \quad \text{or} \quad Y = \mu 1_n^\top + U\Lambda^{1/2}\Gamma + E,$$

where $\mu \in \mathbb{R}^p$ is a fixed vector, $\mathbb{E}\gamma_k = 0$, $\mathrm{Cov}(\gamma_k) = I$, E has independent entries, and Γ has independent columns. We also denote $\bar{X} \in \mathbb{R}^p$, $\bar{E} \in \mathbb{R}^p$, $\bar{\Gamma} \in \mathbb{R}^r$ as the averages of all columns of X, E, and Γ, respectively. Since $\widehat{\Sigma}$ is invariant after any translation on Y, we can assume $\mu = 0$ without loss of generality. The rest of the proof is divided into three steps for the sake of presentation.

Step 1. We define $\widehat{\Sigma}_X = (XX^\top - n\bar{X}\bar{X}^\top)/(n-1)$ as the signal sample covariance. The aim of this step is to develop a concentration inequality for $\widehat{\Sigma} - \Sigma_X$. To this end, we consider

the following decomposition of $n(\widehat{\Sigma} - \widehat{\Sigma}_X)$:

$$
\begin{aligned}
&(n-1)(\widehat{\Sigma} - \widehat{\Sigma}_X) \\
&\quad = (n-1)\widehat{\Sigma} - \bigl(XX^\top - n\bar{X}\bar{X}^\top\bigr) \\
&\quad = YY^\top - n\bar{Y}\bar{Y}^\top - \bigl(XX^\top - n\bar{X}\bar{X}\bigr) \\
&\quad = (X+E)(X+E)^\top - \bigl(XX^\top - n\bar{X}\bar{X}^\top\bigr) - n\bigl(\bar{X}\bar{X}^\top + \bar{X}\bar{E}^\top + \bar{E}\bar{X}^\top + \bar{E}\bar{E}^\top\bigr) \\
&\quad = XE^\top + EX^\top + EE^\top - n\bigl(\bar{X}\bar{E}^\top + \bar{E}\bar{X}^\top + \bar{E}\bar{E}^\top\bigr).
\end{aligned} \tag{26}
$$

We analyze each term of (26) separately as follows. Since E has independent entries and $\operatorname{Var}(E_{ij}) = \sigma_i^2$, the rowwise structured heteroskedastic concentration inequality [8], Theorem 6, implies

$$
\mathbb{E}\bigl\|EE^\top - \mathbb{E}EE^\top\bigr\| \lesssim \sqrt{n}\sigma_{\mathrm{sum}}\sigma_{\max} + \sigma_{\mathrm{sum}}^2. \tag{27}
$$

Since X is deterministic, E is random, and $\mathbb{E}E = 0$, we have $\mathbb{E}EX^\top = 0$. By Lemma 1 in the Supplementary Material,

$$
\begin{aligned}
\mathbb{E}_E\bigl(\bigl\|EX^\top - \mathbb{E}EX^\top\bigr\| \mid X\bigr) \quad &= \quad \mathbb{E}_E\bigl(\bigl\|EX^\top\bigr\| \mid X\bigr) \\
&\lesssim \quad \|X\|\bigl(\sigma_C + r^{1/4}\sigma_{\max}\sigma_C + \sqrt{r}\sigma_{\max}\bigr) \\
&\overset{\text{Cauchy–Schwarz}}{\lesssim} \quad \|X\|(\sigma_{\mathrm{sum}} + \sqrt{r}\sigma_{\max}).
\end{aligned} \tag{28}
$$

Since $\mathbb{E}\|\bar{E}\|_2^2 = \sum_{i=1}^p \mathbb{E}(\bar{E})_i^2 = \sum_{i=1}^p \sigma_i^2/n = \sigma_{\mathrm{sum}}^2/n$, we have

$$
\begin{aligned}
&\mathbb{E}_E\bigl(n\bigl\|\bar{X}\bar{E}^\top + \bar{E}\bar{X}^\top + \bar{E}\bar{E}^\top\bigr\|\bigr) \\
&\quad \le \mathbb{E}_E n\bigl\|\bar{X}\bar{E}^\top\bigr\| + \mathbb{E}_E n\bigl\|\bar{E}\bar{X}^\top\bigr\| + \mathbb{E}_E n\bigl\|\bar{E}\bar{E}^\top\bigr\| \\
&\quad \le \mathbb{E}_E 2n\|\bar{X}\|_2\|\bar{E}\|_2 + \mathbb{E}n\|\bar{E}\|_2^2 \\
&\quad \le 2n\|\bar{X}\|_2 \cdot \bigl(\mathbb{E}\|\bar{E}\|_2^2\bigr)^{1/2} + \mathbb{E}n\|\bar{E}\|_2^2 \le 2n^{1/2}\sigma_{\mathrm{sum}}\|\bar{X}\|_2 + \sigma_{\mathrm{sum}}^2.
\end{aligned} \tag{29}
$$

Combining (27), (28), and (29), we have

$$
\begin{aligned}
&\mathbb{E}_E\bigl\|(n-1)(\widehat{\Sigma} - \widehat{\Sigma}_X) - \mathbb{E}EE^\top\bigr\| \\
&\quad \lesssim \sqrt{n}\sigma_{\mathrm{sum}}\sigma_{\max} + \sigma_{\mathrm{sum}}^2 + \|X\|(\sigma_{\mathrm{sum}} + \sqrt{r}\sigma_{\max}) + n^{1/2}\|\bar{X}\|_2\sigma_{\mathrm{sum}}.
\end{aligned}
$$

Noting that $\mathbb{E}EE^\top = n\operatorname{diag}(\sigma_1^2, \ldots, \sigma_p^2)$ is diagonal and $\Delta(\cdot)$ is the operator that sets all diagonal entries to zero, we further have

$$
\begin{aligned}
&\mathbb{E}_E\bigl\|\Delta\bigl((n-1)(\widehat{\Sigma} - \widehat{\Sigma}_X)\bigr)\bigr\| \\
&\quad = \mathbb{E}_E\bigl\|\Delta\bigl((n-1)(\widehat{\Sigma} - \widehat{\Sigma}_X) - \mathbb{E}EE^\top\bigr)\bigr\| \\
&\quad \overset{\text{Lemma 3}}{\le} 2\mathbb{E}_E\bigl\|(n-1)(\widehat{\Sigma} - \widehat{\Sigma}_X) - \mathbb{E}EE^\top\bigr\| \\
&\quad \lesssim \sqrt{n}\sigma_{\mathrm{sum}}\sigma_{\max} + \sigma_{\mathrm{sum}}^2 + \|X\|(\sigma_{\mathrm{sum}} + \sqrt{r}\sigma_{\max}) + n^{1/2}\|\bar{X}\|_2\sigma_{\mathrm{sum}}.
\end{aligned}
$$

Step 2. Next, we study the expectation of the target function with respect to X. We specifically need to study $\lambda_r(n\widehat{\Sigma}_X)$, $\|X\|$, and $\|\bar{X}\|_2$. Since $\Gamma \in \mathbb{R}^{r\times n}$ has independent columns and each column is isotropic sub-Gaussian distributed, based on the random matrix theory ([55], Corollary 5.35),

$$
\mathbb{P}\bigl(\sqrt{n} + C\sqrt{r} + t \ge \|\Gamma\| \ge \lambda_r(\Gamma) \ge \sqrt{n} - C\sqrt{r} - t\bigr) \le \exp\bigl(-Ct^2/2\bigr).
$$

In addition, $\sqrt{n}\bar{\Gamma} \in \mathbb{R}^r$ is a sub-Gaussian vector with the identity covariance matrix. By the Bernstein-type concentration inequality ([55], Proposition 5.16),

$$\mathbb{P}\big(\|\sqrt{n}\bar{\Gamma}\|_2^2 \geq r + C\sqrt{rx} + Cx\big) \leq C\exp(-cx).$$

If $n \geq Cr$ for some large constant $C > 0$, by setting $t = c\sqrt{n}$ and $x = cn$ in the previous two inequalities, we have

$$2\sqrt{n} \geq \|\Gamma\| \geq \lambda_r(\Gamma) \geq \sqrt{n}/2, \quad \text{and} \quad \|\sqrt{n}\bar{\Gamma}\|_2 \leq \sqrt{n}/3 \tag{30}$$

with probability at least $1 - C\exp(-cn)$. When (30) holds,

$$\begin{aligned}
\lambda_r(n\widehat{\Sigma}_X) &= \lambda_r\big(n\big(XX^\top - n\bar{X}\bar{X}^\top\big)\big) = \lambda_r\big(nU\Lambda^{1/2}\big(\Gamma\Gamma^\top - n\bar{\Gamma}\bar{\Gamma}^\top\big)\Lambda^{1/2}U^\top\big) \\
&\geq \lambda_r(\Lambda)\cdot\lambda_r\big(\Gamma\Gamma^\top - n\bar{\Gamma}\bar{\Gamma}^\top\big) \geq \lambda_r(\Lambda)\big(\lambda_r^2(\Gamma) - \|\sqrt{n}\bar{\Gamma}\|_2^2\big) \\
&\overset{(29)}{\geq} \lambda_r(\Lambda)(n/4 - n/9) \gtrsim n\lambda_r(\Lambda);
\end{aligned} \tag{31}$$

$$\|X\| \leq \big\|U\Lambda^{1/2}\Gamma\big\| \leq \big\|\Lambda^{1/2}\big\|\cdot\|\Gamma\| \overset{(30)}{\leq} 2\sqrt{n}\big\|\Lambda^{1/2}\big\| \lesssim \sqrt{n}\lambda_r^{1/2}(\Lambda),$$

where the last inequality is due to the assumption that $\|\Lambda\|/\lambda_r(\Lambda) \leq C$ for some constant C.

$$\|\bar{X}\|_2 = \big\|U\Lambda^{1/2}\bar{\Gamma}\big\|_2 \leq \big\|\Lambda^{1/2}\big\|\cdot\|\bar{\Gamma}\|_2 \lesssim \lambda_r^{1/2}(\Lambda).$$

Combining the previous three inequalities, we know if (30) holds,

$$\begin{aligned}
&\frac{\mathbb{E}_E\|\Delta((n-1)(\widehat{\Sigma} - \widehat{\Sigma}_X))\|}{\lambda_r((n-1)\widehat{\Sigma}_X)} \wedge 1 \\
&\lesssim \frac{\sqrt{n}\sigma_{\mathrm{sum}}\sigma_{\max} + \sigma_{\mathrm{sum}}^2 + (n\lambda_r(\Lambda))^{1/2}(\sigma_{\mathrm{sum}} + \sqrt{r}\sigma_{\max}) + (n\lambda_r(\Lambda))^{1/2}(\Lambda)\sigma_{\mathrm{sum}}}{n\lambda_r(\Lambda)} \\
&\quad \wedge 1 \\
&\lesssim \left(\frac{\sigma_{\mathrm{sum}} + \sqrt{r}\sigma_{\max}}{(n\lambda_r(\Lambda))^{1/2}} + \frac{\sqrt{n}\sigma_{\mathrm{sum}}\sigma_{\max} + \sigma_{\mathrm{sum}}^2}{n\lambda_r(\Lambda)}\right) \wedge 1 \\
&\lesssim \left(\frac{\sigma_{\mathrm{sum}} + \sqrt{r}\sigma_{\max}}{(n\lambda_r(\Lambda))^{1/2}} + \frac{\sigma_{\mathrm{sum}}\sigma_{\max}}{n^{1/2}\lambda_r(\Lambda)}\right) \wedge 1.
\end{aligned} \tag{32}$$

Here, the last "$\lesssim$" is due to $\sigma_{\mathrm{sum}}^2/(n\lambda_r(\Lambda)) \wedge 1 \leq \sigma_{\mathrm{sum}}/(n\lambda_r(\Lambda))^{1/2} \wedge 1$.

Step 3. Finally, since $\operatorname{rank}(\widehat{\Sigma}_X) \leq r$, the eigenvectors of $\widehat{\Sigma}_X$ are U, and U satisfies the incoherence condition: $I(U) \leq c_I p/r$, the robust $\sin\Theta$ theorem (Theorem 3) for $T = \Omega\big(\log(\frac{n\lambda_r(\Lambda)}{\sigma_{\mathrm{sum}}^2}) \vee 1\big)$ yields

$$\begin{aligned}
&\mathbb{E}_E\big\|\sin\Theta(\widehat{U}, U)\big\| \\
&\lesssim \mathbb{E}_E\left(\frac{\|\Delta((n-1)(\widehat{\Sigma} - \widehat{\Sigma}_X))\|}{\lambda_r((n-1)\widehat{\Sigma}_X)} + 2^{-T}\right) \wedge 1 \\
&\lesssim \left(\frac{\sqrt{n}\sigma_{\mathrm{sum}}\sigma_{\max} + \sigma_{\mathrm{sum}}^2 + \|X\|(\sigma_{\mathrm{sum}} + \sqrt{r}\sigma_{\max}) + n^{1/2}\|\bar{X}\|_2\sigma_{\mathrm{sum}}}{\lambda_r((n-1)\widehat{\Sigma}_X)}\right. \\
&\quad \left. + \frac{\sigma_{\mathrm{sum}}^2}{n\lambda_r(\Lambda)}\right) \wedge 1,
\end{aligned} \tag{33}$$

$$\begin{aligned}
&\mathbb{E}\big\|\sin\Theta(\widehat{U}, U)\big\| \\
&= \mathbb{E}\big\|\sin\Theta(\widehat{U}, U)\big\|1_{\{(30)\text{ holds}\}} + \mathbb{E}\big\|\sin\Theta(\widehat{U}, U)\big\|1_{\{(30)\text{ does not hold}\}}
\end{aligned}$$

$$\overset{(30)}{\lesssim} \left(\frac{\sigma_{\text{sum}} + \sqrt{r}\sigma_{\max}}{(n\lambda_r(\Lambda))^{1/2}} + \frac{\sigma_{\text{sum}}\sigma_{\max}}{n^{1/2}\lambda_r(\Lambda)}\right) \wedge 1 + \mathbb{P}((30) \text{ does not hold})$$

$$\lesssim \left(\frac{\sigma_{\text{sum}} + \sqrt{r}\sigma_{\max}}{(n\lambda_r(\Lambda))^{1/2}} + \frac{\sigma_{\text{sum}}\sigma_{\max}}{n^{1/2}\lambda_r(\Lambda)}\right) \wedge 1 + C\exp(-cn)$$

$$\lesssim \left(\frac{\sigma_{\text{sum}} + \sqrt{r}\sigma_{\max}}{(n\lambda_r(\Lambda))^{1/2}} + \frac{\sigma_{\text{sum}}\sigma_{\max}}{n^{1/2}\lambda_r(\Lambda)}\right) \wedge 1.$$

The last inequality is due to the assumption that $\lambda_r(\Lambda) \geq c\exp(-cn)$. Therefore, we have finished the proof of this theorem. □

6.2. *Proof of Theorem 3.* In this subsection, we prove a more general version of Theorem 3, where the corrupted entries lie in a known set $\mathcal{G} \subset [p] \times [p]$ which need not be the diagonal. Recall the model (14), where we observe a symmetric $p \times p$ matrix $N = M + Z$, where M is a rank-r matrix of interest and Z is the perturbation. Our goal is to estimate $U \in \mathbb{O}_{p,r}$, consisting of the eigenvectors of M. Extending the ideas of Algorithm 1 for HeteroPCA, Algorithm 2 provides a robust estimate of U which iteratively impute the values in the corrupted entries in $\mathcal{G}$. In the special case where $\mathcal{G}$ is the diagonal, that is, $\mathcal{G} = \{(i,i) : 1 \leq i \leq p\}$, Algorithm 2 reduces to Algorithm 1.

Next, we give a performance guarantee for Algorithm 2. For any $H \in \mathbb{R}^{p\times p}$, let $G(H)$ be the matrix H with all entries but those in $\mathcal{G}$ set to zero and $\Gamma(H) = H - G(H)$. Define

$$\eta = \max_{H\in\mathbb{R}^{m\times m}, \text{rank}(H)\leq 2r} \|G(H)\|/\|H\|, \tag{34}$$

which essentially measures the maximum perturbations due to the entries in $\mathcal{G}$ on the singular subspace. We also assume that the set of corrupted entries $\mathcal{G}$ is b-sparse in the sense that

$$\max_i |\{j : (i,j) \in \mathcal{G}\}| \vee \max_j |\{i : (i,j) \in \mathcal{G}\}| \leq b,$$

that is, the number of corrupted entries in each row and each column is at most b. To overcome the "spiky" issue discussed in Remark 1, we again assume the incoherence condition (35). We have the following theoretical results for Algorithm 2.

THEOREM 7 (General robust $\sin\Theta$ theorem). *Assume $\mathcal{G} \in [p] \times [p]$ is b-sparse. Suppose one observes the symmetric matrix $N = M + Z$, where $\text{rank}(M) = r$, Z is any symmetric perturbation, and the eigenvectors of M are $U \in \mathbb{O}_{p,r}$. Let $\widehat{U}^{(t)}$ be the intermediate matrix in Algorithm 1 with t iterations. There exists a constant $c > 0$ such that if the incoherence condition*

$$\frac{I(U)\|M\|}{\lambda_r(M)} \leq \frac{cp}{\eta br(b \wedge r)} \tag{35}$$

Algorithm 2 Generalized HeteroPCA

1: Input: matrix $\widehat{\Sigma}$, rank r, number of iterations T, corruption subset $\mathcal{G} \subseteq [p] \times [p]$.
2: Set $N^{(0)} = \Gamma(N)$.
3: **for** $t = 1, \ldots, T$ **do**
4: Calculate SVD: $N^{(t)} = \sum_i \lambda_i^{(t)} u_i^{(t)} (v_i^{(t)})^\top$, where $\lambda_1^{(t)} \geq \lambda_2^{(t)} \cdots \geq 0$.
5: Let $\widetilde{N}^{(t)} = \sum_{i=1}^r \lambda_i^{(t)} u_i^{(t)} (v_i^{(t)})^\top$.
6: Update corrupted entries: $N^{(t+1)} = G(\widetilde{N}^{(t)}) + \Gamma(\widehat{\Sigma})$.
7: **end for**
8: Output: $\widehat{U} = U^{(T)} = [u_1^{(T)} \cdots u_r^{(T)}]$.

is satisfied and $\eta\|\Gamma(Z)\| \le c\lambda_r(M)$, *then*

$$\|\sin\Theta(\widehat{U}^{(t)}, U)\| \le 4\|\Gamma(Z)\|/\lambda_r(M) + 2^{-(t+3)}/\eta. \tag{36}$$

Here, η *is defined in* (34). *In particular, if* $T = \Omega(\log \frac{\lambda_r(M)}{\eta\|\Gamma(Z)\|} \vee 1)$, *the final outcome* $\widehat{U}$ *of Algorithm* 1 *with corrupted index set* $\mathcal{G}$ *satisfies*

$$\|\sin\Theta(\widehat{U}, U)\| \lesssim \frac{\|\Gamma(Z)\|}{\lambda_r(M)} \wedge 1. \tag{37}$$

REMARK 11. Though calculating the exact value of η can be difficult in general, Lemma 3 in the Supplementary Material shows $\eta \le \sqrt{b \wedge (2r)}$ for all b-sparse $\mathcal{G}$.

PROOFS OF THEOREM 7. To characterize how the proposed procedure refines the estimation by initialization and iterations, we define $T_0 = \|\Gamma(N - M)\| = \|\Gamma(Z)\|$ and $K_t = \|N^{(t)} - M\|$ for $t = 0, 1, \ldots$. Since $H = \Gamma(H) + G(H)$, we have $\|H\| \le \|G(H)\| + \|\Gamma(H)\|$ for any matrix $H \in \mathbb{R}^{p\times p}$.

Step 1. We first analyze the initial error $K_0 = \|N^{(0)} - M\|$. By definition, $N^{(0)} = \Gamma(N)$. To better align $\Gamma(N)$ with M, we decompose $M = \Gamma(M) + G(M)$. Since the singular subspace of M aligns with U, we have $M = P_U M P_U$. Thus,

$$\begin{aligned} K_0 &= \|N^{(0)} - M\| = \|\Gamma(N - M) - G(M)\| \\ &\le \|\Gamma(N - M)\| + \|G(M)\| = \|\Gamma(Z)\| + \|G(P_U M P_U)\| \\ &\overset{(a)}{\le} \|\Gamma(Z)\| + \frac{I(U)rb}{p}\|M\| = T_0 + \frac{I(U)rb}{p}\|M\|. \end{aligned}$$

Here, (a) is due to the contraction property of the map $G(P_U\cdot)$ in Lemma 1. Provided that $\frac{I(U)rb}{p}\|M\| \le \lambda_r(M)/(16\eta)$ in the assumption, we have

$$K_0 \le T_0 + \lambda_r(M)/(16\eta). \tag{38}$$

Step 2. Next, we analyze the evolution of iterations by establishing an upper bound for $\|N^{(t)} - M\|$ based on $\|N^{(t-1)} - M\|$. By definition,

$$N^{(t)} - M = G\big(N^{(t)} - M\big) + \Gamma\big(N^{(t)} - M\big).$$

Since the entries indexed by $\mathcal{G}^c$ in $N^{(t)}$ do not change through iterations, $\|\Gamma(N^{(t)} - M)\| = \|\Gamma(N - M)\|$, which can be bounded by T_0. The analysis for $G(N^{(t)} - M)$ is more complicated. By definition, the entries indexed by $\mathcal{G}$ in $N^{(t)}$ is the same as the ones in $\tilde{N}^{(t-1)} = P_{U^{(t-1)}}N^{(t-1)}$. To align M with $P_{U^{(t-1)}}N^{(t-1)}$, we decompose $M = P_{U^{(t-1)}}M + P_{U_\perp^{(t-1)}}M$. Thus,

$$\begin{aligned} G\big(N^{(t)} - M\big) &= G\big(P_{U^{(t-1)}}N^{(t-1)} - P_{U^{(t-1)}}M - P_{U_\perp^{(t-1)}}M\big) \\ &= G\big(P_{U^{(t-1)}}\big(N^{(t-1)} - M\big)\big) - G\big(P_{U_\perp^{(t-1)}}M\big). \end{aligned}$$

It is still difficult to analyze $G(P_{U^{(t-1)}}(N^{(t-1)} - M))$ due to the complicated connection between $P_{U^{(t-1)}}$ and $N^{(t-1)} - M$. Thus, we decouple them by introducing $P_{U^{(t-1)}} = P_U + (P_{U^{(t-1)}} - P_U)$. Then, we have decomposed $G(N^{(t)} - M)$ into the following three terms:

$$\begin{aligned} G\big(N^{(t)} - M\big) &= G\big(P_U\big(N^{(t-1)} - M\big)\big) - G\big(P_{U_\perp^{(t-1)}}M\big) \\ &\quad + G\big((P_{U^{(t-1)}} - P_U)\cdot\big(N^{(t-1)} - M\big)\big). \end{aligned} \tag{39}$$

Next, we bound these three terms separately. In particular, the upper bound of $\|G(P_U(N^{(t-1)} - M))\|$ and $\|G(P_{U_\perp^{(t-1)}}M)\|$ can be achieved by the application of the contraction property for the map $G(P_U\cdot)$ in Lemma 1; to prove an upper bound for $\|G((P_{U^{(t-1)}} - P_U)\cdot(N^{(t-1)} - M))\|$, we apply the property of $\sin\Theta$ distance to relate $\|(P_{U^{(t-1)}} - P_U)\|$ to $\frac{\|N^{(t-1)}-M\|}{\lambda_r(M)}$. The detailed proofs are provided as follows.

- By Lemma 1,

$$\begin{aligned}\big\|G\big(P_U\big(N^{(t-1)} - M\big)\big)\big\| &\le \sqrt{\frac{I(U)rb(b\wedge r)}{p}}\big\|N^{(t-1)} - M\big\| \\ &= \sqrt{\frac{I(U)rb(b\wedge r)}{p}}K_{t-1}.\end{aligned} \tag{40}$$

- By Lemmas 1 and 5,

$$\begin{aligned}\big\|G(P_{U_\perp^{(t-1)}}M)\big\| &= \big\|G(P_{U_\perp^{(t-1)}}MP_U)\big\| \\ &\le \sqrt{\frac{I(U)rb(b\wedge r)}{p}}\|P_{U_\perp^{(t-1)}}M\| \\ &\le 2\sqrt{\frac{I(U)rb(b\wedge r)}{p}}\big\|N^{(t-1)} - M\big\| \\ &= 2\sqrt{\frac{I(U)rb(b\wedge r)}{p}}K_{t-1}.\end{aligned} \tag{41}$$

- Note that $U^{(t-1)}(U^{(t-1)})^\top$ and $UU^\top$ are both positive semi-definite and $\|U^{(t-1)}(U^{(t-1)})^\top\|\vee\|UU^\top\|\le 1$, we have $\|U^{(t-1)}(U^{(t-1)})^\top - UU^\top\|\le 1$. By Lemma 1 in [10],

$$\begin{aligned}\big\|U^{(t-1)}\big(U^{(t-1)}\big)^\top - UU^\top\big\| &\le 2\big\|\sin\Theta\big(U^{(t-1)},U\big)\big\|\wedge 1 \\ &= 2\big\|\big(U_\perp^{(t-1)}\big)^\top U\big\|\wedge 1 \\ &\le \big(2\big\|\big(U_\perp^{(t-1)}\big)^\top UU^\top M\big\|\cdot\lambda_{\min}^{-1}\big(U^\top M\big)\big)\wedge 1 \\ &\le \big(2\big\|\big(U_\perp^{(t-1)}\big)^\top M\big\|\cdot\lambda_r^{-1}(M)\big)\wedge 1 \\ &\le \bigg(\frac{4\|N^{(t-1)} - M\|}{\lambda_r(M)}\bigg)\wedge 1 = \frac{4K_{t-1}}{\lambda_r(M)}\wedge 1,\end{aligned}$$

where the penultimate step follows from Lemma 5. Note that

$$\begin{aligned}\operatorname{rank}\big((P_{U^{(t-1)}} - P_U)\big(N^{(t-1)} - M\big)\big) &\le \operatorname{rank}(P_{U^{(t-1)}} - P_U) \\ &\le \operatorname{rank}(P_{U^{(t-1)}}) + \operatorname{rank}(P_U) \le 2r,\end{aligned}$$

we have

$$\begin{aligned}&\big\|G\big((P_{U^{(t-1)}} - P_U)\cdot\big(N^{(t-1)} - M\big)\big)\big\| \\ &\quad\le \eta\cdot\|P_{U^{(t-1)}} - P_U\|\cdot\big\|N^{(t-1)} - M\big\| \le \eta K_{t-1}\cdot\bigg(\frac{4K_{t-1}}{\lambda_r(M)}\wedge 1\bigg).\end{aligned} \tag{42}$$

Combining (39)–(42), we have for all $t \geq 1$,

$$\begin{aligned} K_t &\leq \|\Gamma(N^{(t)} - M)\| + \|G(N^{(t)} - M)\| \\ &\leq T_0 + 3\sqrt{\frac{I(U)rb(b \wedge r)}{p}} K_{t-1} + \frac{4\eta}{\lambda_r(M)} K_{t-1}^2. \end{aligned} \tag{43}$$

Step 3. Finally, we use induction to show that, for all $t \geq 0$,

$$K_t \leq 2T_0 + 2^{-(t+4)}\lambda_r(M)/\eta. \tag{44}$$

The base case of $t = 0$ is proved by (38). Next, suppose the statement (44) holds for $t - 1$. Then

$$\begin{aligned} K_t &\overset{\text{(a)}}{\leq} T_0 + 3\sqrt{\frac{I(U)rb(b \wedge r)}{p}} K_{t-1} + \frac{4\eta}{\lambda_r(M)} K_{t-1}^2 \\ &\overset{\text{(b)}}{\leq} T_0 + \frac{K_{t-1}}{4} + K_{t-1}\left(\frac{8\eta T_0}{\lambda_r(M)} + \frac{1}{4}\right) \\ &\overset{\text{(c)}}{\leq} T_0 + \frac{K_{t-1}}{2} \overset{\text{(d)}}{\leq} T_0 + \frac{2T_0 + \lambda_r(M) \cdot (1/2)^{(t-1)+4}/\eta}{2} \\ &= 2T_0 + \lambda_r(M) \cdot (1/2)^{t+4}/\eta, \end{aligned}$$

where (a) is (43); (b) is due to the assumption $144I(U)rb(b \wedge r) \leq p$ and the induction hypothesis; (c) is from the assumption $T_0 \leq \lambda_r(M)/(64\eta)$; (d) is by the induction hypothesis.

Therefore, for all $t \geq \Omega(\log \frac{\lambda_r(M)}{T_0\eta} \vee 1) = \Omega(\log \frac{\lambda_r(M)}{\eta\|\Gamma(Z)\|} \vee 1)$, we have $K_t \leq 3T_0$. Finally, the desired (36) (37) follow from $\sin\Theta$ perturbation bound [37], Theorem 5, $q = \infty$. For completeness, we still provide a proof here. Let $M = USV^\top$ be the singular value decomposition of M, where S is diagonal and $U, V \in \mathbb{O}_{p,r}$. Then

$$\begin{aligned} \|\sin\Theta(\hat{U}^{(t)}, U)\| = \|(\hat{U}_\perp^{(t)})^\top U\| &\overset{(a)}{\leq} \frac{\|(\hat{U}_\perp)^\top USV^\top\|}{\lambda_r(SV^\top)} \overset{(b)}{=} \frac{\|(\hat{U}_\perp)^\top M\|}{\lambda_r(USV^\top)} \\ &\leq \frac{\|(\hat{U}_\perp)^\top(M + N^{(t)} - M)\| + \|(\hat{U}_\perp)^\top(N^{(t)} - M)\|}{\lambda_r(M)} \\ &\overset{(c)}{\leq} \frac{\|\lambda_{r+1}(N^{(t)})\| + \|N^{(t)} - M\|}{\lambda_r(M)} \overset{(d)}{\leq} \frac{\min_{\text{rank}(T)\leq r}\|N^{(t)} - T\| + K_t}{\lambda_r(M)} \\ &\leq \frac{\|N^{(t)} - M\| + K_t}{\lambda_r(M)} = \frac{2K_t}{\lambda_r(M)} = \frac{4T_0}{\lambda_r(M)} + (1/2)^{t+3}/\eta. \end{aligned}$$

Here, (a) holds because for any matrices $A \in \mathbb{R}^{p\times r}$, $B \in \mathbb{R}^{r\times p}$, by defining $x_* = \arg\max_{\|x\|_2=1}\|x^\top A\| = \|A\|$, then

$$\|AB\| = \sup_{\|x\|_2=1}\|x^\top AB\|_2 \geq \|(x_*^\top A)B\|_2 \geq \|x_*^\top A\|_2\lambda_r(B) = \|A\|\lambda_r(B);$$

(b) holds because U has orthonormal columns; (c) is because $(\hat{U})_\perp$ correspond to the $(r+1)$st, …, pth singular vector of $N^{(t)}$; (d) is due to Eckart–Young–Mirsky theorem.[3] Therefore, we have finished the proof of this theorem. □

[3]See a proof in https://en.wikipedia.org/wiki/Low-rank_approximation

REMARK 12. In fact, Theorem 7 implies Theorem 3. To see this, note that if the corruption set $\mathcal{G}$ is the diagonal, $\mathcal{G} = \{(i,i) : 1 \le i \le p\}$, we have

$$b = \max_i \{j : (i,j) \in \mathcal{G}\} \vee \max_j \{i : (i,j) \in \mathcal{G}\} = 1,$$

$$\eta = \max_M \|D(M)\|/\|M\| = \max_M \max_i \frac{|M_{ii}|}{\|M\|} = 1.$$

The next Lemma 1 provides an important technical tool for the proof of robust sin Θ theorem. It essentially shows that the operator norm of the composition of linear maps $G(P_U \cdot)$ is much smaller than the product of individual operator norms $\|G(\cdot)\|$ and $\|P_U\|$, provided that the basis U is incoherent; the same conclusion also applies to $G(\cdot P_V)$.

LEMMA 1. *Assume $\mathcal{G} \subseteq [m_1] \times [m_2]$ is b-sparse, that is, $\max_j\{i : (i,j) \in \mathcal{G}\} \vee \max_i\{j : (i,j) \in \mathcal{G}\} \le b$. Suppose $U \in \mathbb{O}_{m_1,r}$ and $V \in \mathbb{O}_{m_2,r}$. Recall that $G(A)$ is the matrix A with all entries in $\mathcal{G}^c$ set to zero, $I(U) = \frac{m_1}{r}\max_i \|e_i^\top U\|_2^2$, $I(V) = \frac{m_2}{r}\max_i \|e_i^\top V\|_2^2$, $P_U = UU^\top$, and $P_V = VV^\top$. Then for any matrix $A \in \mathbb{R}^{p_1 \times p_2}$, we have*

$$\|G(P_U A)\| \le \sqrt{\frac{I(U) r b (b \wedge r)}{m_1}}\|A\|,$$

$$\|G(A P_V)\| \le \sqrt{\frac{I(V) r b (b \wedge r)}{m_2}}\|A\|, \quad \text{and}$$

$$\|G(P_U A P_V)\| \le \frac{\sqrt{I(U)I(V)} \cdot rb}{\sqrt{m_1 m_2}}\|A\|.$$

In particular, recall that $D(A)$ is the matrix A with all off-diagonal entries set to zero. Suppose $U \in \mathbb{O}_{m,r}$. Then for any matrix $A \in \mathbb{R}^{m \times m}$,

$$\|D(P_U(D(A)))\| \le \frac{I(U) r}{m}\|D(A)\|, \qquad \|D(P_U A)\| \le \sqrt{\frac{I(U) r}{m}}\|A\|.$$

PROOF OF LEMMA 1.

$$\begin{aligned}
\|G(P_U A)\| &= \max_{\|v\|_2=1} \|v^\top G(UU^\top A)\|_2 \\
&= \max_{\|v\|_2=1} \left(\sum_{j=1}^{m_2} (v^\top [G(UU^\top A)]_{\cdot j})^2 \right)^{1/2} \\
&= \max_{\|v\|_2=1} \left(\sum_{j=1}^{m_2} \Big(\sum_{i:(i,j)\in\mathcal{G}} v_i (UU^\top A)_{i,j} \Big)^2 \right)^{1/2} \\
&\le \max_{\|v\|_2=1} \left(\sum_{j=1}^{m_2} \Big(\sum_{i:(i,j)\in\mathcal{G}} v_i^2 \Big) \Big(\sum_{i:(i,j)\in\mathcal{G}} (UU^\top A)_{i,j}^2 \Big) \right)^{1/2},
\end{aligned}$$

where the inequality is due to Cauchy–Schwarz. Now, for any $1 \le j \le m_2$,

$$\begin{aligned}
\sum_{i:(i,j)\in\mathcal{G}} (UU^\top A)_{ij}^2 &\le \sum_{i:(i,j)\in\mathcal{G}} (U_{i\cdot} \cdot (U^\top A)_{\cdot j})^2 = \sum_{i:(i,j)\in\mathcal{G}} \|U_{i\cdot}\|_2^2 \cdot \|(U^\top A)_{\cdot j}\|_2^2 \\
&\le \sum_{i:(i,j)\in\mathcal{G}} \frac{I(U) r}{m_1}\|A\|^2 \le \frac{I(U) r b}{m_1}\|A\|^2.
\end{aligned}$$

Thus,

$$\begin{aligned}\big\|G(P_U A)\big\| &\le \sqrt{\frac{I(U)rb}{m_1}}\|A\| \max_{\|v\|_2=1}\left(\sum_{j=1}^{m_2}\Big(\sum_{i:(i,j)\in\mathcal{G}} v_i^2\Big)\right)^{1/2}\\ &= \sqrt{\frac{I(U)rb}{m_1}}\|A\| \cdot \max_{\|v\|_2=1}\left(\sum_{i=1}^{m_1}\sum_{j:(i,j)\in\mathcal{G}} v_i^2\right)^{1/2}\\ &\le \sqrt{\frac{I(U)rb}{m_1}}\|A\| \cdot \max_{\|v\|_2=1}\left(\sum_{i=1}^{m_1} v_i^2 b\right)^{1/2} \le \sqrt{\frac{I(U)rb^2}{m_1}}\|A\|.\end{aligned}$$

Additionally, since $\operatorname{rank}(A)\le r$ and $\mathcal{G}$ is b-sparse,

$$\begin{aligned}\big\|G(P_U A)\big\|^2 &\le \big\|G(P_U A)\big\|_F^2 = \sum_{i,j}\big(G\big(UU^\top A\big)\big)_{ij}^2\\ &= \sum_{j=1}^{m_2}\sum_{i:(i,j)\in\mathcal{G}}\big(U_{i\cdot}\big(U^\top A\big)_{\cdot j}\big)^2\\ &\le \sum_{j=1}^{m_2}\sum_{i:(i,j)\in\mathcal{G}}\|U_{i\cdot}\|_2^2\cdot\big\|U^\top A_{\cdot j}\big\|_2^2 \le \sum_{j=1}^{m_2}\frac{I(U)rb}{m_1}\cdot\big\|U^\top A_{\cdot j}\big\|_2^2\\ &= \frac{I(U)rb}{m_1}\cdot\big\|U^\top A\big\|_F^2 \le \frac{I(U)rb}{m_1}\cdot r\big\|U^\top A\big\|^2 \le \frac{I(U)r^2 b}{m_1}\|A\|^2.\end{aligned}$$

Combining previous two inequalities, we have

$$\big\|G(P_U A)\big\| \le \sqrt{I(U)rb(r\wedge b)/m_1}\|A\|.$$

The proof for $\|G(AP_V)\| \le \sqrt{I(V)rb(b\wedge r)/m_2}\|A\|$ similarly follows. Next, for any $u\in\mathbb{R}^{m_1}$, $v\in\mathbb{R}^{m_2}$ such that $\|u\|_2=\|v\|_2=1$, we have

$$\begin{aligned}u^\top G(P_U A P_V)v &= u^\top G\big(UU^\top AVV^\top\big)v = \sum_{(i,j)\in\mathcal{G}} u_i v_j\big[UU^\top AVV^\top\big]_{ij}\\ &\le \sum_{(i,j)\in\mathcal{G}}|u_i v_j|\|U_{i\cdot}\|_2\cdot\big\|U^\top AV\big\|\cdot\|V_{j\cdot}\|_2\\ &\le \sum_{(i,j)\in\mathcal{G}}|u_i v_j|\cdot\sqrt{\frac{rI(U)}{m_1}}\cdot\|A\|\cdot\sqrt{\frac{rI(V)}{m_2}}\\ &\le \frac{\sqrt{I(U)I(V)}r\|A\|}{\sqrt{m_1 m_2}}\sum_{(i,j)\in\mathcal{G}}\frac{u_i^2+v_j^2}{2}\\ &\le \frac{\sqrt{I(U)I(V)}r\|A\|}{\sqrt{m_1 m_2}}\Big(\sum_i\sum_{j:(i,j)\in\mathcal{G}}\frac{u_i^2}{2}+\sum_j\sum_{i:(i,j)\in\mathcal{G}}\frac{v_j^2}{2}\Big)\\ &\le \frac{br\sqrt{I(U)I(V)}}{\sqrt{m_1 m_2}}\|A\|,\end{aligned}$$

which means $\|G(P_U AP_V)\| \le br\sqrt{I(U)I(V)/(m_1 m_2)}\|A\|$.

For the diagonal operator $D(\cdot)$, since $D(A)$ is a diagonal matrix, we have $D(A)e_i = D(A)_{ii}e_i$ and

$$\begin{aligned}\big\|D\big(P_U\big(D(A)\big)\big)\big\| &= \max_i \big|\big\{P_U\big(D(A)\big)\big\}_{ii}\big| = \max_i |e_i^\top P_U D(A) e_i| \\ &= \max_i |e_i^\top P_U e_i \cdot A_{ii}| = \max_i \big\|U^\top e_i\big\|_2^2 \cdot |A_{ii}| \le \frac{I(U)r}{m}\big\|D(A)\big\|, \\ \big\|D\big(P_U(A)\big)\big\| &= \max_i |(P_U A)_{ii}| = \max_i |e_i^\top U U^\top A e_i| \\ &\le \big\|e_i^\top U\big\|_2 \cdot \|A\| \le \sqrt{\frac{I(U)r}{m}}\|A\|. \end{aligned}$$

□

Funding. The research of Anru Zhang was supported in part by NSF CAREER award DMS-1944904, NSF Grant DMS-1811868, and NIH Grant R01-GM131399-01.

The research of Tony Cai was supported in part by NSF Grants DMS-1712735 and DMS-2015259 and NIH Grants R01-GM129781 and R01-GM123056.

The research of Yihong Wu was supported in part by the NSF Grant CCF-1527105, an NSF CAREER award CCF-1651588, and an Alfred Sloan fellowship.

SUPPLEMENTARY MATERIAL

Supplement to "Heteroskedastic PCA: Algorithm, optimality, and applications" (DOI: 10.1214/21-AOS2074SUPP; .pdf). The supplementary material contains additional discussions on the robust $\sin\Theta$ theorem and additional technical contents of this paper.

REFERENCES

[1] Bai, J. and Li, K. (2012). Statistical analysis of factor models of high dimension. *Ann. Statist.* **40** 436–465. MR3014313 https://doi.org/10.1214/11-AOS966

[2] Bai, Z. and Yao, J. (2012). On sample eigenvalues in a generalized spiked population model. *J. Multivariate Anal.* **106** 167–177. MR2887686 https://doi.org/10.1016/j.jmva.2011.10.009

[3] Baik, J., Ben Arous, G. and Péché, S. (2005). Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Ann. Probab.* **33** 1643–1697. MR2165575 https://doi.org/10.1214/009117905000000233

[4] Baik, J. and Silverstein, J. W. (2006). Eigenvalues of large sample covariance matrices of spiked population models. *J. Multivariate Anal.* **97** 1382–1408. MR2279680 https://doi.org/10.1016/j.jmva.2005.08.003

[5] Bartlett, M. S. (1937). The statistical conception of mental factors. *Br. J. Health Psychol.* **28** 97.

[6] Bubeck, S. (2015). Convex optimization: Algorithms and complexity. *Found. Trends Mach. Learn.* **8** 231–357.

[7] Cai, J.-F. and Wei, K. (2018). Exploiting the structure effectively and efficiently in low-rank matrix recovery. In *Processing, Analyzing and Learning of Images, Shapes, and Forms. Part* 1. *Handb. Numer. Anal.* **19** 21–51. North-Holland, Amsterdam. MR3839040

[8] Cai, T. T., Han, R. and Zhang, A. R. (2020). On the non-asymptotic concentration of heteroskedastic Wishart-type matrix. ArXiv preprint. Available at arXiv:2008.12434.

[9] Cai, T. T. and Zhang, A. (2016). Minimax rate-optimal estimation of high-dimensional covariance matrices with incomplete data. *J. Multivariate Anal.* **150** 55–74. MR3534902 https://doi.org/10.1016/j.jmva.2016.05.002

[10] Cai, T. T. and Zhang, A. (2018). Rate-optimal perturbation bounds for singular subspaces with applications to high-dimensional statistics. *Ann. Statist.* **46** 60–89. MR3766946 https://doi.org/10.1214/17-AOS1541

[11] Candès, E. J. and Recht, B. (2009). Exact matrix completion via convex optimization. *Found. Comput. Math.* **9** 717–772. MR2565240 https://doi.org/10.1007/s10208-009-9045-5

[12] Candès, E. J., Sing-Long, C. A. and Trzasko, J. D. (2013). Unbiased risk estimates for singular value thresholding and spectral estimators. *IEEE Trans. Signal Process.* **61** 4643–4657. MR3105401 https://doi.org/10.1109/TSP.2013.2270464

[13] CANDÈS, E. J. and TAO, T. (2010). The power of convex relaxation: Near-optimal matrix completion. *IEEE Trans. Inf. Theory* **56** 2053–2080. MR2723472 https://doi.org/10.1109/TIT.2010.2044061
[14] CAO, Y., ZHANG, A. and LI, H. (2020). Multisample estimation of bacterial composition matrices in metagenomics data. *Biometrika* **107** 75–92. MR4064141 https://doi.org/10.1093/biomet/asz062
[15] CHATTERJEE, S. (2015). Matrix estimation by universal singular value thresholding. *Ann. Statist.* **43** 177–214. MR3285604 https://doi.org/10.1214/14-AOS1272
[16] COLLINS, M., DASGUPTA, S. and SCHAPIRE, R. E. (2002). A generalization of principal components analysis to the exponential family. In *Advances in Neural Information Processing Systems* 617–624.
[17] DAVIS, C. and KAHAN, W. M. (1970). The rotation of eigenvectors by a perturbation. III. *SIAM J. Numer. Anal.* **7** 1–46. MR0264450 https://doi.org/10.1137/0707001
[18] DE LATHAUWER, L., DE MOOR, B. and VANDEWALLE, J. (2000). On the best rank-1 and rank-$(R_1, R_2, \ldots, R_N)$ approximation of higher-order tensors. *SIAM J. Matrix Anal. Appl.* **21** 1324–1342. MR1780276 https://doi.org/10.1137/S0895479898346995
[19] DONOHO, D. and GAVISH, M. (2014). Minimax risk of matrix denoising by singular value thresholding. *Ann. Statist.* **42** 2413–2440. MR3269984 https://doi.org/10.1214/14-AOS1257
[20] DONOHO, D., GAVISH, M. and JOHNSTONE, I. (2018). Optimal shrinkage of eigenvalues in the spiked covariance model. *Ann. Statist.* **46** 1742–1778. MR3819116 https://doi.org/10.1214/17-AOS1601
[21] FLORESCU, L. and PERKINS, W. (2016). Spectral thresholds in the bipartite stochastic block model. In *Conference on Learning Theory* 943–959.
[22] FORTUNATO, S. (2010). Community detection in graphs. *Phys. Rep.* **486** 75–174. MR2580414 https://doi.org/10.1016/j.physrep.2009.11.002
[23] GAVISH, M. and DONOHO, D. L. (2017). Optimal shrinkage of singular values. *IEEE Trans. Inf. Theory* **63** 2137–2152. MR3626861 https://doi.org/10.1109/TIT.2017.2653801
[24] GHOSH, J. and DUNSON, D. B. (2009). Default prior distributions and efficient posterior computation in Bayesian factor analysis. *J. Comput. Graph. Statist.* **18** 306–320. MR2749834 https://doi.org/10.1198/jcgs.2009.07145
[25] HONG, D., BALZANO, L. and FESSLER, J. A. (2016). Towards a theoretical analysis of PCA for heteroscedastic data. In *Communication, Control, and Computing (Allerton)*, 2016 54*th Annual Allerton Conference on* 496–503. IEEE, Los Alamitos.
[26] HONG, D., BALZANO, L. and FESSLER, J. A. (2018). Asymptotic performance of PCA for high-dimensional heteroscedastic data. *J. Multivariate Anal.* **167** 435–452. MR3830656 https://doi.org/10.1016/j.jmva.2018.06.002
[27] HONG, D., BALZANO, L. and FESSLER, J. A. (2018). Asymptotic performance of PCA for high-dimensional heteroscedastic data. *J. Multivariate Anal.* **167** 435–452. MR3830656 https://doi.org/10.1016/j.jmva.2018.06.002
[28] HOTELLING, H. (1936). Relations between two sets of variates. *Biometrika* **28** 321–377.
[29] JAIN, P., NETRAPALLI, P. and SANGHAVI, S. (2013). Low-rank matrix completion using alternating minimization (extended abstract). In *STOC'*13*—Proceedings of the* 2013 *ACM Symposium on Theory of Computing* 665–674. ACM, New York. MR3210828 https://doi.org/10.1145/2488608.2488693
[30] JOHNSTONE, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.* **29** 295–327. MR1863961 https://doi.org/10.1214/aos/1009210544
[31] KESHAVAN, R. H. (2012). Efficient algorithms for collaborative filtering Ph.D. thesis Stanford Univ.
[32] KESHAVAN, R. H., MONTANARI, A. and OH, S. (2010). Matrix completion from a few entries. *IEEE Trans. Inf. Theory* **56** 2980–2998. MR2683452 https://doi.org/10.1109/TIT.2010.2046205
[33] KESHAVAN, R. H., MONTANARI, A. and OH, S. (2010). Matrix completion from noisy entries. *J. Mach. Learn. Res.* **11** 2057–2078. MR2678022
[34] LAWLEY, D. N. and MAXWELL, A. E. (1962). Factor analysis as a statistical method. *J. R. Stat. Soc., Ser. D, Stat.* **12** 209–229.
[35] LIU, L. T., DOBRIBAN, E. and SINGER, A. (2018). *e*PCA: High dimensional exponential family PCA. *Ann. Appl. Stat.* **12** 2121–2150. MR3875695 https://doi.org/10.1214/18-AOAS1146
[36] LOUNICI, K. (2014). High-dimensional covariance matrix estimation with missing observations. *Bernoulli* **20** 1029–1058. MR3217437 https://doi.org/10.3150/12-BEJ487
[37] LUO, Y., HAN, R. and ZHANG, A. R. (2020). A Schatten-q matrix perturbation theory via perturbation projection error bound. ArXiv preprint. Available at arXiv:2008.01312.
[38] MARTIN, A. D., QUINN, K. M. and PARK, J. H. (2011). MCMCpack: Markov chain Monte Carlo in R. *J. Stat. Softw.* **42**.
[39] MAZUMDER, R., HASTIE, T. and TIBSHIRANI, R. (2010). Spectral regularization algorithms for learning large incomplete matrices. *J. Mach. Learn. Res.* **11** 2287–2322. MR2719857
[40] MELAMED, D. (2014). Community structures in bipartite networks: A dual-projection approach. *PLoS ONE* **9** e97823. https://doi.org/10.1371/journal.pone.0097823

[41] MOHAMED, S., GHAHRAMANI, Z. and HELLER, K. A. (2009). Bayesian exponential family PCA. In *Advances in Neural Information Processing Systems* 1089–1096.
[42] NADAKUDITI, R. R. (2014). OptShrink: An algorithm for improved low-rank signal matrix denoising by optimal, data-driven singular value shrinkage. *IEEE Trans. Inf. Theory* **60** 3002–3018. MR3200641 https://doi.org/10.1109/TIT.2014.2311661
[43] NADLER, B. (2008). Finite sample approximation results for principal component analysis: A matrix perturbation approach. *Ann. Statist.* **36** 2791–2817. MR2485013 https://doi.org/10.1214/08-AOS618
[44] OWEN, A. B. and WANG, J. (2016). Bi-cross-validation for factor analysis. *Statist. Sci.* **31** 119–139. MR3458596 https://doi.org/10.1214/15-STS539
[45] PAUL, D. (2007). Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statist. Sinica* **17** 1617–1642. MR2399865
[46] RECHT, B. (2011). A simpler approach to matrix completion. *J. Mach. Learn. Res.* **12** 3413–3430. MR2877360
[47] ROBIN, G., JOSSE, J., MOULINES, É. and SARDY, S. (2019). Low-rank model with covariates for count data with missing values. *J. Multivariate Anal.* **173** 416–434. MR3946894 https://doi.org/10.1016/j.jmva.2019.04.004
[48] SALMON, J., HARMANY, Z., DELEDALLE, C.-A. and WILLETT, R. (2014). Poisson noise reduction with non-local PCA. *J. Math. Imaging Vision* **48** 279–294. MR3152105 https://doi.org/10.1007/s10851-013-0435-6
[49] THOMSON, G. (1939). The factorial analysis of human ability. *Br. J. Educ. Psychol.* **9** 188–195.
[50] TIPPING, M. E. and BISHOP, C. M. (1999). Probabilistic principal component analysis. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **61** 611–622. MR1707864 https://doi.org/10.1111/1467-9868.00196
[51] VASWANI, N. and GUO, H. (2016). Correlated-PCA: Principal components' analysis when data and noise are correlated. In *Advances in Neural Information Processing Systems* 1768–1776.
[52] VASWANI, N. and NARAYANAMURTHY, P. (2017). Finite sample guarantees for PCA in non-isotropic and data-dependent noise. In 2017 55*th Annual Allerton Conference on Communication, Control, and Computing* (*Allerton*) 783–789. IEEE, Los Alamitos.
[53] VASWANI, N. and NARAYANAMURTHY, P. (2018). PCA in sparse data-dependent noise. In 2018 *IEEE International Symposium on Information Theory* (*ISIT*) 641–645. IEEE, Los Alamitos.
[54] VASWANI, N. and NARAYANAMURTHY, P. (2020). Fast robust subspace tracking via PCA in sparse data-dependent noise. *IEEE J. Sel. Areas Inf. Theory* **1** 723–744.
[55] VERSHYNIN, R. (2012). Introduction to the non-asymptotic analysis of random matrices. In *Compressed Sensing* 210–268. Cambridge Univ. Press, Cambridge. MR2963170
[56] WANG, W. and FAN, J. (2017). Asymptotics of empirical eigenstructure for high dimensional spiked covariance. *Ann. Statist.* **45** 1342–1374. MR3662457 https://doi.org/10.1214/16-AOS1487
[57] WEDIN, P. (1972). Perturbation bounds in connection with singular value decomposition. *BIT* **12** 99–111. MR0309968 https://doi.org/10.1007/bf01932678
[58] YAO, J., ZHENG, S. and BAI, Z. (2015). *Large Sample Covariance Matrices and High-Dimensional Data Analysis*. *Cambridge Series in Statistical and Probabilistic Mathematics* **39**. Cambridge Univ. Press, New York. MR3468554 https://doi.org/10.1017/CBO9781107588080
[59] YU, Y., WANG, T. and SAMWORTH, R. J. (2015). A useful variant of the Davis–Kahan theorem for statisticians. *Biometrika* **102** 315–323. MR3371006 https://doi.org/10.1093/biomet/asv008
[60] ZHANG, A., CAI, T. T. and WU, Y. (2022). Supplement to "Heteroskedastic PCA: Algorithm, optimality, and applications." https://doi.org/10.1214/21-AOS2074SUPP
[61] ZHANG, A. and XIA, D. (2018). Tensor SVD: Statistical and computational limits. *IEEE Trans. Inf. Theory* **64** 7311–7338. MR3876445 https://doi.org/10.1109/TIT.2018.2841377
[62] ZHOU, Z. and AMINI, A. A. (2020). Optimal bipartite network clustering. *J. Mach. Learn. Res.* **21** 40. MR4073773