Journal of Statistical Planning and Inference 213 (2021) 50–71

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Optimal sparse eigenspace and low-rank density matrix estimation for quantum systems

Tony Cai [a], Donggyu Kim [b], Xinyu Song [c,*], Yazhen Wang [d]

[a] Department of Statistics, The Wharton School, University of Pennsylvania, United States of America
[b] College of Business, Korea Advanced Institute of Science and Technology, Republic of Korea
[c] School of Statistics and Management, Shanghai University of Finance and Economics, China
[d] Department of Statistics, University of Wisconsin-Madison, United States of America

ARTICLE INFO

Article history:
Received 4 November 2019
Received in revised form 21 September 2020
Accepted 1 November 2020
Available online 17 November 2020

Keywords:
Iterative thresholding
Minimax estimation
Pauli matrix
Principal component analysis
Quantum state tomography

ABSTRACT

Quantum state tomography, which aims to estimate quantum states that are described by density matrices, plays an important role in quantum science and quantum technology. This paper examines the eigenspace estimation and the reconstruction of large low-rank density matrix based on Pauli measurements. Both ordinary principal component analysis (PCA) and iterative thresholding sparse PCA (ITSPCA) estimators of the eigenspace are studied, and their respective convergence rates are established. In particular, we show that the ITSPCA estimator is rate-optimal. We present the reconstruction of the large low-rank density matrix and obtain its optimal convergence rate by using the ITSPCA estimator. A numerical study is carried out to investigate the finite sample performance of the proposed estimators.

1. Introduction

In quantum science and quantum technology, we often need to learn and engineer quantum systems. A prominent example is quantum information science (Nielsen and Chuang, 2010; Wang, 2011, 2012; Wang and Song, 2020). A quantum system is described by its state, therefore for its study, we need to reconstruct the quantum state. In the literature, researchers often characterize the quantum state by a complex matrix that is the so-called density matrix and refer to the reconstruction of the quantum state by quantum state tomography. Traditionally, quantum state tomography employs classical statistical models and methods to deduce quantum states from quantum measurements that are observations obtained from measuring identically prepared quantum systems. Because of the exponential complexity of the quantum system and the exponential growth of its corresponding density matrix, quantum state tomography often needs to reconstruct the density matrix of high-dimension. It is known that classical statistical methods are neither efficient nor effective in recovering the large density matrix. In this paper, we employ modern high-dimensional statistics to investigate the reconstruction of large density matrix.

Cai et al. (2016) studied the estimation of large sparse density matrix represented by Pauli matrices and established the optimal convergence rate of its estimator. However, the sparsity condition assumed in Cai et al. (2016) may not be very reasonable. For example, a low-rank density matrix does not satisfy the condition under the Pauli representation. Koltchinskii and Xia (2015) studied the reconstruction of low-rank density matrix and investigated its optimal estimation. We note that the estimation of eigenspace also plays an important role in the reconstruction of the low-rank density

* Correspondence to: 777 Guo Ding Road, Yang Pu District, Shanghai 200433, China.
E-mail address: song.xinyu@mail.shufe.edu.cn (X. Song).

https://doi.org/10.1016/j.jspi.2020.11.002
0378-3758/

matrix, and modern high-dimensional statistics suggest that the optimal estimation of the eigenspace depends on the sparse structure of the eigenvectors. See Birgé (2001), Cai et al. (2013, 2015), Johnstone and Lu (2009), Ma (2013) and Vu and Lei (2013) for related research works. Koltchinskii and Xia (2015) considered a broad class of low-rank density matrices that may not be suitable for density matrices with sparse eigenvectors, and as a result, their optimal rate is not sharp for density matrices with sparse eigenvectors.

This paper considers the eigenspace estimation problem for a quantum spin system based on Pauli measurements. As all Pauli matrices have ± 1 eigenvalues, Pauli measurements also take binary values 1 and -1 while their distributions correspond to shifted and rescaled binomial distributions (Cai et al., 2016; Wang, 2013). Thus, the eigenspace estimation problem studied in this paper is a high-dimensional statistics problem with binomial distributions, where both the matrix size and sample size are allowed to go infinity. To be specific, we analyze the asymptotic behaviors of the principal component analysis (PCA) estimators and establish their convergence rates under both dense and sparse eigenvector settings. Under the sparse eigenvector condition, we derive the minimax lower bound for the eigenspace estimation procedure and demonstrate that the iterative thresholding sparse PCA (ITSPCA) proposed by Ma (2013) can achieve the minimax lower bound, and therefore the ITSPCA is rate-optimal. The convergence rates and minimax lower bound in this paper are obtained by asymptotic analysis with binomial distributions instead of usual normal distributions. With the ITSPCA eigenspace estimator, we can estimate the corresponding eigenvalues and reconstruct the large density matrix. We show that the constructed low-rank density matrix is also rate-optimal.

The rest of this paper proceeds as follows: Section 2 briefly reviews the quantum state and the density matrix that is represented through Pauli matrices with its estimation. Section 3 describes the iterative thresholding estimation algorithm and defines a sparsity condition for eigenvectors. Given the sparsity condition, Section 4 establishes the asymptotic theory for the iterative thresholding estimator and derives the minimax lower bound for eigenspace estimation under spectral and Frobenius norms, where both the matrix size and sample size are allowed to go to infinity. Section 5 proposes the eigenvalue and low-rank density matrix estimators and derives their convergence rates. Section 6 features numerical studies to illustrate the finite sample performances of the proposed estimators. All proofs are collected in Section 7 and further technical details are presented in the Appendix.

2. Review for quantum state tomography

2.1. Quantum state and density matrix

For a d-dimensional quantum system, we describe its quantum state by a density matrix $\boldsymbol{\rho}$ on the d-dimensional complex space $\mathbb{C}^d$, where the density matrix $\boldsymbol{\rho}$ is a d-by-d complex matrix satisfying (1) Hermitian, that is, $\boldsymbol{\rho}$ is equal to its conjugate transpose; (2) positive semi-definite; (3) unit trace. The density matrix $\boldsymbol{\rho}$ can be expressed by the d-dimensional Pauli matrices. To be specific, let

$$\boldsymbol{\sigma}_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \ \boldsymbol{\sigma}_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \ \boldsymbol{\sigma}_2 = \begin{pmatrix} 0 & -\sqrt{-1} \\ \sqrt{-1} & 0 \end{pmatrix}, \ \text{and} \ \boldsymbol{\sigma}_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

where $\boldsymbol{\sigma}_1$, $\boldsymbol{\sigma}_2$, and $\boldsymbol{\sigma}_3$ are called Pauli matrices. High-dimensional Pauli matrices are defined through Tensor products. Let $d = 2^b$ for some integer b. We form b-fold tensor products of $\boldsymbol{\sigma}_0$, $\boldsymbol{\sigma}_1$, $\boldsymbol{\sigma}_2$, and $\boldsymbol{\sigma}_3$ to obtain the d-dimensional Pauli matrices

$$\mathcal{G}_P = \left\{ \mathbf{B}_j = \boldsymbol{\sigma}_{\ell_1} \otimes \boldsymbol{\sigma}_{\ell_2} \otimes \cdots \otimes \boldsymbol{\sigma}_{\ell_b}, \quad (\ell_1, \ell_2, \ldots, \ell_b) \in \{0, 1, 2, 3\}^b \right\},$$

and the cardinality of $\mathcal{G}_P$ is $p = 4^b$. We set $\mathbf{B}_1 = \mathbf{I}_d$, where $\mathbf{I}_d$ is the d-dimensional identity matrix. Denote by $\mathbb{C}^{d\times d}$ the space of all d-by-d complex matrices equipped with the Frobenius norm. Proposition 1 in Cai et al. (2016) showed that all Pauli matrices $\mathbf{B}_1, \ldots, \mathbf{B}_p$ form an orthogonal basis for complex Hermitian matrices in $\mathbb{C}^{d\times d}$, and any density matrix $\boldsymbol{\rho}$ can be expanded under the Pauli basis as follows:

$$\boldsymbol{\rho} = d^{-1} \left(\mathbf{I}_d + \sum_{j=2}^{p} \beta_j \mathbf{B}_j \right),$$

where coefficients β_j's satisfy $\beta_j = \mathrm{tr}(\boldsymbol{\rho}\mathbf{B}_j)$ and $|\beta_j| \le 1$.

2.2. Pauli measurements and density matrix estimation

The Pauli matrices are widely used in quantum physics and quantum information science to perform quantum measurements, and quantum measurements are often based on observables, where an observable is defined as a Hermitian matrix on $\mathbb{C}^d$. To be specific, suppose that an experiment is conducted to perform measurements on each Pauli observable $\mathbf{B}_j$ independently for n quantum systems that are identically prepared under the same quantum state $\boldsymbol{\rho}$. As each $\mathbf{B}_j$ has eigenvalues ± 1, the theory of quantum mechanics indicates that the Pauli measurements take values 1 and -1, and thus are Bernoulli trials. Denote by N_j the average of the n measurement outcomes obtained from measuring $\mathbf{B}_j$, $j = 2, \ldots, p$.

Then $n(N_j + 1)/2$ obeys a binomial distribution with n trials and cell probability $(1 + \beta_j)/2$, where $\mathrm{E}(N_j) = \beta_j$ and $\mathrm{Var}(N_j) = (1 - \beta_j^2)/n$ (Cai et al., 2016). The goal of this paper is to estimate eigenspace of $\boldsymbol{\rho}$ based on data $N_2, \ldots, N_p$.

To estimate the eigenspace of the density matrix $\boldsymbol{\rho}$, we first need an initial estimator of $\boldsymbol{\rho}$ based on the Pauli measurements $N_2, \ldots, N_p$. Given the binomial distribution, we easily derive that each N_j is the MLE and UMVUE of β_j. Thus, a natural estimator of $\boldsymbol{\rho}$ is given by

$$\widehat{\boldsymbol{\rho}} = (\widehat{\rho}_{ij})_{i,j=1,\ldots,d} = \frac{1}{d}\left(\mathbf{I}_d + \sum_{j=2}^{p} \widehat{\beta}_j \mathbf{B}_j\right), \tag{2.1}$$

where $\widehat{\beta}_j = N_j$.

3. Eigenspace estimation

3.1. Eigen-decomposition of density matrix

Assume that a density matrix $\boldsymbol{\rho}$ has finite rank r. By the spectral decomposition, we have

$$\boldsymbol{\rho} = \sum_{\nu=1}^{r} \lambda_\nu \mathbf{q}_\nu \mathbf{q}_\nu^\dagger, \tag{3.1}$$

where λ_ν's are eigenvalues such that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > 0$ and $\sum_{\nu=1}^{r} \lambda_\nu = 1$, moreover, $\mathbf{q}_1, \ldots, \mathbf{q}_r \in \mathbb{C}^d$ are their corresponding eigenvectors.

In this paper, we consider estimation of the eigenspace spanned by the first m eigenvectors of $\boldsymbol{\rho}$, that is, our aim is to estimate the eigenspace generated by $\mathbf{Q} = (\mathbf{q}_1, \ldots, \mathbf{q}_m) \in \mathbb{C}^{d\times m}$, where m is a given integer. To make the eigenspace estimation problem well defined, we need to assume that $m \leq r$ and $\lambda_m - \lambda_{m+1} > C_\lambda$ for some generic positive constant C_λ free of n and d, that is, there is a gap between eigenvalues λ_m and λ_{m+1} so that the corresponding eigenspaces are well separated for investigating asymptotic properties of the eigenspace estimation.

3.2. Ordinary PCA

We define the eigenspace estimator of $\mathbf{Q}$ by the eigenspace spanned by the first m eigenvectors of the density matrix estimator $\widehat{\boldsymbol{\rho}}$ in (2.1). As the m eigenvectors are from ordinary PCA, the defined eigenspace estimator is called the ordinary PCA estimator and is denoted by $\widehat{\mathbf{Q}}$. Before investigating its asymptotic properties, we first fix some notations. For $\mathbf{x} = (x_1, \ldots, x_d)^\top \in \mathbb{C}^d$ and $\mathbf{A} = (A_{ij}) \in \mathbb{C}^{d\times d}$, define the ℓ_α-norms,

$$\|\mathbf{x}\|_\alpha = \left(\sum_{i=1}^{d} |x_i|^\alpha\right)^{1/\alpha}, \quad \|\mathbf{A}\|_\alpha = \sup\{\|\mathbf{A}\mathbf{x}\|_\alpha, \|\mathbf{x}\|_\alpha = 1\}, \quad 1 \leq \alpha \leq \infty.$$

Then the matrix spectral norm $\|\mathbf{A}\|_2$ is equal to square root of the largest eigenvalue of $\mathbf{A}\mathbf{A}^\dagger$. Moreover, note that

$$\|\mathbf{A}\|_1 = \max_{1\leq j\leq d} \sum_{i=1}^{d} |A_{ij}|, \quad \|\mathbf{A}\|_\infty = \max_{1\leq i\leq d} \sum_{j=1}^{d} |A_{ij}|.$$

and we have the following inequality,

$$\|\mathbf{A}\|_2^2 \leq \|\mathbf{A}\|_1 \|\mathbf{A}\|_\infty.$$

The matrix Frobenius norm is denoted by $\|\mathbf{A}\|_F = \sqrt{\mathrm{tr}(\mathbf{A}^\dagger\mathbf{A})}$. For a symmetric or complex Hermitian matrix $\mathbf{A}$, $\|\mathbf{A}\|_F$ is the square root of the sum of squared eigenvalues, $\|\mathbf{A}\|_2$ is equal to its largest absolute eigenvalue, and $\|\mathbf{A}\|_2 \leq \|\mathbf{A}\|_1 = \|\mathbf{A}\|_\infty$. Denote by C a positive generic constant whose values are free of n and p and may change from appearance to appearance. For positive sequences $\varphi_{n,d}$ and $\psi_{n,d}$ depend on n and d, we use $\varphi_{n,d} \asymp \psi_{n,d}$ to denote that their ratio $\varphi_{n,d}/\psi_{n,d}$ is asymptotically bounded by positive generic constants from both below and above as $n, d \to \infty$.

To measure the performances of proposed eigenspace estimators, we define a notation for the distance between eigenspaces. To be specific, the distance between two eigenspaces spanned by $\mathbf{Q}_1$ and $\mathbf{Q}_2$ is defined by

$$\|\sin(\mathbf{Q}_1, \mathbf{Q}_2)\|_F^2 = \|\mathbf{Q}_1\mathbf{Q}_1^\dagger(\mathbf{Q}_2\mathbf{Q}_2^\dagger)^\perp\|_F^2 \tag{3.2}$$

and

$$\|\sin(\mathbf{Q}_1, \mathbf{Q}_2)\|_2^2 = \|\mathbf{Q}_1\mathbf{Q}_1^\dagger(\mathbf{Q}_2\mathbf{Q}_2^\dagger)^\perp\|_2^2, \tag{3.3}$$

where $\mathbf{Q}_1\mathbf{Q}_1^\dagger$ and $\mathbf{Q}_2\mathbf{Q}_2^\dagger$ are projection matrices on eigenspaces $\mathbf{Q}_1$ and $\mathbf{Q}_2$, respectively, and for a given projection matrix $\mathbf{P}$, we have $\mathbf{P}^\perp = \mathbf{I}_d - \mathbf{P}$. The distances refer to the canonical angles between $\mathbf{Q}_1$ and $\mathbf{Q}_2$ that generalize the notion of angles between lines.

The following theorem establishes the convergence rate of the ordinary PCA estimator.

Theorem 1. *Suppose that one of the following conditions is satisfied,*

(i) $n^{\alpha_1} \le d \le \exp(n^{\alpha_2})$ *for some* $\alpha_1 > 1/2$ *and* $\alpha_2 < 1$*;*
(ii) $d \le n^{1/2}$ *and* $\frac{\log n}{d} \to 0$ *as* $n \to \infty$.

Then we have

$$\sup_{\mathbf{Q}\in\mathbb{V}_{d,m}} E\left[\|\sin(\mathbf{Q}, \widehat{\mathbf{Q}})\|_2^2\right] \le \sup_{\mathbf{Q}\in\mathbb{V}_{d,m}} E\left[\|\sin(\mathbf{Q}, \widehat{\mathbf{Q}})\|_F^2\right] \le Cn^{-1}, \tag{3.4}$$

where $\mathbb{V}_{d,m} = \{\mathbf{Q} \in \mathbb{C}^{d\times m} : \mathbf{Q}^\dagger\mathbf{Q} = \mathbf{I}_d\}$ *is the complex Stiefel manifold of d-by-m orthonormal matrices, and C is a generic constant free of n and d.*

Remark 1. Theorem 1 shows that the convergence rate for the ordinary PCA estimator is $n^{-1/2}$ regardless of sparsity condition on eigenvectors. As the PCA approach does not utilize any sparse eigenvectors, it can achieve only $n^{-1/2}$ convergence rate even for a density matrix with sparse eigenvectors. We will show later that this convergence rate is suboptimal for the sparse case. Also due to the proof techniques used we leave some gaps for d in the conditions (i) and (ii), that is, d is between $n^{1/2}$ and n^{α_1} or below $\log n$. However, as α_1 can be very close to $1/2$, and the classical PCA theory indicates that the theorem is true for the case of fixed d, the gaps are very small. Of course future work may resolve this issue.

3.3. Iterative thresholding sparse PCA

As the complexity of a quantum system increases exponentially with its components, the dimension d of the density matrix grows exponentially and is often very large. In usual high-dimensional statistics, we may impose sparsity condition on the eigenvectors of the density matrix and estimate the eigenspace spanned by the first m sparse eigenvectors accordingly. For $\mathbf{A} \in \mathbb{C}^{d\times m}$, $\mathbf{A}_{IJ}$ denotes the submatrix of $\mathbf{A}$ formed by rows and columns whose indices are in I and J, respectively, where I and J are subsets of $\{1, \ldots, d\}$. When I or J includes all the indices, we replace them with dot. For example, $\mathbf{A}_{\cdot J}$ is the submatrix of $\mathbf{A}$ with all rows and columns indexed by J.

We now impose the sparsity condition on the first r eigenvectors of $\boldsymbol{\rho}$ defined in (3.1) as follows: For each $\nu = 1, \ldots, r$, assume that for some $\delta \in [0, 2)$,

$$\mathbf{q}_\nu \in \Xi_\delta(\pi(d)) \stackrel{\text{def}}{=} \left\{\mathbf{a} = (a_1, \ldots, a_d) : \sum_{\nu=1}^d |a_\nu|^\delta \le \pi(d) \text{ and } \sum_{\nu=1}^d |a_\nu|^2 = 1\right\}, \tag{3.5}$$

where $\pi(d)$ is a deterministic function of d that diverges slowly such as $\log d$. Sparsity conditions are often employed in high-dimensional statistics, including sparse covariance matrix estimation (Bickel et al., 2008; Cai and Liu, 2011; Cai and Zhou, 2012), sparse integrated volatility matrix estimation (Kim et al., 2018, 2016; Tao et al., 2013a,b; Wang and Zou, 2010), and sparse PCA (Birnbaum et al., 2013; Kim and Wang, 2016; Johnstone and Lu, 2009; Ma, 2013; Vu and Lei, 2013; Vu et al., 2013).

The orthogonal iteration may be used to compute the leading eigenspace of a given Hermitian matrix (Golub and Van Loan, 1996), which yields the ordinary PCA estimator. As we have shown in Section 3.2, the ordinary PCA estimator has the convergence rate of $n^{-1/2}$. However, the ordinary PCA approach may not be the best for the sparse eigenvector estimation in terms of mean squared error (MSE). To obtain better eigenspace estimators under the sparsity condition in (3.5), we employ iterative thresholding algorithm known as the iterative thresholding sparse PCA (ITSPCA) proposed by Ma (2013) and described in Algorithm 1.

As presented in Algorithm 1, the ITSPCA method has three steps: multiplication, thresholding, and QR factorization. Without the thresholding step, the ITSPCA method returns to the ordinary orthogonal iteration method. The thresholding step removes weak signal elements of $\mathbf{T}^{(k)}$ with a user-specified thresholding function $\mathcal{T}$ which satisfies

$$|\mathcal{T}(t, \gamma) - t| \le \gamma \quad \text{and} \quad \mathcal{T}(t, \gamma)\mathbf{1}_{(|t|\le\gamma)} = 0 \quad \text{for all } t \text{ and all } \gamma > 0, \tag{3.6}$$

where $\mathbf{1}_E$ denotes the indicator function of an event E. We note that both hard thresholding rule $\mathcal{T}_H(t, \gamma) = t\,\mathbf{1}_{(|t|>\gamma)}$ and soft thresholding rule $\mathcal{T}_S(t, \gamma) = e^{\sqrt{-1}\theta}\max(0, |t| - \gamma)$ satisfy (3.6), where $t = |t|e^{\sqrt{-1}\theta}$, and θ is the phase of complex number t.

To harness the ITSPCA algorithm in Algorithm 1, we need an appropriate initial orthonormal matrix $\widehat{\mathbf{Q}}^{(0)}$. Johnstone and Lu (2009) introduced a diagonal thresholding sparse PCA (DTSPCA) method to estimate the eigenspace and showed its consistency. We propose to use the DTSPCA described in Algorithm 2 to obtain $\widehat{\mathbf{Q}}^{(0)}$. Given the output $\widehat{\mathbf{Q}}_S = (\widehat{\mathbf{q}}_1, \ldots, \widehat{\mathbf{q}}_{|S|})$, we may take its first m columns as the initial orthogonal matrix $\widehat{\mathbf{Q}}^{(0)} = (\widehat{\mathbf{q}}_1, \ldots, \widehat{\mathbf{q}}_m)$ for Algorithm 1. We choose $c_q = 0.001$ in the numerical study.

Algorithm 1 Iterative thresholding sparse PCA (ITSPCA)

Input:

(1) Estimated density matrix $\widehat{\boldsymbol{\rho}}$;
(2) Target subspace dimension m;
(3) Initial orthonormal matrix $\widehat{\mathbf{Q}}^{(0)}$;
(4) Thresholding function $\mathcal{T}(t, \gamma)$, and threshold levels $\gamma_{nj}, j = 1, \ldots, m$.

1: **repeat**
2: Multiplication: $\mathbf{T}^{(k)} = (t_{vj}^{(k)}) = \widehat{\boldsymbol{\rho}}\widehat{\mathbf{Q}}^{(k-1)}$;
3: Thresholding: $\widehat{\mathbf{T}}^{(k)} = (\widehat{t}_{vj}^{(k)})$, with $\widehat{t}_{vj}^{(k)} = \mathcal{T}(t_{vj}^{(k)}, \gamma_{nj})$;
4: QR factorization: $\widehat{\mathbf{Q}}^{(k)}\widehat{\mathbf{R}}^{(k)} = \widehat{\mathbf{T}}^{(k)}$;
5: **until** $\|\widehat{\mathbf{Q}}^{(k)} - \widehat{\mathbf{Q}}^{(k-1)}\|_F \le c_q$ for some pre-chosen small c_q.

Algorithm 2 Diagonal thresholding sparse PCA (DTSPCA)

Input:

(1) Estimated density matrix $\widehat{\boldsymbol{\rho}}$;
(2) Diagonal thresholding parameter α_n.

Output: Orthonormal matrix $\widehat{\mathbf{Q}}_S$.

1: Selection: select the set S of coordinates:

$$S = \{\nu : \widehat{\rho}_{\nu\nu} \ge \alpha_n\};$$

2: Reduced PCA: compute the eigenvectors, $\widehat{\mathbf{q}}_1^S, \ldots, \widehat{\mathbf{q}}_{|S|}^S$, of the submatrix $\widehat{\boldsymbol{\rho}}_{SS}$;
3: Zero-padding: construct $\widehat{\mathbf{Q}}_S = (\widehat{\mathbf{q}}_1, \ldots, \widehat{\mathbf{q}}_{|S|})$ such that

$$\widehat{\mathbf{q}}_{jS} = \widehat{\mathbf{q}}_j^S, \quad \widehat{\mathbf{q}}_{jS^c} = 0, \quad j = 1, \ldots, |S|.$$

4. Asymptotic theory for the eigenspace estimation

4.1. Convergence rates of PCA estimators

Assume that density matrix $\boldsymbol{\rho}$ belongs to the following class,

$$\mathcal{F}_\delta(\pi(d)) = \left\{ \boldsymbol{\rho} = \sum_{\nu=1}^{r} \lambda_\nu \mathbf{q}_\nu \mathbf{q}_\nu^\dagger : \mathbf{q}_\nu \in \Xi_\delta(\pi(d)) \text{ for all } \nu \in \{1, \ldots, r\} \right\}, \tag{4.1}$$

where $\Xi_\delta(\pi(d))$ is defined in (3.5). For the eigenspace $\mathbf{Q}$ of the density matrix $\boldsymbol{\rho}$, we consider the ITSPCA estimator $\widehat{\mathbf{Q}}^{(R_s)}$ where R_s is the number of iterations after which Algorithm 1 stops for a theoretical study where

$$R_s = \left\lceil \frac{1.1\ell_1^S}{\ell_m^S - \ell_{m+1}^S} (\log n + 0.5 \log(d \vee n)) \right\rceil,$$

$\lceil\ \rceil$ denotes ceiling, $\ell_j^S = \ell_j(\widehat{\boldsymbol{\rho}}_{SS}) \vee 0$, and $\ell_j(\widehat{\boldsymbol{\rho}}_{SS})$ is the jth largest eigenvalue of $\widehat{\boldsymbol{\rho}}_{SS}$.

The following theorem establishes the convergence rate of the eigenspace estimator $\widehat{\mathbf{Q}}^{(R_s)}$ obtained from Algorithm 1.

Theorem 2. *Assume the density matrix $\boldsymbol{\rho}$ given by model (3.1) belongs to $\mathcal{F}_\delta(\pi(d))$ defined in (4.1) so that for some $\delta \in (0, 2/3)$,*

$$\pi(d) \le C\tau_n^{3\delta/4 - 1/2}, \tag{4.2}$$

where $\tau_n = \sqrt{\frac{\log(d \vee n)}{nd}}$, and C is a constant free of d and n. Take $\alpha_n = C_\alpha \tau_n$ in Algorithm 2 and $\gamma_{nj} = C_\gamma \sqrt{\ell_j^S} \tau_n$ in Algorithm 1 for some constant C_α and C_γ free of n and d, and let

$$R = \left\lceil \frac{\lambda_1}{\lambda_m - \lambda_{m+1}} (\log n + 0.5 \log(d \vee n)) \right\rceil.$$

Then there exist constants C_0 and C_u such that for $(n, d, \pi(d))$ satisfying (4.2), uniformly over $\mathcal{F}_\delta(\pi(d))$, with probability at least $1 - C_0(d \vee n)^{-2}$, and $R_s \in [R, 2R]$, we have

$$\| \sin(\mathbf{Q}, \widehat{\mathbf{Q}}^{(R_s)})\|_2^2 \le \| \sin(\mathbf{Q}, \widehat{\mathbf{Q}}^{(R_s)})\|_F^2 \le C_u \pi(d) \tau_n^{2-\delta}.$$

Remark 2. Due to the proof techniques used we impose the restriction $\delta \in (0, 2/3)$ in Theorem 2. The restriction is largely due to the facts that we need to represent large density matrices by Pauli matrices and handle Pauli representation structures and approximation errors of the truncated representation. We expect the restriction can be relaxed or even removed in future.

Remark 3. The sparsity condition (4.2) is required to obtain the consistency of the proposed estimator. Similar conditions are often imposed by asymptotic analysis in high dimensional statistics such as large covariance matrix estimation where $\pi(d)$ usually grows very slowly in d with an example of $\log d$. Condition (4.2) is not very restrictive in the sense that $\pi(d) = \log d$ satisfies the condition, and in fact, when $\delta < 2/3$, the condition (4.2) indicates that $\pi(d)$ is at most of order d with some positive power.

The result of Theorem 2 can be extended to an upper bound for the MSE. Note that $(d \vee n)^{-2} = o\left(\pi(d)\tau_n^{2-\delta}\right)$ and the loss functions, (3.2) and (3.3), are bounded by r and 1, respectively. The following corollary is a direct consequence of Theorem 2.

Corollary 1. *Under the conditions of Theorem 2, we have*

$$\sup_{\rho \in \mathcal{F}_\delta(\pi(d))} E\left[\| \sin(\mathbf{Q}, \widehat{\mathbf{Q}}^{(R_s)})\|_2^2\right] \le \sup_{\rho \in \mathcal{F}_\delta(\pi(d))} E\left[\| \sin(\mathbf{Q}, \widehat{\mathbf{Q}}^{(R_s)})\|_F^2\right] \le C_u \pi(d) \tau_n^{2-\delta}.$$

Remark 4. Since in high-dimensional statistics, there is a constant C free of n and d such that $\log(d \vee n) \le C \log d$, Theorem 2 and Corollary 1 show that the ITSPCA estimator has the convergence rate of $\pi(d)^{1/2}\left[n^{-1}d^{-1}\log d\right]^{1/2-\delta/4}$ under the Frobenius and spectral norms. As d is often much larger than n, this convergence rate is faster than $n^{-1/2}$, the convergence rate in the ordinary PCA case.

Although our main objective is to estimate the eigenspace, when individual eigenvector $\mathbf{q}_k$ is identifiable, it is interesting to examine whether the ITSPCA method can estimate $\mathbf{q}_k$ well. The corollary below shows that the kth column of $\widehat{\mathbf{Q}}^{(R_s)}$ provides a good estimator for the well-separated $\mathbf{q}_k$.

Corollary 2. *Suppose that for some $k \le m$, $\lambda_k - \lambda_{k+1} \ge C_{\lambda 1}$ and $\lambda_{k-1} - \lambda_k \ge C_{\lambda 2}$ for some positive constants $C_{\lambda 1}$ and $C_{\lambda 2}$ free of n and d. Under the conditions of Theorem 2, we have that the kth column $\widehat{\mathbf{q}}_k^{(R_s)}$ of $\widehat{\mathbf{Q}}^{(R_s)}$ satisfies*

$$\sup_{\rho \in \mathcal{F}_\delta(\pi(d))} E\left[\| \sin(\mathbf{q}_k, \widehat{\mathbf{q}}_k^{(R_s)})\|_2^2\right] \le \sup_{\rho \in \mathcal{F}_\delta(\pi(d))} E\left[\| \sin(\mathbf{q}_k, \widehat{\mathbf{q}}_k^{(R_s)})\|_F^2\right] \le C_u \pi(d) \tau_n^{2-\delta}.$$

4.2. Optimality of ITSPCA estimator

This section establishes the minimax lower bound for the problem of estimating the eigenspace spanned by $\mathbf{Q}$ under model (3.1), uniformly over $\mathcal{F}_\delta(\pi(d))$, and shows that the ITSPCA estimator achieves the minimax lower bound, and thus its convergence rate is optimal.

The theorem below provides a minimax lower bound for eigenspace estimation under the Frobenius and spectral norms.

Theorem 3. *For model (3.1), suppose that for some $\delta \in [0, 2)$, as $d, n \to \infty$,*

$$\pi(d) \asymp d^{(1-\delta/2)-\mathcal{N}} n^{-\delta/2} \log^{\delta/2} d, \tag{4.3}$$

where $\mathcal{N} \in (0, 1)$ is a constant free of n and d. Then there exists a positive constant C_L free of n and d such that for $(n, d, \pi(d))$ satisfying (4.3),

$$\inf_{\check{\mathbf{Q}}} \sup_{\rho \in \mathcal{F}_\delta(\pi(d))} E\left[\| \sin(\mathbf{Q}, \check{\mathbf{Q}})\|_2^2\right] \ge C_L \pi(d) \left[\frac{\log d}{nd}\right]^{1-\delta/2},$$

and

$$\inf_{\check{\mathbf{Q}}} \sup_{\rho \in \mathcal{F}_\delta(\pi(d))} E\left[\| \sin(\mathbf{Q}, \check{\mathbf{Q}})\|_F^2\right] \ge C_L \pi(d) \left[\frac{\log d}{nd}\right]^{1-\delta/2},$$

where $\check{\mathbf{Q}}$ denotes any estimator of $\mathbf{Q}$ based on $N_2, \ldots, N_p$.

Remark 5. The lower bound in Theorem 3 matches the convergence rate of the ITSPCA estimator in Theorem 2, and so we conclude that the ITSPCA estimator achieves the optimal convergence rate under the Frobenius and spectral norms (especially when $d \geq n$). That is, under the sparsity condition, the convergence rate, $\pi(d)^{1/2}\left[n^{-1}d^{-1}\log d\right]^{1/2-\delta/4}$, of the ITSPCA estimator is optimal, while the convergence rate, $n^{-1/2}$, of the ordinary PCA estimator is sub-optimal. On the other hand, without the sparsity assumption on the eigenspace, that is, $\pi(d) = d$ and $\delta = 0$, we can show that the minimax lower bound for estimating the eigenspace of $\boldsymbol{\rho}$ is $n^{-1/2}$. Thus, the upper bound of the ordinary PCA estimator in Theorem 1 is the optimal rate for the dense eigenspace case.

Remark 6. Cai et al. (2016) investigated the optimality of the density matrix estimation in the usual matrix sparsity framework. This paper considers the estimation of large low-rank density matrices and studies the associated optimality of eigenspace estimation. Thus, to derive the lower bound in Theorem 3, we consider a special subclass of $\mathbf{Q}$, and take $\boldsymbol{\rho} = m^{-1}\mathbf{Q}\mathbf{Q}^{\dagger}$, then as usual we apply Fano's lemma to obtain the minimax lower bound (Birnbaum et al., 2013; Vu and Lei, 2013). The key difference between our approach and those in the literature is that our observations are characterized by binomial distributions instead of the usual normal distributions, as a result, different proof arguments are needed to obtain the minimax lower bounds (see Section 7 for more details).

Remark 7. When $\delta = 0$, condition (4.3) becomes $\pi(d) \asymp d^{1-\mathcal{N}}$ with $\mathcal{N} > 0$, and the minimax lower bounds hold for $\pi(d)$ very close to d. Consider $\delta > 0$, and that d typically grows polynomially or exponentially in n. If d grows exponentially in n, that is, $d = e^{n^{\kappa}}$, then $\pi(d) \asymp d^{(1-\delta/2)-\mathcal{N}}(\log d)^{\delta/2(1-1/\kappa)}$, and $\mathcal{N}$ could be chosen very small value such that $\pi(d)$ is of order d with some positive power. In the case of $d = n^{\kappa}$, as quantum systems often have large d, we may consider the case of $d \geq n$ and take $\kappa \geq 1$, and thus $\pi(d) \asymp d^{(1-\delta/2)-\mathcal{N}-\delta/(2\kappa)}(\log d)^{\delta/2}$, which is of order d with some positive power. Therefore, condition (4.3) is feasible. Also the conditions (4.2) and (4.3) are compatible under reasonable setting such as that $[\log d/(nd)]^{\delta/8+1/4}d^{1-\mathcal{N}}$ is bounded by a generic constant. Of course the condition (4.3) can be relaxed, and future work may make it less restrictive.

Remark 8. Koltchinskii and Xia (2015) investigated the optimal convergence rate for estimating a low-rank density matrix belonging to a general low-rank density matrix class. For example, under the Pauli basis, Theorem 10 in Koltchinskii and Xia (2015) shows that the optimal rate of estimating low-rank density matrices is $n^{-1/2}$ which we can obtain by the ordinary PCA estimator (see Theorem 5 in Section 5). Their low-rank class includes both the dense and sparse cases, and thus the minimax rate is determined by the sub-class with dense eigenvectors. However, as we have shown in Theorems 3 and 4, the rate $n^{-1/2}$ is not optimal under the sparse condition (3.5). Also their analysis focused on estimating a low-rank density matrix itself. On the other hand, this paper is devoted to investigating the eigenspace estimation particularly under the sparse condition (3.5). Our analysis implies that the optimal rate of estimating low-rank density matrices under the sparse condition is $\pi(d)^{1/2}\left[n^{-1}d^{-1}\log d\right]^{1/2-\delta/4}$ (see Theorem 4 in Section 5).

5. Large low-rank density matrix estimation

This section proposes low-rank density matrix estimators using the ordinary PCA and the ITSPCA methods. We first develop estimators for eigenvalues of the low-rank density matrix $\boldsymbol{\rho}$ as follows:

$$\widehat{\lambda}_{\nu}^{(R_s)} = \frac{\widetilde{\lambda}_{\nu}^{(R_s)}}{\sum_{j=1}^{r}\widetilde{\lambda}_{j}^{(R_s)}} \quad \text{and} \quad \widehat{\lambda}_{\nu}^{*} = \frac{\widetilde{\lambda}_{\nu}}{\sum_{j=1}^{r}\widetilde{\lambda}_{j}} \quad \text{for } \nu = 1, \ldots, r,$$

where

$$\widetilde{\lambda}_{\nu}^{(R_s)} = \max\left[(\widehat{\mathbf{q}}_{\nu}^{(R_s)})^{\dagger}\widehat{\boldsymbol{\rho}}\,\widehat{\mathbf{q}}_{\nu}^{(R_s)}, 0\right], \qquad \widetilde{\lambda}_{\nu} = \max\left[\widehat{\mathbf{q}}_{\nu}^{\dagger}\widehat{\boldsymbol{\rho}}\,\widehat{\mathbf{q}}_{\nu}, 0\right],$$

and $\widehat{\mathbf{q}}_{\nu}^{(R_s)}$ and $\widehat{\mathbf{q}}_{\nu}$ are the νth column of $\widehat{\mathbf{Q}}^{(R_S)}$ and $\widehat{\mathbf{Q}}$, respectively. Note that $\widehat{\lambda}_{\nu}^{(R_s)}$ and $\widetilde{\lambda}_{\nu}$ are non-negative, and the sum of each set of estimated eigenvalues is 1. Using the eigenvalue and eigenspace estimators, we can reconstruct the low-rank density matrix as follows:

$$\widehat{\boldsymbol{\rho}}^{(R_s)} = \sum_{\nu=1}^{r}\widehat{\lambda}_{\nu}^{(R_s)}\widehat{\mathbf{q}}_{\nu}^{(R_s)}(\widehat{\mathbf{q}}_{\nu}^{(R_s)})^{\dagger} \quad \text{and} \quad \widehat{\boldsymbol{\rho}}^{*} = \sum_{\nu=1}^{r}\widehat{\lambda}_{\nu}^{*}\widehat{\mathbf{q}}_{\nu}\widehat{\mathbf{q}}_{\nu}^{\dagger}.$$

These two estimators are well-defined density matrices given in Section 2.1.

The following theorems provide the convergence rates of eigenvalue estimators $\widehat{\lambda}_{\nu}^{(R_s)}$ and $\widehat{\lambda}_{\nu}^{*}$, and low-rank density matrix estimators $\widehat{\boldsymbol{\rho}}^{(R_s)}$ and $\widehat{\boldsymbol{\rho}}^{*}$.

Theorem 4. *Under the assumptions of Theorem 2 for the ITSPCA, we have for $\nu = 1, \ldots, r$,*

$$E\left[|\widehat{\lambda}_{\nu}^{(R_s)} - \lambda_{\nu}|\right] \leq C\pi(d)^{1/2}\tau_n^{1-\delta/2} \tag{5.1}$$

and

$$E\left[\|\widehat{\boldsymbol{\rho}}^{(R_S)} - \boldsymbol{\rho}\|_F\right] \le C\pi(d)^{1/2}\tau_n^{1-\delta/2}, \tag{5.2}$$

where C is a generic constant free of n and d.

Theorem 5. *Under the assumptions of Theorem 1 for the ordinary PCA, we have for* $\nu = 1, \ldots, r$,

$$E\left[|\widehat{\lambda}_\nu^* - \lambda_\nu|\right] \le C(n^{-1} \vee (nd)^{-1/2}) \quad \textit{and} \quad E\left[\|\widehat{\boldsymbol{\rho}}^* - \boldsymbol{\rho}\|_F\right] \le Cn^{-1/2},$$

where C is a generic constant free of n and d.

Remark 9. When $\delta = 0$, the convergence rate for $E\left[\|\widehat{\boldsymbol{\rho}}^{(R_S)} - \boldsymbol{\rho}\|_F\right]$ is $\pi(d)^{1/2}d^{-1/2}\left(\frac{\log(d\vee n)}{n}\right)^{1/2}$, which is the same as the convergence rate of the optimal density matrix estimator under the sparse representation in Theorem 1 of Cai et al. (2016). Also under the sparse condition (3.5), the minimax lower bound of estimating low-rank density matrices is $\pi(d)^{1/2}\tau_n^{1-\delta/2}$, which can be established using the same sub-class employed in the proof of Theorem 3.

Remark 10. The threshold density matrix estimator has the convergence rate $(d/n)^{1/2}$ under the Frobenius norm (see Lemma 3 of Cai et al. (2016)). On the other hand, the low-rank density matrix estimator has convergence rate $n^{-1/2}$ which is the optimal rate given the general low-rank density matrix class (Koltchinskii and Xia, 2015).

Remark 11. The proposed low-rank density matrix estimation procedure needs to know the true rank r. In practice r is unknown, and we may estimate r from the data to implement the procedure. For example, Kim and Wang (2017) established asymptotic distributions for the eigenvectors of density matrices, and developed some preliminary methods to choose the rank r. This paper focuses on the sparse eigenvector estimation with known r. We may investigate the selection of rank r in the future study.

6. A numerical study

We conducted simulations to check the finite sample performance of the proposed estimators in Sections 6.1 and 6.2, and studied their empirical performance in Section 6.3.

6.1. Simulation for a rank one case

We first considered the case where the density matrix $\boldsymbol{\rho}$ in (3.1) has $r = 1$, and

$$\boldsymbol{\rho} = \mathbf{Q}\mathbf{Q}^\dagger = d^{-1}\left(\mathbf{I}_d + \sum_{j=2}^{p}\beta_j\mathbf{B}_j\right),$$

where $\mathbf{Q} \in \mathbb{C}^d$ and $\beta_j = \mathrm{tr}(\boldsymbol{\rho}\mathbf{B}_j)$ for $j = 1, \ldots, d^2$. The eigenvectors $\mathbf{Q}$ were generated as follows: First, its $\pi(d)$ components were generated by $\pi(d)$ i.i.d. random variables from $U_1 + U_2\sqrt{-1}$, where U_j's are i.i.d. uniform distributions on $[-1, 1]$, and the rest $d - \pi(d)$ components were set to be zero. The generated vector was then normalized by dividing its ℓ_2-norm so that the generated $\mathbf{Q}$ satisfies $\|\mathbf{Q}\|_2 = 1$. We varied $\pi(d)$ from $5\log(d)$ to $d - 1$ with $d = 64{,}128$. The whole procedure was repeated for 200 times.

For each simulated dataset, we estimated $\mathbf{Q}$ using the ITSPCA with hard threshold (ITS-H), ITSPCA with soft threshold (ITS-S), DTSPCA, and ordinary PCA algorithms. The MSEs of the eigenspace estimator $\widehat{\mathbf{Q}}$ and low-rank density matrix estimator $\widehat{\boldsymbol{\rho}}$, $E\|\sin(\widehat{\mathbf{Q}}, \mathbf{Q})\|_F^2$ and $E\|\widehat{\boldsymbol{\rho}} - \boldsymbol{\rho}\|_F^2$, were calculated by averaging the corresponding squared norms of $\widehat{\mathbf{Q}}$ and $\widehat{\boldsymbol{\rho}}$ over 200 simulations. For the ITSPCA and DTSPCA algorithms in Algorithms 1 and 2, respectively, we set tuning parameters (C_α, C_γ) to be (0.1, 2), (0.5, 1), and (0, 1) for the ITS-H, the ITS-S, and the DTSPCA, respectively, by searching in the range of $\{3, 2.5, \ldots, 0.5, 0.1\}^2$ for minimizing MSE. We used hard thresholding rule $\mathcal{T}_H(t, \gamma) = t\mathbf{1}_{(|t|>\gamma)}$ and soft thresholding rule $\mathcal{T}_S(t, \gamma) = e^{\sqrt{-1}\theta}\max(0, |t| - \gamma)$ for the thresholding step in Algorithm 1 for the ITS-H and ITS-S, respectively, where $t = |t|e^{\sqrt{-1}\theta}$. We stopped iterating once $\|\sin(\widehat{\mathbf{Q}}^{(k)}, \widehat{\mathbf{Q}}^{(k-1)})\|_2 \le n^{-1}d^{-1}$.

Table 1 summarizes the MSEs for the eigenspace and density matrix estimators. Regarding the eigenspace estimators, Fig. 1 plots the MSEs against $\pi(d)$ for different n and d values while Fig. 2 plots the relative efficiencies of the ITS-H, ITS-S, DTSPCA, and PCA estimators with respect to the PCA estimator against the sample size n for different d and $\pi(d)$ values. The numerical results show that the MSEs usually decrease in sample size n. The MSEs of the ITSPCA and DTSPCA estimators become worse as $\pi(d)$ increases while the performance of the PCA estimator is robust against $\pi(d)$. For the sparse eigenvectors with $\pi(d) = 5\log d$ or $5d^{1/2}$, the ITSPCA estimators often have superior performance over the DTSPCA and PCA estimators while for the non-sparse case with $\pi(d) = d - 1$, the PCA estimator overall presents the best

Table 1
The MSEs of ITS-H, ITS-S, DTSPCA, and PCA estimators and their corresponding low-rank density matrix estimators when $d = 64, 128$ and $n = 100, 200, 500, 1000, 2000$ (we make the smallest MSE bold).

d	$\pi(d)$	n	MSE (eigenspace) $\times 10^2$				MSE (density matrix) $\times 10^2$				
			ITS-H	ITS-S	DTSPCA	PCA	ITS-H	ITS-S	DTSPCA	PCA	$\widehat{\rho}$
64	$5\log(d)$	100	**0.3008**	0.4383	1.1629	0.9627	**0.6016**	0.8767	2.3258	1.9254	63.0741
		200	**0.1453**	0.2155	0.5489	0.4909	**0.2906**	0.4310	1.0977	0.9819	31.5377
		500	**0.0577**	0.0878	0.2251	0.1929	**0.1153**	0.1757	0.4501	0.3858	12.6377
		1000	**0.0284**	0.0443	0.1163	0.0948	**0.0568**	0.0885	0.2327	0.1895	6.3216
		2000	**0.0142**	0.0224	0.0438	0.0485	**0.0284**	0.0447	0.0877	0.0970	3.1531
	$5d^{1/2}$	100	**0.8961**	1.0046	4.8367	0.9680	**1.7921**	2.0093	9.6733	1.9359	63.0148
		200	**0.4082**	0.5295	2.4086	0.4855	**0.8165**	1.0590	4.8173	0.9709	31.5230
		500	**0.1347**	0.2129	1.0878	0.1941	**0.2693**	0.4259	2.1756	0.3882	12.6007
		1000	**0.0593**	0.1085	0.4885	0.0961	**0.1186**	0.2170	0.9771	0.1922	6.2871
		2000	**0.0304**	0.0576	0.2433	0.0484	**0.0607**	0.1151	0.4866	0.0969	3.1456
	$d-1$	100	**1.3800**	1.5670	12.3886	0.9876	**2.7600**	3.1340	24.7772	1.9751	63.0856
		200	**0.6485**	0.7903	6.3141	0.4871	**1.2970**	1.5806	12.6282	0.9742	31.5273
		500	**0.1952**	0.3325	2.4912	0.1930	**0.3905**	0.6649	4.9825	0.3860	12.5754
		1000	**0.0957**	0.1713	1.3645	0.0971	**0.1914**	0.3427	2.7290	0.1943	6.3095
		2000	**0.0485**	0.0881	0.5835	0.0493	**0.0970**	0.1762	1.1671	0.0985	3.1518
128	$5\log(d)$	100	**0.2084**	0.3270	1.4792	0.9979	**0.4169**	0.6539	2.9584	1.9958	127.2219
		200	**0.0883**	0.1600	0.6500	0.4916	**0.1765**	0.3200	1.3001	0.9832	63.4730
		500	**0.0360**	0.0691	0.2996	0.1954	**0.0721**	0.1382	0.5991	0.3909	25.3742
		1000	**0.0182**	0.0359	0.1593	0.0990	**0.0364**	0.0718	0.3186	0.1979	12.6988
		2000	**0.0090**	0.0188	0.0775	0.0498	**0.0180**	0.0377	0.1550	0.0995	6.3408
	$5d^{1/2}$	100	**0.5461**	0.7029	4.8951	0.9819	**1.0921**	1.4058	9.7902	1.9638	126.9431
		200	**0.2097**	0.3514	2.2088	0.4856	**0.4195**	0.7029	4.4177	0.9711	63.5284
		500	**0.0841**	0.1468	0.9327	0.1968	**0.1682**	0.2937	1.8654	0.3936	25.4182
		1000	**0.0430**	0.0760	0.5597	0.0988	**0.0860**	0.1519	1.1195	0.1977	12.7027
		2000	**0.0216**	0.0388	0.2378	0.0496	**0.0433**	0.0776	0.4756	0.0993	6.3566
	$d-1$	100	1.6628	**1.4878**	19.1328	0.9926	3.3256	**2.9755**	38.2656	1.9851	127.0068
		200	**0.7028**	0.7814	11.4092	0.4903	1.4056	1.5629	22.8184	**0.9807**	63.5389
		500	**0.2354**	0.3478	5.1377	0.1964	0.4708	0.6957	10.2754	**0.3929**	25.4193
		1000	**0.1144**	0.1838	2.5418	0.0979	0.2287	0.3676	5.0837	**0.1957**	12.7010
		2000	**0.0553**	0.0962	1.2805	0.0488	0.1106	0.1924	2.5610	**0.0976**	6.3562

performance. The numerical results of the density matrix estimators summarized in Table 1 show similar behaviors to the results of the eigenspace estimators, while $\widehat{\rho}$ in (2.1) presents much worse performance than the PCA type estimators.

6.2. Simulation for a rank four case

We now simulated density matrix ρ using (3.1) with $r = 4$ and chose arbitrary eigenspace $\mathbf{Q}_0 \in \mathbb{V}_{\pi(d),4}$, where $\mathbb{V}_{h,k}$ is the Stiefel manifold of h-by-k orthonormal matrices. First, we generated a $\pi(d)$-by-$\pi(d)$ positive definite Hermitian matrix from uniform random variables, to be specific, the diagonal elements of the matrix took value 1 and for the off-diagonal elements, the (h, k)th and (k, h)th elements were $U_1 + \sqrt{-1}U_2$, where U_i's follow uniform distributions on $(-\sqrt{0.5}, \sqrt{0.5})$. We then formed d-by-4 matrix $\mathbf{Q} = (\mathbf{Q}_0^T, 0)^T$. Eigenvalues $\mathbf{\Lambda}$ were chosen from $(0.25, 0.25, 0.25, 0.25)$, $(0.4, 0.3, 0.2, 0.1)$, $(0.5, 0.3, 0.19, 0.01)$. The density matrix ρ was obtained in the following:

$$\rho = \sum_{\nu=1}^{4} \lambda_\nu \mathbf{q}_\nu \mathbf{q}_\nu^\dagger = d^{-1} \left(\mathbf{I}_d + \sum_{j=2}^{d^2} \beta_j \mathbf{B}_j \right),$$

where $\mathbf{Q} = (\mathbf{q}_1, \ldots, \mathbf{q}_4)$ and $d = 2^7$. With the ρ above, we computed $\beta_j = \mathrm{tr}(\rho \mathbf{B}_j)$ for $j = 1, \ldots, 2^{14}$, where $\mathbf{B}_j$'s are Pauli matrices. For each simulated dataset, we estimated $\mathbf{Q}$ for $m = 4$ and used the same scheme as the rank one case.

Table 2 summarizes the MSEs for the eigenspace and density matrix estimators and Fig. 3 plots the MSEs of the eigenspace estimators against $\pi(d)$ for different sample size n and eigenvalues $\mathbf{\Lambda}$. Fig. 4 further plots the relative efficiencies of the ITS-H, ITS-S, DTSPCA, and PCA estimators with respect to the PCA estimator against the sample size n for different $\pi(d)$ and eigenvalues $\mathbf{\Lambda}$. We note that the effects of n and $\pi(d)$ levels are similar to the rank one case so that we focus on the effects of the magnitude of the fourth eigenvalue λ_4 in $\mathbf{\Lambda}$. The numerical results show that the magnitude of the λ_4 plays an important role. When λ_4 is large such as 0.25, the MSEs are relatively small, and our methods show better performance than the benchmarks. When λ_4 is small such as 0.01, all estimators present poor performance with large MSEs, and their relative efficiencies are close.

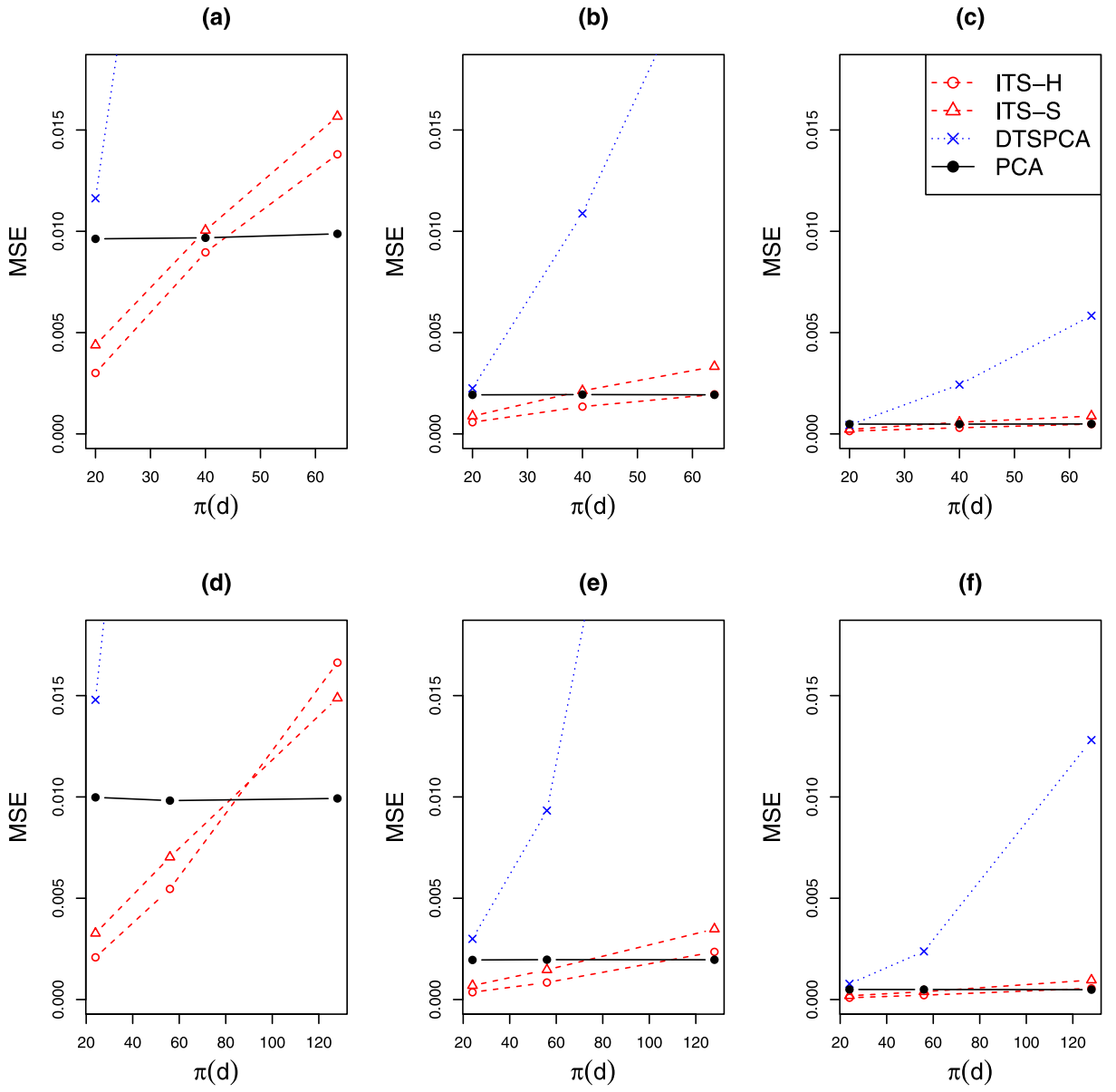

Fig. 1. Plots of MSEs against $\pi(d)$ for the ITS-H, ITS-S, DTSPCA, and PCA estimators with $n = 100, 500, 2000$ and $d = 64,128$. (a)–(c) are plots of MSEs based on the Frobenius norm for $n = 100, 500, 2000$, respectively, with $d = 64$. (d)–(f) are plots of MSEs based on the Frobenius norm for $n = 100, 500, 2000$, respectively, with $d = 128$.

6.3. A real data example

In this section, we conducted a Monte Carlo simulation to analyze the density matrices that were estimated by Häffner et al. (2005). We considered two density matrices with $d = 2^7$ and 2^8, and denoted them by $\boldsymbol{\rho}_7$ and $\boldsymbol{\rho}_8$, respectively. Based on each density matrix $\boldsymbol{\rho}$, we first calculated $\beta_j = \text{tr}(\boldsymbol{\rho}\mathbf{B}_j)$, where $\mathbf{B}_j$'s are the Pauli matrices and then, generated n Pauli measurements for each Pauli matrix. Given the generated Pauli measurements, we estimated $\boldsymbol{\rho}$ by the ITS-H, ITS-S, DTSPCA, and PCA. We selected tuning parameters (0.1, 2), (0.5, 1), and (0, 1) for the ITS-H, ITS-S, and DTSPCA estimators, respectively, and used the results of the rank test proposed in Kim and Wang (2017) to determine the rank r. We varied n from 100 to 2000 and repeated the whole procedure by 200 times.

Fig. 5 plots the absolute values for the elements of eigenvectors corresponding to the first six eigenvectors of $\boldsymbol{\rho}_7$ and $\boldsymbol{\rho}_8$, while the first six eigenvalues of $\boldsymbol{\rho}_7$ is (0.7825, 0.0605, 0.0445, 0.0324, 0.023, 0.0167) and of $\boldsymbol{\rho}_8$ is (0.7514, 0.0609, 0.0456, 0.04, 0.0233, 0.0189). Thus, the density matrices, $\boldsymbol{\rho}_7$ and $\boldsymbol{\rho}_8$, have the low-rank structure with sparse eigenvectors, which

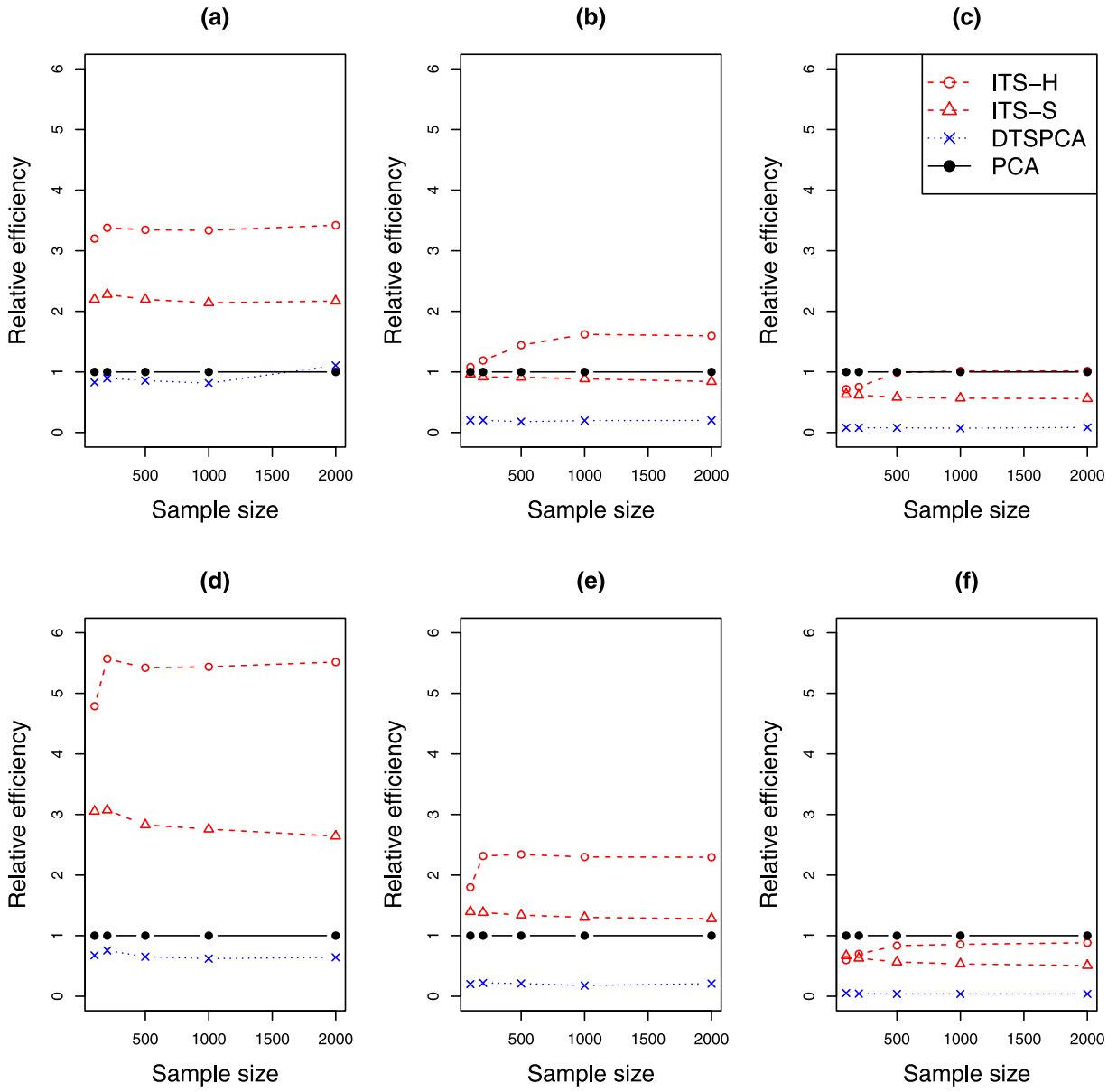

Fig. 2. Plots of relative efficiencies against the sample size n for the ITS-H, ITS-S, DTSPCA, and PCA estimators with respect to the PCA estimator for $\pi(d) = 5\log(d), 5d^{1/2}, d-1$ with $d = 64$ and 128. (a)–(c) are plots of relative efficiencies based on the Frobenius norm for $\pi(d) = 5\log(d), 5d^{1/2}, d-1$, respectively, with $d = 64$. (d)–(f) are plots of relative efficiencies based on the Frobenius norm for $\pi(d) = 5\log(d), 5d^{1/2}, d-1$, respectively, with $d = 128$.

satisfy the assumptions imposed in this paper. It follows that the iterative thresholding estimators such as the ITSPCA and the DTSPCA may present good performance.

Table 3 presents MSEs of the ITS-H, ITS-S, DTSPCA, and PCA density matrix estimators. Fig. 6 plots the relative efficiencies with respect to the PCA estimator against the sample size n and for $d = 128, 256$. From Table 3 and Fig. 6, we can see that the MSEs decrease in the sample size n while the iterative thresholding methods usually have smaller MSEs than the PCA density matrix estimator or the estimator $\widehat{\rho}$ in (2.1).

7. Proofs

Denote by C and C_1 the generic constants whose values are free of n and p and may change from appearance to appearance.

Table 2
The MSEs of the ITS-H, ITS-S, DTSPCA, and PCA estimators and their corresponding low-rank density matrix estimators for $n = 100, 200, 500, 1000, 2000$ and $\mathbf{\Lambda} = (0.25, 0.25, 0.25, 0.25)$, $(0.4, 0.3, 0.2, 0.1)$, $(0.5, 0.3, 0.19, 0.01)$, and $\pi(d) = 5\log d, 5d^{1/2}, d-1$ with $d = 128$ (we make the smallest MSE bold).

$\pi(d)$	$\mathbf{\Lambda}$	n	MSE (eigenspace)				MSE (density matrix)				
			ITS-H	ITS-S	DTSPCA	PCA	ITS-H	ITS-S	DTSPCA	PCA	$\widehat{\boldsymbol{\rho}}$
$5\log d$	(0.25, 0.25, 0.25, 0.25)	100	**0.1861**	0.1941	0.3532	0.6215	**0.0240**	0.0248	0.0448	0.0781	1.2792
		200	**0.0722**	0.0925	0.1724	0.3099	**0.0095**	0.0119	0.0220	0.0391	0.6386
		500	**0.0231**	0.0353	0.0648	0.1240	**0.0031**	0.0046	0.0083	0.0157	0.2555
		1000	**0.0109**	0.0175	0.0310	0.0622	**0.0015**	0.0023	0.0040	0.0079	0.1277
		2000	**0.0052**	0.0085	0.0151	0.0309	**0.0007**	0.0011	0.0019	0.0039	0.0638
	(0.4, 0.3, 0.2, 0.1)	100	0.5935	**0.5626**	0.7734	1.2608	**0.0309**	0.0345	0.0519	0.0902	1.2787
		200	**0.2444**	0.2491	0.3923	0.6903	**0.0128**	0.0150	0.0249	0.0436	0.6384
		500	**0.0846**	0.0903	0.1488	0.2758	**0.0044**	0.0055	0.0092	0.0165	0.2553
		1000	**0.0415**	0.0440	0.0721	0.1379	**0.0019**	0.0027	0.0044	0.0080	0.1276
		2000	**0.0195**	0.0211	0.0344	0.0687	**0.0009**	0.0013	0.0020	0.0040	0.0638
	(0.5, 0.3, 0.19, 0.01)	100	**1.1132**	1.1137	1.2272	1.4019	**0.0257**	0.0321	0.0527	0.0851	1.2782
		200	**1.0371**	1.0486	1.0997	1.1951	**0.0143**	0.0170	0.0275	0.0455	0.6378
		500	**1.0031**	1.0125	1.0219	1.0736	**0.0070**	0.0078	0.0117	0.0194	0.2552
		1000	**0.9887**	0.9975	0.9873	1.0310	**0.0040**	0.0044	0.0057	0.0100	0.1275
		2000	0.9741	0.9817	**0.9514**	1.0041	**0.0022**	0.0024	0.0029	0.0051	0.0637
$5d^{1/2}$	(0.25, 0.25, 0.25, 0.25)	100	0.6163	**0.4233**	0.5953	0.6201	0.0775	**0.0533**	0.0747	0.0779	1.2799
		200	0.2571	**0.2169**	0.3132	0.3085	0.0325	**0.0274**	0.0394	0.0389	0.6385
		500	**0.0813**	0.0882	0.1343	0.1239	**0.0104**	0.0112	0.0170	0.0157	0.2553
		1000	**0.0347**	0.0442	0.0683	0.0620	**0.0044**	0.0056	0.0086	0.0079	0.1276
		2000	**0.0152**	0.0220	0.0336	0.0308	**0.0020**	0.0028	0.0042	0.0039	0.0638
	(0.4, 0.3, 0.2, 0.1)	100	1.1828	**0.9932**	1.0824	1.2713	0.0779	**0.0631**	0.0839	0.0899	1.2790
		200	0.5289	**0.4585**	0.5603	0.6914	0.0329	**0.0306**	0.0420	0.0435	0.6382
		500	0.1773	**0.1701**	0.2300	0.2732	**0.0113**	0.0117	0.0182	0.0164	0.2553
		1000	**0.0815**	0.0840	0.1172	0.1369	**0.0050**	0.0058	0.0098	0.0080	0.1276
		2000	**0.0385**	0.0421	0.0622	0.0688	**0.0021**	0.0029	0.0054	0.0040	0.0638
	(0.5, 0.3, 0.19, 0.01)	100	1.3123	**1.2543**	1.3840	1.4003	0.1199	**0.1089**	0.1554	0.1630	1.3185
		200	**1.0565**	1.1265	1.1983	1.1961	**0.0685**	0.0723	0.1259	0.1260	0.6779
		500	**0.9264**	1.0494	1.0758	1.0763	**0.0493**	0.0508	0.1095	0.1074	0.2952
		1000	**0.8809**	1.0208	1.0348	1.0346	**0.0440**	0.0452	0.1082	0.1037	0.1676
		2000	**0.8658**	1.0051	1.0188	1.0145	**0.0412**	0.0422	0.1088	0.1035	0.1037
$d-1$	(0.25, 0.25, 0.25, 0.25)	100	1.5437	**0.8133**	1.1214	0.6187	0.1930	0.1020	0.1401	**0.0778**	1.2761
		200	0.6946	**0.4366**	0.6917	0.3079	0.0870	0.0549	0.0865	**0.0389**	0.6394
		500	0.2316	0.1886	0.3512	**0.1239**	0.0291	0.0237	0.0439	**0.0157**	0.2553
		1000	0.0996	0.0986	0.1968	**0.0618**	0.0125	0.0124	0.0246	**0.0078**	0.1278
		2000	0.0432	0.0503	0.1020	**0.0309**	0.0055	0.0063	0.0128	**0.0039**	0.0638
	(0.4, 0.3, 0.2, 0.1)	100	2.0636	1.4548	1.6928	**1.2713**	0.2020	0.1127	0.1640	**0.0901**	1.2750
		200	1.1570	0.8458	1.0975	**0.6977**	0.0902	0.0605	0.1019	**0.0437**	0.6392
		500	0.4181	0.3307	0.5206	**0.2738**	0.0287	0.0244	0.0507	**0.0165**	0.2552
		1000	0.1899	0.1702	0.2761	**0.1372**	0.0120	0.0124	0.0275	**0.0080**	0.1278
		2000	0.0875	0.0887	0.1406	**0.0698**	0.0050	0.0063	0.0139	**0.0040**	0.0637
	(0.5, 0.3, 0.19, 0.01)	100	1.9442	1.5199	1.7585	**1.3988**	0.1611	0.0965	0.1579	**0.0847**	1.2743
		200	1.4469	1.2706	1.4723	**1.1967**	0.0792	0.0535	0.1027	**0.0456**	0.6389
		500	1.1390	1.1081	1.2426	**1.0724**	0.0271	0.0241	0.0553	**0.0195**	0.2552
		1000	1.0472	1.0482	1.1359	**1.0301**	0.0123	0.0128	0.0330	**0.0099**	0.1277
		2000	1.0068	1.0109	1.0639	**1.0034**	0.0055	0.0068	0.0185	**0.0051**	0.0637

7.1. Proofs of Theorems 1–2

7.1.1. Proof of Theorem 1

Proof of Theorem 1. By Davis-Kahn's $\sin\theta$ theorem (Theorem 3.1 in Li (1998b)), we obtain the following inequality to establish (3.4),

$$\|\sin(\mathbf{Q}, \widehat{\mathbf{Q}})\|_F^2 \le \frac{\|(\boldsymbol{\rho} - \widehat{\boldsymbol{\rho}})\mathbf{Q}\|_F^2}{(\lambda_m - \widehat{\lambda}_{m+1})^2}, \tag{7.1}$$

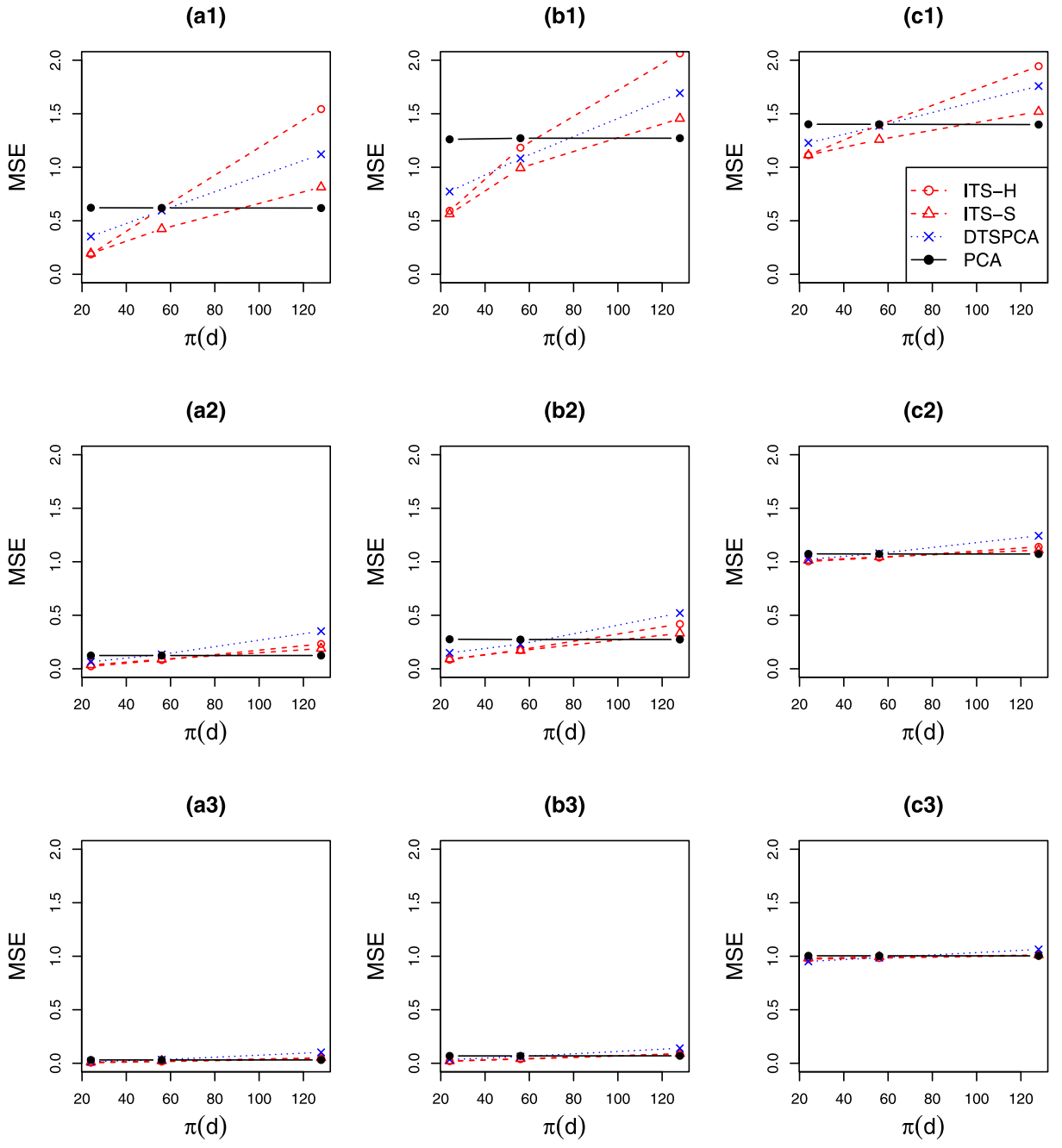

Fig. 3. Plots of MSEs against $\pi(d)$ for the ITS-H, ITS-S, DTSPCA, and PCA estimators for $n = 100, 500, 2000$ and $d = 128$. (a1)–(a3) are plots of MSEs based on the Frobenius norm for $n = 100, 500, 2000$, respectively, with $\mathbf{\Lambda} = (0.25, 0.25, 0.25, 0.25)$. (b1)–(b3) are plots of MSEs based on the Frobenius norm for $n = 100, 500, 2000$, respectively, with $\mathbf{\Lambda} = (0.4, 0.3, 0.2, 0.1)$. (c1)–(c3) are plots of MSEs based on the Frobenius norm for $n = 100, 500, 2000$, respectively, with $\mathbf{\Lambda} = (0.5, 0.3, 0.19, 0.01)$.

where $\widehat{\lambda}_m$ is the mth eigenvalue of $\widehat{\boldsymbol{\rho}}$. For the denominator on the right hand side of (7.1), as $\lambda_m - \lambda_{m+1}$ is bounded below from a generic constant C_λ, we need to study only $\widehat{\lambda}_{m+1} - \lambda_{m+1}$. By Weyl's theorem (Theorem 4.3 in Li (1998a)), we have

$$|\widehat{\lambda}_{m+1} - \lambda_{m+1}| \leq \max_{1 \leq \nu \leq d} |\widehat{\lambda}_\nu - \lambda_\nu| \leq \|\widehat{\boldsymbol{\rho}} - \boldsymbol{\rho}\|_2^2.$$

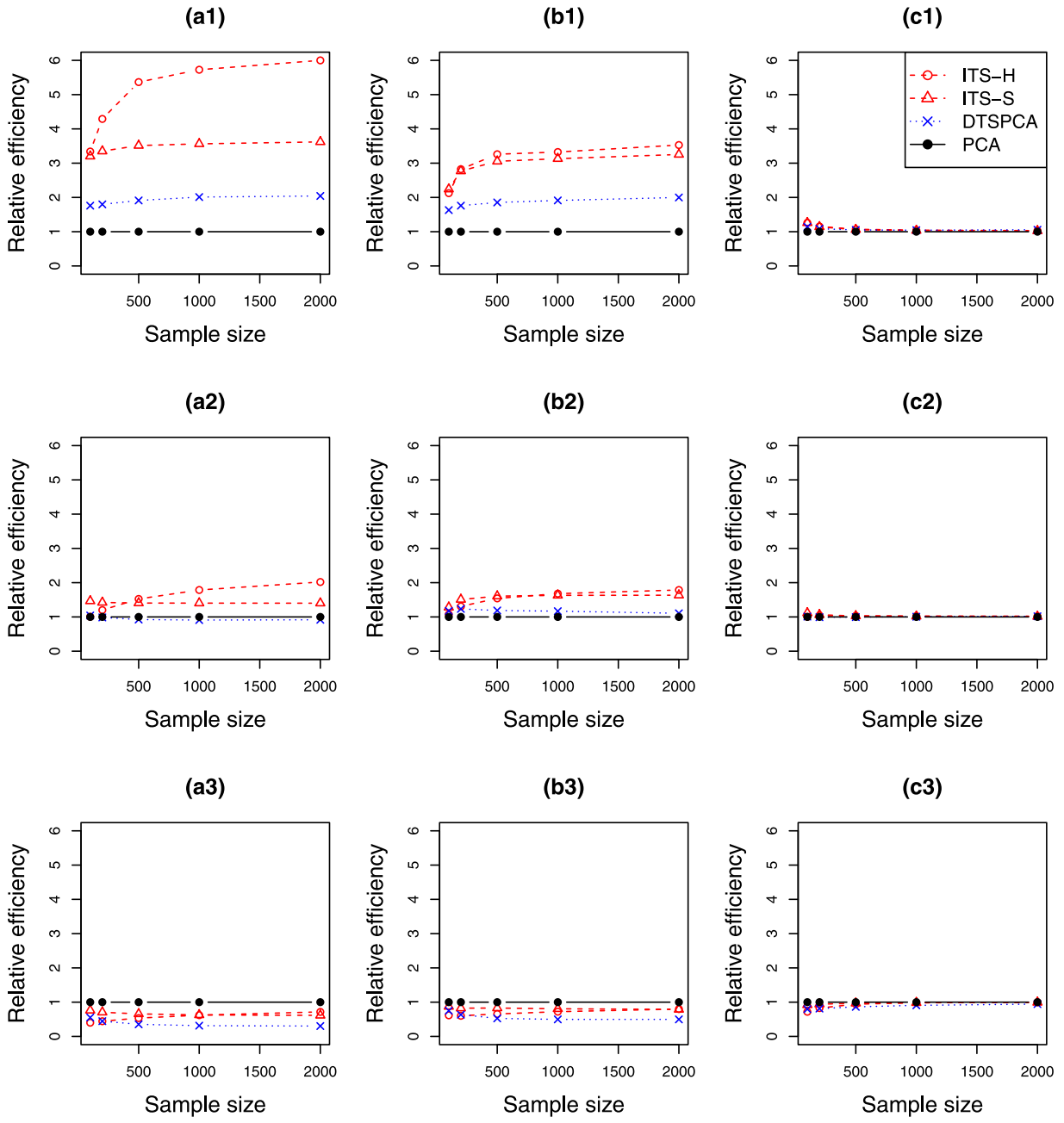

Fig. 4. Plots of relative efficiencies against sample size n for the ITS-H, ITS-S, DTSPCA, and PCA estimators with respect to the PCA estimator for $\pi(d) = 5\log d, 5d^{1/2}, d-1$ with $d = 128$. (a1)–(a3) are plots of relative efficiencies based on the Frobenius norm for $\pi(d) = 5\log(d), 5d^{1/2}, d-1$, respectively, with $\mathbf{\Lambda} = (0.25, 0.25, 0.25, 0.25)$. (b1)–(b3) are plots of relative efficiencies based on the Frobenius norm for $\pi(d) = 5\log(d), 5d^{1/2}, d-1$, respectively, with $\mathbf{\Lambda} = (0.4, 0.3, 0.2, 0.1)$. (c1)–(c3) are plots of relative efficiencies based on the Frobenius norm for $\pi(d) = 5\log(d), 5d^{1/2}, d-1$, respectively, with $\mathbf{\Lambda} = (0.5, 0.3, 0.19, 0.01)$.

Simple algebraic manipulations show

$$\max_j \left\| d^{-1}(\widehat{\beta}_j - \beta_j)\mathbf{B}_j \right\|_2 \leq \frac{2}{d}$$

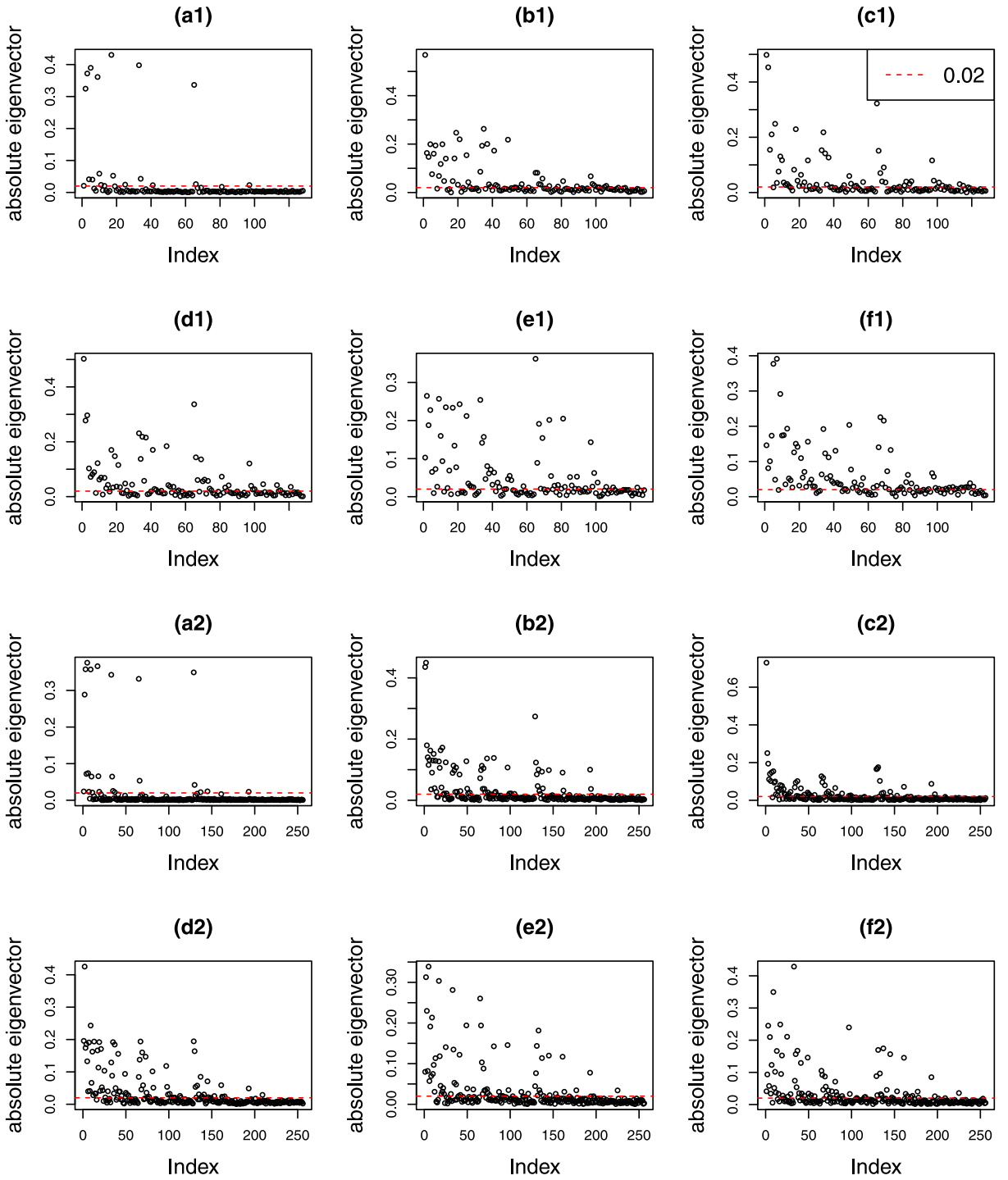

Fig. 5. Plots for the absolute elements of the eigenvectors corresponding to the first 6 eigenvalues. (a1)–(f1) are plots for $\boldsymbol{\rho}_7$. (a2)–(f2) are plots for $\boldsymbol{\rho}_8$.

and

$$\left\| d^{-2} \sum_{j=2}^{p} E\left[(\widehat{\beta}_j - \beta_j)^2 \mathbf{B}_j^T \mathbf{B}_j \right] \right\|_2 \leq \frac{1}{n}.$$

Table 3
MSEs in Frobenius norm for the ITS-H, ITS-S, DTSPCA, and PCA density matrix estimators with $d = 128, 256$ and $n = 100, 200, 500, 1000, 2000$ (We make the smallest MSE bold).

d	n	ITS-H	ITS-S	DTSPCA	PCA	$\widehat{\rho}$
128	100	**0.04672**	0.04975	0.05157	0.06837	1.27381
	200	**0.03360**	0.03632	0.03347	0.04897	0.63686
	500	0.02060	0.02060	**0.01781**	0.02704	0.25442
	1000	0.01233	0.01222	**0.01056**	0.01557	0.12727
	2000	0.00750	0.00706	**0.00630**	0.00781	0.06376
256	100	**0.04529**	0.05616	0.05988	0.09043	2.55323
	200	**0.03612**	0.03778	0.03966	0.05278	1.27709
	500	0.01995	0.01970	**0.01868**	0.02876	0.51098
	1000	0.01246	0.01187	**0.01041**	0.01663	0.25544
	2000	0.00796	0.00758	**0.00649**	0.00978	0.12770

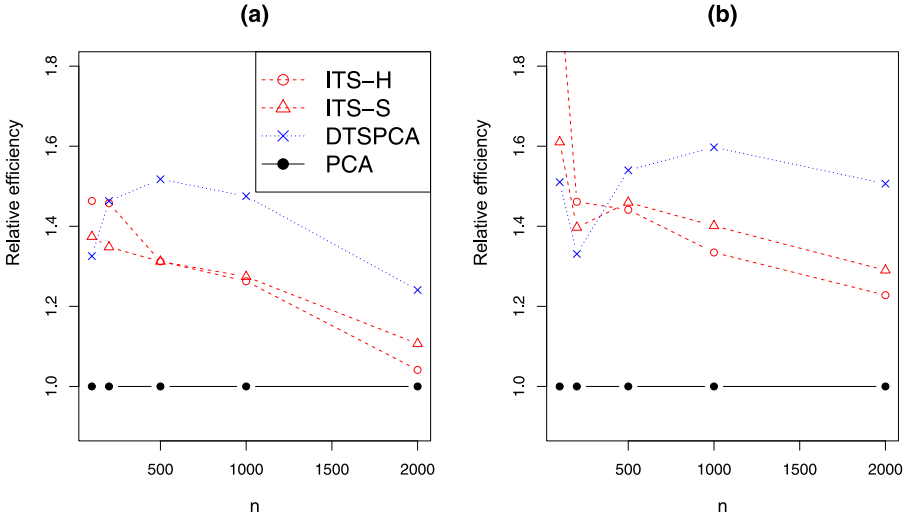

Fig. 6. Plots of relative efficiencies against the sample size n for the ITS-H, ITS-S, DTSPCA, and PCA estimators with respect to the PCA estimator. (a)–(b) are plots of relative efficiencies based on the Frobenius norm with $d = 128$ and 256, respectively.

Then, by the Matrix Bernstein inequality (Theorem 6.1 in Tropp (2012)), we get

$$P\left(\|\widehat{\boldsymbol{\rho}} - \boldsymbol{\rho}\|_2 \geq t\right) = P\left(\left\|d^{-1}\sum_{j=2}^{p}(\widehat{\beta}_j - \beta_j)\mathbf{B}_j\right\|_2 \geq t\right) \leq 2d\exp\left(-\frac{t^2/2}{n^{-1} + 2t/(3d)}\right).$$

First consider the condition (i). We take $t = \sqrt{6\log(d \vee n)/n}$ and then obtain

$$P\left(\|\widehat{\boldsymbol{\rho}} - \boldsymbol{\rho}\|_2 \geq \sqrt{6\log(d \vee n)/n}\right) \leq 2(d \vee n)^{-2}. \tag{7.2}$$

Consider the numerator on the right hand side of (7.1). For any $\mathbf{a} = (a_1, \ldots, a_d) \in \mathbb{C}^d$ such that $\|\mathbf{a}\|_2^2 = 1$, since $(\widehat{\beta}_j - \beta_j)$'s are independent with mean zero, we have

$$\begin{aligned} E\left[\|(\widehat{\boldsymbol{\rho}} - \boldsymbol{\rho})\mathbf{a}\|_2^2\right] &= \frac{1}{d^2}\sum_{j=2}^{p} E\left[(\widehat{\beta}_j - \beta_j)^2\right]\|\mathbf{B}_j\mathbf{a}\|_2^2 \\ &= \frac{1}{d^2}\sum_{j=2}^{p}\frac{1-\beta_j^2}{n} \\ &= \frac{1}{n} - \frac{\sum_{v=1}^{r}\lambda_v^2}{dn}, \end{aligned}$$

which along with $\|\mathbf{q}_\nu\|_2^2 = 1$ imply

$$E\left[\|(\boldsymbol{\rho}-\widehat{\boldsymbol{\rho}})\mathbf{Q}\|_F^2\right] = \sum_{\nu=1}^{m} E\left[\|(\boldsymbol{\rho}-\widehat{\boldsymbol{\rho}})\mathbf{q}_\nu\|_2^2\right] = m\left(\frac{1}{n} - \frac{\sum_{\nu=1}^{r}\lambda_\nu^2}{dn}\right). \tag{7.3}$$

Finally since $\|\sin(\widehat{\mathbf{Q}}, \mathbf{Q})\|_F^2 \le m$, we conclude

$$\begin{aligned} E\left[\|\sin(\widehat{\mathbf{Q}}, \mathbf{Q})\|_F^2\right] &= E\left[\|\sin(\widehat{\mathbf{Q}}, \mathbf{Q})\|_F^2 \mathbf{1}_E\right] + E\left[\|\sin(\widehat{\mathbf{Q}}, \mathbf{Q})\|_F^2 \mathbf{1}_{E^c}\right] \\ &\le \frac{2m}{(d\vee n)^2} + E\left[\|(\widehat{\boldsymbol{\rho}}-\boldsymbol{\rho})\mathbf{Q}\|_F^2\right]\left(\lambda_m - \lambda_{m+1} - \sqrt{\frac{6\log(d\vee n)}{n}}\right)^{-2} \\ &\le \frac{2m}{n} + \frac{m}{n}\left(\lambda_m - \lambda_{m+1} - \sqrt{\frac{6\log(d\vee n)}{n}}\right)^{-2} \\ &= O\left(\frac{n^{-1}}{(\lambda_m - \lambda_{m+1})^2}\right), \end{aligned}$$

where $E = \{\max_{1\le\nu\le d}|\widehat{\lambda}_\nu - \lambda_\nu| \ge \sqrt{\frac{6\log(d\vee n)}{n}}\}$, and the second and third inequalities are due to (7.2) and (7.3), respectively. We prove the theorem in the condition (i).

For the case of the condition (*ii*), we take $t = 2\log n/d$ and replace (7.2) by

$$P\left(\|\widehat{\boldsymbol{\rho}} - \boldsymbol{\rho}\|_2 \ge 2\log n/d\right) \le 2n^{-2}. \tag{7.4}$$

The same argument can be used to prove the theorem as follows:

$$\begin{aligned} E\left[\|\sin(\widehat{\mathbf{Q}}, \mathbf{Q})\|_F^2\right] &= E\left[\|\sin(\widehat{\mathbf{Q}}, \mathbf{Q})\|_F^2 \mathbf{1}_E\right] + E\left[\|\sin(\widehat{\mathbf{Q}}, \mathbf{Q})\|_F^2 \mathbf{1}_{E^c}\right] \\ &\le \frac{4m}{n^2} + E\left[\|(\widehat{\boldsymbol{\rho}}-\boldsymbol{\rho})\mathbf{Q}\|_F^2\right]\left(\lambda_m - \lambda_{m+1} - \frac{2\log n}{d}\right)^{-2} \\ &\le \frac{4m}{n} + \frac{m}{n}\left(\lambda_m - \lambda_{m+1} - \frac{2\log n}{d}\right)^{-2} \\ &= O\left(\frac{n^{-1}}{(\lambda_m - \lambda_{m+1})^2}\right), \end{aligned}$$

where $E = \{\max_{1\le\nu\le d}|\widehat{\lambda}_\nu - \lambda_\nu| \ge 2\log n/d\}$, and the second and third inequalities are due to (7.4) and (7.3), respectively. ■

7.1.2. Proof of Theorem 2

Proof of Theorem 2. Define the set of high signal coordinates,

$$H = H(\tau) = \{\nu : |q_{\nu j}| \ge C_\tau \tau_n, \text{ for some } 1 \le j \le r\},$$

where C_τ is a constant. Then, similar to the proof of Lemma 3.1 in Ma (2013), we can show

$$r \le |H| \le C\pi(p)\tau_n^{-\delta}. \tag{7.5}$$

In addition, let $L = \{1, \ldots, d\} \setminus H$. Here and after, we use an extra superscript "o" to indicate oracle quantities. That is, let

$$\boldsymbol{\rho} = \begin{bmatrix} \boldsymbol{\rho}_{HH} & \boldsymbol{\rho}_{HL} \\ \boldsymbol{\rho}_{LH} & \boldsymbol{\rho}_{LL} \end{bmatrix} \quad \text{and} \quad \boldsymbol{\rho}^o = \begin{bmatrix} \boldsymbol{\rho}_{HH} & 0 \\ 0 & 0 \end{bmatrix}.$$

$\widehat{\boldsymbol{\rho}}$ and $\widehat{\boldsymbol{\rho}}^o$ are estimators for $\boldsymbol{\rho}$ and $\boldsymbol{\rho}^o$, respectively. Specifically,

$$\widehat{\boldsymbol{\rho}} = (\widehat{\rho}_{ij})_{i,j=1,\ldots,p} \quad \text{and} \quad \widehat{\boldsymbol{\rho}}^o = \begin{bmatrix} \widehat{\boldsymbol{\rho}}_{HH} & 0 \\ 0 & 0 \end{bmatrix}.$$

Using Algorithm 1, we construct an oracle sequence of d-by-m orthonormal matrices $\{\widehat{\mathbf{Q}}^{(k),o}, k \ge 1\}$ with the initial $\widehat{\mathbf{Q}}^{(0),o}$. To construct $\widehat{\mathbf{Q}}^{(0),o}$, we use an oracle version of Algorithm 2. Specifically, $S^o = S \cap H$. This ensures that $\widehat{\mathbf{Q}}^{(0),o}_{L\cdot} = 0$.

With probability at least $1 - C_0(d \vee n)^{-2}$, we have

$$\begin{aligned}
&\| \sin(\mathbf{Q}, \widehat{\mathbf{Q}}^{(R_s)})\|_F^2 \\
&\le C \left\{ \| \sin(\mathbf{Q}, \mathbf{Q}^o)\|_F^2 + \| \sin(\mathbf{Q}^o, \widehat{\mathbf{Q}}^o)\|_F^2 + \| \sin(\widehat{\mathbf{Q}}^o, \widehat{\mathbf{Q}}^{(R_s),o})\|_F^2 + \| \sin(\widehat{\mathbf{Q}}^{(R_s),o}, \widehat{\mathbf{Q}}^{(R_s)})\|_F^2 \right\} \\
&\le C \frac{\pi(d)\tau_n^{2-\delta}}{(\lambda_m - \lambda_{m+1})^2},
\end{aligned}$$

where the fist inequality is due to the triangular inequality and Jensen's inequality, and the last inequality is from Propositions 1–4. ■

Proposition 1. *Under assumptions of Theorem 2, we have*

$$\| \sin(\mathbf{Q}, \mathbf{Q}^o)\|_F^2 \le C \frac{\pi(d)\tau_n^{2-\delta}}{(\lambda_m - \lambda_{m+1})^2}.$$

Proposition 2. *Under assumptions of Theorem 2, we have with probability at least* $1 - C_0(d \vee n)^{-2}$

$$\| \sin(\mathbf{Q}^o, \widehat{\mathbf{Q}}^o)\|_F^2 \le C \frac{\pi(d)\tau_n^{2-\delta}}{(\lambda_m - \lambda_{m+1})^2}.$$

Proposition 3. *Under assumptions of Theorem 2, we have with probability at least* $1 - C_0(d \vee n)^{-2}$,

$$\| \sin(\widehat{\mathbf{Q}}^o, \widehat{\mathbf{Q}}^{(R_s),o})\|_F^2 \le C \frac{\pi(d)\tau_n^{2-\delta}}{(\lambda_m - \lambda_{m+1})^2}.$$

Proposition 4. *Under assumptions of Theorem 2, we have with probability at least* $1 - C_0(d \vee n)^{-2}$,

$\widehat{\mathbf{Q}}^{(k),o} = \widehat{\mathbf{Q}}^{(k)}$ *for* $k \ge 0$.

The proofs of Propositions 1–4 are given in the Appendix.

7.2. Proof of Theorem 3

To obtain the lower bound, we consider the real valued density matrix, $\boldsymbol{\rho}$. That is, β_j's corresponding to complex valued Pauli matrices are zero.

We use the following Fano's lemma (Lemma A.5 in Birnbaum et al. (2013)).

Lemma 1 (*Fano's Lemma*). *Denote by* $\{P_\theta : \theta \in \Theta\}$ *a family of probability distribution on a common measurable space, where* Θ *is an arbitrary parameter set. Then, for any finite subset* $\mathcal{G} = \{\theta_1, \ldots, \theta_M\}$ *of* Θ, *we have*

$$\inf_T \sup_{\theta \in \Theta} P_\theta(T \ne \theta) \ge 1 - \inf_F \frac{M^{-1} \sum_{k=1}^M D(P_k \parallel F) + \log 2}{\log M},$$

where F *is an arbitrary probability distribution,* $P_k = P_{\theta_k}$, T *denotes an arbitrary estimator of* θ *with values in* Θ, *and* $D(P_k \parallel F)$ *is the Kullback–Leibler (KL) divergence of* F *from* P_k.

Lemma 2. *For* $k = 1, 2$, *let*

$$\boldsymbol{\rho}_k = \frac{1}{d}\mathbf{B}_1 + \frac{1}{d}\sum_{j=2}^p \beta_j^{(k)} \mathbf{B}_j$$

and P_k *be the product of the binomial probability measures,* $B(n, \frac{1+\beta_2^{(k)}}{2}), \ldots, B(n, \frac{1+\beta_p^{(k)}}{2})$. *Then we have*

$$D(P_1 \parallel P_2) \le n \sum_{j=2}^p \frac{(\beta_j^{(1)} - \beta_j^{(2)})^2}{1 - (\beta_j^{(2)})^2}.$$

Lemma 3. *For* $\epsilon \in [0, 1]$, *the function* $\mathbf{A}_\epsilon : \mathbb{V}_{d-m,m} \mapsto \mathbb{V}_{d,m}$ *is defined in block form as*

$$\mathbf{A}_\epsilon(\mathbf{J}) = \begin{pmatrix} (1-\epsilon^2)^{1/2}\mathbf{I}_m \\ \epsilon \mathbf{J} \end{pmatrix},$$

where $\mathbb{V}_{d,h} = \{\mathbf{Q} \in \mathbb{R}^{d \times h} : \mathbf{Q}^\dagger \mathbf{Q} = \mathbf{I}\}$ *is the Stiefel manifold of d-by-h orthonormal matrices. For* $\mathbf{J}_1, \mathbf{J}_2 \in \mathbb{V}_{d-m,m}$, *we have*

$$\| \sin(\mathbf{A}_\epsilon(\mathbf{J}_1), \mathbf{A}_\epsilon(\mathbf{J}_2))\|_2^2 \ge \epsilon^2(1-\epsilon^2)\|\mathbf{J}_1 - \mathbf{J}_2\|_2^2,$$

and

$$\epsilon^2(1-\epsilon^2)\|\mathbf{J}_1-\mathbf{J}_2\|_F^2 \le \|\sin(\mathbf{A}_\epsilon(\mathbf{J}_1),\mathbf{A}_\epsilon(\mathbf{J}_2))\|_F^2 \le \epsilon^2\|\mathbf{J}_1-\mathbf{J}_2\|_F^2.$$

Proof. Similar to the proof of Lemma 3 (Kim and Wang, 2016), we can show this statement. ■

Lemma 4. *Let h be an integer satisfying $e \le h$, and let $s \in [1,h]$. There exists a subset $\{\mathbf{J}_1,\ldots,\mathbf{J}_M\} \subset \mathbb{V}_{h,1}$ satisfying the following properties:*

(1) $\|\mathbf{J}_j-\mathbf{J}_{j'}\|_2^2 \ge 1/4$ for all $j \ne j'$;
(2) $\|\mathbf{J}_j\|_0 \le s$ for all j;
(3) $\log M \ge \max\{cs[1+\log(h/s)], \log h\}$, where $c > 1/30$ is an absolute constant.

Proof. See the proof of Lemma A.5 in Vu and Lei (2013). ■

Proof of Theorem 3. Since Pauli matrices form an orthogonal basis for all complex Hermitian matrices, for any given $\mathbf{A} \in \mathbb{V}_{d,m}$, where $\mathbb{V}_{d,m}$ is the Stiefel manifold of d-by-m orthonormal matrices, there are $\beta_j's$ such that

$$\boldsymbol{\rho}(\mathbf{A}) = d^{-1}\left(\mathbf{I}_d + \sum_{j=2}^{p}\beta_j\mathbf{B}_j\right) = m^{-1}\mathbf{A}\mathbf{A}^T.$$

We consider the subclass of $\mathbf{A}$ as follows: Let

$$\mathbf{A}_\epsilon(\mathbf{J}) = \begin{pmatrix}(1-\epsilon^2)^{1/2}\mathbf{I}_m \\ \epsilon\mathbf{J}\end{pmatrix}, \tag{7.6}$$

where $\mathbf{I}_m$ is a m-by-m identity matrix, and $\epsilon \in [0,1]$, and $\mathbf{J} \in \mathbb{V}_{d-m,m}$. Using Lemma 4, we construct the packing set of $\mathbf{J}$ as follows: Define $\mathcal{G}_\tau = \{\mathbf{J}_1,\ldots,\mathbf{J}_{M'}\}$ with $h = \lfloor (d-m)/m \rfloor$ and $s = \varrho h$, where $\varrho \in (1/h, 1)$. Then, from Lemma 4, (i) $\log M' \ge c\max\{d\varrho[1-\log\varrho], \log d\}$ for some constant c free n and p; (ii) $\|\mathbf{J}_i\|_0 \le s$ for all $j = 1,\ldots,M'$; (iii) $\|\mathbf{J}_j-\mathbf{J}_{j'}\|_2^2 \ge 1/4$ for all $j \ne j'$. Choose $\mathbf{J}$ in (7.6) as follows:

$$\mathbf{J}(\mathbf{a}_1,\ldots,\mathbf{a}_m) = \begin{pmatrix}\mathbf{a}_1 & 0 & \cdots & 0 \\ 0 & \mathbf{a}_2 & \cdots & 0 \\ \vdots & 0 & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{a}_m\end{pmatrix},$$

where $\mathbf{a}_j \in \mathcal{G}_\tau$ for all j. Let $\mathcal{G}(\mathbf{J}) = \{\mathbf{J}(\mathbf{a}_1,\ldots,\mathbf{a}_m), \mathbf{a}_j \in \mathcal{G}_\tau \text{ for } j = 1,\ldots,m\}$. Then, from the construction of $\mathcal{G}_\tau$, $\mathcal{G}(\mathbf{J}) \subset \mathbb{V}_{d-m,m}$, and the cardinality of $\mathcal{G}(\mathbf{J})$ is $M = (M')^m$. Note that $\log M \ge mc\max\{d\varrho[1-\log\varrho], \log d\}$, and for any $\mathbf{J}_k \in \mathcal{G}(\mathbf{J})$, there exist $\beta_j^{(k)}$'s such that

$$\boldsymbol{\rho}(\mathbf{J}_k) = d^{-1}\left(\mathbf{I}_d + \sum_{j=2}^{p}\beta_j^{(k)}\mathbf{B}_j\right) = m^{-1}\mathbf{A}_\epsilon(\mathbf{J}_k)\mathbf{A}_\epsilon(\mathbf{J}_k)^T.$$

Without loss of generality, we assume that the first d Pauli matrices, $\mathbf{B}_j$'s, correspond to the diagonal Pauli matrices. Define P_0 the product of the binomial probability measures, $B(n, \frac{1+\beta_2^{(0)}}{2}), \ldots, B(n, \frac{1+\beta_p^{(0)}}{2})$ with $\beta_j^{(0)}$'s determined as follows:

$$\beta_{d+1}^{(0)} = \cdots = \beta_p^{(0)} = 0$$

and $\beta_1^{(0)},\ldots,\beta_d^{(0)}$ are a solution of the following equation,

$$\boldsymbol{\rho}_0 = \frac{1}{d}\sum_{j=1}^{d}\beta_j^{(0)}\mathbf{B}_j = m^{-1}\begin{pmatrix}(1-\epsilon^2)\mathbf{I}_m & 0 \\ 0 & \frac{m\epsilon^2}{d-m}\mathbf{I}_{d-m}\end{pmatrix}.$$

Let $\boldsymbol{\beta}^{(0)} = (\beta_1^{(0)},\ldots,\beta_d^{(0)})^T$ and $\boldsymbol{\beta}^{(k)} = (\beta_1^{(k)},\ldots,\beta_d^{(k)})^T$, and $\mathbf{H} = (\mathbf{b}_1,\ldots,\mathbf{b}_d)$, where $\mathbf{b}_j = \mathrm{diag}(\mathbf{B}_j)$ for $j = 1,\ldots,d$. Then, by the construction of the Pauli matrices, $\mathbf{H}$ is d-by-d Hadamard matrix. We have

$$\boldsymbol{\beta}^{(0)} = \mathbf{H}^T\mathrm{diag}(\boldsymbol{\rho}_0) \quad \text{and} \quad \boldsymbol{\beta}^{(k)} = \mathbf{H}^T\mathrm{diag}(\boldsymbol{\rho}(\mathbf{J}_k)).$$

Then

$$\begin{aligned}\sum_{j=2}^{d}|\beta_j^{(k)}-\beta_j^{(0)}|^2 &= \|\mathbf{H}^T[\mathrm{diag}(\boldsymbol{\rho}(\mathbf{J}_k))-\mathrm{diag}(\boldsymbol{\rho}_0)]\|_2^2\\ &= d[\mathrm{diag}(\boldsymbol{\rho}(\mathbf{J}_k))-\mathrm{diag}(\boldsymbol{\rho}_0)]^T[\mathrm{diag}(\boldsymbol{\rho}(\mathbf{J}_k))-\mathrm{diag}(\boldsymbol{\rho}_0)]\\ &\le 2m^{-1}d\epsilon^4, \end{aligned} \tag{7.7}$$

where the second equality is established by the fact that $\mathbf{H}^T\mathbf{H}=d\,\mathbf{I}_d$. Note that $|\beta_j^{(0)}|\le 1-\epsilon^2/2$ for all $j=2,\dots,d$. For off-diagonal terms, we have for any $k=1,\dots,M$,

$$\begin{aligned}\|\boldsymbol{\rho}(\mathbf{J}_k)-\boldsymbol{\rho}_0\|_F^2 &= d\sum_{j=1}^{p}|\beta_j^{(k)}-\beta_j^{(0)}|^2\\ &= m^{-2}\left\|\begin{pmatrix}0 & (1-\epsilon^2)^{1/2}\epsilon\mathbf{J}_k^T\\ (1-\epsilon^2)^{1/2}\epsilon\mathbf{J}_k & \epsilon^2\mathbf{J}_k\mathbf{J}_k^T-\frac{m\epsilon^2}{d-m}\mathbf{I}_{d-m}\end{pmatrix}\right\|_F^2\\ &= m^{-1}[2(1-\epsilon^2)\epsilon^2+\epsilon^4+m\epsilon^4/(d-m)]\le 2m^{-1}\epsilon^2.\end{aligned}$$

So, we have

$$\sum_{j=d+1}^{p}|\beta_j^{(k)}-\beta_j^{(0)}|^2\le 2m^{-1}d\epsilon^2. \tag{7.8}$$

Then, by Lemma 2, we can obtain the upper bound for the KL divergence as follows:

$$\begin{aligned}D(P_k\parallel P_0) &\le n\sum_{j=2}^{p}\frac{(\beta_j^{(k)}-\beta_j^{(0)})^2}{1-(\beta_j^{(0)})^2}=n\left[\sum_{j=2}^{d}\frac{(\beta_j^{(k)}-\beta_j^{(0)})^2}{1-(\beta_j^{(0)})^2}+\sum_{j=d+1}^{p}\frac{(\beta_j^{(k)}-\beta_j^{(0)})^2}{1-(\beta_j^{(0)})^2}\right]\\ &\le n\left[\sum_{j=2}^{d}\frac{(\beta_j^{(k)}-\beta_j^{(0)})^2}{1-(1-\epsilon^2/2)^2}+\sum_{j=d+1}^{p}(\beta_j^{(k)}-\beta_j^{(0)})^2\right]\\ &\le n\left[\frac{4dm^{-1}\epsilon^4}{\epsilon^2}+2m^{-1}d\epsilon^2\right]=6m^{-1}nd\epsilon^2,\end{aligned} \tag{7.9}$$

where the third inequality is due to (7.7) and (7.8).

By Lemmas 3 and 4, we have for any $k\ne k'$,

$$\|\sin(\mathbf{A}_\epsilon(\mathbf{J}_k),\mathbf{A}_\epsilon(\mathbf{J}_{k'}))\|_2^2\ge\epsilon^2(1-\epsilon^2)\|\mathbf{J}_k-\mathbf{J}_{k'}\|_2^2\ge\frac{1}{4}\epsilon^2(1-\epsilon^2). \tag{7.10}$$

By Chebyshev's inequality and Lemma 1, we have for all $\epsilon^2\in[0,1/2]$,

$$\begin{aligned}\max_k E_{P_k}\|\sin(\widehat{\mathbf{A}},\mathbf{A}_\epsilon(\mathbf{J}_k))\|_2^2 &\ge \frac{\epsilon^2(1-\epsilon^2)}{16}\left[1-\frac{6m^{-1}dn\epsilon^2+\log 2}{mc\max\{d\varrho[1-\log\varrho],\log d\}}\right]\\ &\ge \frac{\epsilon^2(1-\epsilon^2)}{16}\left[1-\frac{6dn\epsilon^2}{cm^2d\varrho[1-\log\varrho]}-\frac{\log 2}{mc\log d}\right]\\ &\ge \frac{\epsilon^2}{32}\left[\frac{1}{2}-\frac{6dn\epsilon^2}{cm^2d\varrho[1-\log\varrho]}\right],\end{aligned}$$

where the first inequality is due to (7.9) and (7.10). Take

$$\epsilon^2=\frac{cm^2}{24}\frac{\varrho d[1-\log\varrho]}{dn}=\frac{cm^2}{24}\frac{\varrho[1-\log\varrho]}{n}.$$

Then

$$\max_k E_{P_k}\|\sin(\widehat{\mathbf{A}},\mathbf{A}_\epsilon(\mathbf{J}_k))\|_2^2\ge\frac{1}{128}\epsilon^2. \tag{7.11}$$

To ensure that $\boldsymbol{\rho}(\mathbf{A}_\epsilon(\mathbf{J}_k))$'s are in the sparse subspace, $\mathcal{F}_\delta(\pi(d))$, we need the following condition

$$1+\epsilon^\delta s^{(2-\delta)/2}\le\pi(d). \tag{7.12}$$

Take

$$\varrho=c_\varrho\pi(d)d^{-1}\left(\frac{\log d}{nd}\right)^{-\delta/2},$$

where $c_\varrho = \frac{1}{\sqrt{2}}\left(\frac{cm^2}{24}\right)^{-\delta/2}$. Then (4.3) implies

$$\varrho \asymp \pi(d)d^{-1}\left(\frac{\log d}{nd}\right)^{-\delta/2} \asymp d^{-\mathcal{N}}, \quad \mathcal{N} \in (0, 1),$$

while $1/h = m/(d-m) \asymp d^{-1}$. Thus, asymptotically we have $\varrho \in (1/h, 1]$. Also

$$\begin{aligned}\epsilon^2 &\le \frac{c_\varrho cm^2}{24}\pi(d)d^{-1}n^{-1}\left(\frac{\log d}{nd}\right)^{-\delta/2}\left[1+\frac{1}{2}(1-\delta/2)\log d+\delta/2\log\log d\right]\\ &\le c_\varrho\frac{cm^2}{24}\pi(d)\left(\frac{\log d}{nd}\right)^{1-\delta/2} \le 1/2,\end{aligned}$$

where the last inequality is due to the fact that for $(n, d, \pi(d))$ satisfying (4.3), $\pi(d)\left(\frac{\log d}{nd}\right)^{1-\delta/2}$ is of order $n^{-1}d^{-\mathcal{N}}\log d$ which is asymptotically negligible.

Simple algebras show

$$\begin{aligned}\epsilon^{2\delta}s^{(2-\delta)} &\le c_\varrho^2\left(\frac{cm^2}{24}\right)^{\delta}\left(\pi(d)\left(\frac{\log d}{nd}\right)^{1-\delta/2}\right)^{\delta}\left(\pi(d)\left(\frac{\log d}{nd}\right)^{-\delta/2}\right)^{2-\delta}\\ &= \frac{1}{2}\pi(d)^2\left(\frac{\log d}{nd}\right)^{-\delta}\left(\frac{\log d}{nd}\right)^{\delta}\\ &= \frac{1}{2}\pi(d)^2.\end{aligned}$$

Thus, (7.12) holds. Now, from (7.11), we have

$$\begin{aligned}\max_k E_{P_k}\|\sin(\widehat{\mathbf{A}}, \mathbf{A}_\epsilon(\mathbf{J}_k))\|_2^2 &\ge C\pi(d)n^{-1}d^{-1}\left(\frac{\log d}{nd}\right)^{-\delta/2}\\ &\quad\times\left[1+\mathcal{N}\log d-\log\left(c_\varrho\pi(d)d^{\mathcal{N}-1}\left(\frac{\log d}{nd}\right)^{-\delta/2}\right)\right]\\ &\ge C\pi(d)n^{-1}d^{-1}\left(\frac{\log d}{nd}\right)^{-\delta/2}\log d\\ &= C\pi(d)\left(\frac{\log p}{nd}\right)^{1-\delta/2}, \qquad (7.13)\end{aligned}$$

where the second inequality is due to (4.3).

For the Frobenius norm, by Lemmas 3 and 4, we have for any $k \neq k'$,

$$\|\sin(\mathbf{A}_\epsilon(\mathbf{J}_k), \mathbf{A}_\epsilon(\mathbf{J}_{k'}))\|_F^2 \ge \epsilon^2(1-\epsilon^2)\|\mathbf{J}_k - \mathbf{J}_{k'}\|_F^2 \ge \frac{m}{4}\epsilon^2(1-\epsilon^2).$$

Then, similar to the proof of (7.13), we can show

$$\max_k E_{P_k}\|\sin(\widehat{\mathbf{A}}, \mathbf{A}_\epsilon(\mathbf{J}_k))\|_F^2 \ge C\pi(d)\left(\frac{\log d}{nd}\right)^{1-\delta/2}. \quad \blacksquare$$

Acknowledgments

The research of Tony Cai was supported in part by National Science Foundation, United States of America Grant DMS-1712735 and NIH, United States of America grants R01-GM129781 and R01-GM123056. The research of Donggyu Kim was supported in part by KAIST, Republic of Korea Settlement/Research Subsidies for Newly-hired Faculty grant G04170049 and KAIST, Republic of Korea Basic Research Funds by Faculty (A0601003029). The research of Xinyu Song was supported by the Fundamental Research Funds for the Central Universities, China (2018110128), and National Natural Science Foundation of China (Grant No. 11871323). The research of Yazhen Wang was supported in part by National Science Foundation, United States of America Grants DMS-15-28375, DMS-17-07605, and DMS-19-13149.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jspi.2020.11.002.

References

Bickel, P.J., Levina, E., et al., 2008. Covariance regularization by thresholding. Ann. Statist. 36 (6), 2577–2604.
Birgé, L., 2001. A new look at an old result: Fano's lemma.
Birnbaum, A., Johnstone, I.M., Nadler, B., Paul, D., 2013. Minimax bounds for sparse PCA with noisy high-dimensional data. Ann. Statist. 41 (3), 1055.
Cai, T., Kim, D., Wang, Y., Yuan, M., Zhou, H.H., 2016. Optimal large-scale quantum state tomography with Pauli measurements. Ann. Statist. 44 (2), 682–712.
Cai, T., Liu, W., 2011. Adaptive thresholding for sparse covariance matrix estimation. J. Amer. Statist. Assoc. 106 (494), 672–684.
Cai, T., Ma, Z., Wu, Y., 2013. Sparse PCA: Optimal rates and adaptive estimation. Ann. Statist. 41 (6), 3074–3110.
Cai, T., Ma, Z., Wu, Y., 2015. Optimal estimation and rank detection for sparse spiked covariance matrices. Probab. Theory Related Fields 161 (3–4), 781–815.
Cai, T.T., Zhou, H.H., 2012. Minimax estimation of large covariance matrices under l1-norm. Statist. Sinica 22 (4), 1319–1349.
Golub, G., Van Loan, C., 1996. Matrix Computations, third ed. The John-Hopkins University Press.
Häffner, H., Hänsel, W., Roos, C., Benhelm, J., Chwalla, M., Körber, T., Rapol, U., Riebe, M., Schmidt, P., Becher, C., 2005. Scalable multiparticle entanglement of trapped ions. Nature 438 (7068), 643.
Johnstone, I.M., Lu, A.Y., 2009. On consistency and sparsity for principal components analysis in high dimensions. J. Amer. Statist. Assoc. 104 (486), 682–693.
Kim, D., Kong, X.-B., Li, C.-X., Wang, Y., 2018. Adaptive thresholding for large volatility matrix estimation based on high-frequency financial data. J. Econometrics 203 (1), 69–79.
Kim, D., Wang, Y., 2016. Sparse PCA-based on high-dimensional Itô processes with measurement errors. J. Multivariate Anal. 152, 172–189.
Kim, D., Wang, Y., 2017. Hypothesis tests for large density matrices of quantum systems based on Pauli measurements. Physica A 469, 31–51.
Kim, D., Wang, Y., Zou, J., 2016. Asymptotic theory for large volatility matrix estimation based on high-frequency financial data. Stochastic Process. Appl. 126 (11), 3527–3577.
Koltchinskii, V., Xia, D., 2015. Optimal estimation of low rank density matrices. J. Mach. Learn. Res. 16 (53), 1757–1792.
Li, R.-C., 1998a. Relative perturbation theory: I. Eigenvalue and singular value variations. SIAM J. Matrix Anal. Appl. 19 (4), 956–982.
Li, R.-C., 1998b. Relative perturbation theory: II. Eigenspace and singular subspace variations. SIAM J. Matrix Anal. Appl. 20 (2), 471–492.
Ma, Z., 2013. Sparse principal component analysis and iterative thresholding. Ann. Statist. 41 (2), 772–801.
Nielsen, M.A., Chuang, I.L., 2010. Quantum Computation and Quantum Information. Cambridge Univ. Press.
Tao, M., Wang, Y., Chen, X., 2013a. Fast convergence rates in estimating large volatility matrices using high-frequency financial data. Econometric Theory 29 (4), 838–856.
Tao, M., Wang, Y., Zhou, H.H., 2013b. Optimal sparse volatility matrix estimation for high-dimensional Itô processes with measurement errors. Ann. Statist. 41 (4), 1816–1864.
Tropp, J.A., 2012. User-friendly tail bounds for sums of random matrices. Found. Comput. Math. 12 (4), 389–434.
Vu, V.Q., Cho, J., Lei, J., Rohe, K., 2013. Fantope projection and selection: A near-optimal convex relaxation of sparse PCA. Adv. Neural Inf. Process. Syst. 2, 2670–2678.
Vu, V.Q., Lei, J., 2013. Minimax sparse principal subspace estimation in high dimensions. Ann. Statist. 41 (6), 2905–2947.
Wang, Y., 2011. Quantum Monte Carlo simulation. Ann. Appl. Stat. 5 (2A), 669–683.
Wang, Y., 2012. Quantum computation and quantum information. Statist. Sci. 27 (3), 373–394.
Wang, Y., 2013. Asymptotic equivalence of quantum state tomography and noisy matrix completion. Ann. Statist. 41 (5), 2462–2504.
Wang, Y., Song, X., 2020. Quantum science and quantum technology. Statist. Sci. 35 (1), 51–74.
Wang, Y., Zou, J., 2010. Vast volatility matrix estimation for high-frequency financial data. Ann. Statist. 38 (2), 943–978.