Manifold Principle Component Analysis for Large-Dimensional Matrix Elliptical Factor Model

ZeYu Li § Yong He † Xinbing Kong ‡ Xinsheng Zhang *

March 29, 2022

Abstract

Matrix factor model has been growing popular in scientific fields such as econometrics, which serves as a two-way dimension reduction tool for matrix sequences. In this article, we for the first time propose the matrix elliptical factor model, which can better depict the possible heavy-tailed property of matrix-valued data especially in finance. Manifold Principle Component Analysis (MPCA) is for the first time introduced to estimate the row/column loading spaces. MPCA first performs Singular Value Decomposition (SVD) for each “local” matrix observation and then averages the local estimated spaces across all observations, while the existing ones such as 2-dimensional PCA first integrates data across observations and then does eigenvalue decomposition of the sample covariance matrices. We propose two versions of MPCA algorithms to estimate the factor loading matrices robustly, without any moment constraints on the factors and the idiosyncratic errors. Theoretical convergence rates of the corresponding estimators of the factor loading matrices, factor score matrices and common components matrices are derived under mild conditions. We also propose robust estimators of the row/column factor numbers based on the eigenvalue-ratio idea, which are proven to be consistent. Numerical studies and real example on financial returns data check the flexibility of our model and the validity of our MPCA methods.

Keywords: Factor Model; Grassmann manifold; Matrix elliptical distribution; Principle component analysis.

1 Introduction

Factor models have been a classical dimension reduction tool in statistics, which is popular for its ability to summarize information in large data sets. More importantly, factor models characterize many economic problems, e.g., the Arbitrage Pricing Theory of Ross (1976). In the last two decades large-dimensional approximate factor model is growing popular as we embrace the big data era where more and more variables are recorded and stored, see the seminal work by Bai and Ng (2002) and Stock and Watson (2002), and some representative work by Bai (2003), Onatski (2009), Ahn and Horenstein
(2013), Fan et al. (2013), Bai and Li (2012), Bai and Li (2016) and Trapani (2018). The aforementioned papers all require the fourth moments (or even higher moments) of factors and idiosyncratic errors exist, which may be constrictive in research areas such as finance. He et al. (2022) for the first time propose a Robust Two Step (RTS) procedure to do factor analysis without any moment constraints, under the framework of elliptical distributions, see also the endeavors by Yu et al. (2019), He et al. (2020) and Chen et al. (2021a).

The modern data collected are usually well-structured in a matrix form, such as time list of tables recording several macroeconomic variables across a number of countries; a series of customers’ ratings on a large number of items in an online platform, see Chen and Fan (2021) for further examples of well-structured matrix observations. In the last few years, matrix factor model has drawn growing attention as an important two-way dimension reduction tool for matrix sequences. Wang et al. (2019) for the first time proposed the following matrix factor model:

\[
X_t \in \mathbb{R}^{p \times q} = R_{p \times p_0} \times F_t \in \mathbb{R}^{p_0 \times q_0} \times C_{q_0 \times q}^\top + E_t \in \mathbb{R}^{p \times q},
\]

where \( \{X_t, 1 \leq t \leq T\} \) are matrix observations of dimension \( p \times q \), \( R \) is the row factor loading matrix exploiting the variations of \( X_t \) across the rows, \( C \) is the \( q \times q_0 \) column factor loading matrix reflecting the differences across the columns of \( X_t \), \( F_t \) is the common factor matrix for all cells in \( X_t \) and \( E_t \) is the idiosyncratic components. A naive way to do factor analysis for matrix observations is to first vectorize the data \( X_t \), and then to adopt the classical well-developed vector factor models techniques. However, when data genuinely have a matrix factor structure as in (1.1), this naive approach would lead to sub-optimal inference (Chen and Fan, 2021; He et al., 2021c). Two different types of matrix factor model assumptions are adopted in the existing literature. One type of models assumes that the factors accommodate all dynamics, making the idiosyncratic noise “white” with no autocorrelation but allowing substantial contemporary cross-correlation among the error process, and the estimation of the loading space is done by an eigen-analysis of the nonzero lag autocovariance matrices, see for example Wang et al. (2019). The other type of models assumes that a common factor must have impact on almost all (defined asymptotically) of the matrix time series, but allows the idiosyncratic noise to have weak cross-correlations and weak autocorrelations, and principle component analysis (PCA) of the sample covariance matrix is typically used to estimate the spaces spanned by the row/column loading matrices, see for example Chen and Fan (2021); Yu et al. (2021); He et al. (2021a). As far as we know, all the existing work on matrix factor model assumes that the fourth moments (or even higher moments) of factors and idiosyncratic errors exist, which could be restrictive in real applications such as in finance. Figure 1 depicts the boxplots of the row factor loading estimation errors based on 100 replications, from which we can see that the \((2D)^2\)-PCA by Zhang and Zhou (2005) and the Projection Estimation (PE) method by Yu et al. (2021) results in bigger biases and higher dispersions as the distribution tails become heavier.

In the current work, we for the first time propose a freshly new and flexible model, named as the Matrix Elliptical Factor Model (MEFM), which assumes that the factor matrix \( F_t \) and the idiosyncratic errors matrix \( E_t \) follow an joint Elliptical Matrix Distribution (EMD), which covers a large class of heavy-tailed matrix distributions such as matrix \( t \)-distribution. To estimate the row (column) loading
space \( \text{Span}(R) \) (\( \text{Span}(C) \)) of MEFM robustly, we propose a Manifold Principle Component Analysis (MPCA) method for the first time. In essence, for each data matrix \( X_t \), assume that \( p_0 = q_0 = r_0 \) for better illustration, the MPCA first finds the best \( r_0 \)-dimensional row (column) loading space estimator \( \text{Span}(\hat{R}_t) \) (\( \text{Span}(\hat{C}_t) \)), which can be viewed as an element in the Grassmann manifold \( G(p_0, p) \) (\( G(q_0, q) \)), where the Grassmann manifold \( G(p_0, p) \) is the set of \( p_0 \)-dimensional linear subspaces of the \( \mathbb{R}^p \) (Ham and Lee, 2008). Then MPCA looks for the “centers” of all row/column loading space estimators within their Grassmann manifolds respectively. According to the way of finding the best linear row/column space for each matrix observation, the MPCA then has two versions, MPCA\(_{op}\) and MPCA\(_{F}\), corresponding to the optimization problem (2.2) under matrix operator norm and matrix Frobenius norm respectively. For the MPCA\(_F\), the projection technique in Yu et al. (2021) happens to be taken into account, which increases the signal-to-noise ratio. Now, let us come back to Figure 1, in which we also presented the results using the MPCA\(_{op}\) and MPCA\(_{F}\) methods. It can be seen that MPCA\(_F\) always performs well under various distributions, and is even not sensitive to the elliptical assumption (noting that \( \alpha \)-stable distribution is not elliptical). The MPCA\(_{op}\) method also exhibits an extent of robustness, while it is inferior to the MPCA\(_F\) in all cases. This indicates that for matrix factor model, the projection technique is always preferred as it can increase the signal-to-noise ratio.

![Figure 1: Boxplot of the distance between the estimated row loading space \( \hat{R} \) and the true row loading space \( R \) by MPCA\(_{op}\), MPCA\(_{F}\), (2D)\(^2\)PCA and PE methods under different distributions (normal, \( t_3 \), \( \alpha \)-stable with \( \alpha = 1.8 \) and \( t_1 \)), \( p = q = 100, T = 300 \). The noises are scaled to get comparable performance under various distributions.](image)

To do matrix factor analysis, the first step is to determine the pair of factor numbers. As for the Elliptical Matrix Factor Model (MFM), both the row and column factor numbers should be predetermined. Wang et al. (2019) proposed to estimate the pair of factor numbers by the ratios of consecutive eigenvalues of auto-covariance matrices; Chen and Fan (2021) proposed an \( \alpha \)-PCA based eigenvalue-ratio method and Yu et al. (2021) further proposed a projection-based iterative eigenvalue-ratio method, all borrowing the eigenvalue ratio idea from Ahn and Horenstein (2013). He et al. (2021a) is the only work that determines the pair of factor numbers from the perspective of sequential hypothesis testing. In this article we also propose similar eigenvalue ratio methods based on the MPCA approach. The proposed estimators of the pair of factor numbers are proven to be consistent under mild conditions and performs much better than the existing ones when the matrix-valued data are heavy-tailed shown in the simulation study.
The contributions of the current work lie in the following aspects. Firstly, we for the first time propose a flexible matrix elliptical factor model for matrix observations which is adaptive to their tail properties. Secondly, to estimate the loading spaces of MEFM robustly, we for the first time introduce a freshly new principle component analysis method, named as the Manifold Principle Component Analysis (MPCA), which is computationally efficient and easy to implement. The MPCA is completely different from the traditional PCAs (e.g., the $\alpha$-PCA by Chen and Fan (2021) and $(2D)^2$-PCA by Zhang and Zhou (2005)) method in the sense that MPCA first performs Singular Value Decomposition (SVD) for each “local” matrix observation and then integrates/averages the local spaces, while the traditional PCAs first integrates the matrix observations and then finds the principle eigenvectors of the sample covariance matrices. Clearly, the MPCA would show great computational advantage especially in online updating problems. Thirdly, the theoretical guarantee of MPCA relies heavily on the properties of the expected projection matrices, which has not aroused much attention in existing literatures. We show that the expected projection matrices contain adequate subspace information, which is of independent interest. At last, the theoretical analysis shows that the proposed MPCA estimators are consistent without any moment constraints on the underlying distributions of the factors and the idiosyncratic errors, which generalize the methods’ applicability to heavy-tailed datasets such as financial returns.

The remainder of the article is organized as follows. In Section 2, we first introduce the proposed matrix elliptical factor model. Then we introduce the Manifold Principle Component Analysis (MPCA) method, and take the special degenerated case $q = 1$ to illustrate the intuition of its robustness. At last, we introduce the best subspace approximation for each matrix observation $X_t$ under both the matrix operator norm and the matrix Frobenius norm, by which we further propose the two versions of MPCA algorithms. In Section 3, we investigate the theoretical properties of the estimators by MPCA, including the factor loadings, factor scores and common components matrices. In Section 4, we further discuss the selection of the pair of factor numbers and propose two procedures based on the eigenvalue ratio idea. In section 5, we conduct thorough numerical studies to illustrate the advantages/robustness of the MPCA method and the corresponding eigenvalue-ratio factor number estimation methods over the state-of-the-art methods. In Section 6, we analyze a financial returns dataset to illustrate the practical value of the proposed methods. We conclude the article and discuss the limitation of the current work and possible future research directions in Section 7. The proofs of the main theorems and additional details are collected in the supplementary materials.

To end this section, we introduce some notations throughout the study. For matrix $A$, $||A||_{op}$ and $||A||_F$ represent the operator norm and Frobenius norm, $\text{tr}(A)$ denotes the trace of $A$, and $\text{Vec}(A)$ means vectorization, $\sigma_i(A)$ denotes the $i$-th largest singular value of $A$. Moreover, if $A$ is symmetric, $\lambda_i(A)$ denotes the $i$-th largest eigenvalue of $A$, and let $d_i(A) = (\lambda_i - \lambda_{i+1})(A)$ be the $i$-th eigengap. For vector $a$, denote $\|a\|_2$ as the $l^2$-norm as $\|a\|_2$. The notation $\langle x \rangle$ means rounding $x$ to the nearest integer. The constants $c$ and $C$ may not be identical in different lines.
2 Methodology

2.1 Matrix Elliptical Factor Model

For $p \times q$ matrix-variate sequences $\{X_t, t = 1, \ldots, T\}$, the centered matrix factor model is introduced by Wang et al. (2019) as follows:

$$X_t = RF_tC^\top + E_t, \quad t = 1, \ldots, T,$$

where $R$ is the $p \times p_0$ row factor loading matrix, $C$ is the $q \times q_0$ column factor loading matrix, $F_t$ is the common factor matrix and $E_t$ is the idiosyncratic component with $\mathbb{E}(F_t) = 0$ and $\mathbb{E}(E_t) = 0$. This model is suited for well-structured tables of macroeconomic indicators, financial characteristics, and frames of pictures etc. In this paper, we are interested in recovering the loading spaces $\text{Span}(R)$ and $\text{Span}(C)$. Without loss of generality, we assume $R^\top R/p = I_{p_0}$ and $C^\top C/q = I_{q_0}$. The projection matrices onto $\text{Span}(R)$ and $\text{Span}(C)$ are then naturally $P_R = RR^\top / p$ and $P_C = CC^\top / q$.

Prior to the introduction of Matrix Elliptical Factor Model (MEFM), we first take a look at matrix elliptical distributions. A random matrix $X$ of size $p \times q$ is matrix elliptical distributed if its characteristic function has the form $\varphi_X(T) = \exp \left\{ \text{tr} \left( iT^\top M \right) \right\} \psi \left( \text{tr} \left( T^\top \Sigma T \Omega \right) \right)$ with $T: p \times q$, $M: p \times q$, $\Sigma: p \times p$, $\Omega: q \times q$, $\Sigma \succcurlyeq O$, $\Omega \succcurlyeq O$ and $\psi: [0, \infty) \rightarrow \mathbb{R}$. This distribution is denoted by $E_{p,q}(M, \Sigma \otimes \Omega, \psi)$, see Gupta and Nagar (2018) for details. An important observation given in Gupta and Varga (1994) shows that for $\text{rank}(\Sigma) = m$, $\text{rank}(\Omega) = n$, the random matrix $X \sim E_{p,q}(M, \Sigma \otimes \Omega, \psi)$ if and only if:

$$X \overset{d}{=} rAUB^\top + M,$$

where $U: m \times n$ and $\text{Vec}(U)$ is uniformly distributed on the unit sphere in $\mathbb{R}^{mn}$, $r$ is a nonnegative random variable independent of $U$, $\Sigma = AA^\top$ and $\Omega = BB^\top$ are rank factorizations of $\Sigma$ and $\Omega$. The matrix Gaussian distributions and matrix t-distributions belong to the class of matrix elliptical distributions. In the article, in the definition of MEFM, we assume the factor $F_t$ and noise $E_t$ are from joint matrix elliptical distribution as in He et al. (2022), which is:

$$\begin{pmatrix} \text{Vec}(F_t) \\ \text{Vec}(E_t) \end{pmatrix} = r_t \begin{pmatrix} \Sigma_1^{1/2} \otimes \Sigma_2^{1/2} \\ 0 \end{pmatrix} \begin{pmatrix} 0 \\ \Omega_2^{1/2} \otimes \Omega_1^{1/2} \end{pmatrix} \frac{Z_t}{\|Z_t\|_2}, \quad (2.1)$$

where $Z_t$ is a $(pq + p_0q_0)$-dimensional isotropic Gaussian vector, while $r_t$ is a positive random variable independent of $Z_t$. It is sometimes more convenient to separate the joint model into:

$$F_t = \frac{r_t}{\|Z_t\|_2} \Sigma_1^{1/2} Z_t^E \Sigma_2^{1/2}, \quad E_t = \frac{r_t}{\|Z_t\|_2} \Omega_2^{1/2} Z_t^E \Omega_1^{1/2},$$

with positive-definite transformation matrices $\Sigma_1$ of size $p_0 \times p_0$, $\Sigma_2$ of size $q_0 \times q_0$, $\Omega_1$ of size $p \times p$ and $\Omega_2$ of size $q \times q$. $Z_t^E$ is a $p_0 \times q_0$ random matrix by taking the leading $p_0q_0$ elements of $Z_t$, while $Z_t^E$ of size $p \times q$ consists of all the elements left. It is not hard to verify that $F_t$ and $E_t$ are matrix elliptical distributed, since $\|Z_t\|_2^2 = \|Z_t^E\|_F^2 + \|Z_t^E\|_F^2$ is independent of both $Z_t^E / \|Z_t^E\|_F$ and $Z_t^E / \|Z_t^E\|_F$.

Remark 2.1. Assuming the joint matrix elliptical distribution of $F_t$ and $E_t$ is to ensure distribution-free
signal-to-noise conditions. For example, if \( E_t \) has i.i.d. standard Gaussian elements, then \( \| E_t \|_{op} = O_p((p \vee q)^{1/2}) \). On the other hand, if \( E_t \) has i.i.d. \( t(1) \) elements, then \( \| E_t \|_{op} \geq \| E_t \|_{\infty} \approx pq \).

Assuming joint matrix elliptical distribution ensures simplicity, otherwise, the signal-to-noise conditions are distribution-dependent, and higher signal-to-noise ratio is naturally required for heavier-tailed noise case. See the same joint (vector) elliptical distribution assumption in Fan et al. (2018); He et al. (2022).

### 2.2 Manifold Principle Component Analysis

In this section, we introduce our Manifold Principle Component Analysis (MPCA) method for MEFM estimation. As a simple heuristic argument to see the robustness of MPCA, first consider \( q = 1 \) and then the matrix factor model degenerates to the vector case. The classical vector factor model would be written as:

\[
y_t = Af_t + \epsilon_t, \quad t = 1, \ldots, T,
\]

where \( y_t \) is the \( p \times 1 \) observed vector, \( A \) is the \( p \times p_0 \) loading matrix, \( f_t \) is the \( p_0 \times 1 \) latent factor vector and \( \epsilon_t \) is the \( p \times 1 \) noise vector. The classical PCA seeks the leading \( p_0 \) eigenvectors of the sample covariance matrix \( \hat{\Sigma} = \sum_t y_t y_t^\top / T \), which is easily influenced by outliers, as those \( y_t \) with larger norm naturally have larger influence on \( \hat{\Sigma} \).

It would be more robust to treat all \( y_t \) equally, in a sense that each \( y_t \) provides the same amount of subspace information. The Manifold PCA (MPCA) first finds the best subspace estimation \( \hat{A}_t \) for each \( y_t \), which is the first eigenvector of the rank one matrix \( y_t y_t^\top \). Then it seeks the “center” of all \( \hat{A}_t \), which would be the leading eigenvectors of the average projection matrix \( \hat{\Sigma} = \sum_t \hat{A}_t \hat{A}_t^\top / T \). It is exactly a distance-weighted sample covariance, namely:

\[
\tilde{\Sigma} = \sum_t \hat{A}_t \hat{A}_t^\top / T = \sum_t \frac{y_t y_t^\top}{y_t y_t^\top / T}.
\]

Such degeneration towards vector factor model provides some basic insights for the robustness of our MPCA methods against heavy-tailed noises. If the data set \( \{y_t, 1 \leq t \leq T\} \) is augmented to \( \{y_t - y_j, 1 \leq i < j \leq T\} \), our method is then equivalent to calculating the leading eigenvectors of the multivariate Kendall’s \( \tau \) matrix, which is also valid without moment conditions, see He et al. (2022) for details.

We then introduce the MPCA methods for matrix variate data. Assume first \( p_0 = q_0 = r_0 \), for each data matrix \( X_t \), we give the best linear row/column space estimator for \( X_t \), denoted by orthogonal matrices \( \hat{R}_t \) and \( \hat{C}_t \) respectively, which are the representatives of their own equivalent classes on the Grassmann manifolds \( \mathcal{G}(p_0, p) \) and \( \mathcal{G}(q_0, q) \). Then MPCA finds the “centers” of all \( \text{Span}(\hat{R}_t) \) and \( \text{Span}(\hat{C}_t) \) within their Grassmann manifolds respectively.

In the following section, we will show that for each \( X_t \), the best linear subspace estimator \( \hat{R}_t \) and \( \hat{C}_t \) are the leading \( r_0 \) eigenvectors of \( X_t X_t^\top \) and \( X_t^\top X_t \) with respect to operator norm \( \| \cdot \|_{op} \) loss in (2.2) below. Similarly, \( \hat{R}_t \) and \( \hat{C}_t \) are leading \( r_0 \) eigenvectors of \( X_t P_C X_t^\top \) and \( X_t^\top P_R X_t \) under Frobenius norm \( \| \cdot \|_F \) loss if \( C \) and \( R \) are given respectively.

Then, for given linear space estimators \( \hat{R}_t \) and \( \hat{C}_t \) for each \( X_t \), it is natural to find their “centers” on the Grassmann manifolds \( \mathcal{G}(p_0, p) \) and \( \mathcal{G}(q_0, q) \) as the final estimators. However, Grassmann manifolds
admit highly non-linear structures and direct sample averaging is not admissible. Fortunately, we show that the leading $r_0$ eigenvectors of the average projection matrices $\bar{P}_{R_t} = \frac{1}{t} \sum_t \hat{R}_t \hat{R}_t^\top / T$ and $\bar{P}_{C_t} = \frac{1}{t} \sum_t \hat{C}_t \hat{C}_t^\top / T$, denoted by $\hat{R}/\sqrt{p}$ and $\hat{C}/\sqrt{q}$, could serve as the representatives of the Manifolds.

**Remark 2.2.** Although such manifold center intuition seems vivid, we have to be more careful with those degenerated cases where the best linear subspace approximation for each single matrix data is of lower dimensional than expected. For instance, if the factor matrix $F_t$ is of dimension $p_0 \times q_0$ with $p_0 > q_0$, and we wish to find a $q_0$-dimensional subspace Span($\hat{R}$). The signal part $RF_tC_t^\top$ would be of rank $r_0 = q_0$, so it would be more natural to set $\hat{R}_t$ as the leading $q_0$ left singular vectors of $X_t$, instead of $p_0$. The algorithm could be slightly modified and performs equally well, but the intuition of manifold “center” no longer makes sense, as each $\hat{R}_t$ is $q_0$-dimensional while the $\hat{R}$ we seek is $p_0$-dimensional. That is to say, we are finding the $p_0$-dimensional “centered” subspace among some $q_0$-dimensional subspaces, hence we adopt the term degeneration.

### 2.3 Best Subspace Approximations

First, consider the best subspace approximation for a single matrix data $X_t$. We aim to find basis matrices $\hat{R}_t$ and $\hat{C}_t$ of dimension $p \times p_0$ and $q \times q_0$ respectively, and the $p_0 \times q_0$ compressed factor matrix $\hat{F}_t$ such that $\hat{R}_t \hat{F}_t \hat{C}_t^\top$ is sufficiently close to the original $X_t$. It is then natural to solve the following optimization problem for some matrix norm $\| \cdot \|$, 

$$
(\hat{R}_t, \hat{F}_t, \hat{C}_t) = \arg \min_{R_t \in O_p, F_t \in O_{p_0}, C_t \in O_{q_0}} \| X_t - R_t F_t C_t^\top \|.
$$  \hspace{1cm} (2.2)

For the Frobenius norm $\| \cdot \|_F$, it is well-known that if $\hat{R}_t$ and $\hat{C}_t$ are given, $\hat{F}_t$ would simply be the projected value $\hat{R}_t^\top X_t \hat{C}_t$, see He et al. (2021b). After simple matrix manipulation, $\hat{R}_t$ would be the leading eigenvectors of $X_t \hat{C}_t^\top X_t^\top$ and $\hat{C}_t$ would be the leading eigenvectors of $X_t^\top \hat{C}_t \hat{R}_t \hat{R}_t^\top X_t$.

For the operator norm $\| \cdot \|_{op}$, unfortunately, a close form solution for $\hat{F}_t$ even if $R_t$ and $C_t$ are given is vacant. However, the optimized value $\mathcal{M}(R_t, C_t) = \| X_t - R_t \hat{F}_t C_t^\top \|_{op}$ has a closed form for $\hat{F}_t = \arg \min_{F_t} \| X_t - R_t F_t C_t^\top \|_{op}$, if the largest singular values of $X_t - R_t \hat{F}_t C_t^\top$ do not coincide so that the operator norm $\| \cdot \|_{op}$ is differentiable, which is:

$$
\mathcal{M}(R_t, C_t) = \| X_t - R_t \hat{F}_t C_t^\top \|_{op} = \sigma_{R_t} \vee \sigma_{C_t},
$$

where $\sigma_{R_t}^2$ and $\sigma_{C_t}^2$ are respectively the largest singular values of the matrices $\Sigma_{R_t}$ and $\Sigma_{C_t}$, defined as 

$$
\Sigma_{R_t} = (I - P_{R_t})X_t, \quad \Sigma_{C_t} = (I - P_{C_t})X_t^\top.
$$

We could minimize $\sigma_{R_t}$ and $\sigma_{C_t}$ separately. It is then straightforward that $\hat{R}_t$ and $\hat{C}_t$ are the leading $p_0$ left and $q_0$ right singular vectors of the matrix data $X_t$. In the end, as the optimization problem is sufficiently continuous and operator norm $\| \cdot \|_{op}$ is differentiable almost everywhere except for a zero Lebesgue measure set, the above solutions would be numerically valid.
2.4 MPCA algorithms for MEFM

In this section, we give the details of the MPCA algorithms. We first discuss the operator loss approximation case, naming this variant as MPCA\(_{op}\). As discussed earlier, MPCA\(_{op}\) first acquires the best linear subspace approximation \(\hat{R}_t\) and \(\hat{C}_t\) for each data matrix \(X_t\) by singular value decompositions. For non-degenerated cases namely \(r_0 = p_0 = q_0\), \(\hat{R}_t\) and \(\hat{C}_t\) would be the leading \(r_0\) left and right singular vectors of \(X_t\), which are the representatives of elements on the Grassmann manifolds \(G(p_0, p)\) and \(G(q_0, q)\) respectively. Then, MPCA\(_{op}\) finds the centers by minimizing the projection metrics on Grassmann manifolds:

\[
\hat{R}_{op}/\sqrt{p} = \arg\min_{R^T R = I_{p_0}} \sum_{t=1}^{T} \| RR^T - R_t \hat{R}_t^T \|_F^2, \quad \hat{C}_{op}/\sqrt{q} = \arg\min_{C^T C = I_{q_0}} \sum_{t=1}^{T} \| CC^T - \hat{C}_t \hat{C}_t^T \|_F^2,
\]

which is analogous to the physical notion of barycenter. Without loss of generality, we only focus on discussion of \(\hat{R}\) here. Denote \(P_R = RR^T\) and \(P_{\hat{R}_t} = \hat{R}_t \hat{R}_t^T\), then we have:

\[
\sum_t \| RR^T - R_t \hat{R}_t^T \|_F^2 = \sum_t \text{tr}(P_R) + \sum_t \text{tr}(P_{\hat{R}_t}) - 2 \text{tr} \left[ P_R (\sum_t P_{\hat{R}_t}) \right].
\]

Algorithm 1 MPCA algorithm under the operator norm loss.

Require:
- The set of all data matrices, \(\{X_t\}\):
- Compression dimensions, \(p_0 \leq r\) and \(q_0 \leq r\);

Ensure:
- Estimators by MPCA\(_{op}\), \(\hat{R}_{op}\) and \(\hat{C}_{op}\);

1: Acquire the best linear subspace estimations \(\hat{R}_t\) and \(\hat{C}_t\) for each \(X_t\). For MPCA\(_{op}\), \(\hat{R}_t\) and \(\hat{C}_t\) are the leading \(r_0 = p_0 \wedge q_0\) eigenvectors of \(X_t X_t^\top\) and \(X_t^\top X_t\);
2: The \(\hat{R}_{op}/\sqrt{p}\) and \(\hat{C}_{op}/\sqrt{q}\) are the leading \(p_0\) and \(q_0\) eigenvectors of the average projection matrices \(\sum_t \hat{R}_t \hat{R}_t^T / T\) and \(\sum_t \hat{C}_t \hat{C}_t^T / T\);
3: return \(\hat{R}_{op}, \hat{C}_{op}\).

The first two terms on the right hand side are fixed, so we are actually maximizing the last term, namely \(\mathcal{L}(R) = \text{tr} \left[ R^\top (\sum_t P_{\hat{R}_t}) R \right]\). It is a classical eigenvalue problem and \(\hat{R}_{op}/\sqrt{p}\) would be the leading \(p_0\) eigenvectors of the average projection matrix \(\sum_t P_{\hat{R}_t} / T\).

As for the degenerated case where \(p_0 \neq q_0\), without loss generality we assume \(p_0 > q_0\), then \(\hat{R}_t\) and \(\hat{C}_t\) are the leading \(r_0 = p_0 \wedge q_0\) left and right singular vectors of \(X_t\). Solve the same optimization problem and \(\hat{R}_{op}/\sqrt{p}\), \(\hat{C}_{op}/\sqrt{q}\) are still the leading \(p_0, q_0\) eigenvectors of the average projection matrices \(\sum_t P_{\hat{R}_t} / T\), \(\sum_t P_{\hat{C}_t} / T\) respectively. It is easy to see that \(\hat{C}_{op}/\sqrt{q}\) is exactly the same as in the non-degenerated case. As for \(\hat{R}_t\), it is of size \(p \times q_0\). It is no longer a representative of some element in \(G(p_0, p)\), thus the manifold center intuition no longer holds. However similar geometric interpretation is still somehow valid: each \(\hat{R}_t\) corresponds to a \(q_0\)-dimensional subspace, and with some principal angle related arguments, \(\text{tr}(P_{\hat{R}_t} P_R)\) still gives the magnitude of deviation of \(q_0\)-dimensional Span(\(\hat{R}_t\)) from the \(p_0\)-dimensional Span(\(R\)). It is obvious that MPCA\(_{op}\) needs \(p_0 \lor q_0 \leq r = p \wedge q\), as we are seeking
for lower dimensional column and row subspaces, and the right hand side is the maximal rank of the original data matrices \( \{X_t\} \). The detailed procedures for MPCA\(_{op}\) is summarized in Algorithm 1.

We now discuss the Frobenius loss case, naming this variant as MPCA\(_F\). Under Frobenius loss, \( \hat{R}_t \) would be the leading \( r_0 \) eigenvectors of \( X_tP_CC_t^\top \) if \( C \) is given and \( \hat{C}_t \) would be the leading \( r_0 \) eigenvectors of \( X_t^\top P_R X_t \) if \( R \) is given. Then MPCA\(_F\) takes the leading eigenvectors of the average projection matrices similarly. Note that iterative procedure is necessary as \( C \) and \( R \) is unavailable at the beginning. For initial value, the estimators by MPCA\(_{op}\) can be adopted as a warm start. The detailed procedures for MPCA\(_F\) is summarized in Algorithm 2.

### Algorithm 2

**MPCA algorithm under the Frobenius norm loss.**

**Require:**

The set of all data matrices, \( \{X_t\} \);

Compression dimensions, \( p_0 \leq r \) and \( q_0 \leq r \);

**Ensure:** Estimators by MPCA\(_F\), \( \hat{R}_F \) and \( \hat{C}_F \);

1. Use the result of MPCA\(_{op}\) as a warm start, denoted by \( \hat{R}_F^{(0)} \) and \( \hat{C}_F^{(0)} \);
2. Assume we have acquired \( \hat{R}_t^{(i)} \) and \( \hat{C}_t^{(i)} \), then \( \hat{R}_t^{(i+1)} \) and \( \hat{C}_t^{(i+1)} \) are the leading \( r_0 = p_0 \wedge q_0 \) eigenvectors of \( X_tP_{C_t}X_t^\top \) and \( X_t^\top P_{R_t}X_t \) respectively;
3. Then \( \hat{R}_t^{(i+1)}/\sqrt{p} \) and \( \hat{C}_t^{(i+1)}/\sqrt{q} \) are the leading \( p_0 \) and \( q_0 \) eigenvectors of the average projection matrices \( \sum_t \hat{R}_t^{(i+1)}(\hat{R}_t^{(i+1)})^\top/T \) and \( \sum_t \hat{C}_t^{(i+1)}(\hat{C}_t^{(i+1)})^\top/T \) respectively; iterate until convergence to \( \hat{R}_F, \hat{C}_F \);
4. return \( \hat{R}_F, \hat{C}_F \).

### 3 Theoretical Results

In this section, we present the theoretical properties of the estimators by MPCA, and throughout this section, the number of factors \( p_0 \) and \( q_0 \) are treated as given. The determination of factor numbers \( p_0, q_0 \) are left to Section 4.

#### 3.1 Expected Projection Matrix

Prior to presenting the consistency of our MPCA estimators, we first give some intuitions on why the algorithms work. For clearer illustration, we only analyze \( \hat{R}_{op} \) by MPCA\(_{op}\). Recall that the model is \( X_t = RF_tC^\top + E_t \), while \( \hat{R}_t \) is the leading \( r_0 \) left singular vectors of \( X_t \). In the end, \( \hat{R}_{op}/\sqrt{p} \) is acquired by taking the leading \( p_0 \) eigenvectors of the average projection matrix \( P_{\hat{R}_t} \). Obviously, the algorithm relies heavily on the concentration of average projection matrix \( E_{P_{\hat{R}_t}} \). The algorithms could be justified as long as we show that:

1. The average projection matrix \( E_{P_{\hat{R}_t}} \) converges to the expected version \( \mathbb{E}P_{\hat{R}_t} \) at a rate of \( \sqrt{T} \).
2. The leading eigenvectors of the expected version \( \mathbb{E}P_{\hat{R}_t} \) give us \( \text{Span}(R) \) as desired.

As for the first part, with the help of matrix concentration inequalities in Tropp (2012), we have:
Lemma 3.1 ($\sqrt{T}$-Convergence). For i.i.d. random projection matrices $P_{\tilde{R}_t}$ with dimension $p$, for all $x \geq 0$, the following concentration inequality holds,

$$
P \left\{ \left\| \sum_{t=1}^{T} (P_{\tilde{R}_t} - \mathbb{E}P_{\tilde{R}_t}) \right\|_{op} \geq x \right\} \leq p \cdot e^{-x^2/8T}.
$$

Although we are quite satisfied with this $\sqrt{T}$-consistency result for finite-dimensional matrix data, the haunting dimensional factor $p$ of matrix concentration inequalities would give exploding bounds if the dimension $p$ tends to infinity. Fortunately, in this case of random projection matrices, we are able to shrink the dimensional factor $p$ to $r_0$ via intrinsic dimension arguments. As $r_0$ remains fixed as $p$ goes to infinity, dimension-free convergence could be acquired.

As for the second part, we claim that $\text{Span}(R)$ and $\text{Span}(R^\perp)$ are invariant subspaces of the expected projection matrix $\mathbb{E}P_{\tilde{R}_t}$ if the noise $E_t$ is left spherical. Matrix spherical distribution can be viewed as a special case of matrix elliptical distribution. The random matrix $X$ is left spherical if $X \sim E_{p,q}(0, \Sigma \otimes \mathcal{I}, \psi_l)$, right spherical if $X \sim E_{p,q}(0, \mathcal{I} \otimes \Omega, \psi_r)$ and spherical if $X \sim E_{p,q}(0, \mathcal{I} \otimes \mathcal{I}, \psi_s)$. Right spherical and spherical distributions have similar properties accordingly. Random matrices with i.i.d. centered Gaussian or $t_v$ elements are matrix spherically distributed, see Gupta and Nagar (2018) for details.

Lemma 3.2 (Invariant Subspaces). For joint matrix elliptical data $X_t = RF_tC^T + E_t$ as in 2.1, let $P_{\tilde{R}_t} = \tilde{R}_t\tilde{R}_t^T$, where $\tilde{R}_t$ is the leading $r_0 = p_0 \wedge q_0$ eigenvectors of $X_tX_t^T$. If $E_t$ is left spherical, then $\text{Span}(R)$ and $\text{Span}(R^\perp)$ are invariant subspaces of $\mathbb{E}P_{\tilde{R}_t}$.

The joint matrix elliptical model here is more of a burden instead of blessing. In fact, the conclusion is more straightforward if $F_t$ and $E_t$ are independent and could be of independent interest. The combination of Lemmas 3.1 and 3.2 theoretically justifies the validity of the proposed MPCA methods: the expected projection matrix contains adequate subspace information, while the matrix concentration to it is guaranteed by the compactness of the projection matrices.

3.2 Technical Assumptions

In this section, we give some technical assumptions to establish the convergence rates of the estimators by MPCA.

Assumption A (Joint Matrix Elliptical Model). We assume matrix elliptical factor model as:

$$
X_t = RF_tC^T + E_t, \quad t = 1, \ldots, T,
$$

$$
\begin{pmatrix}
\text{Vec}(F_t) \\
\text{Vec}(E_t)
\end{pmatrix} = r_t \begin{pmatrix}
\Sigma_2^{1/2} \otimes \Sigma_1^{1/2} & 0 \\
0 & \Omega_2^{1/2} \otimes \Omega_1^{1/2}
\end{pmatrix} \frac{Z_t}{\|Z_t\|_2},
$$

where $Z_t$ is a $(pq + p_0q_0)$-dimensional isotropic Gaussian vector, $r_t$ is a positive random variable independent of $Z_t$, with $(pq)^{-1/2}r_t = O_p(1)$ as $p, q \to \infty$. It is sometimes more convenient to separate the
joint model into:

\[ F_t = \frac{r_t}{\|Z_t\|_2} \Sigma_1^{1/2} Z_t^\top \Sigma_2^{1/2}, \quad E_t = \frac{r_t}{\|Z_t\|_2} \Omega_1^{1/2} Z_t^\top \Omega_2^{1/2}, \]

with \( \Sigma_1 \) of size \( p_0 \times p_0 \), \( \Sigma_2 \) of size \( q_0 \times q_0 \), \( \Omega_1 \) of size \( p \times p \) and \( \Omega_2 \) of size \( q \times q \). \( Z_t^F \) is a \( p_0 \times q_0 \) random matrix made by leading \( p_0q_0 \) elements of \( Z_t \), while \( Z_t^E \) of size \( p \times q \) consists of all the elements left.

**Assumption B** (Strong Factor Conditions). We assume \( R^\top R/p = I_{p_0} \) and \( C^\top C/q = I_{q_0} \). In addition, there exist positive constants \( c_1 \) and \( C_1 \) such that \( c_1 \leq \lambda_{p_0}(\Sigma_1) \leq \lambda_1(\Sigma_1) \leq C_1 \), \( c_1 \leq \lambda_{q_0}(\Sigma_2) \leq \lambda_1(\Sigma_2) \leq C_1 \) for \( p, q \rightarrow \infty \).

**Assumption C** (Regular Noise Conditions). We assume there exist positive constants \( c_2 \) and \( C_2 \) such that \( c_2 \leq \lambda_p(\Omega_1) \leq \lambda_1(\Omega_1) \leq C_2 \), \( c_2 \leq \lambda_q(\Omega_2) \leq \lambda_1(\Omega_2) \leq C_2 \) as \( p, q \rightarrow \infty \).

The convergence relies heavily on matrix concentration results. The independence between \( \{X_t\} \) in Assumption A could extend readily to weak dependence by matrix concentration results such as matrix Azuma inequality, see Tropp (2012), Tropp (2015) for details. Assumption B and C are standard in large-dimensional factor models. In addition, the joint matrix elliptical distribution assumption is only for the convenience of theoretical analysis, while empirical experiments show that the MPCA methods are not sensitive to the elliptical assumption.

### 3.3 Consistency of Manifold PCA

**Theorem 3.3** (Consistency of MPCA_{op}). For MPCA_{op}, under Assumption A to C, there exist \( p_0 \times p_0 \) orthonormal matrix \( H_R \) and \( q_0 \times q_0 \) orthonormal matrix \( H_C \) such that:

\[
\|\hat{R}_{op} - RH_R\|_F^2/p = O_p(T^{-1} + p^{-1/2} + q^{-1/2}),
\]

\[
\|\hat{C}_{op} - CH_C\|_F^2/q = O_p(T^{-1} + p^{-1/2} + q^{-1/2}).
\]

As MPCA methods rely heavily on the concentration of \( T \) projection matrices, the convergence rate would be at most \( T^{-1} \). It is slower than \( \alpha \)-PCA from Chen and Fan (2021) and PE from Yu et al. (2021) with the rate \( (Tq)^{-1} \) (or \( (Tp)^{-1} \) when estimating \( R \) (or \( C \)) under strong signal conditions, which are comparable to taking each column (or row) as individual observation. The inefficiency of MPCA methods comes from the fact that by taking each projection matrix as individual observation would lose information especially when each matrix observation \( X_t \) is of large dimensions. However, as we will see in the simulation study, MPCA performs comparably to the classical methods for Gaussian noise, while the compactness of the projection matrices ensures the good performance of MPCA even for noises without any moments. As a result, they could be potential replacements of classical methods for data with heavier-tailed noise.

For MPCA_{F} method, we further discuss how the information of column factor loading matrix \( C \) would help to estimate \( R \), which is named as the projection effect in this work. Estimation of \( C \) can be discussed in a similar way. Recall that for MPCA_{F}, \( \hat{R}_{i}^{(1)} \) would be the leading \( r_0 = p_0 \wedge q_0 \) eigenvalues of \( X_tP_{C^{(i-1)}}X_t^\top \) if \( \hat{C}^{(i-1)} \) is given. Let \( C^{(i-1)} = \hat{C}^{(i-1)}/\sqrt{q} \) for notational simplicity. We focus on the projected matrix model \( X_tC^{(i-1)} = RF_tC^\top C^{(i-1)} + E_tC^{(i-1)} \), and \( \hat{R}_{i}^{(1)} \) would exactly be the estimator by the MPCA_{op} to the projected data set \( \{X_tC^{(i-1)}\} \).
The difference with/without projection lies in the signal-to-noise ratio level. Consider the extreme case where the true value $C$ is known, for $X_tC/\sqrt{q} = q^{1/2}RF_t + E_tC/\sqrt{q}$, the projection does no harm to the signal size while compressing the noise to lower-dimensional $E_tC/\sqrt{q}$. It is then foreseeable that we could increase the signal-to-noise ratio via projection by some $\hat{C}^{(i-1)}$ sufficiently close to $C$, keeping the signal size almost unchanged. It is ensured by assuming $\sigma_{q_0}(C^T\hat{C}^{(i-1)})/q = c > 0$, which is a rather mild condition.

**Theorem 3.4 (Projection Effect of MPCA$_F$).** For each iteration step of MPCA$_F$, under Assumption A to C, given $\hat{C}^{(i-1)}$, if we assume that $\sigma_{q_0}(C^T\hat{C}^{(i-1)})/q > 0$, then there exists $p_0 \times p_0$ orthonormal matrix $H_R$ such that:

$$\|\hat{R}^{(i)} - RH_R\|^2_F/p = Op\left(T^{-1} + q^{-1/2}\right).$$

Similarly, given $\hat{R}^{(i-1)}$, if we assume that $\sigma_{p_0}(R^T\hat{R}^{(i-1)})/p > 0$, then there exists $q_0 \times q_0$ orthonormal matrix $H_C$ such that:

$$\|\hat{C}^{(i)} - CH_C\|^2_F/p = Op\left(T^{-1} + p^{-1/2}\right).$$

As we take the estimator MPCA$_{op}$ as the initial estimate, which is shown to be consistent under Assumption A to C by Theorem 3.3, we have $\sigma_{q_0}(C^T\hat{C}^{(0)})/q > 0$ and $\sigma_{p_0}(R^T\hat{R}^{(0)})/p > 0$ with probability tending to 1.

### 3.4 Factor and Common Component Matrices

After the loading matrices being determined, the factor matrix $F_t$ can be naturally estimated by $\hat{F}_t = \hat{R}^TX_tC/\sqrt{pq}$, and the common component matrix $S_t = RF_tC^T$ be estimated by $\hat{S}_t = \hat{R}\hat{F}_t\hat{C}^T$.

**Corollary 3.1 (Consistency of Factor and Common Component Matrices).** Suppose there exist $p_0 \times p_0$ orthonormal matrix $H_R$ and $q_0 \times q_0$ orthonormal matrix $H_C$ such that for $\varepsilon_R = \hat{R} - RH_R$ and $\varepsilon_C = \hat{C} - CH_C$, we have $\|\varepsilon_R/\sqrt{p}\|_{op} = o_p(1)$ and $\|\varepsilon_C/\sqrt{q}\|_{op} = o_p(1)$, then:

$$\|\hat{F}_t - H_R F_tC\|_{op} = Op\left(\|\varepsilon_R/\sqrt{p}\|_{op} + \|\varepsilon_C/\sqrt{q}\|_{op} + (pq)^{-1/2}\right),$$

$$\|\hat{S}_t - S_tC/\sqrt{pq}\|_{op} = Op\left(\|\varepsilon_R/\sqrt{p}\|_{op} + \|\varepsilon_C/\sqrt{q}\|_{op} + (pq)^{-1/2}\right).$$

Here $\|\varepsilon_R/\sqrt{p}\|_{op} = o_p(1)$ and $\|\varepsilon_C/\sqrt{q}\|_{op} = o_p(1)$ are direct consequences of Theorem 3.3 under Assumptions A to C, so we claim the consistency of factor and common component matrices.

### 4 Determining the Factor Numbers

In the last section, the factor number is assumed to be known in advance, while in practice, the factor numbers $p_0$ and $q_0$ need to be determined. We propose a natural criterion by calculating eigenvalue-ratios (ER) of the average projection matrices, under both MPCA$_{op}$ and MPCA$_F$. The corresponding algorithms are named as MER$_{op}$ and MER$_F$ respectively. Unlike existing eigenvalue-ratio methods
based on covariance-type matrices as in Chen and Fan (2021) and Yu et al. (2021), MER\(_{\text{op}}\) and MER\(_{F}\) are clearly free of moment-constraints. For MER\(_{\text{op}}\), first determine the compression rank \(\hat{r}_0\) by averaging \(\hat{r}_{0,t}\) acquired from each data matrix \(X_t\), that is:

\[
\hat{r}_{0,t} = \arg \max_{j \leq r_{\text{max}}} \frac{\sigma_j(X_t)}{\sigma_{j+1}(X_t)}, \quad \hat{r}_0 = \lfloor \sum_{t=1}^T \hat{r}_{0,t}/T + \frac{1}{2} \rfloor,
\]

where \(\lfloor x + \frac{1}{2} \rfloor\) means rounding \(x\) to the nearest integer. Then \(p_0\) and \(q_0\) are estimated by:

\[
\hat{p}_0 = \arg \max_{j \leq r_{\text{max}}} \frac{\lambda_j(P_{\hat{R}_t})}{\lambda_{j+1}(P_{\hat{R}_t})}, \quad \hat{q}_0 = \arg \max_{j \leq r_{\text{max}}} \frac{\lambda_j(P_{\hat{C}_t})}{\lambda_{j+1}(P_{\hat{C}_t})}, \tag{4.1}
\]

where \(r_{\text{max}}\) is predetermined value larger than \(p_0, q_0\), while \(P_{\hat{R}_t}, P_{\hat{C}_t}\) are the average projection matrices by taking the leading \(\hat{r}_0\) left and right singular vectors of each \(X_t\) respectively.

**Remark 4.1.** In fact, accurate estimation of \(\hat{r}_0\) is not necessary for MER\(_{\text{op}}\), we could still get comparable results from 4.1 even if \(\hat{r}_0 \neq r_0\). However, since each data matrix contains at most \(r_0\)-dimensional subspace information, pre-estimation of \(r_0\) could stabilize the algorithm. Once \(r_0\) has been correctly estimated, \(\hat{R}_t\) would be exactly \(R_t\) in MPCA\(_{\text{op}}\). Further analysis in supplementary materials ensures that \(\lambda_{p_0}(E_{\hat{R}_t}) \geq c > 0, \lambda_{p_0+1}(E_{\hat{R}_t}) \to 0\) while \(\|P_{\hat{R}_t} - E_{\hat{R}_t}\|_{\text{op}} \to 0\) as \(T, p, q \to \infty\) under Assumption A to C, which theoretically justifies MER\(_{\text{op}}\).

**Algorithm 3** MER\(_{\text{op}}\) estimators of the pair of the factor numbers

**Require:**
- The set of all data matrices, \(\{X_t\}\);
- Maximum number, \(r_{\text{max}}\);

**Ensure:**
- MER\(_{\text{op}}\) estimators, \(\hat{p}_0\) and \(\hat{q}_0\);

1: Acquire compression dimension \(\hat{r}_0\) by averaging \(\hat{r}_{0,t}\) from each data matrix \(X_t\), where \(\hat{r}_{0,t} = \arg \max_{j \leq r_{\text{max}}} \sigma_j(X_t)/\sigma_{j+1}(X_t)\) and \(\hat{r}_0 = \lfloor \sum_{t=1}^T \hat{r}_{0,t}/T + \frac{1}{2} \rfloor\);
2: Acquire the best linear subspace estimations \(\hat{R}_t\) and \(\hat{C}_t\) for each \(X_t\), which are the leading \(\hat{r}_0\) eigenvectors of \(X_tX_t^\top\) and \(X_t^\top X_t\);
3: Calculate the average projection matrices \(P_{\hat{R}_t}, P_{\hat{C}_t}\) from \(\hat{R}_t\) and \(\hat{C}_t\). Determine \(\hat{p}_0\) and \(\hat{q}_0\) by finding the largest eigenvalue-ratio from \(\lambda_j(P_{\hat{R}_t})/\lambda_{j+1}(P_{\hat{R}_t})\) and \(\lambda_j(P_{\hat{C}_t})/\lambda_{j+1}(P_{\hat{C}_t})\) for \(j \leq r_{\text{max}}\);
4: return \(\hat{p}_0, \hat{q}_0\).

**Theorem 4.2** (Consistency of MER\(_{\text{op}}\) estimators). Under Assumption A to C, assume the maximum number \(r_{\text{max}} \geq p_0 \lor q_0\), as \(T, p, q \to \infty\),

\[
P(\hat{p}_0 = p_0) \to 1, \quad P(\hat{q}_0 = q_0) \to 1.
\]

Similarly, we could use the average projection matrices from MPCA\(_F\) to increase accuracy. Now.
that estimating \( p_0 \) requires information of \( C \), and estimating \( q_0 \) requires information of \( R \), iterations are naturally needed, and the result from \( \text{MER}_\text{op} \) estimators could be used as a warm start.

**Algorithm 4** \( \text{MER}_F \) estimators of the pair of the factor numbers

**Require:**
- The set of all data matrices, \( \{X_t\} \);
- Maximum number, \( r_{\text{max}} \);

**Ensure:**
- \( \text{MER}_F \) estimators, \( \hat{p}_0^F \) and \( \hat{q}_0^F \);
- Given the \( \text{MER}_\text{op} \) estimators as a warm start, denoted by \( \hat{p}_0^{(0)}, \hat{q}_0^{(0)} \);
- Acquire the best linear subspace estimations \( \tilde{R}_t^{(i+1)} \) and \( \tilde{C}_t^{(i+1)} \) for each \( X_t \), which are the leading \( \hat{p}_0^{(i)} \) and \( \hat{q}_0^{(i)} \) eigenvectors of \( X_t P_{C_t^{(i)}} X_t^\top \) and \( X_t^\top P_{C_t^{(i)}} X_t \) respectively;
- Calculate the average projection matrices \( \bar{R}_t, \bar{C}_t \) from \( \tilde{R}_t^{(i+1)} \) and \( \tilde{C}_t^{(i+1)} \). Determine \( \hat{p}_0^{(i+1)} \) and \( \hat{q}_0^{(i+1)} \) by finding the largest eigenvalue-ratio from \( \lambda_j(\bar{R}_t)/\lambda_j(\bar{C}_t) \) and \( \lambda_j(\bar{R}_t)/\lambda_{j+1}(\bar{C}_t) \) for \( j \leq r_{\text{max}} \); iterate until convergence to \( \hat{p}_0^F, \hat{q}_0^F \);
- \text{return} \( \hat{p}_0^F, \hat{q}_0^F \).

Theoretical analysis of \( \text{MER}_F \) is challenging due to the iteration procedure, thankfully the initial step taken from \( \text{MER}_\text{op} \) has already been consistent under Assumption A to C. As shown in simulations, \( \text{MER}_F \) benefits from the same projection technique as in \( \text{MPCA}_F \) and turns out to be more accurate than \( \text{MER}_\text{op} \) in finite-sample performances.

### 5 Simulation Results

In this section, we investigate the finite-sample performances of MPCA algorithms by generating synthetic datasets. The observed data matrices are generated as \( X_t = RF_t C^\top + E_t \) from moderate noise regime, by rescaling signal and noise to the same scale.

#### 5.1 Data Generation

To generate observations from the model \( X_t = RF_t C^\top + E_t, t = 1, \ldots, T \), we set \( p_0 = q_0 = 3 \), draw the entries of \( R \) and \( C \) from independent standard Gaussian distribution, and let:

\[
F_t = \Omega \times F_{t-1} + \sqrt{1 - \delta^2} \times U_t, \quad U_t \overset{i.i.d.}{\sim} \mathcal{MN} (0, I_{p_0}, I_{q_0}). \]

\[
E_t = \psi \times E_{t-1} + \sqrt{1 - \psi^2} \times s_E \times \Omega_1^{1/2} \Omega_2^{1/2},
\]

where \( V_t = \gamma W_t \) and the elements of \( W_t \) are generated by independent standard centered Gaussian, \( t_v \), skewed-\( t_v \), and \( \alpha \)-stable distributions. The rescaling constant \( \gamma \) is to get comparable signal and noise level. In fact, if \( W_t \) consists of independent standard centered Gaussian random variables, the signal part \( \|RF_t C^\top\|_{\text{op}} \asymp (pq)^{1/2} \) while \( \|W_t\|_{\text{op}} \asymp (p \lor q)^{1/2} \), thus by setting \( \gamma = (p \land q)^{1/2} \) we get comparable
signal and noise level. On the other hand, if $W_t$ consists of independent $t_1$ random variables, it holds $\|W_t\|_{op} \geq pq$, then we need to set $\gamma = (pq)^{-1/2}$.

The pair of dimensions $(p, q)$ are chosen from the set \{(20, 20), (20, 100), (100, 100)\}, the sample size $T$ is set to be $3(pq)^{1/2}$ and $\Omega_1, \Omega_2$ are set to be matrices with ones on the diagonal, and $1/p, 1/q$ on the off-diagonal respectively. In addition, the parameters $\phi$ and $\psi$ control temporal correlation and are set as $\phi = \psi = 1/10$, while $s_E$ is the noise scaling constant chosen from \{1, 1.5, 2\}. We only show the cases with $s_E = 1$ in this section, and the rest are left to the supplementary materials. Elements of $V_t$ are drawn independently from standard Gaussian, $t_3, t_1, \alpha$-stable distribution ($\alpha = 1.8$, skewness parameter $\beta = 0$), and skewed-$t_3$ distribution (standard deviation $\sigma = \sqrt{3}$, skewness parameter $\nu = 2$) respectively, while we set $\gamma = (pq)^{-1/2}$ for $t_1$ distribution and $\gamma = (p \wedge q)^{1/2}$ for the rest distributions. We generate $\alpha$-stable distribution by Python package scipy.stats, and skewed-$t_3$ distribution by Python package ssstudent. All simulation results reported here are based on 100 replications.

### 5.2 Estimation of Loading Spaces

We first compare the performances of MPCA algorithms with those of $(2D)^2$-PCA by Zhang and Zhou (2005) and PE method by Yu et al. (2021) in terms of estimating loading spaces. In fact, $(2D)^2$-PCA is equivalent to $\alpha$-PCA from Chen and Fan (2021) with $\alpha = -1$, whose empirical performances corresponding to $\alpha \in \{-1, 0, 1\}$ are comparable as shown in Yu et al. (2021). To measure the difference between the estimated $\hat{R}$ and the true loading $R$, we used the scaled projection metric as in Yu et al. (2021), which is defined as:

$$D(\hat{R}, R) = \left(1 - \frac{1}{\mu_0} \text{tr}(P_R^T P_R)\right)^{1/2} = (2\mu_0)^{-1/2}\|P_R^T - P_R\|_F,$$

so it is straightforward that $D(\hat{R}, R)$ is always between 0 (corresponding to $\text{Span}(\hat{R}) = \text{Span}(R)$) and 1 (corresponding to $\text{Span}(\hat{R})$ and $\text{Span}(R)$ are orthogonal). $D(\hat{C}, C)$ can be defined similarly.

Table 1 shows the averaged estimation errors of factor loadings with standard deviations in parentheses under different noise distributions with $s_E = 1$. Simulation results with $s_E \in \{1.5, 2\}$ are reported in Table 6 and 7 in the supplementary materials. First, it is observed that the projection effect of MPCA$^F$ and PE is obvious for large-dimensional matrix factor analysis in finite-samples. The projected MPCA$^F$ and PE almost always show advantages over MPCA$_{op}$ and $(2D)^2$-PCA, which in fact correspond to their non-projected versions. For the cases with Gaussian noise, the PE method achieves the best performances, while MPCA$^F$ performs comparably. However, for heavy-tailed noises it is a completely different picture. MPCA$^F$ shows great advantage over PE under $t_3, t_1$ and $\alpha$-stable noises. It is foreseeable since MPCA methods require no moment conditions. Under relatively small noise scale, MPCA$_{op}$ shows comparable performances as MPCA$^F$, but the latter benefits greatly from the projection effect and is more robust against larger noise scale, as shown in Table 6 and Table 7. In addition, by comparing the results under $t_3$ and skewed-$t_3$ noise, we observe that skewness does almost no harm to MPCA methods, while increasing the estimation errors of $(2D)^2$-PCA and PE. To summarize, both MPCA$^F$ and PE benefit greatly from the projection effect, which is essential in large-dimensional matrix factor analysis. MPCA methods perform comparably with $(2D)^2$-PCA and PE under light-tailed noises, but
Table 1: Means and standard deviations (in parentheses) of $D(\hat{R}, R)$ and $D(\hat{C}, C)$ over 100 replications with $s_E = 1$ and $T = 3(pq)^{1/2}$. Here MPCA_{op} and MPCA_{F} stands for Manifold PCA methods; $(2D)^2$-PCA is from Zhang and Zhou (2005), it is equivalent to $\alpha$-PCA by Chen and Fan (2021) with $\alpha = -1$; PE stands for the projected estimation by Yu et al. (2021).

| Distribution | Evaluation | $p$ | $q$ | MPCA_{op} | MPCA_{F} | $(2D)^2$-PCA | PE |
|--------------|------------|----|----|-----------|-----------|-------------|----|
| Gaussian     | $D(\hat{R}, R)$ | 20 20 | (0.3024, 0.1272) | (0.1154, 0.0233) | (0.2007, 0.1147) | (0.0833, 0.0179) |
|              | 20 100 | (0.4510, 0.0739) | (0.0402, 0.0046) | (0.1375, 0.0756) | (0.0234, 0.0030) |
|              | 100 100 | (0.0878, 0.0226) | (0.0426, 0.0025) | (0.0632, 0.0173) | (0.0337, 0.0024) |
|              | $D(\hat{C}, C)$ | 20 20 | (0.0694, 0.0073) | (0.0665, 0.0058) | (0.0545, 0.0067) | (0.0521, 0.0059) |
|              | 20 100 | (0.0305, 0.1237) | (0.1170, 0.0239) | (0.1807, 0.1061) | (0.0838, 0.0182) |
|              | 100 100 | (0.0694, 0.0073) | (0.0665, 0.0058) | (0.0545, 0.0067) | (0.0521, 0.0059) |

much more robustly under heavy-tailed and skewed noises, and as a result are more suitable for financial and econometrical applications.

5.3 Estimation Errors for Common Components

In this section, we compare the performances of of MPCA algorithms with those of $(2D)^2$-PCA by Zhang and Zhou (2005) and PE method by Yu et al. (2021) in terms of estimating the common component matrices. We evaluate the performances by mean squared error (MSE) and maximum operator loss (opMax), which are defined as:

$$\text{MSE} = \frac{1}{Tpq} \sum_{t=1}^{T} \left\| \hat{S}_t - S_t \right\|_F^2,$$

$$\text{opMax} = \frac{1}{(pq)^{1/2}} \max_{1 \leq t \leq T} \left\| \hat{S}_t - S_t \right\|_{op}.$$
where $\hat{S}_t = P_tX_tP_C^t$ refers to the estimated common component matrix and $S_t$ is the true value.

Table 2: Means and standard deviations (in parentheses) of MSE and opMax over 100 replications with $s_E = 1$ and $T = 3(pq)^{1/2}$. Here $\text{MPCA}_{op}$ and $\text{MPCA}_F$ stands for Manifold PCA methods; $(2D)^2$-PCA is from Zhang and Zhou (2005), it is equivalent to $\alpha$-PCA by Chen and Fan (2021) with $\alpha = -1$; PE stands for the projected estimation by Yu et al. (2021).

| MSE Distribution | $p$ | $q$ | $\text{MPCA}_{op}$ | $\text{MPCA}_F$ | $(2D)^2$-PCA | PE |
|------------------|-----|-----|---------------------|-----------------|--------------|----|
| Gauss            | 20  | 20  | (0.0782,0.0222)    | (0.0327,0.0030) | (0.0463,0.0141) | (0.0280,0.0024) |
|                  | 100 |     | (0.0023,0.0005)    | (0.0012,0.0000) | (0.0016,0.0003) | (0.0011,0.0000) |

| $t_3$            | 20  | 20  | (0.2381,0.0299)    | (0.1012,0.0158) | (0.3367,0.1668) | (0.2206,0.2185) |
|                  | 100 |     | (0.0405,0.0095)    | (0.0886,0.0008) | (0.0767,0.2454) | (0.0466,0.2472) |

| $t_1$            | 20  | 20  | (46.248,271.94)    | (42.431,252.56) | (1968.3,13474)  | (1968.4,13474)  |
|                  | 100 |     | (4.6400,37.238)    | (4.0799,31.791) | (2180.1,19330)  | (2180.1,19330)  |

| $\alpha$-stable  | 20  | 20  | (0.6030,0.9509)    | (0.2638,0.3138) | (7.0910,29.585) | (7.1591,29.592) |
|                  | 100 |     | (0.0613,0.0313)    | (0.0243,0.0288) | (1.9485,3.6707) | (1.9865,3.6689) |

| skewed-$t_3$     | 20  | 20  | (0.2273,0.0504)    | (0.0979,0.0232) | (0.3977,0.2926) | (0.2918,0.3304) |
|                  | 100 |     | (0.0397,0.0089)    | (0.0085,0.0007) | (0.0747,0.1179) | (0.0442,0.1293) |

| opMax Distribution | $p$ | $q$ | $\text{MPCA}_{op}$ | $\text{MPCA}_F$ | $(2D)^2$-PCA | PE |
|-------------------|-----|-----|---------------------|-----------------|--------------|----|
| Gauss             | 20  | 20  | (0.0820,0.0207)    | (0.0501,0.0047) | (0.0600,0.0132) | (0.0490,0.0047) |
|                  | 100 |     | (0.0067,0.0011)    | (0.0047,0.0003) | (0.0053,0.0006) | (0.0047,0.0003) |

| $t_3$            | 20  | 20  | (0.1554,0.0377)    | (0.1166,0.0398) | (0.3523,0.2749) | (0.3187,0.3208) |
|                  | 100 |     | (0.0558,0.0145)    | (0.0252,0.0148) | (0.1291,0.2865) | (0.0927,0.2944) |

| $t_1$            | 20  | 20  | (3.2653,11.248)    | (3.1082,10.779) | (21.583,73.268) | (21.583,73.268) |
|                  | 100 |     | (0.8514,3.6067)    | (0.8207,3.3761) | (16.575,78.617) | (16.575,78.617) |

| $\alpha$-stable  | 20  | 20  | (0.1554,0.0377)    | (0.1166,0.0398) | (0.3523,0.2749) | (0.3187,0.3208) |
|                  | 100 |     | (0.0558,0.0145)    | (0.0252,0.0148) | (0.1291,0.2865) | (0.0927,0.2944) |

Table 2 reports the means and standard deviations of MSEs and opMaxs with $s_E = 1$. Simulation results with $s_E \in \{1.5, 2\}$ are reported in Table 8 and 9 in the supplementary materials. Similar as the conclusions drawn for factor loadings, both $\text{MPCA}_F$ and PE also benefit from the projection effect. MPCA methods are comparable with $(2D)^2$-PCA and PE under light-tailed noises, but are much more robust under the heavy-tailed and skewed noises.

5.4 Estimation of Factor Numbers

Accurate estimation of the pair of factor numbers is of vital importance in matrix factor analysis. In this section, we compare the empirical performances of our MER$_{op}$, MER$_F$ algorithms with IterER
method by Yu et al. (2021) and \( (2D)^2 \)-ER method, which is equivalent to the eigenvalue-ratio method in Chen and Fan (2021) with \( \alpha = -1 \).

Table 3: Frequencies of exact estimation and underestimation (in parentheses) of factor numbers over 100 replications with \( s_E = 1 \) and \( T = 3(pq)^{1/2} \). Here MER\(_{op}\) and MER\(_F\) stands for Manifold eigenvalue-ratio methods; \( (2D)^2 \)-ER is equivalent to the ER method in Chen and Fan (2021) with \( \alpha = -1 \); IterER is from Yu et al. (2021).

| Distribution | \( p \) | \( q \) | MER\(_{op}\) | MER\(_F\) | \( (2D)^2 \)-ER | IterER |
|--------------|--------|--------|-------------|-------------|----------------|--------|
| Gaussian     | 20     | 20     | (0.37,0.19) | (0.95,0.04) | (0.12,0.73)    | (0.94,0.06) |
|              | 20     | 100    | (0.98,0.00) | (1.00,0.00) | (0.19,0.37)    | (1.00,0.00) |
|              | 100    | 100    | (1.00,0.00) | (1.00,0.00) | (0.36,0.00)    | (1.00,0.00) |
| \( t_3 \)    | 20     | 20     | (0.13,0.65) | (0.53,0.43) | (0.04,0.83)    | (0.33,0.63) |
|              | 20     | 100    | (0.17,0.13) | (1.00,0.00) | (0.05,0.50)    | (0.80,0.02) |
|              | 100    | 100    | (0.04,0.02) | (1.00,0.00) | (0.00,0.24)    | (0.51,0.03) |
| \( t_1 \)    | 20     | 20     | (0.99,0.01) | (1.00,0.00) | (0.02,0.85)    | (0.01,0.76) |
|              | 20     | 100    | (1.00,0.00) | (1.00,0.00) | (0.08,0.82)    | (0.08,0.72) |
|              | 100    | 100    | (1.00,0.00) | (1.00,0.00) | (0.07,0.86)    | (0.05,0.88) |
| \( \alpha \)-stable | 20     | 20     | (0.05,0.85) | (0.37,0.63) | (0.02,0.95)    | (0.01,0.94) |
|              | 20     | 100    | (0.26,0.47) | (1.00,0.00) | (0.03,0.92)    | (0.04,0.85) |
|              | 100    | 100    | (0.06,0.93) | (0.95,0.05) | (0.07,0.87)    | (0.07,0.91) |
| skewed-\( t_3 \) | 20     | 20     | (0.08,0.83) | (0.61,0.38) | (0.03,0.91)    | (0.33,0.59) |
|              | 20     | 100    | (0.23,0.12) | (1.00,0.00) | (0.16,0.47)    | (0.75,0.03) |
|              | 100    | 100    | (0.07,0.06) | (1.00,0.00) | (0.00,0.37)    | (0.43,0.02) |

Table 3 reports the frequencies of exact estimation and underestimation with \( s_E = 1 \). Simulation results with \( s_E \in \{1.5, 2\} \) are reported in Table 10 and Table 11 in the supplementary materials. We set \( r_{\text{max}} = 8 \) for all the algorithms. It is observed that both MER\(_F\) and IterER benefit from the projection effect. In addition, MER\(_F\) is no worse than IterER for Gaussian noise, and outperforms IterER by a large margin for the heavy-tailed and skewed noises. As a result, MER\(_F\) can be used as a safe replacement of IterER in financial and econometrical applications.

### 6 Real Data Analysis

In this section, we apply the proposed algorithms on a financial portfolio dataset as in Wang et al. (2019), Yu et al. (2021). The dataset consists of monthly returns of 100 portfolios from January 1964 to December 2019, covering 672 months. The portfolios are constructed into \( 10 \times 10 \) data matrices, whose rows correspond to market capital size (S1-S10), and columns correspond to book-to-equity ratio (BE1-BE10). Detailed information could be found on the website http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

Following Wang et al. (2019) and Yu et al. (2021), we first subtract the corresponding monthly excess returns and impute the missing data by linear interpolation. The augmented Dickey-Fuller tests indicate stationarity of all time series. We apply eigenvalue-ratio algorithms on the full dataset to determine the pair of factor numbers \( p_0 \) and \( q_0 \), where MER\(_{op}\), \( (2D)^2 \)-ER suggest \( p_0 = q_0 = 1 \), while MER\(_F\), IterER
suggest \( p_0 = 1 \) and \( q_0 = 2 \). As the latter two projected algorithms are more stable under moderate noise shown in simulation study, we take \( p_0 = 1 \) and \( q_0 = 2 \) for further analysis.

Table 4: Factor loadings with \( p_0 = 1 \) and \( q_0 = 2 \) for Fama-French dataset after varimax rotation and scaling by 30. Here MPCA\(_F\) and MPCA\(_{op}\) stands for Manifold PCA methods; PE stands for the projected estimation by Yu et al. (2021); \((2D)^2\)-PCA is from Zhang and Zhou (2005), it is equivalent to \( \alpha \)-PCA by Chen and Fan (2021) with \( \alpha = -1 \).

| Size          | Method | Factor | S1 | S2 | S3 | S4 | S5 | S6 | S7 | S8 | S9 | S10 |
|---------------|--------|--------|----|----|----|----|----|----|----|----|----|-----|
| MPCA\(_{op}\) | 1      | -10    | -10| -11| -11| -11| -9 | -9 | -6 | 0  |     |     |
| MPCA\(_F\)   | 1      | -11    | -12| -11| -11| -11| -10| -9 | -8 | -5 | 0  |     |
| \((2D)^2\)-PCA| 1      | -10    | -11| -12| -11| -11| -10| -9 | -8 | -6 | 0  |     |
| PE           | 1      | -11    | -12| -11| -11| -10| -8 | -7 | -5 | 0  |     |     |

| Book-to Equity | Method | Factor | BE1 | BE2 | BE3 | BE4 | BE5 | BE6 | BE7 | BE8 | BE9 | BE10 |
|----------------|--------|--------|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|
| MPCA\(_{op}\) | 1      | 3      | 0   | -4  | -6  | -10 | -11 | -13 | -12 | -12 | -12 |      |
|               | 2      | 17     | 18  | 13  | 9   | 4   | 2   | -1  | -1  | -3  | -3  |      |
| MPCA\(_F\)   | 1      | 3      | 0   | -4  | -6  | -9  | -11 | -12 | -13 | -13 | -12 |      |
|               | 2      | 19     | 17  | 11  | 8   | 5   | 1   | -2  | -1  | -3  | -3  |      |
| \((2D)^2\)-PCA| 1      | 3      | -1  | -5  | -8  | -10 | -12 | -12 | -12 | -12 | -11 |      |
|               | 2      | 19     | 18  | 12  | 7   | 3   | 0   | -2  | -3  | -2  | -2  |      |
| PE           | 1      | 3      | -2  | -5  | -8  | -10 | -11 | -12 | -12 | -13 | -11 |      |
|               | 2      | 21     | 16  | 11  | 7   | 3   | 0   | -2  | -3  | -2  | -1  |      |

The estimated loading matrices after varimax rotation and scaling are reported in Table 4. It is observed that MPCA\(_{op}\), MPCA\(_F\), \((2D)^2\)-PCA and PE methods lead to similar estimated loadings. The small size portfolios load heavily on the front loading. The two factors in the back loading separate portfolios well from the perspective of book-to-equity, with large BE portfolios loading mainly on the first factor, and small BE portfolios loading mainly on the second.

Figure 2 shows the time series plots of the 100 series, while Figure 3 shows the estimated factors by MPCA\(_{op}\) and MPCA\(_F\) with \( p_0 = 1 \) and \( q_0 = 2 \), which show similar patterns and further indicate the estimated factors could potentially replace the original data matrices for further analysis.

To further compare these methods, we apply similar rolling-validation procedures as in Wang et al. (2019) and Yu et al. (2021). For each year \( t \) from 1996 to 2019, we take \( n \) (bandwidth) years before \( t \) as the training set, which is used to fit matrix factor models. The estimated loadings are then used to estimate factors and corresponding residuals of the testing set, consisting of the 12 months next year. Specifically, let \( Y_{it} \) and \( \hat{Y}_{it} \) be the observed and predicted price matrix of month \( i \) in year \( t \), we focus on the errors \( \text{MSE}_t \) and \( \text{opMax}_t \) defined as:

\[
\text{MSE}_t = \frac{1}{12 \times 10 \times 10} \sum_{i=1}^{12} \left\| Y_{it} - \hat{Y}_{it} \right\|_F^2, \quad \text{opMax}_t = \max_{1 \leq i \leq 12} \left\| Y_{it} - \hat{Y}_{it} \right\|_{op}.
\]

Table 5 reports the means and standard deviations of \( \text{MSE}_t \) and \( \text{opMax}_t \) by MPCA\(_{op}\), MPCA\(_F\), \((2D)^2\)-PCA and PE methods. The reported errors of different methods are very close, but MPCA\(_F\) performs slightly better under almost all bandwidths \( n \), in terms of both \( \text{MSE}_t \) and \( \text{opMax}_t \). Financial
Figure 2: Time series plots of Fama-French 10 by 10 series.

Figure 3: Plots of $\hat{F}_t$ estimated by MPCA$_{op}$ and MPCA$_F$ respectively after varimax rotation.
Table 5: Rolling validation with $p_0 = 1$ and $q_0 = 2$ for Fama-French dataset, the sample size of training set is $12n$. We report the means and standard deviations (in parentheses) of MSE$_t$ and opMax$_t$. Here MPCA$_F$ and MPCA$_{op}$ stands for Manifold PCA methods; PE stands for the projected estimation by Yu et al. (2021); $(2D)^2$-PCA is from Zhang and Zhou (2005), which is equivalent to $\alpha$-PCA by Chen and Fan (2021) with $\alpha = -1$.

| n  | MPCA$_{op}$       | MPCA$_F$       | $(2D)^2$-PCA | PE            |
|----|------------------|----------------|--------------|---------------|
| 5  | (0.7423,0.7372)  | (0.7405,0.7490)| (0.7410,0.7291)| (0.7378,0.7310) |
| 10 | (0.7430,0.7413)  | (0.7431,0.7544)| (0.7512,0.7377)| (0.7472,0.7431) |
| 15 | (0.7488,0.7568)  | (0.7456,0.7637)| (0.7524,0.7525)| (0.7470,0.7529) |
| 20 | (0.7452,0.7528)  | (0.7417,0.7578)| (0.7499,0.7534)| (0.7451,0.7534) |
| 25 | (0.7444,0.7501)  | (0.7407,0.7559)| (0.7492,0.7574)| (0.7450,0.7566) |

| n  | MPCA$_{op}$       | MPCA$_F$       | $(2D)^2$-PCA | PE            |
|----|------------------|----------------|--------------|---------------|
| 5  | (0.7599,0.4774)  | (0.7509,0.4714)| (0.7568,0.4764)| (0.7509,0.4708) |
| 10 | (0.7600,0.4784)  | (0.7482,0.4694)| (0.7683,0.4660)| (0.7538,0.4591) |
| 15 | (0.7597,0.4768)  | (0.7482,0.4654)| (0.7647,0.4655)| (0.7515,0.4564) |
| 20 | (0.7621,0.4839)  | (0.7497,0.4726)| (0.7666,0.4701)| (0.7528,0.4600) |
| 25 | (0.7627,0.4788)  | (0.7488,0.4682)| (0.7625,0.4626)| (0.7493,0.4557) |

Data is well-known to be heavy-tailed, and thus the more robust MPCA$_F$ is always preferred.

7 Conclusions and Discussions

Data in real world such as financial returns are well-known to be heavy-tailed, and robust factor modelling is indispensable as the traditional PCA estimation method would result in bigger biases and higher dispersions as the distribution tails become heavier (He et al., 2022, 2021b). In this article, we for the first time propose a flexible Matrix Elliptical Factor Model (MEFM) for better modelling heavy-tailed matrix-valued data, which can be viewed as an extension of the matrix factor model by Wang et al. (2019). We also propose robust Manifold Principle Component Analysis (MPCA) procedures to estimate the factor loading, scores, and common components matrices without any moment constraint under the framework of Matrix Elliptical Distributions (MED). We explore two versions of MPCA algorithms, denoted as MPCA$_F$ and MPCA$_{op}$, by considering the optimization problems in (2.2) under matrix operator norm and matrix frobenius norm respectively. Theoretical convergence rates of the estimators are derived for both versions. However, the MPCA$_F$ method is not only robust to heavy-tailed data, but also enjoys the nice property of the projection technique, thus performs the best in finite-sample experiments. In addition, we also proposed two robust versions to estimate the pair of factor numbers, by calculating eigenvalue-ratios (ER) of the average projection matrices corresponding to MPCA$_F$ and MPCA$_{op}$. We prove that the estimators of the pair of factor numbers are consistent. We conduct extensive numerical studies to validate the empirical performance of the proposed robust methods and an application to a Fama-French financial portfolios dataset illustrates the practical value of the current work. In the theoretical analysis of the MPCA$_F$, we assume that either R or C is given

18
to establish the convergence rate of $C$ or $R$, which is not quite satisfying. As a future work, we will establish the convergence rates of estimators from the iterative procedure, which is more challenging as both statistical error and computational error should be taken into account.

Acknowledgements

He’s work is supported by National Science Foundation (NSF) of China (12171282,11801316), National Statistical Scientific Research Key Project (2021LZ09), Young Scholars Program of Shandong University, Project funded by China Postdoctoral Science Foundation (2021M701997) and the Fundamental Research Funds of Shandong University. Kong’s work is partially supported by NSF China (71971118 and 11831008) and the WRJH-QNBJ Project and Qinglan Project of Jiangsu Province. Zhang’s work is supported by NSF China (11971116).

References

Ahn, S.C., Horenstein, A.R., 2013. Eigenvalue ratio test for the number of factors. Econometrica 81, 1203–1227.

Bai, J., 2003. Inferential theory for factor models of large dimensions. Econometrica 71, 135–171.

Bai, J., Li, K., 2012. Statistical analysis of factor models of high dimension. The Annals of Statistics 40, 436–465.

Bai, J., Li, K., 2016. Maximum likelihood estimation and inference for approximate factor models of high dimension. Review of Economics and Statistics 98, 298–309.

Bai, J., Ng, S., 2002. Determining the number of factors in approximate factor models. Econometrica 70, 191–221.

Chen, E.Y., Fan, J., 2021. Statistical inference for high-dimensional matrix-variate factor models. Journal of the American Statistical Association, 1–18.

Chen, L., Dolado, J.J., Gonzalo, J., 2021a. Quantile factor models. Econometrica 89, 875–910.

Chen, Y., Chi, Y., Fan, J., Ma, C., et al., 2021b. Spectral methods for data science: A statistical perspective. Foundations and Trends® in Machine Learning 14, 566–806.

Edelman, A., 1988. Eigenvalues and condition numbers of random matrices. SIAM journal on matrix analysis and applications 9, 543–560.

Fan, J., Liao, Y., Mincheva, M., 2013. Large covariance estimation by thresholding principal orthogonal complements. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 75, 603–680.

Fan, J., Liu, H., Wang, W., 2018. Large covariance estimation through elliptical factor models. The Annals of Statistics: An Official Journal of the Institute of Mathematical Statistics 46, 1383–1414.
Gupta, A., Varga, T., 1994. A new class of matrix variate elliptically contoured distributions. Journal of the Italian Statistical Society 3, 255–270.

Gupta, A.K., Nagar, D.K., 2018. Matrix variate distributions. volume 104. CRC Press.

Ham, J., Lee, D.D., 2008. Grassmann discriminant analysis: a unifying view on subspace-based learning, in: Machine Learning, Twenty-fifth International Conference, Helsinki, Finland, June.

He, Y., Kong, X., Trapani, L., Yu, L., 2021a. Vector factor model or matrix factor model? A strong rule helps! arXiv: arXiv:2110.01008.

He, Y., Kong, X., Yu, L., Zhang, P., 2020. Learning quantile factors for large-dimensional time series with statistical guarantee. arXiv:2006.08214.

He, Y., Kong, X., Yu, L., Zhang, X., 2022. Large-dimensional factor analysis without moment constraints. Journal of Business & Economic Statistics 40, 302–312.

He, Y., Kong, X., Yu, L., Zhang, X., Zhao, C., 2021b. Statistical inference for large-dimensional matrix factor model from least squares and huber loss points of view. arXiv:2112.04186.

He, Y., Kong, X.B., Trapani, L., Yu, L., 2021c. Online change-point detection for matrix-valued time series with latent two-way factor structure. arXiv:2112.13479.

Onatski, A., 2009. Testing hypotheses about the number of factors in large factor models. Econometrica 77, 1447–1479.

Ross, S.A., 1976. The arbitrage theory of capital asset pricing. Journal of Finance 13, 341–360.

Rudelson, M., Vershynin, R., 2010. Non-asymptotic theory of random matrices: extreme singular values, in: Proceedings of the International Congress of Mathematicians 2010 (ICM 2010) (In 4 Volumes) Vol. I: Plenary Lectures and Ceremonies Vols. II–IV: Invited Lectures, World Scientific. pp. 1576–1602.

Stock, J.H., Watson, M.W., 2002. Forecasting using principal components from a large number of predictors. Journal of the American statistical association 97, 1167–1179.

Trapani, L., 2018. A randomised sequential procedure to determine the number of factors. Journal of the American Statistical Association 113, 1341–1349.

Tropp, J.A., 2012. User-friendly tail bounds for sums of random matrices. Foundations of computational mathematics 12, 389–434.

Tropp, J.A., 2015. An introduction to matrix concentration inequalities. arXiv preprint arXiv:1501.01571.

Wang, D., Liu, X., Chen, R., 2019. Factor models for matrix-valued high-dimensional time series. Journal of econometrics 208, 231–248.
Yu, L., He, Y., Kong, X., Zhang, X., 2021. Projected estimation for large-dimensional matrix factor models. Journal of Econometrics, in press. doi:10.1016/j.jeconom.2021.04.001.

Yu, L., He, Y., Zhang, X., 2019. Robust factor number specification for large-dimensional elliptical factor model. Journal of Multivariate analysis 174, 104543.

Yu, Y., Wang, T., Samworth, R.J., 2015. A useful variant of the davis–kahan theorem for statisticians. Biometrika 102, 315–323.

Zhang, D., Zhou, Z.H., 2005. (2d) 2pca: Two-directional two-dimensional pca for efficient face representation and recognition. Neurocomputing 69, 224–231.
Supplementary Materials for “Manifold Principle Component Analysis for Large-Dimensional Matrix Elliptical Factor Model”

ZeYu Li * ‡, Yong He † ‡, Xinbing Kong ‡, Xinsheng Zhang *

This document provides detailed proofs and additional simulation results of the main paper.

A Proof of Lemma 3.1

It is a direct consequence of the matrix Hoeffding inequality from Tropp (2012), we only need to verify that $|P_{\hat{R}_t} - EP_{\hat{R}_t}|^2 \ll I_p$ almost surely. Since $0 \leq \langle v, P_{\hat{R}_t}v \rangle \leq 1$ almost surely and thus $0 \leq \langle v, EP_{\hat{R}_t}v \rangle \leq 1$ for all $\|v\|_2 = 1$, we have $\|P_{\hat{R}_t} - EP_{\hat{R}_t}\|_{op} \leq 1$ almost surely. The rest would be straightforward.

B Proof of Lemma 3.2

For joint matrix elliptical data $X_t = RF_tC^T + E_t$, since $E_t$ is left spherical, we could write $F_t = r_t\Sigma_1^{1/2}Z^T\Sigma_2^{1/2}/\|Z\|_2$ and $E_t = r_t\Sigma_1^{1/2}Z_t\Sigma_2^{1/2}/\|Z_t\|_2$ under model 2.1. Let $Z_t^E = (2P_R - I)Z_t^F$, then $\|Z_t\|_2^2 = \|Z_t^F\|_2^2 + \|Z_t^E\|_F^2 = \|Z_t^F\|_F^2 + \|Z_t^E\|_F^2 = \|Z_t^E\|_F^2$ almost surely, where $Z_t^E$ is defined as $(\text{Vec}(Z_t^F)^\top, \text{Vec}(Z_t^E)^\top)$. Due to rotational invariance of $Z_t^E$, $Z_t$ is identically distributed to $Z_t^E$, the latter generates $X_t' = RF_t'C^T + E_t' \overset{d}{=} X_t$. Now that $F_t' = r_t\Sigma_1^{1/2}Z_t^F\Sigma_2^{1/2}/\|Z_t\|_2 = F_t$ almost surely, while $E_t' = r_tZ_t^E\Sigma_2^{1/2}/\|Z_t\|_2 = (2P_R - I)E_t$ almost surely, we have $X_t' = (2P_R - I)X_t$ almost surely. Since $X_t$ and $X_t'$ are identically distributed, so do their best linear subspace estimations $\hat{R}_t$ and $\hat{R}_t'$, so that:

$$\mathbb{E}\left[P_{\hat{R}_t}\right] = \mathbb{E}\left[P_{\hat{R}_t} + P_{\hat{R}_t'}\right]/2.$$

From $X_t' = (2P_R - I)X_t$ we know that the column vectors of $X_t'$ and $X_t$ are symmetric in respect of $\text{Span}(R)$, so should their leading left singular vectors $\hat{R}_t'$ and $\hat{R}_t$, which gives $P_{\hat{R}_t'} = (2P_R - I)P_{\hat{R}_t}(2P_R - I)$. That is to say, $\forall u \in \text{Span}(R)$, $(P_{\hat{R}_t} + P_{\hat{R}_t'})u = 2P_RP_{\hat{R}_t}u \in \text{Span}(R)$. Similarly, $\forall v \in \text{Span}(R^\perp)$, $(P_{\hat{R}_t} + P_{\hat{R}_t'})v = 2(I - P_R)P_{\hat{R}_t}v \in \text{Span}(R^\perp)$. In the end, since $\mathbb{E}$ is linear, $\mathbb{E}\left[P_{\hat{R}_t} + P_{\hat{R}_t'}\right]u \in \text{Span}(R)$ and $\mathbb{E}\left[P_{\hat{R}_t} + P_{\hat{R}_t'}\right]v \in \text{Span}(R^\perp)$, we claim the proof.

C Proof of Theorem 3.3

Without loss of generality, we only prove $\text{Span}(R)$ here. For notation simplicity, here we let $\hat{R}$ to be the result from MPCA, instead of $\hat{R}_{op}$. Before the proof, recall a well-known fact that for $P_{\hat{R}} = \hat{R}\hat{R}^\top/p$ and $P_R = RR^\top/p$:

$$\|P_{\hat{R}} - P_R\|_F^2 \asymp \min_{H_R \in \mathcal{H}_{op,p}} \|\hat{R} - RH_R\|_F^2/p.$$
where the left hand side is known as the projection metric on Grassmann manifolds. So it is equivalent to study the term \( \| P_R - \hat{P}_R \|_F \) instead, see Chen et al. (2021b) for details.

### C.1 Spherical Neighbour

From lemma 3.2, it is ideal if \( E_t \) is left spherical when estimating \( \text{Span}(R) \), right spherical when estimating \( \text{Span}(C) \), and spherical when estimating both. Unfortunately, let \( \zeta = r_t / \| Z_t \|_2 \), the noise \( E_t = \zeta \Omega_1^{1/2} Z_r^T \Omega_2^{1/2} \) is elliptically transformed by \( \Omega_1 \) and \( \Omega_2 \). It is then natural to evaluate how much harm would deviating to non-spherical noise do. A spherical neighbour argument is applied where we construct a desirable \( \hat{E}_t \) sufficiently close to \( E_t \), controlling the difference by matrix perturbation results. Since we are currently dealing with \( \text{Span}(R) \), \( \hat{E}_t \) needs to be left spherical. Let \( \omega_1 = \arg \min \omega \| \Omega_1^{1/2} - \omega^{1/2} I_p \|_{op} \), define \( \hat{E}_t = \zeta \omega_1^{1/2} Z_r^T \Omega_2^{1/2} \). Denote \( \hat{X}_t = RF_t C^\top + \hat{E}_t, \hat{P}_R \) as the empirical projection matrix from each \( \hat{X}_t \), and \( \hat{P}_R \) as the MPCA\(_{op} \) result from \( \{ \hat{X}_t \} \), then by triangular inequality we have:

\[
\| P_R - \hat{P}_R \|_F \leq \| P_R - \hat{P}_R \|_F + \| \hat{P}_R - \hat{P}_R \|_F. \tag{C.1}
\]

We first focus on the term \( \| P_R - \hat{P}_R \|_F \), which comes from noise \( E_t \) being non-spherical. Define \( d_i(A) = (\lambda_i - \lambda_{i+1})(A) \) as the \( i \)-th eigengap of matrix \( A \), where \( \lambda_i \) is the \( i \)-th non-increasing eigenvalue of \( A \). Consider the matrix perturbation \( \sum_{t=1}^T P_{\hat{R}_t}/T = \sum_{t=1}^T \hat{P}_{\hat{R}_t}/T + \sum_{t=1}^T (P_{\hat{R}_t} - \hat{P}_{\hat{R}_t})/T \), by YWS’s inequality from Yu et al. (2015) and Jensen’s inequality:

\[
\| P_R - \hat{P}_R \|_F \leq \frac{2 \sqrt{2} \| \sum_{t=1}^T (P_{\hat{R}_t} - \hat{P}_{\hat{R}_t})/T \|_F}{d_{p_0}(\sum_{t=1}^T \hat{P}_{\hat{R}_t}/T)} \leq \frac{2 \sqrt{2} \sum_{t=1}^T \| P_{\hat{R}_t} - \hat{P}_{\hat{R}_t} \|_F}{T d_{p_0}(\sum_{t=1}^T \hat{P}_{\hat{R}_t}/T)}. \tag{C.2}
\]

We are going to show that \( d_{p_0}(\mathbb{E} \hat{P}_{\hat{R}_t}) > 0 \) and \( \| \sum_{t=1}^T \hat{P}_{\hat{R}_t}/T - \mathbb{E} \hat{P}_{\hat{R}_t} \|_{op} \to 0 \) as \( T, p, q \to \infty \) in later analysis. By Weyl’s inequality the eigengap \( d_{p_0}(\sum_{t=1}^T \hat{P}_{\hat{R}_t}/T) \to d_{p_0}(\mathbb{E} \hat{P}_{\hat{R}_t}) > 0 \). So we focus on the term \( \sum_{t=1}^T \| P_{\hat{R}_t} - \hat{P}_{\hat{R}_t} \|_F/T \). Consider the perturbation \( X_t = \hat{X}_t + (E_t - \hat{E}_t) \), by Wedin’s theorem and Weyl’s inequality:

\[
\| P_{\hat{R}_t} - \hat{P}_{\hat{R}_t} \|_F \leq \frac{2 \sqrt{r_0} \| E_t - \hat{E}_t \|_{op}}{\sigma_{t,r_0} - 2 \| E_t - \hat{E}_t \|_{op}} \wedge \sqrt{2}r_0, \tag{C.3}
\]

where \( \sigma_{t,r_0} \) is the \( r_0 \)-th singular value of the signal part \( S_t = RF_t C^\top \), and by sub-multiplicivity of operator norm:

\[
\| E_t - \hat{E}_t \|_{op} = \| (\Omega_1^{1/2} - \omega_1^{1/2} I_p) \hat{E}_t \|_{op} \leq \frac{\| \Omega_1^{1/2} - \omega_1^{1/2} I_p \|_{op} \| \hat{E}_t \|_{op}}{\omega_1^{1/2}}. \tag{C.4}
\]

Then under Assumption A to C, we have

\[
\mathbb{E} \| P_{\hat{R}_t} - \hat{P}_{\hat{R}_t} \|_F = O(p^{-1/4} + q^{-1/4}). \tag{C.5}
\]

Intuitively, under strong factor model, \( \sigma_{t,r_0} \gtrless (pq)^{1/2} \zeta_t \sigma_{r_0}(Z_t^F) \) and \( \| \hat{E}_t \|_{op} \lesssim (p \vee q)^{1/2} \zeta_t \), so that \( \mathbb{E} \| P_{\hat{R}_t} - \hat{P}_{\hat{R}_t} \|_F \) tends to zero as \( p, q \to \infty \) from C.3 and C.4. To be more specific, here without loss of generality assume \( p_0 = r_0 \leq q_0 \), consider the matrix \( S_t S_t^\top = q RF_t F_t^\top R_t^\top \), whose \( r_0 \)-th eigenvalue
that is to say, the influence of $E_{\text{Span}} \sigma \dot{\sigma}$ so the convergence of scalar concentration to $C$.2, $C$.5 and:

$$\| \dot{E}_t \|_{op} \preceq \zeta_t \| Z^F \|_{op} \text{ is straightforward by sub-multiplicativity of operator norm.}$$

Consider the ratio in C.3, under joint matrix elliptical model in Assumption A, the shared $\zeta_t$ is cancelled out, we only need to focus on the expectation of $(pq)^{-1/2} \| Z^F \|_{op} / \sigma_{r_0}(Z^F)$, whose numerator and denominator are independent.

Non-asymptotic random matrix theory asserts that $\sigma_{r_0}(Z^F) \asymp \sqrt{r_0} - \sqrt{\rho_0}$, so $\mathbb{E}\| P_{R_t} - \hat{P}_{R_t} \|_F = O(p^{-1/2} + q^{-1/2})$ unless $p_0 = q_0 = r_0$, only then $\sigma_{r_0}(Z^F)$ has a larger probability towards 0, leading to invertibility problems. In this case, we could use the fact from Edelman (1988) that:

$$\mathbb{P}\left( \sigma_{r_0}(Z^F) \leq \varepsilon r_0^{-1/2} \right) \leq \varepsilon, \quad \forall \varepsilon \geq 0,$$

see Rudelson and Vershynin (2010) for details on extreme singular values of random matrices. If $p \geq q$, let $\varepsilon = q^{-1/4}$, and $\mathbb{P}\left( \sigma_{r_0}(Z^F) \leq q^{-1/4} \right) \lesssim q^{-1/4}$. Then we have:

$$\mathbb{E}\| P_{R_t} - \hat{P}_{R_t} \|_F = \mathbb{E}\left( \| P_{R_t} - \hat{P}_{R_t} \|_F I_{\sigma_{r_0}(Z^F) \leq q^{-1/4}} \right) + \mathbb{E}\left( \| P_{R_t} - \hat{P}_{R_t} \|_F I_{\sigma_{r_0}(Z^F) > q^{-1/4}} \right) \lesssim \sqrt{2r_0} \mathbb{P}\left( \sigma_{r_0}(Z^F) \leq q^{-1/4} \right) + q^{-1/4} = O(q^{-1/4}),$$

where $\mathbb{E}(\| P_{R_t} - \hat{P}_{R_t} \|_F I_{\sigma_{r_0}(Z^F) > q^{-1/4}}) \lesssim q^{-1/4}$ comes from the fact that when $\sigma_{r_0}(Z^F) > q^{-1/4}$, we have $\sigma_{t,r_0} \geq c_1 \gamma t^{1/2} q^{1/4}$, and $\| P_{R_t} - \hat{P}_{R_t} \|_F \lesssim \| \dot{E}_t \|_{op} / \sigma_{t,r_0} \lesssim q^{-1/4}$ from C.3. Similarly, if $p < q$, $\mathbb{E}\| P_{R_t} - \hat{P}_{R_t} \|_F = O(p^{-1/4})$. In the end, since $\| P_{R_t} - \hat{P}_{R_t} \|_F \leq \sqrt{2r_0}$ almost surely, apply scalar concentration to C.2, C.5 and:

$$\| P_{R_t} - \hat{P}_{R_t} \|_F \lesssim \frac{\sum_{t=1}^{T} \| P_{R_t} - \hat{P}_{R_t} \|_F}{T} = O_p(T^{-1/2} + p^{-1/4} + q^{-1/4}),$$

that is to say, the influence of $E_t$ being non-spherical is ignorable as $T, p, q \to \infty$.

We then focus on the second term $\| \hat{P}_R - P_R \|_F$ in C.1, the convergence of MPCA$_{op}$ under spherical noise $\hat{E}_t$. With slight abuse of notation but no loss of generality, the model could be reset as $X_t = RFtC^T + E_t$, where $\Omega_1 = \omega_1 I_p$ and $E_t$ is left spherical. The second term is then $\| P_R - \hat{P}_R \|_F$. Since $\text{Span}(R)$ is the invariant subspace of $\mathbb{E}P_{R_t}$ according to lemma 3.2, consider the matrix perturbation $\sum_{t=1}^{T} P_{R_t}/T = \mathbb{E}P_{\hat{R}_t} + (\sum_{t=1}^{T} P_{R_t}/T - \mathbb{E}P_{R_t})$, by YWS’s inequality we have:

$$\| P_R - \hat{P}_R \|_F \leq \frac{2\sqrt{2p_0} \sum_{t=1}^{T} P_{R_t}/T - \mathbb{E}P_{R_t}}{d_{op}(\mathbb{E}P_{R_t})} \wedge \sqrt{2p_0},$$

so the convergence of $\| P_R - \hat{P}_R \|_F$ naturally depends on the expected projection matrix $\mathbb{E}P_{R_t}$, the denominator, and on matrix concentration, the numerator.
C.2 Non-degenerated Case

We first show convergence of \( \|P_R - P_{\hat{R}}\|_F \) in the non-degenerated case, where \( p_0 = r_0 \leq q_0 \). After spherical neighbour arguments, the model is reset as \( X_t = RF_tC^\top + E_t \) with \( E_t \) left spherical.

C.2.1 Expected Projection Matrix

We first take a look at \( \mathbb{E}P_{R_t} \), by lemma 3.2, \( \mathbb{E}P_{\hat{R}_t} \) could be decomposed into two separate parts:

\[
\mathbb{E}P_{R_t} = \frac{p}{i=1} \lambda_i u_i u_i^\top = \sum_{i=1}^{r_0} \lambda_i u_i u_i^\top + \sum_{j=r_0+1}^{p} \lambda_j u_j u_j^\top,
\]

where \( u_i, i \in \{1, \ldots, r_0\} \), generate \( \text{Span}(R) \) while \( u_j, j \in \{r_0 + 1, \ldots, p\} \), generate \( \text{Span}(R^\perp) \).

Lemma C.1 (Subspace Variance). For \( E_t \) left spherical, let \( \{\theta_i\} \) be the principal angles between \( \text{Span}(\hat{R}_t) \) and \( \text{Span}(R) \), the following equality holds:

\[
\text{tr}(N) = \mathbb{E}\left( \sum_{i=1}^{r_0} \sin^2 \theta_i \right) = \mathbb{E}\|P_{\hat{R}_t} - P_R\|_F^2/2.
\]

Proof. Now that \( \mathbb{E}P_{\hat{R}_t} \) and \( P_R \) share eigenspace, It is easy to see that:

\[
\text{tr}(N) = \text{tr}\left( \mathbb{E}P_{\hat{R}_t}(I - P_R) \right) = r_0 - \mathbb{E}\text{tr}(P_{\hat{R}_t}P_R).
\]

Since \( \text{tr}(P_{\hat{R}_t}P_R) = \sum_{i=1}^{r_0} \cos^2 \theta_i \) by definition of principal angles, we have acquired the proof. \( \square \)

Lemma C.2 (Subspace Deviation). For matrix model \( X_t = RF_tC^\top + E_t \), denote \( \sigma_{t,r_0} \) as the \( r_0 \)-th singular value of the signal part \( RF_tC^\top \), we have:

\[
\|P_{\hat{R}_t} - P_R\|_F/\sqrt{2} \leq \frac{\sqrt{2r_0}\|E_t\|_{op}}{\sigma_{t,r_0} - \|E_t\|_{op}} \wedge \sqrt{r_0}.
\]

Proof. It is a straightforward corollary of Wedin’s theorem for perturbation \( X_t = RF_tC^\top + E_t \). \( \square \)

Since \( \text{tr}(\mathbb{E}P_{\hat{R}_t}) = \mathbb{E}\text{tr}(P_{\hat{R}_t}) = r_0 \), from lemma C.1, C.2 and Assumption A to C, by truncation method as in C.6, we have \( \mathbb{E}\|P_{\hat{R}_t} - P_R\|_F^2 = O(p^{-1/3} + q^{-1/3}) \), which means \( \text{tr}(S) = \sum_{i=1}^{r_0} \lambda_i \to r_0 \) and \( \text{tr}(N) = \sum_{j=r_0+1}^{p} \lambda_j \to 0 \) as \( p, q \to \infty \). Since \( 0 \leq \lambda_i \leq 1 \) for \( i \in \{1, 2, \ldots, p\} \), we have \( \lambda_{r_0} \to 1 \), \( \lambda_{r_0+1} \to 0 \) and naturally \( d_{r_0}(\mathbb{E}P_{\hat{R}_t}) \to 1 \) as \( p, q \to \infty \).

C.2.2 Matrix Concentration

We then turn to the matrix concentration problem on the numerator of C.8. Taking direct advantage of lemma 3.1 would give \( \|P_{\hat{R}_t} - P_R\|_F = O_p(\sqrt{\log p/T}) \). This dimensional factor is inherent in existing matrix concentration results. Alternatively, by triangular inequality and taking expectation on both sides:

\[
\mathbb{E}\|P_{\hat{R}_t} - \mathbb{E}P_{\hat{R}_t}\|_{op} \leq \mathbb{E}\|P_{\hat{R}_t} - P_R\|_{op} + \|P_R - \mathbb{E}P_{\hat{R}_t}\|_{op},
\]
where \( \|P_R - \mathbb{E}P_{R_i}\|_{op} \leq \text{tr}(N) = O(p^{-1/3} + q^{-1/3}) \), while \( \mathbb{E}\|P_{R_i} - P_R\|_{op} \leq \mathbb{E}\|P_{R_i} - P_R\|_F = O(p^{-1/4} + q^{-1/4}) \) by applying truncation method as in C.6 to lemma C.2. In the end, since \( \|P_{R_i} - \mathbb{E}P_{R_i}\|_{op} \leq 1 \) almost surely, by applying Jensen’s inequality and scalar concentration to C.8:

\[
\|P_R - P_R\|_F \leq \frac{2\sqrt{T_0} \sum_{i=1}^T \|P_{R_i} - \mathbb{E}P_{R_i}\|_{op}}{Td_{r_0}(\mathbb{E}P_{R_i})} = O_p(T^{-1/2} + p^{-1/4} + q^{-1/4}),
\]

which could be absorbed into the non-spherical deviation term from the previous section.

### C.3 Degenerated Case

Then we discuss the degenerated case where \( p_0 > q_0 = r_0 \). For each data matrix \( X_t = RF_tC^\top + E_t \), the signal part \( RF_tC^\top \) is at most rank \( r_0 \). It is then natural to set \( \hat{R}_t \) to be the leading \( r_0 \) left singular vectors of \( X_t \). Then, MP\( \text{CA}_{op} \) calculates the \( p_0 \) leading eigenvectors of average projection matrices \( \sum_t P_{R_i}/T \), denoted as \( \hat{R}/\sqrt{p} \). As we discussed earlier, the manifold center intuition no longer holds under degeneration. Fortunately, the previous arguments on non-degenerated cases could be transferred readily to degenerated ones with only slight adjustments.

#### C.3.1 Expected Projection Matrix

By lemma 3.2, \( \mathbb{E}P_{R_i} \) could still be decomposed into two separate parts:

\[
\mathbb{E}P_{R_i} = \sum_{i=1}^p \lambda_i u_i u_i^\top = \sum_{i=1}^{p_0} \lambda_i u_i u_i^\top + \sum_{j=p_0+1}^p \lambda_j u_j u_j^\top, \tag{C.9}
\]

where \( u_i, i \in \{1, \ldots, p_0\} \), generate \( \text{Span}(R) \) while \( u_j, j \in \{p_0 + 1, \ldots, p\} \), generate \( \text{Span}(R^\perp) \). As in the non-degenerated case, \( \text{tr}(S) = \mathbb{E}\text{tr}(P_{R_i} P_{R_i}) \) and \( \text{tr}(N) = r_0 - \mathbb{E}\text{tr}(P_{R_i} P_{R_i}) \). Take a further look at \( \hat{R}_t \), the leading \( r_0 \) left singular vectors of \( X_t = RF_tC^\top + E_t \). It is actually the perturbation of \( R_t \), the \( r_0 \) left singular vectors of the signal part \( RF_tC^\top \). Actually, \( \text{Span}(R_t) \) is a \( r_0 \)-dimensional random subspace of \( \text{Span}(R) \), the randomness comes from signal \( F_t \). Then \( P_R \) could be decomposed into two parts as \( P_R = P_{R_i} + P_{R_i}^\perp \), the second term corresponds to the \( (p_0 - r_0) \)-dimensional subspace left. We have:

\[
\text{tr}(S) = \mathbb{E}\text{tr}(P_{R_i} P_{R_i}) + \mathbb{E}\text{tr}(P_{R_i} P_{R_i}^\perp) \geq \mathbb{E}\text{tr}(P_{R_i} P_{R_i}),
\]

where \( \mathbb{E}\text{tr}(P_{R_i} P_{R_i}) = r_0 - \mathbb{E}\|P_{R_i} - P_{R_i}\|_F^2/2 \) and \( \mathbb{E}\|P_{R_i} - P_{R_i}\|_F^2/2 \) could be viewed as subspace variance and controlled by arguments as in lemma C.2. Since \( \text{tr}(S) \geq \text{tr}(\mathbb{E}P_{R_i}) \), we have \( \text{tr}(S) = \sum_{i=1}^{p_0} \lambda_i \rightarrow r_0 \) and \( \text{tr}(N) = \sum_{j=p_0+1}^p \lambda_j \rightarrow 0 \) as \( p, q \rightarrow \infty \). The problem is that now \( \text{tr}(S) \rightarrow r_0 \) is distributed within \( p_0 \) eigenvalues, which gives chance for really small eigenvalue \( \lambda_{p_0} \), while the eigengap \( d_{p_0}(\mathbb{E}P_{R_i}) = \lambda_{p_0} - \lambda_{p_0+1} \) needs to be positive to ensure convergence. Fortunately, positive eigengap is justified by the following lemma.

**Lemma C.3 (Positive Eigengap).** Under Assumption A to C, there exists \( c > 0 \) free of \( p \) and \( q \) such that the eigengap \( d_{p_0}(\mathbb{E}P_{R_i}) = \lambda_{p_0} - \lambda_{p_0+1} \geq c \) as \( p, q \rightarrow \infty \).
Proof. First we prove there exists $c > 0$ such that $d_{p_0} (\mathbb{E} P_{R_t}) > c$, where $R_t$ would be the $r_0$ leading left singular vectors of the signal part $RF_t C^\top = \zeta_t R \Sigma_1^{1/2} Z_t \Sigma_2^{1/2} C^\top$. By stochastic representation of matrix spherical distributions given in Gupta and Nagar (2018), we could decompose $Z_t^F$ into three independent parts, namely $Z_t^F = U_t D_t V_t^\top$. Here $U_t$ of shape $p_0 \times q_0$ and $V_t$ of shape $q_0 \times q_0$ are uniformly distributed orthonormal matrices, while $D_t$ is diagonal. Clearly, Span$(R_t) = \text{Span}(R \Sigma_1^{1/2} U_t)$ and:

$$P_{R_t} = (R \Sigma_1^{1/2} U_t) \left( U_t \Sigma_1 U_t \right)^{-1} \left( U_t \Sigma_1^{1/2} R^\top \right)/p. \quad (C.10)$$

Here $U_t \Sigma_1 U_t$ is invertible since it is a $q_0 \times q_0$ sub-matrix of a $p_0 \times p_0$ positive definite matrix $\hat{U}_t \Sigma_1 \hat{U}_t$, where $\hat{U}_t$ is acquired by filling $U_t$ to $p_0 \times p_0$ orthonormal. In fact, the smallest eigenvalue of $(U_t \Sigma_1 U_t)^{-1}$ is larger than $1/C_1$ almost surely. Since Span$(R_t)$ is random subspace of Span$(R)$, the $(p_0 + 1)$-th eigenvalue of $\mathbb{E} P_{R_t}$ is clearly 0, we only need to show that $\langle u, \mathbb{E} P_{R_t} u \rangle > c$ for all $u \in \text{Span}(R)$, $\|u\|_2 = 1$. For $u \in \text{Span}(R)$, $\|u\|_2 = 1$, it is not hard to verify that $R^\top u/\sqrt{p}$ would be a unit vector in $\mathbb{R}^{p_0}$, so that $\|\Sigma_1^{1/2} R^\top u/\sqrt{p}\|_2 \geq \sqrt{c_1}$ under Assumption B. For unit vector $v = (\Sigma_1^{1/2} R^\top u)/\|\Sigma_1^{1/2} R^\top u\|_2$, from C.10 we have $\langle u, \mathbb{E} P_{R_t} u \rangle = c_1 \mathbb{E} \langle U_t^\top v, (U_t \Sigma_1 U_t)^{-1} U_t^\top v \rangle$ almost surely, and the right hand side is free of $p$ and $q$. Now that $\lambda_{p_0} \left( (U_t \Sigma_1 U_t)^{-1} \right) \geq 1/C_1$ almost surely, if we assume that $\mathbb{E} \langle U_t^\top v, (U_t \Sigma_1 U_t)^{-1} U_t^\top v \rangle = 0$, we should have $U_t^\top v = 0$ almost surely, which leads to contradiction.

In the end since $\|\mathbb{E} P_{R_t} - \mathbb{E} P_{R_t}\|_{op} \leq \|P_{R_t} - P_{R_t}\|_{op}$ by Jensen’s inequality, while the latter goes to 0 when $p, q \to \infty$ as we discussed earlier, apply Weyl’s inequality to acquire the proof. 

\[ \square \]

C.3.2 Matrix Concentration

We then turn to the matrix concentration problem on the numerator of C.8. Again taking direct advantage of lemma 3.1 would give $\|P_{R} - P_{R}\|_F = O_p(\sqrt{\log p/T})$. Fortunately, in this case of random projection matrices, we are able to shrink the dimensional factor $p$ to $r_0$ via intrinsic dimension arguments. According to matrix Bernstein inequality in Hermitian case with intrinsic dimension from Tropp (2015), we only need to focus on the independent centered term $P_t = P_{R_t} - \mathbb{E} P_{R_t}$. First, $\|P_t\|_{op} \leq 1$ almost surely. Then, consider the matrix variance $\mathbb{E} P^2_t = \mathbb{E} P_{R_t} - (\mathbb{E} P_{R_t})^2$. Under left spherical $E_t$, the eigenvalues of $\mathbb{E} P^2_t$ are precisely $\lambda_i - \lambda^2_j$, $i \in \{1, 2, \ldots, p_0\}$, and $\lambda_j - \lambda^2_j \to 0$ for $j \in \{p_0 + 1, p_0 + 2, \ldots, p\}$, where $\lambda_i$ and $\lambda_j$ are eigenvalues of $S$ and $N$ from C.9 respectively. Since there exists $c > 0$ such that $\lambda_{p_0} \geq c$ according to lemma C.3, while $\lambda_{p_0} \leq r_0/p_0$ automatically, there exists $C > 0$ free of $p$ and $q$ such that $C \leq \|\mathbb{E} P^2_t\|_{op} \leq 1/4$. So the intrinsic dimension of $\mathbb{E} P^2_t$, namely $\text{intdim}(\mathbb{E} P^2_t) = \text{tr}(\mathbb{E} P^2_t)/\|\mathbb{E} P^2_t\|_{op} \leq r_0$ does not grow with matrix dimension, and we could replace $p$ in lemma 3.1 with $r_0$ as:

$$\mathbb{P} \left\{ \left\| \sum_{t=1}^T (P_{R_t} - \mathbb{E} P_{R_t}) \right\|_{op} \geq x \right\} \lesssim r_0 \cdot \exp \left\{ \frac{-x^2/2}{T \|\mathbb{E} P^2_t\|_{op} + x/3} \right\}. $$

By applying dimension-free convergence, $\|P_{R} - P_{R}\|_F = O_p(T^{-1/2})$. In the end, take the deviation from noise being non-spherical into account, we claim the proof.
D  Proof of Theorem 3.4

We only discuss $\text{Span}(R)$ here due to symmetry. Recall that for MPCA$_F$, $\hat{R}_i^{(i)}$ would be the leading $r_0 = p_0 \wedge q_0$ eigenvalues of $X_tP_{C(i-1)}X_t^T$ if $\hat{C}^{(i-1)}$ is given. Let $C^{(i-1)} = \hat{C}^{(i-1)}/\sqrt{q_i}$ for notational simplicity. We focus on the projected matrix model $X_tC^{(i-1)} = RF_tC^\top C^{(i-1)} + E_tC^{(i-1)}$, and $\hat{R}_i^{(i)}$ would exactly be the result of applying MPCA$_{op}$ to the projected data set $\{X_tC^{(i-1)}\}$. It is worth mentioning that multiplying $C^{(i-1)}$ on the right does not affect the left properties we need in section C: for instance, if $E_t$ is left spherical, then $E_tC^{(i-1)}$ is still left spherical, so the proof from section C could adjust to the projected data set readily.

The difference lies in the signal-to-noise ratio. In section C, let $\sigma_{t,r_0}$ be the $r_0$-th singular value of the signal part $S_t = RF_tC^\top$, then $\sigma_{t,r_0} \geq (pq)^{1/2} \zeta_t \sigma_{r_0}(Z_t^F)$ and $\|E_t\|_{op} \lesssim \zeta_t \|Z_t^F\|_{op}$. It is foreseeable that we could increase the signal-to-noise ratio via projection by some $\hat{C}^{(i-1)}$ sufficiently close to $C$, keeping the signal size almost unchanged. It is ensured by $\sigma_{r_0}(C^\top \hat{C}^{(i-1)})/q = c > 0$.

In essence, let $\sigma_{r_0}^{(i-1)}$ be the $r_0$-th singular value of the projected signal part $S_tC^{(i-1)} = RF_tC^\top C^{(i-1)}$. If $p_0 = r_0 \leq q_0$, consider the matrix $S_tP_{C(i-1)}S_t^\top = RF_tC^\top P_{C(i-1)}F_tC^\top R_tC^\top$, whose $r_0$-th eigenvalue is $(\sigma_{r_0}^{(i-1)})^2$. Let $v = Ru/\sqrt{p}$, $\|v\|_2 = 1$ be the unit vector of $r_0$-dimensional subspace $\text{Span}(R)$, then $\langle v, S_tP_{C(i-1)}S_t^\top v \rangle = \langle p(u, F_tC^\top P_{C(i-1)}F_tC^\top u) \rangle \geq p\sigma_{r_0}^2(F_tC^\top C^{(i-1)}) \geq pq(c\zeta_t\sigma_{r_0}(Z_t^F))$, $\forall v \in \text{Span}(R)$, the last inequality comes from $\|\Sigma_{C}^{1/2} Z_t^F \Sigma_{C}^{1/2} C^{(i-1)} w \|_2 \geq q^{1/2}c_1\zeta_c\sigma_{r_0}(Z_t^F)$ for all $\|w\|_2 = 1$. In the end, by Courant–Fischer’s minimax theorem, $\sigma_{r_0}^{(i-1)} \geq (pq)^{1/2}c_1\zeta_c\sigma_{r_0}(Z_t^F)$. If $p_0 > q_0 = r_0$, then similarly, since $\|(C^{(i-1)})^\top \Sigma_{C}^{1/2} \Sigma_{C}^{1/2} (Z_t^F) (Z_t^F)^\top R^\top - \|w\|_2 \geq (pq)^{1/2}c_1\zeta_c\sigma_{r_0}(Z_t^F)$ for all $w \in \text{Span}(R)$ and $\|w\|_2 = 1$, we still have $\sigma_{r_0}^{(i-1)} \geq (pq)^{1/2}c_1\zeta_c\sigma_{r_0}(Z_t^F)$.

As for the noise part, let $E_t = \zeta_t \Omega^{1/2} Z_t^F \Omega^{1/2}$ with $\zeta_t = r_t/\|Z_t\|_2$, we need to prove that:

$$\|E_t C^{(i-1)}\|_{op} = \sup_{\|v\|_2 = 1} \|E_t C^{(i-1)} v\|_2 \lesssim \zeta_t p^{1/2}. $$

First, $u = \Omega^{1/2} C^{(i-1)} v$ spans a $q_0$-dimensional subspace of the $q$-dimensional space, with $c_2^{1/2} \leq \|u\|_2 \leq C_2^{1/2}$ under Assumption C. For $E_t C^{(i-1)} v = r_t \Omega^{1/2} Z_t^F u/\|Z_t\|_2$, since $Z_t^F$ is rotation invariant while $\|Z_t\|_2$ remains unchanged under rotation as shown in section B, there is no loss of generality if we rotate $\text{Span}(u)$ to be $\text{Span}(e_1, e_2, \ldots, e_{q_0})$, where $\{e_1, e_2, \ldots, e_{q_0}\}$ are the first $q_0$ Euclidean basis vectors. It is equivalent to say that only the first $q_0$ elements in vector $u$ can be non-zero. That is to say, we should only take the first $q_0$ columns of $Z_t^F$ into account when maximizing $\zeta_t \Omega^{1/2} Z_t^F u$, which is a $p \times q_0$ random matrix with i.i.d. standard Gaussian elements, and directly $\|E_t C^{(i-1)}\|_{op} \lesssim \zeta_t p^{1/2}$.

The proof is then identical to section C, since projection effects the signal-to-noise ratio, the convergence rate for $\hat{C}^{(i)}$ given $\hat{C}^{(i-1)}$ would be $O_p(T^{-1} + q^{-1/2}).$

E  Proof of Corollary 3.1

For $\hat{F}_t = \hat{R}^\top X_t \hat{C}/(pq)$, plug in $X_t = RF_tC^\top + E_t$ and get:

$$\hat{F}_t = \hat{R}^\top R F_t C^\top \hat{C}/(pq) + \hat{R}^\top E_t \hat{C}/(pq). $$

(E.1)

Recall that $\zeta_t = r_t/\|Z_t\|_2 = O_p(1)$ under Assumption B, let $\varepsilon_R = \hat{R} - RH_R$ and $\varepsilon_C = \hat{C} - HC_C$,
for the latter term we have:

\[
\tilde{R}^T E_t \tilde{C} = (R \epsilon_R + \epsilon_C)^T E_t (C \epsilon_C + \epsilon_C)
\]

\[
= H_R^T R^T E_t C \epsilon_C + H_R^T R^T E_t \epsilon_C + \epsilon_C^T E_t C \epsilon_C + \epsilon_C^T E_t \epsilon_C,
\]

(E.2)

where \(\|H_R^T R^T E_t C \epsilon_C\|_op = O_p((pq)^{1/2})\) by similar projection arguments as in section D, it is consistent with results in Yu et al. (2021). Take \(\|\cdot\|_op\) on both sides of E.2, since \(E_t = O_p((pq)^{1/2})\), \(\|\epsilon_C^T \sqrt{p}\|_op = o_p(1)\), \(\|\epsilon_C^T \sqrt{q}\|_op = o_p(1)\), by sub-multiplicativity of operator norm and triangular inequality we have:

\[
\|\tilde{R}^T E_t \tilde{C}\|_op/(pq) = O_p(\|\epsilon_C^T \sqrt{p}\|_op + \|\epsilon_C^T \sqrt{q}\|_op + (pq)^{-1/2}).
\]

(E.3)

As for the former term on the right hand side of E.1, we have:

\[
\tilde{R}^T R F_t C^T \tilde{C} = (R \epsilon_R + \epsilon_C)^T R F_t C^T (C \epsilon_C + \epsilon_C)
\]

\[
= q \epsilon_R H_R^T F_t H + p \epsilon_R H_R^T F_t C^T \epsilon_C + q \epsilon_R^T R F_t H \epsilon_C + \epsilon_C^T R F_t C^T \epsilon_C.
\]

(E.4)

By arranging E.1, E.3, E.4 and taking \(\|\cdot\|_op\) on both sides, due to sub-multiplicativity of operator norm and triangular inequality, we have:

\[
\|\tilde{R} - H_R^T F_t H\|_op = O_p(\|\epsilon_C^T \sqrt{p}\|_op + \|\epsilon_C^T \sqrt{q}\|_op + (pq)^{-1/2}).
\]

Similarly, let \(\tilde{S}_t - S_t = \tilde{R} \tilde{F}_t \tilde{C} - R F_t C^T\), we have:

\[
\tilde{S}_t - S_t = \tilde{R} \tilde{F}_t \tilde{C} - R F_t C^T
\]

\[
= (R \epsilon_R) (H_R^T F_t H \epsilon_C + \epsilon_C)^T - R F_t C^T
\]

\[
= R F_t H \epsilon_C^T + R H_R \epsilon_C^T F_t H \epsilon_C + R H_R \epsilon_C^T F_t C^T + \epsilon_C^T F_t H \epsilon_C + \epsilon_C^T F_t C^T + \epsilon_C^T F_t \epsilon_C.
\]

Again, take \(\|\cdot\|_op\) on both sides, by sub-multiplicativity of operator norm and triangular inequality:

\[
\|\tilde{S}_t - S_t\|_op/\sqrt{pq} = O_p(\|\epsilon_C^T \sqrt{p}\|_op + \|\epsilon_C^T \sqrt{q}\|_op + (pq)^{-1/2}).
\]

**F Proof of Theorem 4.2**

Here we only prove \(\tilde{p}_0\) due to symmetry. Now that:

\[
P(\tilde{p}_0 = p_0) \geq P(\tilde{p}_0 = p_0, \tilde{r}_0 = r_0) = P(\tilde{r}_0 = r_0) P(\tilde{p}_0 = p_0 | \tilde{r}_0 = r_0),
\]

it suffices to prove that \(P(\tilde{r}_0 = r_0) \rightarrow 1\) and \(P(\tilde{p}_0 = p_0 | \tilde{r}_0 = r_0) \rightarrow 1\). As for the first part, \(P(\hat{r}_{0,t} = r_0) \rightarrow 1\) under Assumption A to C, since \(\hat{r}_{0,t}\) is determined by each data matrix \(X_t = RF_t C^T + E_t\) and the signal part goes to infinity faster than noise. So directly \(P(\tilde{r}_0 = r_0) \rightarrow 1\).

As for the second part, under condition \(\tilde{r}_0 = r_0\), meaning that true compression rank \(r_0\) is acquired, \(\tilde{R}_t\) would be exactly \(\hat{R}_t\) in MPCA\(_{op}\). Under spherical neighbour arguments in section C, apply
Jensen’s inequality to C.7 to get \( \| \sum_{t=1}^{T} (P_{R_t} - \hat{P}_{R_t}) / T \|_{op} = o_p(1) \), to C.5 to get \( \| \mathbb{E} P_{R_t} - \mathbb{E} \hat{P}_{R_t} \|_{op} = o(1) \). Then, take \( \| \sum_{t=1}^{T} \hat{P}_{R_t} / T - \mathbb{E} \hat{P}_{R_t} \|_{op} = o_p(1) \) from section C, by triangular inequality we have \( \| \sum_{t=1}^{T} P_{R_t} / T - \mathbb{E} P_{R_t} \|_{op} = o_p(1) \) as \( T, p, q \to \infty \). In addition, now that \( \lambda_{p_0} (\mathbb{E} \hat{P}_{R_t}) \geq c > 0 \) while \( \lambda_{p_0+1} (\mathbb{E} \hat{P}_{R_t}) \to 0 \) as \( p, q \to \infty \), by Weyl’s inequality, \( \| \mathbb{E} P_{R_t} - \mathbb{E} \hat{P}_{R_t} \|_{op} = o(1) \) and \( \| \sum_{t=1}^{T} P_{R_t} / T - \mathbb{E} P_{R_t} \|_{op} = o_p(1) \) we have \( \mathbb{P} ( \tilde{p}_0 = p_0 \mid \hat{r}_0 = r_0 ) \to 1 \).
Table 6: Means and standard deviations (in parentheses) of $D(\hat{R}, R)$ and $D(\hat{C}, C)$ over 100 replications with $s_E = 1.5$ and $T = 3(pq)^{1/2}$. Here MPCA$_{op}$ and MPCA$_F$ stands for Manifold PCA methods; $(2D)^2$-PCA is from Zhang and Zhou (2005), it is equivalent to $\alpha$-PCA by Chen and Fan (2021) with $\alpha = -1$; PE stands for the projected estimation by Yu et al. (2021).

| Distribution | Evaluation | $p$ | $q$ | MPCA$_{op}$ | MPCA$_F$ | $(2D)^2$-PCA | PE |
|--------------|------------|-----|-----|-------------|-------------|----------------|-----|
| Gaussian     | $D(\hat{R}, R)$ | 20 20 | | $(0.4859,0.0885)$ | $(0.1702,0.0412)$ | $(0.4314,0.1167)$ | $(0.1372,0.0377)$ |
|              |            | 20 100 | | $(0.5261,0.0579)$ | $(0.0535,0.0082)$ | $(0.4293,0.1207)$ | $(0.0361,0.0064)$ |
|              |            | 100 100 | | $(0.2646,0.0994)$ | $(0.0591,0.0040)$ | $(0.2982,0.1353)$ | $(0.0514,0.0040)$ |
|              | $D(\hat{C}, C)$ | 20 20 | | $(0.4898,0.0952)$ | $(0.1756,0.0468)$ | $(0.4410,0.1271)$ | $(0.1419,0.0521)$ |
|              |            | 20 100 | | $(0.1026,0.0115)$ | $(0.0910,0.0082)$ | $(0.0863,0.0113)$ | $(0.0784,0.0089)$ |
|              |            | 100 100 | | $(0.2625,0.0886)$ | $(0.0586,0.0039)$ | $(0.2932,0.1267)$ | $(0.0509,0.0039)$ |
| $t_3$        | $D(\hat{R}, R)$ | 20 20 | | $(0.6169,0.0693)$ | $(0.3567,0.1261)$ | $(0.6974,0.0892)$ | $(0.5756,0.1938)$ |
|              |            | 20 100 | | $(0.5560,0.0391)$ | $(0.0829,0.0135)$ | $(0.6107,0.0748)$ | $(0.2065,0.2175)$ |
|              |            | 100 100 | | $(0.6091,0.0222)$ | $(0.0968,0.0077)$ | $(0.7433,0.1208)$ | $(0.4291,0.3020)$ |
|              | $D(\hat{C}, C)$ | 20 20 | | $(0.6232,0.0649)$ | $(0.3723,0.1308)$ | $(0.6880,0.0866)$ | $(0.5782,0.2010)$ |
|              |            | 20 100 | | $(0.2285,0.0479)$ | $(0.1454,0.0165)$ | $(0.3655,0.1833)$ | $(0.2800,0.2107)$ |
|              |            | 100 100 | | $(0.6092,0.0277)$ | $(0.0977,0.0072)$ | $(0.7398,0.1204)$ | $(0.4247,0.3019)$ |
| $t_1$        | $D(\hat{R}, R)$ | 20 20 | | $(0.0700,0.0177)$ | $(0.0522,0.0076)$ | $(0.8259,0.1215)$ | $(0.8402,0.1488)$ |
|              |            | 20 100 | | $(0.0415,0.0096)$ | $(0.0231,0.0030)$ | $(0.8677,0.0953)$ | $(0.8793,0.1189)$ |
|              |            | 100 100 | | $(0.0147,0.0010)$ | $(0.0149,0.0009)$ | $(0.9769,0.0240)$ | $(0.9807,0.0253)$ |
|              | $D(\hat{C}, C)$ | 20 20 | | $(0.0689,0.0145)$ | $(0.0507,0.0075)$ | $(0.8304,0.1164)$ | $(0.8432,0.1339)$ |
|              |            | 20 100 | | $(0.0296,0.0027)$ | $(0.0316,0.0025)$ | $(0.9519,0.0850)$ | $(0.9578,0.1005)$ |
|              |            | 100 100 | | $(0.0147,0.0011)$ | $(0.0151,0.0009)$ | $(0.9762,0.0292)$ | $(0.9831,0.0198)$ |
| $\alpha$-stable | $D(\hat{R}, R)$ | 20 20 | | $(0.6460,0.0612)$ | $(0.4506,0.1279)$ | $(0.8692,0.0598)$ | $(0.8834,0.0781)$ |
|              |            | 20 100 | | $(0.5714,0.0449)$ | $(0.0988,0.0171)$ | $(0.8873,0.0437)$ | $(0.9080,0.0454)$ |
|              |            | 100 100 | | $(0.9013,0.0449)$ | $(0.1188,0.0091)$ | $(0.9846,0.0071)$ | $(0.9854,0.0070)$ |
|              | $D(\hat{C}, C)$ | 20 20 | | $(0.6502,0.0592)$ | $(0.4507,0.1374)$ | $(0.8713,0.0613)$ | $(0.8887,0.0740)$ |
|              |            | 20 100 | | $(0.3284,0.0861)$ | $(0.1654,0.0174)$ | $(0.9606,0.0400)$ | $(0.9787,0.0287)$ |
|              |            | 100 100 | | $(0.9022,0.0543)$ | $(0.1188,0.0087)$ | $(0.9834,0.0074)$ | $(0.9843,0.0069)$ |
| skewed-$t_3$ | $D(\hat{R}, R)$ | 20 20 | | $(0.6189,0.0643)$ | $(0.3869,0.1410)$ | $(0.7403,0.0909)$ | $(0.6967,0.1768)$ |
|              |            | 20 100 | | $(0.5480,0.0465)$ | $(0.0813,0.0150)$ | $(0.6245,0.0882)$ | $(0.2654,0.2657)$ |
|              |            | 100 100 | | $(0.6299,0.0296)$ | $(0.0951,0.0066)$ | $(0.8033,0.1176)$ | $(0.5844,0.3226)$ |
|              | $D(\hat{C}, C)$ | 20 20 | | $(0.6070,0.0768)$ | $(0.3616,0.1408)$ | $(0.7421,0.0955)$ | $(0.6961,0.1930)$ |
|              |            | 20 100 | | $(0.2161,0.0469)$ | $(0.1375,0.0150)$ | $(0.4176,0.2039)$ | $(0.3341,0.2542)$ |
|              |            | 100 100 | | $(0.6265,0.0239)$ | $(0.0945,0.0068)$ | $(0.8072,0.1158)$ | $(0.5877,0.3136)$ |
Table 7: Means and standard deviations (in parentheses) of $\mathcal{D}(\hat{R}, R)$ and $\mathcal{D}(\hat{C}, C)$ over 100 replications with $s_E = 2$ and $T = 3(pq)^{1/2}$. Here MPCA$_{op}$ and MPCA$_F$ stands for Manifold PCA methods; $(2D)^2$-PCA is from Zhang and Zhou (2005), it is equivalent to $\alpha$-PCA by Chen and Fan (2021) with $\alpha = -1$; PE stands for the projected estimation by Yu et al. (2021).

| Distribution | Evaluation | $p$ | $q$ | MPCA$_{op}$ Mean | MPCA$_{op}$ Std | MPCA$_F$ Mean | MPCA$_F$ Std | $(2D)^2$-PCA Mean | $(2D)^2$-PCA Std | PE Mean | PE Std |
|--------------|------------|-----|-----|-----------------|----------------|----------------|----------------|-------------------|-----------------|---------|--------|
| Gaussian     | $\mathcal{D}(\hat{R}, R)$ | 20  | 20  | (0.5463,0.0752) | (0.2518,0.0795) | (0.5284,0.0818) | (0.2209,0.0882) |
|              |            | 20  | 100 | (0.5409,0.0430) | (0.0660,0.0110) | (0.5195,0.0635) | (0.0484,0.0096) |
|              |            | 100 | 100 | (0.5380,0.0466) | (0.0755,0.0052) | (0.5580,0.0304) | (0.0687,0.0049) |
|              | $\mathcal{D}(\hat{C}, C)$ | 20  | 20  | (0.5648,0.0592) | (0.2668,0.0984) | (0.5447,0.0670) | (0.2315,0.0953) |
|              |            | 20  | 100 | (0.1422,0.0212) | (0.1163,0.0125) | (0.1261,0.0215) | (0.1048,0.0129) |
|              |            | 100 | 100 | (0.5393,0.0460) | (0.0762,0.0052) | (0.5585,0.0273) | (0.0692,0.0051) |
| $t_3$        | $\mathcal{D}(\hat{R}, R)$ | 20  | 20  | (0.7081,0.0763) | (0.5596,0.1189) | (0.7817,0.0755) | (0.7694,0.1298) |
|              |            | 20  | 100 | (0.5696,0.0457) | (0.1143,0.0248) | (0.6824,0.0920) | (0.3751,0.2600) |
|              |            | 100 | 100 | (0.7410,0.0510) | (0.1294,0.0104) | (0.8591,0.1006) | (0.7143,0.2917) |
|              | $\mathcal{D}(\hat{C}, C)$ | 20  | 20  | (0.7051,0.0693) | (0.5471,0.1226) | (0.7738,0.0829) | (0.7675,0.1385) |
|              |            | 20  | 100 | (0.4192,0.1217) | (0.1890,0.0203) | (0.6013,0.1564) | (0.4481,0.2478) |
|              |            | 100 | 100 | (0.7408,0.0522) | (0.1296,0.0103) | (0.8587,0.0990) | (0.7108,0.2969) |
| $t_1$        | $\mathcal{D}(\hat{R}, R)$ | 20  | 20  | (0.0821,0.0237) | (0.0594,0.0095) | (0.8697,0.0809) | (0.8957,0.0811) |
|              |            | 20  | 100 | (0.0569,0.0135) | (0.0267,0.0033) | (0.9039,0.0399) | (0.9091,0.0428) |
|              |            | 100 | 100 | (0.0179,0.0014) | (0.0170,0.0009) | (0.9842,0.0067) | (0.9855,0.0062) |
|              | $\mathcal{D}(\hat{C}, C)$ | 20  | 20  | (0.0849,0.0193) | (0.0592,0.0090) | (0.8706,0.0829) | (0.8795,0.0703) |
|              |            | 20  | 100 | (0.3631,0.0037) | (0.0364,0.0025) | (0.6013,0.1564) | (0.4481,0.2478) |
|              |            | 100 | 100 | (0.0176,0.0013) | (0.0169,0.0010) | (0.9840,0.0090) | (0.9851,0.0064) |
| $\alpha$-stable | $\mathcal{D}(\hat{R}, R)$ | 20  | 20  | (0.7404,0.0672) | (0.6178,0.1109) | (0.8953,0.0390) | (0.9104,0.0408) |
|              |            | 20  | 100 | (0.6226,0.0442) | (0.1385,0.0335) | (0.9046,0.0370) | (0.9174,0.0360) |
|              |            | 100 | 100 | (0.7900,0.0157) | (0.2533,0.2459) | (0.9836,0.0081) | (0.9843,0.0076) |
|              | $\mathcal{D}(\hat{C}, C)$ | 20  | 20  | (0.7507,0.0702) | (0.6170,0.1215) | (0.8959,0.0456) | (0.9115,0.0376) |
|              |            | 20  | 100 | (0.5948,0.1134) | (0.2171,0.0262) | (0.9774,0.0156) | (0.9842,0.0063) |
|              |            | 100 | 100 | (0.9681,0.0177) | (0.2295,0.0222) | (0.9839,0.0075) | (0.9846,0.0073) |
| skewed-$t_3$ | $\mathcal{D}(\hat{R}, R)$ | 20  | 20  | (0.6981,0.0697) | (0.5468,0.1087) | (0.8213,0.0619) | (0.8328,0.1005) |
|              |            | 20  | 100 | (0.5842,0.0532) | (0.1137,0.0260) | (0.7048,0.0897) | (0.4515,0.2845) |
|              |            | 100 | 100 | (0.8091,0.0651) | (0.1281,0.0103) | (0.9191,0.0655) | (0.8814,0.1871) |
|              | $\mathcal{D}(\hat{C}, C)$ | 20  | 20  | (0.7142,0.0755) | (0.5531,0.1118) | (0.8218,0.0735) | (0.8444,0.0990) |
|              |            | 20  | 100 | (0.4068,0.1141) | (0.1905,0.0268) | (0.6639,0.1673) | (0.5066,0.2803) |
|              |            | 100 | 100 | (0.8226,0.0577) | (0.1269,0.0090) | (0.9247,0.0666) | (0.9092,0.1498) |

11
Table 8: Means and standard deviations (in parentheses) of MSE and opMax over 100 replications with $s_E = 1.5$ and $T = 3(pq)^{1/2}$. Here MPCA$_{\text{op}}$ and MPCA$_F$ stands for Manifold PCA methods; $(2D)^2$-PCA is from Zhang and Zhou (2005), it is equivalent to $\alpha$-PCA by Chen and Fan (2021) with $\alpha = -1$; PE stands for the projected estimation by Yu et al. (2021).

| MSE Distribution | $p$ | $q$ | MPCA$_{\text{op}}$ | MPCA$_F$ | $(2D)^2$-PCA | PE |
|------------------|-----|-----|---------------------|----------|---------------|----|
| Gauss           | 20  | 20  | (0.1961,0.0260)     | (0.0758,0.0081) | (0.1670,0.0296) | (0.0683,0.0083) |
|                 | 20  | 100 | (0.0387,0.0108)     | (0.0666,0.0003) | (0.0266,0.0075) | (0.0059,0.0003) |
|                 | 100 | 100 | (0.0134,0.0047)     | (0.0027,0.0001) | (0.0161,0.0078) | (0.0025,0.0001) |
| $t_3$           | 20  | 20  | (0.4953,0.0590)     | (0.2977,0.0688) | (0.7513,0.3589) | (0.6813,0.4534) |
|                 | 20  | 100 | (0.0628,0.0082)     | (0.0193,0.0013) | (0.1273,0.1098) | (0.0806,0.1168) |
|                 | 100 | 100 | (0.0575,0.0036)     | (0.0079,0.0003) | (0.1415,0.1468) | (0.1068,0.1579) |
| $t_1$           | 20  | 20  | (6.3133,36.191)     | (4.5986,23.124) | (2492.6,23518) | (2492.7,23518) |
|                 | 20  | 100 | (19.744,156.67)     | (16.525,126.47) | (5107.7,43364) | (5107.7,43364) |
| $\alpha$-stable | 20  | 20  | (3.4060,16.984)     | (3.1931,18.262) | (35.941,250.45) | (36.095,250.43) |
|                 | 20  | 100 | (1.2566,0.0966)     | (0.0515,0.0413) | (5.3870,18.945) | (5.4352,18.946) |
| skewed-$t_3$    | 20  | 20  | (0.4887,0.0708)     | (0.3088,0.0789) | (0.1264,0.0170) | (0.0757,0.0073) |
|                 | 20  | 100 | (0.0622,0.0087)     | (0.0187,0.0013) | (0.1576,0.1347) | (0.1165,0.1595) |
|                 | 100 | 100 | (0.0609,0.0043)     | (0.0078,0.0003) | (0.2176,0.5886) | (0.1974,0.5937) |

| opMax Distribution | $p$ | $q$ | MPCA$_{\text{op}}$ | MPCA$_F$ | $(2D)^2$-PCA | PE |
|--------------------|-----|-----|---------------------|----------|---------------|----|
| Gauss              | 20  | 20  | (0.1264,0.0170)     | (0.0758,0.0065) | (0.1133,0.0159) | (0.0757,0.0073) |
|                    | 20  | 100 | (0.0545,0.0110)     | (0.0155,0.0012) | (0.0433,0.0091) | (0.0153,0.0012) |
|                    | 100 | 100 | (0.0187,0.0047)     | (0.0071,0.0004) | (0.0207,0.0066) | (0.0071,0.0005) |
| $t_3$              | 20  | 20  | (0.2551,0.1117)     | (0.2119,0.0883) | (0.6100,0.4355) | (0.6526,0.4516) |
|                    | 20  | 100 | (0.0598,0.0099)     | (0.0370,0.0136) | (0.2376,0.2590) | (0.1992,0.2808) |
|                    | 100 | 100 | (0.0385,0.0070)     | (0.0160,0.0057) | (0.3002,0.2869) | (0.2864,0.3061) |
| $t_1$              | 20  | 20  | (1.7080,3.8917)     | (1.5531,3.2567) | (17.793,84.060) | (17.794,84.060) |
|                    | 20  | 100 | (1.6269,7.4722)     | (1.5267,6.8267) | (25.729,120.29) | (25.729,120.29) |
|                    | 100 | 100 | (0.3822,1.1203)     | (0.3740,1.0838) | (19.831,63.344) | (19.832,63.344) |
| $\alpha$-stable    | 20  | 20  | (1.1161,2.5846)     | (0.9556,2.6376) | (3.9438,9.1424) | (3.9717,9.1340) |
|                    | 20  | 100 | (0.2179,0.2049)     | (0.1697,0.1533) | (2.3220,2.6888) | (2.3308,2.6843) |
|                    | 100 | 100 | (0.3336,0.3474)     | (0.1577,0.2429) | (4.4859,5.0260) | (4.4867,5.0255) |
| skewed-$t_3$       | 20  | 20  | (0.2826,0.1318)     | (0.2507,0.1286) | (0.9194,0.6804) | (0.9961,0.6567) |
|                    | 20  | 100 | (0.0649,0.0163)     | (0.0399,0.0143) | (0.2827,0.2790) | (0.2557,0.3077) |
|                    | 100 | 100 | (0.0401,0.0150)     | (0.0179,0.0066) | (0.3965,0.4447) | (0.3925,0.4546) |
Table 9: Means and standard deviations (in parentheses) of MSE and opMax over 100 replications with $s_E = 2$ and $T = 3(pq)^{1/2}$. Here $\text{MPCA}_{op}$ and $\text{MPCA}_F$ stands for Manifold PCA methods; $(2D)^2$-PCA is from Zhang and Zhou (2005), it is equivalent to $\alpha$-PCA by Chen and Fan (2021) with $\alpha = -1$; PE stands for the projected estimation by Yu et al. (2021).

| Distribution | $p$ | $q$ | $\text{MPCA}_{op}$ | $\text{MPCA}_F$ | $(2D)^2$-PCA | PE |
|--------------|-----|-----|---------------------|-----------------|--------------|----|
| Gauss        | 20  | 20  | (0.3133,0.0297)     | (0.1549,0.0276) | (0.2971,0.0298) | (0.1432,0.0303) |
|              | 20  | 100 | (0.0485,0.0106)     | (0.0116,0.0006) | (0.0443,0.0090) | (0.0107,0.0006) |
|              | 100 | 100 | (0.0439,0.0044)     | (0.0047,0.0001) | (0.0460,0.0047) | (0.0045,0.0001) |
| $t_3$        | 20  | 20  | (0.8438,0.01168)    | (0.6879,0.1074) | (1.4007,0.7806) | (1.4703,0.8140) |
|              | 20  | 100 | (0.1073,0.0162)     | (0.0360,0.0036) | (0.3363,0.4292) | (0.2782,0.4445) |
|              | 100 | 100 | (0.0882,0.0048)     | (0.0144,0.0005) | (0.2006,0.1075) | (0.1919,0.1296) |
| $t_1$        | 20  | 20  | (12.0691,70.742)    | (10.7265,63.910) | (651.55,4839.8) | (651.66,4839.9) |
|              | 20  | 100 | (184.19,1275.9)     | (155.09,1063.5) | (80440,663049)  | (80440,663049)  |
|              | 100 | 100 | (0.6902,2.9519)     | (0.6519,2.7665) | (1434,6985.8)   | (1434,6985.8)   |
| $\alpha$-stable | 20  | 20  | (4.1815,15.561)     | (3.1335,12.560) | (32.625,162.79) | (32.842,162.80) |
|              | 20  | 100 | (0.4038,0.8881)     | (0.1229,0.2451) | (13.323,69.263) | (13.383,69.264) |
|              | 100 | 100 | (1.6554,8.3149)     | (0.0981,0.1326) | (37,704,149.03) | (37,749,149.03) |
| skewed-$t_3$ | 20  | 20  | (0.8543,0.2492)     | (0.6999,0.2006) | (1.8161,1.3214) | (1.9703,1.3554) |
|              | 20  | 100 | (0.1038,0.0125)     | (0.0353,0.0029) | (0.3024,0.2596) | (0.2616,0.2885) |
|              | 100 | 100 | (0.0986,0.0071)     | (0.0143,0.0066) | (0.2907,0.2318) | (0.3079,0.2351) |

| Distribution | $p$ | $q$ | $\text{MPCA}_{op}$ | $\text{MPCA}_F$ | $(2D)^2$-PCA | PE |
|--------------|-----|-----|---------------------|-----------------|--------------|----|
| Gauss        | 20  | 20  | (0.1536,0.0146)     | (0.1104,0.0130) | (0.1506,0.0142) | (0.1110,0.0148) |
|              | 20  | 100 | (0.0594,0.0109)     | (0.0207,0.0015) | (0.0566,0.0098) | (0.0205,0.0015) |
|              | 100 | 100 | (0.0348,0.0039)     | (0.0095,0.0006) | (0.0359,0.0040) | (0.0095,0.0006) |
| $t_3$        | 20  | 20  | (0.3410,0.01820)    | (0.3184,0.1199) | (0.9777,0.6471) | (1.0549,0.6193) |
|              | 20  | 100 | (0.0804,0.00445)    | (0.0542,0.0298) | (0.4899,0.5037) | (0.4745,0.5275) |
|              | 100 | 100 | (0.0442,0.0147)     | (0.0217,0.0060) | (0.4244,0.2420) | (0.4308,0.2528) |
| $t_1$        | 20  | 20  | (2.2703,5.4880)     | (2.1404,5.1760) | (15,821,40.776) | (15,824,40.775) |
|              | 20  | 100 | (4.6084,22.901)     | (4.2820,21.001) | (85,414,480.86) | (85,415,480.86) |
|              | 100 | 100 | (0.5571,1.3087)     | (0.5456,1.2697) | (20,645,61.759) | (20,645,61.759) |
| $\alpha$-stable | 20  | 20  | (1.5571,2.5314)     | (1.2625,2.1515) | (4.8993,7.9365) | (4.9220,7.9273) |
|              | 20  | 100 | (0.4417,0.6399)     | (0.2548,0.3278) | (3.2412,4.8479) | (3.2514,4.8433) |
|              | 100 | 100 | (0.7617,1.9625)     | (0.2440,0.2686) | (6.2133,7.8045) | (6.2145,7.8038) |
| skewed-$t_3$ | 20  | 20  | (0.3958,0.2649)     | (0.3815,0.2394) | (1.2478,0.7846) | (1.3338,0.7403) |
|              | 20  | 100 | (0.0781,0.0280)     | (0.0557,0.0208) | (0.4686,0.3850) | (0.4641,0.4085) |
|              | 100 | 100 | (0.0536,0.0238)     | (0.0262,0.0120) | (0.5649,0.3378) | (0.5866,0.3244) |
Table 10: Frequencies of exact estimation and underestimation (in parentheses) of factor numbers over 100 replications with $s_E = 1.5$ and $T = 3(pq)^{1/2}$. Here $\text{MER}_{op}$ and $\text{MER}_F$ stands for Manifold eigenvalue-ratio methods; $(2D)^2$-$\text{ER}$ is equivalent to the ER method in Chen and Fan (2021) with $\alpha = -1$; IterER is from Yu et al. (2021).

| Distribution | $p$ | $q$ | $\text{MER}_{op}$ | $\text{MER}_F$ | $(2D)^2$-$\text{ER}$ | IterER |
|--------------|-----|-----|-------------------|----------------|----------------------|--------|
| Gaussian     | 20  | 20  | (0.09,0.39)       | (0.77,0.22)    | (0.05,0.76)          | (0.68,0.32) |
|              | 20  | 100 | (0.29,0.06)       | (1.00,0.00)    | (0.07,0.28)          | (1.00,0.00) |
|              | 100 | 100 | (0.70,0.00)       | (1.00,0.00)    | (0.00,0.00)          | (1.00,0.00) |
| $t_3$        | 20  | 20  | (0.00,0.97)       | (0.12,0.87)    | (0.00,0.99)          | (0.05,0.94) |
|              | 20  | 100 | (0.15,0.75)       | (1.00,0.00)    | (0.03,0.95)          | (0.70,0.11) |
|              | 100 | 100 | (0.00,0.98)       | (0.87,0.13)    | (0.02,0.98)          | (0.66,0.33) |
| $t_1$        | 20  | 20  | (0.95,0.05)       | (1.00,0.00)    | (0.09,0.79)          | (0.04,0.75) |
|              | 20  | 100 | (0.98,0.02)       | (1.00,0.00)    | (0.08,0.82)          | (0.07,0.80) |
|              | 100 | 100 | (1.00,0.00)       | (1.00,0.00)    | (0.09,0.73)          | (0.10,0.76) |
| $\alpha$-stable | 20  | 20  | (0.01,0.97)       | (0.04,0.95)    | (0.03,0.97)          | (0.04,0.96) |
|              | 20  | 100 | (0.01,0.99)       | (0.91,0.09)    | (0.09,0.88)          | (0.05,0.91) |
|              | 100 | 100 | (0.00,1.00)       | (0.34,0.66)    | (0.09,0.81)          | (0.02,0.94) |
| skewed-$t_3$ | 20  | 20  | (0.00,0.99)       | (0.11,0.88)    | (0.00,0.99)          | (0.05,0.94) |
|              | 20  | 100 | (0.18,0.71)       | (0.98,0.02)    | (0.06,0.92)          | (0.66,0.19) |
|              | 100 | 100 | (0.00,1.00)       | (0.87,0.13)    | (0.02,0.98)          | (0.41,0.58) |

Table 11: Frequencies of exact estimation and underestimation (in parentheses) of factor numbers over 100 replications with $s_E = 2$ and $T = 3(pq)^{1/2}$. Here $\text{MER}_{op}$ and $\text{MER}_F$ stands for Manifold eigenvalue-ratio methods; $(2D)^2$-$\text{ER}$ is equivalent to the ER method in Chen and Fan (2021) with $\alpha = -1$; IterER is from Yu et al. (2021).

| Distribution | $p$ | $q$ | $\text{MER}_{op}$ | $\text{MER}_F$ | $(2D)^2$-$\text{ER}$ | IterER |
|--------------|-----|-----|-------------------|----------------|----------------------|--------|
| Gaussian     | 20  | 20  | (0.08,0.72)       | (0.39,0.59)    | (0.08,0.78)          | (0.34,0.62) |
|              | 20  | 100 | (0.12,0.10)       | (1.00,0.00)    | (0.15,0.40)          | (0.99,0.01) |
|              | 100 | 100 | (0.00,0.00)       | (1.00,0.00)    | (0.00,0.02)          | (1.00,0.00) |
| $t_3$        | 20  | 20  | (0.00,1.00)       | (0.00,1.00)    | (0.00,1.00)          | (0.00,1.00) |
|              | 20  | 100 | (0.00,1.00)       | (0.93,0.07)    | (0.00,1.00)          | (0.48,0.44) |
|              | 100 | 100 | (0.00,1.00)       | (0.25,0.75)    | (0.03,0.96)          | (0.14,0.86) |
| $t_1$        | 20  | 20  | (0.88,0.12)       | (1.00,0.00)    | (0.01,0.91)          | (0.02,0.88) |
|              | 20  | 100 | (0.98,0.02)       | (1.00,0.00)    | (0.10,0.77)          | (0.08,0.79) |
|              | 100 | 100 | (1.00,0.00)       | (1.00,0.00)    | (0.11,0.77)          | (0.07,0.80) |
| $\alpha$-stable | 20  | 20  | (0.00,1.00)       | (0.00,1.00)    | (0.04,0.95)          | (0.05,0.94) |
|              | 20  | 100 | (0.00,1.00)       | (0.50,0.50)    | (0.01,0.95)          | (0.01,0.95) |
|              | 100 | 100 | (0.00,1.00)       | (0.02,0.98)    | (0.03,0.85)          | (0.02,0.90) |
| skewed-$t_3$ | 20  | 20  | (0.00,1.00)       | (0.00,1.00)    | (0.00,1.00)          | (0.00,1.00) |
|              | 20  | 100 | (0.00,1.00)       | (0.89,0.11)    | (0.00,1.00)          | (0.42,0.50) |
|              | 100 | 100 | (0.00,1.00)       | (0.18,0.82)    | (0.03,0.93)          | (0.04,0.96) |