1 Introduction

Principal Component Analysis (PCA) is a fundamental technique widely used to uncover hidden low-rank structures from high-dimensional noisy datasets. This is typically achieved by modeling the data through a low-rank-plus-noise structure, such as the spiked covariance model, and applying PCA to identify the leading signal components.

However, traditional PCA has notable limitations, particularly under heteroskedasticity noise cases, where the variance of the noise differs across dimensions. Classical PCA inherently assumes noise has uniform variance, an assumption that is frequently violated in practical applications. Under heteroskedastic conditions, the sample covariance matrix becomes significantly biased, particularly on its diagonal, which can dominate its spectral norm and result in inaccurate subspace recovery.

To address this issue, we turn to Heteroskedastic PCA, which is an iterative, diagonal-debiasing technique that remains accurate even when noise variances differ across coordinates. By repeatedly zeroing the sample-covariance diagonal and projecting onto a low-rank (rank $\leq r$) positive-semidefinite matrix set, HeteroPCA recovers the principal subspace with substantially smaller error than classical PCA under heteroskedastic noise. Beyond its core algorithm and theory, the method has proved adaptable: it extends naturally to matrices with missing entries and to non-Gaussian settings such as Poisson observations, and subsequent work has further equipped it with row-wise and entry-wise inferential tools.

2 Bound Comparison for Heteroskedastic PCA and Standard PCA

Bound for Standard PCA Method

We first discuss the upper bound on the angular error between the estimated subspace and the true subspace using the standard PCA method. Consider the spiked covariance model:

$$Y_k = X_k + \varepsilon_k, \quad X_k \sim \mathcal{N}(0, \Sigma_0), \quad \varepsilon_k \sim \mathcal{N}(0, D)$$

where $D = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_p^2)$ represents heteroskedastic noise. Classical PCA ignores diagonal bias, which leads to distorted estimates of eigenvalues and eigenvectors.

According to the Davis–Kahan $\sin\Theta$ theorem, let the true covariance matrix Σ admit the eigendecomposition:

$$\Sigma = U\Lambda U^\top + U^\perp \Lambda^\perp (U^\perp)^\top$$

and let $\hat{\Sigma}$ be the perturbed version. Define the eigengap as $\delta = \lambda_r(\Lambda) - \lambda_{r+1}(\Lambda)$. Then the angular error satisfies:

$$\|\sin\Theta(\tilde{U}, U)\| \leq \frac{\|\hat{\Sigma} - \Sigma\|}{\delta}$$

In our setting, set $\Sigma = \Sigma_0 + \beta I_p$. Then the bound becomes:

$$\|\sin\Theta(\tilde{U}, U)\| \lesssim \frac{\|\hat{\Sigma} - (\Sigma_0 + \beta I_p)\|}{\lambda_r(\Lambda)} \wedge 1$$

Here, $\lambda_r(\Lambda)$ denotes the smallest non-zero eigenvalue of the low-rank signal component Σ_0, which dominates the eigengap. Σ_0 trailing eigenvalues are all zero:

$$\lambda_1(\Lambda) \geq \cdots \geq \lambda_r(\Lambda) > 0, \quad \lambda_{r+1}(\Lambda) = \cdots = \lambda_p(\Lambda) = 0$$

Therefore, the eigengap is $\delta = \lambda_r(\Lambda) - 0 = \lambda_r(\Lambda)$

Also the use of $\wedge 1$ ensures that the bound does not exceed 1.

Bound for Hetero PCA Method

We now force on the upper bound on the angular error between the estimated subspace and the true subspace using the Hetero PCA method. There are three steps in total:

Step 1: Upper bounding the numerator

Our goal is to bound the deviation between the noisy and clean sample covariance matrices:

$$\left\|\hat{\Sigma} - \hat{\Sigma}_X\right\| = \left\|\frac{1}{n-1}(YY^\top - n\bar{Y}\bar{Y}^\top) - \frac{1}{n-1}(XX^\top - n\bar{X}\bar{X}^\top)\right\|.$$

This is equivalent to bound:

$$(n-1)(\hat{\Sigma}-\hat{\Sigma}_X) = YY^\top - n\bar{Y}\bar{Y}^\top - (XX^\top - n\bar{X}\bar{X}^\top).$$

Using the model $Y = X + E$, we expand:

$$(n-1)(\hat{\Sigma}-\hat{\Sigma}_X) = XE^\top + EX^\top + EE^\top - n(\bar{X}\bar{E}^\top + \bar{E}\bar{X}^\top + \bar{E}\bar{E}^\top).$$

Applying heteroskedastic matrix concentration inequalities [2] also applying Theorem 6 and Lemma 1 [3], we can control the leading error terms and obtain:

$$\mathbb{E}_E\big\|\Delta\big((n-1)(\hat{\Sigma}-\hat{\Sigma}_X)\big)\big\| \lesssim \sqrt{n}\,\sigma_{\text{sum}}\sigma_{\max} + \sigma_{\text{sum}}^2 + \|X\|\,(\sigma_{\text{sum}} + \sqrt{r}\sigma_{\max}) + n^{1/2}\|\bar{X}\|_2\,\sigma_{\text{sum}}.$$

Step 2: Lower bounding the denominator
We aim to estimate the smallest nonzero eigenvalue of the sample signal covariance matrix $\hat{\Sigma}_X$ to lower bound $\lambda_r(n\hat{\Sigma}_X)$.
Let $\Gamma \in \mathbb{R}^{n\times r}$ be a matrix with independent, isotropic sub-Gaussian columns. Based on matrix concentration theory [4], we have:

$$\mathbb{P}\left(\sqrt{n} + C\sqrt{r} + t \geq \|\Gamma\| \geq \sqrt{n} - C\sqrt{r} - t\right) \geq 1 - \exp(-Ct^2/2).$$

Choosing $t = \sqrt{n}$, we obtain the following with at least $1 - Cexp(cn)$ probability:

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

Given that $X = U\Lambda^{1/2}\Gamma^\top$, we express the sample covariance matrix as:

$$n\hat{\Sigma}_X = n(XX^\top - n\bar{X}\bar{X}^\top),$$

and its eigenvalue satisfies:

$$\lambda_r(n\hat{\Sigma}_X) \geq \lambda_r(\Lambda)\cdot\lambda_r(\Gamma\Gamma^\top - n\bar{\Gamma}\bar{\Gamma}^\top) \gtrsim n\lambda_r(\Lambda).$$

Hence, by applying random matrix theory, we obtain a nontrivial lower bound on the effective signal strength.
Step 3: Estimating subspace error by sinΘ theorem
From robust perturbation theory:

$$\mathbb{E}\|\sin\Theta(\hat{U}, U)\| \lesssim \frac{\mathbb{E}\|\Delta((n-1)(\hat{\Sigma}-\hat{\Sigma}_X))\|}{\lambda_r((n-1)\hat{\Sigma}_X)}$$

Thus, using the bound from Step 1 and Step 2:

$$\mathbb{E}\|\sin\Theta(\hat{U}, U)\| \lesssim \underbrace{\frac{\sigma_{\text{sum}} + \sqrt{r}\sigma_{\max}}{(n\lambda_r(\Lambda))^{1/2}}}_{\text{noise fluctuation}} + \underbrace{\frac{\sigma_{\text{sum}}\sigma_{\max}}{n^{1/2}\lambda_r(\Lambda)}}_{\text{higher-order terms}}$$

Why the Heteroskedastic PCA Bound is Tighter

1. **Diagonal bias removal.**
   Classical PCA keeps the deterministic diagonal bias $\text{diag}(\sigma_1^2, \ldots, \sigma_p^2)$ inside the perturbation norm $\|\hat{\Sigma} - (\Sigma_0 + \beta I_p)\|$; HeteroPCA applies the operator $\Delta(\cdot)$ that zeroes out all diagonals, so the numerator becomes $\big\|\Delta(\hat{\Sigma} - \Sigma_X)\big\|$ and no longer scales with the largest noise variance $\sigma_{\max}$.

2. **Signal eigenvalue is preserved.**
   Both bounds divide by a signal strength (eigengap). HeteroPCA shows $\lambda_r\big((n-1)\hat{\Sigma}_X\big) \sim n\lambda_r(\Lambda)$ in probability, hence the denominator stays of the same order as in the classical bound.

3. **Fluctuation vs. worst–case.**
   The new numerator depends on $\sigma_{\text{sum}} + \sqrt{r}\,\sigma_{\max}$ (root–sum fluctuation), whereas the classical one is dominated by the single largest $\sigma_{\max}$. The resulting leading term is

$$\frac{\sigma_{\text{sum}} + \sqrt{r}\,\sigma_{\max}}{(n\lambda_r(\Lambda))^{1/2}},$$

   strictly smaller whenever $\sigma_{\text{sum}} < p\,\sigma_{\max}$, which is the generic heteroskedastic case.

4. **Expectation-level guarantee.**
   HeteroPCA reports the *expected* subspace error, reflecting an average–case behaviour; the classical Davis–Kahan bound is worst–case in operator norm, hence looser in practice.

Consequently, by isolating diagonal bias and focusing on true random fluctuations, the Heteroskedastic PCA bound delivers a strictly tighter (often order–optimal) guarantee for subspace recovery under heteroskedastic noise.

3 Key Assumptions and Limiations of the HeteroPCA Algorithm

Entrywise vs. Columnwise Heteroskedasticity

For the heteroskedastic noise model, Zhang et al. give the following expression relating the expectation of the sample gram matrix $N = YY^\intercal$ and the gram matrix of the underlying data matrix $M = XX^\intercal$:

$$\mathbb{E}N_{i,j} = M_{i,j} + \sum_{k=1}^{p_2} \text{Var}(E_{i,k})$$

However, if furhter assume that $(\forall k \in [p_2]) \quad \text{Var}(E_{i,k}) = \text{Var}(E_{j,k})$ (i.e. variance is equal in a given column).

$$\delta = \sum_{k=1}^{p_2} \text{Var}(E_{i,k}) = \sum_{k=1}^{p_2} \text{Var}(E_{j,k})$$

$$\mathbb{E}N = M + \delta I$$

Since M is symmetric, we can take the diagonal decomposition

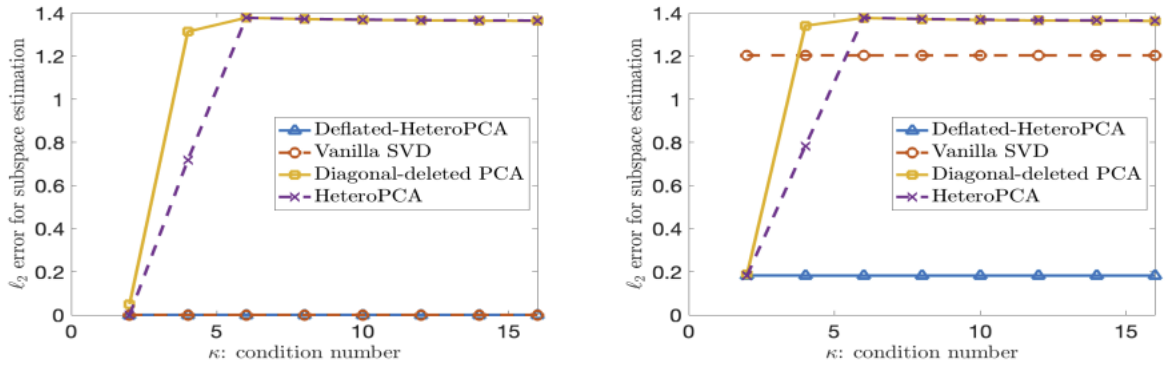
$$M = U\Lambda U^\intercal$$

$$\mathbb{E}N = U(\Lambda + \delta I)U^\intercal$$

Most critically, this very simple proof demonstrates that weakening the assumption of entry-wise heteroskedasticity to only column-wise heteroskedasticity guarantees that the eigenvectors of $\mathbb{E}N$ (and their corresponding order by size) are identical to those of M. This is meaningful because this bias is the initial motivation the paper cites for the insufficiency of both diagonal deletion and traditional SVD for the hetroskedastic noise problem.

The impact of the distinction between entrywise and columnwise heteroskedasticity is further shown numerically in the experiments conducted by Hong et al. in "HePPCAT: Probabilistic PCA for Data With Heteroscedastic Noise" [9]. This paper focuses on addressing heteroskedasticity across samples, which correspond to columns in the example above. This means that our above assumption of equal variance in within each column is an assumption for this paper.

In the experiment, matrices are generated under noise that is heteroskedastic across samples. Then, the subspaces of these matrices are estimated using weighted PCA (inverse noise variance and square inverse noise variance), HeteroPCA, and HePPCAT. In this numerical experiment, HeteroPCA was the worst performing under every metric shown: normalized subspace estimation error and recovery of the first three principle components. It performed particularly poorly in the recovery of the first two principle components as the noise variance grew larger. This experiment demonstrates that HePPCAT may offer superior recovery to HeteroPCA in situations where heteroskedasticity only holds across samples rather than across all entries.



(a) Noiseless case

(b) Noisy case

Impact of the Condition Number on HeteroPCA

One limitation of a few of the bounds given in the HeteroPCA paper is their assumption in an at most constant condition number. For example, this assumption is made in remark 3, a formula describing the optimal estimation error rate, Theorem 4, which provides the main bound for the subspace estimation error for the SVD under heteroskedastic noise problem, and a similar bound placing $\frac{||\Lambda||}{\lambda_r(\Lambda)} \leq C$ for Theorem 1, which generates the bound demonstrated earlier in this paper.

Specifically, this assumption is invoked in Theorem 1 is used to show:

$$||\bar{X}||_2 \lesssim (n\lambda_r(\Lambda))^{1/2}$$

This quantity ends up forming the numerator of our noise fluctuation term shown at the conclusion in Step 3 of the first proof in this paper. This shows that this bound on the condition number is very critical to the final bounds generated.

The problem of HeteroPCA's dependence on a controlled condition number is further explored in Zhou et al.'s paper "Deflated HeteroPCA: Overcoming the curse of ill-conditioning in heteroskedastic PCA" [10]. They begin the paper by demonstrating the insufficiency of HeteroPCA in circumstances where the condition number grows large by an experiment. By generating random low rank underlying truth matrices and then applying heteroskedastic nosie to one version but leaving the first entirely without random noise, Zhou et al. generated the following figures comparing the ℓ_2 subspace estimation error across a variety of methods as κ increases:

Beyond simply showing that larger condition numbers can cause HeteroPCA to underperform on data with heteroskedastic noise, the left plot demonstrates that a large condition number can actually cause both diagonal delted PCA and HeteroPCA to perform worse under an increasing condition number even when the dataset is entirely noiseless. Meanwhile, in the noisy case, both HeteroPCA and diagonal deletion are outperformed by vanilla SVD once the condition number becomes large enough. These results show that, beyond simply comprimising the assumptions behind our bounds, large condition numbers can cause the performance of HeteroPCA to significantly degrade. In order to remedy this shortcoming, the authors of the paper designed a novel algorithm, called Deflated-HeteroPCA. In short, this algorithm divides a given matrix, which may have a large condition number, into well conditioned subblocks where HeteroPCA can be properly applied. This algorithm also achieves the same minimax bounds of the original HeteroPCA algorithm. Therefore, in cases where a large condition number is a possible concern, Deflated-HeteroPCA will be much more robust to these matrices than the original algorithm.

Reference List:

[1] Yu, Y., Wang, T., and Samworth, R. J. (2015). A useful variant of the Davis–Kahan theorem for statisticians. *Biometrika*, 102(2), 315–323.
[2] CAI,T.T.,HAN,R.andZHANG, A. R. (2020). On the non-asymptotic concentration of heteroskedastic Wishart-type matrix. ArXiv preprint. Available at arXiv:2008.12434.
[3] Zhang, A. R., Cai, T. T. and Wu, Y. (2022). Heteroskedastic PCA: Algorithm, Optimality, and Applications. The Annals of Statistics, 50(1), 53–80. https://doi.org/10.1214/21-AOS2074
[4] VERSHYNIN, R. (2012). Introduction to the non-asymptotic analysis of random matrices. In Compressed Sensing 210–268. Cambridge Univ. Press, Cambridge. MR2963170
[5] Bao, Z., Ding, X., Wang, J., and Wang, K. (2022). Statistical inference for principal components of spiked covariance matrices. The Annals of Statistics, 50(2):1144–1169.
[6] Yan, Y., Chen, Y., and Fan, J. (2024). Inference for heteroskedastic PCA with missing data. The Annals of Statistics, 52(2), 729–756. https://doi.org/10.1214/24-AOS2366
[7] Wahba, G. (1965). A least squares estimate of satellite attitude. SIAM review, 7(3):409–409.
[8] Cai, C., Li, G., Chi, Y., Poor, H. V., and Chen, Y. (2021). Subspace estimation from unbalanced and incomplete data matrices: $\ell_{2,\infty}$ statistical guarantees. The Annals of Statistics, 49(2):944–967.
[9] Hong, D., Gilman, K., Balzano, L., and Fessler, J. A. (2021). HePPCAT: Probabilistic PCA for Data With Heteroscedastic Noise. IEEE Transactions on Signal Processing, vol. 69, pp. 4819-4834.
[10] Zhou, Y., Chen, Y. (2024). Deflated HeteroPCA: Overcoming the curse of ill-conditioning in heteroskedastic PCA. Annals of Statistics, 53(1):91-116.

Appendix

A Introduction

Principal Component Analysis (PCA) is a foundational tool in modern data analysis, commonly used for dimensionality reduction, noise filtering, and feature extraction. Classical PCA assumes homoskedastic noise and complete data, which are often unrealistic assumptions in high-dimensional settings. Recent work has focused on modifying PCA to account for heteroskedastic noise—that is, noise with non-uniform variance across observations—as well as missing or corrupted entries.

While several methods have been proposed to address these limitations, including diagonal-deletion and de-biasing techniques, they often suffer from instability, reduced accuracy, or lack of theoretical guarantees under realistic data conditions. One notable contribution, *Heteroskedastic PCA: Algorithm, Optimality, and Applications*, introduced a framework for consistent subspace estimation under heteroskedasticity by modifying the spectral structure of the sample covariance matrix. However, that paper focuses primarily on estimating the principal subspace and does not provide elementwise uncertainty quantification or robust procedures for inference on individual matrix entries.

Inference for Heteroskedastic PCA with Missing Data addresses these gaps by proposing a new framework that not only recovers the underlying low-rank structure in heteroskedastic noise but also quantifies uncertainty at the entrywise level. This is particularly relevant in applications where individual matrix entries carry scientific or operational significance, such as in recommender systems, genomics, and signal processing. The paper introduces novel estimators and confidence intervals designed to handle noise heterogeneity, backed by non-asymptotic theoretical guarantees and empirical results demonstrating superior performance over prior methods.

B Motivation and Problem Setup

Classical PCA, despite its widespread utility, is known to perform poorly when applied to data with heteroskedastic noise or missing entries—both of which are common in high-dimensional applications. In such cases, the empirical covariance matrix becomes a biased and inconsistent estimator of the population covariance, especially in the presence of non-uniform variances across rows or columns.

The paper *Heteroskedastic PCA: Algorithm, Optimality, and Applications* provides a significant step forward by proposing a spectral method, HeteroPCA, that consistently estimates the principal components under heteroskedastic noise. Although missing data is not an explicit focus of the paper, HeteroPCA is able to implicitly handle missing entries through its iterative structure. Specifically, each iteration of HeteroPCA performs a best rank-r approximation via singular value decomposition (SVD), which functions as a matrix completion step. This enables the algorithm to refine the low-rank approximation of the signal matrix while filling in unobserved entries without requiring explicit imputation.

However, while HeteroPCA recovers the principal subspace effectively, it does not provide elementwise statistical inference, such as confidence intervals for individual entries of the signal matrix. In many real-world applications—including recommender systems, genomics, and spatio-temporal signal analysis—it is crucial not only to estimate a low-rank signal matrix but also to quantify uncertainty on a per-entry basis, especially when decisions or predictions are made at the individual level.

Inference for Heteroskedastic PCA with Missing Data is motivated by this need for robust and fine-grained inference under practical noise and sampling constraints. The paper considers the model

$$Y = X + E,$$

where X is an unknown low-rank signal matrix and E is a noise matrix with independent, mean-zero, heteroskedastic entries. Critically, only a subset of the entries of Y is observed. The goal is to estimate entries of X and construct entrywise confidence intervals, while accounting for both missing data and noise variability.

This formulation captures a wide range of high-dimensional problems where data is incomplete and noisy. The key challenge is to disentangle the heteroskedastic noise structure from the low-rank signal and the randomness induced by missingness, in order to enable statistically valid inference at the entry level.

C Summary of Contributions

Inference for Heteroskedastic PCA with Missing Data makes several key contributions to the study of low-rank matrix estimation under realistic noise and sampling conditions:

- **Entrywise Inference Framework:** The paper introduces a principled framework for constructing entrywise confidence intervals for the elements of a low-rank matrix corrupted by heteroskedastic noise and partially observed data. This addresses a major gap in prior work, including *Heteroskedastic PCA: Algorithm, Optimality, and Applications*, which focuses on subspace recovery rather than inference on individual entries.
- **Theoretical Guarantees:** The paper provides non-asymptotic theoretical guarantees for the proposed confidence intervals. These results hold under a random missingness model and do not require uniform noise assumptions, making them broadly applicable.
- **Empirical Validation:** Through simulations and real-data experiments, the authors demonstrate that their method outperforms previous approaches in terms of coverage accuracy and robustness under noise heterogeneity. The results highlight both the practical utility and theoretical soundness of the method.

Taken together, these contributions provide a substantial advancement in the development of statistically valid, entrywise inference tools for heteroskedastic and incomplete data settings. The work bridges a critical gap between low-rank matrix recovery and uncertainty quantification.

D Technical Framework

D.1 Model and Assumptions

The paper considers the standard matrix denoising model with heteroskedastic and incomplete observations:

$$Y = X + E,$$

where $Y \in \mathbb{R}^{d\times n}$ is the observed data matrix, $X \in \mathbb{R}^{d\times n}$ is the unknown low-rank signal matrix of interest, and $E \in \mathbb{R}^{d\times n}$ is a noise matrix with independent, mean-zero entries. The matrix X is assumed to have rank $r \ll \min(d, n)$.

A key feature of this model is the heteroskedasticity of the noise: the entries of E are assumed to satisfy

$$\mathbb{E}[E_{ij}] = 0, \quad \mathbb{E}[E_{ij}^2] = \omega_{ij}^2,$$

where the variances ω_{ij}^2 may vary arbitrarily across i and j, subject to mild moment conditions. This contrasts with homoskedastic models, where the noise variance is constant across entries.

In addition to noise heterogeneity, the paper assumes that only a random subset of the entries of Y is observed. Let $\Omega \subset [d] \times [n]$ denote the set of observed indices. The sampling process is modeled as uniform random sampling, where each entry is observed independently with probability p. Define the sampling mask $\mathbf{1}_{(i,j)\in\Omega}$, and denote the observed data matrix as

$$\tilde{Y}_{ij} = \begin{cases} Y_{ij}, & \text{if } (i,j) \in \Omega, \\ \text{unobserved}, & \text{otherwise.} \end{cases}$$

The goal is to estimate the underlying signal matrix X and construct valid confidence intervals for each entry X_{ij} using only the partially observed, heteroskedastic data. This requires carefully disentangling the effects of noise and missingness, while leveraging the assumed low-rank structure of X.

D.2 Proposed Algorithm 1

The paper introduces a procedure to construct confidence regions for the latent row factors $U^\star_{l,\cdot}$ of the low-rank signal matrix $X = U^\star \Sigma^\star V^{\star\top}$, where the goal is to quantify the uncertainty in estimating each row of the left singular vector matrix $U^\star$. The algorithm assumes as input the output of a low-rank approximation (from Hetero PCA Algorithm), specifically the estimated matrices (U, Σ, S), the sampling rate p, and the desired coverage level $1 - \alpha$. The steps are summarized below.

Step 1: Estimate noise levels. For each row $l = 1, \ldots, d$, the noise variance $\hat{\omega}_l^2$ is estimated from the observed entries in that row:

$$\hat{\omega}_l^2 = \frac{\sum_{j=1}^n y_{l,j}^2 \, \mathbf{1}_{(l,j)\in\Omega}}{\sum_{j=1}^n \mathbf{1}_{(l,j)\in\Omega}} - S_{l,l}.$$

Here, $S_{l,l}$ is the l-th diagonal entry of the matrix S, which captures the estimated contribution of the signal matrix in row l. It is computed based on the output of the low-rank approximation in Hetero PCA Algorithm. This subtraction isolates the noise variance by removing the estimated signal component from the observed energy in each row.

Step 2: Estimate covariance of row factor. The algorithm constructs a covariance estimate $\hat{\Sigma}_{U,l}$ for the vector $U^\star_{l,\cdot}$, which incorporates both the noise estimate and the sampling variability:

$$\hat{\Sigma}_{U,l} = \left(\tfrac{1-p}{np} \|U_{l,\cdot}\Sigma\|_2^2 + \tfrac{\hat{\omega}_l^2}{np} \right) \Sigma^{-2} + \tfrac{2(1-p)}{np} U_{l,\cdot}^\top U_{l,\cdot} + \Sigma^{-2} U^\top \operatorname{diag}(\{d_{l,i}\}) \, U \, \Sigma^{-2},$$

where the diagonal elements $\{d_{l,i}\}_{i=1}^d$ are defined as

$$d_{l,i} = \tfrac{1}{np^2} \left[\hat{\omega}_l^2 + (1-p)\|U_{l,\cdot}\Sigma\|_2^2 \right] \left[\omega_i^2 + (1-p)\|U_{i,\cdot}\Sigma\|_2^2 \right] + \tfrac{2(1-p)^2}{np^2} S_{l,i}^2.$$

This step captures higher-order variance contributions due to the interaction of heteroskedastic noise and subsampling, using a plug-in approach based on the current estimates.

Step 3: Construct confidence region. Let $\tau_{1-\alpha}$ denote the $(1-\alpha)$-quantile of the chi-squared distribution with r degrees of freedom, i.e.,

$$\mathcal{B}_{1-\alpha} = \{ \, z \in \mathbb{R}^r : \|z\|_2^2 \le \tau_{1-\alpha} \, \}.$$

Then, the confidence region for $U^\star_{l,\cdot}$ is given by

$$\mathrm{CR}_{U,l}^{1-\alpha} = U_{l,\cdot} + \left(\hat{\Sigma}_{U,l} \right)^{1/2} \mathcal{B}_{1-\alpha}.$$

This defines an ellipsoidal region in $\mathbb{R}^r$ centered at the estimated row vector $U_{l,\cdot}$, with shape determined by the estimated covariance.

Interpretation. This algorithm provides high-probability confidence sets for each row factor $U^\star_{l,\cdot}$, accounting for both heteroskedasticity and missingness. The construction relies on variance decomposition that captures signal strength, noise level, and subsampling effects, and the result is a data-driven region in which the true row vector lies with probability at least $1 - \alpha$, under mild conditions.

D.3 Theoretical Guarantees for Algorithm 1

The reliability of the proposed inference method is established through two main results: Theorem 1 and Theorem 2. These theorems provide a row-wise Gaussian approximation for the subspace estimator returned by Algorithm 2, and validate the coverage accuracy of the constructed confidence regions.

Theorem 1. Assume that each column of the ground truth matrix X is independently generated from $\mathcal{N}(0, S^\star)$, and that the sampling set Ω follows the random sampling model described in Section 1.1. Let $p < 1 - \delta$ for some constant $0 < \delta < 1$, or $p = 1$, and assume $\kappa, \mu, r, \kappa_\omega \asymp 1$. Suppose that Assumption 1 holds, $d \gtrsim \log^5 n$, and the sample size and noise satisfy:

$$\frac{\omega_{\max}^2}{p\sigma_r^{\star 2}}\sqrt{\frac{d}{n}} \lesssim \frac{1}{\log^{7/2}(n+d)}, \quad \frac{\omega_{\max}}{\sigma_r^\star}\sqrt{\frac{d}{np}} \lesssim \frac{1}{\log^3(n+d)},$$

$$ndp^2 \gtrsim \log^9(n+d), \quad np \gtrsim \log^7(n+d),$$

and that the number of iterations t_0 satisfies:

$$t_0 \gtrsim \log\left[\left(\frac{\log^2(n+d)}{\sqrt{ndp}} + \frac{\omega_{\max}^2}{p\sigma_r^{\star 2}}\sqrt{\frac{d}{n}}\log(n+d) + \frac{\log(n+d)}{\sqrt{np}} + \frac{\omega_{\max}}{\sigma_r^\star}\sqrt{\frac{d\log(n+d)}{np}}\right)^{-1}\right].$$

Let $R = \mathrm{sgn}(U^\top U^\star)$ be the $r \times r$ rotation matrix aligning the estimated subspace with the true one. Then the estimate U returned by Algorithm 2 satisfies:

$$\sup_{C \in \mathcal{C}_r} \left|\mathbb{P}\left([UR - U^\star]_{l,\cdot} \in C\right) - \mathbb{P}\left(\mathcal{N}(0, \Sigma_{U,l}^\star) \in C\right)\right| = o(1), \quad \text{for all } 1 \le l \le d.$$

Interpretation and Importance: Theorem 1 shows that each row of the estimated subspace U, after alignment, admits a nearly tight Gaussian approximation with a closed-form covariance matrix $\Sigma_{U,l}^\star$. This result provides the theoretical basis for constructing confidence regions around each estimated subspace row.

Theorem 2. Suppose that the conditions of Theorem 1 hold. Then there exists a rotation matrix $R = \mathrm{sgn}(U^\top U^\star)$ such that the confidence regions $\mathrm{CR}_{U,l}^{1-\alpha}$, for $1 \le l \le d$, computed in Algorithm 3, satisfy:

$$\sup_{1 \le l \le d} \left|\mathbb{P}\left(U_{l,\cdot}^\star R^\top \in \mathrm{CR}_{U,l}^{1-\alpha}\right) - (1-\alpha)\right| = o(1).$$

Interpretation and Importance: Theorem 2 ensures that the confidence regions constructed from the plug-in estimator of $\Sigma_{U,l}^\star$ achieve asymptotically valid coverage uniformly across all rows. This confirms that the procedure not only estimates subspace directions but also provides statistically valid uncertainty quantification in high-dimensional, heteroskedastic, and incomplete data settings.

D.4 Proposed Algorithm 2

The paper introduces a second algorithm to construct entrywise confidence intervals for the elements of the signal covariance matrix $S^\star$. The procedure uses the output of HeteroPCA, which provides an estimate S of $S^\star$, and estimates the variance of each entry to construct confidence intervals.

Inputs: The algorithm takes as input the estimated low-rank factors (U, Σ, S) from Hetero PCA Algorithm, the sampling rate p, and the desired confidence level $1 - \alpha$.

Step 1: Estimate noise levels. For each row l, estimate the noise variance using:

$$\omega_l^2 := \frac{\sum_{j=1}^n y_{l,j}^2 \cdot \mathbf{1}_{(l,j)\in\Omega}}{\sum_{j=1}^n \mathbf{1}_{(l,j)\in\Omega}} - S_{l,l}.$$

Step 2: Estimate entrywise variances. For each entry (i, j), a variance estimator $v^\star_{i,j}$ is constructed using the plug-in formula derived in the paper:

- For off-diagonal terms $(i \neq j)$, the variance is estimated by:

$$\begin{aligned} v^\star_{i,j} := &\frac{2-p}{np} S^\star_{i,i} S^\star_{j,j} + \frac{4-3p}{np} (S^\star_{i,j})^2 + \frac{1}{np} (\omega_i^{\star 2} S^\star_{j,j} + \omega_j^{\star 2} S^\star_{i,i}) \\ &+ \frac{1}{np^2} \sum_{k=1}^{d} [\omega_i^{\star 2} + (1-p) S^\star_{i,i}][\omega_k^{\star 2} + (1-p) S^\star_{k,k}] \\ &+ \frac{1}{np^2} \sum_{k=1}^{d} [\omega_j^{\star 2} + (1-p) S^\star_{j,j}][\omega_k^{\star 2} + (1-p) S^\star_{k,k}] \\ &+ 2(1-p)^2 \sum_{k=1}^{d} (S^\star_{i,k})^2 (U^\star_k \cdot U^\star_j)^2 + (S^\star_{j,k})^2 (U^\star_k \cdot U^\star_i)^2. \end{aligned}$$

- For diagonal terms $(i = j)$, the variance becomes:

$$\begin{aligned} v^\star_{i,i} := &\frac{12-9p}{np} (S^\star_{i,i})^2 + \frac{4}{np} \omega_i^{\star 2} S^\star_{i,i} \\ &+ \frac{4}{np^2} \sum_{k=1}^{d} [\omega_i^{\star 2} + (1-p) S^\star_{i,i}][\omega_k^{\star 2} + (1-p) S^\star_{k,k}] \\ &+ 2(1-p)^2 \sum_{k=1}^{d} (S^\star_{i,k})^2 (U^\star_k \cdot U^\star_i)^2. \end{aligned}$$

Step 3: Construct confidence intervals. Finally, output the confidence interval:

$$\mathrm{CI}^{1-\alpha}_{i,j} = \left[S_{i,j} \pm \Phi^{-1}(1-\alpha/2) \cdot \sqrt{v_{i,j}} \right],$$

where $\Phi^{-1}(\cdot)$ is the inverse CDF of the standard Gaussian distribution.

This procedure enables inference for each entry $S^\star_{i,j}$, incorporating estimated heteroskedastic noise and subsampling uncertainty in a fully data-driven way.

D.5 Theoretical Guarantees for Algorithm 2

The theoretical validity of the confidence intervals constructed in Algorithm 2 is grounded in Theorem 3 and Theorem 4 of the paper.

Theorem 3. Suppose that $p < 1-\delta$ for some arbitrary constant $0 < \delta < 1$ or $p = 1$, and that $\kappa, \mu, r, \kappa_\omega \asymp 1$. Assume that $U^\star$ is μ-incoherent and satisfies:

$$\|U^\star_{i,\cdot}\|_2 + \|U^\star_{j,\cdot}\|_2 \gtrsim \left[\frac{\omega_{\max}}{\sigma^\star_r} \sqrt{\frac{d \log^5(n+d)}{np}} + \frac{\omega^2_{\max}}{p \sigma^{\star 2}_r} \sqrt{\frac{d \log^5(n+d)}{n}} + \sqrt{\frac{\log^7(n+d)}{ndp^2}} \right] \cdot \frac{1}{\sqrt{d}}.$$

Also assume:

$$d \gtrsim \log^5 n, \quad np \gtrsim \log^7(n+d), \quad ndp^2 \gtrsim \log^7(n+d),$$

$$\frac{\omega_{\max}}{\sigma^\star_r} \sqrt{\frac{d}{np}} \lesssim \frac{1}{\log^3(n+d)}, \quad \frac{\omega^2_{\max}}{p \sigma^{\star 2}_r} \sqrt{\frac{d}{n}} \lesssim \frac{1}{\log^{7/2}(n+d)},$$

and that the number of iterations satisfies the lower bound in equation (3.6). Then, for any entry (i, j), the matrix S computed by Algorithm 2 satisfies:

$$\sup_{t \in \mathbb{R}} \left| \mathbb{P}\left(\frac{S_{i,j} - S^\star_{i,j}}{\sqrt{v^\star_{i,j}}} \leq t \right) - \Phi(t) \right| = o(1),$$

where $\Phi(t)$ denotes the CDF of the standard Gaussian distribution.

Interpretation: This result shows that the estimation error for each entry $S_{i,j}$ is asymptotically normal, centered at zero, with variance $v^\star_{i,j}$. This is the key technical result that justifies using Gaussian quantiles in the construction of entrywise confidence intervals.

Theorem 4. *Suppose the conditions in Theorem 3 hold. Assume that* $ndp^2 \gtrsim \log^8(n+d)$. *Then the confidence interval computed in Algorithm 4 satisfies:*

$$\mathbb{P}\left(S^\star_{i,j} \in \mathrm{CI}^{1-\alpha}_{i,j}\right) = 1 - \alpha + o(1).$$

Interpretation: Theorem 4 confirms the validity of the confidence intervals produced by Algorithm 4. Despite the presence of heteroskedastic noise and missing data, the constructed intervals for each entry of $S^\star$ attain the nominal coverage level $1-\alpha$ asymptotically.

E Empirical Results

To validate the proposed inference framework, the authors conduct a comprehensive set of Monte Carlo simulations. These experiments test both the estimation accuracy and inferential validity of the algorithms introduced in the paper under a controlled heteroskedastic and partially observed setting.

Setup. The simulations fix the dimension $d = 100$ and the number of samples $n = 2000$. The true covariance matrix is generated as $S^\star = U^\star U^{\star\top}$, where $U^\star \in \mathbb{R}^{n\times r}$ is drawn uniformly from the Haar distribution on the Grassmann manifold. To model heteroskedasticity, the noise variance $\omega^{\star 2}_l$ for each row l is sampled independently from Uniform$[0.1\omega^\star, 2\omega^\star]$, and noise entries $\eta_{l,j}$ are then drawn from $\mathcal{N}(0, \omega^{\star 2}_l)$.

Subspace and Covariance Estimation Accuracy. The authors compare HeteroPCA against both SVD-based PCA and diagonal-deletion PCA. Results are evaluated using various relative error metrics, such as:

$$\frac{\|UR - U^\star\|}{\|U^\star\|}, \quad \frac{\|S - S^\star\|}{\|S^\star\|}.$$

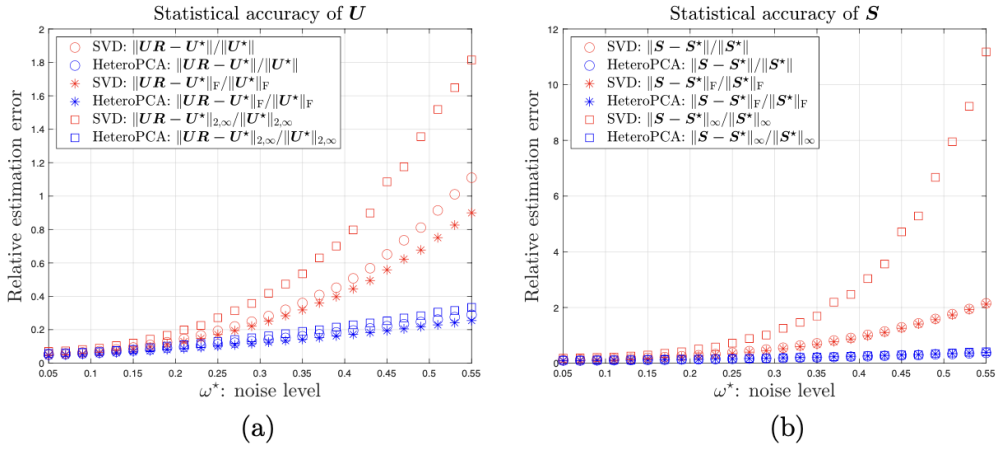


Figure 1: Relative estimation errors of U and S under varying noise levels $\omega^\star$, comparing HeteroPCA and SVD-based PCA.

Figure 1 shows that HeteroPCA uniformly outperforms the SVD-based estimator across all error metrics as noise level $\omega^\star$ increases (with $r = 3$ and $p = 0.6$). Figure 2 confirms this superiority across different missing probabilities p, highlighting the robustness of HeteroPCA to both noise heterogeneity and missingness.

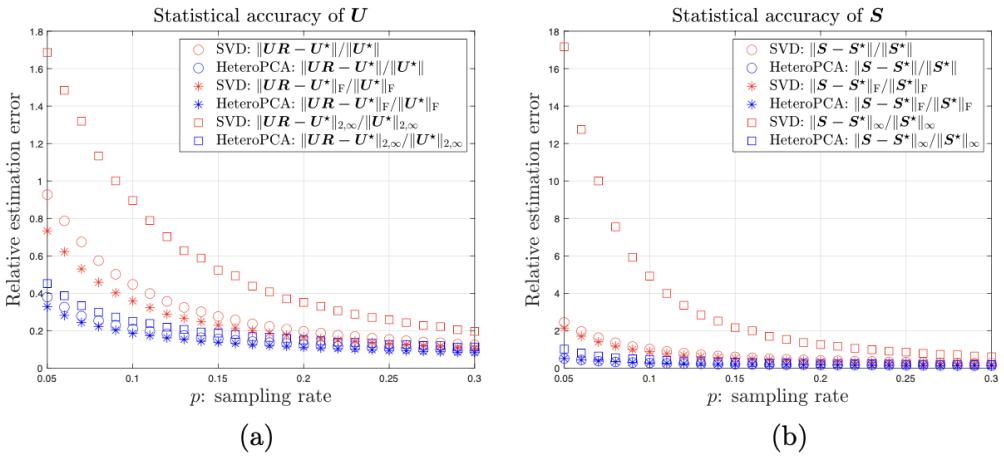


Figure 2: Relative estimation errors of U and S under varying sampling rates p, showing the performance of HeteroPCA vs. SVD-based PCA.

In comparison with diagonal-deletion PCA, HeteroPCA also achieves substantially better performance, particularly when the signal strength is not too low. These results support the efficacy of the iterative spectral correction used in HeteroPCA.