

An improved robust ADMM algorithm for quantum state tomography

Kezhi Li^{1,2} · Hui Zhang¹ · Sen Kuang¹ ·
Fangfang Meng³ · Shuang Cong¹

Received: 2 October 2015 / Accepted: 23 February 2016 / Published online: 15 March 2016
© Springer Science+Business Media New York 2016

Abstract In this paper, an improved adaptive weights alternating direction method of multipliers algorithm is developed to implement the optimization scheme for recovering the quantum state in nearly pure states. The proposed approach is superior to many existing methods because it exploits the low-rank property of density matrices, and it can deal with unexpected sparse outliers as well. The numerical experiments are provided to verify our statements by comparing the results to three different optimization algorithms, using both adaptive and fixed weights in the algorithm, in the cases of with and without external noise, respectively. The results indicate that the improved algorithm has better performances in both estimation accuracy and robustness to external noise. The further simulation results show that the successful recovery rate increases when more qubits are estimated, which in fact satisfies the compressive sensing theory and makes the proposed approach more promising.

Keywords Quantum state estimation · Compressive sensing · ADMM algorithm · Adaptive weights

1 Introduction

Quantum state estimation is an important problem in quantum information processing and has many applications [1, 2]. Usually an unknown quantum state is estimated based

✉ Shuang Cong
scong@ustc.edu.cn

¹ Department of Automation, University of Science and Technology of China, Hefei 230027, People's Republic of China

² Imperial College London, London, UK

³ Hefei University, Hefei 230022, People's Republic of China

on the data obtained from the measuring process of a quantum system. This process can be carried out by quantum state tomography [3,4], in which a large number of measurements of the quantum state must be collected in physical experiments. For an n -qubit quantum system, the quantum state can be described by a density matrix of size $d \times d = 2^n \times 2^n = 4^n$, and the dimension of the density matrix d increases exponentially with the number of qubits n in consideration. This makes the quantum state estimation expensive and impractical for identifying the state of merely several qubits by conventional experiments. In recent years, compressive sensing (CS) has attracted lots of attention [5,6] for the reason that it may require less sampling rate than what Shannon–Nyquist sampling Theorem needs to recover the original signal. Aiming for achieving a good estimation performance with fewer measurements, researchers have applied compressive sensing into the quantum state estimation and achieved a significant improvement [7,8].

In practical quantum state tomography, people are often interested in states consisting of n qubits which can be represented as the probabilistic combination of r pure states. Under such circumstances, the density matrix can be a low-rank Hermitian matrix, which satisfies the prior conditions of using the compressive sensing theory for low-rank matrices reconstruction [9,10]. The quantum state estimation based on compressive sensing can exploit these low-dimensional measurements from projections of high-dimensional density matrix and reconstruct the density matrix accurately. In this case, much fewer measurements [e.g., $O(rd \log d)$ instead of $O(d^2)$] are needed to reconstruct all elements of the density matrix comparing with the conventional estimation method. Normally half of the elements at most are needed to estimate the density matrix of pure states based on matrix transformation [11]. While using CS, one is able to recover the values of pure states by using even fewer measurements [12], because the quantum state estimation can be converted to a convex optimization problem with quantum characteristic constraints and be solved by leveraging the rank constraints [13]. There are some common optimization algorithms that have been proposed to estimate the quantum states. Smith summarized the least square problem and solved it in quantum state estimation through MATLAB toolbox [1]. Liu adopted Dantzig algorithm to estimate the density matrix [14]. In recent years, the alternating direction method of multipliers (ADMM) algorithm [15–17] has been widely adopted due to its fast computation and good robustness. This method reconstructs the quantum state by minimizing the nuclear norm of the density matrices with the prior information by leveraging the nearly pure property [8,18–20]. Li applied ADMM algorithm to quantum tomography using CS, established the framework of the optimization problem, and tested the proposed approach in the case of 5 qubits [21]. However, one major problem of the existed ADMM algorithm is that it has the difficulties in choosing the proper weight parameters. It is not trivial applying this approach in practice with fixed weights, and currently these weights are determined mostly relying on the experience [22].

In this paper, we firstly formulate the quantum state tomography problem with outlier measurements in density matrices to mathematical forms and propose an optimization equation to solve it. Then, a novel algorithmic solver is developed to solve this optimization problem by using an improved adaptive ADMM technique, and its theoretical analysis is provided. This analysis is established in the quantum estima-

tion combined with CS for the first time. This method can adjust the weight values according to the error in each iteration automatically and overcome the difficulties of choosing proper parameters. Finally, we compare the results of the reconstructing density matrix based on CS to the results of three other algorithms, including the least square method, Dantzig and conventional/adaptive ADMM algorithms. Simulation results reveal that the improved adaptive ADMM algorithm has superiority in quantum state estimation both with less convergence time and with higher accuracy.

The paper is organized as follows: Sect. 2 describes the idea of quantum state estimation based on compressive sensing, optimization algorithms, and an improved adaptive weighted ADMM algorithm proposed. Experiments in four cases are implemented, and the results are analyzed in Sect. 3. Finally, the conclusion is summarized in Sect. 4.

2 Quantum state estimation based on compressive sensing and optimization algorithms

2.1 Problem description

In a quantum system, the density matrix can be denoted as ρ of size $d \times d$, $\rho = \sum_{i=1}^r p_i |\psi_i\rangle\langle\psi_i|$, where $|\psi_i\rangle$ are different quantum pure states, which consists of the quantum state vector ψ . Denote p_i as the probability of each pure state $|\psi_i\rangle$ in ψ , then the state vector Ψ_r can be rewritten as $\Psi_r = |\psi_i\rangle = [\sqrt{p_1}|\psi_1\rangle, \sqrt{p_2}|\psi_2\rangle, \dots, \sqrt{p_r}|\psi_r\rangle]$, Ψ_r has size $d \times r$, so the rank of the density matrix of a nearly pure quantum system is at most r due to the rank property of multiplication of two matrices, where r is much smaller than d . The dimension of n qubits density matrix is $d = 2^n$ when the Pauli matrices are selected as orthogonal bases.

The observation matrix \mathbf{O}_i^* in quantum state estimation for the i th measurement can be calculated as $\sigma_{l_1} \otimes \sigma_{l_2} \otimes \dots \otimes \sigma_{l_n}$ [14]. Here $\sigma_{l_1}, \sigma_{l_2}, \dots, \sigma_{l_n}$ represent a random selection from a unit matrix \mathbf{I}_2 and Pauli matrices $\sigma_1, \sigma_2, \sigma_3$, specifically

$$\mathbf{I}_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \sigma_1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \sigma_2 = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \sigma_3 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}. \quad (1)$$

Define the system measurement operator $\mathcal{A} : \mathbb{C}^{d \times d} \rightarrow \mathbb{C}^M$; the observation matrix for one random selection of Pauli matrices set is $\mathbf{O}_i^* \in \mathbb{C}^{d \times d}$. Then the corresponding sampling matrix $\mathbf{A} \in \mathbb{C}^{M \times d^2}$ is the normalized operator whose i th row is the concatenation of \mathbf{O}_i^* 's rows. It has been proven that Pauli matrices satisfy the rank-restricted isometry property (RIP); hence, by using Pauli matrices as sampling matrices, we are able to recover the unique density matrix with sufficient measurements via compressive sensing approach [7, 23]. The size of the sampling matrix \mathbf{A} directly decides the compressive rate $\eta = M/d^2$, where M is the number of measurements.

Since in CS the number of samples we need for accurate recovery is much less than the number of elements in the density matrix with $M \ll d^2$, the original number of measurements can be significantly reduced. The process of measurement is randomly

projecting ρ to the sampling matrix \mathcal{A} and finally obtains measurements \mathbf{y}_i . The values \mathbf{y}_i can be represented as:

$$\begin{aligned} \mathbf{y}_i &= (\mathcal{A}(\rho))_i + e_i = c \cdot \text{tr}(\mathbf{O}_i^* \rho) + e_i, \quad i = 1, \dots, m, \quad \text{or} \\ \mathbf{y} &= \mathbf{A} \text{vec}(\rho) + \mathbf{e}, \end{aligned} \quad (2)$$

where $\text{tr}(\cdot)$ is the trace operator and $\text{vec}(\cdot)$ represents the transformation from matrix to vector, $\mathbf{y} \in \mathcal{C}^{M \times 1}$, $\rho \in \mathcal{C}^{d \times d}$, $\text{vec}(\rho) \in \mathcal{C}^{d^2 \times 1}$; c is a normalized parameter that could be $\frac{d}{\sqrt{M}}$, and \mathbf{e} is the measurement error.

Generally, the normalized error is used to examine the reconstruction performance as:

$$\text{error} = \frac{\|\rho^* - \hat{\rho}\|_2^2}{\|\rho^*\|_2^2}, \quad (3)$$

where $\hat{\rho}$ is the estimated quantum density matrix and ρ^* represents the real quantum density matrix which can be generated by the following formula [24]

$$\rho^* = \frac{\Psi_{\mathbf{r}} \Psi_{\mathbf{r}}^*}{\text{tr}(\Psi_{\mathbf{r}} \Psi_{\mathbf{r}}^*)}. \quad (4)$$

The estimation error is the normalization error of the difference between ρ^* and $\hat{\rho}$.

In the actual experiments, the noise introduced will affect the estimation accuracy of system. Normally the noise is supposed to satisfy a distribution, such as Gaussian noise. However, sometimes due to the unexpected external noise or bad samples, the noises may be large but only very few, which can be shown as outliers in the system. These outliers of course do not satisfy the Gaussian distribution yet can be reflected as sparse entries in $\hat{\rho}$. In this case, the average system measured values can be expressed as:

$$\mathbf{y}_i = (\mathcal{A}(\rho + \mathbf{S}))_i + e_i = c \cdot \text{tr}(\mathbf{O}_i^* (\rho + \mathbf{S})) + e_i, \quad i = 1, \dots, m. \quad (5)$$

where $\mathbf{S} \in \mathcal{C}^{d \times d}$ is a sparse external noise matrix, which has the same dimension as density matrix, but the value of \mathbf{S} is much smaller than ρ .

Given the vector \mathbf{y} , one can estimate the density matrix ρ by using (2) or (5). This process is equivalent to solving a problem of d^2 unknown variables from M equations. Because $M \ll d^2$, (2) and (5) may have non-unique solutions. Fortunately, the density matrix ρ can be estimated by solving an optimization problem which minimizes the measurement error within the prior low-rank constraints, so a unique solution of ρ might be achieved. Next, we will give the formulation of the quantum state estimation optimization problem based on compressive sensing.

2.2 Quantum state estimation based on compressive sensing

Quantum state estimation is a method based on statistical information to reconstruct the density matrix of quantum systems through measuring the same unknown quantum states many times. The computation time of quantum state estimation is very heavy due to the large number of unknown elements. Since compressive sensing can

recover sparse or low-rank matrix with less measurements, it is natural to combine compressive sensing and quantum state estimation and expect a shorter computation time. The core idea of compressive sensing is to randomly project the density matrix to obtain a small number of measured values. Since the degree of $\rho \in \mathbb{C}^{d \times d}$ is $\mathcal{O}(d^2)$, it requires $\mathcal{O}(d^2)$ measurements to determine the only system state. When ρ can be factorized as the matrices of size of m by n , the nuclear norm $\|\rho\|_*$ is defined by $\|\rho\|_* = \text{tr}(\sqrt{\rho^* \rho}) = \sum_{i=1}^{\min\{m, n\}} \sigma_i$, which is a convex function that can be optimized effectively. Considering the optimization process given the measurements, the priori information of the low-rank property and other quantum constraints on density matrices:

$$\hat{\rho} = \arg \min_{\rho} \|\rho\|_*, \quad \text{s.t. } \|\mathbf{y} - \text{Avec}(\rho)\|_2^2 \leq \epsilon, \rho^* = \rho, \rho \geq 0, \quad (6)$$

where $\arg \min_{\rho}$ is denoted as the value of variable ρ when $\|\rho\|_*$ has the minimum value, $\rho \geq 0$ means ρ is a positive semidefinite (p.s.d) matrix. When there exists an external large noise added to a few entries in the density matrix of the quantum system, we formulate these outlier noises as a sparse matrix \mathbf{S} , and then, one can compute the estimation of true density matrix ρ^* by minimizing

$$\hat{\rho} = \arg \min_{\rho} (\|\rho\|_* + \|\mathbf{S}\|_1), \quad \text{s.t. } \|\mathbf{y} - \text{Avec}(\rho + \mathbf{S})\|_2^2 \leq \epsilon, \rho^* = \rho, \rho \geq 0, \quad (7)$$

where sparse noise $\mathbf{S} \in \mathbb{C}^{d \times d}$ is added directly on density matrix ρ .

There are two advantages for combining compressive sensing and quantum state estimation: First, it only needs a smaller number of random measurements of the original signal. Second, in terms of the lost entries of the original density matrix or large external noise, the method has good robustness.

2.3 Optimization algorithms in quantum state estimation

From previous subsections, we know that the quantum state estimation can be solved by minimizing (6) or (7). In order to solve the problem, here we introduce three different algorithms, including the least square, Dantzig algorithm and alternating direction method of multipliers algorithm (ADMM). The Least Square algorithm (LS) is a method of finding a vector that is a local minimizer to a sum of square functions subject to certain constraints. One can use the convex optimization toolbox [25] to solve the problem, which is a popular convex optimization solving algorithm toolbox concerned with the optimization of objective functions. The matrix to be minimized is the density matrix ρ which is positive semidefinite. Thus, the LS algorithm can turn the density matrix optimization into solving M objective functions, which is $\sum_i [y_i - c \cdot \text{tr}(\mathbf{O}_i^* \rho)]^2 \leq \epsilon$. Another effective optimization algorithm was proposed by Dantzig. Usually in the LS method when the number of qubit increases, the number of function increases a lot, which leads the difficulty to solve a large number of objective functions. The Dantzig algorithm applies the LS method into the situation of solving the sparse or low-rank matrix. By using this prior information to transform the matrix,

Dantzig algorithm can reduce the number of objective functions and effectively deal with the large computation. One can also implement the Dantzig algorithm to optimize the low-rank density matrix by using the convex optimization toolbox [25].

The ADMM algorithm is known as an efficient approach in distributed convex optimization, particularly in the large-scale problems with a very large number of elements. In order to solve the optimization problems (6) or (7) using ADMM algorithm, we need to change the problem into the Lagrange form:

$$L_{\lambda_1}(\boldsymbol{\rho}, \mathbf{S}, \mathbf{u}) = \|\boldsymbol{\rho}\|_* + \|\mathbf{S}\|_1 + \mathbf{u}^T (\mathbf{Avec}(\boldsymbol{\rho}) + \mathbf{Avec}(\mathbf{S}) - \mathbf{y}) + \frac{\lambda_1}{2} \|\mathbf{Avec}(\boldsymbol{\rho}) + \mathbf{Avec}(\mathbf{S}) - \mathbf{y}\|_2^2, \tag{8}$$

where $\lambda_1 > 0$ is a constant weight value, which affects the convergence rate and the number of iterations. We then combine the linear with the quadratic terms in (8) and get:

$$L_{\lambda_1}(\boldsymbol{\rho}, \mathbf{S}, \mathbf{u}) = \|\boldsymbol{\rho}\|_* + \|\mathbf{S}\|_1 + \frac{\lambda_1}{2} \|\mathbf{Avec}(\boldsymbol{\rho}) + \mathbf{Avec}(\mathbf{S}) - \mathbf{y} + \mathbf{u}\|_2^2, \tag{9}$$

with $\mathbf{u} = (1/\lambda_1)\mathbf{u}'$.

The iteration of ADMM algorithm consists of the following three steps:

1. $\boldsymbol{\rho}$ minimization: $\boldsymbol{\rho}^{k+1} = \arg \min_{\boldsymbol{\rho}} L_{\lambda_1}(\boldsymbol{\rho}, \mathbf{S}^k), \mathbf{u}^k$;
2. \mathbf{S} minimization: $\mathbf{S}^{k+1} = \arg \min_{\mathbf{S}} L_{\lambda_1}(\boldsymbol{\rho}^{k+1}, \mathbf{S}^k, \mathbf{u}^k)$;
3. \mathbf{u} update: $\mathbf{u}^{k+1} = \mathbf{u}^k + \lambda_1 (\mathbf{Avec}(\boldsymbol{\rho}^{k+1}) + \mathbf{Avec}(\mathbf{S}^{k+1}) - \mathbf{y})$.

When the error is less than the prefixed value ϵ , the iterations stop.

2.4 An improved adaptive weighted ADMM algorithm

The value λ_1 in (9) plays a learning rate role. In the above conventional method, λ_1 is fixed in the experiments. Both too large or too small value of fixed λ_1 are not good in the optimization procedure. Usually in order to get a good result, one can only select a smaller value of λ_1 which leads to more iterations. Here we propose an improved method of adaptive weight value λ_1 , which can effectively self-tune the value λ_1 according to the requirements. The improved adaptive weighted ADMM algorithm proposed can be formulated as:

$$\lambda_1^{k+1} = \begin{cases} 1.05\lambda_1^k & \text{error}^k < \text{error}^{k-1} \\ 0.7\lambda_1^k & \text{error}^k > \text{error}^{k-1} \\ \lambda_1^k & \text{others} \end{cases}, \tag{10}$$

where error^k means the current estimation error, error^{k-1} means the last iteration estimation error, λ_1^{k+1} and λ_1^k represent the next and current iteration weight value λ_1 .

The advantages of the adaptive ADMM algorithm proposed are as follows: If the estimation error is $\text{error}^k < \text{error}^{k-1}$, which means the estimation error is decreasing,

so the weight value changes as $1.05\lambda_1^k$ to enlarge the amplitude of error decreased. If $\text{error}^k > \text{error}^{k-1}$, it means the weight value is too large to cause the overshoot control, and then, the weight value reduces to $0.7\lambda^k$. By adjusting the value itself, the estimation error can maintain larger speed to decrease, which uses less time and the number of iteration to satisfy the accuracy.

The iteration of the adaptive ADMM algorithm consists of the following steps:

1. ρ minimization: $\rho^{k+1} = \arg \min_{\rho} L_{\lambda_1^k}(\rho^k, \mathbf{S}^k), \mathbf{u}^k$;
2. \mathbf{S} minimization: $\mathbf{S}^{k+1} = \arg \min_{\mathbf{S}} L_{\lambda_1^k}(\rho^{k+1}, \mathbf{S}^k, \mathbf{u}^k)$;
3. λ_1 update: if $\text{error}^k < \text{error}^{k-1}$, $\lambda_1^{k+1} = 1.05\lambda_1^k$, else if $\text{error}^k > \text{error}^{k-1}$, $\lambda_1^{k+1} = 0.7\lambda_1^k$, else $\lambda_1^{k+1} = \lambda_1^k$;
4. \mathbf{u} update: $\mathbf{u}^{k+1} = \mathbf{u}^k + \lambda_1 (\text{Avec}(\rho^{k+1}) + \text{Avec}(\mathbf{S}^{k+1})) - \mathbf{y}$.

Step 1 minimizes the low-rank matrix ρ^{k+1} matrix with fixed \mathbf{S}^k and \mathbf{u}^k :

$$\rho^{k+1} = \arg \min_{\rho} \left\{ \|\rho\|_* + \frac{\lambda_1}{2} \|\text{Avec}(\rho) + \text{Avec}(\mathbf{S}^k) - \mathbf{y} + \mathbf{u}^k\|_2^2 \right\}, \tag{11}$$

where k is the number of iterations.

First, we minimize the unconstrained quadratic function in terms of ρ . The analytic solution to the least square estimation can be written as

$$\rho_1^{k+1} = \text{mat} \left((\mathbf{A}^* \mathbf{A})^{-1} \mathbf{A}^* (\mathbf{y} - \mathbf{u}^k - \text{Avec}(\mathbf{S})) \right). \tag{12}$$

Second, projecting ρ_1^{k+1} on to the constraints set \mathcal{C} at the same time with low rank, and denoting the result as ρ_2^{k+1} , i.e.,

$$\rho_2^{k+1} = \Pi_{\mathcal{C}}(\rho_1^{k+1}), \tag{13}$$

where $\Pi_{\mathcal{C}}$ denotes the Euclidean projection onto \mathcal{C} and at the same time with low rank. For the particular constraint set of quantum state, \mathcal{C} is a proper cone of the Hermitian p.s.d. matrices. We will show the projection process in Sect. 2.5 with efficient approach.

Step 2 updates the sparse matrix \mathbf{S} with fixed $\rho^{k+1} = \rho_2^{k+1}$ and \mathbf{u}^k .

$$\mathbf{S}^{k+1} := \arg \min_{\mathbf{S}} \left\{ \|\mathbf{S}\|_1 + \frac{\lambda_1}{2} \|\text{Avec}(\rho^{k+1}) + \text{Avec}(\mathbf{S}) - \mathbf{y} + \mathbf{u}^k\|_2^2 \right\}. \tag{14}$$

It is a conventional LASSO problem and can be solved by iterations. Here we follow the approaches in [21] and adopt a shrink operator to calculate a solution efficiently.

In detail, the least square estimate \mathbf{S} can be approximated by

$$\mathbf{S}_1^{k+1} = \text{mat} \left(\left((\mathbf{A}^* \mathbf{A})^{-1} \mathbf{A}^* \left(\mathbf{y} - \mathbf{u}^k - \mathbf{A} \text{vec}(\boldsymbol{\rho}^{k+1}) \right) \right) \right) \tag{15}$$

Then we apply a shrink operator to \mathbf{S} in order to achieve a sparse solution

$$\mathbf{S}_2^{k+1} = \mathcal{S}_{\tau'}(\mathbf{s}) = \text{sgn}[\mathbf{s}] \max(|\mathbf{s}| - \tau' \mathbf{1}, \mathbf{0}) \tag{16}$$

where \mathcal{S} is the shrink operator which is also adopted in the projection sub-step in step 1, $\mathbf{s} = \text{vec}(\mathbf{S}_1^{k+1})$, τ' is a shrink parameter that depends on the sparsity level of \mathbf{S} .

Step 3 is to calculate the estimation error and update the weight value λ_1^{k+1}

$$\text{error} = \|\mathbf{y} - \mathbf{A} \text{vec}(\boldsymbol{\rho}^k + \mathbf{S}^k)\|_2^2. \tag{17}$$

In this step, the value of λ depends on the current and last time estimation error as shown in (10).

Step 4 is the dual-update step:

$$\mathbf{u}^{k+1} = \mathbf{u}^k + \lambda^{k+1} (\mathbf{y} - \mathbf{A} \text{vec}(\boldsymbol{\rho}^{k+1}) - \mathbf{A} \text{vec}(\mathbf{S}^{k+1})). \tag{18}$$

The algorithm follows the steps 1–4 to update each component. Relatively small numbers of iterations, like 30–40, are sufficient to achieve a good accuracy in practice. There are several stopping criterions, e.g., adopting bounds, we have

$$\begin{aligned} \|\mathbf{y} - \mathbf{A} \text{vec}(\boldsymbol{\rho}^k + \mathbf{S}^k)\|_2^2 &\leq \varepsilon_1 \|\mathbf{y}\|_2, \\ \|\boldsymbol{\rho}^k - \boldsymbol{\rho}^{k-1}\|_2 &\leq \varepsilon_2, \quad \|\mathbf{S}^k - \mathbf{S}^{k-1}\|_2 \leq \varepsilon_3. \end{aligned} \tag{19}$$

where $\varepsilon_1, \varepsilon_2, \varepsilon_3$ are parameters need to be tuned. Some methods of tuning parameters of alternating direction methods are indicated in [20, 26].

2.5 Projection onto the constraint set with low rank

We leverage the *positive eigenvalue thresholding operator* \mathcal{D}_τ to calculate $\boldsymbol{\rho}_2^{k+1}$. Let $\mathcal{S}_\tau : \mathcal{R}^d \rightarrow \mathcal{R}^d$ denote the shrink operator such that

$$\mathcal{S}_\tau(\mathbf{x}) = \text{sgn}[\mathbf{x}] \max(|\mathbf{x}| - \tau \mathbf{1}, \mathbf{0}) \tag{20}$$

where $\mathbf{1}$ represents a vector with all elements 1. This definition can also be extended to the matrix form. Then the positive eigenvalue thresholding operator \mathcal{D}_τ is defined as

$$\boldsymbol{\rho}_2^{k+1} = \mathcal{D}_\tau(\boldsymbol{\rho}_1^{k+1}) = \mathbf{V} \mathcal{S}_\tau(\boldsymbol{\Sigma}^+) \mathbf{V}^* \tag{21}$$

$\boldsymbol{\Sigma}^+$ only keeps the positive part of the eigenvalues where $\boldsymbol{\Sigma}^+ = \max(\boldsymbol{\Sigma}, \mathbf{0})$, $\mathcal{S}_\tau(\boldsymbol{\Sigma}^+)$ is a shrink operator on the diagonal matrix $\boldsymbol{\Sigma}^+$ which has eigenvalues as entries, $\tau =$

$1/\lambda_1$, where Σ, \mathbf{V} are obtained from the eigenvalue decomposition of a symmetrized matrix $1/2(\rho_1^{k+1} + \rho_1^{k+1*})$,

$$\mathbf{V}\Sigma\mathbf{V}^* = 1/2(\rho_1^{k+1} + \rho_1^{k+1*}), \tag{22}$$

This approach can be derived from its Karush–Kuhn–Tucker (KKT) conditions of the optimal projection from ρ_2^{k+1} to set \mathcal{C} with least square errors. Taking the indicator function $I_{\mathcal{C}}(\rho)$ for instance, under mild assumptions on a proper cone \mathcal{C} , the KKT conditions of

$$\begin{aligned} &\text{minimize } \|\bar{\rho} - \rho\|_2^2 \\ &\text{s.t. } \bar{\rho} \in I_{\mathcal{C}} \end{aligned} \tag{23}$$

are given by

$$\begin{aligned} \bar{\rho} &\in I_{\mathcal{C}}, & \bar{\rho} - \rho &= \theta, \\ \theta &\in I_{\mathcal{C}}, & \theta^* \bar{\rho} &= 0. \end{aligned} \tag{24}$$

The third term in (24) is because the positive semidefinite cone is self-dual. Then the Euclidean projection can be derived by decomposing ρ to the subtraction of two orthogonal elements: one with nonnegative eigenvalues and the other with negative part. Thus a solution satisfying low-rank constraints will be obtained after the shrink operator. In addition, if given the prior information, then the objective quantum state is the probabilistic linear combination of less than or equal to r pure states, and we may project ρ to the set of r -rank matrices by selecting the maximum r positive eigenvalues in Σ^+ in (21). For the details of the derivation, the readers may refer to [25].

2.6 Remark

1. The proposed robust algorithm is suitable for reconstructing near-pure quantum state. Here, “near pure” we mean that the density matrix ρ falls in one of the following 3 situations:

Situation 1 the quantum state consisting of n qubits is the probabilistic combination of r pure states, which means:

$$\hat{\rho} = \sum_{i=1}^r p_i |\psi_i\rangle\langle\psi_i|. \tag{25}$$

In this case, one can prove that its density matrix ρ with size $d \times d$ has rank not larger than r , $d = 2^n$ [21].

Situation 2 the quantum state consisting of n qubits is the probabilistic combination of r pure states plus random noises, which means

$$\left\| \hat{\rho} - \sum_{i=1}^r p_i |\psi_i\rangle\langle\psi_i| \right\|_F \leq \epsilon_e. \tag{26}$$

In this case, the density matrix ρ can even be not low rank, but we can still recover the near-pure state roughly within an error bound as long as there exists very few large eigenvalues and the non-low-rank noises are very small [7, 18]. This error bound depends on the magnitudes of ϵ_e . Similar circumstance happens in compressive sensing that the number of nonzero values of the original signal can be not sparse (e.g., exponential decaying); however, CS can still recover the original signal within an error bound [27, 28].

Situation 3 the density matrix $\hat{\rho}$ is the probabilistic combination of r pure states plus impulse outliers.

$$\hat{\rho} = \sum_{i=1}^r p_i |\psi_i\rangle\langle\psi_i| + \mathbf{S}. \tag{27}$$

The proposed robust algorithm explained is for Situation 3 plus measurement errors in \mathbf{y} . The algorithm for Situation 1 and 2 can be simplified accordingly.

2. Regarding the convergence of ADMM and error bounds of recovering low-rank matrix from its measurements, the readers may refer to [15, 17, 19, 25]. If there is no analytical solution to ADMM steps, we may also use the semidefinite programs. Details and the software can be found in [29].

3. In practice, the observable \mathbf{O}_i is not necessarily the tensor product of Pauli matrices. For instance, in [1] the author developed a device to proceed the quantum state tomography by continuous measurements where \mathbf{O}_i is affected by outer radio frequency magnetic fields. In this case, we can still use the proposed algorithm to recover the quantum state, as long as that \mathbf{O}_i satisfies the rank RIP and number of measurements are sufficient large. Regarding the details of rank RIP and the measurement number of the compressive quantum tomography approach, please refer to the ‘‘Appendix.’’

4. If the dataset is large, our algorithm equipped with ADMM technique can be extended to a distributed manner as a consensus optimization problem. Assume N agents can communicate with each other, and denote each cost function $f_i(\cdot)$, $i = 1, 2, \dots$ as in (2), in this case the steps in iterations turns to

$$\begin{aligned} \mathbf{x}_i^{k+1} &= \arg \min_{\mathbf{x}_i} \left(f_i(\mathbf{x}_i) + \mathbf{y}_i^{kT} (\mathbf{x}_i - \bar{\mathbf{x}}_i^k) + \frac{\lambda}{2} \|\mathbf{x}_i - \bar{\mathbf{x}}_i^k\|_2^2 \right), \\ \mathbf{y}_i^{k+1} &= \mathbf{y}_i^k + \lambda(\mathbf{x}_i^{k+1} - \bar{\mathbf{x}}_i^{k+1}), \end{aligned} \tag{28}$$

where $\bar{\mathbf{x}}_i^k = 1/n_i \sum_{i=1}^{n_i} \mathbf{x}_i^k$ represents the average of n neighbors of agent i . Generally speaking, we gather \mathbf{x}_i^k from outside and scatter $\bar{\mathbf{x}}^k$ to processors and then update $\mathbf{x}_i, \mathbf{y}_i$ in each processor locally in parallel. Finally, each agent can achieve a consensus about the quantum state. See details of consensus optimization via ADMM in [15].

3 Experiment results and analyses

In this section, we verify the superiority and robustness of the proposed adaptive ADMM algorithm in all three algorithms based on compressive sensing in quantum state estimation, which is analyzed from four aspects: 1. Better performance: the

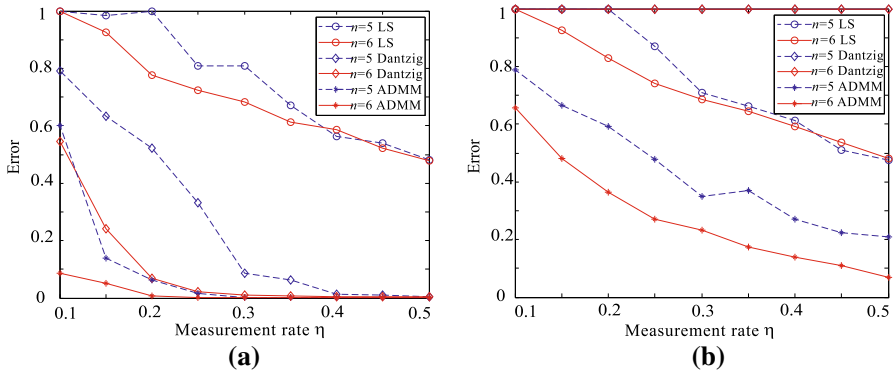


Fig. 1 Effects of measurement rate to the error of quantum state estimation among three algorithms without and with noises in the cases of qubit $n = 5$ and $n = 6$

estimation error comparisons among three different algorithms which are the least square, Dantzig and ADMM algorithms, respectively; 2. Robustness: performance comparisons of ADMM algorithm without and with external noise in the different measurement rate; 3. Superiority: performance comparisons of ADMM algorithm under different quantum qubits; and 4. Improvement: performance comparisons between the fixed and adaptive weight values of ADMM algorithm.

3.1 Performance comparisons among three different algorithms

First, we investigate the effects of measurement rate η to the error of quantum state estimation among three different algorithms without and with noise in the cases of qubits $n = 5$ and $n = 6$. In the experiments, algorithms used are the least square method, Dantzig algorithm and ADMM algorithm; the measurement rate ranges from $\eta = 0.1$ to $\eta = 0.5$ with 0.05 difference between each data point. The effects to estimation errors of three algorithms in the cases of qubit $n = 5$ and $n = 6$ without external noise are shown in Fig. 1a, from which one can see that the estimation errors of three different algorithms all gradually decrease in different measurement rates, where the errors of LS algorithm decrease most slowly; when the measurement rate is $\eta = 0.45$, the error is still error > 0.5 . Meanwhile, the errors of Danzig and ADMM algorithms are error < 0.01 and error < 0.001 , respectively. The ADMM algorithm error is the least in all three algorithms in any measurement rate η .

In the experiments with noise, the measured vector y is (5), the number of external noises $N_{o,S} = 0.01 \times 4^6 \approx 41$. If the error is larger than 1, we record it as 1. The experiment results are shown in Fig. 1b, from which one can see that the LS algorithm is robust toward external noises, but the estimation error is larger and decreases slowly. Dantzig algorithm has poor robust performance; it cannot handle the case of external noises, because the estimation error is always unable to converge. ADMM algorithm works best among three algorithms in error convergence. As the measurement rate continues to increase, the estimation accuracy with noises can reach above 90 %.

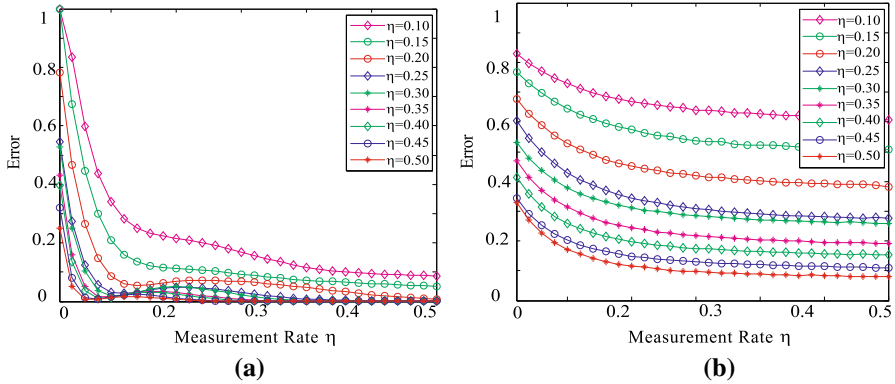


Fig. 2 Effects of measurement rate to the error of quantum state estimation of ADMM algorithm without and with noise in the case $n = 6$

3.2 Performance comparisons of ADMM algorithm without and with noise

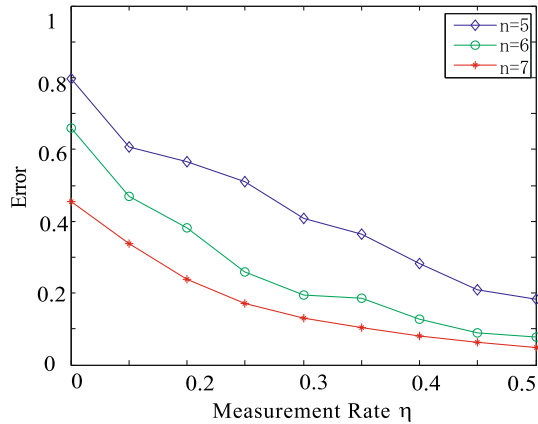
In this subsection, we verify the robustness of ADMM algorithm by comparing and analyzing the performance without and with noise in the case of qubits $n = 6$. In the experiments, the measurement rate ranges from $\eta = 0.1$ to $\eta = 0.5$, and step size of η is 0.05. The iteration number of ADMM algorithm is 30.

1. *Without noise* We first study the effects of the iteration number to the estimation error at different measurement rates without noise. With the different numbers of iteration, the experiment results of the estimation errors in different measurement rate η are shown in Fig. 2a, from which one can see that: When the iteration numbers increase, the errors of different measurement rates decrease quickly, all of which decrease below 0.01 in 20 iterations. When the measurement rates η are above 0.2, the errors only need several (< 5) iterations to decrease below 0.01, whose estimation accuracies are above 99%. Especially when $\eta = 0.45$, 99% accuracy can be reached in only 3 iterations; while $\eta = 0.10$ it needs at least 25 iterations. So one can conclude that: The larger the measurement rate η , the less the iteration number which needs to achieve the estimation accuracy.

2. *With external noises* In order to testify the robustness of ADMM algorithm, we do the experiments on the effects of measurement rate to the estimation error with noises. The number of external noise is selected 10% of \mathbf{O}^* , which is $No.S = 0.01 \times 4^6 \approx 41$, and the magnitudes of \mathbf{S} are ± 0.01 . The effects of measurement rate η to the error of quantum state estimation of ADMM algorithms without noise are shown in Fig. 2b, from which one can see that: As the iteration numbers increase, all the errors in different measurement rates gradually decrease, which can achieve the 90% estimation accuracy in the case of $\eta > 0.40$. When the measurement rate $\eta > 0.45$, it needs 15 iterations to achieve 90% accuracy.

We can conclude from the analysis of the experimental results that the ADMM algorithm has good robustness against the external noises. It can reach beyond 90% estimation accuracy at higher measurement rates, even when there is a larger external noise. The estimation errors can decrease quickly without noise and decrease gradually

Fig. 3 Effects of different qubits to the estimation errors of ADMM algorithm



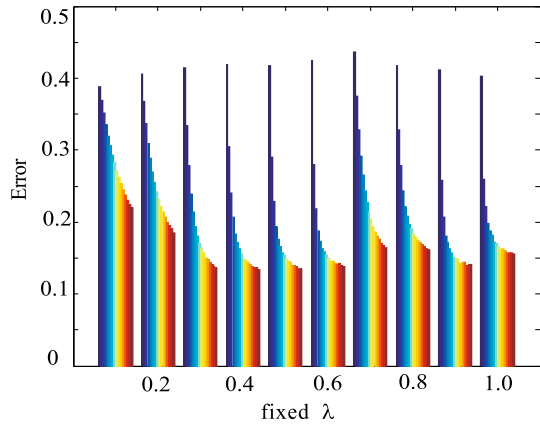
in the presence of external noises. When the estimation accuracy is fixed, the higher the measurement rate, the less the iteration number which needs to meet the accuracy; when the measurement rate η is fixed, the more the iteration number, the higher the estimation accuracy.

3.3 Performance comparisons of ADMM under different quantum qubits

Theoretically, when the number of qubits n increases, the element number of density matrix dramatically increases, which will cost expensive computation and make it much more difficult to estimate the quantum state. For example, when $n = 5$, the number of elements is $d \times d = 2^5 \cdot 2^5 = 1024$, and the dimension of observation matrices is $d^2 \times d^2 = 4^5 \cdot 4^5 \approx 1.05 \times 10^6$. As the number of qubits increases to $n = 7$, the number of elements and observation matrices increases significantly to $2^7 \cdot 2^7 = 16384$ and $4^7 \cdot 4^7 \approx 2.68 \times 10^8$, which dramatically increase the amount of computation. In this subsection, we study the effects of the number of qubits to the estimation errors with different measurement rates in the cases of $n = 5, 6$ and 7 with noise. The experimental results are shown in Fig. 3, from which one can see that in general, the error in the case $n = 5$ is the largest at any measurement rate, while the error with $n = 7$ is the smallest. When the number of qubit increases from $n = 5$ to $n = 7$ at the measurement rate $\eta = 0.45$, the errors decrease from 0.2153 to 0.0781, and the estimation accuracies increase from 78.47 to 93.19%. If the estimation accuracy is prefixed as 90%, then $n = 7$ needs smaller measurement rate than $n = 5$ to meet the accuracy.

In conclusion, the ADMM algorithm has significant advantage in the estimation of large amount of elements, and the superiority of ADMM demonstrates more significance as the number of qubits becomes larger.

Fig. 4 Effects of different fixed weight values to the estimation errors



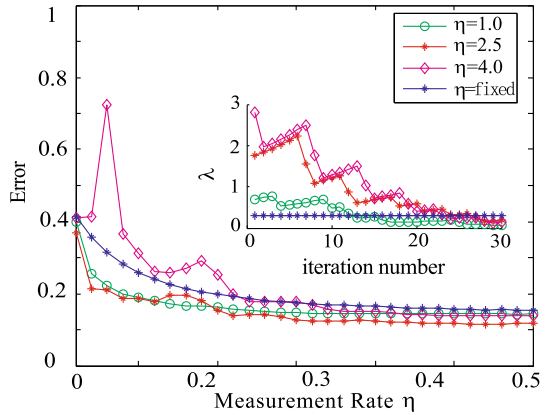
3.4 Performance comparisons between the fixed and adaptive weight values of ADMM

In this subsection, we verify the improvement of the proposed adaptive ADMM algorithm by comparing and analyzing the performances between fixed and adaptive weight values.

1. *Determination of the best performance with fixed weight values:* In the experiments, the effects of different fixed weight values λ_1 in (9) to the estimation error are studied at measurement rate $\eta = 0.4$ in the cases of qubits $n = 6$. We select 10 different fixed weight values λ_1 from $\lambda_1 = 0, 1$ to $1.0, \Delta\lambda_1 = 0.1$. Each value λ_1 is for one individual experiment. We compare the effects of the different weight values λ_1 to corresponding estimation errors. The experimental results are shown in Fig. 4, in which each of the histogram represents an estimation error decreasing with a fixed λ_1 in 30 iterations. The color of the histograms changing from blue (left) to red (right) means the numbers of iterations increase from 1 to 30. Neither too large nor too small λ_1 is suitable for obtaining the least estimation error. We can draw a conclusion that the selection of weight value λ_1 determines the final estimation error. The best suitable value in this experiment is $\lambda_{fixed} = 0.3$, in this case the minimum error = 0.1240.

2. *Comparisons between adaptive and fixed weight values:* To validate the improvement of the proposed adaptive weight value ADMM algorithm, we do the comparisons of the estimation errors between different initial adaptive weight values and fixed optimal weight value $\lambda_{fixed} = 0.3$ at measurement rate $\eta = 0.4$ in the case of qubits $n = 6$. The initial values of adaptive weights are set as $\lambda_1 = 1.0, 2.5$ and 4.0 , respectively. The experimental results are shown in Fig. 5, from which one can see that: (a) The estimation error of adaptive initial weight value $\lambda_1 = 2.5$ rapidly decreases to 0.2009 only in 2 iterations, while the estimation error of fixed $\lambda_{fixed} = 0.3$ uses at least 10 iterations to slowly decrease to 0.2050. (b) The estimation error of adaptive $\lambda_1 = 2.5$ is the least in all three cases, whose estimation accuracy is $\Delta = 3.6\%$ higher than the fixed $\lambda_{fixed} = 0.3$. (c) Regardless of the initial values of the adaptive weight, all the estimation errors of the adaptive weight values are less than the fixed ones. The estimation accuracy of adaptive ADMM algorithm is higher.

Fig. 5 Effects of different initial adaptive and fixed values to the estimation errors



We can draw a conclusion All the estimation accuracies using adaptive weight values are higher than the fixed ones. The initial values of the adaptive λ_1 affect little in adaptive method. By using adaptive weight value, one can get the expect estimation accuracy at less iterations.

4 Conclusion

In this paper, we estimated quantum system states based on the theory of compressive sensing by using the Least Square method, Dantzig algorithm and ADMM algorithm. Experimental results indicated that by using compressive sensing, one is able to estimate higher dimensional density matrix in the same number of iteration, especially when the number of system qubits becomes larger, the adaptive weight ADMM algorithm can not only obtain smaller estimation error but also have higher robustness.

Acknowledgments This work was supported by the National Natural Science Foundation of China (61573330).

Appendix

Definition 1 (Rank RIP) [14, 23] The \mathcal{A} satisfies the rank-restricted isometry property (RIP) if for all $d \times d \mathbf{X}$, we have

$$(1 - \delta)\|\mathbf{X}\|_F \leq \|\mathcal{A}(\mathbf{X})\|_2 \leq (1 + \delta)\|\mathbf{X}\|_F \tag{29}$$

where some constant $0 < \delta < 1$.

References

1. Smith, A., Riofro, C., Anderson, B., Martinez, H., Deutsch, I., Jessen, P.: Quantum state tomography by continuous measurement and compressed sensing. *Phys. Rev. A* **87**, 030102 (2013)
2. Wu, X., Xu, K.: Partial standard quantum process tomography. *Quantum Inf. Process.* **12**(2), 1379–1393 (2013). doi:[10.1007/s11128-012-0473-9](https://doi.org/10.1007/s11128-012-0473-9)

3. Heinosaari, T., Mazzarella, L., Wolf, M.: Quantum tomography under prior information. *Commun. Math. Phys.* **318**(2), 355–374 (2013). doi:[10.1007/s00220-013-1671-8](https://doi.org/10.1007/s00220-013-1671-8)
4. Wu, L.A., Byrd, M.: Self-protected quantum algorithms based on quantum state tomography. *Quantum Inf. Process.* **8**(1), 1–12 (2009). doi:[10.1007/s11128-008-0090-9](https://doi.org/10.1007/s11128-008-0090-9)
5. Baraniuk, R.: Compressive sensing. *IEEE Signal Process. Mag.* **24**(4), 118–121 (2007). doi:[10.1109/MSP.2007.4286571](https://doi.org/10.1109/MSP.2007.4286571)
6. Donoho, D.L.: Compressed sensing. *IEEE Trans. Inf. Theory* **52**, 1289–1306 (2006)
7. Gross, D., Liu, Y., Flammia, S.T., Becker, S., Eisert, J.: Quantum state tomography via compressed sensing. *Phys. Rev. Lett.* **105**(15), 150401 (2010)
8. Schwemmer, C., Tóth, G., Niggelbaum, A., Moroder, T., Gross, D., Gühne, O., Weinfurter, H.: Experimental comparison of efficient tomography schemes for a six-qubit state. *Phys. Rev. Lett.* **113**(5), 0401503 (2014)
9. Lloyd, S., Mohseni, M., Rebentrost, P.: Quantum principal component analysis. *Nat. Phys.* **10**(9), 631–633 (2014)
10. Shabani, A., Kosut, R.L., Mohseni, M., Rabitz, H., Broome, M.A., Almeida, M.P., Fedrizzi, A., White, A.G.: Efficient measurement of quantum dynamics via compressive sensing. *Phys. Rev. Lett.* **106**(4), 100401 (2011). doi:[10.1103/PhysRevLett.106.100401](https://doi.org/10.1103/PhysRevLett.106.100401)
11. Flammia, S.T., Gross, D., Liu, Y.K., Eisert, J.: Quantum tomography via compressed sensing: error bounds, sample complexity and efficient estimators. *New J. Phys.* **14**(9), 095022 (2012)
12. Miosso, C., von Borries, R., Argaez, M., Velazquez, L., Quintero, C., Potes, C.: Compressive sensing reconstruction with prior information by iteratively reweighted least-squares. *IEEE Trans. Signal Process.* **57**(6), 2424–2431 (2009). doi:[10.1109/TSP.2009.2016889](https://doi.org/10.1109/TSP.2009.2016889)
13. Kosut, R., Lidar, D.: Quantum error correction via convex optimization. *Quantum Inf. Process.* **8**(5), 443–459 (2009). doi:[10.1007/s11128-009-0120-2](https://doi.org/10.1007/s11128-009-0120-2)
14. Liu, Y.: Universal low-rank matrix recovery from pauli measurements. In: *Proceedings of Advances in Neural Information Processing Systems*, pp. 1638–1646 (2011)
15. Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J.: Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.* **3**(1), 1–22 (2011)
16. He, B., Yang, H., Wang, S.: Alternating direction method with self-adaptive penalty parameters for monotone variational inequalities. *J. Optim. Theory Appl.* **106**(2), 337–356 (2000). doi:[10.1023/A:1004603514434](https://doi.org/10.1023/A:1004603514434)
17. Lin, Z., Liu, R., Su, Z.: Linearized alternating direction method with adaptive penalty for low rank representation. In: *Proceedings of Advances in Neural Information Processing Systems*, pp. 612–620 (2011)
18. Gross, D.: Recovering low-rank matrices from few coefficients in any basis. *IEEE Trans. Inf. Theory* **57**(3), 1548–1566 (2011). doi:[10.1109/TIT.2011.2104999](https://doi.org/10.1109/TIT.2011.2104999)
19. Wright, J., Ganesh, A., Min, K., Ma, Y.: Compressive principal component pursuit. *J. IMA* **2**, 32–68 (2013)
20. Yuan, X.M., Yang, J.: Sparse and low-rank matrix decomposition via alternating direction methods. *Pac. J. Optim.* (2009)
21. Li, K., Cong, S.: A robust compressive quantum state tomography algorithm using admm. In: *The 19th World Congress of the International Federation of Automatic Control*, pp. 6878–6883 (2014)
22. Cong S., Z.H., K., L.: An improved quantum state estimation algorithm via compressive sensing. In: *2014 IEEE international conference on Robio and Biomimetics*, 5–10, pp. 2238–2343 (2014)
23. Recht, B., Fazel, M., Parillo, P.: Guaranteed minimum rank solution of matrix equations via nuclear norm minimization. *SIAM Rev.* **52**, 471–501 (2007)
24. Zyczkowski, K., Penson, K.A., Nechita, I., Collins, B.: Generating random density matrices. *J. Math. Phys.* **52**(6), 062201 (2011)
25. Boyd, S., Vandenberghe, L.: *Convex Optimization*. Cambridge Univ. Press, Cambridge, U.K. (2004)
26. Candès, E.J., Li, X., Ma, Y., Wright, J.: Robust principal component analysis? *J. ACM* **58**(3), 1–37 (2011)
27. Candès, E., Romberg, J.: Quantitative robust uncertainty principles and optimally sparse decompositions. *Found. Comput. Math.* **6**(8), 227–254 (2006)
28. Candès, E., Romberg, J.: Sparsity and incoherence in compressive sampling. *Inverse Probl.* **23**, 969–985 (2007)
29. Sturm, J.F.: Using sedumi 1.02, a matlab toolbox for optimization over symmetric cones. *Optim. Methods Softw.* **11–12**, 625–653 (1999)