Preprints of the 19th World Congress
The International Federation of Automatic Control
Cape Town, South Africa. August 24-29, 2014

A Robust Compressive Quantum State Tomography Algorithm Using ADMM ⋆

Kezhi Li * **Shuang Cong** **

* *ACCESS Linnaeus Centre, Royal Institute of Technology (KTH), SE 10044 Stockholm, Sweden (e-mail: kezhi@kth.se).*
** *Dept. of Automation, Univ. of Science and Technology of China, Hefei, 230027, China (Corresponding authors email: scong@ustc.edu.cn)*

Abstract: The possible state space dimension increases exponentially with respect to the number of qubits. This feature makes the quantum state tomography expensive and impractical for identifying the state of merely several qubits. The recent developed approach, compressed sensing, gives us an alternative to estimate the quantum state with fewer measurements. It is proved that the estimation then can be converted to a convex optimization problem with quantum mechanics constraints. In this paper we present an alternating augmented Lagrangian method for quantum convex optimization problem aiming to recover pure or near pure quantum states corrupted by sparse noise given observables and the expectation values of the measurements. The proposed algorithm is much faster, robust to outlier noises (even very large for some entries) and can solve the reconstruction problem distributively. The simulations verify the superiority of the proposed algorithm and compare it to the conventional least square and compressive quantum tomography using the Dantzig method.

Keywords: Quantum state tomography, ADMM, rank minimization, convex optimization and regularization

1. INTRODUCTION

The interests of applying control theory and signal processing techniques to quantum mechanics have increased dramatically in recent decades. One objective is to develop a series of systematic methods for the active manipulation and control of quantum systems. The foundation of such theory lies in the fact that we are capable to prepare and measure a given quantum state efficiently. It is not trivial since the microscopic quantum systems have their unique features, on account of which they significantly differ from the classic world. In practice, people often use the measurement data to estimate an unknown quantum state. In mathematics, a quantum pure state $|\psi\rangle$ can be described as a vertical vector with the size d in a Hilbert space. This vector is called a state vector and it theoretically contains the statistical information about the quantum system. For the mixed state that corresponds to a probabilistic mixture of pure states, a state vector is not enough. It usually requires a $d \times d$ density matrix $\boldsymbol{\rho}$ to depict the quantum state by giving the probabilities in each possible state, which implies that $O(d^2)$ parameters are needed to describe an arbitrary quantum state in a d dimensional Hilbert space. Various reconstruction algorithms have been proposed and experimentally tested (Lvovsky and Raymer [2009], Liu et al. [2012]), such as methods based on inverse linear transformation, maximum-likelihood and maximum-entropy.

However, in fact most states people are interested in real life are nearly pure. Here nearly pure means the quantum state is a mixed state that can be represented as the probabilistic combination of equal to or less than r pure states. Suppose that the unknown mixed state is a probabilistic mixture of $r \sim O(1)$ pure states, then it means that its density matrix $\boldsymbol{\rho}$ has rank not larger than r. This prior information enables us to reduce the number of parameters to identify a quantum system (Gross et al. [2010]). By using a novel signal processing technique called compressed sensing (CS) (Donoho [2006]) that has been widely investigated in last a few years, researchers are able to obtain good estimates of nearly pure quantum states with $O(rd \log d)$ expectations and corresponding observables (Gross et al. [2010], Liu [2011]). Thus the required number of identifying a quantum state can be reduced dramatically by solving an optimization problem when d goes to large, and its effectiveness has been verified by a series of experiments such as in Smith et al. [2013]. While in current literatures of quantum state tomography via CS, this problem cannot be efficiently handled by generic optimization solvers because of the large number of involving variables. For instance in Smith et al. [2013] the authors summarized the estimation to a least squares (LS) problem or a compressed sensing (CS) problem, and solved them by using the prevalent convex optimization toolbox. This paper addresses the problem of state recovery using the low rank information. Our work is

⋆ This work was partially supported by the Swedish Research Council, the Linnaeus Center ACCESS at KTH, the European Research Council under the advanced grant LEARN, contract 267381, and the China National Key Basic Research Program under Grant No. 2011CBA00200

inspired by the alternating augmented Lagrangian method (ADMM) which has received much attentions from the optimization community (Boyd and Vandenberghe [2004], Boyd et al. [2011]) though originally it was developed in the 1970s. In this paper we reformulate the quantum state tomography to an optimization problem and design a fast algorithm based on ADMM where in each iteration we try to optimize the density matrix and project it onto the constraint set of quantum states according to the Karush-Kuhn–Tucker (KKT) conditions, and the algorithm finally reaches a solution with good accuracy. Due to the alternating properties of the proposed approach, we are able to run it distributively, and normally obtain the quantum state with most purified result by minimizing the nuclear norm of the density matrix, which can be seen as a heuristic for minimizing the rank (Candes and Plan [2011]). People have proved that many matrices bases, including Pauli matrices, satisfy the rank restricted isometry property (RIP) introduced in Recht et al. [2007], hence by using Pauli matrices as sensing matrices we are able to recover the unique density matrix with sufficient measurements via compressive sensing approach. This work can be deemed as the reconstruction part of the compressive quantum state tomography, which gives a solution to identifying the density matrices accurately and efficiently for standard tomography as well as continuous tomography (Smith et al. [2013]).

This paper is organized as follows. In Section 2, we will explain the idea of quantum state tomography via compressive sensing and the framework of ADMM. In Section 3 the compressive state tomography with quantum constraints is formulated formally, and the proposed algorithm is introduced and analyzed in detail. Simulations verify the effectiveness of the proposed approach in Section 4, and finally the conclusion is summarized in Section 5. Moreover, some necessary supplementary knowledge is explained in the Appendix.

Notation: Bold letters are used to denote a vector or a matrix. For vectors, $||\cdot||_1, ||\cdot||_2$ represent the l_1, l_2 norm, respectively. For matrix, $\mathbf{A}^T$ and $\mathbf{A}^*$ denote the transpose and Hermitian transpose of $\mathbf{A}$, respectively. $||\cdot||_p$ denotes the Schatten p-norm with $||\mathbf{A}||_p = \left(\sum_i \sigma_i(\mathbf{A}^p)^{1/p}\right)$, where $\sigma_i(\mathbf{A})$ are the singular values of $\mathbf{A}$. Specifically, $||\cdot||_*$ is the nuclear norm and $||\cdot||_F$ represents the Frobenius norm. $\mathrm{tr}(\cdot)$ is the operator to calculate the trace. $\mathrm{vec}(\mathbf{A})$ represents the vertical vector concatenates $\mathbf{A}$'s columns, and "mat" is its inverse operator to convert a vector to a matrix. Bra-ket notations $|\psi_i\rangle$ are used to denote quantum states. $\mathbf{A} \succeq 0$ means $\mathbf{A}$ is a positive semi-definite (p.s.d) matrix.

2. COMPRESSIVE QUANTUM STATE TOMOGRAPHY AND ADMM

2.1 Quantum State Tomography Via Compressive Sensing

The task of quantum state tomography is to reconstruct the quantum states processed and produced by physical systems. Due to the special characteristics of the quantum mechanics, a $d \times d, d = 2^q$ density matrix ρ, a quantum-mechanical analogue to a phase-space probability measure, is used to describe a quantum system, where q denotes the number of qubits under consideration. Since the degrees of the freedom of ρ are $d \times d$, usually people need the number of measurements increasing with exponential growth regarding the state space dimension d in order to identify $\boldsymbol{\rho}$. If we make the measurements discretely and denote the observable matrix $\mathbf{O}_i$, the expectation of measurements $\mathbf{y}_i \in \mathbb{R}^m$, and measuring operator $\mathcal{A} : \mathcal{C}^{d\times d\to m}$, then

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

where $\mathbf{A} \in \mathbb{C}^{m\times d^2}$ is the normalized operator whose ith row is the concatenation of $\mathbf{O}_i^*$'s rows, $\mathbf{e} \in \mathbb{R}^m$ represents the noise caused by the system or measuring process. c is some normalized constant. If we set $E(\mathcal{A}^*\mathcal{A}) = \mathcal{I}$ where E represents the expectation over all $\mathcal{A}$, c would be $\frac{d}{\sqrt{m}}$. Conventionally, people use the least square approach to estimate $\boldsymbol{\rho}$

$$\begin{aligned}\hat{\boldsymbol{\rho}} = \arg\min_{\boldsymbol{\rho}} \sum_i \left[\mathbf{y}_i - c\cdot \mathrm{tr}(\mathbf{O}_i^*\boldsymbol{\rho})\right]^2, \\ \text{s.t. } \boldsymbol{\rho}^* = \boldsymbol{\rho},\ \boldsymbol{\rho} \succeq 0,\ \mathrm{tr}(\boldsymbol{\rho}) = 1.\end{aligned} \tag{2}$$

Because the degrees of freedom of $\boldsymbol{\rho}$ are $O(d^2)$, normally $O(d^2)$ measurements are needed to identify the unique state.

Yet if we assume that the underlying quantum system is pure or nearly pure, $\boldsymbol{\rho}$ becomes a probabilistic weighted combination of equal to or less than r rank-1 matrices derived from a series pure states (see details in the Appendix). When r is small, people have suggested that $O(rd\log d)$ settings would possibly suffice instead of d^2. Minimizing the rank of a matrix belongs to NP-hard problems, so alternatively people pursuit the solution by minimize the nuclear norm of the density matrix $||\boldsymbol{\rho}||_* = \mathrm{tr}\left(\sqrt{\boldsymbol{\rho}^*\boldsymbol{\rho}}\right) = \sum_{i=1}^{\min\{m,n\}}\sigma_i$, which is a convex function that can be optimized efficiently. The nuclear norm has been proved as the best convex approximation of the rank function over the unit ball (Recht et al. [2007]), so minimizing $||\boldsymbol{\rho}||_*$ is a heuristic for minimizing the rank (Gross et al. [2010]). Thus $\boldsymbol{\rho}$ with low rank can be estimated by compressed sensing approaches such as in Liu [2011]:

$$\begin{aligned}&\text{Dantzig with quantum constraints:} \\ &\hat{\boldsymbol{\rho}} = \arg\min_{\boldsymbol{\rho}} ||\boldsymbol{\rho}||_* \\ &\text{s.t. } \sum_i \left[\mathbf{y}_i - c\cdot \mathrm{tr}(\mathbf{O}_i^*\boldsymbol{\rho})\right]^2 \le \epsilon, \boldsymbol{\rho}^* = \boldsymbol{\rho},\ \boldsymbol{\rho} \succeq 0, \quad \text{or} \\ &\text{LASSO with quantum constraints:} \\ &\hat{\boldsymbol{\rho}} = \arg\min_{\boldsymbol{\rho}} \frac{1}{2}||\mathbf{y}_i - c\cdot \mathrm{tr}(\mathbf{O}_i^*\boldsymbol{\rho})||_2^2 + \mu||\boldsymbol{\rho}||_* \\ &\text{s.t. } \boldsymbol{\rho}^* = \boldsymbol{\rho},\ \boldsymbol{\rho} \succeq 0,\end{aligned} \tag{3}$$

where ϵ, μ are parameters. In this paper we develop a convex optimization algorithm based on ADMM to solve above problems corrupted by sparse outliers with quantum constraints.

2.2 Alternating Direction Method of Multipliers (ADMM)

ADMM is an optimization method equipped with good robustness and can support decomposition. Consider an optimization problem:

$$\text{minimize } \ f(\mathbf{x}) + g(\mathbf{z}) \ \text{ s.t. } \ \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z} = \mathbf{c} \tag{4}$$

for some variable $\mathbf{x}, \mathbf{z} \in \mathbb{R}^n$, where $f, g : \mathbb{R}^n \to \mathbb{R}$ are two convex functions. The augmented Lagrangian of (4) is defined as:

$$\begin{aligned} L_\lambda(\mathbf{x}, \mathbf{z}, \mathbf{w}) &= f(\mathbf{x}) + g(\mathbf{z}) \\ &+ \mathbf{w}^T(\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z} - \mathbf{c}) + \frac{\lambda}{2}||\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z} - \mathbf{c}||_2^2. \end{aligned} \tag{5}$$

where $\lambda > 0$ is a tunable parameter. Then the kth iteration of ADMM algorithm consists of three steps as follows:

$$\begin{aligned} &1)\ \mathbf{x}^{k+1} = \arg\min_{\mathbf{x}} L_\lambda(\mathbf{x}, \mathbf{z}^k, \mathbf{w}^k) \quad // \ \mathbf{x}\text{-minimization} \\ &2)\ \mathbf{z}^{k+1} = \arg\min_{\mathbf{z}} L_\lambda(\mathbf{x}^{k+1}, \mathbf{z}^k, \mathbf{w}^k) \quad // \ \mathbf{z}\text{-minimization} \\ &3)\ \mathbf{w}^{k+1} = \mathbf{w}^k + \lambda(\mathbf{A}\mathbf{x}^{k+1} + \mathbf{B}\mathbf{z}^{k+1} - \mathbf{c}) \quad // \text{ dual-update} \end{aligned} \tag{6}$$

From above steps one can see that if we minimize over $\mathbf{x}$ and $\mathbf{z}$ jointly, the approach reduces to the classic method of multipliers. Instead, people split the augmented Lagrangian and minimize over $\mathbf{x}$ with $\mathbf{z}$ fixed and vice versa. The three steps are repeated until the convergence. Certain stopping criteria is made to decide when the algorithm is thought to achieve a convergence. For instance, the algorithm is iterated until the primal and dual residuals are bounded

$$\begin{aligned} ||\mathbf{A}\mathbf{x}^k + \mathbf{B}\mathbf{z}^k - \mathbf{c}||_2^2 &\le \varepsilon_{\text{pri}}, \\ ||\mathbf{x}^k - \mathbf{x}^{k-1}||_2^2 + ||\mathbf{z}^k - \mathbf{z}^{k-1}||_2^2 &\le \varepsilon_{\text{dual}}, \end{aligned} \tag{7}$$

where $\varepsilon_{\text{pri}} > 0, \varepsilon_{\text{dual}} > 0$ are tolerance parameters. For more details and a complete convergence analysis, people who have interests may refer to Boyd et al. [2011].

3. PROBLEM FORMULATION AND METHOD

In this section, we formulate the problem of the robust quantum state tomography and derive an efficient optimization algorithm using ADMM. Here "robust" means the algorithm fits for the circumstance of the existence of not only small random noises, but also sparse outlier noises involved in the density matrix.

3.1 Robust Compressive Quantum State Tomography

During the measuring process of quantum state tomography, noises are involved due to the system or measurement errors. Normally we assume $\mathbf{e}$ satisfying certain distribution (like Gaussian) and it can be minimized with least square techniques (2), similar in Danzig or LASSO (3). However there exist abnormal circumstances in the measuring process that may cause the perturbation in the density matrix, and it can be reflected by sparse outlier entries in $\boldsymbol{\rho}$ and of course these outlier entries do not satisfy the Gaussian distribution. We formulate these outlier entries as a sparse matrix $\mathbf{S} \in \mathbb{C}^{d\times d}$, then (1) becomes

$$\begin{aligned} \mathbf{y}_i = (\mathcal{A}(\boldsymbol{\rho} + \mathbf{S}))_i + e_i = c \cdot \text{tr}(\mathbf{O}_i^*(\boldsymbol{\rho} + \mathbf{S})) + e_i, \\ i = 1, \cdots, m. \end{aligned} \tag{8}$$

In this case the result of the least square method in (2) will change significantly sometimes because of the existence of outliers. In addition, given the information that $\boldsymbol{\rho}$ is relatively pure which implies that it has low rank, the Dantzig/LASSO solver in (3) with low rank constraints based on truncated Singular Value Decomposition (SVD) might also fail because the sparse outliers effect the classic principle component analysis (PCA) dramatically in the process of dimensionality reduction. To reduce the influence of the noise to the rank estimation, we may reformulate a robust Dantzig solver with sparse outliers and quantum constraints to

$$\begin{aligned} &\text{minimize} \ \ ||\boldsymbol{\rho}||_* + ||\mathbf{S}||_1 \\ &\text{s.t.} \ ||\mathbf{y} - \mathbf{A}\text{vec}(\boldsymbol{\rho} + \mathbf{S})||_2^2 \le \epsilon, \boldsymbol{\rho}^* = \boldsymbol{\rho}, \ \boldsymbol{\rho} \succeq 0, \end{aligned} \tag{9}$$

where $\mathbf{A}$ is with the same definition in (1). The idea of minimizing sparse noises can also be found in E. J. Candés and Wright [2011], Zhou et al. [2010], Kyrillidis and Cevher [2012] and has many applications in the face recognition area, etc. While in most previous papers the authors aimed at solving a matrix completion problem however here we want to recover the density matrix from observable measurements with special constraints on $\boldsymbol{\rho}$. To involve the quantum constraints in ADMM, we rewrite (9) as

$$\begin{aligned} &\text{minimize} \ \ ||\boldsymbol{\rho}||_* + I_{\mathcal{C}}(\boldsymbol{\rho}) + ||\mathbf{S}||_1 \\ &\text{s.t.} \ ||\mathbf{y} - \mathbf{A}\text{vec}(\boldsymbol{\rho} + \mathbf{S})||_2^2 \le \epsilon, \end{aligned} \tag{10}$$

where $I_{\mathcal{C}}(\boldsymbol{\rho})$ is the indictor function on a convex set $\mathcal{C}$ with $I_{\mathcal{C}}(\boldsymbol{\rho}) = 0$ for $\boldsymbol{\rho} \in \mathcal{C}$, and $I_{\mathcal{C}}(\boldsymbol{\rho}) = \infty$ for $\boldsymbol{\rho} \notin \mathcal{C}$, $\mathcal{C}(\boldsymbol{\rho})$ here is the Hermitian p.s.d. set whose entries satisfy $\boldsymbol{\rho}^* = \boldsymbol{\rho}, \ \boldsymbol{\rho} \succeq 0$. So we have obtained two sets of variables with separable objectives. Then the augmented Lagrangian can be derived from (5) as

$$\begin{aligned} L_{\lambda_1}(\boldsymbol{\rho}, \mathbf{S}, \mathbf{w}) &= (||\boldsymbol{\rho}||_* + I_{\mathcal{C}}(\boldsymbol{\rho})) + ||\mathbf{S}||_1 \\ &+ \mathbf{w}^T(\mathbf{A}\text{vec}(\boldsymbol{\rho}) + \mathbf{A}\text{vec}(\mathbf{S}) - \mathbf{y}) \\ &+ \frac{\lambda_1}{2}||\mathbf{A}\text{vec}(\boldsymbol{\rho}) + \mathbf{A}\text{vec}(\mathbf{S}) - \mathbf{y}||_2^2, \end{aligned} \tag{11}$$

where λ_1 is a parameter that can effect the rate of convergence and the number of iterations. Alternatively we may combine the linear and quadratic terms in (11) and it becomes

$$\begin{aligned} L_{\lambda_1}(\boldsymbol{\rho}, \mathbf{S}, \mathbf{u}) &= (||\boldsymbol{\rho}||_* + I_{\mathcal{C}}(\boldsymbol{\rho})) + ||\mathbf{S}||_1 \\ &+ \frac{\lambda_1}{2}||\mathbf{A}\text{vec}(\boldsymbol{\rho}) + \mathbf{A}\text{vec}(\mathbf{S}) - \mathbf{y} + \mathbf{u}||_2^2, \end{aligned} \tag{12}$$

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

3.2 ADMM Steps

We carry out the following steps in each iteration of the ADMM algorithm to solve (10).

Step 1 In the $\boldsymbol{\rho}$ minimization step, we update the low rank $\boldsymbol{\rho}$ matrix with fixed $\mathbf{S}, \mathbf{u}$:

$$\begin{aligned} \boldsymbol{\rho}^{k+1} :=& \arg\min_{\boldsymbol{\rho}} \Big\{ ||\boldsymbol{\rho}||_* + I_{\mathcal{C}}(\boldsymbol{\rho}) \\ &+ \frac{\lambda_1}{2}||\mathbf{A}\text{vec}(\boldsymbol{\rho}) + \mathbf{A}\text{vec}(\mathbf{S}^k) - \mathbf{y} + \mathbf{u}^k||_2^2 \Big\}, \end{aligned} \tag{13}$$

where k is the number of iterations. First, we minimize the unconstrained quadratic function in terms of $\boldsymbol{\rho}$. The analytic solution to the least square estimation can be written as

$$\boldsymbol{\rho}_1^{k+1} = \text{mat}\left((\mathbf{A}^*\mathbf{A})^{-1} \mathbf{A}^* \left(\mathbf{y} - \mathbf{u}^k - \mathbf{A}\text{vec}(\mathbf{S}) \right) \right). \tag{14}$$

Second, projecting $\boldsymbol{\rho}_1^{k+1}$ on to the constraints set $\mathcal{C}$ at the same time with low rank, and denoting the result as $\boldsymbol{\rho}_2^{k+1}$, *i.e.*

$$\boldsymbol{\rho}_2^{k+1} = \Pi_{\mathcal{C}}(\boldsymbol{\rho}_1^{k+1}), \tag{15}$$

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 Section 3.3 with efficient approach.

Step 2 In the $\mathbf{S}$ minimization step, we update the sparse matrix $\mathbf{S}$ with fixed $\boldsymbol{\rho}^{k+1} = \boldsymbol{\rho}_2^{k+1}, \mathbf{u}$.

$$\mathbf{S}^{k+1} := \arg\min_{\mathbf{S}} \left\{ ||\mathbf{S}||_1 + \frac{\lambda_1}{2} ||\mathbf{A}\mathrm{vec}(\boldsymbol{\rho}^{k+1}) + \mathbf{A}\mathrm{vec}(\mathbf{S}) - \mathbf{y} + \mathbf{u}^k||_2^2 \right\}. \tag{16}$$

It is a conventional LASSO problem and can be solved by iterations. However here we avoid solving it by a sequence of convex programs and adopt a shrink operator to calculate a solution efficiently. In detail, the least square estimate $\mathbf{S}$ can be approximated by

$$\mathbf{S}_1^{k+1} = \mathrm{mat}\left(\left(\mathbf{A}^*\mathbf{A}\right)^{-1} \mathbf{A}^* \left(\mathbf{y} - \mathbf{u}^k - \mathbf{A}\mathrm{vec}(\boldsymbol{\rho}^{k+1})\right)\right), \tag{17}$$

and then we shrink the magnitude to achieve a sparse solution

$$\mathbf{S}_2^{k+1} = \mathcal{S}_{\tau'}(\mathbf{s}) = \mathrm{sgn}[\mathbf{s}] \max(|\mathbf{s}| - \tau'\mathbf{1}, \mathbf{0}) \tag{18}$$

where $\mathcal{S}$ is the shrink operator which is also adopted in Section 3.3, $\mathbf{s} = \mathrm{vec}(\mathbf{S}_1^{k+1})$, τ' is a shrink parameter depends on the sparsity level of $\mathbf{S}$.

Step 3 At last we proceed the dual update step:

$$\mathbf{u}^{k+1} = \mathbf{u}^k + (\mathbf{y} - \mathbf{A}\mathrm{vec}(\boldsymbol{\rho}^{k+1}) - \mathbf{A}\mathrm{vec}(\mathbf{S}^{k+1})). \tag{19}$$

This step is to record the alternative update direction and prepare for the next step.

Stop Criteria and Parameter Settings The algorithm follows the steps 1-3 to carry out the updating information iteratively. In practice, relatively small numbers of iterations, like 30-40, are sufficient to achieve a good accuracy. There are several stopping criterions, e.g., adopting bounds in (7) we have

$$\begin{aligned} &||\mathbf{y} - \mathbf{A}\mathrm{vec}(\boldsymbol{\rho}^k + \mathbf{S}^k)||_2^2 \le \varepsilon_1 ||\mathbf{y}||_2, \\ &||\boldsymbol{\rho}^k - \boldsymbol{\rho}^{k-1}||_2 \le \varepsilon_2, \quad ||\mathbf{S}^k - \mathbf{S}^{k-1}||_2 \le \varepsilon_3. \end{aligned} \tag{20}$$

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 Yuan and Yang [2009], E. J. Candés and Wright [2011].

3.3 Projection onto the Constraint Set with Low Rank

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

$$\mathcal{S}_\tau(\mathbf{x}) = \mathrm{sgn}[\mathbf{x}] \max(|\mathbf{x}| - \tau\mathbf{1}, \mathbf{0}) \tag{21}$$

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

$$\boldsymbol{\rho}_2^{k+1} = \mathcal{D}_\tau(\boldsymbol{\rho}_1^{k+1}) = \mathbf{V}\mathcal{S}_\tau(\boldsymbol{\Sigma}^+)\mathbf{V}^* \tag{22}$$

where $\boldsymbol{\Sigma}, \mathbf{V}$ are obtained from the eigenvalue decomposition of a symmetrized matrix $1/2(\boldsymbol{\rho}_1^{k+1} + \boldsymbol{\rho}_1^{k+1^*})$,

$$\mathbf{V}\boldsymbol{\Sigma}\mathbf{V}^* = 1/2(\boldsymbol{\rho}_1^{k+1} + \boldsymbol{\rho}_1^{k+1^*}), \tag{23}$$

$\boldsymbol{\Sigma}^+$ 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, $\tau = 1/\lambda_1$. This approach can be derived from its Karush-Kuhn-Tucker (KKT) conditions of the optimal projection from $\boldsymbol{\rho}_2^{k+1}$ to set $\mathcal{C}$ with least square errors. Taking the indicator function $I_{\mathcal{C}}(\boldsymbol{\rho})$ for instance, under mild assumptions on a proper cone $\mathcal{C}$, the KKT conditions of

$$\begin{aligned} &\text{minimize} \quad ||\bar{\boldsymbol{\rho}} - \boldsymbol{\rho}||_2^2 \\ &\text{s.t.} \quad \bar{\boldsymbol{\rho}} \in I_{\mathcal{C}} \end{aligned} \tag{24}$$

are given by

$$\begin{aligned} &\bar{\boldsymbol{\rho}} \in I_{\mathcal{C}}, \quad \bar{\boldsymbol{\rho}} - \boldsymbol{\rho} = \theta, \\ &\theta \in I_{\mathcal{C}}, \quad \theta^* \bar{\boldsymbol{\rho}} = 0. \end{aligned} \tag{25}$$

The third term is because the positive semidenite cone is self-dual. Then the Euclidean projection can be derived by decomposing $\boldsymbol{\rho}$ into the difference of two orthogonal elements: one with nonnegative eigenvalues and one with negative part. After that the shrink operator leads to a solution satisfying low rank constraints. In addition, if given the information that the objective quantum state is the probabilistic linear combination of less than or equal to r pure states, we may project $\boldsymbol{\rho}$ to the set of r-rank matrices by selecting the maximum r positive eigenvalues in $\boldsymbol{\Sigma}^+$ in (22). For the details of the derivation the readers may refer to (Boyd and Vandenberghe [2004]).

Remark:

1) Regarding the convergence of ADMM and error bounds of recovering low rank matrix from its measurements the readers may refer to Boyd and Vandenberghe [2004], Boyd et al. [2011], J. Wright and Ma [2013], Lin et al. [2011]. If there is no analytical solution to (6), we may also use the semidefinite programs. Details and the software can be found in Sturm [1999].

2) In practice, the observable $\mathbf{O}_i$ is not necessary the tensor product of Pauli matrices. For instance, in Smith et al. [2013] 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 $\mathbf{O}_i$ satisfy 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.

3) 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 $f_i(\cdot)$, $i = 1, 2, \cdots$ as in (1), in this case (6) turns to

$$\begin{aligned} &\mathbf{x}_i^{k+1} = \\ &\arg\min_{\mathbf{x}_i} \left(f_i(\mathbf{x}_i) + \mathbf{y}_i^{k^T}(\mathbf{x}_i - \bar{\mathbf{x}}_i^k) + \frac{\lambda}{2}||\mathbf{x}_i - \bar{\mathbf{x}}_i^k||_2^2 \right), \\ &\mathbf{y}_i^{k+1} = \mathbf{y}_i^k + \lambda(\mathbf{x}_i^{k+1} - \bar{\mathbf{x}}_i^{k+1}), \end{aligned} \tag{26}$$

where $\bar{\mathbf{x}}_i^k = 1/n_i \sum_{i=1}^{n_i} \mathbf{x}_i^k$ represents the average of n neighbours of agent i. Generally speaking, we gather $\mathbf{x}_i^k$ from outside and scatter $\bar{\mathbf{x}}^k$ to processors, 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 Boyd et al. [2011].

4. NUMERICAL EXAMPLES

In the following we demonstrate the reconstruction performance of the proposed algorithm for quantum state tomography. Two experiments are carried out to show the superior of the proposed algorithm. Consider a quantum state consisting of $q = 5$ qubits, its density matrix $\boldsymbol{\rho}$ has size $d \times d$, $d = 2^5$. Let the true quantum state as $\boldsymbol{\rho}^*$, $\boldsymbol{\rho}^*$ is generated from normalized Wishart random matrices with form as (Zyczkowski et al. [2011])

$$\boldsymbol{\rho}^* = \frac{\boldsymbol{\Psi}_\mathbf{r}\boldsymbol{\Psi}_\mathbf{r}^*}{\mathrm{tr}(\boldsymbol{\Psi}_\mathbf{r}\boldsymbol{\Psi}_\mathbf{r}^*)}, \tag{27}$$

where $\boldsymbol{\Psi}_\mathbf{r}$ is a complex $d \times r$ matrix with i.i.d. complex random Gaussian entries, the denominator is constructed due to the trace 1 constraint of the density matrix. We construct $\mathbf{A}$ as a $M \times d^2$ sampling matrix whose M rows are chosen randomly without replacement from an $d^2 \times d^2$ matrix whose rows are the set of all vecterized tensor product of Pauli matrices. Maltab R2012b version is used to run the numerical simulations and each data in figures is recorded after averaging 200 experiments.

At first we consider the scenario when the system has small random noises. Here we set e_i in (1) satisfies random Gaussian distribution $\mathcal{N}(0, 0.001||\boldsymbol{\rho}||_2)$. In this case there are two terms in (10) without $\mathbf{S}$, then the problem is simplified to

$$\begin{aligned} &\text{minimize} ||\mathbf{y} - \mathbf{A}\mathrm{vec}(\boldsymbol{\rho})||_2 + I_{\mathcal{C}}(\mathbf{z}), \\ &\text{s.t.} \quad \boldsymbol{\rho} = \mathbf{z}, \end{aligned} \tag{28}$$

where $\mathcal{C}$ represents the low rank Hermitian p.s.d. matrix set. We may update the quadratic term and $I_{\mathcal{C}}(\boldsymbol{\rho})$ iteratively using ADMM. Specifically, the iteration steps are

$$\begin{aligned} &\boldsymbol{\rho}^{k+1} = \mathrm{mat}\left(\arg\min_{\boldsymbol{\rho}} \left\{||\mathbf{y} - \mathbf{A}\boldsymbol{\rho}||_2^2 + \lambda/2||\boldsymbol{\rho} - \mathbf{z}^k + \mathbf{u}^k||_2^2\right\}\right), \\ &\mathbf{z}^{k+1} = \Pi_{\mathcal{C}}(\boldsymbol{\rho}^{k+1} + \mathbf{u}^k), \\ &\mathbf{u}^{k+1} = \mathbf{u}^k + (\boldsymbol{\rho}^{k+1} - \mathbf{z}^{k+1}). \end{aligned} \tag{29}$$

The $\boldsymbol{\rho}$ updating step can be completed by calculating its analytic solution,

$$\boldsymbol{\rho} = \left(\mathbf{A}^*\mathbf{A} + \lambda\mathbf{I}\right)^{-1}\left(\mathbf{A}^*\mathbf{y} + \lambda(\mathbf{z}^k - \mathbf{u}^k)\right). \tag{30}$$

The projection process in step 2 follows the explanation in Section 3.3 that exploits the shrink or truncated eigenvalue decomposition as formulated in (22) (23). In addition, we set rank $r = 2$ in the generation of true state $\boldsymbol{\rho}$ in (27), $\lambda = 1$ in (29). The reconstruction performances are evaluated by the error defined as

$$\text{error} = \frac{||\boldsymbol{\rho}^* - \hat{\boldsymbol{\rho}}||_2^2}{||\boldsymbol{\rho}^*||_2^2}, \tag{31}$$

where $\boldsymbol{\rho}^*$ and $\hat{\boldsymbol{\rho}}$ denote the true state and the estimate state, respectively. The error is calculated verses the increasing measurement rate $\eta = M/d^2$. Fig.1 depicts the reconstruction errors with increasing η. From Fig.1 one can observe that given the low rank information as priori knowledge, the number of measurements is dramatically reduced. Specifically the Dantzig using cvx performs better than the least square approach, and our simplified algorithm using ADMM has smaller errors comparing to the Dantzig given the same number of measurements M.

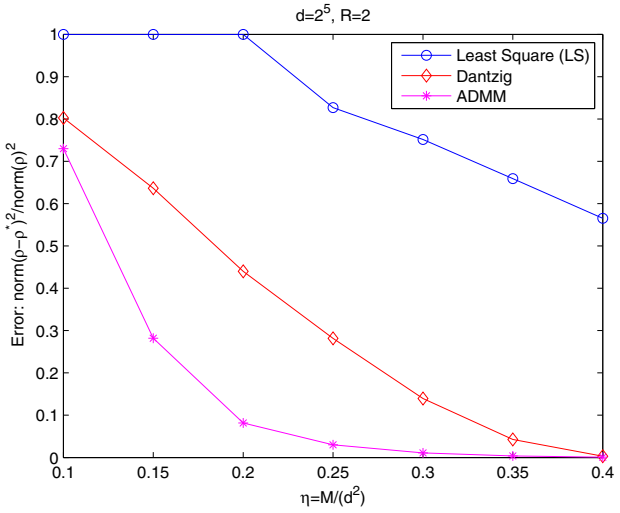

Fig. 1. The comparison of reconstruction performances of different algorithms, including the least square method in (2) using cvx toolbox, compressive quantum tomography solving Dantzig in (3) using cvx toolbox, and compressive quantum tomography solving (28) using ADMM.

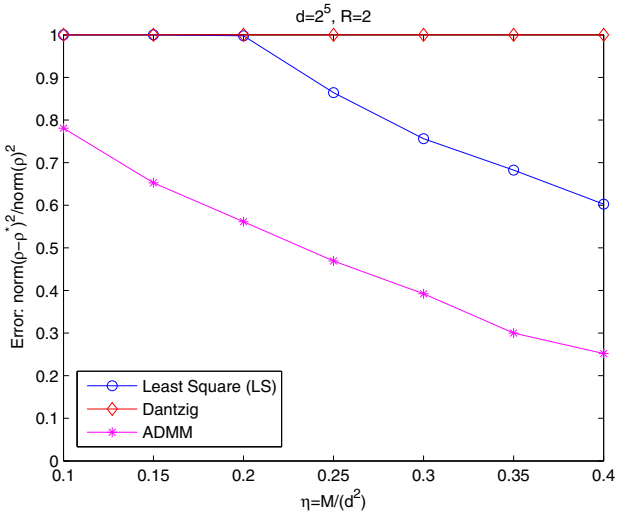

Fig. 2. The comparison of reconstruction performances with sparse outlier noises, including the least square using cvx toolbox, Dantzig using cvx toolbox, and compressive quantum tomography using ADMM.

In the second simulation we add the outlier noises in the density matrix. We set the measurements $\mathbf{y} = \mathcal{A}(\boldsymbol{\rho} + \mathbf{S}) + \mathbf{e}$ where $\mathbf{S} \in \mathcal{C}^{d\times d}$ has $(0.01d^2)$ nonzero values located uniform randomly with magnitudes satisfying Gaussian distribution $\mathcal{N}(0, 0.1||\boldsymbol{\rho}||_2)$. If the error is larger than 1, we record it as 1. The results are demonstrated in Fig. 2. Fig. 2 shows that the Dantzig approach fails under such scenario since the nuclear norm minimization is influenced significantly by the large outliers, however the proposed method may overcome this shortage and still lead to a recovery much better than the result of the LS method.

5. CONCLUSION

After reviewing several existing algorithms of the compressive quantum state tomography and the ADMM method, this paper proposed an alternating augmented Lagrangian method for quantum convex optimization problem aiming to recover pure or nearly pure state with sparse outlier noises. The algorithm updated the density matrix and the sparse estimate noises iteratively and finally obtained a reconstruction result efficiently. Simulations showed that the proposed algorithm achieves a better recovery accuracy comparing to the conventional least square and compressive Dantzig method with the same number of measurements. For the case of existence of sparse outlier noises, the proposed algorithm beat the Dantzig method due to the fact that the influence of outliers had been reduced. In the future, we will test our algorithm in the real experiment and upgrade our computation capability to deal with more qubits e.g., 8-10 qubits.

6. APPENDIX

Proposition 1. When the quantum state consisting of q qubits is the probabilistic combination of r pure states, then its density matrix $\boldsymbol{\rho}$ with size $d \times d$ has rank not larger than r, $d = 2^q$.

Proof: The proof is simple however it seldom appears in literatures and it lays the foundation of the compressive quantum tomography, so we give a proof here. Suppose the quantum state under consideration

$$\hat{\boldsymbol{\rho}} = \sum_{i=1}^{r} p_i |\psi_i\rangle\langle\psi_i|, \tag{32}$$

which means the quantum system may be found in state $|\psi_i\rangle$ with probability p_i, $i = 1, 2, \cdots, r$. If we concatenate the column vectors $|\psi_i\rangle$ as a matrix, then (32) is equivalent to

$$\begin{aligned} &\hat{\boldsymbol{\rho}} = \boldsymbol{\Psi}_{\mathbf{r}} \cdot \boldsymbol{\Psi}_{\mathbf{r}}{}^*, \text{where} \\ &\boldsymbol{\Psi}_{\mathbf{r}} = \left[\sqrt{p_1}|\psi_1\rangle, \sqrt{p_2}|\psi_2\rangle, \cdots, \sqrt{p_r}|\psi_r\rangle\right], \end{aligned} \tag{33}$$

$\boldsymbol{\Psi}_{\mathbf{r}}$ is of size $d \times r$. Thus the density matrix $\hat{\boldsymbol{\rho}}$ has rank at most r due to the rank property of multiplication of two matrices.

Definition 1. (**Rank RIP**). Recht et al. [2007], Liu [2011] The $\mathcal{A}$ satisfies the rank restricted isometry property (RIP) if for all $d \times d$ $\mathbf{X}$, we have

$$(1-\delta)||\mathbf{X}||_F \leq ||\mathcal{A}(\mathbf{X})||_2 \leq (1+\delta)||\mathbf{X}||_F \tag{34}$$

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

Proposition 2. When we formulate the measurement process as equations in (1), and the observable $\mathbf{O}_i$ are the tensor/Kronecker product of a series of complex and unitary elemental 2×2 Pauli matrices $\mathbf{P}_i$ chosen from the four possibilities randomly,

$$\begin{aligned} &\mathbf{I}_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \\ &\sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \end{aligned} \tag{35}$$

Then the sampling operator $\mathcal{A}$ satisfies the rank RIP and we are able to recover the rank-r density matrix $\boldsymbol{\rho}$ by using number of measurements $m \geq c \cdot rd \log^6 d$ for some absolute constant c with high probability.

Proof: The details of the proof can be found in Gross [2011], Liu [2011].

REFERENCES

Boyd, S., Parikh, N., Chu, E., Peleato, B., and Eckstein, J. (2011). Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3(1), 1122.

Boyd, S. and Vandenberghe, L. (2004). *Convex Optimization*. Cambridge Univ. Press, Cambridge, U.K.

Candes, E. and Plan, Y. (2011). Tight oracle inequalities for low-rank matrix recovery from a minimal number of noisy random measurements. 57(4), 2342–2359.

Donoho, D.L. (2006). Compressed sensing. 52, 1289–1306.

E. J. Candés, X. Li, Y.M. and Wright, J. (2011). Robust principal component analysis? *J. ACM*, 58(3), 1–37.

Gross, D. (2011). Recovering low-rank matrices from few coefficients in any basis. 57(3), 1548–1566.

Gross, D., Liu, Y., Flammia, S.T., Becker, S., and Eisert, J. (2010). Quantum state tomography via compressed sensing. *Phys. Rev. Lett.*, 105(15), 150401150404.

J. Wright, A. Ganesh, K.M. and Ma, Y. (2013). Compressive principal component pursuit. *A Journal of the IMA*, (2), 3268.

Kyrillidis, A. and Cevher, V. (2012). Matrix alps: Accelerated low rank and sparse matrix reconstruction. In *IEEE Statistical Signal Processing Workshop (SSP) 2012*, 185–188.

Lin, Z., Liu, R., and Su, Z. (2011). Linearized alternating direction method with adaptive penalty for low-rank representation. In *NIPS*, 612–620.

Liu, W.T., Zhang, T., Liu, J.Y., Chen, P.X., and Yuan, J.M. (2012). Experimental quantum state tomography via compressed sampling. *Physical Review Letters*, 108, 170403.

Liu, Y. (2011). Universal low-rank matrix recovery from pauli measurements. *NIPS*, 1638–1646.

Lvovsky, A. and Raymer, M. (2009). Continuous-variable optical quantum-state tomography. *Reviews of Mordern Physics*, 81, 299–328.

Recht, B., Fazel, M., and Parillo, P. (2007). Guaranteed minimum rank solution of matrix equations via nuclear norm minimization. *SIAM Rev.*, 52, 471501.

Smith, A., Riofro, C., Anderson, B., Martinez, H., Deutsch, I., and Jessen, P. (2013). Quantum state tomography by continuous measurement and compressed sensing. *Phys. Rev. A*, 030102(87).

Sturm, J.F. (1999). Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones. *Optim. Methods Softw.*, 11-12, 625653.

Yuan, X.M. and Yang, J. (2009). Sparse and low-rank matrix decomposition via alternating direction methods. *Pacific Journal of Optimization*.

Zhou, Z., Li, X., Wright, J., Candes, E., and Ma, Y. (2010). Stable principal component pursuit. In *IEEE International Symposium on Information Theory Proceedings (ISIT), 2010*, 1518–1522.

Zyczkowski, K., Penson, K.A., Nechita, I., and Collins, B. (2011). Generating random density matrices. *J. Math. Phys*, 52, 062201.