The Annals of Statistics
2016, Vol. 44, No. 2, 682–712
DOI: 10.1214/15-AOS1382

OPTIMAL LARGE-SCALE QUANTUM STATE TOMOGRAPHY WITH PAULI MEASUREMENTS

BY TONY CAI[*,1], DONGGYU KIM[†], YAZHEN WANG[†,2], MING YUAN[†,3] AND HARRISON H. ZHOU[‡,4]

University of Pennsylvania[*], *University of Wisconsin-Madison*[†] *and Yale University*[‡]

Quantum state tomography aims to determine the state of a quantum system as represented by a density matrix. It is a fundamental task in modern scientific studies involving quantum systems. In this paper, we study estimation of high-dimensional density matrices based on Pauli measurements. In particular, under appropriate notion of sparsity, we establish the minimax optimal rates of convergence for estimation of the density matrix under both the spectral and Frobenius norm losses; and show how these rates can be achieved by a common thresholding approach. Numerical performance of the proposed estimator is also investigated.

1. Introduction. For a range of scientific studies including quantum computation, quantum information and quantum simulation, an important task is to learn and engineer quantum systems [Aspuru-Guzik et al. (2005), Benenti, Casati and Strini (2004, 2007), Brumfiel (2012), Jones (2013), Lanyon et al. (2010), Nielsen and Chuang (2000), and Wang (2011, 2012)]. A quantum system is described by its state characterized by a density matrix, which is a positive semidefinite Hermitian matrix with unit trace. Determining a quantum state, often referred to as quantum state tomography, is an important but difficult task [Alquier et al. (2013), Artiles, Gill and Guţă (2005), Aubry, Butucea and Meziani (2009), Butucea, Guţă and Artiles (2007), Guţă and Artiles (2007), Häffner et al. (2005), Wang (2013), and Wang and Xu (2015)]. It is often inferred by performing measurements on a large number of identically prepared quantum systems.

More specifically, we describe a quantum spin system by the d-dimensional complex space $\mathbb{C}^d$ and its quantum state by a complex matrix on $\mathbb{C}^d$. When mea-

Received January 2015; revised August 2015.

[1]Supported in part by NSF Grants DMS-12-08982 and DMS-14-03708, and NIH Grant R01 CA127334.

[2]Supported in part by NSF Grants DMS-10-5635, DMS-12-65203 and DMS-15-28375.

[3]Supported in part by NSF Career Award DMS-13-21692 and FRG Grant DMS-12-65202.

[4]Supported in part by NSF Grant DMS-12-09191.

MSC2010 subject classifications. Primary 62H12, 81P50; secondary 62C20, 62P35, 81P45, 81P68.

Key words and phrases. Compressed sensing, density matrix, Pauli matrices, quantum measurement, quantum probability, quantum statistics, sparse representation, spectral norm, minimax estimation.

suring the quantum system by performing measurements on some observables which can be represented by Hermitian matrices, we obtain the measurement outcomes for each observable, where the measurements take values at random from all eigenvalues of the observable, with the probability of observing a particular eigenvalue equal to the trace of the product of the density matrix and the projection matrix onto the eigenspace corresponding to the eigenvalue. To handle the up and down states of particles in a quantum spin system, a common approach is to employ the well-known Pauli matrices as observables to perform measurements and obtain the so-called Pauli measurements [Britton et al. (2012), Johnson et al. (2011), Liu (2011), Sakurai and Napolitano (2010), Shankar (1994), and Wang (2012, 2013)]. Since all Pauli matrices have ± 1 eigenvalues, Pauli measurements takes discrete values 1 and -1, and the resulted measurement distributions can be characterized by binomial distributions. The goal is to estimate the density matrix based on the Pauli measurements.

Traditional quantum tomography employs classical statistical models and methods to deduce quantum states from quantum measurements. These approaches are designed for the setting where the size of a density matrix is greatly exceeded by the number of quantum measurements, which is almost never the case even for moderate quantum systems in practice because the dimension of the density matrix grows exponentially in the size of the quantum system. For example, the density matrix for b spin-$\frac{1}{2}$ quantum systems is of size $2^b \times 2^b$. In this paper, we aim to effectively and efficiently reconstruct the density matrix for a large-scale quantum system with a relatively limited number of quantum measurements.

Quantum state tomography is fundamentally connected to the problem of recovering a high-dimensional matrix based on noisy observations [Wang (2013)]. The latter problem arises naturally in many applications in statistics and machine learning and has attracted considerable recent attention. When assuming that the unknown matrix of interest is of (approximately) low-rank, many regularization techniques have been developed. Examples include Candès and Recht (2009), Candès and Tao (2010), Candès and Plan (2009, 2011), Keshavan, Montanari and Oh (2010), Recht, Fazel and Parrilo (2010), Bunea, She and Wegkamp (2011, 2012), Klopp (2011, 2012), Koltchinskii (2011), Koltchinskii, Lounici and Tsybakov (2011), Negahban and Wainwright (2011), Recht (2011), Rohde and Tsybakov (2011), and Cai and Zhang (2015), among many others. Taking advantage of the low-rank structure of the unknown matrix, these approaches can often be applied to estimate unknown matrices of high dimensions. Yet these methods do not fully account for the specific structure of quantum state tomography. As demonstrated in a pioneering article, Gross et al. (2010) argued that, when considering quantum measurements characterized by the Pauli matrices, the density matrix can often be characterized by the sparsity with respect to the Pauli basis. Built upon this connection, they suggested a compressed sensing [Donoho (2006)] strategy for quantum state tomography [Gross (2011) and Wang (2013)]. Although promising, their proposal assumes exact measurements, which is rarely the case in practice, and adopts the constrained nuclear norm minimization method, which may not be an appropriate matrix completion approach for estimating a density matrix with unit trace (or unit nuclear norm). We specifically address such challenges in the present paper. In particular, we establish the minimax optimal rates of convergence for the density matrix estimation under both the spectral and Frobenius norm losses when assuming that the true density matrix is approximately sparse under the Pauli basis. Furthermore, we show that these rates could be achieved by carefully thresholding the coefficients with respect to the Pauli basis. Because the quantum Pauli measurements are characterized by the binomial distributions, the convergence rates and minimax lower bounds are derived by asymptotic analysis with manipulations of binomial distributions instead of the usual normal distribution based calculations.

The rest of paper proceeds as follows. Section 2 gives some background on quantum state tomography and introduces a thresholding based density matrix estimator. Section 3 develops theoretical properties for the density matrix estimation problem. In particular, the convergence rates of the proposed density matrix estimator and its minimax optimality with respect to both the spectral and Frobenius norm losses are established. Section 4 features a simulation study to illustrate finite sample performance of the proposed estimators. All technical proofs are collected in Section 5.

2. Quantum state tomography with Pauli measurements. In this section, we first review the quantum state and density matrix and introduce Pauli matrices and Pauli measurements. We also develop results to describe density matrix representations through Pauli matrices and characterize the distributions of Pauli measurements via binomial distribution before introducing a thresholding based density matrix estimator.

2.1. *Quantum state and measurements*. For a d-dimensional quantum system, we describe its quantum state by a density matrix $\boldsymbol{\rho}$ on d dimensional complex space $\mathbb{C}^d$, where 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 semidefinite; (3) unit trace, that is, $\mathrm{tr}(\boldsymbol{\rho}) = 1$.

For a quantum system, it is important but difficult to know its quantum state. Experiments are conducted to perform measurements on the quantum system and obtain data for studying the quantum system and estimating its density matrix. In physics literature, quantum state tomography refers to reconstruction of a quantum state based on measurements for the quantum systems. Statistically, it is the problem of estimating the density matrix from the measurements. Common quantum measurements are on observable $\mathbf{M}$, which is defined as a Hermitian matrix on $\mathbb{C}^d$. Assume that the observable $\mathbf{M}$ has the following spectral decomposition:

$$\mathbf{M} = \sum_{a=1}^{r} \lambda_a \mathbf{Q}_a, \tag{2.1}$$

where λ_a are r different real eigenvalues of $\mathbf{M}$, and $\mathbf{Q}_a$ are projections onto the eigenspaces corresponding to λ_a. For the quantum system prepared in state $\boldsymbol{\rho}$, we need a probability space $(\Omega, \mathcal{F}, P)$ to describe measurement outcomes when performing measurements on the observable $\mathbf{M}$. Denote by R the measurement outcome of $\mathbf{M}$. According to the theory of quantum mechanics, R is a random variable on $(\Omega, \mathcal{F}, P)$ taking values in $\{\lambda_1, \lambda_2, \dots, \lambda_r\}$, with probability distribution given by

$$P(R = \lambda_a) = \operatorname{tr}(\mathbf{Q}_a \boldsymbol{\rho}), \qquad a = 1, 2, \dots, r, \qquad E(R) = \operatorname{tr}(\mathbf{M}\boldsymbol{\rho}). \tag{2.2}$$

We may perform measurements on an observable for a quantum system that is identically prepared under the state and obtain independent and identically distributed observations. See Holevo (1982), Sakurai and Napolitano (2010), and Wang (2012).

2.2. *Pauli measurements and their distributions.* The Pauli matrices as observables are widely used in quantum physics and quantum information science to perform quantum measurements. Let

$$\boldsymbol{\sigma}_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad \boldsymbol{\sigma}_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

$$\boldsymbol{\sigma}_2 = \begin{pmatrix} 0 & -\sqrt{-1} \\ \sqrt{-1} & 0 \end{pmatrix}, \qquad \boldsymbol{\sigma}_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

where $\boldsymbol{\sigma}_1$, $\boldsymbol{\sigma}_2$ and $\boldsymbol{\sigma}_3$ are called the two-dimensional Pauli matrices. Tensor products are used to define high-dimensional Pauli matrices. 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 d dimensional Pauli matrices

$$\boldsymbol{\sigma}_{\ell_1} \otimes \boldsymbol{\sigma}_{\ell_2} \otimes \cdots \otimes \boldsymbol{\sigma}_{\ell_b}, \qquad (\ell_1, \ell_2, \dots, \ell_b) \in \{0, 1, 2, 3\}^b. \tag{2.3}$$

We identify index $j = 1, \dots, d^2$ with $(\ell_1, \ell_2, \dots, \ell_b) \in \{0, 1, 2, 3\}^b$. For example, $j = 1$ corresponds to $\ell_1 = \cdots = \ell_b = 0$. With the index identification we denote by $\mathbf{B}_j$ the Pauli matrix $\boldsymbol{\sigma}_{\ell_1} \otimes \boldsymbol{\sigma}_{\ell_2} \otimes \cdots \otimes \boldsymbol{\sigma}_{\ell_b}$, with $\mathbf{B}_1 = \mathbf{I}_d$. We have the following theorem to describe Pauli matrices and represent a density matrix by Pauli matrices.

PROPOSITION 1. (i) *Pauli matrices* $\mathbf{B}_2, \ldots, \mathbf{B}_{d^2}$ *are of full rank and have eigenvalues* ± 1. *Denote by* $\mathbf{Q}_{j\pm}$ *the projections onto the eigen-spaces of* $\mathbf{B}_j$ *corresponding to eigenvalues* ± 1, *respectively. Then for* $j, j' = 2, \ldots, d^2$,

$$\operatorname{tr}(\mathbf{Q}_{j\pm}) = \frac{d}{2}, \qquad \operatorname{tr}(\mathbf{B}_{j'}\mathbf{Q}_{j\pm}) = \begin{cases} \pm\dfrac{d}{2}, & \text{if } j = j', \\ 0, & \text{if } j \neq j'. \end{cases}$$

(ii) *Denote by* $\mathbb{C}^{d\times d}$ *the space of all* d *by* d *complex matrices equipped with the Frobenius norm. All Pauli matrices defined by* (2.3) *form an orthogonal basis for all complex Hermitian matrices. Given a density matrix* $\boldsymbol{\rho}$, *we can expand it under the Pauli basis as follows*:

$$\boldsymbol{\rho} = \frac{\mathbf{I}_d}{d} + \sum_{j=2}^{d^2} \beta_j \frac{\mathbf{B}_j}{d}, \tag{2.4}$$

where β_j *are coefficients. For* $j = 2, \ldots, d^2$,

$$\operatorname{tr}(\boldsymbol{\rho}\mathbf{Q}_{j\pm}) = \frac{1 \pm \beta_j}{2}.$$

Suppose that an experiment is conducted to perform measurements on Pauli observable $\mathbf{B}_j$ independently for n quantum systems which are identically prepared in the same quantum state $\boldsymbol{\rho}$. As $\mathbf{B}_j$ has eigenvalues ± 1, the Pauli measurements take values 1 and -1, and thus the average of the n measurements for each $\mathbf{B}_j$ is a sufficient statistic. Denote by N_j the average of the n measurement outcomes obtained from measuring $\mathbf{B}_j$, $j = 2, \ldots, d^2$. Our goal is to estimate $\boldsymbol{\rho}$ based on $N_2, \ldots, N_{d^2}$.

The following proposition provides a simple binomial characterization for the distributions of N_j.

PROPOSITION 2. *Suppose that* $\boldsymbol{\rho}$ *is given by* (2.4). *Then* $N_2, \ldots, N_{d^2}$ *are independent with*

$$E(N_j) = \beta_j, \qquad \operatorname{Var}(N_j) = \frac{1 - \beta_j^2}{n},$$

and $n(N_j + 1)/2$ *follows a binomial distribution with* n *trials and cell probabilities* $\operatorname{tr}(\boldsymbol{\rho}\mathbf{Q}_{j+}) = (1 + \beta_j)/2$, *where* $\mathbf{Q}_{j+}$ *denotes the projection onto the eigenspace of* $\mathbf{B}_j$ *corresponding to eigenvalue* 1, *and* β_j *is the coefficient of* $\mathbf{B}_j$ *in the expansion of* $\boldsymbol{\rho}$ *in* (2.4).

2.3. *Density matrix estimation.* Since the dimension of a quantum system grows exponentially with its components such as the number of particles in the system, the matrix size of $\boldsymbol{\rho}$ tends to be very large even for a moderate quantum system. We need to impose some structure such as sparsity on $\boldsymbol{\rho}$ in order to make it consistently estimable. Suppose that $\boldsymbol{\rho}$ has a sparse representation under the Pauli basis, following wavelet shrinkage estimation we construct a density matrix estimator of $\boldsymbol{\rho}$. Assume that representation (2.4) is sparse in a sense that there is only a relatively small number of coefficients β_k with large magnitudes. Formally, we specify sparsity by assuming that coefficients $\beta_2, \dots, \beta_{d^2}$ satisfy

$$\sum_{k=2}^{d^2} |\beta_k|^q \leq \pi_n(d), \tag{2.5}$$

where $0 \leq q < 1$, and $\pi_n(d)$ is a deterministic function with slow growth in d such as $\log d$.

Pauli matrices are used to describe the spins of spin-$\frac{1}{2}$ particles along different directions, and density matrix $\boldsymbol{\rho}$ in (2.4) represents a mixture of quantum states with spins along many directions. Sparsity assumption (2.5) with $q = 0$ indicates the mixed state involving spins along a relatively small number of directions corresponding to those Pauli matrices with nonzero β_k. The sparsity reduces the complexity of mixed states. Sparse density matrices often occur in quantum systems where particles have sparse interactions such as location interactions. Examples include many quantum systems in quantum information and quantum computation [Berry et al. (2014), Boixo et al. (2014), Britton et al. (2012), Flammia et al. (2012), Senko et al. (2014), and Wang (2011, 2012)].

Since N_k are independent, and $E(N_k) = \beta_k$. We naturally estimate β_k by N_k and threshold N_k to estimate large β_k, ignoring small β_k, and obtain

$$\begin{aligned} &\hat{\beta}_k = N_k 1\big(|N_k| \geq \varpi\big) \quad \text{or} \\ &\hat{\beta}_k = \operatorname{sign}(N_k)\big(|N_k| - \varpi\big)_+, \qquad k = 2, \dots, d^2, \end{aligned} \tag{2.6}$$

and then we use $\hat{\beta}_k$ to construct the following estimator of $\boldsymbol{\rho}$,

$$\hat{\boldsymbol{\rho}} = \frac{\mathbf{I}_d}{d} + \sum_{k=2}^{d^2} \hat{\beta}_k \frac{\mathbf{B}_k}{d}, \tag{2.7}$$

where the two estimation methods in (2.6) are called hard and soft thresholding rules, and ϖ is a threshold value which, we reason below, can be chosen to be $\varpi = \hbar\sqrt{(4/n)\log d}$ for some constant $\hbar > 1$. The threshold value is designed such that for small β_k, N_k must be bounded by threshold ϖ with overwhelming probability, and the hard and soft thresholding rules select only those N_k with large signal components β_k.

As $n(N_k+1)/2 \sim \text{Bin}(n, (1+\beta_k)/2)$, an application of Bernstein's inequality leads to that for any $x > 0$,

$$P(|N_k - \beta_k| \ge x) \le 2\exp\left(-\frac{nx^2}{2(1-\beta_k^2 + x/3)}\right) \le 2\exp\left(-\frac{nx^2}{2(1+x/3)}\right),$$

and

$$\begin{aligned} &P\Big(\max_{2\le k\le d^2} |N_k - \beta_k| \le \varpi\Big) \\ &\quad = \prod_{k=2}^{d^2} P(|N_k - \beta_k| \le \varpi) \\ &\quad \ge \left[1 - 2\exp\left(-\frac{n\varpi^2}{2(1+\varpi/3)}\right)\right]^{d^2-1} = \left[1 - 2d^{-2\hbar/(1+o(1))}\right]^{d^2-1} \to 1, \end{aligned}$$

as $d \to \infty$ and $\varpi \to 0$, that is, with probability tending to one, $|N_k| \le \varpi$ uniformly for $k = 2, \dots, d^2$. Thus, we can select $\varpi = \hbar\sqrt{(4/n)\log d}$ to threshold N_k and obtain $\hat{\beta}_k$ in (2.6).

3. Asymptotic theory for the density matrix estimator.

3.1. *Convergence rates.* We fix matrix norm notation for our asymptotic analysis. Let $\mathbf{x} = (x_1, \dots, x_d)^T$ be a d-dimensional vector and $\mathbf{A} = (A_{ij})$ be a d by d matrix, and define their ℓ_α norms

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

Denote by $\|\mathbf{A}\|_F = \sqrt{\text{tr}(\mathbf{A}^\dagger\mathbf{A})}$ the Frobenius norm of $\mathbf{A}$.

For the case of matrix, the ℓ_2 norm is called the matrix spectral norm or operator norm. $\|\mathbf{A}\|_2$ is equal to the square root of the largest eigenvalue of $\mathbf{A}\mathbf{A}^\dagger$,

$$\|\mathbf{A}\|_1 = \max_{1\le j\le d}\sum_{i=1}^d |A_{ij}|, \qquad \|\mathbf{A}\|_\infty = \max_{1\le i\le d}\sum_{j=1}^d |A_{ij}|, \tag{3.1}$$

and

$$\|\mathbf{A}\|_2^2 \le \|\mathbf{A}\|_1\|\mathbf{A}\|_\infty. \tag{3.2}$$

For a real symmetric or complex Hermitian matrix $\mathbf{A}$, $\|\mathbf{A}\|_2$ is equal to the largest absolute eigenvalue of $\mathbf{A}$, $\|\mathbf{A}\|_F$ is the square root of the sum of squared eigenvalues, $\|\mathbf{A}\|_F \le \sqrt{d}\|\mathbf{A}\|_2$, and (3.1)–(3.2) imply that $\|\mathbf{A}\|_2 \le \|\mathbf{A}\|_1 = \|\mathbf{A}\|_\infty$.

The following theorem gives the convergence rates for $\hat{\boldsymbol{\rho}}$ under the spectral and Frobenius norms.

THEOREM 1. *Denote by* Θ *the class of density matrices satisfying the sparsity condition* (2.5). *Assume* $n^{c_0} \le d \le e^{n^{c_1}}$ *for some constants* $c_0 > 0$ *and* $c_1 < 1$. *For density matrix estimator* $\hat{\boldsymbol{\rho}}$ *defined by* (2.6)–(2.7) *with threshold* $\varpi = \hbar\sqrt{(4/n)\log d}$ *for some constant* $\hbar > 1$, *we have*

$$\sup_{\boldsymbol{\rho}\in\Theta} E\big[\|\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}\|_2^2\big] \le c_2 \pi_n^2(d)\frac{1}{d^2}\left(\frac{\log d}{n}\right)^{1-q},$$

$$\sup_{\boldsymbol{\rho}\in\Theta} E\big[\|\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}\|_F^2\big] \le c_3 \pi_n(d)\frac{1}{d}\left(\frac{\log d}{n}\right)^{1-q/2},$$

where c_2 *and* c_3 *are constants free of* n *and* d.

REMARK 1. Theorem 1 shows that $\hat{\boldsymbol{\rho}}$ achieves the convergence rate $\pi_n(d) \times d^{-1}(n^{-1}\log d)^{1-q/2}$ under the squared Frobenius norm loss and the convergence rate $\pi_n^2(d)d^{-2}(n^{-1}\log d)^{1-q}$ under the squared spectral norm loss. Both rates will be shown to be optimal in the next section. Similar to the optimal convergence rates for large covariance and volatility matrix estimation [Cai and Zhou (2012) and Tao, Wang and Zhou (2013)], the optimal convergence rates here have factors involving $\pi_n(d)$ and $\log d/n$. However, unlike the covariance and volatility matrix estimation case, the convergence rates in Theorem 1 have factors d^{-1} and d^{-2} for the squared spectral and Frobenius norms, respectively, and go to zero as d approaches to infinity. In particular, the result implies that MSEs of the proposed estimator get smaller for large d. This is quite contrary to large covariance and volatility matrix estimation where the traces are typically diverge, the optimal convergence rates grow with the logarithm of matrix size, and the corresponding MSEs increase in matrix size. The new phenomenon may be due to the unit trace constraint on density matrix and that the density matrix representation (2.4) needs a scaling factor d^{-1} to satisfy the constraint. Also for finite sample $\hat{\boldsymbol{\rho}}$ may not be positive semidefinite, we may project $\hat{\boldsymbol{\rho}}$ onto the cone formed by all density matrices under a given matrix norm $\|\cdot\|$, and obtain a positive semidefinite density matrix estimator $\tilde{\boldsymbol{\rho}}$. Since the underlying true density matrix $\boldsymbol{\rho}$ is positive semidefinite with unit trace, and the representation (2.7) ensures that $\hat{\boldsymbol{\rho}}$ has unit trace, the projection implies $\|\tilde{\boldsymbol{\rho}} - \hat{\boldsymbol{\rho}}\| \le \|\boldsymbol{\rho} - \hat{\boldsymbol{\rho}}\|$. Thus, $\|\tilde{\boldsymbol{\rho}} - \boldsymbol{\rho}\| \le \|\tilde{\boldsymbol{\rho}} - \hat{\boldsymbol{\rho}}\| + \|\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}\| \le 2\|\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}\|$. Taking $\|\cdot\|$ as the spectral norm or the Frobenius norm and using Theorem 1, we conclude that $\tilde{\boldsymbol{\rho}}$ has the same convergence rates as $\hat{\boldsymbol{\rho}}$.

3.2. *Optimality of the density matrix estimator.* The following theorem establishes a minimax lower bound for estimating $\boldsymbol{\rho}$ under the spectral norm.

THEOREM 2. *We assume that* $\pi_n(d)$ *in the sparsity condition* (2.5) *satisfies*

$$\pi_n(d) \le \aleph d^v (\log d)^{q/2}/n^{q/2}, \tag{3.3}$$

for some constant $\aleph > 0$ *and* $0 < v < 1/2$. *Then*

$$\inf_{\check{\boldsymbol{\rho}}} \sup_{\boldsymbol{\rho}\in\Theta} E\big[\|\check{\boldsymbol{\rho}} - \boldsymbol{\rho}\|_2^2\big] \geq c_4 \pi_n^2(d) \frac{1}{d^2}\left(\frac{\log d}{n}\right)^{1-q},$$

where $\check{\boldsymbol{\rho}}$ *denotes any estimator of* $\boldsymbol{\rho}$ *based on measurement data* $N_2, \dots, N_{d^2}$, *and* c_4 *is a constant free of* n *and* d.

REMARK 2. The lower bound in Theorem 2 matches the convergence rate of $\hat{\boldsymbol{\rho}}$ under the spectral norm in Theorem 1, so we conclude that $\hat{\boldsymbol{\rho}}$ achieves the optimal convergence rate under the spectral norm. To establish the minimax lower bound in Theorem 2, we construct a special subclass of density matrices and then apply Le Cam's lemma. Assumption (3.3) is needed to guarantee the positive definiteness of the constructed matrices as density matrix candidates and to ensure the boundedness below from zero for the total variation of related probability distributions in Le Cam's lemma. Assumption (3.3) is reasonable in a sense that if the right-hand side of (3.3) is large enough, (3.3) will not impose very restrictive condition on $\pi_n(d)$. We evaluate the dominating factor $n^{-q/2}d^v$ on the right-hand side of (3.3) for various scenarios. First, consider $q = 0$, the assumption becomes $\pi_n(d) \leq \aleph d^v$, $v < 1/2$, and so assumption (3.3) essentially requires $\pi_n(d)$ grows in d not faster than $d^{1/2}$, which is not restrictive at all as $\pi_n(d)$ usually grows slowly in d. The asymptotic analysis of high-dimensional statistics usually allows both d and n go to infinity. Typically, we may assume d grows polynomially or exponentially in n. If d grows exponentially in n, that is, $d \sim \exp(b_0 n)$ for some $b_0 > 0$, then $n^{q/2}$ is negligible in comparison with d^v, and $n^{-q/2}d^v$ behavior like d^v. The assumption in this case is not very restrictive. For the case of polynomial growth, that is, $d \sim n^{b_1}$ for some $b_1 > 0$, then $n^{-q/2}d^v \sim d^{v-q/(2b_1)}$. If $v - q/(2b_1) > 0$, $n^{-q/2}d^v$ grows in d like some positive power of d. Since we may take v arbitrarily close to $1/2$, the positiveness of $v - q/(2b_1)$ essentially requires $b_1 > q$, which can often be quite realistic given that q is usually very small.

The theorem below provides a minimax lower bound for estimating $\boldsymbol{\rho}$ under the Frobenius norm.

THEOREM 3. *We assume that* $\pi_n(d)$ *in the sparsity condition* (2.5) *satisfies*

$$\pi_n(d) \leq \aleph' d^{v'}/n^q, \tag{3.4}$$

for some constants $\aleph' > 0$ *and* $0 < v' < 2$. *Then*

$$\inf_{\check{\boldsymbol{\rho}}} \sup_{\boldsymbol{\rho}\in\Theta} E\big[\|\check{\boldsymbol{\rho}} - \boldsymbol{\rho}\|_F^2\big] \geq c_5 \pi_n(d) \frac{1}{d}\left(\frac{\log d}{n}\right)^{1-q/2},$$

where $\check{\boldsymbol{\rho}}$ *denotes any estimator of* $\boldsymbol{\rho}$ *based on measurement data* $N_2, \dots, N_{d^2}$, *and* c_5 *is a constant free of* n *and* d.

REMARK 3. The lower bound in Theorem 3 matches the convergence rate of $\hat{\boldsymbol{\rho}}$ under the Frobenius norm in Theorem 1, so we conclude that $\hat{\boldsymbol{\rho}}$ achieves the optimal convergence rate under the Frobenius norm. Similar to the Remark 2 after Theorem 2, we need to apply Assouad's lemma to establish the minimax lower bound in Theorem 3, and assumption (3.4) is used to guarantee the positive definiteness of the constructed matrices as density matrix candidates and to ensure the boundedness below from zero for the total variation of related probability distributions in Assouad's lemma. Also the appropriateness of (3.4) is more relaxed than (3.3), as $v' < 2$ and the right-hand side of (3.4) has main powers more than the square of that of (3.3).

It is interesting to consider density matrix estimation under a Schatten norm, where given a matrix $\mathbf{A}$ of size d, we define its Schatten s-norm by

$$\|\mathbf{A}\|_{*s} = \left(\sum_{j=1}^{d} |\lambda_j|^s\right)^{1/s},$$

and $\lambda_1, \dots, \lambda_d$ are the eigenvalues of the square root of $\mathbf{A}^\dagger\mathbf{A}$. Spectral norm and Frobenius norm are two special cases of the Schatten s-norm with $s = 2$ and $s = \infty$, respectively, and the nuclear norm corresponds to the Schatten s-norm with $s = 1$. The following result provides the convergence rate for the proposed thresholding estimator under the Schatten s-norm loss for $1 \le s \le \infty$.

PROPOSITION 3. *Under the assumptions of Theorem 1, the density matrix estimator $\hat{\boldsymbol{\rho}}$ defined by (2.6)–(2.7) with threshold $\varpi = \hbar\sqrt{(4/n)\log d}$ for some constant $\hbar > 1$ satisfies*

$$\sup_{\boldsymbol{\rho}\in\Theta} E\big[\|\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}\|_{*s}^2\big] \le c\big[\pi_n(d)\big]^{2-2/\max(s,2)} \frac{1}{d^{2-2/s}} \left(\frac{\log d}{n}\right)^{1-q+q/\max(s,2)} \tag{3.5}$$

for $1 \le s \le \infty$, where c is a constant not depending on n and d.

The upper bound in (3.5) matches the minimax convergence rates for both the spectral norm and Frobenius norm. Moreover, for the case of the nuclear norm corresponding to the Schatten s-norm with $s = 1$, (3.5) leads to an upper bound with the convergence rate $\pi_n(d)(\frac{\log d}{n})^{1-q/2}$. We conjecture that the upper bounds in (3.5) are rate-optimal under the Schatten s-norm loss for all $1 \le s \le \infty$. However, establishing a matching lower bound for the general Schatten norm loss is a difficult task, and we believe that a new approach is needed for studying minimax density matrix estimation under the Schatten s-norm, particularly the nuclear norm.

REMARK 4. The Pauli basis expansion (2.4) is orthogonal with respect to the usual Euclidean inner product, and as in the proof of Lemma 3 we have

$$\|\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}\|_F^2 = \sum_{k=2}^{d^2} |\hat{\beta}_k - \beta_k|^2/d,$$

where $\hat{\beta}$ and $\hat{\boldsymbol{\rho}}$ are threshold estimators of β and $\boldsymbol{\rho}$, respectively. The sparse vector estimation problem is well studied under the Gaussian or sub-Gaussian noise case [Donoho and Johnstone (1994) and Zhang (2012)] and can be used to recover the minimax result for density matrix estimation under the Frobenius norm loss, because of orthogonality. In fact, our relatively simple proof of the minimax results for the Frobenius norm loss is essentially the same as the sparse vector estimation approach. However, such an equivalence between sparse density matrix estimation and sparse vector estimation breaks down for the general Schatten norm loss such as the commonly used spectral norm and nuclear norm losses. For the spectral norm loss, Lemma 3 in Section 5 provides a sharp upper bound for $E[\|\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}\|_2^2]$ through the ℓ_1-norm of $(\hat{\beta}_k - \beta_k)$, and the proof of the minimax lower bound in Theorem 2 relies on the property that the spectral norm is determined by the largest eigenvalue only. Such a special property allows us to reduce the problem to a simple subproblem and establish the lower bound under the spectral norm loss. The arguments cannot be applied to the case of the general Schatten norm loss in particular the nuclear norm loss. Moreover, instead of directly applying Lemma 3 and Remark 5 in Section 5 to derive upper bounds for the general Schatten norm loss, we use the obtained sharp upper bounds for the spectral norm and Frobenius norm losses together with moment inequalities to derive sharper upper bounds in Proposition 3. However, similar lower bounds are not available. Our analysis leads us to believe that it is not possible to use sparse vector estimation to recover minimax lower bound results for the general Schatten norm loss in particular for the spectral norm loss.

4. A simulation study. A simulation study was conducted to investigate the performance of the proposed density matrix estimator for the finite sample. We took $d = 32, 64, 128$ and generated a true density matrix $\boldsymbol{\rho}$ for each case as follows. $\boldsymbol{\rho}$ has an expansion over the Pauli basis

$$\boldsymbol{\rho} = d^{-1}\left(\mathbf{I}_d + \sum_{j=2}^{d^2} \beta_j \mathbf{B}_j\right), \tag{4.1}$$

where $\beta_j = \operatorname{tr}(\boldsymbol{\rho}\mathbf{B}_j)$, $j = 2, \dots, d^2$. From $\beta_2, \dots, \beta_{d^2}$, we randomly selected $[6 \log d]$ coefficients β_j and set the rest of β_j to be zero. We simulated $[6 \log d]$ values independently from a uniform distribution on $[-0.2, 0.2]$, assigned the simulated values at random to the selected β_j, and then constructed $\boldsymbol{\rho}$ from (4.1). The constructed $\boldsymbol{\rho}$ always has unit trace but may not be positive semi-definite. The procedure was repeated until we generated a positive semi-definite $\boldsymbol{\rho}$. We took it as the true density matrix. The simulation procedure guarantees the obtained $\boldsymbol{\rho}$ is a density matrix and has a sparse representation under the Pauli basis.

For each true density matrix $\boldsymbol{\rho}$, as described in Section 2.2 we simulated data N_j from a binomial distribution with cell probability β_j and the number of cells $n = 100, 200, 500, 1000, 2000$. We constructed coefficient estimators $\hat{\beta}_j$ by (2.6) and obtained density matrix estimator $\hat{\boldsymbol{\rho}}$ using (2.7). The whole estimation procedure is repeated 200 times. The density matrix estimator is measured by the mean squared errors (MSE), $E\|\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}\|_2^2$ and $E\|\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}\|_F^2$, that are evaluated by the average of $\|\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}\|_2^2$ and $\|\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}\|_F^2$ over 200 repetitions, respectively. Three thresholds were used in the simulation study: the universal threshold $1.01\sqrt{4 \log d/n}$ for all β_j, the individual threshold $1.01\sqrt{4(1 - N_j^2)\log d/n}$ for each β_j, and the optimal threshold for all β_j, which minimizes the computed MSE for each corresponding hard or soft threshold method. The individual threshold takes into account the fact in Theorem 2 that the mean and variance of N_j are β_j and $(1-\beta_j^2)/n$, respectively, and the variance of N_j is estimated by $(1 - N_j^2)/n$.

Figures 1 and 2 plot the MSEs of the density matrix estimators with hard and soft threshold rules and its corresponding density matrix estimator without thresh-

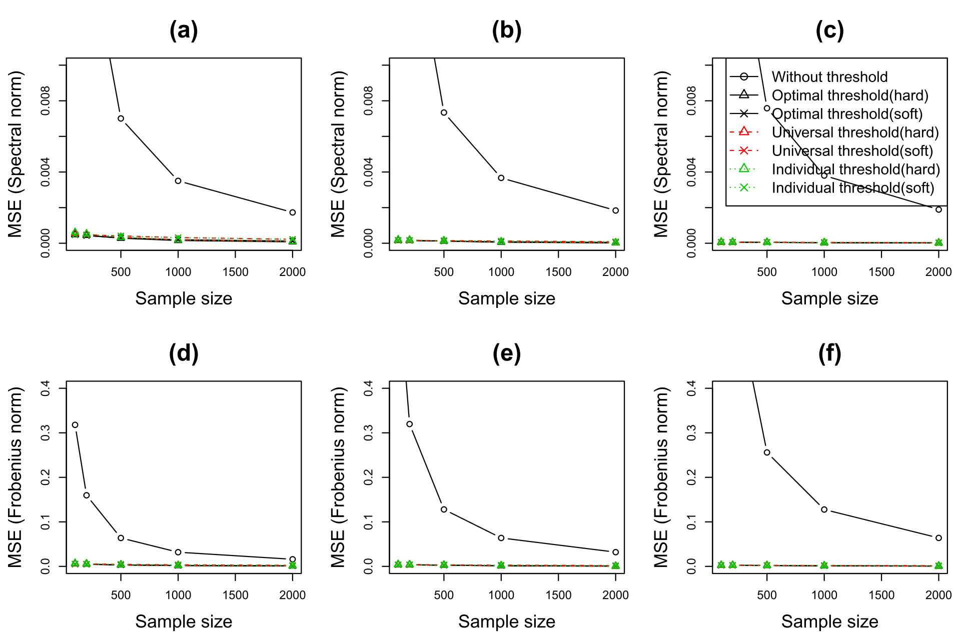

FIG. 1. *The MSE plots against sample size for the proposed density estimator with hard and soft threshold rules and its corresponding estimator without thresholding for* $d = 32, 64, 128$. (a)–(c) *are plots of MSEs based on the spectral norm for* $d = 32, 64, 128$, *respectively, and* (d)–(f) *are plots of MSEs based on the Frobenius norm for* $d = 32, 64, 128$, *respectively.*

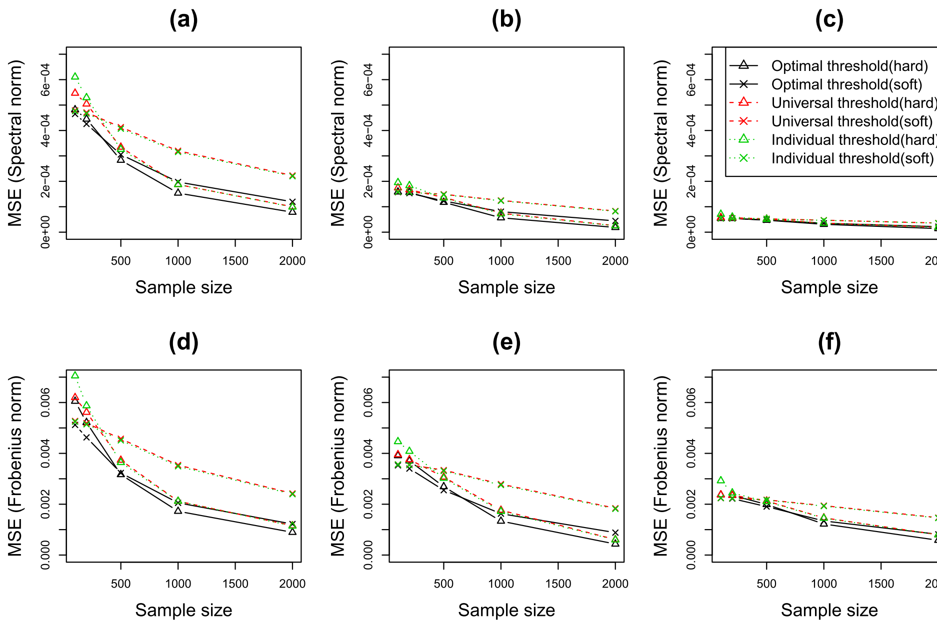

FIG. 2. *The MSE plots against sample size for the proposed density estimator with hard and soft threshold rules for* $d = 32, 64, 128$. (a)–(c) *are plots of MSEs based on the spectral norm for* $d = 32, 64, 128$, *respectively, and* (d)–(f) *are plots of MSEs based on the Frobenius norm for* $d = 32, 64, 128$, *respectively.*

olding [i.e., β_j are estimated by N_j in (2.7)] against the sample size n for different matrix size d, and Figures 3 and 4 plot their MSEs against matrix size d for different sample size. The numerical values of the MSEs are reported in Table 1. Figures 1 and 2 show that the MSEs usually decrease in sample size n, and the thresholding density matrix estimators enjoy superior performances than that the density matrix estimator without thresholding even for $n = 2000$; while all threshold rules and threshold values yield thresholding density matrix estimators with very close MSEs, the soft threshold rule with individual and universal threshold values produce larger MSEs than others for larger sample size such as $n = 1000, 2000$ and the soft threshold rule tends to give somewhat better performance than the hard threshold rule for smaller sample size like $n = 100, 200$. Figures 3 and 4 demonstrate that while the MSEs of all thresholding density matrix estimators decrease in the matrix size d, but if we rescale the MSEs by multiplying it with d^2 for the spectral norm case and d for the Frobenius norm case, the rescaled MSEs slowly increase in matrix size d. The simulation results largely confirm the theoretical findings discussed in Remark 1.

5. Proofs. Let $p = d^2$. Denote by C's generic constants whose values are free of n and p and may change from appearance to appearance. Let $u \vee v$

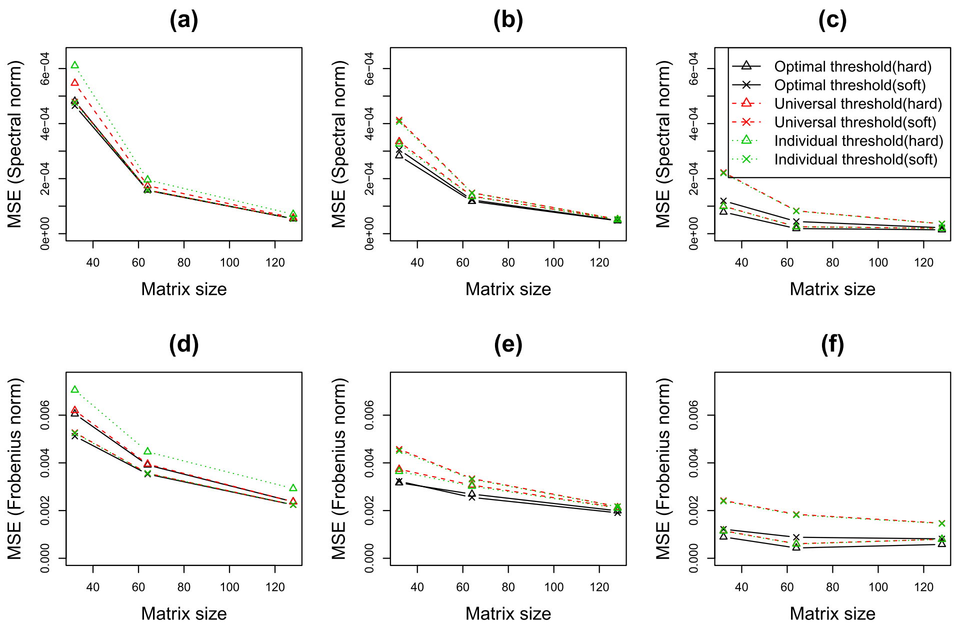

FIG. 3. *The MSE plots against matrix size for the proposed density estimator with hard and soft threshold rules for* $n = 100, 500, 2000$. (a)–(c) *are plots of MSEs based on the spectral norm for* $n = 100, 500, 2000$, *respectively, and* (d)–(f) *are plots of MSEs based on the Frobenius norm for* $n = 100, 500, 2000$, *respectively.*

and $u \wedge v$ be the maximum and minimum of u and v, respectively. For two sequences $u_{n,p}$ and $v_{n,p}$, we write $u_{n,p} \sim v_{n,p}$ if $u_{n,p}/v_{n,p} \to 1$ as $n, p \to \infty$, and write $u_{n,p} \asymp v_{n,p}$ if there exist positive constants C_1 and C_2 free of n and p such that $C_1 \le u_{n,p}/v_{n,p} \le C_2$. Without confusion we may write $\pi_n(d)$ as $\pi_n(p)$.

5.1. *Proofs of Propositions* 1 *and* 2.

PROOF OF PROPOSITION 1. In two dimensions, Pauli matrices satisfy $\mathrm{tr}(\boldsymbol{\sigma}_0) = 2$, and $\mathrm{tr}(\boldsymbol{\sigma}_1) = \mathrm{tr}(\boldsymbol{\sigma}_2) = \mathrm{tr}(\boldsymbol{\sigma}_3) = 0$, $\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2, \boldsymbol{\sigma}_3$ have eigenvalues ± 1, the square of a Pauli matrix is equal to the identity matrix, and the multiplications of any two Pauli matrices are equal to the third Pauli matrix multiplying by $\sqrt{-1}$, for example, $\boldsymbol{\sigma}_1\boldsymbol{\sigma}_2 = \sqrt{-1}\boldsymbol{\sigma}_3$, $\boldsymbol{\sigma}_2\boldsymbol{\sigma}_3 = \sqrt{-1}\boldsymbol{\sigma}_1$, and $\boldsymbol{\sigma}_3\boldsymbol{\sigma}_1 = \sqrt{-1}\boldsymbol{\sigma}_2$.

For $j = 2, \dots, p$, consider $\mathbf{B}_j = \boldsymbol{\sigma}_{\ell_1} \otimes \boldsymbol{\sigma}_{\ell_2} \otimes \cdots \otimes \boldsymbol{\sigma}_{\ell_b}$. $\mathrm{tr}(\mathbf{B}_j) = \mathrm{tr}(\boldsymbol{\sigma}_{\ell_1}) \times \mathrm{tr}(\boldsymbol{\sigma}_{\ell_2}) \cdots \mathrm{tr}(\boldsymbol{\sigma}_{\ell_b}) = 0$, and $\mathbf{B}_j$ has eigenvalues ± 1, $\mathbf{B}_j^2 = \mathbf{I}_d$.

For $j, j' = 2, \dots, p$, $j \ne j'$, $\mathbf{B}_j = \boldsymbol{\sigma}_{\ell_1} \otimes \boldsymbol{\sigma}_{\ell_2} \otimes \cdots \otimes \boldsymbol{\sigma}_{\ell_b}$ and $\mathbf{B}_{j'} = \boldsymbol{\sigma}_{\ell_1'} \otimes \boldsymbol{\sigma}_{\ell_2'} \otimes \cdots \otimes \boldsymbol{\sigma}_{\ell_b'}$,

$$\mathbf{B}_j\mathbf{B}_{j'} = [\boldsymbol{\sigma}_{\ell_1}\boldsymbol{\sigma}_{\ell_1'}] \otimes [\boldsymbol{\sigma}_{\ell_2}\boldsymbol{\sigma}_{\ell_2'}] \otimes \cdots \otimes [\boldsymbol{\sigma}_{\ell_b}\boldsymbol{\sigma}_{\ell_b'}],$$

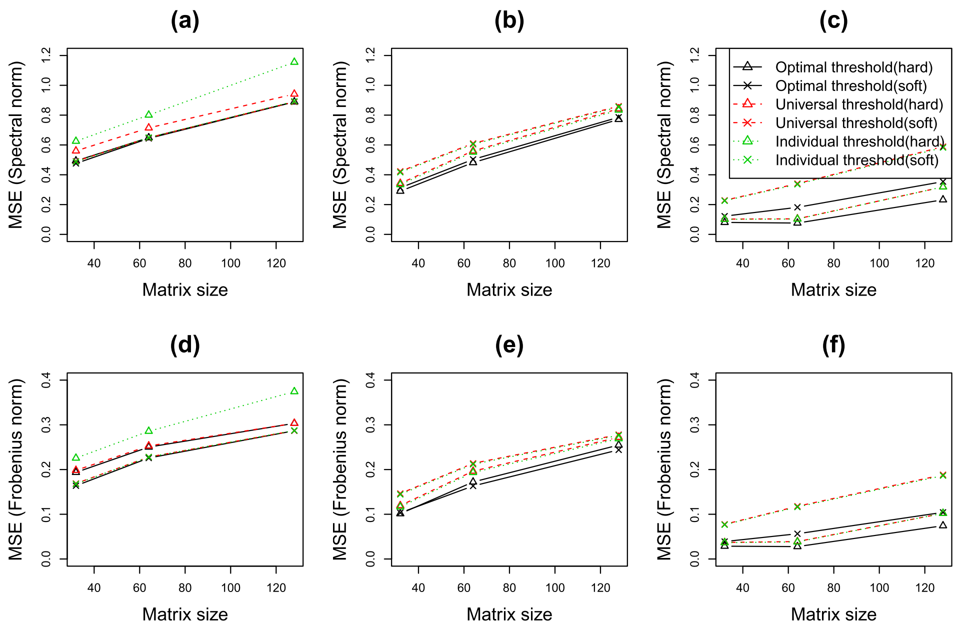

FIG. 4. *The plots of MSEs multiplying by d or d^2 against matrix size d for the proposed density estimator with hard and soft threshold rules for $n = 100, 500, 2000$.* (a)–(c) *are plots of d^2 times of MSEs based on the spectral norm for $n = 100, 500, 2000$, respectively, and* (d)–(f) *are plots of d times of MSEs based on the Frobenius norm for $n = 100, 500, 2000$, respectively.*

is equal to a d dimensional Pauli matrix multiplying by $(\sqrt{-1})^b$, which has zero trace. Thus, $\mathrm{tr}(\mathbf{B}_j\mathbf{B}_{j'}) = 0$, that is, $\mathbf{B}_j$ and $\mathbf{B}_{j'}$ are orthogonal, and $\mathbf{B}_1, \dots, \mathbf{B}_p$ form an orthogonal basis. $\mathrm{tr}(\boldsymbol{\rho}\mathbf{B}_j/d) = \beta_k \,\mathrm{tr}(\mathbf{B}_j^2)/d = \beta_k$. In particular $\mathbf{B}_1 = \mathbf{I}_d$, and $\beta_1 = \mathrm{tr}(\boldsymbol{\rho}\mathbf{B}_1) = \mathrm{tr}(\boldsymbol{\rho}) = 1$.

Denote by $\mathbf{Q}_{j\pm}$ the projections onto the eigenspaces corresponding to eigenvalues ± 1, respectively. Then for $j = 2, \dots, p$,

$$\mathbf{B}_j = \mathbf{Q}_{j+} - \mathbf{Q}_{j-}, \qquad \mathbf{B}_j^2 = \mathbf{Q}_{j+} + \mathbf{Q}_{j-} = \mathbf{I}_d, \qquad \mathbf{B}_j\mathbf{Q}_{j\pm} = \pm\mathbf{Q}_{j\pm}^2 = \pm\mathbf{Q}_{j\pm},$$
$$0 = \mathrm{tr}(\mathbf{B}_j) = \mathrm{tr}(\mathbf{Q}_{j+}) - \mathrm{tr}(\mathbf{Q}_{j-}), \qquad d = \mathrm{tr}(\mathbf{I}_d) = \mathrm{tr}(\mathbf{Q}_{j+}) + \mathrm{tr}(\mathbf{Q}_{j-}),$$

and solving the equations we get

$$\mathrm{tr}(\mathbf{Q}_{j\pm}) = d/2, \qquad \mathrm{tr}(\mathbf{B}_j\mathbf{Q}_{j\pm}) = \pm\,\mathrm{tr}(\mathbf{Q}_{j\pm}) = \pm d/2. \tag{5.1}$$

For $j \neq j'$, $j, j' = 2, \dots, p$, $\mathbf{B}_j$ and $\mathbf{B}_{j'}$ are orthogonal,

$$0 = \mathrm{tr}(\mathbf{B}_{j'}\mathbf{B}_j) = \mathrm{tr}(\mathbf{B}_{j'}\mathbf{Q}_{j+}) - \mathrm{tr}(\mathbf{B}_{j'}\mathbf{Q}_{j-}),$$

and

$$\mathbf{B}_{j'}\mathbf{Q}_{j+} + \mathbf{B}_{j'}\mathbf{Q}_{j-} = \mathbf{B}_{j'}(\mathbf{Q}_{j+} + \mathbf{Q}_{j-}) = \mathbf{B}_{j'},$$
$$\mathrm{tr}(\mathbf{B}_{j'}\mathbf{Q}_{j+}) + \mathrm{tr}(\mathbf{B}_{j'}\mathbf{Q}_{j-}) = \mathrm{tr}(\mathbf{B}_{j'}) = 0,$$

TABLE 1

MSEs based on spectral and Frobenius norms of the density estimator defined by (2.6) *and* (2.7) *and its corresponding density matrix estimator without thresholding, and threshold values used for* $d = 32, 64, 128$, *and* $n = 100, 200, 500, 1000, 2000$

		MSE (Spectral norm) $\times 10^4$							**Threshold value (ϖ) $\times 10^2$**		
		Without threshold	**Optimal threshold**		**Universal threshold**		**Individual threshold**		**Universal**	**Optimal**	
d	***n***	**Density estimator**	**Hard**	**Soft**	**Hard**	**Soft**	**Hard**	**Soft**	**Universal**	**Hard**	**Soft**
32	100	348.544	4.816	4.648	5.468	4.790	6.104	4.762	24.782	15.180	0.619
	200	175.034	4.449	4.257	5.043	4.708	5.293	4.667	17.524	7.739	0.562
	500	70.069	2.831	3.054	3.344	4.130	3.260	4.071	11.083	2.397	0.373
	1000	35.028	1.537	1.974	1.875	3.201	1.875	3.155	7.837	1.099	0.212
	2000	17.307	0.785	1.195	1.001	2.230	0.989	2.200	5.541	0.551	0.116
64	100	368.842	1.583	1.572	1.744	1.583	1.954	1.586	27.148	16.660	0.395
	200	183.050	1.565	1.534	1.669	1.575	1.833	1.571	19.196	9.252	0.376
	500	73.399	1.175	1.228	1.367	1.490	1.347	1.476	12.141	2.900	0.307
	1000	36.692	0.566	0.807	0.747	1.249	0.722	1.233	8.585	1.308	0.177
	2000	18.402	0.186	0.443	0.255	0.832	0.251	0.820	6.070	0.657	0.061
128	100	381.032	0.543	0.542	0.574	0.543	0.705	0.545	29.323	17.500	0.237
	200	190.113	0.541	0.539	0.570	0.542	0.594	0.542	20.734	10.246	0.235
	500	75.824	0.471	0.480	0.514	0.525	0.509	0.522	13.114	3.547	0.213
	1000	38.010	0.309	0.350	0.355	0.470	0.354	0.466	9.273	1.613	0.146
	2000	18.907	0.142	0.216	0.194	0.359	0.194	0.356	6.557	0.725	0.080

TABLE 1
(*Continued*)

		MSE (Frobenius norm) $\times 10^3$							Threshold value (ϖ) $\times 10^2$		
		Without threshold	**Optimal threshold**		**Universal threshold**		**Individual threshold**		**Universal**	**Optimal**	
d	***n***	**Density estimator**	**Hard**	**Soft**	**Hard**	**Soft**	**Hard**	**Soft**	**Universal**	**Hard**	**Soft**
32	100	317.873	6.052	5.119	6.195	5.274	7.050	5.246	24.782	11.004	9.936
	200	159.679	5.217	4.629	5.616	5.187	5.874	5.143	17.524	5.681	3.771
	500	63.823	3.165	3.229	3.732	4.575	3.642	4.512	11.083	2.286	0.954
	1000	31.856	1.722	2.053	2.119	3.540	2.119	3.492	7.837	1.100	0.401
	2000	15.967	0.894	1.219	1.155	2.424	1.141	2.394	5.541	0.546	0.182
64	100	641.437	3.909	3.528	3.951	3.563	4.463	3.562	27.148	13.719	13.234
	200	319.720	3.706	3.401	3.755	3.548	4.082	3.536	19.196	7.042	5.515
	500	127.958	2.691	2.551	3.069	3.342	3.023	3.309	12.141	2.800	1.275
	1000	63.845	1.335	1.628	1.765	2.791	1.717	2.756	8.585	1.277	0.548
	2000	31.952	0.433	0.882	0.610	1.842	0.596	1.817	6.070	0.647	0.258
128	100	1283.182	2.370	2.240	2.370	2.242	2.924	2.245	29.323	15.989	16.128
	200	639.556	2.349	2.219	2.354	2.238	2.444	2.238	20.734	8.218	7.799
	500	255.954	1.990	1.906	2.125	2.172	2.102	2.160	13.114	3.355	1.773
	1000	127.714	1.221	1.341	1.463	1.943	1.448	1.924	9.273	1.546	0.729
	2000	63.921	0.581	0.815	0.798	1.471	0.798	1.456	6.557	0.719	0.327

which imply

$$\operatorname{tr}(\mathbf{B}_{j'}\mathbf{Q}_{j\pm}) = 0, \qquad j \neq j',\, j, j' = 2, \dots, p.$$

For a density matrix $\boldsymbol{\rho}$ with representation (2.4) under the Pauli basis (2.3), from (5.1) we have $\operatorname{tr}(\mathbf{Q}_{k\pm}) = d/2$ and $\operatorname{tr}(\mathbf{B}_k\mathbf{Q}_{k\pm}) = \pm d/2$, and thus

$$\begin{aligned}\operatorname{tr}(\boldsymbol{\rho}\mathbf{Q}_{k\pm}) &= \frac{1}{d}\operatorname{tr}(\mathbf{Q}_{k\pm}) + \sum_{j=2}^{p} \frac{\beta_j}{d}\operatorname{tr}(\mathbf{B}_j\mathbf{Q}_{k\pm}) \\ &= \frac{1}{2} + \frac{\beta_k}{d}\operatorname{tr}(\mathbf{B}_k\mathbf{Q}_{k\pm}) = \frac{1 \pm \beta_k}{2}.\end{aligned} \tag{5.2}$$

□

PROOF OF PROPOSITION 2. We perform measurements on each Pauli observable $\mathbf{B}_k$ independently for n quantum systems that are identically prepared under state $\boldsymbol{\rho}$. Denote by $R_{k1}, \dots, R_{kn}$ the n measurement outcomes for measuring $\mathbf{B}_k$, $k = 2, \dots, p$.

$$N_k = (R_{k1} + \cdots + R_{kn})/n, \tag{5.3}$$

$R_{k\ell}$, $k = 2, \dots, p$, $\ell = 1, \dots, n$, are independent, and take values ± 1, with distributions given by

$$P(R_{k\ell} = \pm 1) = \operatorname{tr}(\boldsymbol{\rho}\mathbf{Q}_{k\pm}), \qquad k = 2, \dots, p, \ell = 1, \dots, n. \tag{5.4}$$

As random variables $R_{k1}, \dots, R_{kn}$ are i.i.d. and take eigenvalues ± 1, $n(N_k + 1)/2 = \sum_{\ell=1}^{n}(R_{k\ell} + 1)/2$ is equal to the total number of random variables $R_{k1}, \dots, R_{kn}$ taking eigenvalue 1, and thus $n(N_k + 1)/2$ follows a binomial distribution with n trials and cell probability $P(R_{k1} = 1) = \operatorname{tr}(\boldsymbol{\rho}\mathbf{Q}_{k+})$. From (5.3)–(5.4) and Proposition 1, we have for $k = 2, \dots, p$,

$$\begin{aligned}&\operatorname{tr}(\boldsymbol{\rho}\mathbf{Q}_{k+}) = \frac{1 + \beta_k}{2}, \qquad E(N_k) = E(R_{k1}) = \operatorname{tr}(\boldsymbol{\rho}\mathbf{B}_k) = \beta_k \operatorname{tr}(\mathbf{B}_k^2)/d = \beta_k, \\ &\operatorname{Var}(N_k) = \frac{1 - \beta_k^2}{n}.\end{aligned}$$

□

5.2. *Proof of Theorem 1: Upper bound.*

LEMMA 1. *If β_j satisfy sparsity condition (2.5), then for any a,*

$$\sum_{j=2}^{p} |\beta_j| 1\big(|\beta_j| \leq a\varpi\big) \leq a^{1-q}\pi_n(p)\varpi^{1-q},$$

$$\sum_{j=2}^{p} 1\big(|\beta_j| \geq a\varpi\big) \leq a^{-q}\pi_n(p)\varpi^{-q}.$$

PROOF. Simple algebraic manipulation shows

$$\sum_{j=2}^{p} |\beta_j| 1\big(|\beta_j| \le a\varpi\big) \le (a\varpi)^{1-q} \sum_{j=2}^{p} |\beta_j|^q 1\big(|\beta_j| \le a\varpi\big)$$
$$\le a^{1-q} \pi_n(p) \varpi^{1-q},$$

and

$$\sum_{j=2}^{p} 1\big(|\beta_j| \ge a\varpi\big) \le \sum_{j=2}^{p} \big[|\beta_j|/(a\varpi)\big]^q 1\big(|\beta_j| \ge a\varpi\big)$$
$$\le (a\varpi)^{-q} \sum_{j=2}^{p} |\beta_j|^q \le a^{-q} \pi_n(p) \varpi^{-q}. \qquad \square$$

LEMMA 2. *With* $\varpi = \hbar n^{-1/2}\sqrt{2\log p}$ *for some positive constant* $\hbar$, *we have for any* $a \ne 1$,

$$P\big(N_j - \beta_j \le -|a-1|\varpi\big) \le 2p^{-(\hbar^2|a-1|^2)/(1+o(1))},$$
$$P\big(N_j - \beta_j \ge |a-1|\varpi\big) \le 2p^{-(\hbar^2|a-1|^2)/(1+o(1))}.$$

PROOF. From Proposition 2 and (5.3)–(5.4), we have that N_j is the average of $R_{j1}, \ldots, R_{jn}$, which are i.i.d. random variables taking values ± 1, $P(R_{j1} = \pm 1) = (1 \pm \beta_j)/2$, $E(R_{j1}) = \beta_j$ and $\mathrm{Var}(R_{j1}) = 1 - \beta_j^2$. Applying Bernstein's inequality, we obtain for any $x > 0$,

$$P\big(|N_j - \beta_j| \ge x\big) \le 2\exp\biggl(-\frac{nx^2}{2(1-\beta_j^2 + x/3)}\biggr) \le 2\exp\biggl(-\frac{nx^2}{2(1+x/3)}\biggr).$$

Both $P(N_j - \beta_j \le -|a-1|\varpi)$ and $P(N_j - \beta_j \ge |a-1|\varpi)$ are less than $P(|N_j - \beta_j| \ge |a-1|\varpi)$, which is bounded by

$$2\exp\biggl(-\frac{n|a-1|^2\varpi^2}{2(1+|a-1|\varpi/3)}\biggr) = 2\exp\biggl(-\frac{\hbar^2|a-1|^2\log p}{1+o(1)}\biggr)$$
$$= 2p^{-(\hbar^2|a-1|^2)/(1+o(1))}. \qquad \square$$

LEMMA 3.

$$E\|\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}\|_F^2 = p^{-1/2} \sum_{j=2}^{p} E|\hat{\beta}_j - \beta_j|^2, \tag{5.5}$$

$$p^{1/2} E\|\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}\|_2 \le \sum_{j=2}^{p} E|\hat{\beta}_j - \beta_j|, \tag{5.6}$$

$$
\begin{aligned}
pE\|\hat{\boldsymbol{\rho}}-\boldsymbol{\rho}\|_2^2 \le \sum_{j=2}^{p} E\big[|\hat{\beta}_j-\beta_j|^2\big] + \Biggl\{\sum_{j=2}^{p} E\big[|\hat{\beta}_j-\beta_j|\big]\Biggr\}^2 \\
- \sum_{j=2}^{p}\bigl\{E\bigl(|\hat{\beta}_j-\beta_j|\bigr)\bigr\}^2.
\end{aligned} \tag{5.7}
$$

PROOF. Since Pauli matrices $\mathbf{B}_j$ are orthogonal with respect to the usual Euclidean inner product, with $\|\mathbf{B}_j\|_F = d^{1/2}$, and $\|\mathbf{B}_j\|_2 = 1$, we have

$$
\begin{aligned}
\|\hat{\boldsymbol{\rho}}-\boldsymbol{\rho}\|_F^2 &= \Biggl\|\sum_{j=2}^{p}(\hat{\beta}_j-\beta_j)\mathbf{B}_j\Biggr\|_F^2 \Big/ d^2 = \sum_{j=2}^{p}|\hat{\beta}_j-\beta_j|^2\|\mathbf{B}_j\|_F^2/d^2 \\
&= \sum_{j=2}^{p}|\hat{\beta}_j-\beta_j|^2/d,
\end{aligned} \tag{5.8}
$$

$$
\begin{aligned}
p^{1/2}\|\hat{\boldsymbol{\rho}}-\boldsymbol{\rho}\|_2 &= \Biggl\|\sum_{j=2}^{p}(\hat{\beta}_j-\beta_j)\mathbf{B}_j\Biggr\|_2 \le \sum_{j=2}^{p}|\hat{\beta}_j-\beta_j|\|\mathbf{B}_j\|_2 \\
&= \sum_{j=2}^{p}|\hat{\beta}_j-\beta_j|,
\end{aligned} \tag{5.9}
$$

$$
\begin{aligned}
p\|\hat{\boldsymbol{\rho}}-\boldsymbol{\rho}\|_2^2 &= \Biggl\|\sum_{j=2}^{p}(\hat{\beta}_j-\beta_j)\mathbf{B}_j\Biggr\|_2^2 \\
&\le \sum_{j=2}^{p}|\hat{\beta}_j-\beta_j|^2\|\mathbf{B}_j\|_2^2 + 2\sum_{i<j}^{p}\bigl|(\hat{\beta}_i-\beta_i)(\hat{\beta}_j-\beta_j)\bigr|\|\mathbf{B}_i\mathbf{B}_j\|_2 \\
&\le \sum_{j=2}^{p}|\hat{\beta}_j-\beta_j|^2\|\mathbf{B}_j\|_2^2 + 2\sum_{i<j}^{p}\bigl|(\hat{\beta}_i-\beta_i)(\hat{\beta}_j-\beta_j)\bigr|\|\mathbf{B}_i\|_2\|\mathbf{B}_j\|_2 \\
&= \sum_{j=2}^{p}|\hat{\beta}_j-\beta_j|^2 + 2\sum_{i<j}^{p}\bigl|(\hat{\beta}_i-\beta_i)(\hat{\beta}_j-\beta_j)\bigr|.
\end{aligned} \tag{5.10}
$$

As $N_2,\dots,N_p$ are independent, $\hat{\beta}_2,\dots,\hat{\beta}_p$ are independent. Thus, from (5.8)–(5.10) we obtain (5.5)–(5.6), and

$$
\begin{aligned}
&pE\|\hat{\boldsymbol{\rho}}-\boldsymbol{\rho}\|_2^2 \\
&\quad \le \sum_{j=2}^{p} E|\hat{\beta}_j-\beta_j|^2 + 2\sum_{i<j}^{p} E\bigl|(\hat{\beta}_i-\beta_i)(\hat{\beta}_j-\beta_j)\bigr|
\end{aligned}
$$

$$= \sum_{j=2}^{p} E|\hat{\beta}_j - \beta_j|^2 + 2\sum_{i<j}^{p} E|\hat{\beta}_i - \beta_i| E|\hat{\beta}_j - \beta_j|$$

$$= \sum_{j=2}^{p} E\big[|\hat{\beta}_j - \beta_j|^2\big] + \left\{\sum_{j=2}^{p} E\big[|\hat{\beta}_j - \beta_j|\big]\right\}^2 - \sum_{j=2}^{p}\big\{E\big(|\hat{\beta}_j - \beta_j|\big)\big\}^2. \qquad \square$$

REMARK 5. Since Pauli matrices $\mathbf{B}_j$ have eigenvalues ± 1, the Schatten s-norm $\|\mathbf{B}_j\|_{*s} = d^{1/s}$. Similar to (5.9)–(5.10), we obtain that

$$p^{1/2}\|\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}\|_{*s} \le \sum_{j=2}^{p} |\hat{\beta}_j - \beta_j| \|\mathbf{B}_j\|_{*s} = d^{1/s} \sum_{j=2}^{p} |\hat{\beta}_j - \beta_j|, \tag{5.11}$$

$$\begin{aligned} p\|\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}\|_{*s}^2 &\le d^{2/s}\left[\sum_{j=2}^{p} |\hat{\beta}_j - \beta_j|\right]^2 \\ &= d^{2/s}\left[\sum_{j=2}^{p} |\hat{\beta}_j - \beta_j|^2 + 2\sum_{i<j}^{p} \big|(\hat{\beta}_i - \beta_i)(\hat{\beta}_j - \beta_j)\big|\right]. \end{aligned} \tag{5.12}$$

LEMMA 4.

$$\sum_{j=2}^{p} E|\hat{\beta}_j - \beta_j| \le C_1 \pi_n(d) \varpi^{1-q}, \tag{5.13}$$

$$\sum_{j=2}^{p} \big[E|\hat{\beta}_j - \beta_j|\big]^2 \le \sum_{j=2}^{p} E\big[|\hat{\beta}_j - \beta_j|^2\big] \le C_2 \pi_n(d) \varpi^{2-q}. \tag{5.14}$$

PROOF. Using (2.6), we have

$$\begin{aligned} &E|\hat{\beta}_j - \beta_j| \\ &\quad \le E\big[\big(|N_j - \beta_j| + \varpi\big)1\big(|N_j| \ge \varpi\big)\big] + |\beta_j| P\big(|N_j| \le \varpi\big) \\ &\quad \le \big[E|N_j - \beta_j|^2 P\big(|N_j| \ge \varpi\big)\big]^{1/2} + \varpi P\big(|N_j| \ge \varpi\big) + |\beta_j| P\big(|N_j| \le \varpi\big) \\ &\quad \le \big[n^{-1}(1 - \beta_j^2) P\big(|N_j| \ge \varpi\big)\big]^{1/2} + \varpi P\big(|N_j| \ge \varpi\big) + |\beta_j| P\big(|N_j| \le \varpi\big) \\ &\quad \le 2\varpi\big[P\big(|N_j| \ge \varpi\big)\big]^{1/2} + |\beta_j| P\big(|N_j| \le \varpi\big) \\ &\quad = 2\varpi\big[P\big(|N_j| \ge \varpi\big)\big]^{1/2}\big\{1\big(|\beta_j| > a_1\varpi\big) + 1\big(|\beta_j| \le a_1\varpi\big)\big\} \\ &\qquad + |\beta_j| P\big(|N_j| \le \varpi\big)\big\{1\big(|\beta_j| > a_2\varpi\big) + 1\big(|\beta_j| \le a_2\varpi\big)\big\} \\ &\quad \le 2\varpi 1\big(|\beta_j| > a_1\varpi\big) + 2\varpi\big[P\big(|N_j| \ge \varpi\big)\big]^{1/2} 1\big(|\beta_j| \le a_1\varpi\big) \\ &\qquad + P\big(|N_j| \le \varpi\big)1\big(|\beta_j| > a_2\varpi\big) + |\beta_j| 1\big(|\beta_j| \le a_2\varpi\big), \end{aligned}$$

where a_1 and a_2 are two constants satisfying $a_1 < 1 < a_2$ whose values will be chosen later, and

$$
\begin{aligned}
\sum_{j=2}^{p} E|\hat{\beta}_j - \beta_j| &\le 2\varpi \sum_{j=2}^{p} 1(|\beta_j| > a_1\varpi) \\
&\quad + 2\varpi \sum_{j=2}^{p} [P(|N_j| \ge \varpi)]^{1/2} 1(|\beta_j| \le a_1\varpi) \qquad (5.15)\\
&\quad + \sum_{j=2}^{p} P(|N_j| \le \varpi) 1(|\beta_j| > a_2\varpi) + \sum_{j=2}^{p} |\beta_j| 1(|\beta_j| \le a_2\varpi).
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
&[E(|\hat{\beta}_j - \beta_j|)]^2 \\
&\quad \le E[|\hat{\beta}_j - \beta_j|^2] \\
&\quad \le E[2(|N_j - \beta_j|^2 + \varpi^2) 1(|N_j| \ge \varpi)] + |\beta_j|^2 P(|N_j| \le \varpi) \\
&\quad \le 2[E|N_j - \beta_j|^4 P(|N_j| \ge \varpi)]^{1/2} \\
&\qquad + 2\varpi^2 P(|N_j| \ge \varpi) + |\beta_j|^2 P(|N_j| \le \varpi) \\
&\quad \le c\varpi^2 [P(|N_j| \ge \varpi)]^{1/2} + |\beta_j|^2 P(|N_j| \le \varpi) \\
&\quad = c\varpi^2 [P(|N_j| \ge \varpi)]^{1/2} \{1(|\beta_j| > a_1\varpi) + 1(|\beta_j| \le a_1\varpi)\} \\
&\qquad + |\beta_j|^2 P(|N_j| \le \varpi)[1(|\beta_j| > a_2\varpi) + 1(|\beta_j| \le a_2\varpi)] \\
&\quad \le c\varpi^2 1(|\beta_j| > a_1\varpi) + c\varpi^2 [P(|N_j| \ge \varpi)]^{1/2} 1(|\beta_j| \le a_1\varpi) \\
&\qquad + P(|N_j| \le \varpi) 1(|\beta_j| > a_2\varpi) + |\beta_j|^2 1(|\beta_j| \le a_2\varpi),
\end{aligned}
$$

and

$$
\begin{aligned}
&\sum_{j=2}^{p} E[|\hat{\beta}_j - \beta_j|^2] \\
&\quad \le c\varpi^2 \sum_{j=2}^{p} 1(|\beta_j| > a_1\varpi) \\
&\qquad + c\varpi^2 \sum_{j=2}^{p} [P(|N_j| \ge \varpi)]^{1/2} 1(|\beta_j| \le a_1\varpi) \qquad (5.16)\\
&\qquad + \sum_{j=2}^{p} P(|N_j| \le \varpi) 1(|\beta_j| > a_2\varpi) + \sum_{j=2}^{p} |\beta_j|^2 1(|\beta_j| \le a_2\varpi).
\end{aligned}
$$

By Lemma 1, we have

$$\sum_{j=2}^{p} |\beta_j| 1(|\beta_j| \le a_2\varpi) \le a_2^{1-q} \pi_n(d) \varpi^{1-q}, \tag{5.17}$$

$$\begin{aligned} &\sum_{j=2}^{p} |\beta_j|^2 1(|\beta_j| \le a_2\varpi) \\ &\quad \le (a_2\varpi)^{2-q} \sum_{j=2}^{p} |\beta_j|^q 1(|\beta_j| \le a_2\varpi) \le a_2^{2-q} \pi_n(d) \varpi^{2-q}, \end{aligned} \tag{5.18}$$

$$\varpi \sum_{j=2}^{p} 1(|\beta_j| \ge a_1\varpi) \le a_1^{-q} \pi_n(d) \varpi^{1-q}. \tag{5.19}$$

On the other hand,

$$\begin{aligned} &\sum_{j=2}^{p} P(|N_j| \le \varpi) 1(|\beta_j| > a_2\varpi) \\ &\quad \le \sum_{j} P(-\varpi - \beta_j \le N_j - \beta_j \le \varpi - \beta_j) 1(|\beta_j| > a_2\varpi) \\ &\quad \le \sum_{j=2}^{p} [P(N_j - \beta_j \le -|a_2 - 1|\varpi) + P(N_j - \beta_j \ge |a_2 - 1|\varpi)] \\ &\quad \le 4p^{1-(\hbar^2|a_2-1|^2)/(1+o(1))} = 4p^{-1-(2-q)/(2c_0)} \le 4p^{-1} n^{-(q-2)/2} \\ &\quad = o(\pi_n(d)\varpi^{2-q}), \end{aligned} \tag{5.20}$$

where the third inequality is from Lemma 2, the first equality is due the fact that we take $a_2 = 1 + \{2 + (2-q)/(2c_0)\}^{1/2}(1+o(1))^{1/2}/\hbar$ so that $\hbar^2(1-a_2)^2/(1+o(1)) = 2 + (2-q)/(2c_0)$, and c_0 is the constant in assumption $p \ge n^{c_0}$. Finally, we can show

$$\begin{aligned} &\varpi \sum_{j=2}^{p} [P(|N_j| \ge \varpi)]^{1/2} 1(|\beta_j| \le a_1\varpi) \\ &\quad \le \varpi \sum_{j=2}^{p} [P(N_j - \beta_j \le -\varpi - \beta_j) \\ &\qquad + P(N_j - \beta_j \ge \varpi - \beta_j)]^{1/2} 1(|\beta_j| \le a_1\varpi) \\ &\quad \le \varpi \sum_{j=2}^{p} [P(N_j - \beta_j \le -|1 - a_1|\varpi) + P(N_j - \beta_j \ge |1 - a_1|\varpi)]^{1/2} \\ &\quad \le 2\varpi p^{1-\hbar^2(1-a_1)^2/(2(1+o(1)))} = 2\varpi p^{-1} = o(\pi_n(d)\varpi^{1-q}), \end{aligned} \tag{5.21}$$

where the third inequality is from Lemma 2, and the first equality is due to the fact that we take $a_1 = 1 - 2(1 + o(1))^{1/2}/\hbar$ so that $\hbar^2(1 - a_1)^2 = 4$. Plugging (5.17)–(5.21) into (5.15) and (5.16), we prove the lemma. □

PROOF OF THEOREM 1. Combining Lemma 4 and (5.5)–(5.6) in Lemma 3, we easily obtain

$$E[\|\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}\|_2] \le C_1 \frac{\pi_n(d)}{p^{1/2}} \left(\frac{\log p}{n}\right)^{(1-q)/2},$$

$$E[\|\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}\|_F^2] \le C_0 \pi_n(d) \frac{1}{d} \left(\frac{\log p}{n}\right)^{1-q/2}.$$

Using Lemma 4 and (5.7) in Lemma 3, we conclude

$$\begin{aligned} E[\|\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}\|_2^2] &\le C_2 \left[\pi_n^2(d) \frac{1}{p} \left(\frac{\log p}{n}\right)^{1-q} + \pi_n(d) \frac{1}{p} \left(\frac{\log p}{n}\right)^{1-q/2}\right] \\ &\le C \frac{\pi_n^2(d)}{d^2} \left(\frac{\log p}{n}\right)^{1-q}, \end{aligned} \tag{5.22}$$

where the last inequality is due to the fact that the first term on the right-hand side of (5.22) dominates its second term. □

PROOF OF PROPOSITION 3. Applying the Lyapunov's moment inequality to the Schatten s-norm, we have for $s \in [1, 2]$ and $\boldsymbol{\rho} \in \Theta$,

$$\begin{aligned} E[\|\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}\|_{*s}^2] &\le d^{-1+2/s} E[\|\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}\|_{*2}^2] \\ &= d^{-1+2/s} E[\|\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}\|_F^2] \\ &\le c_1 \pi_n(d) d^{-2+2/s} \left(\frac{\log d}{n}\right)^{1-q/2}, \end{aligned}$$

where the last inequality is due to Theorem 1. On the other hand, applying Hölder's inequality by interpolating between Schatten s-norms with $s = 2$ and $s = \infty$, we obtain for $s \in [2, \infty]$ and $\boldsymbol{\rho} \in \Theta$,

$$\begin{aligned} E[\|\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}\|_{*s}^2] &\le E[\|\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}\|_{*2}^{4/s} \|\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}\|_{*\infty}^{2-4/s}] \\ &\le [E\|\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}\|_{*2}^2]^{2/s} [E\|\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}\|_{*\infty}^2]^{1-2/s} \\ &\le c_7 \pi_n^{2-2/s}(d) d^{-2+2/s} \left(\frac{\log d}{n}\right)^{1-q+q/s}, \end{aligned}$$

where the last inequality is due to Theorem 1, and $c_7 = c_1^{(s-2)/s} c_2^{2/s}$. The result follows by combining the above two inequalities together. □

5.3. *Proofs of Theorems 2 and 3: Lower bound.*

PROOF OF THEOREM 2 (for the lower bound under the spectral norm). We first define a subset of the parameter space Θ. It will be shown later that the risk upper bound under the spectral norm is sharp up to a constant factor, when the parameter space is sufficiently sparse. Consider a subset of the Pauli basis, $\{\boldsymbol{\sigma}_{l_1} \otimes \boldsymbol{\sigma}_{l_2} \otimes \cdots \otimes \boldsymbol{\sigma}_{l_b}\}$, where $\boldsymbol{\sigma}_{l_1} = \boldsymbol{\sigma}_0$ or $\boldsymbol{\sigma}_3$. Its cardinality is $d = 2^b = p^{1/2}$. Denote each element of the subset by $\mathbf{B}_j$, $j = 1, 2, \ldots, d$, and let $\mathbf{B}_1 = \mathbf{I}_d$. We will define each element of Θ as a linear combination of $\mathbf{B}_j$. Let $\gamma_j \in \{0, 1\}$, $j \in \{1, 2, \ldots, d\}$, and denote $\eta = \sum_j \gamma_j = \|\boldsymbol{\gamma}\|_0$. The value of η is either 0 or K, where K is the largest integer less than or equal to $\pi_n(d)/(\frac{\log p}{n})^{q/2}$. By assumption (3.3), we have

$$1 \le K = O(d^v) \qquad \text{with } v < 1/2. \tag{5.23}$$

Let $\varepsilon^2 = (1 - 2v)/4$ and set $a = \varepsilon\sqrt{\frac{\log p}{n}}$. Now we are ready to define Θ,

$$\Theta = \left\{\boldsymbol{\rho}(\boldsymbol{\gamma}) : \boldsymbol{\rho}(\boldsymbol{\gamma}) = \frac{\mathbf{I}_d}{d} + a \sum_{j=2}^{d} \gamma_j \frac{\mathbf{B}_j}{d}, \text{ and } \eta = 0 \text{ or } K\right\}. \tag{5.24}$$

Note that Θ is a subset of the parameter space, since

$$\sum_{j=2}^{d} (a\gamma_j)^q \le K a^q \le \varepsilon^q \pi_n(d) \le \pi_n(d),$$

and its cardinality is $1 + \binom{d-1}{K}$.

We need to show that

$$\inf_{\hat{\boldsymbol{\rho}}} \sup_{\Theta} E\|\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}\|_2^2 \gtrsim \pi_n^2(d) \frac{1}{p} \left(\frac{\log p}{n}\right)^{1-q}.$$

Note that for each element in Θ, its first entry $\boldsymbol{\rho}_{11}$ may take the form $1/d + a\sum_{j=2}^{d} \gamma_j/d = 1/d + (a/d)\eta$. It can be shown that

$$\inf_{\hat{\boldsymbol{\rho}}} \sup_{\Theta} E\|\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}\|_2^2 \ge \inf_{\hat{\rho}_{11}} \sup_{\Theta} E(\hat{\rho}_{11} - \rho_{11})^2 \ge \frac{a^2}{d^2} \inf_{\hat{\eta}} \sup_{\Theta} E(\hat{\eta} - \eta)^2.$$

It is then enough to show that

$$\inf_{\hat{\eta}} \sup_{\Theta} E(\hat{\eta} - \eta)^2 \gtrsim K^2, \tag{5.25}$$

which immediately implies

$$\inf_{\hat{\boldsymbol{\rho}}} \sup_{\Theta} E\|\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}\|_2^2 \gtrsim K^2 \frac{a^2}{d^2} \gtrsim \pi_n^2(d) \frac{1}{p} \left(\frac{\log p}{n}\right)^{1-q}.$$

We prove equation (5.25) by applying Le Cam's lemma. From observations N_j, $j = 2, \dots, d$, we define $\tilde{N}_j = n(N_j + 1)/2$, which is Binomial$(n, \frac{1+a\gamma_j}{2})$. Let $\mathbb{P}_{\boldsymbol{\gamma}}$ be the joint distribution of independent random variables $\tilde{N}_2, \tilde{N}_3, \dots, \tilde{N}_d$. The cardinality of $\{\mathbb{P}_{\boldsymbol{\gamma}}\}$ is $1 + \binom{d-1}{K}$. For two probability measures $\mathbb{P}$ and $\mathbb{Q}$ with density f and g with respect to any common dominating measure μ, write the total variation affinity $\|\mathbb{P} \wedge \mathbb{Q}\| = \int f \wedge g \, d\mu$, and the Chi-square distance $\chi^2(\mathbb{P}, \mathbb{Q}) = \int \frac{g^2}{f} - 1$. Define

$$\bar{\mathbb{P}} = \binom{d-1}{K}^{-1} \sum_{\|\boldsymbol{\gamma}\|_0 = K} \mathbb{P}_{\boldsymbol{\gamma}}.$$

The following lemma is a direct consequence of Le Cam's lemma [cf. Le Cam (1973) and Yu (1997)]. □

LEMMA 5. *Let $\hat{\eta}$ be any estimator of η based on an observation from a distribution in the collection $\{\mathbb{P}_{\boldsymbol{\gamma}}\}$, then*

$$\inf_{\hat{k}} \sup_{\Theta} E(\hat{\eta} - \eta)^2 \geq \frac{1}{4} \|\mathbb{P}_{\mathbf{0}} \wedge \bar{\mathbb{P}}\|^2 \cdot K^2.$$

We will show that there is a constant $c > 0$ such that

$$\|\mathbb{P}_{\mathbf{0}} \wedge \bar{\mathbb{P}}\| \geq C, \tag{5.26}$$

which, together with Lemma 5, immediately imply equation (5.25).

LEMMA 6. *Under conditions* (5.23) *and* (5.24), *we have*

$$\inf_{\hat{\boldsymbol{\rho}}} \sup_{\Theta} E(\hat{\eta} - \eta)^2 \gtrsim K^2,$$

which implies

$$\inf_{\hat{\boldsymbol{\rho}}} \sup_{\Theta} E\|\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}\|_2^2 \gtrsim \pi_n^2(d) \frac{1}{p} \left(\frac{\log p}{n}\right)^{1-q}.$$

PROOF. It is enough to show that

$$\chi^2(\mathbb{P}_{\mathbf{0}}, \bar{\mathbb{P}}) \to 0,$$

which implies $\|\mathbb{P}_{\mathbf{0}} - \bar{\mathbb{P}}\|_{\mathrm{TV}} \to 0$, then we have $\|\mathbb{P}_{\mathbf{0}} \wedge \bar{\mathbb{P}}\| \to 1$. Let $J(\boldsymbol{\gamma}, \boldsymbol{\gamma}')$ denote the number of overlapping nonzero coordinates between $\boldsymbol{\gamma}$ and $\boldsymbol{\gamma}'$. Note that

$$\begin{aligned} \chi^2(\mathbb{P}_{\mathbf{0}}, \bar{\mathbb{P}}) &= \int \frac{(d\bar{\mathbb{P}})^2}{d\mathbb{P}_{\mathbf{0}}} - 1 \\ &= \binom{d-1}{K}^{-2} \sum_{0 \leq j \leq K} \sum_{J(\boldsymbol{\gamma}, \boldsymbol{\gamma}') = j} \left(\int \frac{d\mathbb{P}_{\boldsymbol{\gamma}} \cdot d\mathbb{P}_{\boldsymbol{\gamma}'}}{d\mathbb{P}_{\mathbf{0}}} - 1 \right). \end{aligned}$$

When $J(\boldsymbol{\gamma},\boldsymbol{\gamma}')=j$, we have

$$\begin{aligned}\int\frac{d\mathbb{P}_{\boldsymbol{\gamma}}\cdot d\mathbb{P}_{\boldsymbol{\gamma}'}}{d\mathbb{P}_{\mathbf{0}}}&=\Biggl(\sum_{l=0}^{n}\biggl[\binom{n}{l}\frac{1}{2^l}\frac{1}{2^{n-l}}\cdot(1+a)^{2l}(1-a)^{2n-2l}\biggr]\Biggr)^j\\&=\Biggl(\sum_{l=0}^{n}\biggl[\binom{n}{l}\biggl(\frac{(1+a)^2}{2}\biggr)^l\biggl(\frac{(1-a)^2}{2}\biggr)^{n-l}\biggr]\Biggr)^j\\&=\biggl(\frac{(1+a)^2}{2}+\frac{(1-a)^2}{2}\biggr)^{nj}\\&=(1+a^2)^{nj}\le\exp(na^2j),\end{aligned}$$

which implies

$$\begin{aligned}\chi^2(\mathbb{P}_{\mathbf{0}},\bar{\mathbb{P}})&\le\binom{d-1}{K}^{-2}\sum_{0\le j\le K}\sum_{J(\boldsymbol{\gamma},\boldsymbol{\gamma}')=j}\bigl(\exp(na^2j)-1\bigr)\\&\le\binom{d-1}{K}^{-2}\sum_{1\le j\le K}\sum_{J(\boldsymbol{\gamma},\boldsymbol{\gamma}')=j}\exp(na^2j)\\&=\sum_{1\le j\le K}\frac{\binom{K}{j}\binom{d-1-K}{K-j}}{\binom{d-1}{K}}d^{2\varepsilon^2j}.\end{aligned}$$

Since

$$\begin{aligned}\frac{\binom{K}{j}\binom{d-1-K}{K-j}}{\binom{d-1}{K}}&=\frac{[K\cdot\ldots\cdot(K-j+1)]^2\cdot(d-1-K)\cdot\ldots\cdot(d-2K+j)}{j!\cdot(d-1)\cdot\ldots\cdot(d-K)}\\&\le\frac{K^{2j}(d-1-K)^{K-j}}{(d-K)^K}\le\biggl(\frac{K^2}{d-K}\biggr)^j,\end{aligned}$$

and $\varepsilon^2=(1-2v)/4$, we then have

$$\chi^2(\mathbb{P}_{\mathbf{0}},\bar{\mathbb{P}})\le\sum_{1\le j\le K}\biggl[\frac{K^2}{d-K}d^{2\varepsilon^2}\biggr]^j\le\sum_{1\le j\le K}\biggl[\frac{d^{2v+(1-2v)/2}}{d-K}\biggr]^j\to0.\qquad\square$$

PROOF OF THEOREM 3 (for the lower bound under the Frobenius norm). Recall that Θ is the collection of density matrices such that

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

where

$$\sum_{j=2}^{p}|\beta_j|^q\le\pi_n(p).$$

Apply Assouad's lemma, and we show below that

$$\inf_{\check{\boldsymbol{\rho}}} \sup_{\boldsymbol{\rho}\in\Theta} E\big[\|\check{\boldsymbol{\rho}} - \boldsymbol{\rho}\|_F^2\big] \geq C\pi_n(p)\frac{1}{d}\left(\frac{\log p}{n}\right)^{1-q/2},$$

where $\check{\boldsymbol{\rho}}$ denotes any estimator of $\boldsymbol{\rho}$ based on measurement data $N_2, \dots, N_p$, and C is a constant free of n and p.

To this end, it suffices to construct a collection of $M+1$ density matrices $\{\boldsymbol{\rho}_0 = \mathbf{I}_d/d, \boldsymbol{\rho}_1, \dots, \boldsymbol{\rho}_M\} \subset \Theta$ such that (i) for any distinct k and k_0,

$$\|\boldsymbol{\rho}_k - \boldsymbol{\rho}_{k_0}\|_F^2 \geq C_1\pi_n(p)\frac{1}{d}\left(\frac{\log p}{n}\right)^{1-q/2},$$

where C_1 is a constant; (ii) there exists a constant $0 < C_2 < 1/8$ such that

$$\frac{1}{M}\sum_{k=1}^{M} D_{\mathrm{KL}}(P_{\boldsymbol{\rho}_k}, P_{\boldsymbol{\rho}_0}) \leq C_2 \log M,$$

where D_{KL} denotes the Kullback–Leibler divergence.

By the Gilbert–Varshamov bound [cf. Nielsen and Chuang (2000)], we have that for any $h < p/8$, there exist M binary vectors $\boldsymbol{\gamma}_k = (\gamma_{k2}, \dots, \gamma_{kp})' \in \{0,1\}^{p-1}$, $k = 1, \dots, M$, such that (i) $\|\boldsymbol{\gamma}_k\|_1 = \sum_{j=2}^p |\gamma_{kj}| = h$, (ii) $\|\boldsymbol{\gamma}_k - \boldsymbol{\gamma}_{k_0}\|_1 = \sum_{j=2}^p |\gamma_{kj} - \gamma_{k_0 j}| \geq h/2$, and (iii) $\log M > 0.233 h \log(p/h)$. Let

$$\boldsymbol{\rho}_k = \frac{1}{d}\left(\mathbf{I}_d + \epsilon \sum_{j=2}^{p} \gamma_{kj}\mathbf{B}_j\right),$$

where

$$\epsilon = C_3\left(\frac{\pi_n(p)}{h}\right)^{1/q}.$$

Since $\sum_{j=2}^p |\epsilon\gamma_{kj}|^q = \epsilon^q h = C_3\pi_n(p)$, $\boldsymbol{\rho}_k \in \Theta$ whenever $C_3 \leq 1$. Moreover,

$$d\|\boldsymbol{\rho}_k - \boldsymbol{\rho}_{k_0}\|_F^2 = \epsilon^2\|\boldsymbol{\gamma}_k - \boldsymbol{\gamma}_{k_0}\|_1 \geq \frac{\epsilon^2 h}{4}.$$

On the other hand,

$$\begin{aligned} D_{\mathrm{KL}}(P_{\boldsymbol{\rho}_k}, P_{\boldsymbol{\rho}_0}) &= h D_{\mathrm{KL}}\left(\mathrm{Bin}\left(n, \frac{1+\epsilon}{2}\right), \mathrm{Bin}\left(n, \frac{1}{2}\right)\right) \\ &= hn\frac{\epsilon}{2}\log\frac{1/2+\epsilon}{1/2-\epsilon} \leq C_4 hn\epsilon^2. \end{aligned}$$

Now the lower bound can be established by taking

$$h = \pi_n(p)\left(\frac{\log p}{n}\right)^{-q/2},$$

and then

$$\epsilon = C_3\left(\frac{\log p}{n}\right)^{1/2}, \qquad \frac{\epsilon^2 h}{4} = C_3\pi_n(p)\left(\frac{\log p}{n}\right)^{1-q/2},$$

$$C_4 h n\epsilon^2 = C_4 h \log p, \qquad h\log(p/h) = h\log p - h\log h,$$

$$\log h \sim \log \pi_n(p) + \frac{q}{2}\log n - \frac{q}{2}\log\log p,$$

which are allowed by the assumption $\log\pi_n(p) + \frac{q}{2}\log n < v'\log p$ for $v' < 1$. $\square$

REFERENCES

ALQUIER, P., BUTUCEA, C., HEBIRI, M. and MEZIANI, K. (2013). Rank penalized estimation of a quantum system. *Phys. Rev. A* **88** 032133.

ARTILES, L. M., GILL, R. D. and GUŢĂ, M. I. (2005). An invitation to quantum tomography. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **67** 109–134. MR2136642

ASPURU-GUZIK, A., DUTOI, A. D., LOVE, P. J. and HEAD-GORDON, M. (2005). Simulated quantum computation of molecular energies. *Science* **309** 1704–1707.

AUBRY, J.-M., BUTUCEA, C. and MEZIANI, K. (2009). State estimation in quantum homodyne tomography with noisy data. *Inverse Probl.* **25** 015003, 22. MR2465335

BENENTI, G., CASATI, G. and STRINI, G. (2004). *Principles of Quantum Computation and Information. Vol. I: Basic Concepts*. World Scientific, River Edge, NJ. MR2053465

BENENTI, G., CASATI, G. and STRINI, G. (2007). *Principles of Quantum Computation and Information. Vol. II: Basic Tools and Special Topics*. World Scientific, Hackensack, NJ. MR2317159

BERRY, D. W., CHILDS, A. M., CLEVE, R., KOTHARI, R. and SOMMA, R. D. (2014). Exponential improvement in precision for simulating sparse Hamiltonians. In *STOC'*14*—Proceedings of the* 2014 *ACM Symposium on Theory of Computing* 283–292. ACM, New York. MR3238954

BOIXO, S., RNNOW, T. F., ISAKOV, S. V., WANG, Z., WECKER, D., LIDAR, D. A., MARTINIS, J. M. and TROYER, M. (2014). Evidence for quantum annealing with more than one hundred qubits. *Nature Physics* **10** 218–224.

BRITTON, J. W., SAWYER, B. C., KEITH, A., WANG, C.-C. J., FREERICKS, J. K., UYS, H., BIERCUK, M. J. and BOLLINGER, J. J. (2012). Engineered 2D Ising interactions on a trapped-ion quantum simulator with hundreds of spins. *Nature* **484** 489–492.

BRUMFIEL, G. (2012). Simulation: Quantum leaps. *Nature* **491** 322–324.

BUNEA, F., SHE, Y. and WEGKAMP, M. H. (2011). Optimal selection of reduced rank estimators of high-dimensional matrices. *Ann. Statist.* **39** 1282–1309. MR2816355

BUNEA, F., SHE, Y. and WEGKAMP, M. H. (2012). Joint variable and rank selection for parsimonious estimation of high-dimensional matrices. *Ann. Statist.* **40** 2359–2388. MR3097606

BUTUCEA, C., GUŢĂ, M. and ARTILES, L. (2007). Minimax and adaptive estimation of the Wigner function in quantum homodyne tomography with noisy data. *Ann. Statist.* **35** 465–494. MR2336856

CAI, T. T. and ZHANG, A. (2015). ROP: Matrix recovery via rank-one projections. *Ann. Statist.* **43** 102–138. MR3285602

CAI, T. T. and ZHOU, H. H. (2012). Optimal rates of convergence for sparse covariance matrix estimation. *Ann. Statist.* **40** 2389–2420. MR3097607

CANDÈS, E. J. and PLAN, Y. (2009). Matrix completion with noise. *Proceedings of the IEEE* **98** 925–936.

CANDÈS, E. J. and PLAN, Y. (2011). Tight oracle inequalities for low-rank matrix recovery from a minimal number of noisy random measurements. *IEEE Trans. Inform. Theory* **57** 2342–2359. MR2809094

CANDÈS, E. J. and RECHT, B. (2009). Exact matrix completion via convex optimization. *Found. Comput. Math.* **9** 717–772. MR2565240

CANDÈS, E. J. and TAO, T. (2010). The power of convex relaxation: Near-optimal matrix completion. *IEEE Trans. Inform. Theory* **56** 2053–2080. MR2723472

DONOHO, D. L. (2006). Compressed sensing. *IEEE Trans. Inform. Theory* **52** 1289–1306. MR2241189

DONOHO, D. L. and JOHNSTONE, I. M. (1994). Minimax risk over l_p-balls for l_q-error. *Probab. Theory Related Fields* **99** 277–303. MR1278886

FLAMMIA, S. T., GROSS, D., LIU, Y. K. and EISERT, J. (2012). Quantum tomography via compressed sensing: Error bounds, sample complexity and efficient estimators. *New J. Phys.* **14** 095022.

GROSS, D. (2011). Recovering low-rank matrices from few coefficients in any basis. *IEEE Trans. Inform. Theory* **57** 1548–1566. MR2815834

GROSS, D., LIU, Y.-K., FLAMMIA, S. T., BECKER, S. and EISERT, J. (2010). Quantum state tomography via compressed sensing. *Phys. Rev. Lett.* **105** 150401.

GUŢĂ, M. and ARTILES, L. (2007). Minimax estimation of the Wigner function in quantum homodyne tomography with ideal detectors. *Math. Methods Statist.* **16** 1–15. MR2319467

HÄFFNER, H., HÄNSEL, W., ROOS, C. F., BENHELM, J., CHEK-AL-KAR, D., CHWALLA, M., KÖRBER, T., RAPOL, U. D., RIEBE, M., SCHMIDT, P. O., BECHER, C., GÜHNE, O., DÜR, W. and BLATT, R. (2005). Scalable multiparticle entanglement of trapped ions. *Nature* **438** 643–646.

HOLEVO, A. S. (1982). *Probabilistic and Statistical Aspects of Quantum Theory*. *North-Holland Series in Statistics and Probability* **1**. North-Holland, Amsterdam. MR0681693

JOHNSON, M. W., AMIN, M. H. S., GILDERT, S., LANTING, T., HAMZE, F., DICKSON, N., HARRIS, R., BERKLEY, A. J., JOHANSSON, J., BUNYK, P., CHAPPLE, E. M., ENDERUD, C., HILTON, J. P., KARIMI, K., LADIZINSKY, E., LADIZINSKY, N., OH, T., PERMINOV, I., RICH1, C., THOM, M. C., TOLKACHEVA, E., TRUNCIK, C. J. S., UCHAIKIN, S., WANG, J., WILSON, B. and ROSE, G. (2011). Quantum annealing with manufactured spins. *Nature* **473** 194–198.

JONES, N. (2013). Computing: The quantum company. *Nature* **498** 286–288.

KESHAVAN, R. H., MONTANARI, A. and OH, S. (2010). Matrix completion from noisy entries. *J. Mach. Learn. Res.* **11** 2057–2078. MR2678022

KLOPP, O. (2011). Rank penalized estimators for high-dimensional matrices. *Electron. J. Stat.* **5** 1161–1183. MR2842903

KLOPP, O. (2012). Noisy low-rank matrix completion with general sampling distribution. Manuscript.

KOLTCHINSKII, V. (2011). Von Neumann entropy penalization and low-rank matrix estimation. *Ann. Statist.* **39** 2936–2973. MR3012397

KOLTCHINSKII, V., LOUNICI, K. and TSYBAKOV, A. B. (2011). Nuclear-norm penalization and optimal rates for noisy low-rank matrix completion. *Ann. Statist.* **39** 2302–2329. MR2906869

LANYON, B. P., WHITFIELD, J. D., GILLETT, G. G., GOGGIN, M. E., ALMEIDA, M. P., KASSAL, I., BIAMONTE, J. D., MOHSENI, M., POWELL, B. J., BARBIERI, M., ASPURU-GUZIK, A. and WHITE, A. G. (2010). Towards quantum chemistry on a quantum computer. *Nat. Chem.* **2** 106–111.

LE CAM, L. (1973). Convergence of estimates under dimensionality restrictions. *Ann. Statist.* **1** 38–53. MR0334381

LIU, Y. K. (2011). Universal low-rank matrix recovery from Pauli measurements. Unpublished manuscript.

NEGAHBAN, S. and WAINWRIGHT, M. J. (2011). Estimation of (near) low-rank matrices with noise and high-dimensional scaling. *Ann. Statist.* **39** 1069–1097. MR2816348

NIELSEN, M. A. and CHUANG, I. L. (2000). *Quantum Computation and Quantum Information*. Cambridge Univ. Press, Cambridge. MR1796805

RECHT, B. (2011). A simpler approach to matrix completion. *J. Mach. Learn. Res.* **12** 3413–3430. MR2877360

RECHT, B., FAZEL, M. and PARRILO, P. A. (2010). Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM Rev.* **52** 471–501. MR2680543

ROHDE, A. and TSYBAKOV, A. B. (2011). Estimation of high-dimensional low-rank matrices. *Ann. Statist.* **39** 887–930. MR2816342

SAKURAI, J. J. and NAPOLITANO, J. (2010). *Modern Quantum Mechanics*, 2nd ed. Addison-Wesley, Reading, MA.

SENKO, C., SMITH, J., RICHERME, P., LEE, A., CAMPBELL, W. C. and MONROE, C. (2014). Coherent imaging spectroscopy of a quantum many-body spin system. *Science* **345** 430–433. MR3241225

SHANKAR, R. (1994). *Principles of Quantum Mechanics*, 2nd ed. Plenum Press, New York. MR1343488

TAO, M., WANG, Y. and ZHOU, H. H. (2013). Optimal sparse volatility matrix estimation for high-dimensional Itô processes with measurement errors. *Ann. Statist.* **41** 1816–1864. MR3127850

WANG, Y. (2011). Quantum Monte Carlo simulation. *Ann. Appl. Stat.* **5** 669–683. MR2840170

WANG, Y. (2012). Quantum computation and quantum information. *Statist. Sci.* **27** 373–394. MR3012432

WANG, Y. (2013). Asymptotic equivalence of quantum state tomography and noisy matrix completion. *Ann. Statist.* **41** 2462–2504. MR3127872

WANG, Y. and XU, C. (2015). Density matrix estimation in quantum homodyne tomography. *Statist. Sinica* **25** 953–973.

YU, B. (1997). Assouad, Fano, and Le Cam. In *Festschrift for Lucien Le Cam* 423–435. Springer, New York. MR1462963

ZHANG, C. (2012). Minimax ℓ_q risk in ℓ_p balls. *Inst. Math. Stat. Collect.* **8** 78–89.

T. CAI
DEPARTMENT OF STATISTICS
THE WHARTON SCHOOL
UNIVERSITY OF PENNSYLVANIA
PHILADELPHIA, PENNSYLVANIA 19104
USA
E-MAIL: tcai@wharton.upenn.edu

D. KIM
Y. WANG
M. YUAN
DEPARTMENT OF STATISTICS
UNIVERSITY OF WISCONSIN-MADISON
MADISON, WISCONSIN 53706
USA
E-MAIL: kimd@stat.wisc.edu
yzwang@stat.wisc.edu
myuan@stat.wisc.edu

H. H. ZHOU
DEPARTMENT OF STATISTICS
YALE UNIVERSITY
NEW HAVEN, CONNECTICUT 06511
USA
E-MAIL: huibin.zhou@yale.edu