Proceedings of the 6th International Conference on Statistics: Theory and Applications (ICSTA'24)
Barcelona, Spain – August 19 - 21, 2024
Paper No. 175
DOI: 10.11159/icsta24.175

Robust Shrinkage Estimator for Perturbed Covariance Matrices

Greeshma Balabhadra
Stony Brook University
New York, USA
Greeshma.balabhadra@stonybrook.edu

Abstract - In high-dimensional statistics and finance, accurately estimating covariance matrices is crucial, especially when dealing with noisy or perturbed data. Traditional estimators often falter in these scenarios, particularly in high-dimensional, low sample size contexts. To address these challenges, we propose a new method that combines quadratic shrinkage for eigenvalues with James-Stein Estimation (JSE) shrinkage for eigenvectors. This dual approach enhances the robustness and accuracy of covariance matrix estimation by effectively mitigating the impact of noise and preserving the matrix's structural integrity. Our method significantly improves key metrics, including Tracking Error, Variance Forecast Ratio, and True Variance Ratio. These metrics consistently outperform those obtained with the standard JSE, particularly across various perturbation levels. The results underscore the potential of our approach to deliver more reliable and stable estimations. This makes it highly relevant for applications in finance and other fields where accurate covariance matrix estimation is essential. Our findings open further research into optimizing this combined shrinkage technique.

***Keywords*:** High-Dimensional Statistics, Shrinkage Estimator, Perturbed Covariance Matrices, Robust estimator

1. Introduction

Covariance matrices are integral to various applications in statistics and machine learning. Many recent studies have focused on estimating covariance matrices through their eigenvectors. Estimating the covariance matrix becomes even more challenging in high dimensions when p >> n, where p is the dimension of the covariance matrix and n is the sample size. It becomes even more challenging as statistical noises and sampling errors often perturb these matrices. We propose a robust shrinkage estimator for the eigenvectors and eigenvalues of the sample covariance matrices in this high-dimensional-low sample setting.

Various studies have attacked the above problem in different directions. [3] studies a special set of covariance matrices called spiked covariance matrices where the sample eigenvalues follow the Marchenko-Pastur or "quarter-circle" law, shifted upwards, and the top eigenvectors are inconsistent. [3] also discusses various loss functions and derives unique admissible shrinkers of eigenvalues for each loss function. [2] expands these convergence limits of the individual shrinkers for all the loss functions to a generalized spiked covariance model where the perturbance is a positive semi-definite matrix. [2] also discusses the asymptotic eigenvector distributions for sample covariance matrices under general assumptions. [6] and [7] discuss the dispersion bias in the leading sample eigenvector in the factor-based covariance model. They propose James-Stein for eigenvectors, a data-driven eigenvector shrinkage model, and show that the proposed eigenvector performs well in variance-minimizing problems compared to the eigenvalue shrinkage model. [5] studies the perturbation bounds of eigenvectors in low-rank incoherent matrices. More references on eigenvalue shrinkage are [1] [4] [8] [10]. All these models either discuss the shrinkage of eigenvalue or eigenvector in perturbed covariance matrices. We propose a method that finds the optimal shrinkage for both the eigenvector and eigenvalue for the underlying perturbed sample covariance matrix under general assumptions.

The shrinkage estimators for both eigenvectors and eigenvalues have substantially impacted estimation errors in minimum variance portfolios separately. We use this to our advantage and get an optimal shrinkage estimator regardless of the loss function for both eigenvector and eigenvalue. We propose a data-driven approach and evaluate the performance using the metrics used in [7] and [5], like optimization bias and variance forecast ratio. We finally evaluate the forecast performance of the proposed estimator using a minimum variance portfolio. We compare our performance with the state-of-the-art shrinkage estimators like [6], [12]. We can use this robust estimator in approximate factor model analysis, which has wide applications exploring correlation structure in finance, economics, genomics, etc, with various types and sources of perturbations in the underlying true covariance matrix

2. Main Model

Consider the following data-generating process

$$y_t = Bf_t + \epsilon_t$$

where $y_t \in R^p$, and $f_t \in R^k$ is a vector of latent factors, B is matrix of corresponding factor loading coefficients, and ϵ_t is idiosyncratic part that factors cannot explain. The covariance of the model is

$$\Sigma = BB^T + \Sigma_\epsilon$$

- Σ is the covariance matrix of the observed data.
- $A = BB^T$ is the low-rank component representing the signal, with B as the factor loading matrix.
- $\Sigma_\epsilon = var(\epsilon)$ is the covariance matrix of the idiosyncratic errors, assumed to be sparse. It can be decomposed as $\Sigma_\epsilon = S + N$ where S is a sparse matrix representing sparse contamination or systematic errors and N is a random matrix representing purely idiosyncratic noise or estimation error.

The singular value decomposition of A can be written as

$$A = \sum_{i=1}^{r} \sigma_i u_i v_i^T$$

where r is the rank of A and the singular values are $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$. From [5] we have the eigenvalue gap γ_0 is defined as:

$$\gamma_0 = \min_{i=1,\ldots,r} (\sigma_i - \sigma_{i+1}) \tag{1}$$

where σ_i are the singular values of the matrix A, with $\sigma_{r+1} = 0$. The perturbation $E = S + N$ is measured using the following rescaled norms:

$$\tau_0 = \max\left(\sqrt{d_2/d_1}\left||E|\right|_1, \sqrt{d_1/d_2}\left||E|\right|_\infty\right) \tag{2}$$

where $\left||E|\right|_1 = \max_j \sum_{i=1}^{d_1} |E_{ij}|$, and $\left||E|\right|_\infty = \max_j \sum_{i=1}^{d_2} |E_{ij}|$

The perturbation of the eigenvectors (or singular vectors) is bounded asymptotically as given in [5]:

$$\max_{i=1,\ldots,r} \left||\widetilde{u_i} - u_i|\right|_\infty \leq C(r, \mu_0)\,(\tau_0/\,\gamma_0)\sqrt{d_1}$$

where:

- $C(r, \mu_0)$ is a constant dependent on the rank r and the coherence μ_0 of the matrix A.
- μ_0 is the maximum coherences of the left and right singular vectors of A.

Shrinkage for Leading Eigenvector and Eigenvalues

Let S be the sample covariance matrix of a p-dimensional random vector based on n observations. The spectral decomposition of S is given by

$$S = \sum_{i=1}^{p} \lambda_i h_i h_i^T \tag{3}$$

where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$ are the eigenvalues, and h_i are the corresponding orthonormal eigenvectors. Our interest lies in the leading eigenvalue λ_1 and the leading eigenvector h_1 when this sample covariance matrix is perturbed.

James-Stein Shrinkage for the Leading Eigenvector

The James-Stein estimator (JSE) introduced in [6], [7], [11], is a shrinkage method designed to reduce the excess dispersion of the leading sample eigenvector in the high-dimension low-sample size (HL) regime. The JSE shrinks the entries of the leading eigenvector h1 toward their average:

$$h_{JSE} = m(h)1 + c_{JSE}(h1 - m(h)1), \tag{4}$$

where $m(h)$ is the average of the entries of h_1, 1 is the p-dimensional vector of ones, and the shrinkage constant c_{JSE} is defined as:

$$c_{JSE} = 1 - \frac{v^2}{s^2(h_1)}$$

where $s^2(h_1)$ is the sample variance of the entries of h_1,:

$$s^2(h_1) = \frac{1}{p}\sum_{i=1}^{p}\left(h_{1i} - m(h)\right)^2$$

and v^2 is the average of the non-leading eigenvalues:

$$v^2 = \frac{tr(S) - \lambda_1}{p - 1}$$

Ledoit-Wolf Shrinkage for the Leading Eigenvalue.
The Ledoit-Wolf (LW) shrinkage estimator introduced in [9] shrinks the sample eigenvalues to improve the covariance matrix estimation. The LW shrinkage for the leading eigenvalue λ1 is given by

$$\lambda_{1,LW} = (1 - \alpha)\lambda_1 + \alpha\lambda_{target}, \quad (5)$$

where λ_{target} is a target eigenvalue, often chosen as the average of the sample eigenvalues $\lambda_{target} = \sum_{i=1}^{p}\lambda_i$, and α is the shrinkage intensity, typically estimated by minimizing the mean squared error between the sample covariance matrix and the true covariance matrix. [12] extended this shrinkage estimator to the quadratic shrinkage estimator for the inverse eigenvalues, which is expressed as:

$$\delta_i^{-1} = \left(1 - \frac{p}{n}\right)\lambda_i^{-1} + 2\,\frac{p}{n}\left(1 - \frac{p}{n}\right)\lambda_i^{-1}\theta\left(\lambda_i^{-1}\right) + \left(\frac{p}{n}\right)^2 \lambda_i^{-1}A\left(\lambda_i^{-1}\right) \quad (6)$$

Here, λ_i are the eigenvalues of the sample covariance matrix S, and $\theta\left(\lambda_i^{-1}\right)$ is the smoothed Stein shrinker defined as:

$$\theta\left(\lambda_i^{-1}\right) = \frac{1}{p}\sum_{j=1}^{p}\frac{\lambda_j^{-1}}{\left(\lambda_j^{-1} - \lambda_i^{-1}\right)^2 + h^2\lambda_j^{-2}}$$

The squared amplitude $A^2\left(\lambda_i^{-1}\right)$ is given by:

$$A^2\left(\lambda_i^{-1}\right) = \left[\frac{1}{p}\sum_{j=1}^{p}\frac{\lambda_j^{-1}}{\left(\lambda_j^{-1} - \lambda_i^{-1}\right)^2 + h^2\lambda_j^{-2}}\right]^2 + \left[\frac{1}{p}\sum_{j=1}^{p}\frac{h\lambda_j^{-1}}{\left(\lambda_j^{-1} - \lambda_i^{-1}\right)^2 + h^2\lambda_j^{-2}}\right]^2$$

where h is the smoothing parameter, controlling the degree of smoothing applied to the Stein shrinker.

The quadratic shrinkage method outperforms the linear shrinkage approach, particularly when the covariance matrix is perturbed due to noise or other external factors. The quadratic nature of the shrinkage allows the estimator to adapt better to the concentration ratio $\frac{p}{n}$, offering more accurate estimates of the true eigenvalues, thereby improving key metrics like the true variance ratio, variance forecast ratio, and reducing tracking errors. This improvement is especially significant when there is a high level of perturbation in the covariance matrix, as the quadratic shrinkage helps mitigate the impact of noise, leading to more robust portfolio optimization and risk management decisions.

Let $\Sigma_{perturbed}$ be the perturbed covariance matrix, which can be decomposed as:

$$\Sigma_{perturbed} = U\Lambda U^T + E \quad (7)$$

where U is the matrix of eigenvectors, $\Lambda = diag\left(\lambda_1, \lambda_2, \dots, \lambda_p\right)$ is the diagonal matrix of eigenvalues, and E represents the perturbation or noise matrix. The eigenvalues λ_i of the perturbed covariance matrix $\Sigma_{perturbed}$ are shrunk using the quadratic shrinkage formula as given in (6). The eigenvectors u_i of the perturbed covariance matrix $\Sigma_{perturbed}$ are shrunk using the James-Stein shrinkage technique. The shrunken eigenvector u_i^{shr} is given by (5). The final covariance matrix estimator Σ_{final} is constructed by combining the shrunken eigenvalues and eigenvectors:

$$\Sigma_{final} = \sum_{i=1}^{p}\delta_i\, u_i^{shr}\left(u_i^{shr}\right)^T \quad (8)$$

where δ_i are the shrunken eigenvalues obtained from the quadratic shrinkage method, and u_i^{shr} are the shrunken eigenvectors obtained from the James-Stein shrinkage.

3. Performance metrics

We consider three performance metrics like [6] to compare the performance of the proposed estimator with the JSE estimator when the covariance matrices are perturbed.

Variance Forecast Ratio (VFR). The Variance Forecast Ratio (VFR) is defined as the ratio of the estimated variance to the true variance of an optimized portfolio. It measures the accuracy of the variance forecast provided by the estimated covariance matrix.

$$VFR(w^*) = \frac{{w^*}^T \breve{\Sigma} w^*}{{w^*}^T \Sigma w^*} \tag{9}$$

Here, w^* is the optimized portfolio weight vector, $\breve{\Sigma}$ is the estimated covariance matrix, and Σ is the true covariance matrix. The VFR is important because it assesses how well the estimated covariance matrix predicts the variance of the optimized portfolio. A VFR close to 1 indicates that the estimation is accurate, while deviations from 1 suggest under- or overestimation of portfolio risk.

True Variance Ratio (TVR). The True Variance Ratio (TVR) is the ratio of the true variance of the true optimal portfolio to the true variance of the portfolio derived from the estimated covariance matrix.

$$TVR(w^*) = \frac{w^T \Sigma w}{{w^*}^T \Sigma w^*} \tag{10}$$

Here, w is the true optimal portfolio weight vector. The TVR measures the efficiency of the optimized portfolio relative to the true optimal portfolio. A TVR greater than 1 indicates that the estimated portfolio is suboptimal, with a higher variance than the true optimal portfolio.

Tracking Error (TE). Tracking Error (TE) quantifies the deviation of the returns of the estimated optimal portfolio from the returns of the true optimal portfolio.

$$TE^2(w^*) = (w^* - w)^T \Sigma (w^* - w) \tag{11}$$

TE is widely used in portfolio management to assess how closely the estimated portfolio tracks the true optimal portfolio. Lower TE indicates better performance in matching the true optimal portfolio's returns.

These evaluation metrics are essential for assessing the performance of covariance matrix estimators in the HL regime. They provide insights into the accuracy of variance forecasts, the efficiency of the estimated portfolio, and the overall tracking performance. In settings where perturbations or noise affect the covariance matrix, these metrics become even more critical for understanding the effectiveness of shrinkage techniques, such as the James-Stein estimator (JSE) for eigenvectors and quadratic shrinkage for eigenvalues.

4. Experimental results

We set simulations to compare the proposed estimator with the James-Stein Estimator (JSE) and different shrinkage methods across various perturbation levels. The simulation parameters are:

- The maximum number of assets $p = 500$
- The number of experiments to be run for each perturbation level $m = 10000$
- The number of time periods considered in the simulation $n = 252$
- Perturbation levels applied to the covariance matrix are given by $l = (0.0125, 0.0525, 0.1252, 0.2252)$. These values represent increasing levels of noise or perturbation in the covariance matrix.

We simulate asset returns for each experiment, and the sample covariance matrix is computed. We compare the proposed covariance estimator with the JSE estimator by [7] and evaluated the metrics - tracking error, variance forecast ratio, and true variance ratio in table Table 1 using the (9), (10) and (11). All analytics show improvement across the different perturbation levels, with a particularly significant enhancement in the variance forecast ratio at all levels. However, no clear pattern emerges in the improvements as perturbation levels increase, suggesting a potential area for future research.

Table 1: Improvement of the proposed estimator over the James-Stein Estimator (JSE) in terms of Tracking Error, Variance Forecast Ratio, and True Variance Ratio, presented as percentage improvements.

Perturbation level	Tracking error	Variance forecast ratio	True variance ratio
0.0125	4.2743	28.0973	4.2355
0.0525	5.1159	36.9158	5.0705
0.1252	4.1736	37.2269	4.0236
0.2252	1.5526	14.2780	1.4939

5. Conclusions

Applying quadratic shrinkage to eigenvalues, combined with James-Stein shrinkage of eigenvectors, significantly enhances the accuracy and robustness of covariance matrix estimation, especially in high-dimensional, low-sample size scenarios. Our results show that this combined approach outperforms the James-Stein estimator alone, particularly in metrics like variance forecast ratio, true variance ratio, and tracking error. The quadratic shrinkage effectively mitigates the impact of noise and perturbations in the covariance matrix, leading to more reliable and stable estimations. This synergy between eigenvalue and eigenvector shrinkage improves the alignment with the true covariance matrix. It opens new avenues for future research in handling various perturbation models and developing adaptive shrinkage techniques. This method represents a significant advancement in the field, with substantial implications for high-dimensional statistical and financial applications.

References

[1] M. J. Daniels and R. E. Kass, Shrinkage estimators for covariance matrices, Biometrics, 57 (2001), pp. 1173–1184.

[2] X. Ding, Y. Li, and F. Yang, Eigenvector distributions and optimal shrinkage estimators for large covariance and precision matrices, arXiv preprint arXiv:2404.14751, (2024).

[3] D. L. Donoho, M. Gavish, and I. M. Johnstone, Optimal shrinkage of eigenvalues in the spiked covariance model, Annals of statistics, 46 (2018), p. 1742.

[4] D. L. Donoho, M. Gavish, and I. M. Johnstone, Optimal shrinkage of eigenvalues in the spiked covariance model, Annals of statistics, 46 (2018), p. 1742.

[5] J. Fan, W. Wang, and Y. Zhong, An $\ell\infty$ eigenvector perturbation bound and its application to robust covariance estimation, Journal of machine learning research: JMLR, 18 (2018).

[6] L. R. Goldberg and A. N. Kercheval, James–stein for the leading eigenvector, Proceedings of the National Academy of Sciences, 120 (2023), p. e2207046120.

[7] L. R. Goldberg, A. Papanicolaou, and A. Shkolnik, The dispersion bias, SIAM Journal on Financial Mathematics, 13 (2022), pp. 521–550.

[8] C. Lam, P. Feng, and C. Hu, Nonlinear shrinkage estimation of large integrated covariance matrices, Biometrika, 104 (2017), pp. 481–488.

[9] O. Ledoit and M. Wolf, A well-conditioned estimator for large-dimensional covariance matrices, Journal of multivariate analysis, 88 (2004), pp. 365–411.

[10] O. Ledoit and M. Wolf, Nonlinear shrinkage estimation of large-dimensional covariance matrices, The Annals of Statistics, 40 (2012), pp. 1024 – 1060.

[11] O. Ledoit and M. Wolf, Shrinkage estimation of large covariance matrices: Keep it simple, statistician?, Journal of Multivariate Analysis, 186 (2021), p. 104796.

[12] O. Ledoit and M. Wolf, Quadratic shrinkage for large covariance matrices, Bernoulli, 28 (2022), pp. 1519–1547.