Low-rank Matrix Recovery from Noisy, Quantized and Erroneous Measurements

Pengzhi Gao, *Student member, IEEE,* Ren Wang, *Student Member, IEEE,* Meng Wang, *Member, IEEE*, Joe H. Chow, *Fellow, IEEE*

***Abstract*—This paper proposes a communication-reduced, cyber-resilient, and information-preserved data collection framework. Random noise and quantization are applied to the measurements before transmission to compress data and enhance data privacy. Leveraging the low-rank property of the data, we develop novel methods to recover the original data from quantized measurements even when partial measurements are corrupted. The data recovery is achieved through solving a constrained maximum likelihood estimation problem. The recovery error is proven to be order-wise optimal and decays in the same order as that of the state-of-the-art method when there is no corruption. The data accuracy is thus maintained while the data privacy is enhanced. A proximal algorithm with convergence guarantee is proposed to solve the non-convex problem. The analyses are extended to situations when some measurements are lost, or when multiple copies of noisy, quantized measurements are sent for each data point. A new application of this framework for data privacy in power systems is discussed. Experiments on synthetic data and real synchrophasor data in power systems demonstrate the effectiveness of our method.**

***Index Terms*—low-rank matrix, quantization, maximum likelihood, data privacy.**

I. Introduction

Modern devices like Phasor Measurement Units (PMUs) [29] are installed in power systems, providing voltage and current phasors at a rate of 30 samples per second. Due to the mismatch with communication networks that were not designed to carry high-speed PMU data, data losses and data quality degradations happen quite often, especially in the Eastern interconnection [33]. On the other hand, data privacy in smart grids is receiving increasing attentions. Despite the existing legal agreement on data privacy [28], the technical guarantee of data privacy is limited. PMU measurements are owned by utility companies and considered to be private and sensitive. To the best of our knowledge, only [35] considered protecting the privacy of PMU data through encryption. Privacy of smart meter data has been formulated as a lossy source coding problem [32]. A rate-distortion bound is derived in [32], but no discussion on decoder design was included there.

This paper proposes a simple signal processing approach to achieve data compression and data privatization. Random noise are added to the measurements to protect data privacy, and quantization is applied afterwards to reduce the communication cost. A similar idea appeared recently for sensor networks [36]. Quantization can also protect the privacy of smart meter data [30], and adding noise is known to enhance the privacy level of individual users, measured by the differential privacy [15]. Such data privacy is, however, achieved at a cost of reduced data accuracy. One major contribution of this paper is a data recovery method utilizing the obtained quantized measurements such that the recovery error diminishes when the data dimension increases. Therefore, the reduced data transmission, privacy enhancement of individual utilities, and the information accuracy for the central operator are achieved simultaneously under the proposed data collection framework.

The proposed recovery method leverages the property that the spatial-temporal blocks of PMU data are low-rank, which has been exploited to recover PMU data losses [17] and correct erroneous measurements [16]. Our problem formulation falls in line with recent research on low-rank matrix recovery from quantized measurements [1]–[3], [7], [11], [13], [21], [24]–[26], motivated by collaborative filtering, sensor network location, and image processing. The common feature of the existing recovery methods is to relate the quantized measurements with the underlying low-rank matrix L with a probabilistic model and estimate L by solving a constrained maximum likelihood optimization problem. Refs. [7], [13] provided the recovery guarantee when measurements are binary and partially lost. Refs. [3], [21] extended to multi-level quantized measurements and provided the bound of the recovery error. Ref. [2] considered the special case that the low-rank matrix is also positive semidefinite, motivated by the sensor localization problem.

The existing methods rely on the assumption that the measurement matrix is exactly low-rank, which may not hold when some measurements are corrupted. For instance, bad data exist in PMU measurements, and bad data detection has always been an important issue for power system monitoring. To the best of our knowledge, only [25] considered the matrix recovery problem when partial measurements are corrupted, but no theoretical analysis of the recovery accuracy was reported.

This paper provides the first theoretical study of low-rank matrix recovery from quantized, partially corrupted, and partially lost measurements. The recovery error using partially corrupted measurements by our method is in the same order as the state-of-the-art result when no corruption exists [1], [3]. When there is no missing data, our recovery error is further proven to be in the same order as the minimum possible error from unquantized measurements. The data recovery is achieved by solving a nonconvex constrained maximum likelihood optimization problem. A proximal algorithm with a global convergence guarantee is developed to solve the

The first two authors contributed equally to this paper. The authors are with the Dept. of Electrical, Computer, and Systems Engineering, Rensselaer Polytechnic Institute, Troy, NY. Email: {gaop, wangr8, wangm7, chowj}@rpi.edu. Partial and preliminary results appeared in [18].

proposed non-convex problem.

Moreover, this paper considers a general scenario where multiple copies of quantized measurements with independent noise could be generated for the same ground-truth value to improve the data recovery accuracy, while the existing work all focused on a single copy of measurement. A data recovery method with theoretical guarantee is developed and proved to be order-wise optimal in reducing the recovery error.

The rest of the paper is organized as follows. The problem formulation and motivation are introduced in Section II. Section III discusses the data recovery with one quantized measurement per value. Section III-C extends to cases with multiple quantized measurements per value. Section V extends to cases with missing data. Section VI records our numerical experiments. Section VII concludes the paper. All the proofs are deferred to the Appendix.

II. Problem Formulation and Motivation

Let $L^* \in \mathbb{R}^{m\times n}$ denote the actual data, and let $C^* \in \mathbb{R}^{m\times n}$ denote the sparse additive errors in the measurements. Let $M^* = L^* + C^*$ denote the measurements that are partially erroneous[1]. The rank of L^* is r ($r \ll n$), and C^* has at most s nonzero entries. Let $\|\|_\infty$ and $\|\|_F$ denote entry-wise infinity norm and matrix Frobenius norm. We further assume $\|L^*\|_\infty \le \alpha$ and $\|C^*\|_\infty \le \alpha$ for some constant α. Let $N \in \mathbb{R}^{m\times n}$ denote the noise matrix which has i.i.d. entries with known cumulative distribution function $\Phi(z)$. Given a positive constant K, let $[K]$ denote the set $\{1, ..., K\}$. A K-level quantized noisy measurement Y_{ij} based on M^*_{ij} satisfies

$$Y_{ij} = \mathcal{Q}(L^*_{ij} + C^*_{ij} + N_{ij}),\ \forall (i,j), \tag{1}$$

where the operator $\mathcal{Q}$ maps a real number to one of the K labels. Given the quantization boundaries $\omega_0 < \omega_1 < ... < \omega_K$, we have

$$\mathcal{Q}(x) = l \quad \text{if } \omega_{l-1} < x \le \omega_l,\ l \in [K]. \tag{2}$$

Then, one can check that

$$Y_{ij} = l \text{ with probability } f_l(M^*_{ij}),\ \forall (i,j), \tag{3}$$

where $\sum_{l=1}^K f_l(M^*_{ij}) = 1$, and

$$f_l(M^*_{ij}) = P(Y_{ij} = l | M^*_{ij}) = \Phi(\omega_l - M^*_{ij}) - \Phi(\omega_{l-1} - M^*_{ij}). \tag{4}$$

The data recovery problem is stated as follows.

(P1) Given quantized observations Y, noise distribution Φ, and quantization boundaries $\{\omega_0, ..., \omega_K\}$, can we recover the actual data L^*?

(P1) lies within the framework of low-rank matrix estimation problem that finds applications in image processing, collaborative filtering, and environmental monitoring [8], [10]. A special case of (P1) where C^* is a zero matrix has been studied in [1], [3], [7], [11], [13], [21], [24], [26] for collaborative filtering and commodity selection [20], where Y represents users' responses chosen from a discrete set, and L^* represents users' preferences. Ref. [25] further considered nonzero C^*, but no theoretical analysis was reported.

Only the noise distribution Φ instead of the actual noise N is available for data recovery. Two common choices of Φ are: **(i)** logistic model with $\Phi(x) = \Phi_{\log}(x/\sigma) = \frac{1}{1+e^{-x/\sigma}}$, $\sigma > 0$; **(ii)** Probit model with $\Phi(x) = \Phi_{\text{norm}}(x/\sigma)$, $\sigma > 0$, where Φ_{norm} is the cumulative distribution function of the standard normal distribution $\mathcal{N}(0,1)$. Other noise models such as Poisson noise [31] and general exponential noise families [19] have been studied for the low-rank matrix estimation problem.

This paper proposes a novel application of (P1) in communication-reduced, cyber-resilient, and information-preserved data collection. Each row of L^* represents the measurements of one sensor across time instants. C^* represents additive errors in the measurements resulting from sensor malfunctions. As shown in Fig. 1, instead of transmitting the raw data to the fusion center, we add noise and apply quantization to the observations in the sensors. Only highly noisy and quantized values are transmitted to reduce the communication cost and increase data security to eavesdroppers. The fusion center will recover the data by solving (P1).

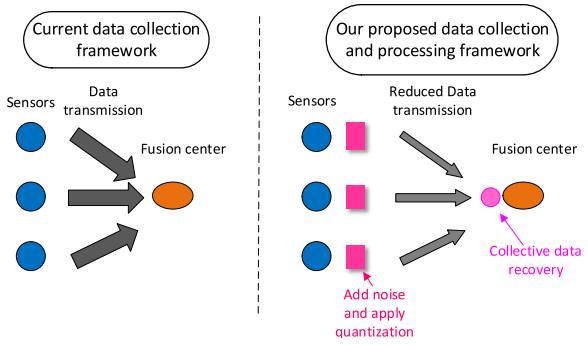


Fig. 1. Our proposed data collection framework.

One application of the proposed system is the collection of PMU data, where each sensor corresponds to each PMU, and the fusion center corresponds to the system operator. As we will show through Theorems 1, 3 and 5, the data recovery will be accurate only if m and n are large enough. The operator with sufficient number of measurements from many PMUs can recover the data accurately, while an eavesdropper with access to a few PMUs cannot obtain much useful information even if it uses the same method as the operator. Thus, the reduction of communication, data privacy against eavesdroppers, and accurate information recovery for the operator are achieved simultaneously. Another application is load monitoring of power systems using data collected from smart meters [27], where the privacy of individual users are protected while the load estimation accuracy is maintained.

Note that in this framework, the error correction is not performed at each individual sensor before quantization because of the limited computational resources at sensors and the data privacy concern. For each sensor to correct errors in its own observations without knowing measurements from other sensors, some additional prior information is required. For instance, L^* can be decomposed as $L^* = U^* V^{*T}$ with $U \in \mathbb{R}^{m\times r}$ and $V \in \mathbb{R}^{n\times r}$. If V^* is known to a sensor, and

[1]We assume $m \ge n$ throughout the paper for notational simplicity. The results can be extended to cases that $n \ge m$ with very minor modifications.

the percentage of errors is small, the error correction problem at each sensor can be formulated as a compressed sensing problem and solved by methods like ℓ_1 minimization [9], [14]. A central node needs to collect raw data from multiple sensors, compute V^*, and then communicate V^* to all the sensors. Since V^* might change over time, the data collection and estimation should happen continuously. Then the privacy of this data communication is still an issue, which is what this paper wants to address.

Motivated by the proposed data collection application, this paper also considers a more general setup where a sensor could send W ($W \geq 1$) copies of noisy quantized measurements based on the same ground-truth M^*_{ij} instead of just sending one copy. Each measurement Y^t_{ij} ($t \in [W]$) is obtained by adding independent noise and applying quantization, i.e.,

$$Y^t_{ij} = \mathcal{Q}(L^*_{ij} + C^*_{ij} + N^t_{ij}),\ \forall (i,j), \tag{5}$$

A generalized data recovery problem is stated as follows

(P2) Given quantized observations Y^t for all $t \in [W]$, noise distribution Φ, and $\omega_l, l \in [K]$, can we recover L^*?

(P2) has never been studied before. Intuitively, the data recovery accuracy can be improved by increasing K and/or W. We will address (P1) (i.e., $W = 1$) in Section III and extend the idea to solve (P2) in Section IV.

Another generalization of (P1) is that partial measurements are lost during the transmission. Only some of entries of Y are available for data recovery. Let Ω denote the set of indices of measurements that arrive at the fusion center. The data recovery problem with data losses is stated as follows

(P3) Given partial quantized observations Y_{ij}'s for $(i,j) \in \Omega$, noise distribution Φ, and $\omega_l, l \in [K]$, can we recover L^*?

We will study (P3) in Section V.

III. Quantized Matrix Recovery with Erroneous Measurements

To solve (P1), we propose to estimate the unknown (L^*, C^*) using a constrained maximum likelihood approach. The negative log-likelihood function is given by

$$F_Y(X) = -\sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{l=1}^{K}(\mathbf{1}_{[Y_{ij}=l]}\log(f_l(X_{ij}))), \tag{6}$$

where $\mathbf{1}_{[A]}$ denotes the indicator function that takes value '1' if A is true and value '0' otherwise. (6) is a convex function in X when the function f_l is log-concave in X_{ij}. We estimate (L^*, C^*) by $(\hat{L}, \hat{C})$, where

$$\begin{aligned}(\hat{L}, \hat{C}) = \arg\min_{L,C} &-\sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{l=1}^{K}\mathbf{1}_{[Y_{ij}=l]}\log(f_l(L_{ij}+C_{ij})),\\ &\text{s.t. } L + C \in \mathcal{S}_f,\end{aligned} \tag{7}$$

and the feasible set $\mathcal{S}_f$ is defined as

$$\begin{aligned}\mathcal{S}_f := \{X \in \mathbb{R}^{m\times n} : X = L + C, \|L\|_\infty \leq \alpha, \|C\|_\infty \leq \alpha,\\ \text{rank}(L) \leq r, \sum_{i=1}^{m}\sum_{j=1}^{n}\mathbf{1}_{[C_{ij}\neq 0]} \leq s\}.\end{aligned} \tag{8}$$

(7) is nonconvex due to the nonconvexity of $\mathcal{S}_f$.

The recovery accuracy of the global minimizer of (7) is discussed in Section III-A. We provide the information theoretical limit of solving (P1) in Section III-B and demonstrate that the recovery accuracy of the global minimizer of (7) is order-wise optimal. An approximate algorithm to solve (7) is introduced in Section III-C.

A. Theoretical recovery guarantee of our approach

Our theoretical bound is built upon [1], in which C^* is a zero matrix. We generalize the analysis to an arbitrary sparse matrix C^*. We first define two constants γ_α and L_α as follows,

$$\gamma_\alpha = \min_{l\in[K]}\inf_{|x|\leq 2\alpha}\{\frac{\dot{f}^2_l(x)}{f^2_l(x)} - \frac{\ddot{f}_l(x)}{f_l(x)}\}, \tag{9}$$

$$L_\alpha = \max_{l\in[K]}\sup_{|x|\leq 2\alpha}\{|\dot{f}_l(x)|/f_l(x)\}, \tag{10}$$

where $\dot{f}_l(x)$ and $\ddot{f}_l(x)$ are the first and second order derivatives with respect to x. γ_α and L_α depend on α, Φ, and ω_i's ($i \in \{0, 1, ..., K\}$) but independent of m and n. Since $f_l(x)$ is log-concave if and only if $(\dot{f}_l(x))^2 \geq \ddot{f}_l(x)f_l(x)$ [6], $\gamma_\alpha \geq 0$ for log-concave f_l and $\gamma_\alpha > 0$ for strictly log-concave f_l. One can check that $\gamma_\alpha > 0$ for logistic and Gaussian noises [1].

The following theorem applies to strictly log-concave f_l's including logistic and Gaussian noise models.

Theorem 1. *Suppose that $M^* \in \mathcal{S}_f$, $m \geq n$, and $f_l(x)$ is strictly log-concave in x, $\forall l \in [K]$. Then, with probability at least $1 - C_1e^{-C_2 m}$, any global minimizer $(\hat{L}, \hat{C})$ of (7) satisfies*

$$\frac{\|(\hat{L}+\hat{C}) - (L^*+C^*)\|_F}{\sqrt{mn}} \leq \min(2\alpha(1+\sqrt{\frac{s}{mn}}), U_\alpha), \tag{11}$$

where

$$U_\alpha = \max(\frac{8.04L_\alpha\sqrt{2r}}{\gamma_\alpha\sqrt{n}}, \sqrt{\frac{16.08\alpha L_\alpha\sqrt{2rms} + 8\alpha s L_\alpha}{\gamma_\alpha mn}}), \tag{12}$$

for some positive constants C_1 and C_2. Furthermore, with the same probability, it holds that

$$\frac{\|\hat{L} - L^*\|_F}{\sqrt{mn}} \leq \min(2\alpha(1+2\sqrt{\frac{s}{mn}}), U_\alpha + 2\alpha\sqrt{\frac{s}{mn}}). \tag{13}$$

Theorem 1 establishes the recovery error when the measurements are partially corrupted. As long as s is at most $\Theta(m)$, i.e., the number of corrupted measurements per row is bounded[2], then

$$\|(\hat{L}+\hat{C}) - (L^*+C^*)\|_F/\sqrt{mn} \leq O(\sqrt{2r/n}), \tag{14}$$

and

$$\|\hat{L} - L^*\|_F/\sqrt{mn} \leq O(\sqrt{2r/n}). \tag{15}$$

When there is no corruption, i.e., $C^* = 0$, the recovery error by the method in [1] is in the same order $O(\sqrt{\frac{r}{n}})$. Thus, our method can handle the additional sparse errors without increasing the recovery error. If s is higher order than m, the

[2]We use the notations $g(n) \in O(h(n))$, $g(n) \in \Omega(h(n))$, or $g(n) = \Theta(h(n))$ if as n goes to infinity, $g(n) \leq c \cdot h(n)$, $g(n) \geq c \cdot h(n)$ or $c_1 \cdot h(n) \leq g(n) \leq c_2 \cdot h(n)$ eventually holds for some positive constants c, c_1 and c_2 respectively.

error bound is $O(\sqrt{\frac{s}{mn}})$. For example, if $s = \Theta((mn)^{1-\alpha})$ for some $\alpha \in (0, 1/2]$, the error bound is $O((mn)^{-\alpha/2})$, which also decays to zero as the dimension increases.

We remark that although it is intuitive and verified by our numerical experiments that the recovery error decreases when the quantization level K increases, the error bound in Theorem 1 does not capture this property explicitly. As shown in Theorem 1, the recovery error decreases when L_α/γ_α decreases, while the dependences of L_α/γ_α on K cannot be characterized in a closed-form expression without specifying Φ and ω_l's. In fact, we show in Appendix-A that when the noise follows the logistic model, L_α/γ_α is minimized when $K = 2$. This counter-intuitive result is due to the artifacts in our proof techniques, which are built upon sufficient but not necessary conditions to study the worst-case performance. We emphasize that the significance of Theorem 1 is that it characterizes the recovery error as a function of m, and as will shown in Theorem 2, this error bound is order-wise optimal in the sense that no other recovery method can achieve an error rate that decays faster in terms of the order of m.

B. Optimality of our proposed method

We next characterize the lower bound of the recovery error of any method even when the measurements are unquantized in Theorem 2. We will show that little is lost with quantization, and our recovery approach is almost optimal.

We consider the case that the noise follows the Gaussian distribution with zero mean and variances σ^2. We consider the region that s is at most $\Theta(mn)$, and m and n are sufficiently large. To simplify the presentation, we quantify this region as

$$s/mn \leq C_0 \text{ and } \frac{r}{n}(mn - s) \geq 64 \tag{16}$$

for a constant C_0.

Theorem 2. *Let $N \in \mathbb{R}^{m\times n}$ contain i.i.d. entries from $\mathcal{N}(0, \sigma^2)$. Assume* (16) *holds. Consider any algorithm that, for any $M \in \mathcal{S}_f$, takes $Y = M + N$ as the input and returns an estimate $\hat{M}$. Then there always exists $M \in \mathcal{S}_f$ such that with probability at least $\frac{3}{4}$,*

$$\frac{\|\hat{M} - M\|_F}{\sqrt{mn}} \geq \min(C_3, C_4\sigma\sqrt{\frac{rm - r\lfloor\frac{s}{n}\rfloor - 64}{mn - n\lfloor\frac{s}{n}\rfloor}}) \tag{17}$$

holds for some fixed constants C_3 and C_4, where $C_3 = \sqrt{\frac{1-2C_0}{8}}\alpha$ and $C_4 < \sqrt{\frac{1-2C_0}{256}}$.

The proof of Theorem 2 borrows the idea from [13] by exploiting the standard arguments in information theory using Fano's inequality. See Appendix-D, E for details. If s is no greater than $\Theta(mn)$, then from (17), we have

$$\|\hat{M} - M\|_F/\sqrt{mn} \geq \Theta(\sqrt{r/n}), \tag{18}$$

i.e., the minimum possible recovery error from unquantized noisy measurements is at least in the order of $O(\sqrt{\frac{r}{n}})$. Note that from (14), the recovery error by solving (7) from quantized measurements is at most $O(\sqrt{\frac{r}{n}})$, as long as s is at most $\Theta(m)$. Combining (14) and (18), we conclude that the recovery error of our method is order optimal when s is at most $\Theta(m)$. Thus, very little information is lost by quantization.

The lower bound of the recovery error in this problem formulation is also studied in [22], where the error is measured by $\sqrt{\frac{\|\hat{L}-L\|_2^2+\|\hat{C}-C\|_2^2}{mn}}$. Ref. [22] shows the error is lower bounded by $O(\sqrt{r/n + s/mn})$, and this rate is minimax optimal. Note that the errors here and in [22] are measured using different norms.

C. A proximal algorithm and its convergence analysis

Algorithm 1 Algorithm for quantized matrix recovery

Input: Quantized matrix $Y \in \mathbb{R}^{m\times n}$, initialization matrices $U^0 \in \mathbb{R}^{m\times r}$, $V^0 \in \mathbb{R}^{n\times r}$ and zero matrix $C^0 \in \mathbb{R}^{m\times n}$, parameters r, β, s, and tol.

for $t = 0, 1, 2, ...,$ until $\|U_t - U_{t-1}\|_F + \|V_t - V_{t-1}\|_F + \|C_t - C_{t-1}\|_F < tol$ **do**
  $U^{t+1} = U^t - \tau_U \nabla_U F_Y(U^t V^{t^T} + C^t)$.
  $V^{t+1} = V^t - \tau_V \nabla_V F_Y(U^{t+1} V^{t^T} + C^t)$.
  $C^{t+1} = C^t - \tau_C \nabla_C F_Y(U^{t+1} V^{t+1^T} + C^t)$.
  if $\sum_{ij} \mathbf{1}_{[C_{ij}^{t+1} \neq 0]} > s$ **then**
    C^{t+1} only keeps s entries with the largest absolute values. Other nonzero entries are set to be zero.
  end if
end for
Return: $L^* = U^* V^{*^T}$ and C^*.

We propose an algorithm to solve the non-convex optimization problem (7) approximately. We factorize the low-rank matrix L into $L = UV^T$ with $U \in \mathbb{R}^{m\times r}$ and $V \in \mathbb{R}^{n\times r}$. We first rewrite our optimization problem as follows.

$$(\hat{U}, \hat{V}, \hat{C}) = \text{argmin}_{U,V,C} F_Y(UV^T + C) + K(C), \tag{19}$$

where

$$K(C) = \begin{cases} 0, & \text{if } \sum_{ij} \mathbf{1}_{[C_{ij}\neq 0]} \leq s, \\ +\infty, & \text{if } \sum_{ij} \mathbf{1}_{[C_{ij}\neq 0]} > s. \end{cases} \tag{20}$$

$K(C)$ guarantees that C has at most s nonzero entries. Here we discard the constant constraint α on the infinity norms of the matrices in the proximal algorithm. The infinity norm constraints simplify the theoretical analysis of the recovery problem. In practical cases that the infinity norm is unknown, one could set α to be a large value, then these constraints are often times not binding. This also happens when L^* has a few large entries while most entries have relatively small absolute values. Moreover, dropping these constraints simplifies the algorithm and is more practical for cases when α is unknown. The numerical results are not affected much in many cases.

In each iteration, we apply alternating gradient descent to update the estimations of U, V and C. A hard thresholding is applied to C afterwards by keeping s entries with the largest absolute values and setting others to zero. The details of our algorithm is summarized in Algorithm 1. We have

$$\begin{aligned} \nabla_U F_Y(UV^T + C) &= \nabla_X F_Y(X)V, \\ \nabla_V F_Y(UV^T + C) &= \nabla_X F_Y(X)^T U, \\ \nabla_C F_Y(UV^T + C) &= \nabla_X F_Y(X), \end{aligned} \tag{21}$$

where

$$[\nabla_X F_Y(X)]_{ij} = \frac{\dot{\Phi}(\omega_{Y_{ij}} - X_{ij}) - \dot{\Phi}(\omega_{Y_{ij}-1} - X_{ij})}{\Phi(\omega_{Y_{ij}} - X_{ij}) - \Phi(\omega_{Y_{ij}-1} - X_{ij})}, \forall (i,j).$$

The step size τ_B's ($B = U, V$, and C) in steps 2-4 of Algorithm 1 are selected via a backtracking line search using Armijo's rule as follows. We fix a parameter $\beta \in (0,1)$ and start with $\tau_B = 1$. We continuously update τ_B by $\beta\tau_B$ until it holds that

$$F_Y(B - \tau_B \nabla_B F_Y) \leq F_Y(B) - \frac{\tau_B}{2}\|\nabla_B F_Y\|_2^2. \quad (22)$$

Lines 2-6 of Algorithm 1 can be equivalently written in the form of proximal regularization as follows.

$$\begin{aligned} U^{t+1} &\in \text{prox}_{\tau_U}(U^t - \tau_U \nabla_U F_Y(U^t V^{t^T} + C^t)), \\ V^{t+1} &\in \text{prox}_{\tau_V}(V^t - \tau_V \nabla_V F_Y(U^{t+1} V^{t^T} + C^t)), \\ C^{t+1} &\in \text{prox}_{\tau_C}(C^t - \tau_C \nabla_C F_Y(U^{t+1} V^{t+1^T} + C^t)), \end{aligned} \quad (23)$$

where the proximal map prox_{τ_B} is defined as

$$\text{prox}_{\tau_B}(B^t - \tau_B \nabla_B F_Y(B^t)) := \text{argmin}_B\{(\langle B - B^t, \nabla_B F_Y(B^t)\rangle) + \frac{1}{2\tau_B}\|B - B^t\|_F^2 + K(C)\}. \quad (24)$$

$K(C)$ in (24) can be achieved by keeping s entries with the largest absolute values and setting others to be zero (Proposition 4 in [5]), which correspond to lines 5-6.

Next we prove the convergence of Algorithm 1. Our algorithm is a special case of Proximal Alternating Linearized Minimization (PALM) algorithms. The convergence of PALM algorithms have been proved in [5]. Based on Proposition 3 in [5], if we can show that ∇F_Y is Lipschitz continuous, and $F_Y(UV^T + C) + K(C)$ satisfies the Kurdyka-Lojasiewicz (KL) property, then our algorithm converges to a critical point of (19) from every initial point. A critical point is a value in the domain where all partial derivatives are zero. Details of KL property and its many applications can be found in [5], [23].

$K(C)$ is an indicator function of semi-algebraic sets. Therefore, it is a KL function according to [5]. Since $F_Y(UV^T + C)$ is differentiable everywhere, or equivalently, real analytic, $F_Y(UV^T + C)$ also has the KL property according to the examples in session 2.2 of [37]. Thus, $F_Y(UV^T + C) + K(C)$ satisfies the KL property. The probability density function of the normal distribution and its derivatives are all bounded. From the definition in (4), $1 > f_l \geq \beta > 0$ for some small positive β. Then f_l^{-1} is bounded. One can check that the gradient of the analytic function ∇F_Y is bounded. Thus, ∇F_Y is Lipschitz continuous. Combining the above arguments, we conclude that every sequence generated by Algorithm 1 converges to a critical point. One can check that the computational complexity for each iteration of Algorithm 1 is in the order of $O(mnr)$.

We remark that Algorithm 1 can be readily revised to handle the infinity norm constraints following the projection step in [3]. Specifically, after line 6, multiply U^{t+1} and V^{t+1} by $\sqrt{\alpha/\|U^{t+1}(V^{t+1})^T\|_\infty}$ if $\|U^{t+1}(V^{t+1})^T\|_\infty > \alpha$. Similarly, multiple C^{t+1} by $\alpha/\|C^{t+1}\|_\infty$ if $\|C^{t+1}\|_\infty > \alpha$. We find that adding these constraints do not change the results much numerically. With the thresholding step of $\|UV^T\|$, the resulting algorithm is no longer a special case of PALM, and the above convergence analysis does not hold any more. The other alternative to handle the infinity norm constraints is to convert these constraints into log-barrier functions in the objective as in [3], when $C^* = 0$. [3] then develops a gradient descent algorithm and establishes the convergence. Here, due to the thresholding step (lines 5-7 in Algorithm 1) to obtain a sparse C, the objective function is not differentiable, and the convergence analysis in [3] does not apply here. Since the log-barrier function is not Lipschitz differentiable, the above analysis based on PALM does not apply here either.

IV. Extension of Multiple Copies of Noisy Quantized Measurements

We here consider the general case that W copies of measurements are generated based on M_{ij}^* by adding W copies of independent noise separately and applying quantization. In order to solve (P2), we define

$$\bar{F}_{\bar{Y}}(X) = -\sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{l=1}^{K}(\sum_{t=1}^{W}\mathbf{1}_{[Y_{ij}^t = l]}) \cdot \log(f_l(L_{ij} + C_{ij})), \quad (25)$$

where $\bar{Y} = [Y^1, ..., Y^W]$. We estimate (L^*, C^*) by $(\hat{L}, \hat{C})$, where

$$(\hat{L}, \hat{C}) = \arg\min_{L+C \in \mathcal{S}_f} \bar{F}_{\bar{Y}}(X). \quad (26)$$

The recovery accuracy of (26) is characterized as follows.

Theorem 3. *Same assumptions as Theorem 1. With probability at least $1 - C_1 e^{-C_2 m/W}$, for some positive constants C_1 and C_2, any global minimizer $(\hat{L}, \hat{C})$ of (26) satisfies*

$$\frac{\|(\hat{L} + \hat{C}) - (L^* + C^*)\|_F}{\sqrt{mn}} \leq \min(2\alpha(1 + \sqrt{\frac{s}{mn}}), U'_\alpha), \quad (27)$$

where

$$U'_\alpha = \max(\frac{8.04 L_\alpha \sqrt{2r}}{\gamma_\alpha \sqrt{nW}}, \sqrt{\frac{16.08\alpha L_\alpha \sqrt{2rms} + 8\alpha L_\alpha s}{\gamma_\alpha \sqrt{W} mn}}), \quad (28)$$

Furthermore, with the same probability, it holds that

$$\frac{\|\hat{L} - L^*\|_F}{\sqrt{mn}} \leq \min(2\alpha(1 + 2\sqrt{\frac{s}{mn}}), U'_\alpha + 2\alpha\sqrt{\frac{s}{mn}}). \quad (29)$$

Comparing Theorems 1 and 3, the recover error is reduced by a ratio of $1/\sqrt{W}$ by having W copies of measurements instead of one. Specifically, when s is no greater than $\Theta(m)$,

$$\|(\hat{L} + \hat{C}) - (L^* + C^*)\|_F/\sqrt{mn} \leq O(\sqrt{2r/nW}), \quad (30)$$

and

$$\|\hat{L} - L^*\|_F/\sqrt{mn} \leq O(\sqrt{2r/nW}). \quad (31)$$

The probability in Theorem 3 increases to 1 at a slower rate if W increases. From numerical experiments in Section VI, one can see that W being a small constant is sufficient to reduce the recovery error in practice. Thus, increasing W would not affect the probability much.

We also characterize the information-theoretic limit of the recovery error from unquantized measurements with Gaussian noise as follows.

Theorem 4. *Let $N^t \in \mathbb{R}^{m\times n}$ for $t \in [W]$ contain i.i.d. entries from $\mathcal{N}(0,\sigma^2)$. Assume* (16) *holds. Consider any algorithm that, for any $M \in \mathcal{S}_f$, takes $Y^t = M + N^t$ for all $t \in [W]$ as the input and returns $\hat{M}$. Then there exists $M \in \mathcal{S}_f$ such that with probability at least $\frac{3}{4}$,*

$$\frac{\|\hat{M}-M\|_F}{\sqrt{mn}} \geq \min(C_3, \frac{C_4\sigma}{\sqrt{W}}\sqrt{\frac{rm - r\lfloor\frac{s}{n}\rfloor - 64}{mn - n\lfloor\frac{s}{n}\rfloor}}) \quad (32)$$

holds for some fixed constants C_3 and C_4, where $C_3 = \sqrt{\frac{1-2C_0}{8}}\alpha$ and $C_4 < \sqrt{\frac{1-2C_0}{256}}$.

From (32), if s is no more than $\Theta(mn)$, we have

$$\|\hat{M}-M\|_F/\sqrt{mn} \geq \Theta(\sqrt{r/nW}). \quad (33)$$

The minimum recovery error from unquantized noisy measurements by any method is at least in the order of $\Theta(\sqrt{\frac{r}{nW}})$. From (30), the recovery error from quantized measurements is $O(\sqrt{\frac{r}{nW}})$ when s is at most $\Theta(m)$. Combining (30) and (33), we know that the recovery error by (26) is order optimal as long as s is no greater than $\Theta(m)$. (26) can also be solved by an approximate projected gradient method. One only needs to replace $F_Y(X)$ with $\bar{F}_{\bar{Y}}(X)$ in Algorithm 1 and others remain the same. We skip the details of the resulting algorithm.

V. Quantized Matrix Completion with Erroneous Measurements

We here study (P3) where only measurements with indices in Ω are available for data recovery. We follow the assumptions proposed in [4] and used in [3]. $\mathcal{G} = (V, E)$ is a bipartite graph with $V = \{1, ..., m\} \cup \{1, ..., n\}$, and $(i, j) \in E$ if and only if $(i, j) \in \Omega$. G denotes the bi-adjacency matrix of $\mathcal{G}$ with $G_{ij} = 1$ for all $(i, j) \in \Omega$. We assume each row of G has d nonzero entries. $\sigma_1(G)$ and $\sigma_2(G)$ denote the largest and the second largest singular values of G, respectively. We assume $\sigma_1(G) \geq d$ and $\sigma_2(G) \leq Q\sqrt{d}$ where Q is a positive constant.

The above assumption is more general than the uniform sampling assumption. As discussed in [4], an Erdős-Rényi random graph with an average degree of $C\log(d)$ satisfies it with high probability. The sampling model also applies to stochastic block models for certain choices of edge connection probabilities that result in nonuniform sampling. Based on this sampling model, let $p = \frac{|\Omega|}{mn}$, we have the following result.

Theorem 5. *Suppose that $\frac{m}{n} = \kappa \geq 1$, and $f_l(x)$ is strictly log-concave in x, $\forall l \in [K]$. Then, when $\gamma_\alpha > 0$, with probability at least $1 - 2(9\alpha\sqrt{mn})^{-r(m+n+1)} - C_1e^{-C_2m}$, any global minimizer $\hat{L}$ of (7) satisfies*

$$\|\hat{L}-L^*\|_F/\sqrt{mn} \leq \min(2\alpha_1, U_1, U_2), \quad (34)$$

where

$$\begin{aligned} U_1 &= C_1'\frac{\kappa r\sqrt{r}}{p^2\sqrt{n}} + C_2'\frac{r\kappa^{1/2}s^{1/4}}{p^{3/2}n^{3/4}} + C_3'\frac{\sqrt{rs}}{pn} \\ U_2 &= D_1'\frac{(\kappa r)^{1/2}(r\log(\sqrt{\kappa}n))^{1/4}}{p^{3/4}n^{1/4}} + D_2'\frac{\sqrt{rs}}{pn}, \end{aligned} \quad (35)$$

for some positive constants C_1, C_2, C_1', C_2', C_3', D_1', and D_2'.

Theorem 5 shows that the error bound depends on p, r, n, and s. For example, when $p = \Theta(n^{-5/16}(\log n)^{-3/8})$, U_2 is smaller than U_1. If s is at most $O(nr^2)$, the recovery error is bounded by $O(r^{3/4}(\log n)^{17/32}n^{-1/64})$, which decays to zero as n increases. When p is a fixed constant, which means that $|\Omega|$ is in the order of mn, U_1 in Theorem 5 is in the order of $O(\sqrt{\frac{r^3}{n}})$, as long as s is no greater than $\Theta(nr^2)$. The first term of U_2 is $O(\sqrt[4]{\frac{r^3\log(n)}{n}})$. The second term of U_2 and the third term of U_1 are in the same order. Note that if we consider the case of full observations by setting $p = 1$ in Theorem 5, the resulting error bound of $O(\sqrt{\frac{r^3}{n}})$ is not as tight as the bound $O(\sqrt{\frac{r}{n}})$ in Theorem 5. This results from the relaxation in the proof of Theorem 5 to handle partial observations.

We also compare Theorem 5 with existing quantized matrix completion methods that considered no corruptions, i.e., $C^* = 0$. In this case, (34) is reduced to

$$\frac{\|\hat{L}-L^*\|_F}{\sqrt{mn}} \leq \min\left(2\alpha_1, O(\frac{1}{p^2}\sqrt{\frac{r^3}{n}}), O(\frac{1}{p^{\frac{3}{4}}}\sqrt[4]{\frac{r^3}{n}})\right) \quad (36)$$

The error bound of data recovery from 1-bit measurements is shown to be

$$\frac{\|\hat{L}-L^*\|_F}{\sqrt{mn}} \leq O(\sqrt[4]{\frac{r}{pn}}), \quad (37)$$

when the constraints are imposed by nuclear norm [13] and max-norm [7] respectively. Using the rank constraint, Ref. [3] obtained an error bound of

$$\frac{\|\hat{L}-L^*\|_F}{\sqrt{mn}} \leq \min(O(\frac{1}{p^2}\sqrt{\frac{r^3}{n}}), O(\sqrt[4]{\frac{r^3\log n}{p^3n}})) \quad (38)$$

when $m = \Theta(n)$. Refs. [21], [24], [34] obtained an error bound of

$$\frac{\|\hat{L}-L^*\|_F}{\sqrt{mn}} \leq O(\sqrt{\frac{r\log m}{pn}}). \quad (39)$$

Comparing (36) with (37)-(39), one can see that if m and n are fixed, (39) by [21], [24], [34] is the tightest bound in terms of p and r, and our bound in (36) is inferior to the one in (39). On the other hand, when both p and r are constants, (39) by [21], [24], [34] is reduced to $O(\sqrt{\frac{\log m}{n}})$. Our bound in (36) and the bound in (38) by [3] are both reduced to $O(\sqrt{\frac{1}{n}})$, which decays slightly faster than $O(\sqrt{\frac{\log m}{n}})$.

One can easily modify Algorithm 1 to handle partial observations. One simply needs to replace $F_Y(X)$ and $\nabla_X F_Y(X)$ with $F_{\Omega,Y}(X)$ and $\nabla_{\Omega,X}F_Y(X)$, respectively, where $[F_{\Omega,Y}(X)]_{ij} = 0$, $[\nabla_X F_{\Omega,Y}(X)]_{ij} = 0$ if $(i,j) \notin \Omega$, and $[F_{\Omega,Y}(X)]_{ij} = F_Y(X)]_{ij}$, $[\nabla_X F_{\Omega,Y}(X)]_{ij} = [\nabla_X F_Y(X)]_{ij}$ if $(i,j) \in \Omega$. We skip the details here.

VI. Simulation

We explore the performance of our method on both synthetic data and actual PMU data from the Central New York (NY) Power System. The recovery performance is measured by the relative recovery error $\|L^* - \hat{L}\|_F^2/\|L^*\|_F^2$, where L^* denotes the actual data, and $\hat{L}$ denotes the recovered data. The corruption rate (s/mn) denotes the fraction of nonzero entries in C^*. tol is set to be 10^{-4} in our method. The simulations run in MATLAB on a computer with 3.4 GHz Intel Core i7.

A. Performance on synthetic data

We first set $m = n = 200$, and $r = 3$. We construct $L^* \in \mathbb{R}^{m\times n}$ as $L^* = AB^T$, where $A \in \mathbb{R}^{m\times r}$ and $B \in \mathbb{R}^{n\times r}$ are matrices with entries drawn i.i.d. from uniform distribution on $[0, 1]$. We then scale L^* such that $\|L^*\|_\infty = 1$. The locations of nonzero entries of the sparse matrix C^* are randomly selected. The entries of N are drawn i.i.d. from $\mathcal{N}(0, 0.18^2)$. K is set to be 5, where $\omega_1 = -0.3$, $\omega_2 = 0$, $\omega_3 = 0.5$, and $\omega_4 = 1.2$.

We compare the recovery performance of our method with the quantized robust PCA (QRPCA) in [25], and a simple method that uses the middle number of two adjacent bin boundaries as the estimation. Nonzero entries of sparse matrix are uniformly selected from $[-0.5, -0.2]$ and $[0.2, 0.5]$. Both r and s are set to be the actual rank and the number of corruptions in our method. The results are averaged over 100 runs. [25] shows that QRPCA achieves a smaller recovery error when the estimation is rescaled by $\|L^*\|_F$. Since $\|L^*\|_F$ is usually unknown in practice, we test QRPCA both with and without rescale. As shown in Fig. 2 (a), our method has better performance compared with the other methods.

We then test our method with larger corruptions. We here select the nonzero entries uniformly from $[-10, 10]$. Fig. 2 (b) shows the box-plot-diagram of relative recovery error with 100 runs. The tops and bottoms of each "box" are the 25th and 75th percentiles of the samples respectively. Our method can handle large corruptions with a very small recovery error.

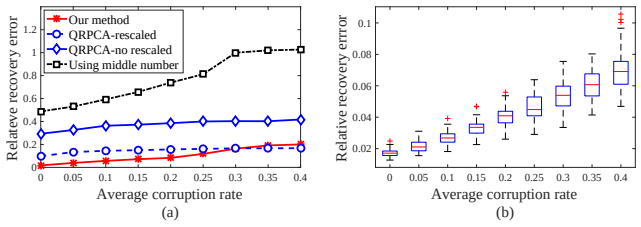


Fig. 2. (a) Relative recovery error of a 200×200 synthetic data. (b) Box-plot-diagram of relative recovery error with 100 runs.

We also test our method when varying the problem setup. Fig. 3 (a) shows that the recovery error of our method decreases when the dimension of matrix increase. When n increases from 200 to 800, the error is reduced by more than half. This coincides with our analytical error bound of $O(\sqrt{r/n})$. In terms of the computational time, when the matrix dimension is 8000×8000, the running time of our method for one test is 5000 seconds on average, while QRPCA takes 30000 seconds. Fig. 3 (b) shows relative recovery error under different noise levels. The noise is drawn from $\mathcal{N}(0, \sigma^2)$ with σ changing from 0.15 to 0.45. We also vary the matrix rank while keeping the matrix size. Let $L^* = AB^T$, and the entries of $A, B \in \mathbb{R}^{500\times r}$ are drawn independently from standard Gaussian distribution. L^* is then scale to be in the range of $[-1, 1]$. K is set to be 5, where $\omega_1 = -0.8$, $\omega_2 = -0.2$, $\omega_3 = 0.2$, and $\omega_4 = 0.8$. Fig. 3 (c) shows relative recovery error when the matrix rank r changes from 1 to 15.

We next test the detection performance when C^* is row-sparse, i.e., a few rows have all nonzero entries. This simulates the scenario that some sensor constantly produce bad data

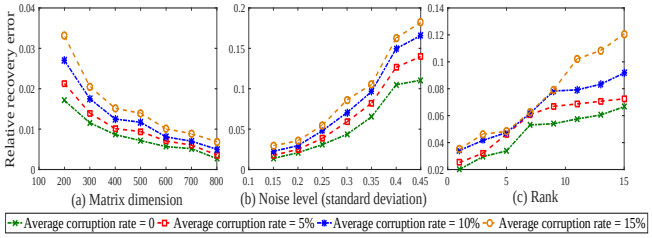


Fig. 3. (a) Relative recovery error when matrix dimension changes. $r = 3$. (b) Relative recovery error when the noise level changes. Matrix dimension 200×200. $r = 3$. (c) Relative recovery error when the matrix rank changes. Matrix dimension 500×500.

due to either device malfunction or cyber attacks. We can locate the affected channels if the identified corruption rate is significantly larger than others. The identified corruption rate in a row is the fraction of locations where $\hat{C}$ has nonzero entries. We randomly select four rows, rows 6, 90, 117, and 134, and pick nonzero values uniformly from $[-10, 10]$ for each entry of these rows of C^*. Fig. 4 (a) shows the identified corruption rate of each row. The identified corruption rate of these four rows are much larger than others. Thus, the corrupted rows can be correctly identified even though the measurements are highly noisy and quantized. The remaining rows can be accurately recovered. We also vary the number of corrupted rows. For each fixed number, we randomly selected rows that are corrupted. All the setups are the same as before. The relative recovery error of uncorrupted rows with different numbers of corrupted rows is shown in Fig. 4 (b). Each result is averaged over 100 runs.

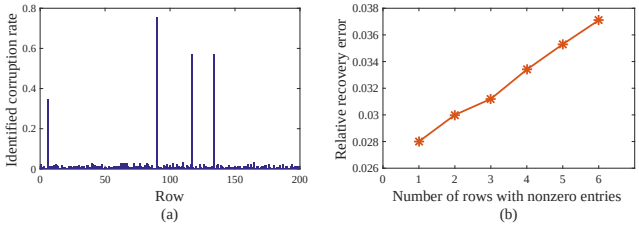


Fig. 4. (a) Identified corruption rate of each row when the corrupted rows are 6, 90, 117, and 134. (b) Relative recovery error of uncorrupted rows when the number of corrupted rows changes.

We then vary K and keep other setups the same. When $K = 2$, we set $\omega_0 = -\infty$, $\omega_1 = 0$, and $\omega_2 = \infty$. Note that the measurement is binary and can be represented in a single bit. When $K = 3$, we set $\omega_0 = -\infty$, $\omega_1 = 0.33$, $\omega_2 = 0.66$, and $\omega_3 = \infty$. When $K = 4$, we set $\omega_0 = -\infty$, $\omega_1 = 0.25$, $\omega_2 = 0.5$, $\omega_3 = 0.75$, and $\omega_4 = \infty$. When $K = 5$, we set $\omega_0 = -\infty$, $\omega_1 = 0.2$, $\omega_2 = 0.4$, $\omega_3 = 0.6$, $\omega_4 = 0.8$, and $\omega_5 = \infty$. Fig. 5 (a) shows the recovery performance of our method according to different K. The results are averaged over 100 runs. We can see that the recovery error reduces significantly in the high-corruption-rate region when K increases from 2 to 3. That is because when $K = 2$ and $W = 1$, binary measurements do not provide enough information to handle a large amount of corruptions.

This suggests that we need at least 3-level quantization when corruption rate is larger than 10%. The recovery performance does not change much when K increases from 4 to 5.

We then consider the general case that $W > 1$. We set $K = 2$ such that each quantized measurement is binary. We keep other parameters as before and increase the number of binary measurements W for each M^*_{ij}. When $W = 1$, it reduces to the binary case, which is the same as $K = 2$ in Fig. 5 (a). Fig. 5 (b) shows the recovery performance of our method with different W. The results are averaged over 100 runs. We can see that the recovery performance is improved with the increase of W.

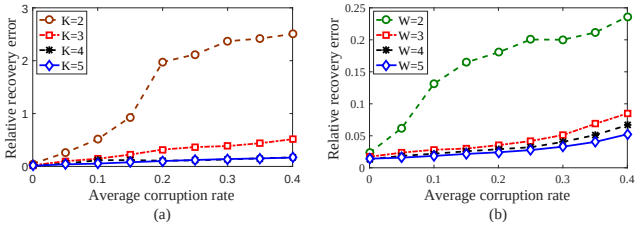


Fig. 5. (a) Recovery performance of our method according to different K. (b) Recovery performance of our method according to different W ($K = 2$).

We also test the data completion performance under different data loss rates. Fig. 6 shows the relative recovery error when the fraction of available observations changes. 10% of available observations are corrupted.

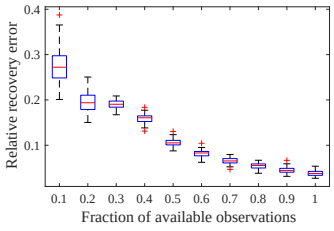


Fig. 6. Relative recovery error according to different number of observations

B. Performance on actual PMU data

We test our method on actual PMU data from the Central New York power system. The PMU data contains five seconds of current magnitudes at 24 different locations with a rate of 30 samples per second. The dimension of the data matrix is 24×151. We normalize the current magnitudes into $[0, 1]$. We generate C^* in the same way as for the synthetic data. The noise N_{ij} is drawn i.i.d. from $\mathcal{N}(0, 0.18^2)$.

We first consider $W = 1$, and $K = 5$ with quantization boundaries $\omega_0 = -\infty$, $\omega_1 = -0.3$, $\omega_2 = 0$, $\omega_3 = 0.5$, $\omega_4 = 1.2$ and $\omega_5 = \infty$. We set $r = 3$ in solving (7). Fig. 7 shows the original data, the quantized value with 5% of corruptions, the recovered data and the data after filtering of channels 3 and 21. Fig. 8 shows the results when we increase the noise to $\mathcal{N}(0, 0.3^2)$). One can see that the details of the time series are masked in the quantized measurements. The recovered data are noisier than the actual data because we add noise to each channel with a noise level comparable to the signal level before quantization. Still, the overall trend of the time series are correctly recovered. If needed, the noise in the recovered data could be removed by applying a low-pass filter. We apply a moving average filter to the recovered data with a window size of 9 in our simulation.

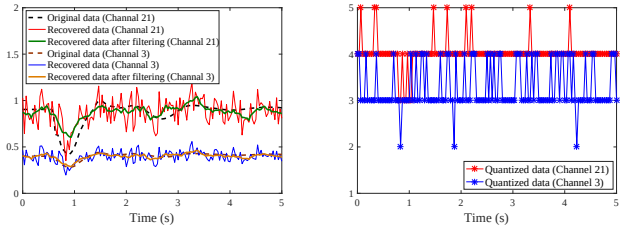


Fig. 7. Original, quantized, recovered data of channels 3 and 21 ($\mathcal{N}(0, 0.18^2)$)

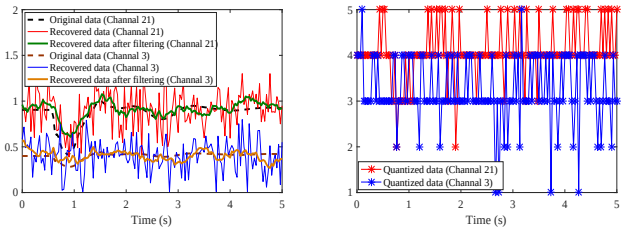


Fig. 8. Original, quantized, recovered data of channels 3 and 21 ($\mathcal{N}(0, 0.3^2)$)

Fig. 9 (a) compares our method with using middle number and QRPCA method. The relative recovery error in Fig. 9 (a) is averaged over 100 tests. Our method has the best performance among these methods. Fig. 9 (b) shows the box-plot-diagram of our relative recovery error with 100 runs.

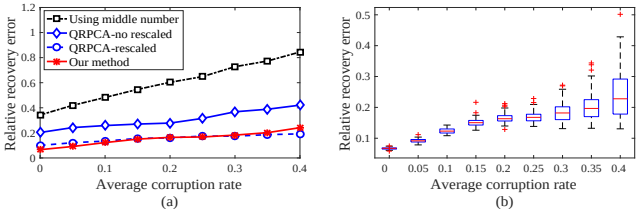


Fig. 9. (a) Relative recovery error on actual PMU data. (b) Box-plot-diagram of relative recovery error with 100 runs on PMU data.

Fig. 10 shows relative recovery error under different noise levels. The y-axis is on a logarithmic scale. The noise is drawn from $\mathcal{N}(0, \sigma^2)$ with σ changing from 0.15 to 0.45. We also vary the average corruption rate. The relative error of the measurements before quantization ($\frac{\|N\|_F}{\|L^*\|_F}$) is included for comparison. The recovery error from the quantized measurements is much less than the error in the original noisy measurements.

We next show the recovery performance by increasing K and W, respectively. We want to study with the same communication cost, which one, K or W, the user should increase to reduce the recovery error. Fig. 11 compares the recovery performance for different K and W. The measurements when $W = 2$, $K = 2$ and when $W = 1$, $K = 4$ both can be represented by 2 bits. The measurements when $W = 3$, $K = 2$ and when $W = 1$, $K = 8$ both can be represented by 3 bits. Increasing K and W can improve the recovery performance.

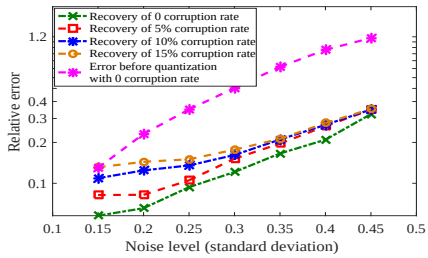


Fig. 10. Relative error when the noise level changes

When the noise level is relatively high, increasing K leads to a smaller recovery error compared with increasing W. That is in this region, increasing the quantization level could provide more information about the original data compared with multiple copies of highly noisy measurements.

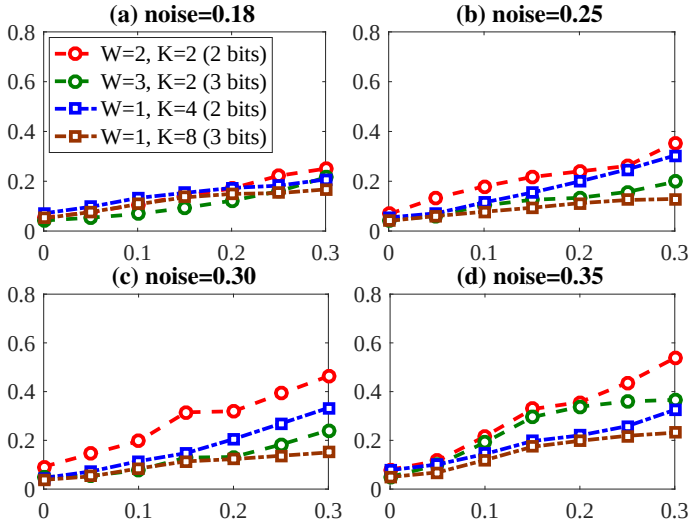


Fig. 11. Comparison of varying W and K with the same communication cost

VII. Conclusion and Discussions

This work provides the first theoretical study of low-rank matrix recovery noisy, quantized, partially corrupted, and partially lost measurements. Exploiting the low-rank property of the data, the recovery error of our developed method diminishes in the same order as that of the best possible recovery method with unquantized measurements. This paper also extends to a more general setup where a ground-truth value could generate multiple copies instead of one copy of the quantized measurements. Proximal algorithms with convergence guarantees are developed and numerically evaluated. We then propose a communicated-reduced, cyber-resilient, and information-preserved data collection framework with applications in synchrophasor data management in power systems. The noisy and quantized measurements reduce data transmission and protect the privacy of individual data sources, while the operator can exploit the proposed method to accurately recover data.

Acknowledgement

This research is supported in part by NSF 1508875, ARO W911NF-17-1-0407, IBM Corporation, and the ERC Program of NSF and DoE under the supplement to NSF Award EEC-1041877 and the CURENT Industry Partnership Program. We thank New York Power Authority for providing historical PMU datasets.

Appendix

A. Discussion on γ_α and L_α based on logistic model

Here we consider γ_α and L_α for the logistic noise with $\Phi_{\log}(x) = 1/(1+e^{-x})$. We consider the case that $\omega_1 = -\zeta$ and $\omega_{K-1} = \zeta$ for the constant $\zeta \geq 0$. Note that $\omega_0 = -\infty$, and $\omega_K = \infty$. From (18) in [3], we have

$$\frac{\dot{f}_l^2(x)}{f_l^2(x)} - \frac{\ddot{f}_l(x)}{f_l(x)} = \frac{1}{\sigma^2}[\Phi_{\log}(\frac{\omega_l - x}{\sigma})(1-\Phi_{\log}(\frac{\omega_l - x}{\sigma})) + \Phi_{\log}(\frac{\omega_{l-1} - x}{\sigma})(1-\Phi_{\log}(\frac{\omega_{l-1} - x}{\sigma}))]. \tag{40}$$

When $l = K$, we have

$$\frac{\dot{f}_K^2(x)}{f_K^2(x)} - \frac{\ddot{f}_K(x)}{f_K(x)} = \frac{1}{\sigma^2}[\Phi_{\log}(\frac{\zeta - x}{\sigma})(1-\Phi_{\log}(\frac{\zeta - x}{\sigma}))] \geq \frac{1}{\sigma^2}[\Phi_{\log}(\frac{\zeta + 2\alpha}{\sigma})(1-\Phi_{\log}(\frac{\zeta + 2\alpha}{\sigma}))] \tag{41}$$

for all x, $|x| \leq 2\alpha$, and the equality holds when $x = -2\alpha$. Note that (41) also holds if we replace with $l = 1$. One can easily check that (40) is larger than the right hand side of (41) for any $|x| \leq 2\alpha$ when $l \in \{2, ..., K-1\}$. We then have

$$\gamma_\alpha = e^{-\frac{\zeta+2\alpha}{\sigma}}/(\sigma^2(1+e^{-\frac{\zeta+2\alpha}{\sigma}})^2). \tag{42}$$

γ_α is maximized when $\zeta = 0$. In this case, K should be 2.

In order to analyze L_α, note that

$$\frac{|\dot{f}_l(x)|}{f_l(x)} = \frac{1}{\sigma}\frac{|\dot{\Phi}_{\log}(\frac{\omega_{l-1}-x}{\sigma}) - \dot{\Phi}_{\log}(\frac{\omega_l - x}{\sigma})|}{\Phi_{\log}(\frac{\omega_l - x}{\sigma}) - \Phi_{\log}(\frac{\omega_{l-1}-x}{\sigma})}. \tag{43}$$

We define a function g of $x_1, x_2 \geq 0$ such that

$$g(x_1, x_2) = \frac{\dot{\Phi}_{\log}(x_1) - \dot{\Phi}_{\log}(x_1 + x_2)}{\Phi_{\log}(x_1 + x_2) - \Phi_{\log}(x_1)}. \tag{44}$$

We then have

$$\frac{\partial g(x_1, x_2)}{\partial x_1} = \frac{\ddot{\Phi}_{\log}(x_1) - \ddot{\Phi}_{\log}(x_1 + x_2)}{\Phi_{\log}(x_1 + x_2) - \Phi_{\log}(x_1)} + \frac{(\dot{\Phi}_{\log}(x_1) - \dot{\Phi}_{\log}(x_1 + x_2))^2}{(\Phi_{\log}(x_1 + x_2) - \Phi_{\log}(x_1))^2}. \tag{45}$$

Note that $\Phi_{\log}(x_1 + x_2) - \Phi_{\log}(x_1)$ is log-concave in x_1 such that

$$(\dot{\Phi}_{\log}(x_1 + x_2) - \dot{\Phi}_{\log}(x_1))^2 - (\ddot{\Phi}_{\log}(x_1 + x_2) - \ddot{\Phi}_{\log}(x_1)) (\Phi_{\log}(x_1 + x_2) - \Phi_{\log}(x_1)) \geq 0. \tag{46}$$

We then have $\frac{\partial g(x_1, x_2)}{\partial x_1} \geq 0$, $\forall x_1, x_2 \geq 0$. From some tedious calculations, one could also verify that $\frac{\partial g(x_1, x_2)}{\partial x_2} \geq 0$,

$\forall x_1, x_2 \geq 0$. We then have $g(x_1, x_2)$ is a monotone increasing function in both x_1 and x_2, $\forall x_1, x_2 \geq 0$. We have that

$$L_\alpha = \max_{l\in[K]} \sup_{|x|\leq 2\alpha} \{\frac{|\dot{f}_l(x)|}{f_l(x)}\} = \frac{1}{\sigma}\frac{\dot{\Phi}_{\log}(\frac{\zeta+2\alpha}{\sigma})}{1-\Phi_{\log}(\frac{\zeta+2\alpha}{\sigma})} \\ = 1/(\sigma(1+e^{-\frac{\zeta+2\alpha}{\sigma}})). \tag{47}$$

L_α is minimized when $\zeta = 0$. In this case, K should be 2. Thus, L_α/γ_α is minimized when $K = 2$ for the logistic model.

B. Proof sketch of Theorem 1

The proof of Theorem 1 follows the same line as the proof of Theorem 3.1 in [1] but is more involved due to the additional sparse matrix C^*.

Let $\theta = \mathrm{vec}(X) \in \mathbb{R}^{mn}$ and $\mathcal{F}_Y(\theta) = F_Y(X)$. By the second-order Taylor's theorem, we have

$$\mathcal{F}_Y(\theta) = \mathcal{F}_Y(\theta^*) + \langle \nabla_\theta \mathcal{F}_Y(\theta^*), \theta - \theta^* \rangle \\ + \frac{1}{2}\left\langle \theta - \theta^*, (\nabla^2_{\theta\theta}\mathcal{F}_Y(\tilde{\theta}))(\theta - \theta^*) \right\rangle, \tag{48}$$

where $\tilde{\theta} = \theta^* + \eta(\theta - \theta^*)$ for some $\eta \in [0, 1]$, with corresponding matrices $\tilde{X} = M^* + \eta(X - M^*)$. Note that for the matrices A and B, we have $\langle A, B\rangle := \mathrm{tr}(\mathrm{A^T B})$, and $|\langle A, B\rangle| \leq \|A\|_2 \|B\|_*$.

Lemma 1. *Let $\theta' = \mathrm{vec}(X')$, $\theta^* = \mathrm{vec}(M^*)$, and X', $M^* \in \mathcal{S}_f$. Then with probability at least $1 - C_1 e^{-C_2 m}$,*

$$|\langle \nabla_\theta \mathcal{F}_Y(\theta^*), \theta' - \theta^* \rangle| \leq 2.01 L_\alpha \sqrt{2rm}(\|X' - M^*\|_F \\ + 2\alpha\sqrt{s}) + 2\alpha s L_\alpha, \tag{49}$$

holds for the positive constants C_1 and C_2.

Please refer to the Appendix-C for the proof of Lemma 1.

Lemma 2 (Lemma A.3 in [1])**.** *Let $\theta' = \mathrm{vec}(X')$, $\theta^* = \mathrm{vec}(M^*)$, and X', $M^* \in \mathcal{S}_f$. Then for any $\tilde{\theta} = \theta^* + \eta(\theta' - \theta^*)$ and any $\eta \in [0, 1]$, we have*

$$\left\langle \theta' - \theta^*, (\nabla^2_{\theta\theta}\mathcal{F}_Y(\tilde{\theta}))(\theta' - \theta^*) \right\rangle \geq \gamma_\alpha \|X' - M^*\|_F^2. \tag{50}$$

Proof of Theorem 1: The first bound follows from the fact that $\hat{M}$, $M^* \in \mathcal{S}_f$. Note that

$$\frac{\|\hat{M} - M^*\|_F}{\sqrt{mn}} \leq \frac{\|\hat{L} - L^*\|_F}{\sqrt{mn}} + \frac{\|\hat{C} - C^*\|_F}{\sqrt{mn}} \\ \leq 2\alpha + \frac{2\alpha\sqrt{s}}{\sqrt{mn}} = 2\alpha(1 + \sqrt{\frac{s}{mn}}). \tag{51}$$

We discuss the second bound as follows. Note that the objective function $F_Y(X)$ is continuous in X, and the set $\mathcal{S}_f$ is compact. $F_Y(X)$ then achieves a minimum in $\mathcal{S}_f$. Suppose that $\hat{M} \in \mathcal{S}_f$ minimizes $F_Y(X)$. Combining (48), Lemma 1 and Lemma 2, we have that

$$F_Y(\hat{M}) \geq F_Y(M^*) - 2.01 L_\alpha \sqrt{2rm}\|\hat{M} - M^*\|_F \\ - 4.02\alpha L_\alpha \sqrt{2rms} - 2\alpha s L_\alpha + \frac{\gamma_\alpha}{2}\|\hat{M} - M^*\|_F^2 \tag{52}$$

holds with probability at least $1 - C_1 e^{-C_2 m}$. Note that $F_Y(\hat{M}) \leq F_Y(M^*)$. We then have

$$\frac{\gamma_\alpha}{2}\|\hat{M} - M^*\|_F^2 \leq 2.01 L_\alpha \sqrt{2rm}\|\hat{M} - M^*\|_F \\ + 4.02\alpha L_\alpha \sqrt{2rms} + 2\alpha s L_\alpha \tag{53}$$

holds with probability at least $1 - C_1 e^{-C_2 m}$. Note that $x \leq \max(2a, 2b)$ holds if $x \leq a + b$. We then have with the same probability

$$\|\hat{M} - M^*\|_F / \sqrt{mn} \leq U_\alpha. \tag{54}$$

C. Proof of Lemma 1

Proof: We need the following result for the proof.

Lemma 3 (Lemma A.1 of [1])**.** *Take any two numbers m and n such that $1 \leq n \leq m$. Suppose that $A = [a_{ij}]_{1\leq i\leq m, 1\leq j\leq n}$ is a matrix whose entries are independent random variables that satisfy, for some $\sigma^2 \in [0, 1]$,*

$$\mathbb{E}[a_{ij}] = 0, \quad \mathbb{E}[a_{ij}^2] \leq \sigma^2, \quad \textit{and} \quad |a_{ij}| \leq 1 \quad a.s. \tag{55}$$

Suppose that $\sigma^2 \geq m^{-1+\epsilon}$ for some $\epsilon > 0$. Then

$$P(\|A\|_2 \geq 2.01\sigma\sqrt{m}) \leq C_1(\epsilon) e^{-C_2 \sigma^2 m}, \tag{56}$$

where $C_1(\epsilon)$ is a constant that depends only on ϵ and C_2 is a positive universal constant.

Consider

$$Z_{ij} := [L_\alpha^{-1}\nabla_X F_Y(M^*)]_{ij} = -L_\alpha^{-1}\sum_{l=1}^{K}\frac{\dot{f}_l(M^*_{ij})}{f_l(M^*_{ij})}\mathbf{1}_{[Y_{ij}=l]}.$$

Using (10) and the fact that $\sum_{l=1}^K f_l(X_{ij}) = 1$, we have $\mathbb{E}[Z_{ij}] = 0$, $|Z_{ij}| \leq 1$, and $\mathbb{E}[Z_{ij}^2] \leq 1$. By Lemma 3, we have

$$\|L_\alpha^{-1}\nabla_X F_Y(M^*)\|_2 \leq 2.01\sqrt{m} \tag{57}$$

holds with probability at least $1 - C_1 e^{-C_2 m}$ for some positive constants C_1 and C_2. Note that

$$\|L' - L^*\|_* \leq \sqrt{2r}\|L' - L^*\|_F. \tag{58}$$

We then have

$$\|\nabla_X F_Y(M^*)\|_2 \|L' - L^*\|_* \leq 2.01 L_\alpha \sqrt{2rm}\|L' - L^*\|_F \tag{59}$$

holds with probability at least $1 - C_1 e^{-C_2 m}$. We also have

$$|\langle \nabla_X F_Y(M^*), C' - C^* \rangle| \leq 2\alpha s L_\alpha. \tag{60}$$

Then,

$$|\langle \nabla_\theta \mathcal{F}_Y(\theta^*), \theta' - \theta^* \rangle| = |\langle \nabla_X F_Y(M^*), X' - M^* \rangle| \\ \leq |\langle \nabla_X F_Y(M^*), L' - L^* \rangle| + |\langle \nabla_X F_Y(M^*), C' - C^* \rangle| \\ \leq \|\nabla_X F_Y(M^*)\|_2 \|L' - L^*\|_* + |\langle \nabla_X F_Y(M^*), C' - C^* \rangle| \\ \leq 2.01 L_\alpha \sqrt{2rm}\|L' - L^*\|_F + 2\alpha s L_\alpha \\ \leq 2.01 L_\alpha \sqrt{2rm}(\|X' - M^*\|_F + \|C' - C^*\|_F) + 2\alpha s L_\alpha \\ \leq 2.01 L_\alpha \sqrt{2rm}(\|X' - M^*\|_F + 2\alpha\sqrt{s}) + 2\alpha s L_\alpha$$

holds with probability at least $1 - C_1 e^{-C_2 m}$, where C_1 and C_2 are positive constants. The third inequality follows from (59) and (60). ∎

D. Supporting Lemmas for the Proof of Theorem 2

Here we first introduce two Lemmas that are useful to Theorem 2 and defer the proof of Theorem 2 to Appendix-E. The proofs of these lemmas are similar to Lemmas 3 and 5 in [13] with modifications for our problem setup.

Lemma 4. *There is a set $\mathcal{X} \subset \mathcal{S}_f$ with*

$$|\mathcal{X}| \geq \exp(\frac{rm - r\lfloor \frac{s}{n} \rfloor}{16}) \tag{61}$$

such that for any $\gamma \in (0, 1]$, the following properties hold:

1. *For all $X \in \mathcal{X}$, $|X_{ij}| = 0$ or $\alpha\gamma$, $\forall (i, j)$.*
2. *For all $X^{(i)}$, $X^{(j)} \in \mathcal{X}$, $i \neq j$,*

$$\|X^{(i)} - X^{(j)}\|_F^2 > \alpha^2\gamma^2(\frac{mn}{2} - s). \tag{62}$$

Proof: We will prove the existence of $\mathcal{X}$ by a probabilistic argument. Consider a set $\mathcal{X}$ of $\lceil \exp(\frac{rm - r\lfloor \frac{s}{n} \rfloor}{16}) \rceil$ random matrices independently generated from the following distribution. Each matrix $X \in \mathbb{R}^{m \times n}$ consists of blocks with dimensions $m \times r$. Consider the first block, i.e., X_{ij} for $i \in [m]$ and $j \in [r]$. Fix the locations of $\lfloor \frac{s}{n} \rfloor$ entries in each column and set the values to zero. The remaining $rm - r\lfloor \frac{s}{n} \rfloor$ entries are i.i.d. symmetric random variables taking values $\pm\alpha\gamma$ with equal probabilities. The remaining entries of X are obtained through repetition. Specifically, for all $i \in [m]$, $j \in \{r+1, ..., n\}$,

$$X_{ij} := X_{ij'}, \text{ where } j' = j (\text{mod } r) + 1. \tag{63}$$

Then one can see that X can be written as $X = L + C$, where $\text{rank}(L) \leq r$, and $\sum_{ij} \mathbf{1}_{[C_{ij} \neq 0]} \leq s$. We further have

$$\|L\|_\infty = \alpha\gamma \leq \alpha \text{ and } \|C\|_\infty = \alpha\gamma \leq \alpha. \tag{64}$$

Thus $\mathcal{X} \in \mathcal{S}_f$. It remains to show that $\mathcal{X}$ satisfies property 2.

Note that the locations of the zero entries are the same for all matrices drawn from the above distribution. Consider two different matrices X and $\hat{X}$ drawn as above, we have

$$\begin{aligned} &\|X - \hat{X}\|_F^2 = \sum_{i,j} (X_{ij} - \hat{X}_{ij})^2 \\ &\geq \lfloor \frac{n}{r} \rfloor \sum_{i=1}^{m} \sum_{j=1}^{r} (X_{ij} - \hat{X}_{ij})^2 = 4\alpha^2\gamma^2 \lfloor \frac{n}{r} \rfloor \sum_{i=1}^{rm - r\lfloor \frac{s}{n} \rfloor} \delta_i, \end{aligned} \tag{65}$$

where δ_i's are independent $0/1$ Bernoulli random variables with mean $1/2$. By Hoeffding's inequality and the union bound, we have

$$\begin{aligned} &P(\min_{X \neq \hat{X} \in \mathcal{X}} \|X - \hat{X}\|_F^2 \leq \alpha^2\gamma^2 \lfloor \frac{n}{r} \rfloor (rm - r\lfloor \frac{s}{n} \rfloor) \\ &\leq \binom{|\mathcal{X}|}{2} \exp(-\frac{rm - r\lfloor \frac{s}{n} \rfloor}{8}). \end{aligned} \tag{66}$$

Note that for $\mathcal{X}$ of the size given in (61), the right-hand side of (66) is less than 1. Thus, the event that

$$\|X - \hat{X}\|_F^2 > \alpha^2\gamma^2 \lfloor \frac{n}{r} \rfloor (rm - r\lfloor \frac{s}{n} \rfloor) \geq \alpha^2\gamma^2(\frac{mn}{2} - s) \tag{67}$$

for all X and $\hat{X} \in \mathcal{X}$ such that $X \neq \hat{X}$ has nonzero probability, where the second inequality uses the fact that $\lfloor x \rfloor \geq x/2$ for all $x \geq 1$. ∎

Suppose $X \in \mathcal{X}$ is chosen uniformly at random. Let $Y = X + N$, where N contains i.i.d. Gaussian entries with zero mean and variance σ^2. We bound the mutual information $I(X, Y)$ in the following lemma.

Lemma 5.

$$I(X, Y) \leq \frac{mn - n\lfloor \frac{s}{n} \rfloor}{2} \log(1 + (\frac{\alpha\gamma}{\sigma})^2) \tag{68}$$

Proof: We have

$$\begin{aligned} &I(X, Y) = H(Y) - H(Y|X) \\ &= H(X + N) - H(X + N|X) = H(X + N) - H(N). \end{aligned} \tag{69}$$

Let ξ denote a matrix of i.i.d. ± 1 entries. We have

$$\begin{aligned} H(X \cdot \xi + N) &= H((X + N) \cdot \xi) \\ &\geq H((X + N) \cdot \xi | \xi) = H(X + N), \end{aligned} \tag{70}$$

where $X \cdot \xi$ denotes the entry-wise product of X and ξ. Let $\tilde{X} = X \cdot \xi$. We then have

$$I(X, Y) \leq H(\tilde{X} + N) - H(N). \tag{71}$$

We treat $\tilde{X} + N$ as a vector, denoted by $\text{vec}(\tilde{X} + N)$, of length mn, and compute the covariance matrix as

$$\Sigma := \mathbb{E}[\text{vec}(\tilde{X} + N)\text{vec}(\tilde{X} + N)^T]. \tag{72}$$

By Theorem 8.6.5 in [12], we have

$$\begin{aligned} H(\tilde{X} + N) &\leq \frac{1}{2} \log((2\pi e)^{mn} \det(\Sigma)) \\ &= \frac{1}{2} \log((2\pi e)^{mn} (\alpha^2\gamma^2 + \sigma^2)^{mn - n\lfloor \frac{s}{n} \rfloor} \sigma^{2n\lfloor \frac{s}{n} \rfloor}), \end{aligned} \tag{73}$$

where the last equality holds since $\tilde{X}$ has $n\lfloor \frac{s}{n} \rfloor$ zero entries. We have that $H(N) = \frac{1}{2} \log((2\pi e)^{mn} \sigma^{2mn})$ and so

$$I(X, Y) \leq \frac{1}{2} \log(\frac{(\alpha^2\gamma^2 + \sigma^2)^{mn - n\lfloor \frac{s}{n} \rfloor} \sigma^{2n\lfloor \frac{s}{n} \rfloor}}{\sigma^{2mn}}), \tag{74}$$

which establishes the lemma. ∎

E. Proof of Theorem 2

Choose ϵ so that

$$\epsilon^2 = \min\{\frac{(1 - 2C_0)\alpha^2}{8}, C_4^2\sigma^2 \frac{rm - r\lfloor \frac{s}{n} \rfloor - 64}{mn - n\lfloor \frac{s}{n} \rfloor}\} \tag{75}$$

for a constant C_4 to be determined later. We will consider running any algorithms on a random element in a set $\mathcal{X} \in \mathcal{S}_f$. For our set $\mathcal{X}$, we will use the set, the existence of which is guaranteed by Lemma 4. We will set γ such that

$$\frac{2\epsilon}{\alpha} \sqrt{\frac{2mn}{mn - 2s}} \leq \gamma \leq \frac{2\epsilon}{\alpha} \sqrt{\frac{2}{1 - 2C_0}} \leq 1. \tag{76}$$

This is possible since $s/mn \leq C_0$. Suppose for the sake of a contradiction that there exists an algorithm such that for any $X \in \mathcal{S}_f$, given Y, returns an $\hat{X}$ such that

$$\|X - \hat{X}\|_F^2 / mn \leq \epsilon^2 \tag{77}$$

with probability at least $1/4$. Let

$$X^* = \arg \min_{X' \in \mathcal{X}} \|X' - \hat{X}\|_F^2. \tag{78}$$

We claim that if (77) holds, then $X^* = X$. To see this, for any $X' \in \mathcal{X}$ with $X' \neq X$, from (62) and (76) we have

$$\|X' - X\|_F > \alpha\gamma\sqrt{mn/2 - s} \geq 2\sqrt{mn}\epsilon. \tag{79}$$

Combine (77) and (79), we then have

$$\begin{aligned}\|X' - \hat{X}\|_F &= \|X' - X + X - \hat{X}\|_F \\ &\geq \|X' - X\|_F - \|X - \hat{X}\|_F > 2\sqrt{mn}\epsilon - \sqrt{mn}\epsilon = \sqrt{mn}\epsilon.\end{aligned} \tag{80}$$

Since $X \in \mathcal{X}$ is a candidate for X^*, we have that

$$\|X^* - \hat{X}\|_F \leq \|X - \hat{X}\|_F \leq \sqrt{mn}\epsilon. \tag{81}$$

Thus, if (77) holds, then $\|X^* - \hat{X}\|_F < \|X' - \hat{X}\|_F$ for any $X' \in \mathcal{X}$ with $X' \neq X$, and hence we must have $X^* = X$. By assumption, (77) holds with probability at least $1/4$, and thus

$$P(X \neq X^*) \leq 3/4. \tag{82}$$

We consider running this algorithm on a matrix X chosen uniformly at random from $\mathcal{X}$. However, by Fano's inequality, the probability that $X \neq X^*$ is at least

$$\begin{aligned}P(X \neq X^*) &\geq \frac{H(X|Y) - 1}{\log|\mathcal{X}|} \\ &= \frac{H(X) - I(X;Y) - 1}{\log|\mathcal{X}|} \geq 1 - \frac{I(X;Y) + 1}{\log|\mathcal{X}|}.\end{aligned} \tag{83}$$

Plugging in $|\mathcal{X}|$ from Lemma 4 and $I(X,Y)$ from Lemma 5, and using the inequality $\log(1+z) \leq z$, we obtain

$$P(X \neq \hat{X}) \geq 1 - \frac{16}{rm - r\lfloor\frac{s}{n}\rfloor}\left(\frac{mn - n\lfloor\frac{s}{n}\rfloor}{2}\left(\frac{\alpha\gamma}{\sigma}\right)^2 + 1\right). \tag{84}$$

Combining (84) with (76) and (82), we obtain

$$\frac{16}{rm - r\lfloor\frac{s}{n}\rfloor}\left((mn - n\lfloor\frac{s}{n}\rfloor)\frac{4}{1-2C_0}\left(\frac{\epsilon}{\sigma}\right)^2 + 1\right) \geq \frac{1}{4}, \tag{85}$$

which implies that

$$\epsilon^2 \geq \frac{(1-2C_0)\sigma^2}{256}\frac{rm - r\lfloor\frac{s}{n}\rfloor - 64}{mn - n\lfloor\frac{s}{n}\rfloor}. \tag{86}$$

Setting $C_4^2 < \frac{1-2C_0}{256}$ leads to a contradiction, hence (77) must fail to hold with probability at least $3/4$, which proves the theorem. Note that $C_3 = \sqrt{\frac{1-2C_0}{8}}\alpha$ and $C_4 < \sqrt{\frac{1-2C_0}{256}}$.

F. Proof of Theorem 3

Proof: The proof is similar to the proof of Theorem 1 with minor modifications to handle W. We will only prove

$$\|\hat{M} - M^*\|_F/\sqrt{mn} \leq U'_\alpha, \tag{87}$$

while other steps are the same as the proof of Theorem 1.

Consider

$$\bar{Z}_{ij} := -W^{-1}L_\alpha^{-1}\sum_{t=1}^{W}\sum_{l=1}^{K}\frac{\dot{f}_l(M^*_{ij})}{f_l(M^*_{ij})}\mathbf{1}_{[Y_{ij}^t = l]}.$$

$\mathbb{E}[\bar{Z}_{ij}] = 0$, $|\bar{Z}_{ij}| \leq 1$. Given M^*_{ij}, Y^t_{ij}'s are independent of each other since N^t_{ij}'s are independent noise. Then one can check that $\mathbb{E}[\bar{Z}^2_{ij}] \leq 1/W$. From Lemma 3, we have

$$\|W^{-1}L_\alpha^{-1}\nabla_X\bar{F}(M^*)\|_2 \leq 2.01\sqrt{m/W} \tag{88}$$

holds with probability at least $1 - C_1'e^{-\frac{C_2'm}{W}}$ for some positive constants C_1' and C_2'.

With arguments similar to those after (60) in the proof of Lemma 1, we obtain

$$\begin{aligned}&|\left\langle\nabla_\theta\bar{\mathcal{F}}_{\bar{Y}}(\theta^*), \theta' - \theta^*\right\rangle| \\ \leq& 2.01L_\alpha\sqrt{2rmW}(\|X' - M^*\|_F + 2\alpha\sqrt{s}) + 2\alpha sL_\alpha\sqrt{W}\end{aligned} \tag{89}$$

holds with probability at least $1 - C_1e^{-\frac{C_2m}{W}}$ for some constants C_1 and C_2.

Following the same steps of the proof of Lemma 2 (Lemma A.3 in [1]), one can check that

$$\left\langle\theta' - \theta^*, (\nabla^2_{\theta\theta}\bar{\mathcal{F}}_{\bar{Y}}(\tilde{\theta}))(\theta' - \theta^*)\right\rangle \geq W\gamma_\alpha\|X' - M^*\|_F^2. \tag{90}$$

Combining (48), (89), and (90), and note that $\bar{\mathcal{F}}_{\bar{Y}}(\hat{M}) \leq \bar{\mathcal{F}}_{\bar{Y}}(M^*)$, we then have

$$\begin{aligned}&\frac{W\gamma_\alpha}{2}\|\hat{M} - M^*\|_F^2 \leq 2.01L_\alpha\sqrt{2rmW}\|\hat{M} - M^*\|_F \\ &+4.02\alpha L_\alpha\sqrt{2rmsW} + 2\alpha L_\alpha s\sqrt{W},\end{aligned} \tag{91}$$

which completes the proof. ∎

G. Proof of Theorem 4

Proof: The proof is similar to the proof of Theorem 2 with minor modifications to handle W. Choose ϵ so that

$$\epsilon^2 = \min\{\frac{(1-2C_0)\alpha^2}{8}, \frac{C_4^2\sigma^2}{W}\frac{rm - r\lfloor\frac{s}{n}\rfloor - 64}{mn - n\lfloor\frac{s}{n}\rfloor}\} \tag{92}$$

for a constant C_4 to be determined later. We will set γ such that

$$\frac{2\epsilon}{\alpha}\sqrt{\frac{2mn}{mn - 2s}} \leq \gamma \leq \frac{\epsilon}{\alpha}\sqrt{\frac{2}{1-2C_0}} \leq 1. \tag{93}$$

Suppose $X \in \mathcal{X}$ is chosen uniformly at random. Let $Y^i = X + N^i$ $(\forall i \in \{1, ..., W\})$ as in our problem formulation. Note that we have

$$I(X;\bar{Y}) \leq \sum_{i=1}^{W} I(X;Y^i) \leq \frac{mn - n\lfloor\frac{s}{n}\rfloor}{2}W\log(1 + (\frac{\alpha\gamma}{\sigma})^2), \tag{94}$$

where the last inequality holds by Lemma 5.

Suppose for the sake of a contradiction that there exists an algorithm such that for any $X \in \mathcal{S}_f$, when given access to the measurements $\bar{Y}$, returns an $\hat{X}$ such that

$$\|X - \hat{X}\|_F^2/mn \leq \epsilon^2 \tag{95}$$

with probability at least $1/4$. We consider running this algorithm on a matrix X chosen uniformly at random from $\mathcal{X}$. Let

$$X^* = \arg\min_{X' \in \mathcal{X}}\|X' - \hat{X}\|_F^2. \tag{96}$$

We claim that if (95) holds, then $X^* = X$. By assumption, (95) holds with probability at least $1/4$, and thus

$$P(X \neq X^*) \leq 3/4. \tag{97}$$

However, by Fano's inequality,

$$\begin{aligned}\mathcal{P}(X \neq X^*) &\geq \frac{H(X|\bar{Y}) - 1}{\log|\mathcal{X}|} \\ &= \frac{H(X) - I(X;\bar{Y}) - 1}{\log|\mathcal{X}|} \geq 1 - \frac{I(X;\bar{Y}) + 1}{\log|\mathcal{X}|}.\end{aligned} \tag{98}$$

Plugging in $|\mathcal{X}|$ from Lemma 4 and $I(X,\bar{Y})$ from (94), and using the inequality $\log(1+z) \leq z$, we obtain

$$\mathcal{P}(X \neq X^*) \geq 1 - \frac{16}{rm - r\lfloor\frac{s}{n}\rfloor}(\frac{mn - n\lfloor\frac{s}{n}\rfloor}{2}W(\frac{\alpha\gamma}{\sigma})^2 + 1). \tag{99}$$

Combining this with (97) and (93), we obtain

$$\frac{16}{rm - r\lfloor\frac{s}{n}\rfloor}(W(mn - n\lfloor\frac{s}{n}\rfloor)\frac{4}{1-2C_0}(\frac{\epsilon}{\sigma})^2 + 1) \geq \frac{1}{4}, \tag{100}$$

which implies that

$$\epsilon^2 \geq \frac{(1-2C_0)\sigma^2}{256W}\frac{rm - r\lfloor\frac{s}{n}\rfloor - 64}{mn - n\lfloor\frac{s}{n}\rfloor}. \tag{101}$$

Setting $C_4^2 < \frac{1-2C_0}{256}$ leads to a contradiction, and hence (95) must fail to hold with probability at least $3/4$, which proves the theorem. Note that $C_3 = \sqrt{\frac{1-2C_0}{8}}\alpha$ and $C_4 < \sqrt{\frac{1-2C_0}{256}}$. ∎

H. Supporting Lemmas for the Proof of Theorem 5

Define $a = \frac{\sqrt{mn}}{\sigma_1(G)}$, $b = \frac{\sqrt{rmn}\sigma_2(G)}{\sigma_1(G)}$, $a_0 = \frac{\sigma_1(G)}{\sqrt{2rmn}}$, $c_g = 2.01L_\alpha\sqrt{2rm}$, $\bar{c}_h = \frac{\gamma_\alpha}{2}$, $c_h = \frac{\sigma_1^2(G)\gamma_\alpha}{4rmn}$, and $\eta = c_g b 2\alpha_1\sqrt{2r} + c_g\frac{2\alpha_2\sqrt{s}}{a_0} + 2\alpha_2 L_\alpha s$. Let $H(|\Omega|)$ denote the number of corruptions in the observed measurements.

Lemma 6. *Suppose that $M^* \in \mathcal{S}_f$, it holds that*

$$\begin{aligned}&|\left\langle \nabla_M \mathcal{F}_Y(M^*), \hat{M} - M^* \right\rangle| \leq \frac{c_g}{a_0}\|(\hat{M} - M^*)_\Omega\|_F + \\ &\frac{c_g}{a_0}2\alpha_2\sqrt{s} + 2\sqrt{2r}\alpha_1 b c_g + 2\alpha_2 L_\alpha s.\end{aligned} \tag{102}$$

Proof: We have

$$\begin{aligned}&|\left\langle \nabla_{\Omega,M}\mathcal{F}_Y(M^*), \hat{M} - M^* \right\rangle| \\ &\leq |\left\langle \nabla_{\Omega,M}\mathcal{F}_Y(M^*), \hat{L} - L^* \right\rangle| + |\left\langle \nabla_{\Omega,M}\mathcal{F}_Y(\theta^*), \hat{C} - C^* \right\rangle| \\ &\overset{(a)}{\leq} c_g\sqrt{2ra}\|(\hat{L} - L^*)_\Omega\|_F + 2c_g\alpha_1 b\sqrt{2r} + \sum_\Omega L_\alpha|\hat{C} - C^*| \\ &\overset{(b)}{\leq} c_g\sqrt{2ra}(\|(\hat{M} - M^*)_\Omega\|_F + \|(\hat{C} - C^*)_\Omega\|_F) \\ &+ 2c_g\sqrt{2r}\alpha_1 b + 2\alpha_2 L_\alpha H(|\Omega|) \\ &\leq \frac{c_g}{a_0}\|(\hat{M} - M^*)_\Omega\|_F + \frac{c_g}{a_0}2\alpha_2\sqrt{s} + 2\sqrt{2r}\alpha_1 b c_g + 2\alpha_2 L_\alpha s.\end{aligned}$$

(a) holds according to (10) and (59). (b) holds because of Lemma 9 in [1]. ∎

Lemma 7. *It holds that*

$$\begin{aligned}&|\left\langle \nabla_{\Omega,X}F_Y(M^*), \hat{M} - M^* \right\rangle| \leq \\ &4L_\alpha(1+\alpha_1)\sqrt{|\Omega|r(m+n+1)\log(9\alpha_1(mn)^{1/2})} + 2L_\alpha\alpha_2 s\end{aligned} \tag{103}$$

with probability at least $1 - 2(9\alpha\sqrt{mn})^{-r(m+n+1)} - C_1\exp(-C_2 m)$.

Proof: We have

$$\begin{aligned}&\langle \nabla_{\Omega,\theta}\mathcal{F}_Y(\theta^*), \theta' - \theta^* \rangle = \left\langle \nabla_{\Omega,X}F_Y(M^*), \hat{M} - M^* \right\rangle \\ &= \left\langle \nabla_{\Omega,X}F_Y(M^*), \hat{L} - L^* \right\rangle + \left\langle \nabla_{\Omega,X}F_Y(M^*), \hat{C} - C^* \right\rangle \\ &\leq \left\langle \nabla_{\Omega,X}F_Y(M^*), \hat{L} - L^* \right\rangle + 2L_\alpha\alpha_2 s \overset{(a)}{\leq} \\ &4L_\alpha(1+\alpha_1)\sqrt{|\Omega|r(m+n+1)\log(9\alpha_1(mn)^{1/2})} + 2L_\alpha\alpha_2 s.\end{aligned} \tag{104}$$

(a) and the holding probability comes from Lemma 6 in [1]. ∎

I. Proof of Theorem 5

From (48), Lemma 2, and Lemma 6, we have

$$\begin{aligned}&0 \geq F_{\Omega,Y}(\hat{M}) - F_{\Omega,Y}(M^*) \\ &\geq -\frac{c_g\|(\hat{M} - M^*)_\Omega\|_F}{a_0} + \bar{c}_h\|(\hat{M} - M^*)_\Omega\|_F^2 - \eta.\end{aligned} \tag{105}$$

By solving (105), we then have

$$\|(\hat{M} - M^*)_\Omega\|_F \leq (c_g + \sqrt{c_g^2 + 4a_0^2\bar{c}_h\eta})/2a_0\bar{c}_h. \tag{106}$$

Define

$$\begin{aligned}\bar{c}_g &= L_\alpha\sqrt{|\Omega|}4(1+\alpha_1)\sqrt{r(m+n+1)\log(9\alpha_1(mn)^{1/2})} \\ &+ 2L_\alpha\alpha_2 s\end{aligned} \tag{107}$$

Applying Lemma 2 and Lemma 7, we have

$$0 \geq F_{\Omega,Y}(\hat{M}) - F_{\Omega,Y}(M^*) \geq -\bar{c}_g + \bar{c}_h\|(\hat{M} - M)_\Omega\|_F^2 \tag{108}$$

holds with probability at least $1 - C_1\exp(-C_2 m) - 2(9\alpha\sqrt{mn})^{-r(m+n+1)}$. By solving (108), we have

$$\|(\hat{M} - M^*)_\Omega\|_F \leq \sqrt{2\bar{c}_g/\gamma_\alpha}. \tag{109}$$

We then have

$$\begin{aligned}&\|\hat{L} - L^*\|_F \overset{(a)}{\leq} (\sqrt{2ra})\|(\hat{L} - L^*)_\Omega\|_F + 2\sqrt{2r}\alpha_1 b \\ &\leq (\sqrt{2ra})(\|(\hat{M} - M^*)_\Omega\|_F + \|(\hat{C} - C^*)_\Omega\|_F) + 2\sqrt{2r}\alpha_1 b \\ &\overset{(b)}{\leq} (\sqrt{2ra})(\frac{c_g + \sqrt{c_g^2 + 4a_0^2\bar{c}_h\eta}}{2a_0\bar{c}_h} + 2\alpha_2\sqrt{s}) + 2\sqrt{2r}\alpha_1 b \\ &\overset{(c)}{\leq} \frac{4.02\sqrt{2}L_\alpha r^{\frac{3}{2}}m^{\frac{3}{2}}n}{d^2} + (\frac{(4.02\sqrt{2})^2L_\alpha^2 r^3 m^3 n^2}{d^4} \\ &+ \frac{64.32QL_\alpha\alpha_1 r^{\frac{3}{2}}m^2 n^{\frac{3}{2}}}{\gamma_\alpha^2 d^{\frac{5}{2}}} + \frac{32.16L_\alpha\alpha_2 r^2 m^2 n^{\frac{3}{2}}\sqrt{s}}{\gamma_\alpha d^3} \\ &+ \frac{8L_\alpha\alpha_2 rmns}{\gamma_\alpha d^2})^{\frac{1}{2}} + \frac{2\alpha_2\sqrt{2mns}}{d} + \frac{2\sqrt{2}Q\alpha_1 r\sqrt{mn}}{\sqrt{d}} \\ &\leq \frac{24Q'L'_\alpha\alpha'_1 r^{\frac{3}{2}}m^{\frac{3}{2}}n}{d^2} + \frac{6(L_\alpha\alpha_2 n)^{\frac{1}{2}}(ns)^{\frac{1}{4}}rm}{d^{\frac{3}{2}}\sqrt{\gamma_\alpha}} + \frac{8L_\alpha'^{\frac{1}{2}}\alpha_2\sqrt{rmns}}{\sqrt{\gamma'_\alpha}d} \\ &= \frac{24Q'L'_\alpha\alpha'_1 r^{\frac{3}{2}}m^{\frac{3}{2}}}{p^2 n} + \frac{6(L_\alpha\alpha_2)^{\frac{1}{2}}(ns)^{\frac{1}{4}}rm}{p^{\frac{3}{2}}n\sqrt{\gamma_\alpha}} + \frac{8L_\alpha'^{\frac{1}{2}}\alpha_2\sqrt{rmns}}{\sqrt{\gamma'_\alpha}pn},\end{aligned} \tag{110}$$

where $Q' = \max(Q, 1)$, $L'_\alpha = \max(L_\alpha, 1)$, $\alpha'_1 = \max(\alpha_1, 1)$, $\gamma'_\alpha = \min(\gamma_\alpha, 1)$. (a) holds because of Lemma 9 in [1]. (b) holds according to (106). (c) holds because of Cauchy-Schwarz inequality and the assumption that $\sigma_1(G) \geq d$ and $\sigma_2(G) \leq Q\sqrt{d}$. Similar to (110), we then have

$$\|\hat{L} - L^*\|_F \leq \frac{8Q'(1+\alpha_1)\sqrt{L'_\alpha}}{\sqrt{\gamma'_\alpha}} \frac{(rm)^{\frac{3}{4}}((m+n)\log(\sqrt{mn}))^{\frac{1}{4}}}{p\sqrt{pn}} + \frac{8\alpha_2\sqrt{L'_\alpha rsm}}{p\sqrt{\gamma'_\alpha n}}. \tag{111}$$

Following (110), (111), and the assumption $\frac{m}{n} = \kappa \geq 1$, we obtain the final results.

REFERENCES

[1] S. A. Bhaskar, "Probabilistic low-rank matrix recovery from quantized measurements: Application to image denoising," in *Proc. Asilomar Conference on Signals, Systems and Computers*, 2015, pp. 541–545.

[2] ——, "Localization from connectivity: A 1-bit maximum likelihood approach," *IEEE/ACM Transactions on Networking*, vol. 24, no. 5, pp. 2939–2953, 2016.

[3] ——, "Probabilistic low-rank matrix completion from quantized measurements," *Journal of Machine Learning Research*, vol. 17, no. 60, pp. 1–34, 2016.

[4] S. Bhojanapalli and P. Jain, "Universal matrix completion," in *Proc. International Conference on Machine Learning*, 2014, pp. 1881–1889.

[5] J. Bolte, S. Shoham, and T. Marc, "Proximal alternating linearized minimization for nonconvex and nonsmooth problems," *Mathematical Programming*, vol. 146, no. 1-2, pp. 459–494, 2014.

[6] S. P. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[7] T. Cai and W.-X. Zhou, "A max-norm constrained minimization approach to 1-bit matrix completion," *The Journal of Machine Learning Research*, vol. 14, no. 1, pp. 3619–3647, 2013.

[8] E. Candès and T. Tao, "The power of convex relaxation: Near-optimal matrix completion," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2053 –2080, 2010.

[9] ——, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.

[10] E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," *Foundations of Computational mathematics*, vol. 9, no. 6, pp. 717–772, 2009.

[11] Y. Cao and Y. Xie, "Categorical matrix completion," in *Proc. IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, 2015.

[12] T. M. Cover and J. A. Thomas, *Elements of information theory*. Wiley New York:, 1991.

[13] M. A. Davenport, Y. Plan, E. van den Berg, and M. Wootters, "1-bit matrix completion," *Information and Inference*, vol. 3, no. 3, pp. 189–223, 2014.

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

[15] C. Dwork and A. Smith, "Differential privacy for statistics: What we know and what we want to learn," *Journal of Privacy and Confidentiality*, vol. 1, no. 2, p. 2, 2010.

[16] P. Gao, M. Wang, J. H. Chow, S. G. Ghiocel, B. Fardanesh, G. Stefopoulos, and M. P. Razanousky, "Identification of successive "unobservable" cyber data attacks in power systems." *IEEE Trans. Signal Process.*, vol. 64, no. 21, pp. 5557–5570, 2016.

[17] P. Gao, M. Wang, S. G. Ghiocel, J. H. Chow, B. Fardanesh, and G. Stefopoulos, "Missing data recovery by exploiting low-dimensionality in power system synchrophasor measurements," *IEEE Trans. Power Syst.*, vol. 31, no. 2, pp. 1006–1013, 2016.

[18] P. Gao, R. Wang, M. Wang, and J. H. Chow, "Low-rank matrix recovery from quantized and erroneous measurements: Accuracy-preserved data privatization in power grids," in *Proc. Asilomar Conference on Signals, Systems, and Computers*, 2016.

[19] S. Gunasekar, P. Ravikumar, and J. Ghosh, "Exponential family matrix completion under structural constraints." in *Proc. Int. Conf. Machine Learning (ICML)*, 2014, pp. 1917–1925.

[20] N. Kallus and M. Udell, "Revealed preference at scale: Learning personalized preferences from assortment choices," in *Proc. ACM Conference on Economics and Computation*. ACM, 2016, pp. 821–837.

[21] O. Klopp, J. Lafond, É. Moulines, and J. Salmon, "Adaptive multinomial matrix completion," *Electronic Journal of Statistics*, vol. 9, no. 2, pp. 2950–2975, 2015.

[22] O. Klopp, K. Lounici, and A. B. Tsybakov, "Robust matrix completion," *Probability Theory and Related Fields*, pp. 1–42, 2016.

[23] K. Kurdyka, "On gradients of functions definable in o-minimal structures," *Annales de l'institut Fourier*, vol. 48, no. 3, pp. 769–783, 1998.

[24] J. Lafond, O. Klopp, E. Moulines, and J. Salmon, "Probabilistic low-rank matrix completion on finite alphabets," in *Advances in Neural Information Processing Systems*, 2014, pp. 1727–1735.

[25] A. S. Lan, C. Studer, and R. G. Baraniuk, "Matrix recovery from quantized and corrupted measurements," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2014, pp. 4973–4977.

[26] A. S. Lan, A. E. Waters, C. Studer, and R. G. Baraniuk, "Sparse factor analysis for learning and content analytics." *Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1959–2008, 2014.

[27] A. Molina-Markham, P. Shenoy, K. Fu, E. Cecchet, and D. Irwin, "Private memoirs of a smart meter," in *Proceedings of the 2nd ACM workshop on embedded sensing systems for energy-efficiency in building*. ACM, 2010, pp. 61–66.

[28] North American Electric Reliability Corperation, "Operating reliability data confidentiality agreement, version 3." 2009. [Online]. Available: http://www.nerc.com/comm/OC/Pages/OperatingReliabilityDataConfidentialityAgreement.aspx

[29] A. Phadke and J. Thorp, *Synchronized phasor measurements and their applications*. Springer, 2008.

[30] A. Reinhardt, F. Englert, and D. Christin, "Enhancing user privacy by preprocessing distributed smart meter data," in *Proc. Sustainable Internet and ICT for Sustainability (SustainIT)*, 2013, pp. 1–7.

[31] J. Salmon, Z. Harmany, C.-A. Deledalle, and R. Willett, "Poisson noise reduction with non-local PCA," *Journal of mathematical imaging and vision*, vol. 48, no. 2, pp. 279–294, 2014.

[32] L. Sankar, S. Kar, R. Tandon, and H. V. Poor, "Competitive privacy in the smart grid: An information-theoretic approach," in *Proc. IEEE International Conference on Smart Grid Communications(SmartGridComm)*, 2011, pp. 220–225.

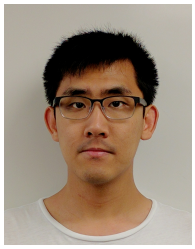
[33] A. Silverstein and J. E. Dagle, "Successes and challenges for synchrophasor technology: An update from the north american synchrophasor initiative," in *Proc. HICSS*, 2012, pp. 2091–2095.

[34] A. Soni, S. Jain, J. Haupt, and S. Gonella, "Noisy matrix completion under sparse factor models," *IEEE Transactions on Information Theory*, vol. 62, no. 6, pp. 3636–3661, 2016.

[35] Y. Tong, J. Sun, and K. Sun, "Privacy-preserving spectral estimation in smart grid," in *Proc. IEEE International Conference on Smart Grid Communications(SmartGridComm)*, 2015, pp. 43–48.

[36] S. Xiong, A. D. Sarwate, and N. B. Mandayam, "Randomized requantization with local differential privacy," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2016, pp. 2189–2193.

[37] Y. Xu and W. Yin, "A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion," *SIAM Journal on imaging sciences*, vol. 6, no. 3, pp. 1758–1789, 2013.



Pengzhi Gao (S'14) received the B.E. degree from Xidian University, Xian, China, in 2011, and the M.S. degree in electrical engineering from University of Pennsylvania, Philadelphia, PA, USA, in 2013.

He is currently working toward the Ph.D. degree in electrical engineering at Rensselaer Polytechnic Institute, Troy, NY, USA. His research interests include signal processing, compressive sensing, lowrank matrix recovery, and power networks.

Ren Wang (S'16) received the B.E. degree and M.S. degree in electrical engineering from Tsinghua University, Beijing, China, in 2013 and 2016.

He is pursuing the Ph.D. degree in electrical engineering at Rensselaer Polytechnic Institute, Troy, NY. His research interests include unsupervised learning, optimization and their applications in power systems.

Meng Wang (M'12) received the Ph.D. degree from Cornell University, Ithaca, NY, USA, in 2012.

She is an Assistant Professor in the department of Electrical, Computer, and Systems Engineering at Rensselaer Polytechnic Institute. Her research interests include high dimensional data analysis and their applications in power systems monitoring and network inference.

Joe H. Chow (F'92) received the M.S. and Ph.D. degrees from the University of Illinois, Urbana-Champaign, Urbana, IL, USA.

After working in the General Electric power system business in Schenectady, NY, USA, he joined Rensselaer Polytechnic Institute, Troy, NY, USA, in 1987, where he is a Professor of Electrical, Computer, and Systems Engineering. His research interests include multivariable control, power system dynamics and control, FACTS controllers, and synchronized phasor data.