An Efficient and Fast Quantum State Estimator With Sparse Disturbance

Jiaojiao Zhang, Shuang Cong, *Senior Member, IEEE*, Qing Ling, *Senior Member, IEEE*, and Kezhi Li

***Abstract*—A pure or nearly pure quantum state can be described as a low-rank density matrix, which is a positive semidefinite and unit-trace Hermitian. We consider the problem of recovering such a low-rank density matrix contaminated by sparse components, from a small set of linear measurements. This quantum state estimation task can be formulated as a robust principal component analysis (RPCA) problem subject to positive semidefinite and unit-trace Hermitian constraints. We propose an efficient and fast inexact alternating direction method of multipliers (I-ADMM), in which the subproblems are solved inexactly and hence have closed-form solutions. We prove global convergence of the proposed I-ADMM, and the theoretical result provides a guideline for parameter setting. Numerical experiments show that the proposed I-ADMM can recover state density matrices of 5 qubits on a laptop in 0.69 s, with 6×10^{-4} accuracy (99.38% fidelity) using 30% compressive sensing measurements, which outperforms existing algorithms.**

***Index Terms*—Alternating direction method of multipliers (ADMM), quantum state estimation (QSE), robust principal component analysis (RPCA).**

I. Introduction

QUANTUM state estimation (QSE), also known as quantum state tomography, is an important technique for quantum information processing, and the estimation accuracy significantly affects subsequent applications, such as quantum computing and quantum communication [1]–[5]. The task of QSE is to recover the density matrix of a quantum state given a series of measurements obtained from physical experiments. An n-qubit quantum state can be fully represented by a density matrix ρ, which is a Hermitian matrix in a $d = 2^n$ dimensional Hilbert space [6], [7]. Since ρ has $\mathcal{O}(d^2)$ unknown elements, its recovery usually requires $\mathcal{O}(d^2)$ linear measurements.

Obtaining a quantum measurement, which is a linear combination of the elements in ρ, is costly. One has to prepare a large number of identical copies of the quantum state, and measure the probability of collapsing through calculating the mean value of observations. Fortunately, the development of compressive sensing provides us a powerful tool to recover structured (sparse, low-rank, etc.) signals from fewer measurements [8]–[10]. In the context of QSE, people are usually interested in a pure or nearly pure state in a practical quantum system, meaning that the corresponding density matrix ρ has rank $r \ll d$ [11], [12]. According to the compressive sensing theory, $\mathcal{O}(rd \log d)$ linear measurements, instead of $\mathcal{O}(d^2)$, are suffice for exact recovery of ρ. This observation motivates recovering the low-rank density matrix ρ through minimizing its nuclear norm, subject to quantum state constraints [13].

The process of measuring a quantum state inevitably involves disturbance, which largely biases the estimation from the optimal one [14]–[16]. The disturbance can appear either in the quantum measurements, or in the quantum state itself. For the former case, the measurement disturbance is often assumed to be Guassian noise, and can be filtered through least-squares techniques [1], [17], [18]. The low-rankness of ρ was used and yielded the matrix least absolute shrinkage and selection operator (LASSO) model [13]. For the latter case, the state disturbance introduced sparse outliers on the elements of ρ and brought difficulties to exact recovery of the true quantum state [19].

In this paper, we focus on the problem of recovering a $d \times d$ low-rank density matrix ρ contaminated by a sparse $d \times d$ matrix S from m linear measurements, where $m = \mathcal{O}(rd \log d) \ll d^2$. Mathematically, it can be formulated as a robust principal component analysis (RPCA) problem that minimizes the nuclear norm of ρ plus the ℓ_1 norm of S, subject to linear measurement constraints [20], [21]. Nevertheless, different to the classical RPCA model, here the optimization variable ρ is complex and subject to quantum state constraints—namely, ρ is a positive semidefinite and unit-trace Hermitian matrix. This additional constraints make solving the QSE problem a more challenging task.

Without the quantum state constraints, the RPCA problem can be solved with the singular value thresholding method [22], [23] or a variant of the augmented Lagrangian method [24]. For the QSE problem with sparse quantum state disturbance, [19] first ignored the quantum state constraints and solved the relaxed RPCA problem with the alternating direction method of multipliers (ADMM), and then projected the solution to the feasible set. This approach cannot guarantee the convergence to the optimal solution. Meanwhile, it

Manuscript received September 19, 2017; revised February 2, 2018 and April 3, 2018; accepted April 15, 2018. This work was supported by the National Natural Science Foundation of China under Grant 61573330 and Grant 61720106009. This paper was recommended by Associate Editor W. X. Zheng. *(Corresponding author: Shuang Cong.)*

J. Zhang and S. Cong are with the Department of Automation, University of Science and Technology of China, Hefei 230027, China (e-mail: scong@ustc.edu.cn).

Q. Ling is with the School of Data and Computer Science, Sun Yat-sen University, Guangzhou 510006, China.

K. Li is with the Department of Electronic and Electrical Engineering, Imperial College London, London SW5 7AZ, U.K. (e-mail: kezhi.li@imperial.ac.uk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCYB.2018.2828498

involves calculating the pseudo inverse of an $m \times d^2$ matrix, whose complexity is $\mathcal{O}(d^6)$. Recent works [25] and [26] considered the special case of $r = 1$ (namely, ρ is rank-one), which significantly simplified the quantum state constraints. Specifically, [25] combined the fixed-point technique with ADMM and developed an algorithm called as FP-ADMM, which improved the accuracy and reduced the computational complexity to $\mathcal{O}(md^4)$ by avoiding the pseudo inverse operation. Further, [26] applied the iterative shrinkage-thresholding operator in solving the subproblems of ADMM, and the resultant IST-ADMM algorithm reduced the computational complexity to $\mathcal{O}(md^2)$. In summary, to solve the QSE problem with sparse quantum state disturbance, the existing algorithms are heuristic through ignoring or simplifying the quantum state constraints, and often with high computational complexity. This motivates us to develop a computationally efficient and provably convergent algorithm, which explicitly considers the quantum state constraints and is able to exactly recover the quantum state from the measurements contaminated by sparse disturbance.

We propose a novel algorithm that exactly solves the QSE with disturbance in this paper. Using the technique of ADMM, we obtain two subproblems, one is a nuclear norm optimization problem subject to quantum state constraints for estimating the quantum state ρ, and the other is an ℓ_1 norm minimization problem for estimating the sparse disturbance matrix S. Since both subproblems do not have closed-form solutions and are intractable when d is large, we instead solve their approximations and thus have the computational complexity of $\mathcal{O}(md^2)$. The resultant algorithm is called as inexact ADMM, or I-ADMM in short. We theoretically establish the global convergence of I-ADMM to the optimal solution. The theoretical result also provides a guideline for the parameter setting. Numerical experiments on a 5-qubit QSE with disturbance demonstrate that the proposed I-ADMM, compared with the existing matrix LASSO [13], ADMM [19], FP-ADMM [25], and IST-ADMM [26] algorithms, are more efficient and faster. It is able to recover quantum states accurately using fewer measurements and less iterations.

The rest of this paper is organized as follows. Section II is the description of the QSE with sparse quantum state disturbance. Section III is the I-ADMM proposed for the QSE, and Section IV is the proof of the global convergence of the I-ADMM proposed. Numerical experiments with performance comparisons of different methods are given in Section V, and Section VI summarizes this paper.

Notations: For a vector, $\|\cdot\|_2$ denotes the ℓ_2 norm. For a matrix, $\|\cdot\|_*$, $\|\cdot\|_1$, $\|\cdot\|_F$ and $\mathrm{tr}(\cdot)$ denote its nuclear, ℓ_1, Frobenius and trace norms. Superscript "H" denotes the conjugate transpose of a complex matrix. For a linear operator $\mathcal{A} : \mathbb{C}^{d\times d} \to \mathbb{C}^m$, $\mathcal{A}^{\mathrm{H}} : \mathbb{C}^m \to \mathbb{C}^{d\times d}$ denotes its Hermitian transpose. $\langle\cdot,\cdot\rangle$ denotes the inner product such that $\langle Z, Z'\rangle = \mathrm{tr}(Z^{\mathrm{H}}Z')$; $G \succeq 0$ means that G is positive semidefinite. We define $\langle Z, Z'\rangle_G = \langle Z, GZ'\rangle$ and the induced norm $\|Z\|_G = \sqrt{\langle Z, Z\rangle_G}$. $\mathrm{vec}(\cdot)$ vectorizes a matrix to a column vector. ∂f denotes the subdifferential of a function f.

II. QSE With Sparse Quantum State Disturbance

Consider a quantum system of n qubits, whose density matrix is given by $\rho \in \mathbb{C}^{d\times d}$ with $d = 2^n$. Quantum state constraints require that the density matrix is a positive semidefinite and unit-trace Hermitian; namely, $\rho^{\mathrm{H}} = \rho$, $\rho \succeq 0$, and $\mathrm{tr}(\rho) = 1$. In addition, for a pure or nearly pure quantum system, ρ is low-rank, meaning that its rank $r \ll d$. The quantum mechanics theory shows that, to observe ρ, one can use d^2 Hermitian operators $\{\omega_i \in \mathbb{C}^{d\times d}, i = 1, \ldots, d^2\}$. If one uses ω_i for instance, the measurement is given by

$$b_i = \mathrm{tr}\big(\omega_i^{\mathrm{H}}\rho\big) = \mathrm{vec}(\omega_i)^{\mathrm{H}}\mathrm{vec}(\rho). \tag{1}$$

Note that in a physical system, such a measurement is obtained by preparing a large number of identical copies of the quantum state and calculating the mean value of observations [6].

The Hermitian operators $\{\omega_i\}$ can be chosen as a series of Pauli bases. Define four complex and unitary 2×2 Pauli matrices

$$\sigma_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},\ \sigma_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},\ \sigma_3 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$$
$$\sigma_4 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

A Pauli basis of an n-qubit quantum system is generated by randomly selecting n Pauli matrices—here we allow replacement—and calculating their Kronecker product followed by normalization.

The facts that the quantum measurement is costly and that the density matrix ρ is low-rank allow us to apply compressive sensing to QSE. Randomly choosing $m = \mathcal{O}(rd\log d) \ll d^2$ Hermitian operators out of the total d^2 and defining the sampling rate as

$$\eta = m/d^2 \tag{2}$$

we obtain m measurements according to (1). For simplicity, write the measurement function in a matrix form $b = \mathcal{A}(\rho)$, where $b \in \mathbb{R}^m$ is the measurement vector that presents the probability of the quantum state collapses onto each basis, and $\mathcal{A} : \mathbb{C}^{d\times d} \to \mathbb{C}^m$ is a linear operator that denotes the measurement process. To be more specific, define a matrix $A \in \mathbb{C}^{m\times d^2}$ where each row corresponds to a Hermitian operator ω_i and is given by $\mathrm{vec}(\omega_i)^{\mathrm{H}}$. Thus, $\mathcal{A}(\rho) \coloneqq A\mathrm{vec}(\rho)$.

In this paper, we consider the situation of a sparse disturbance in ρ, which is a case in the physical measurement process. Define a sparse matrix $S \in \mathbb{R}^{d\times d}$ whose nonzero elements are much less than d^2. The disturbed measurements can be written as

$$b = \mathcal{A}(\rho + S). \tag{3}$$

Therefore, to estimate the low-rank density matrix ρ and the sparse disturbance matrix S from the quantum measurements b, the task of such a QSE design needs to solve the following convex optimization problem

$$\begin{aligned} \min\quad & \|\rho\|_* + \gamma\|S\|_1 \\ \text{s.t.}\quad & \mathcal{A}(\rho + S) = b, \quad \rho^{\mathrm{H}} = \rho, \quad \rho \succeq 0, \quad \mathrm{tr}(\rho) = 1 \end{aligned} \tag{4}$$

where γ is a positive weight factor.

Observe that (4) differs from the classical RPCA model in the quantum state constraints $\rho^{\mathrm{H}} = \rho, \rho \succeq 0, \mathrm{tr}(\rho) = 1$. For simplicity, we define a convex set $C := \{\rho \in \mathbb{C}^{d\times d} | \rho^{\mathrm{H}} = \rho, \rho \succeq 0, \mathrm{tr}(\rho) = 1\}$, introduce an indicator function $I_C(\rho)$ whose value is 0 when $\rho^{\mathrm{H}} = \rho, \rho \succeq 0, \mathrm{tr}(\rho) = 1$ and ∞, otherwise, and rewrite (4) as

$$\begin{aligned} \min \quad & \|\rho\|_* + I_C(\rho) + \gamma\|S\|_1 \\ \text{s.t.} \quad & \mathcal{A}(\rho + S) = b. \end{aligned} \tag{5}$$

III. ADMM AND I-ADMM

In order to design a convergent algorithm for the QSE with disturbance (5) in this paper, we propose an inexact version of ADMM, called as I-ADMM, to solve (5) efficiently. We first introduce ADMM in Section III-A and then propose I-ADMM in Section III-B.

A. ADMM

To solve an optimization problem with a separable cost function and linear constraints, the basic idea of ADMM is to split the optimization variables into two blocks, minimize the corresponding augmented Lagrangian function with respect to the two blocks in an alternating direction manner, and then update the dual variable with dual gradient ascent [27]. It has been shown that ADMM is a powerful tool to solve structured optimization problems [28]–[30]. Its sublinear convergence rate was established in [31], and linear convergence rate was established in [32]–[34].

Before applying ADMM to solve the QSE with disturbance (5), we write its augmented Lagrangian as

$$\begin{aligned} L(\rho, S, y, \alpha) = \|\rho\|_* + I_C(\rho) + \gamma\|S\|_1 - \langle y, \mathcal{A}(\rho + S) - b\rangle \\ + \frac{\alpha}{2}\|\mathcal{A}(\rho + S) - b\|_2^2 \end{aligned} \tag{6}$$

where $y \in \mathbb{R}^m$ is the Lagrange multiplier, and α is a positive penalty parameter. We enforce the Lagrange multiplier y to be a real vector for the following reason. During the optimization process, we shall guarantee that every iteration of ρ stays in the set C such that $\rho^{\mathrm{H}} = \rho$. Thus, we know that $\mathcal{A}(\rho)$ is real, and so as $\mathcal{A}(\rho + S) - b$.

Following the ADMM routine, the resultant algorithm is:

$$\begin{cases} \rho^{k+1} = \arg\min\limits_{\rho} \left\{ \|\rho\|_* + I_C(\rho) \right. \\ \qquad \left. + \dfrac{\alpha}{2}\left\| \mathcal{A}\left(\rho + S^k\right) - b - \dfrac{y^k}{\alpha} \right\|_2^2 \right\} & \text{(7a)} \\ S^{k+1} = \arg\min\limits_{S} \left\{ \gamma\|S\|_1 + \dfrac{\alpha}{2}\left\| \mathcal{A}\left(\rho^{k+1} + S\right) \right. \right. \\ \qquad \left. \left. - b - \dfrac{y^k}{\alpha} \right\|_2^2 \right\} & \text{(7b)} \\ y^{k+1} = y^k - \kappa\alpha\left(\mathcal{A}\left(\rho^{k+1} + S^{k+1}\right) - b\right) & \text{(7c)} \end{cases}$$

where κ is a positive scaling factor for tuning the stepsize of dual gradient ascent. Though $\kappa = 1$ is a normal choice in ADMM, tuning it within the range of $(0, (\sqrt{5}+1)/2)$ can often improve the convergence speed [35].

The main challenge of implementing (7a) is the high computational complexity of solving the subproblems (7a) and (7b). To be specific, (7a) is essentially the minimization of a nonsmooth nuclear norm plus a least-squares term subject to the quantum state constraints, while (7b) minimizes the summation of a nonsmooth ℓ_1 norm and a least-squares term. Both subproblems have no closed-form solutions, except that $\mathcal{A}^{\mathrm{H}}\mathcal{A}$ has a special structure such as diagonal, which is not the case in this paper.

B. I-ADMM

The main idea of I-ADMM is to run one proximal gradient step to approximately solve each of the subproblems (7a) and (7b). The gradient steps work on the smooth least-squares terms, while the proximal steps work on the nuclear norm, the ℓ_1 norm and the indicator function. In the I-ADMM algorithm proposed, the estimated density matrix ρ^{k+1}, sparse disturbance S^{k+1}, and Lagrange multiplier y^{k+1} are, respectively,

$$\begin{cases} \rho^{k+1} = \arg\min\limits_{\rho} \ \|\rho\|_* + I_C(\rho) + \dfrac{\alpha}{2\tau_1}\left\|\rho - \tilde{\rho}^k\right\|_F^2 & \text{(8a)} \\ S^{k+1} = \arg\min\limits_{S} \ \gamma\|S\|_1 + \dfrac{\alpha}{2\tau_2}\left\|S - \tilde{S}^k\right\|_F^2 & \text{(8b)} \\ y^{k+1} = y^k - \kappa\alpha\left(\mathcal{A}\left(\rho^{k+1} + S^{k+1}\right) - b\right). & \text{(8c)} \end{cases}$$

Here, we define the proximal density matrix $\tilde{\rho}^k$ and proximal sparse disturbance $\tilde{S}^k$ as

$$\tilde{\rho}^k := \rho^k - \tau_1\mathcal{A}^H\left(\mathcal{A}\left(\rho^k + S^k\right) - b - \frac{y^k}{\alpha}\right)$$

and

$$\tilde{S}^k := S^k - \tau_2\mathcal{A}^H\left(\mathcal{A}\left(\rho^{k+1} + S^k\right) - b - \frac{y^k}{\alpha}\right).$$

The two positive constants τ_1 and τ_2 are the proximal gradient stepsizes. Observe that $\tilde{\rho}^k$ and $\tilde{S}^k$ come from running gradient steps on the least squares terms in (7a) and (7b), from the current ρ^k and S^k, respectively. Next, we show that the proximal mappings in (8a) and (8b) have closed-form solutions by means of defining Subproblems 1 and 2.

Subproblem 1: For (8a), considering the quantum state constraints $\rho = \rho^{\mathrm{H}}$, $\rho \succeq 0$ and $\mathrm{tr}(\rho) = 1$, we have that $\|\rho\|_* = \mathrm{tr}(\rho) = 1$. Thus, by rewriting the indicator function $I_C(\rho)$, (8a) reduces to solving the following semidefinite programming problem:

$$\begin{aligned} \min \quad & \left\|\rho - \tilde{\rho}^k\right\|_F^2 \\ \text{s.t.} \quad & \rho = \rho^{\mathrm{H}}, \rho \succeq 0, \mathrm{tr}(\rho) = 1 \end{aligned} \tag{9}$$

or equivalently, to solving

$$\begin{aligned} \min \quad & \left\|\rho - \frac{\tilde{\rho}^k + (\tilde{\rho}^k)^{\mathrm{H}}}{2}\right\|_F^2 \\ \text{s.t.} \quad & \rho \succeq 0, \mathrm{tr}(\rho) = 1. \end{aligned} \tag{10}$$

Note that (10) has a closed-form solution [36]. Operate the eigenvalue decomposition to rewrite $(\tilde{\rho}^k + (\tilde{\rho}^k)^{\mathrm{H}})/2$ as $V\mathrm{diag}\{a_i\}V^{\mathrm{H}}$, where $V \in \mathbb{C}^{d\times d}$ is a unitary matrix and $\{a_i, i = 1, \ldots, d\}$ are eigenvalues of $(\tilde{\rho}^k + (\tilde{\rho}^k)^{\mathrm{H}})/2$. Here, we sort the eigenvalues as $a_1 \geq a_2 \geq \cdots \geq a_d$. The optimal solution of the estimated density matrix to (10) is

$$\rho^{k+1} = V\ \mathrm{diag}\{x_i\}V^{\mathrm{H}} \tag{11}$$

where $\{x_i, i = 1, \ldots, d\}$ are the eigenvalues of ρ^{k+1} and solved from

$$\begin{aligned} \min \quad & \frac{1}{2}\sum_{i=1}^{d}(x_i - a_i)^2 \\ \text{s.t.} \quad & \sum_{i=1}^{d} x_i = 1, x_i \geq 0, \forall i. \end{aligned} \tag{12}$$

Write the partial Lagrangian for (12) as

$$\begin{aligned} L(\{x_i\}, \beta) = & \frac{1}{2}\sum_{i=1}^{d}(x_i - a_i)^2 + \beta\left(\sum_{i=1}^{d} x_i - 1\right) \\ \text{s.t.} \quad & x_i \geq 0, \quad \forall i \end{aligned} \tag{13}$$

where $\beta \in \mathbb{R}$ is the Lagrange multiplier. According to the convex optimization theory, if β is the optimal Lagrange multiplier, minimizing $L(\{x_i\}, \beta)$ over $\{x_i\}$ from

$$\begin{aligned} \min \quad & \frac{1}{2}\sum_{i=1}^{d}(x_i - a_i + \beta)^2 \\ \text{s.t.} \quad & x_i \geq 0, \quad \forall i \end{aligned} \tag{14}$$

yields the optimal primal solution of the eigenvalues of ρ^{k+1} as

$$x_i = \max\{a_i - \beta, 0\}, \forall i. \tag{15}$$

Since for the optimal β, we must have $\sum_{i=1}^{d} x_i = 1$. Using the relation (15), it is equivalent to $\sum_{i=1}^{d}\max\{a_i - \beta, 0\} = 1$. This equality suggests us to try $\beta = a_i$ at every $i = 1, \ldots, d$ and decide which interval the optimal β must be within. Suppose that β is within $[a_{t+1}, a_t]$, we know that $x_i = a_i - \beta, \forall i \leq t$ and $x_i = 0, \forall i \geq t+1$. Thus, the optimal β is solved from $\sum_{i=1}^{t}(a_i - \beta) = 1$; namely, $\beta = (\sum_{i=1}^{t} a_i - 1)/t$. Using this optimal β, we calculate $\{x_i\}$ from (15) and consequently ρ^{k+1} from (11). The main computational complexity comes from the calculation of the proximal density matrix $\tilde{\rho}^k$, which is in the order of $\mathcal{O}(md^2)$ and the same as that in IST-ADMM [26].

Subproblem 2: For (8b), according to [37] it has a well-known soft thresholding solution

$$S^{k+1} = \mathrm{shrink}_{\gamma\tau_2/\alpha}\left(S^k - \tilde{S}^k\right) \tag{16}$$

where $\mathrm{shrink}_{\gamma\tau_2/\alpha}$ is an element-wise shrinkage operator. Given a scalar s, $\mathrm{shrink}_{\gamma\tau_2/\alpha}(s) = \max\{|s - \gamma\tau_2/\alpha|, 0\}\mathrm{sign}(s - \gamma\tau_2/\alpha)$.

The proposed I-ADMM is outlined in Algorithm 1.

Algorithm 1 I-ADMM Algorithm

Require: Initialize variables to $\rho^0 = 0, S^0 = 0, y^0 = 0$ and set algorithm parameters $\tau_1, \tau_2, \alpha, \kappa > 0$.

1: **For** $k = 1, 2, \cdots$ **Do**
2: Calculate $\tilde{\rho}^k = \rho^k - \tau_1\mathcal{A}^H(\mathcal{A}(\rho^k + S^k) - b - y^k/\alpha)$.
3: Run eigenvalue decomposition on $(\tilde{\rho}^k + (\tilde{\rho}^k)^{\mathrm{H}})/2$ to obtain $V\ diag\{a_i\}V^{\mathrm{H}}$, where $a_1 \geq a_2 \geq \cdots \geq a_d$.
4: Set β as $a_1, a_2, \cdots$ to find t so that $\sum_{i=1}^{t}\max\{a_i - \beta, 0\} < 1$ and $\sum_{i=1}^{t+1}\max\{a_i - \beta, 0\} > 1$.
5: Find $\beta = (1/t)\sum_{i=1}^{t} a_i$ and $x_i = \max\{a_i - \beta, 0\}, \forall i$.
6: Update $\rho^{k+1} = Vdiag\{x_i\}V^{\mathrm{H}}$.
7: Calculate $\tilde{S}^k = \tau_2\mathcal{A}^H(\mathcal{A}(\rho^{k+1} + S^k) - b - y^k/\alpha)$.
8: Update $S^{k+1} = shrink_{\gamma\tau_2/\alpha}(S^k - \tilde{S}^k)$.
9: Update $y^{k+1} = y^k - \kappa\alpha(\mathcal{A}(\rho^{k+1} + S^{k+1}) - b)$.
10: **End For**

IV. Main Result and Convergence Analysis

In this section, the QSE with disturbance is designed by means of the convergent I-ADMM algorithm proposed in Section III, and the global convergence of the algorithm is proved. Several inexact versions of ADMM have been proved to converge to optimal solutions under certain conditions. For example, [35] designed an inexact ADMM through linearizing one of the two subproblems. The two subproblems were inexactly solved in [38], and the convergence was still guaranteed. Our method proposed in this paper is novel in two aspects. First, [35] and [38] considered inexactly to solving the unconstrained subproblems, while one of the subproblems we inexactly solved is essentially constrained for practical QSE. Second, we introduce a scaling factor κ to tune the stepsize of dual gradient ascent, and our method provides a guideline of how to set κ in order to guarantee the convergence. Our main result is stated in Theorem 1.

Theorem 1: Suppose that the positive I-ADMM parameters τ_1, τ_2, and κ satisfy $\tau_1\lambda_{\max} < 1$ and $\tau_2\lambda_{\max} + \kappa < 2$, where $\lambda_{\max}$ denotes the maximum eigenvalue of $\mathcal{A}^{\mathrm{H}}\mathcal{A}$. For any fixed stepsize $\alpha > 0$, the sequence (ρ^k, S^k, y^k) generated by the I-ADMM algorithm from any starting point (ρ^0, S^0, y^0) converges to (ρ^*, S^*, y^*), which is an optimal primal-dual solution of the QSE with disturbance (5).

Proof: See the Appendix. ■

Theorem 1 suggests that as long as the proximal stepsize τ_1 and τ_2 and the scaling factor κ are properly chosen, the proposed I-ADMM algorithm for the QSE with disturbance (5) guarantees global convergent solution for any positive penalty factor α. It provides a guideline of setting the algorithm parameters τ_1, τ_2, and κ, which is related to the maximum eigenvalue of $\mathcal{A}^{\mathrm{H}}\mathcal{A}$, denoted as $\lambda_{\max}$. The inequality $\tau_1\lambda_{\max} < 1$ suggests that τ_1 must be smaller than $1/\lambda_{\max}$. On the other hand, the two parameters τ_2 and κ are entangled in the inequality $\tau_2\lambda_{\max} + \kappa < 2$. For example, choosing κ within $(0, 1)$ leads to a small dual stepsize $\kappa\alpha < \alpha$, and allows τ_2 to be larger than $1/\lambda_{\max}$.

Remark 1: When the linear operator $\mathcal{A}$ is constructed from the Pauli bases as shown in Section II, we have $\lambda_{\max} = 1$. According to the theoretical results given by Theorem 1, the

restrictions on the parameters τ_1, τ_2, and κ become $\tau_1 < 1$ and $\tau_2 + \kappa < 2$ in this case.

V. Numerical Experiments

In this section, we provide numerical experiments to demonstrate the sampling efficiency and computational superiority of the proposed I-ADMM in recovering quantum states contaminated by the sparse disturbance. In the experiments, a measurement is generated from $b = \mathcal{A}(\ddot{\rho} + \ddot{S})$, where $\ddot{\rho}$ represents the true quantum state to be recovered, and $\ddot{S}$ is the sparse disturbance. We generate $\ddot{\rho}$ by

$$\ddot{\rho} = \frac{\psi_r {\psi_r}^H}{\mathrm{tr}\left(\psi_r {\psi_r}^H\right)} \tag{17}$$

where ψ_r is a complex $d \times r$ Wishart matrix with independent identically distributed complex random Gaussian entries [39].

In the experiments, we set $d = 2^n$ with $n = 5$ and $r = 2$. The disturbance matrix $\ddot{S} \in \mathbb{R}^{d\times d}$ has $d^2/10$ nonzero entries located uniform randomly, with values satisfying Gaussian distribution $\mathcal{N}(0, \|\rho\|_F/100)$. The linear operator $\mathcal{A}$ is constructed from the Pauli bases as shown in Section II such that $\mathcal{A}\mathcal{A}^{\mathrm{H}} = I$. In consequence, we have $\lambda_{\max}(\mathcal{A}^{\mathrm{H}}\mathcal{A}) = 1$, which is useful in setting the parameters for I-ADMM. In the optimization problem (8a), the weight is $\gamma = 1/\sqrt{d}$ as suggested by [40].

The performances of the estimated ρ are evaluated by three typical criteria for quantum states. The first one is the normalized distance $D(\rho, \ddot{\rho})$ between the estimate ρ and the true density matrix $\ddot{\rho}$, defined by

$$D(\rho, \ddot{\rho}) = \|\rho - \ddot{\rho}\|_F^2 / \|\ddot{\rho}\|_F^2. \tag{18}$$

The second one is the normalized distance $D(\rho, \rho^*)$ between the estimate ρ and the optimal solution ρ^*, defined by

$$D\left(\rho, \rho^*\right) = \left\|\rho - \rho^*\right\|_F^2 / \left\|\rho^*\right\|_F^2. \tag{19}$$

The third one is the fidelity defined by

$$\text{fidelity} = \mathrm{tr}\left(\sqrt{\sqrt{\ddot{\rho}}\rho\sqrt{\ddot{\rho}}}\right). \tag{20}$$

Note that the fidelity is within $[0, 1]$, and its value is 1 when the two states are identical [41], [42].

We compare I-ADMM proposed in this paper with three algorithms matrix LASSO [13], ADMM [19], and IST-ADMM [26]. Matrix LASSO is a nuclear-norm regularized least squares model with quantum constraints [13]. For the sake of computational efficiency, we solve the matrix LASSO model with ADMM where the ρ-subproblem is solved inexactly just as what we have done in the I-ADMM update (8a). Different from our model (5) that is solved by I-ADMM, the matrix LASSO model does not consider the impact of the sparse disturbance S on the density matrix ρ. As we have discussed in Section I, ADMM in [19] could not guarantee convergence to the optimal solution, while IST-ADMM in [26] was only able to handle the case of $r = 1$. In the case of $r = 2$ as we set in the experiments, IST-ADMM cannot guarantee the positive semi-definiteness of the solution. Note that another existing algorithm FP-ADMM in [25], which we do not compare below, has the same solution accuracy as IST-ADMM but much higher computational complexity. All the experiments run in MATLAB R2010b on a laptop with 2.5 GHz Intel Core 2 i5-3210M CPU and 6 GB memory.

In the experiments, $\tau_1 = 0.99$, while τ_2 and κ vary to demonstrate the impact on the performance of I-ADMM. The penalty parameter $\alpha = 8$. The parameters of matrix LASSO, ADMM, and IST-ADMM are hand-tuned to the best. All the algorithms are terminated if

$$\|\mathcal{A}\left(\rho^k + S^k\right) - b\|_2 / \|b\|_2 < 10^{-7} \quad \text{or} \quad k > k_{\max}. \tag{21}$$

A. Impact of the Sampling Rate η on I-ADMM, IST-ADMM, ADMM, and Matrix LASSO

The sampling rate η defined in (2) is an important factor to quantify the measurement complexity in QSE. A low sampling rate η means that the number of measurements is much less than d^2, the size of ρ. However, the lower the sampling rate η, the more difficult to accurately recover the density matrix ρ.

We compare the proposed I-ADMM ($\kappa = 1.1$ and $\tau_2 = 0.899$) with IST-ADMM, ADMM, and matrix LASSO, and run them for the same number of iterations $k_{\max} = 20$ under different sampling rates. The sampling rate increases from $\eta = 0.05$ to $\eta = 0.5$ with the incremental step 0.05. Fig. 1 depicts $D(\rho, \ddot{\rho})$, the normalized distance between the estimate ρ and the true density matrix $\ddot{\rho}$, with different η. Observe that I-ADMM and IST-ADMM both require much smaller sampling rates to achieve a target accuracy than matrix LASSO and ADMM. The reasons are that the sparse disturbance S cannot be eliminated by the matrix LASSO model [13], and that ADMM in [19] handles the objective function and constraints of (5) separately, which results in loss of optimality. Among the other two algorithms, I-ADMM outperforms IST-ADMM. To be more specific, when the sampling rate $\eta = 0.2$, the value of $D(\rho, \ddot{\rho})$ of I-ADMM is 0.1901 [the fidelity defined by (20) is 83.15%], and those of IST-ADMM, ADMM, and matrix LASSO are 0.2731 (75.18%), 0.6246 (54.17%), and 0.5632 (58.23%), respectively. When the sampling rate $\eta = 0.3$, the value of $D(\rho, \ddot{\rho})$ of I-ADMM decreases to 0.0019 and the fidelity is 99.66%. The value of $D(\rho, \ddot{\rho})$ (fidelity) is 0.0036 (93.80%) for IST-ADMM, 0.43 (67.65%) for ADMM, and 0.3651 (69.13%) for matrix LASSO. Therefore, this experiment demonstrates the efficiency of the proposed I-ADMM in recovering quantum states in terms of the sampling complexity.

B. Convergence Properties of I-ADMM, IST-ADMM, ADMM, and Matrix LASSO

In order to verify the convergence performance, we compare the proposed I-ADMM with IST-ADMM, ADMM, and matrix LASSO, in terms of the normalized distance $D(\rho, \rho^*)$ between the estimate ρ and the optimal solution ρ^*, as well as the normalized distance $D(\rho, \ddot{\rho})$ between the estimate ρ and the true density matrix $\ddot{\rho}$. We set the sampling rate $\eta = 0.3$ as

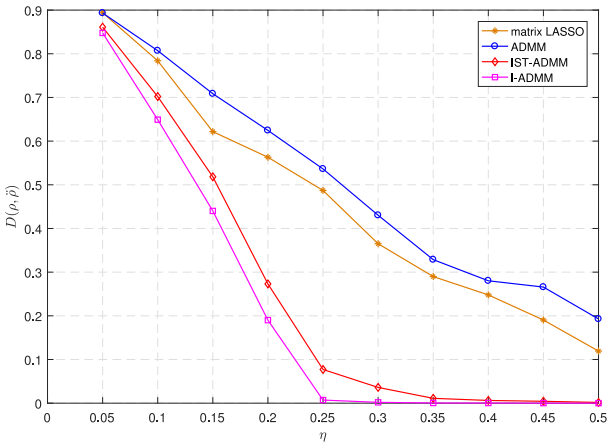


Fig. 1. Comparison of recovery accuracies of I-ADMM, IST-ADMM, ADMM, and matrix LASSO that run for $k_{\max} = 20$ iterations.

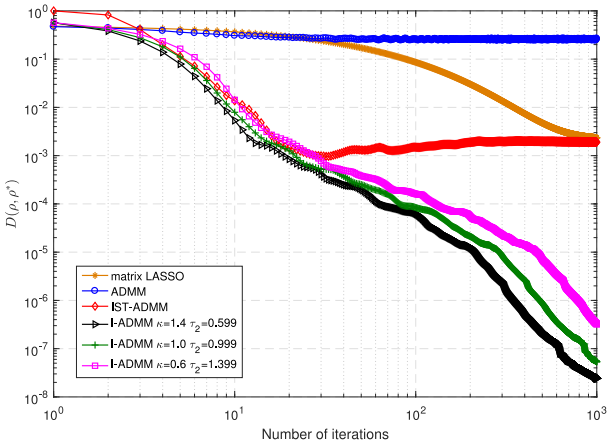


Fig. 2. Comparison of convergence rates of I-ADMM, IST-ADMM, ADMM, and matrix LASSO that run for the sampling rate $\eta = 0.3$.

the previous experiment has shown that it is sufficient for I-ADMM to accurately recover rank-2 density matrices when $n = 5$.

Fig. 2 depicts the evolution of $D(\rho, \rho^*)$ with respect to the number of iterations. For I-ADMM the estimated ρ gradually approaches to the optimal solution ρ^* with a remarkably fast convergence rate. Contrarily, IST-ADMM, ADMM, and matrix LASSO halt after tens of iterations. For I-ADMM, choosing different values of κ and τ_2 leads to slightly different convergence speeds. Empirically we observe that a moderately larger κ (say, $\kappa = 1.4$) is able to accelerate the algorithm, though in this case τ_2 must be smaller (say, $\tau_2 = 0.599$) due to the condition $\kappa + \tau_2 < 2$ given by Theorem 1. However, no matter how to fine-tune κ and τ_2, I-ADMM attains smaller $D(\rho, \rho^*)$ than IST-ADMM, ADMM, and matrix LASSO. When the algorithms run for 1000 iterations, the $D(\rho, \rho^*)$ of I-ADMM ($\kappa = 1.4$ and $\tau_2 = 0.599$), IST-ADMM, ADMM, and matrix LASSO are 2.41×10^{-8}, 1.90×10^{-3}, and 0.266 and 2.36×10^{-3}, respectively. Thus, we verify the proposed I-ADMM own better convergence performance than IST-ADMM, ADMM, and matrix LASSO.

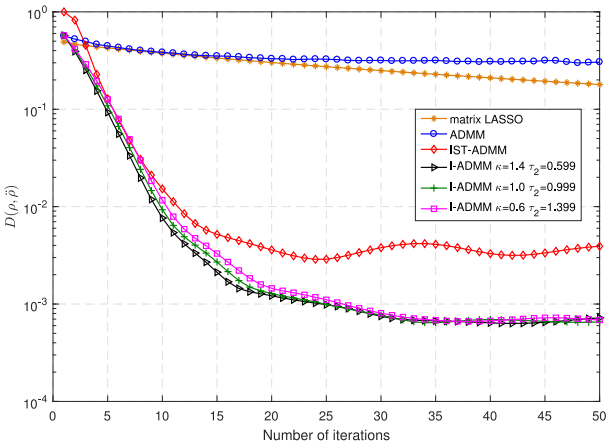


Fig. 3. Comparison of recovery accuracies of I-ADMM, IST-ADMM, ADMM, and matrix LASSO that run for the sampling rate $\eta = 0.3$.

Fig. 3 further shows the evolution of $D(\rho, \ddot{\rho})$ with respect to the number of iterations. After ten iterations, the values of $D(\rho, \ddot{\rho})$ for I-ADMM ($\kappa = 1.4$ and $\tau_2 = 0.599$), IST-ADMM, ADMM, and matrix LASSO are 0.0076 (the fidelity is 99.85%), 0.0152 (99.54%), 0.3842 (67.12%), and 0.3762 (68.63%), respectively. At the end of 50 iterations, the values of $D(\rho, \ddot{\rho})$ for I-ADMM ($\kappa = 1.4$ and $\tau_2 = 0.599$), IST-ADMM, ADMM, and matrix LASSO are 6×10^{-4} (the fidelity is 99.38%), 4×10^{-3} (97.65%), 0.3053 (77.83%), and 0.1797 (81.37%), respectively.

We also compare the running times of 50 iterations: 0.54 s (matrix LASSO), 27.42 s (ADMM), 0.75 s (IST-ADMM), 0.69 s (I-ADMM with $\kappa = 1.4$ and $\tau_2 = 0.599$), 0.72 s (I-ADMM with $\kappa = 1.0$ and $\tau_2 = 0.999$), and 0.68 s (I-ADMM with $\kappa = 0.6$ and $\tau_2 = 1.399$). The iteration complexity of ADMM is significantly higher, while those of I-ADMM, IST-ADMM, and matrix LASSO are similar.

VI. Conclusion

In this paper, we proposed an efficient and fast I-ADMM algorithm to design a convergent quantum state estimator with sparse disturbance based on compressive sensing and proved the convergence. The proposed algorithm solved the two subproblems regarding the density matrix and the sparse disturbance matrix approximately by taking proximal gradient steps, and then used an adjustable stepsize to update the Lagrange multiplier. The comparative simulation results demonstrated that the proposed algorithm allows lower sampling rates, converges faster and achieves higher accuracy compared with the algorithms matrix LASSO, ADMM, FP-ADMM, and IST-ADMM.

Appendix
Proof of Theorem 1

The proof consists of three steps. In the first step, we give a lower bound for the decrease of the distance from the primal-dual pair of I-ADMM to the optimal one. In the second step, we further prove that the decrease is sufficiently large when the parameters are properly chosen. This enables the third step to

show that the limit point of the primal-dual pair of I-ADMM is unique and optimal.

Step 1: The Lagrangian of (5) is

$$L(\rho, S, y) = \|\rho\|_* + I_C(\rho) + \gamma\|S\|_1 - \langle y, \mathcal{A}(\rho + S) - b\rangle. \tag{22}$$

Let (ρ^*, S^*, y^*) be any optimal primal-dual solution of (5). According to the convex optimization theory, its KKT conditions are as follows:

1) *Primal Feasibility:*

$$\mathcal{A}(\rho^* + S^*) = b. \tag{23}$$

2) *Stationarity:*

$$(\rho^*, S^*) = \arg\min_{\rho,S} L(\rho, S, y^*). \tag{24}$$

Observe that in (24), the optimization variables ρ and S are separable. Thus, for S we have

$$S^* = \arg\min_S \gamma\|S\|_1 - \langle y, \mathcal{A}(S)\rangle$$

and consequently its optimality condition

$$\mathcal{A}^{\mathrm{H}}(y^*) \in \gamma\,\partial\|S^*\|_1. \tag{25}$$

For ρ we have

$$\rho^* = \arg\min_\rho \|\rho\|_* + I_C(\rho) - \langle y, \mathcal{A}(\rho)\rangle. \tag{26}$$

Similar to the discussion in Section III-B, the quantum state constraint $\rho \in C$ means that $\rho = \rho^{\mathrm{H}}$, $\rho \succeq 0$ and $\mathrm{tr}(\rho) = 1$, and hence leads to $\|\rho\|_* = \mathrm{tr}(\rho) = 1$. Therefore, we can remove the nuclear norm term from (26) to obtain

$$\rho^* = \arg\min_\rho I_C(\rho) - \langle y, \mathcal{A}(\rho)\rangle$$

and consequently write its optimality condition as

$$\langle \rho^* - \rho, \mathcal{A}^{\mathrm{H}}(y^*)\rangle \geq 0, \quad \forall \rho \in C, \text{ and } \rho^* \in C. \tag{27}$$

In summary, the KKT conditions of (5) are (23), (25), and (27).

For convenience, we let $\hat{\rho} := \rho^{k+1}$, $\hat{S} := S^{k+1}$ and $\hat{y} := y^k - \alpha(\mathcal{A}(\hat{\rho} + \hat{S}) - b)$. As such $y^{k+1} = y^k - \kappa(y^k - \hat{y})$ due to (8c).

Using the fact of $\|\rho\|_* = \mathrm{tr}(\rho) = 1$, (8a) is equivalent to solving

$$\hat{\rho} = \arg\min_\rho \frac{\alpha}{2\tau_1}\left\|\rho - \tilde{\rho}^k\right\|_F^2, \quad \text{s.t. } \rho \in C. \tag{28}$$

Note that the optimality condition of (28) is given by

$$\langle \rho - \hat{\rho}, \hat{\rho} - \tilde{\rho}^k\rangle \geq 0, \quad \forall \rho \in C. \tag{29}$$

Using the definitions $\tilde{\rho}^k := \rho^k - \tau_1\mathcal{A}^H(\mathcal{A}(\rho^k + S^k) - b - y^k/\alpha)$ and $\hat{y} := y^k - \alpha(\mathcal{A}(\hat{\rho} + \hat{S}) - b)$, (29) can be rewritten to

$$\left\langle \rho - \hat{\rho}, \hat{\rho} - \rho^k + \tau_1\mathcal{A}^H \times \left(\mathcal{A}(\rho^k - \hat{\rho}) + \mathcal{A}(S^k - \hat{S}) - \frac{\hat{y}}{\alpha}\right)\right\rangle \geq 0, \quad \forall \rho \in C. \tag{30}$$

Because $\rho^* \in C$, we have

$$\left\langle \rho^* - \hat{\rho}, \hat{\rho} - \rho^k + \tau_1\mathcal{A}^H\left(\mathcal{A}\left(\rho^k - \hat{\rho}\right) + \mathcal{A}\left(S^k - \hat{S}\right) - \frac{\hat{y}}{\alpha}\right)\right\rangle \geq 0. \tag{31}$$

From (27) and $\hat{\rho} \in C$, we also have

$$\left\langle \rho^* - \hat{\rho}, \mathcal{A}^{\mathrm{H}}(y^*)\right\rangle \geq 0. \tag{32}$$

Multiplying (31) with α/τ_1 and summing up with (32), we obtain

$$\left\langle \hat{\rho} - \rho^*, \left(\frac{\alpha I}{\tau_1} - \alpha\mathcal{A}^H\mathcal{A}\right)\left(\rho^k - \hat{\rho}\right) - \alpha\mathcal{A}^H\mathcal{A}\left(S^k - \hat{S}\right) + \mathcal{A}^H\left(\hat{y} - y^*\right)\right\rangle \geq 0. \tag{33}$$

Similarly, the optimality condition for (8b) is given by

$$0 \in \frac{\gamma\tau_2}{\alpha}\partial\left\|\hat{S}\right\|_1 + \hat{S} - \tilde{S}^k. \tag{34}$$

Using the definitions $\tilde{S}^k := S^k - \tau_2\mathcal{A}^H(\mathcal{A}(\rho^{k+1} + S^k) - b - y^k/\alpha)$ and $\hat{y} := y^k - \alpha(\mathcal{A}(\hat{\rho} + \hat{S}) - b)$, (34) can be reduced to

$$0 \in \frac{\gamma\tau_2}{\alpha}\partial\left\|\hat{S}\right\|_1 + \hat{S} - S^k + \tau_2\mathcal{A}^H(\mathcal{A}(S^k - \hat{S}) - y^k/\alpha). \tag{35}$$

Combining (25) with (35) and using the fact that $\partial\|\cdot\|_1$ is a monotone operator, we get

$$\left\langle \hat{S} - S^*, \frac{\alpha}{\tau_2}\left(S^k - \hat{S}\right) - \alpha\mathcal{A}^H\mathcal{A}\left(S^k - \hat{S}\right) + \mathcal{A}^H\left(\hat{y} - y^*\right)\right\rangle \geq 0. \tag{36}$$

Using the primal feasibility condition $\mathcal{A}(\rho^* + S^*) = b$ in (23) and the definition $\hat{y} := y^k - \alpha(\mathcal{A}(\hat{\rho} + \hat{S}) - b)$, we manipulate the summation of (33) and (36) to obtain

$$\left\langle \hat{\rho} - \rho^*, \left(\frac{\alpha I}{\tau_1} - \alpha\mathcal{A}^H\mathcal{A}\right)\left(\rho^k - \hat{\rho}\right)\right\rangle + \frac{\alpha}{\tau_2}\left\langle \hat{S} - S^*, S^k - \hat{S}\right\rangle - \left\langle y^k - \hat{y}, \mathcal{A}\left(S^k - \hat{S}\right)\right\rangle + \frac{1}{\alpha}\left\langle y^k - \hat{y}, \hat{y} - y^*\right\rangle \geq 0. \tag{37}$$

To write (37) in a compact form and facilitate the following derivation, define triplets of variables $u^* := (\rho^*, S^*, y^*)$, $\hat{u} = (\hat{\rho}, \hat{S}, \hat{y})$, and $u^k := (\rho^k, S^k, y^k)$. Also define triplets of matrices $G_0 := (I, I, \kappa)$, $G_1 := ((\alpha/\tau_1)I - \alpha\mathcal{A}^H\mathcal{A}, (\alpha/\tau_2)I, 1/\alpha)$ and $G_2 := ((\alpha/\tau_1)I - \alpha\mathcal{A}^H\mathcal{A}, (\alpha/\tau_2)I, 1/(\alpha\kappa))$ where I denotes the $d \times d$ identity matrix. Observe that in G_0, G_1, and G_2, all the matrices are positive definite because $\kappa, \tau_1, \tau_2 > 0$ and $\tau_1\lambda_{\max} < 1$ by hypothesis. For a triplet of matrices $G = (H_1, H_2, H_3)$ and two triplets of variables (ρ, S, y) and (ρ', S', y'), define the product

$$G(\rho, S, y) = (H_1\rho, H_2S, H_3y)$$

the inner product

$$\left\langle (\rho, S, y), (\rho', S', y')\right\rangle_G = \left\langle \rho, \rho'\right\rangle_{H_1} + \left\langle S, S'\right\rangle_{H_2} + \left\langle y, y'\right\rangle_{H_3}$$

and the norm

$$\|(\rho, S, y)\|_G = \sqrt{\|\rho\|_{H_1}^2 + \|S\|_{H_2}^2 + \|y\|_{H_3}^2}.$$

Since all the matrices in G_0, G_1, and G_2 are positive definite, the norm is well defined.

Thus, (37) can be rewritten as

$$\left\langle \hat{u}-u^*, u^k-\hat{u}\right\rangle_{G_1} \geq \left\langle y^k-\hat{y}, \mathcal{A}(S^k-\hat{S})\right\rangle. \tag{38}$$

Adding $\|u^k-\hat{u}\|^2_{G_1}$ to both sides of (38), we obtain

$$\left\langle u^k-u^*, u^k-\hat{u}\right\rangle_{G_1} \geq \left\|u^k-\hat{u}\right\|^2_{G_1} + \left\langle y^k-\hat{y}, \mathcal{A}(S^k-\hat{S})\right\rangle. \tag{39}$$

According to the definition of $\hat{y}$, it holds $y^{k+1} = y^k - \kappa(y^k-\hat{y})$. Thus, the iteration of u can be written as $u^{k+1} = u^k - G_0(u^k-\hat{u})$, from which we have

$$\begin{aligned}
&\left\|u^{k+1}-u^*\right\|^2_{G_2} = \left\|u^k-u^*-G_0\left(u^k-\hat{u}\right)\right\|^2_{G_2}\\
&= \left\|u^k-u^*\right\|^2_{G_2} - 2\left\langle u^k-u^*, G_0\left(u^k-\hat{u}\right)\right\rangle_{G_2}\\
&\quad + \left\|G_0\left(u^k-\hat{u}\right)\right\|^2_{G_2}\\
&= \left\|u^k-u^*\right\|^2_{G_2} - 2\left\langle u^k-u^*, u^k-\hat{u}\right\rangle_{G_1} + \left\|G_0\left(u^k-\hat{u}\right)\right\|^2_{G_2}.
\end{aligned} \tag{40}$$

Here, the second equality comes from the definitions of G_0, G_1, and G_2. Combining (39) and (40), we get

$$\begin{aligned}
&\left\|u^k-u^*\right\|^2_{G_2} - \left\|u^{k+1}-u^*\right\|^2_{G_2}\\
&= 2\left\langle u^k-u^*, u^k-\hat{u}\right\rangle_{G_1} - \left\|G_0\left(u^k-\hat{u}\right)\right\|^2_{G_2}\\
&\geq 2\left\|u^k-\hat{u}\right\|^2_{G_1} + 2\left\langle y^k-\hat{y}, \mathcal{A}\left(S^k-\hat{S}\right)\right\rangle - \left\|G_0\left(u^k-\hat{u}\right)\right\|^2_{G_2}\\
&= \left\langle \rho^k-\hat{\rho}, \left(\frac{\alpha}{\tau_1}I-\alpha\mathcal{A}^H\mathcal{A}\right)\left(\rho^k-\hat{\rho}\right)\right\rangle + \frac{\alpha}{\tau_2}\left\|S^k-\hat{S}\right\|^2_F\\
&\quad + \frac{2-\kappa}{\alpha}\left\|y^k-\hat{y}\right\|^2_2 + 2\left\langle y^k-\hat{y}, \mathcal{A}\left(S^k-\hat{S}\right)\right\rangle.
\end{aligned} \tag{41}$$

Step 2: For the first term at the right-hand side of (41), we have

$$\begin{aligned}
&\left\langle \rho^k-\hat{\rho}, (\frac{\alpha}{\tau_1}I-\alpha\mathcal{A}^H\mathcal{A})(\rho^k-\hat{\rho})\right\rangle\\
&\qquad \geq \left(\frac{\alpha}{\tau_1}-\alpha\lambda_{\max}\right)\|\rho^k-\hat{\rho}\|^2_F
\end{aligned} \tag{42}$$

where $\lambda_{\max}$ is the largest eigenvalue of $\mathcal{A}^{\mathrm{H}}\mathcal{A}$. For the last term at the right-hand side of (41), observe that for any $\theta > 0$, it holds

$$\begin{aligned}
2\left\langle y^k-\hat{y}, \mathcal{A}(S^k-\hat{S})\right\rangle &\geq -\theta\left\|y^k-\hat{y}\right\|^2_2 - \frac{1}{\theta}\left\|\mathcal{A}(S^k-\hat{S})\right\|^2_F\\
&\geq -\theta\left\|y^k-\hat{y}\right\|^2_2 - \frac{\lambda_{\max}}{\theta}\left\|S^k-\hat{S}\right\|^2_F.
\end{aligned} \tag{43}$$

Combining (41)–(43), we have

$$\begin{aligned}
&\left\|u^k-u^*\right\|^2_{G_2} - \left\|u^{k+1}-u^*\right\|^2_{G_2}\\
&\geq \left(\frac{\alpha}{\tau_1}-\alpha\lambda_{\max}\right)\|\rho^k-\hat{\rho}\|^2_F\\
&\quad + \left(\frac{\alpha}{\tau_2}-\frac{\lambda_{\max}}{\theta}\right)\left\|S^k-\hat{S}\right\|^2_F + \left(\frac{2-\kappa}{\alpha}-\theta\right)\left\|y^k-\hat{y}\right\|^2_2.
\end{aligned} \tag{44}$$

Now, we proceed to show that we can choose a certain $\theta > 0$ such that all the coefficients at the right-hand side of (44) are positive. By hypothesis $\tau_1 < 1/\lambda_{\max}$ and thus $\alpha/\tau_1 - \alpha\lambda_{\max} > 0$. To make both $\alpha/\tau_2 - \lambda_{\max}/\theta$ and $(2-\kappa)/\alpha - \theta$ positive, we only need to choose θ within the range of $(\tau_2\lambda_{\max}/\alpha, (2-\kappa)/\alpha)$, which is nonempty because by hypothesis $\tau_2\lambda_{\max} + \kappa < 2$.

Step 3: Due to the nonnegativity of the right-hand side of (44), we have:

1) $\|u^k-u^*\|^2_{G_2}$ is monotonically decreasing and thus converges;
2) $\|u^k-u^{k+1}\|_{G_2} \to 0$;
3) $\{u^k\}$ lies in a compact region.

It follows from 2) that $\rho^k-\rho^{k+1} \to 0$, $S^k-S^{k+1} \to 0$ and $y^k-y^{k+1} \to 0$ as $k\to\infty$. Then $y^k = y^{k-1}-\kappa\alpha(\mathcal{A}(\rho^k+S^k)-b)$ implies that $\mathcal{A}(\rho^k+S^k)-b \to 0$. From 3), u^k has a subsequence $\{u^{k_j}\}$ that converges to $\bar{u} = (\bar{\rho}, \bar{S}, \bar{y})$. Therefore, $\bar{u} = (\bar{\rho}, \bar{S}, \bar{y})$ is a limit point of $\{u^k = (\rho^k, S^k, y^k)\}$, where $\mathcal{A}(\bar{\rho}+\bar{S})-b = 0$ and $\bar{\rho} \in C$ (because ρ^k stays in C throughout the optimization process). Below we prove that any limit point is an optimal solution to (5).

Consider any limit point $\bar{u} = (\bar{\rho}, \bar{S}, \bar{y})$. Observing (30) where $\hat{\rho} := \rho^{k+1}$, $\hat{S} := S^{k+1}$ and $\hat{y} := y^k - \alpha(\mathcal{A}(\hat{\rho}+\hat{S})-b)$, replacing (ρ^k, S^k, y^k) and $(\rho^{k+1}, S^{k+1}, y^{k+1})$ by $(\bar{\rho}, \bar{S}, \bar{y})$, and using the fact of $\mathcal{A}(\bar{\rho}+\bar{S})-b = 0$, from (30) we have

$$\langle \mathcal{A}^{\mathrm{H}}(\bar{y}), \bar{\rho}-\rho\rangle \geq 0, \quad \forall \rho \in C. \tag{45}$$

Similarly, (35) implies that

$$\mathcal{A}^{\mathrm{H}}(\bar{y}) \in \gamma\,\partial\left\|\bar{S}\right\|_1. \tag{46}$$

Along with $\bar{\rho} \in C$ and $\mathcal{A}(\bar{\rho}+\bar{S})-b = 0$, (45) and (46) imply that $(\bar{\rho}, \bar{S}, \bar{y})$ satisfies the KKT conditions of (5)—see (23), (25) and (27)—and is thus an optimal solution to (5).

To complete the proof, it remains to show that $\{u^k = (\rho^k, S^k, y^k)\}$ has a unique limit point. Let $\bar{u}_1 = (\bar{\rho}_1, \bar{S}_1, \bar{y}_1)$ and $\bar{u}_2 = (\bar{\rho}_2, \bar{S}_2, \bar{y}_2)$ be any two limit points. As we have shown, both $\bar{u}_1$ and $\bar{u}_2$ are optimal solutions to (5). Thus, u^* in (44) can be replaced by $\bar{u}_1$ and $\bar{u}_2$. This results in $\|u^{k+1}-\bar{u}_i\|^2_{G_2} \leq \|u^k-\bar{u}_i\|^2_{G_2}$, $i = 1, 2$, and consequently $\lim_{k\to\infty}\|u^k-\bar{u}_i\|^2_{G_2} = \nu_i$, $i = 1, 2$, where ν_1 and ν_2 are two finite constants. Using the identity

$$\begin{aligned}
&\left\|u^k-\bar{u}_1\right\|^2_{G_2} - \left\|u^k-\bar{u}_2\right\|^2_{G_2}\\
&= -2\left\langle u^k, \bar{u}_1-\bar{u}_2\right\rangle_{G_2} + \|\bar{u}_1\|^2_{G_2} - \|\bar{u}_2\|^2_{G_2}
\end{aligned}$$

and taking limits for $u^k \to \bar{u}_1$ and $u^k \to \bar{u}_2$, respectively, we have

$$\begin{aligned}
\nu_1-\nu_2 &= -2\langle \bar{u}_1, \bar{u}_1-\bar{u}_2\rangle_{G_2} + \|\bar{u}_1\|^2_{G_2} - \|\bar{u}_2\|^2_{G_2}\\
&= -\|\bar{u}_1-\bar{u}_2\|^2_{G_2}
\end{aligned}$$

and

$$\begin{aligned}
\nu_1-\nu_2 &= -2\langle \bar{u}_2, \bar{u}_1-\bar{u}_2\rangle_{G_2} + \|\bar{u}_1\|^2_{G_2} - \|\bar{u}_2\|^2_{G_2}\\
&= \|\bar{u}_1-\bar{u}_2\|^2_{G_2}.
\end{aligned}$$

Therefore, we must have $\|\bar{u}_1 - \bar{u}_2\|_{G_2}^2 = 0$, which indicates that the limit point of $\{(\rho^k, S^k, y^k)\}$ is unique and completes the proof.

REFERENCES

[1] C. Schwemmer *et al.*, "Systematic errors in current quantum state tomography tools," *Phys. Rev. Lett.*, vol. 114, no. 8, 2015, Art. no. 080403.

[2] T. Xin *et al.*, "Quantum state tomography via reduced density matrices," *Phys. Rev. Lett.*, vol. 118, no. 2, 2017, Art. no. 020401.

[3] C. Zu *et al.*, "Experimental realization of universal geometric quantum gates with solid-state spins," *Nature*, vol. 514, no. 7520, pp. 72–75, 2014.

[4] L. Hanzo *et al.*, "Wireless myths, realities, and futures: From 3G/4G to optical and quantum wireless," *Proc. IEEE*, vol. 100, pp. 1853–1888, May 2012.

[5] S. Imre and L. Gyongyosi, *Advanced Quantum Communications: An Engineering Approach*. Hoboken, NJ, USA: Wiley, 2012.

[6] D. Gross, Y.-K. Liu, S. T. Flammia, S. Becker, and J. Eisert, "Quantum state tomography via compressed sensing," *Phys. Rev. Lett.*, vol. 105, no. 15, 2010, Art. no. 150401.

[7] L. Gyongyosi, "Quantum imaging of high-dimensional Hilbert spaces with radon transform," *Int. J. Circuit Theory Appl.*, vol. 45, no. 7, pp. 1029–1046, 2017.

[8] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[9] C. Riofrío *et al.*, "Experimental quantum compressed sensing for a seven-qubit system," *Nat. Commun.*, vol. 8, May 2017, Art. no. 15305.

[10] S. Lloyd, M. Mohseni, and P. Rebentrost, "Quantum principal component analysis," *Nat. Phys.*, vol. 10, no. 9, pp. 631–633, 2014.

[11] W. T. Liu, T. Zhang, J. Y. Liu, P. X. Chen, and J. M. Yuan, "Experimental quantum state tomography via compressed sampling," *Phys. Rev. Lett.*, vol. 108, no. 17, 2012, Art. no. 170403.

[12] D. Suess, A. Eisfeld, and W. T. Strunz, "Hierarchy of stochastic pure states for open quantum system dynamics," *Phys. Rev. Lett.*, vol. 113, no. 15, 2014, Art. no. 150403.

[13] Y.-K. Liu, "Universal low-rank matrix recovery from Pauli measurements," in *Proc. Adv. Neural Inf. Process. Syst.*, 2011, pp. 1638–1646.

[14] N. Yamamoto and L. Bouten, "Quantum risk-sensitive estimation and robustness," *IEEE Trans. Autom. Control*, vol. 54, no. 1, pp. 92–107, Jan. 2009.

[15] Y. Liu, S. Kuang, and C. Shuang, "Lyapunov-based feedback preparation of GHZ entanglement of N-qubit systems," *IEEE Trans. Cybern.*, vol. 47, no. 11, pp. 3827–3839, Nov. 2017.

[16] C. Wu, B. Qi, C. Chen, and D. Dong, "Robust learning control design for quantum unitary transformations," *IEEE Trans. Cybern.*, vol. 47, no. 12, pp. 4405–4417, Dec. 2017.

[17] A. Smith *et al.*, "Quantum state tomography by continuous measurement and compressed sensing," *Phys. Rev. A*, vol. 87, no. 3, 2013, Art. no. 030102.

[18] Z. Hou *et al.*, "Full reconstruction of a 14-qubit state within four hours," *New J. Phys.*, vol. 18, no. 8, 2016, Art. no. 083036.

[19] K. Z. Li and S. Cong, "A robust compressive quantum state tomography algorithm using ADMM," *IFAC Proc. Vol.*, vol. 47, no. 3, pp. 6878–6883, 2014.

[20] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *J. ACM*, vol. 58, no. 3, pp. 1101–1137, 2011.

[21] V. Chandrasekaran, S. Sanghavi, P. A. Parrilo, and A. S. Willsky, "Rank-sparsity incoherence for matrix decomposition," *SIAM J. Optim.*, vol. 21, no. 2, pp. 572–596, 2009.

[22] J.-F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM J. Optim.*, vol. 20, no. 4, pp. 1956–1982, 2010.

[23] Y. Yin, D. Xu, X. Wang, and M. Bai, "Online state-based structured SVM combined with incremental PCA for robust visual tracking," *IEEE Trans. Cybern.*, vol. 45, no. 9, pp. 1988–2000, Sep. 2015.

[24] M. Tao and X. Yuan, "Recovering low-rank and sparse components of matrices from incomplete and noisy observations," *SIAM J. Optim.*, vol. 21, no. 1, pp. 57–81, 2011.

[25] K. Zheng, K. Z. Li, and S. Cong, "A reconstruction algorithm for compressive quantum tomography using various measurement sets," *Sci. Rep.*, vol. 6, Dec. 2016, Art. no. 38497.

[26] J. J. Zhang, K. Z. Li, S. Cong, and H. T. Wang, "Efficient reconstruction of density matrices for high dimensional quantum state tomography," *Signal Process.*, vol. 139, pp. 136–142, Oct. 2017.

[27] D. Gabay and B. Mercier, "A dual algorithm for the solution of nonlinear variational problems via finite element approximation," *Comput. Math. Appl.*, vol. 2, no. 1, pp. 17–40, 1976.

[28] J. Eckstein and D. P. Bertsekas, "On the Douglas–Rachford splitting method and the proximal point algorithm for maximal monotone operators," *Math. Program.*, vol. 55, nos. 1–3, pp. 293–318, 1992.

[29] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends® Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2011.

[30] H. Wang, Y. Gao, Y. Shi, and R. Wang, "Group-based alternating direction method of multipliers for distributed linear classification," *IEEE Trans. Cybern.*, vol. 47, no. 1, pp. 3568–3582, Nov. 2017.

[31] B. S. He and X. M. Yuan, "On the $o(1/n)$ convergence rate of the Douglas–Rachford alternating direction method," *SIAM J. Numer. Anal.*, vol. 50, no. 2, pp. 700–709, 2012.

[32] W. Deng and W. Yin, "On the global and linear convergence of the generalized alternating direction method of multipliers," *J. Sci. Comput.*, vol. 66, no. 3, pp. 889–916, 2016.

[33] M. Hong and Z.-Q. Luo, "On the linear convergence of the alternating direction method of multipliers," *Math. Program.*, vol. 162, nos. 1–2, pp. 165–199, 2017.

[34] P. Giselsson and S. Boyd, "Linear convergence and metric selection for Douglas–Rachford splitting and ADMM," *IEEE Trans. Autom. Control*, vol. 62, no. 2, pp. 532–544, Feb. 2017.

[35] J. F. Yang and Y. Zhang, "Alternating direction algorithms for ℓ_1-problems in compressive sensing," *SIAM J. Sci. Comput.*, vol. 33, no. 1, pp. 250–278, 2009.

[36] D. S. Gonçalves, M. A. Gomes-Ruggiero, and C. Lavor, "A projected gradient method for optimization over density matrices," *Optim. Methods Softw.*, vol. 31, no. 2, pp. 328–341, 2015.

[37] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Commun. Pure Appl. Math.*, vol. 57, no. 11, pp. 1413–1457, 2004.

[38] S. Q. Ma, "Alternating proximal gradient method for convex minimization," *J. Sci. Comput.*, vol. 68, no. 2, pp. 546–572, 2016.

[39] K. Zyczkowski, K. A. Penson, I. Nechita, and B. Collins, "Generating random density matrices," *J. Math. Phys.*, vol. 52, no. 6, p. 255, 2010.

[40] J. Wright, A. Ganesh, K. Min, and Y. Ma, "Compressive principal component pursuit," in *Proc. IEEE Int. Symp. Inf. Theory*, vol. 2. Cambridge, MA, USA, 2012, pp. 1276–1280.

[41] L. Gyongyosi, S. Imre, and H. V. Nguyen, "A survey on quantum channel capacities," *IEEE Commun. Surveys Tuts.*, to be published, doi: 10.1109/COMST.2017.2786748.

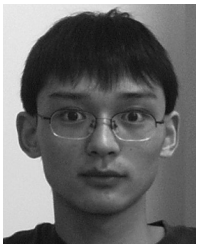
[42] S. T. Flammia and Y. K. Liu, "Direct fidelity estimation from few pauli measurements," *Phys. Rev. Lett.*, vol. 106, no. 106, 2011, Art. no. 230501.

Jiaojiao Zhang received the B.E. degree in automation from the School of Automation, Harbin Engineering University, Harbin, China, in 2015. She is currently pursuing the master's degree in control theory and control engineering from the University of Science and Technology of China, Hefei, China.

Her current research interests include convex optimization and quantum state estimation.

Shuang Cong (M'07–SM'12) received the graduation degree from the Beijing University of Aeronautics and Astronautics, Beijing, China, in 1982 and the Ph.D. degree in system engineering from the University of Rome La Sapienza, Rome, Italy, in 1995.

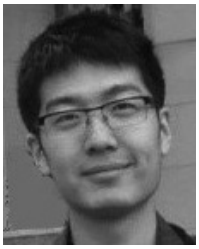
She is currently a Professor with the Department of Automation, University of Science and Technology of China, Hefei, China. Her current research interest includes advanced control strategies for motion control, fuzzy logic control, neural networks design and applications, robotic coordination control, and quantum systems control.



Qing Ling (M'07–SM'15) received the B.E. degree in automation and the Ph.D. degree in control theory and control engineering from the University of Science and Technology of China, Hefei, China, in 2001 and 2006, respectively.

He was a Post-Doctoral Research Fellow with the Department of Electrical and Computer Engineering, Michigan Technological University, Houghton, MI, USA, from 2006 to 2009, and an Associate Professor with the Department of Automation, University of Science and Technology of China from 2009 to 2017. He is currently a Professor with the School of Data and Computer Science, Sun Yat-sen University, Guangzhou, China. His current research interest includes decentralized network optimization and its applications.

Dr. Ling was a recipient of the 2017 IEEE Signal Processing Society Young Author Best Paper Award (as co-author) and the 2017 International Consortium of Chinese Mathematicians Distinguished Paper Award. He is an Associate Editor of the IEEE TRANSACTIONS ON NETWORK AND SERVICE MANAGEMENT and the IEEE SIGNAL PROCESSING LETTERS.



Kezhi Li received the B.Eng. degree in electronic engineering from the University of Science and Technology of China (USTC), Hefei, China, and the Ph.D. degree from Imperial College London, London, U.K.

He was a Research Associate with the University of Cambridge, Cambridge, U.K., a Research Fellow with the Royal Institute of Technology, Stockholm, Sweden, and a Research Assistant with Microsoft Research Asia, Beijing, China, and USTC. He is currently a Research Fellow with the Department of Electrical and Electronic Engineering, Imperial College London. His current research interests include several interdiscipline areas, such as statistical signal processing and their applications in quantum tomography and bio-engineering.