EURASIP Journal on Advances in Signal Processing

RESEARCH Open Access

Achieve data privacy and clustering accuracy simultaneously through quantized data recovery

Ren Wang[1], Meng Wang[1*] and Jinjun Xiong[2]

*Correspondence:
wangm7@rpi.edu
[1]Department of Electrical, Computer, and Systems Engineering, Rensselaer Polytechnic Institute, Troy, NY, USA
Full list of author information is available at the end of the article

Abstract

This paper develops a data collection and processing framework that achieves individual users' data privacy and the operator's information accuracy simultaneously. Data privacy is enhanced by adding noise and applying quantization to the data before transmission, and the privacy of an individual user is measured by information-theoretic analysis. This paper develops a data recovery and clustering method for the operator to extract features from the privacy-preserving, partially corrupted, and partially observed measurements of a large number of users. To prevent cyber intruders from accessing the data of many users, it also develops a decentralized algorithm such that multiple data owners can collaboratively recover and cluster the data without sharing the raw measurements directly. The recovery accuracy is characterized analytically and showed to be close to the fundamental limit of any recovery method. The proposed algorithm is proved to converge to a critical point from any initial point. The method is evaluated on recorded Irish smart meter data and UMass smart microgrid data.

Keywords: Subspace clustering, Quantization, Data recovery, Data privacy, Smart meter

1 Introduction

Smart meters provide fine-grained measurements of power consumption of industrial and residential customers and can enhance the distribution system visibility. Non-intrusive load monitoring (NILM) approaches [1, 2] can identify individual appliances from the high-time-resolution smart meter data of the aggregated power consumption. Intruders can thus extract user behavior, and user privacy is an increasing concern. One way to protect data privacy is by applying additive homomorphic encryption [3]. It requires the network to have tree-like connections and can only decrypt the sum of the load curves. The other way to enhance data privacy is data obfuscation whereby the actual power consumption of each household is masked by adding noise to the smart meter measurements either through signal processing approaches [4, 5] or by physically adding rechargeable batteries to the households [6, 7]. Moreover, the aggregated consumption of the load and the battery can be adjusted to a constant to obfuscate the information further

[8, 9]. Then, applying the NILM to these noisy and quantized measurements, an intruder can no longer accurately identify the patterns of individual appliances and, in turn, the user behavior. The increase in user privacy is achieved, however, at a cost of data distortion and reduced data accuracy for the operating center [10–12]. Although the operating center does not need high-time-resolution information of every individual appliances in each household, it still requires accurate estimation of the aggregated power consumption and the common load patterns among households for forecasting, demand response, and planning. For example, the center clusters customers with similar load patterns and then employs the load pattern of each cluster to enhance the load forecasting accuracy [13] and determine the incentives for demand response [14, 15]. If noise and quantization are added to the data to enhance the privacy, the information accuracy for the operator is effectively reduced.

This paper shows that the data privacy can be protected for each individual user[1] and, *at the same time*, the information accuracy at the operating center about user power consumption and the major patterns among different users are maintained. To the best of our knowledge, this is the first work that achieves data privacy and information accuracy *simultaneously*. In our proposed framework, each user's actual power consumption is masked by first adding noise to the measurements and then quantizing the output to one of a few levels. The privacy of an individual user can be enhanced in this way, from an information-theoretic perspective [16–19]. Once the data is quantized, the variation information is blurred and hence NILM methods fail to identify individual appliances. Although adding noise and quantization have been employed before to enhance privacy (e.g., [6, 20]), this paper, for the first time, shows that such privacy enhancement does not necessarily lead to a reduction in the information accuracy. The central technical contribution of this paper is the development of a data recovery and clustering method, even when the measurements are highly noisy and quantized, contain significant errors, and are partially lost. Our method is proved to provide accurate data recovery and clustering results, as long as the center has measurements from a sufficient number of users. In contrast, a cyber intruder with access to the measurements of a small number of users cannot obtain accurate information even with the same approach. We develop a decentralized algorithm that allows multiple data owners to cooperatively recover and cluster the data without sharing their own raw measurements directly. Then, it is extremely difficult for an intruder to access large amounts of data. Thus, the data privacy of an individual user is enhanced while maintaining the information accuracy for the operating center.

Since the load profiles with similar load patterns can be represented by data points in a low-dimensional subspace in the high-dimensional ambient space, all the load profiles can be characterized by the Union of Subspaces (UoS) model [21], and the load clustering problem can be formulated as a subspace clustering problem. Various subspace clustering techniques have been developed, see e.g., [21–26]. None of these approaches, however, considers the case that the measurements are highly quantized. To the best of our knowledge, only one recent work considered subspace clustering and data recovery from highly noisy and quantized data [27]. This paper follows the mathematical setup of [27] but extends significantly in the following aspects. Ref. [27] does not consider data privacy, while this paper proposes a data collection framework to achieve data privacy and

[1] Throughout this paper, we refer to each household as one user.

information accuracy simultaneously. We characterize the data privacy through mutual information, and such analysis does not exist in [27]. Ref. [27] assumes that all the measurements are available to the center, while this paper considers a more general setup that partial measurements are lost during the transmission and do not arrive at the center. This paper characterizes the data recovery error by our proposed method analytically as a function of data loss percentage. Moreover, this paper characterizes the fundamental limit of the recovery error by any possible recovery method and shows that our method is nearly optimal in reducing the recovery error. All these fundamental analyses do not exist in [27]. Furthermore, only a centralized algorithm Sparse-APA is discussed in [27]. This paper develops a Distributed Sparse Alternative Proximal Algorithm (DSAPA) for multiple data owners to collaboratively solve the subspace clustering and data recovery problem without sharing the measurements with others. Thus, the user data privacy can be further protected. This paper is also related to the quantized matrix recovery problem [28–36], in which the data matrix is assumed to be low rank. The low-rank matrix model is a special case of the UoS model by restricting to one subspace only. In fact, the data matrix of the load profiles can be high rank or even full rank in our setup. Finally, we remark that this paper considers smart meter measurements that measure the aggregated energy consumption in a house, and does not consider applying NILM on the operator side. Distributed smart metering can provide energy consumption of individual electrical appliances in a house [20].

The rest of the paper is organized as follows. Section 2 introduces our proposed framework, problem formulation, related works, and the data privacy enhancement analysis. The theoretical analyses of our recovery and clustering method is presented in Section 3. Section 4 introduces the details of the DSAPA with its convergence guarantee. Section 5 records the numerical experiments of our method on the real smart meter dataset. Section 6 concludes the paper. All the proofs are deferred to Appendix 1, Appendix 2, Appendix 3, Appendix 4, Appendix 5, Appendix 6, and Appendix 7.

2 Our proposed framework of privacy-preserving data collection and information recovery

2.1 Our framework and problem formulation

Figure 1 visualizes our proposed framework of privacy-preserving smart meter data collection and information recovery. To enhance the user data privacy, the actual power consumption is mapped to a few fixed power levels at the output of the smart meter. One can achieve this through signal processing in the smart meter or connecting a rechargeable battery to each household. Thus, the actual consumption is masked in the noisy and quantized smart meter measurements. As shown in Fig. 1, the measurements are collected by W agents disjointly, and agents do not share measurements directly. The agents recover the data and cluster the users with similar consumption patterns collaboratively in a distributed fashion. When $W = 1$, it reduces to the case of one single center.

We defer the discussion of user privacy enhancement through the proposed framework to Section 2.3. We first define the recovery and clustering problem from quantized data mathematically as follows. $L^* \in \mathbb{R}^{m\times n}$ denotes the actual power usages of n users, with each column containing the power usage of one user in m time instants. We assume that users with similar consumption patterns belong to the same group and there are p groups in total. The corresponding columns of the same group belong to a d-dimensional

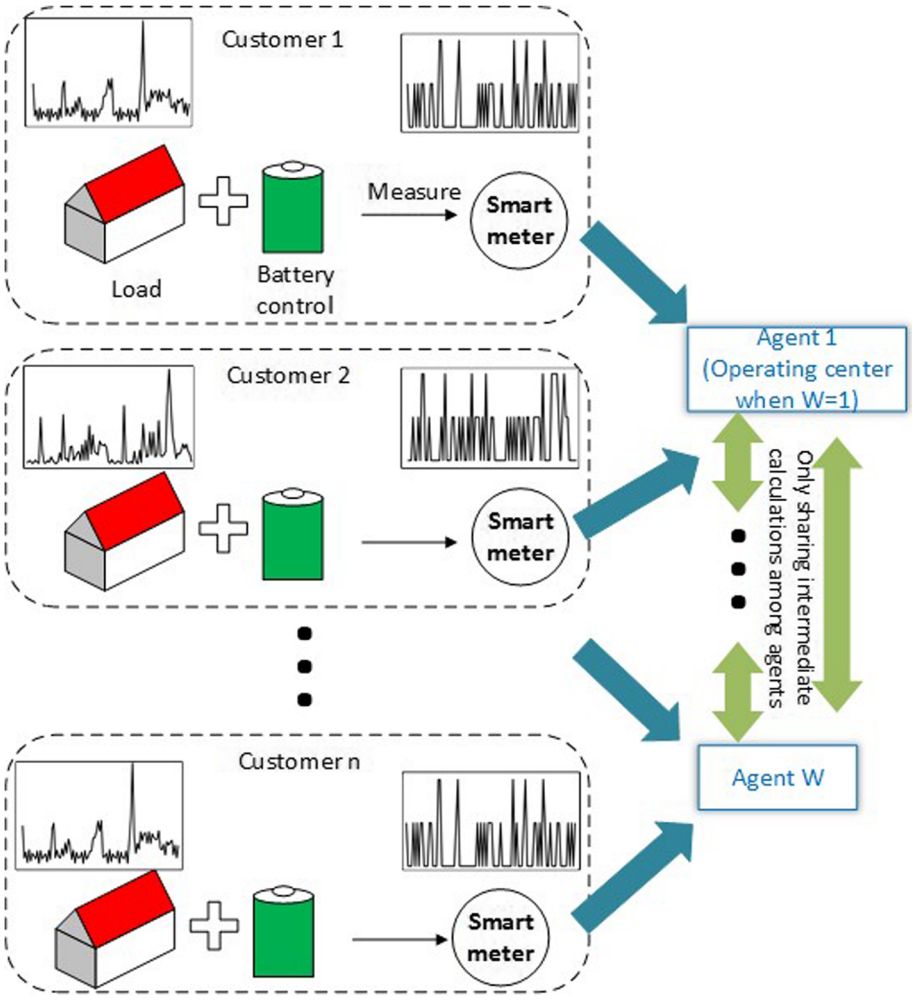

Fig. 1 The process of the quantization, measurement, collection, and transmission

subspace in $\mathbb{R}^m$ with $d \leq m$. Let S_i ($i \in [p]$) denote the ith subspace, and these p subspaces are distinct[2]. Let r denote the rank of L^*, then $r \leq pd$. Let L_i^* denote the submatrix of L^* that contains points in S_i, and let n_i denote the number of columns in L_i^*, i.e., the number of users in group i. We assume $m \leq n_i \leq \xi n/p$ for all i and some positive constant ξ. We further assume $m = n/\kappa p$ for some positive constant κ to simplify the representation of main results.

There exists a coefficient matrix $C^* \in \mathbb{R}^{n\times n}$ such that $L^* = L^*C^*$, $C_{i,i}^* = 0$ for all $i \in [n]$. Moreover, $C_{i,j}^*$ is zero if the ith and jth columns of L^* do not belong to the same subspace [21]. We summarize these two properties as *self-expressive property* and *subspace-preserving property* in Definition 1. These properties have been exploited in the literature of subspace clustering and are summarized as follows.

Definition 1 [27] *A matrix $L \in \mathbb{R}^{m\times n}$ has the self-expressive property if $L = LC$ for some $C \in \mathbb{R}^{n\times n}$, and $C_{i,i} = 0$ for all $i \in [n]$. Moreover, C has the subspace-preserving property of L if $C_{i,j} = 0$ for columns i and j of L belonging to different subspaces.*

[2] S_i's ($i \in [p]$) are distinct provided for any i, j, there always exists some β that belongs to S_i but not S_j.

Let matrix $E^* \in \mathbb{R}^{m\times n}$ denote the additive errors in the measurements. We assume the number of nonzeros s in E^* is much smaller than mn. The partially corrupted measurements can be represented by $X^* = L^* + E^*$. We assume the energy consumption and the errors are bounded, i.e., $\|L^*\|_\infty \le \alpha_1$ and $\|E^*\|_\infty \le \alpha_2$, for some positive constants α_1, α_2, and the infinity norm $\|\cdot\|_\infty$ measures the maximum absolute value.

The quantization process in each household is modeled as follows. The measured energy consumption at each time step is mapped to one of K values in a probabilistic fashion. Figure 2 shows the quantization process. It can be modeled as adding random noise first and then quantizing to K levels. $N \in \mathbb{R}^{m\times n}$ is independent from X^*. Entries of N are i.i.d. generated from a fixed cumulative distribution function (c.d.f.) $\Psi(x)$. The quantization boundaries $\omega_0 < \omega_1 < ... < \omega_{l-1} < \omega_l ... < \omega_K$ and the quantized value $\mathcal{Q}_l, l \in [K]$ for the bin $[\omega_{l-1}, \omega_l)$ are given. Then, the probability of mapping $X^*_{i,j}$ to $Y_{i,j} = \mathcal{Q}_l, \forall i,j$ is represented by

$$\begin{aligned}\varphi_l(X^*_{i,j}) &= P\left(Y_{i,j} = \mathcal{Q}_l | X^*_{i,j}\right) \\ &= \Psi\left(\omega_l - X^*_{i,j}\right) - \Psi\left(\omega_{l-1} - X^*_{i,j}\right),\end{aligned} \tag{1}$$

and $\sum_{l=1}^{K} \varphi_l\left(X^*_{i,j}\right) = 1$. The noise N is introduced to hide the user information. One choice of $\Psi(x)$ is the probit model with $\Psi(x) = \Psi_{\text{norm}}(x/\sigma)$, where Ψ_{norm} is the c.d.f. of the standard Gaussian distribution $\mathcal{N}(0,1)$, and $\sigma > 0$ is the standard deviation. Note that $\Psi\left(\omega_l - X^*_{i,j}\right) \ge \Psi\left(\omega_{l-1} - X^*_{i,j}\right) + \beta$ for some positive β. Then, $1 \ge \varphi_l \ge \beta > 0$.

The quantized measurements Y are sent to the center. Data losses can happen during the communication, visualized by the question marks in Fig. 2. Let set Ω denote the indices of measurements that are not lost. In the general case that the measurements are collected by W agents/nodes separately, we assume for simplicity that each node collects the data from $q = n/W$ users. Node 1 collects the data from the first q users; node 2 collects the next q users and so on. Let $\Phi_i = \{q(i-1)+1, q(i-1)+2, ..., qi\}$, then $L^*_{\Phi_i}$ denotes the submatrix of L^* with column indices in Φ_i. L^* can also written as $\left[L^*_{\Phi_1}, L^*_{\Phi_2}, ..., L^*_{\Phi_W}\right]$. Similarly, $E^* = \left[E^*_{\Phi_1}, E^*_{\Phi_2}, ..., E^*_{\Phi_W}\right]$. Node i collects Y_{Φ_i}.

The data recovery and pattern extraction problem for one center can be stated as follows.

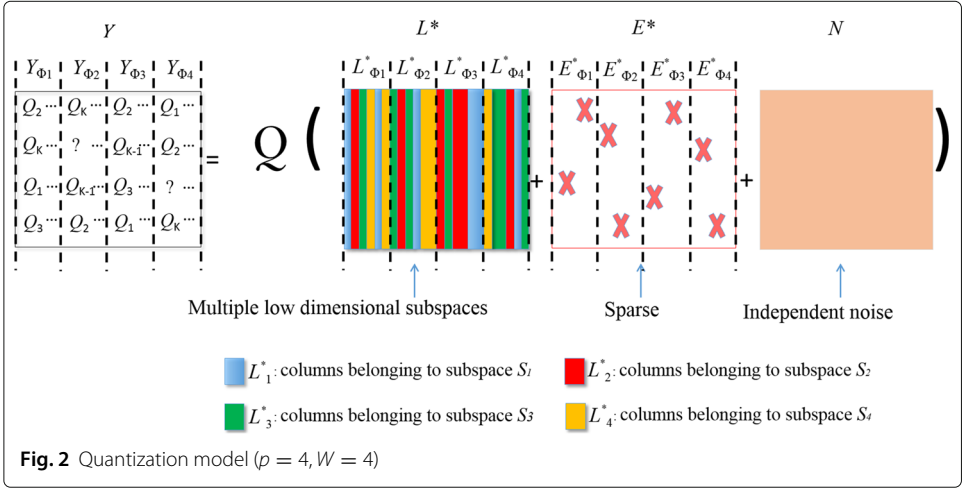

Fig. 2 Quantization model ($p = 4, W = 4$)

(P1) *Given quantized measurement Y_Ω, known boundaries $\omega_0 < \omega_1 < ... < \omega_K$ and noise distribution Ψ, can we recover the real power usages L^* and cluster the users through estimating C^* simultaneously?*

Moreover, if measurements $Y_{i,j}$'s are not shared among W nodes to protect the user privacy,

(P2) *Can we estimate L^* and C^* with W nodes in a decentralized fashion?*

Some notations in this paper are summarized in Table 1.

2.2 Related work

When $p = 1$, i.e., all the users share the same pattern, L^* is approximately a low-rank matrix. Then, (P1) reduces to the problem of low-rank matrix recovery from quantized measurements [28–37], with motivating applications in image processing [38], collaborative filtering [31], and sensor networks [39]. Note that since there is only one subspace in this case, these works do not consider data clustering and only focus on data recovery.

When the quantization process does not exist, the problem (P1) reduces to the conventional subspace clustering problem [21–26, 40]. If the subspace preserving C^* is estimated, one can apply the spectral clustering [41] method to obtain the clustering of the data points. For example, Sparse Subspace Clustering (SSC) [21] is a common choice for subspace clustering, and SSC estimates C^* by solving a convex optimization problem. Other clustering methods exist that cluster data points based on the Euclidean distance. For instance, refs. Lin et al. [42] and Keogh et al. [43] leverage a linear combination of box basis functions to approximate the original data, yet still retain the features of interest.

Reference [27] is the first paper that studies the subspace clustering from quantized measurements when $p \geq 1$. Wang et al. [27] do not consider missing data and develop a centralized data recovery method from full observations. This paper follows the same problem formulation as [27] and extends to the general case of partial observations. We provide both the recovery guarantee of our approach and the fundamental limit of the recovery accuracy by any method. Moreover, a framework of privacy-preserving smart meter data collection is proposed in this paper, and we further enhance the data privacy by developing a decentralized data recovery method.

Our problem formulation and methods apply to other domains such as image and video processing and phasor measurement unit (PMU) data analytics for power systems. In image recovery and image clustering [27], images of the same person with varying illumination belong to the same low-dimensional subspace [44]. Columns of L^* correspond to

Table 1 Notations

S_i	The ith subspace
L^*_i	Columns of matrix L^* belonging to the ith subspace
$L^*_{\star i}$	The ith column of matrix L^*
$L^*_{i\star}$	The ith row of matrix L^*
L^*_{ij}	Entry on the ith row and jth column of matrix L^*
$[p]$	The set $\{1, ..., p\}$
Φ_i	Index set containing $\{q(i-1)+1, q(i-1)+2, ..., qi\}$
C_{Φ_i}	Columns of matrix C belonging to the set Φ_i
$C_{\Phi_i\star}$	Rows of matrix C belonging to the set Φ_i
$(C_{\Phi_i})_{\Phi_j\star}$	Rows of matrix C_{Φ_i} belonging to the set Φ_j
$C^T_{\Phi_i\star}$	The transpose of $C_{\Phi_i\star}$

images of multiple people. The goal is to enhance the image quality and cluster the data using low-resolution images. Similarly, in motion segmentation, each column of L^* represents the trajectory of a reference point. The reference points in the same rigid object belong to the same subspace. The motion segmentation becomes a subspace clustering problem from the observed measurements. In PMU data analytics, the time series of PMUs affected by the same event belong to the same subspace [32, 45]. The event location problem can be solved by subspace clustering.

2.3 Data privacy enhancement in the proposed framework

Various methods have been developed to enhance the privacy of power consumption data. For example, one can use pre-processing techniques like temporal averaging, adding additional noise, and quantization [4, 5, 20] to alter the data. However, directly altering data might affect the accuracy of some applications, e.g., billing and profiling [46]. Alternatively, rechargeable batteries and PV converter can be leveraged to mask the actual power consumption [6, 7, 47]. The noise addition and quantization process in this paper can be achieved by either signal processing or rechargeable batteries.

In general, privacy guarantee can be achieved through either computational hardness [48–50] or information-theoretic analysis [16–19]. The existing analytical results of data privacy only work for specific or simple models and do not easily generalize. For instance, under the setup of communication between two nodes, ref. [17] analyzes the trade-off between data sharing and privacy. Under the assumptions of i.i.d. input load sequence and an i.i.d. energy harvesting process, the minimum information leakage rate is provided with a certain energy management policy in [51]. Some other methods try to analyze data privacy numerically. In [52], the information leakage rate is measured by the relative entropy of the probability measures of the original load data and the modified load data and is calculated by Monte-Carlo method. Refs. [7] and [12] consider measuring the information leakage through mutual information of the original load data and the modified load data. Following the existing work on smart meter data privacy, see, e.g., [19, 52–54], this paper analyzes the data privacy from an information-theoretic perspective. The data privacy of an individual user is analyzed by comparing the original data and the data after privacy enhancement through quantities like the Kullback-Leibler (KL) divergence [52], mutual information, and normalized mutual information [18]. In our framework, the actual energy consumption of user i, denoted by $L^*_{\star i}$, is masked by additive Gaussian noise and quantization, resulting in $Y_{\star i}$. Let $P_{L^*_{\star i}}$ and $P_{Y_{\star i}}$ denote the probability distribution of $L^*_{\star i}$ and $Y_{\star i}$, respectively. The privacy can be measured through the normalized mutual information (NI) between $L^*_{\star i}$ and $Y_{\star i}$ [18], defined as follows:

Definition 2

$$NI\left(L^*_{\star i}, Y_{\star i}\right) = \frac{\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} P_{(L^*_{\star i},Y_{\star i})}(x,y)\log\frac{P_{(L^*_{\star i},Y_{\star i})}(x,y)}{P_{L^*_{\star i}}(x)P_{Y_{\star i}}(y)}}{\sum_{x\in\mathcal{X}} P_{L^*_{\star i}}(x)\log\frac{1}{P_{L^*_{\star i}}(x)}} \tag{2}$$

where spaces $\mathcal{X}$ and $\mathcal{Y}$ are the feasible set of $L^*_{\star i}$ and $Y_{\star i}$, respectively. $P_{(L^*_{\star i},Y_{\star i})}$ is the joint distribution of $L^*_{\star i}$ and $Y_{\star i}$. $P_{L^*_{\star i}}$ and $P_{Y_{\star i}}$ are the marginal distributions of $L^*_{\star i}$ and $Y_{\star i}$, respectively.

The numerator of (2) is the mutual information between $L^*_{\star i}$ and $Y_{\star i}$, and the denominator is the entropy of $L^*_{\star i}$. When $L^*_{\star i}$ and $Y_{\star i}$ are independent of each other, $NI\left(L^*_{\star i}, Y_{\star i}\right)$ reaches its minimum value 0. When $Y_{\star i}$ is exactly the same as $L^*_{\star i}$, $NI\left(L^*_{\star i}, Y_{\star i}\right)$ equals to the maximum value 1. A smaller NI corresponds to a higher level of data privacy of $L^*_{\star i}$ and also indicates more significant difference between $L^*_{\star i}$ and $Y_{\star i}$. Note that rigorously speaking, $L^*_{\star i}$ belongs to the continuous space. However, since all measuring devices have a finite resolution, $L^*_{\star i}$ can be viewed as a discrete random variable. When computing NI in practice, one can divide the range of the values into small regions to compute sample probability distribution.

The above information-theoretic measures show that when the data of individual users are processed separately, a user's data privacy is enhanced at the cost of reduced information accuracy. We need to emphasize that the measures like NI or KL divergence focus on an individual signal and do not characterize the information recovery when multiple signals are processed together. In fact, when the data of multiple users are available, and strong correlations exist among different users' data, such correlation can be leveraged to enhance the data accuracy. As stated in problems (P1) and (P2), the major technical objective of this paper is to develop data recovery and clustering methods from quantized measurements of multiple users, where the data correlations are characterized by data points belonging to the same subspace. As we will show in Section 3 (Theorem 1 and Proposition 1), the asymptotic information accuracy from quantized measurements can be achieved when the number of users increases to the infinity. We need to emphasize that this result does not contradict the data privacy enhancement by adding noise and applying quantization. This is because the asymptotic information accuracy is only achieved when processing the correlated data of a large number of users, while a cyber intruder is very unlikely to have access to the data of so many users. In our proposed decentralized data collection and processing framework (Fig. 1), each agent collects the measurements of a subset of users, and the measurements are not directly shared among the agents. A cyber intruder needs to hack either all these agents or the smart meters of all users to be able to access all the data. Since such attack is very unlikely to happen, the user's data privacy is still protected. Privacy from the recovery perspective will be discussed in details in Section 3.3.

3 Results: theoretical

Here, we consider solving (P1) at a single center and defer the discussion of solving (P2) in a decentralized way through distributed nodes to Section 4. We propose to estimate L^*, C^*, and E^* by the solution $\left(\hat{L}, \hat{E}, \hat{C}\right)$ to the following optimization problem,

$$\min_{L,E\in\mathbb{R}^{m\times n},C\in\mathbb{R}^{n\times n}} F(L,E) \quad \text{s.t.}(L,E,C)\in\mathcal{S}_f, \tag{3}$$

where

$$F(L,E) = -\sum_{(i,j)\in\Omega}\sum_{l=1}^{K}\mathbf{1}_{[Y_{i,j}=\mathcal{Q}_l]}\log\left(\varphi_l\left(L_{i,j}+E_{i,j}\right)\right), \tag{4}$$

$$\begin{aligned}\mathcal{S}_f = \{&(L,E,C) : \|L\|_\infty \le \alpha_1, \|E\|_\infty \le \alpha_2, \|E\|_0 \le s,\\ &\text{rank}(L) \le r, L = LC, \|C_{\star i}\|_0 \le d, C_{i,i} = 0, \forall i \in [n]\}.\end{aligned} \tag{5}$$

$\mathbf{1}_{[A]}$ is the indicator function that takes value 1 if A is true and value 0 otherwise. $\|\cdot\|_0$ measures the number of nonzero entries in a vector or matrix. Data recovery and subspace clustering are achieved simultaneously by solving (3)–(5).

Equations (3)–(5) are a constrained maximum log-likelihood estimation problem that maximizes the likelihood of obtaining Y_Ω when the underlying data matrix is $\hat{L}$, and the error matrix is $\hat{E}$. The formulation follows (8) of [27] by extending from full observations to partial observations in Ω. After obtaining $\hat{C}$, spectral clustering [41] is applied to $\hat{C}$ to obtain group labels.

Equations (3)–(5) are nonconvex due to the nonconvexity of the feasible set $\mathcal{S}_f$ in (5). We first analyze the recovery and clustering performance, assuming that a solution exists. We defer the algorithm to Section 4.

3.1 Data recovery guarantee

Two constants γ_α and L_α are needed for the recovery analysis,

$$\gamma_\alpha = \min_{l\in[K]} \inf_{|x|\le\alpha_1+\alpha_2} \left\{ \frac{\dot{\varphi}_l^2(x)}{\varphi_l^2(x)} - \frac{\ddot{\varphi}_l(x)}{\varphi_l(x)} \right\}, \tag{6}$$

$$L_\alpha = \max_{l\in[K]} \sup_{|x|\le\alpha_1+\alpha_2} \{|\dot{\varphi}_l(x)|/\varphi_l(x)\}, \tag{7}$$

where $\dot{\varphi}_l(x)$ and $\ddot{\varphi}_l(x)$ are the first- and second-order derivatives with respect to x. Note that $\dot{\varphi}_l(x)^2 - \ddot{\varphi}_l(x)\varphi_l(x) > 0$ if φ_l is strictly log-concave. One can check that φ_l is strictly log-concave if Ψ is log-concave, which holds true for Gaussian and logistic distributions [28]. L_α and γ_α are bounded by some fixed constants when α_1, α_2, and φ_l are given.

Since the data recovery performance and the clustering performance are coupled together, we first analyze the recovery performance, assuming that the clustering results are not "arbitrarily bad." We follow the same assumption as [27], which essentially requires that in the estimated clustering results, every cluster contains data points belong to at most a constant number out of p original subspaces. Formally, we have

Assumption 1 [27]*: Columns of $\hat{L}$ belong to $\hat{p}$ subspaces, each of which has a dimension smaller or equal to d. Columns in $\hat{L}$ with indices corresponding to columns of L^* in $S_i (i \in [p])$ belong to at most $(g-1)$ subspaces, where g is a constant larger than 1.*

We follow the assumption in [28] about the location of the observed entries. We make a minor change to handle multiple subspaces instead of one subspace in [28]. Assumption 2 is a generalization of the uniform sampling and includes the uniform sampling as a special case. We define a binary matrix G with $G_{i,j} = 1$ if and only if $(i,j) \in \Omega$, i.e., $Y_{i,j}$ is observed. $G_{i,j} = 0$ otherwise. Let $G_i \in \mathbb{R}^{m\times n_i}$ denote the submatrix of G with columns corresponding to subspace i.

Assumption 2 *Assume each column of G_i has h nonzero entries. Let $\sigma_1(G_i)$ and $\sigma_2(G_i)$ denote the largest and the second largest singular values of G_i, respectively. Assume $\sigma_1(G_i) \ge h$ and $\sigma_2(G_i) \le \mathcal{C}\sqrt{h}$ for $i \in [p]$, where $\mathcal{C}$ is a positive constant.*

Assumption 2 is similar to the sampling assumption in [28]. The difference is that we make the assumption on columns belonging to each subspace instead of the whole matrix. The above assumption is more general than the uniform sampling assumption [28].

Theorem 1 *Suppose that $\varphi_l(x)$ is strictly log-concave in x, $\forall l \in [K]$. Then, under Assumptions 1 and 2, with probability at least $1 - pC_1e^{-C_2\xi n/p}$, any global minimizer $\hat{L}$ to (3)–(5) satisfies*

$$\left\|\hat{L} - L^*\right\|_F / \sqrt{mn} \leq \min\left(2\alpha_1 + 2\alpha_2\sqrt{\frac{s}{mn}}, U_1\right), \tag{8}$$

where

$$\begin{aligned} U_1 =& C_1'\frac{\kappa d\sqrt{d}}{f^2\sqrt{m}} + C_2'\frac{d\kappa^{3/4}}{f^{3/2}m^{1/4}}\left(\frac{s}{mn}\right)^{1/4} \\ &+ C_3'\frac{\sqrt{\kappa d}}{f}\left(\frac{s}{mn}\right)^{1/2} \end{aligned} \tag{9}$$

for some positive constants C_1, C_2, $C_1'(L_\alpha, g, \xi)$, $C_2'(L_\alpha, g, \xi, \alpha_2)$, and $C_3'(L_\alpha, g, \xi, \alpha_2)$. $f = \frac{|\Omega|}{mn} = \frac{h}{m}$ is the data loss rate.

Theorem 1 characterizes the recovery error from partially observed, partially corrupted, and quantized measurements. It can be interpreted from the following aspects.

(1) Correction of corrupted measurements. We first fix the data loss rate f and consider the recovery performance with corrupted measurements. Suppose f is a constant, i.e., a constant fraction of the measurements are available. Then, (8) indicates that as long as the number of corrupted measurements s is at most $\Theta\left(md^2p\right)$, we have[3]

$$\left\|\hat{L} - L^*\right\|_F / \sqrt{mn} \leq \mathcal{O}\left(\sqrt{\frac{d^3}{m}}\right). \tag{10}$$

Thus, the recovery method tolerates a constant number of corrupted per column without degrading the recovery performance.

(2) Asymptotic recovery of the actual data. Since $\mathcal{O}\left(\sqrt{\frac{d^3}{m}}\right)$ decreases to 0 when m increases to infinity, and $\|L^*\|_F$ is in the order of $\sqrt{mn}$, (10) indicates that the relative error between $\hat{L}$ and L^* diminishes asymptotically. Moreover, as long as p is $o(n)$, the failure probability $1 - pC_1e^{-C_2\xi n/p}$ also decays to zero as n increases to infinity. The asymptotic recovery differentiates the operating center and cyber intruders. An operating center with a sufficient number of measurements can recover L^* accurately. In contrast, a cyber intruder with access to a small number of users cannot recover the data even using the same approach (3)–(5).

(3) Tolerance of the missing data. To the best of our knowledge, only refs. [28] and [31] provided the theoretical analysis of low-rank matrix recovery from quantized observations with data losses. No corruptions are considered in [28, 31]. The relative recovery error by [28] is $\mathcal{O}\left(\sqrt{\frac{r^3}{m}}\right)$ under the partial observation case when f is a fixed constant, where r is the rank of the matrix. The relative recovery error by [31] is $\mathcal{O}\left(\frac{r^{1/4}}{m^{1/4}}\right)$ under the partial observation case. Our result in (10) indicates that when f is a constant, the error is at most $\mathcal{O}\left(\sqrt{\frac{d^3}{m}}\right)$ even with corrupted measurements. Note that the rank of L^* can be as large as pd when the subspaces are all orthogonal to each other. If one directly applies the approach in [37] to our setup, the relative recovery error can be as large as

[3] We use the notations $u(n) \in \mathcal{O}(v(n))$, $u(n) \in o(v(n))$, or $u(n) = \Theta(v(n))$ if as n goes to infinity, $u(n) \leq c \cdot v(n)$, $u(n) \geq c \cdot v(n)$ or $c_1 \cdot v(n) \leq u(n) \leq c_2 \cdot v(n)$ eventually holds for some positive constants c, c_1 and c_2, respectively.

$\mathcal{O}\left(\sqrt{\frac{p^3d^3}{m}}\right)$, which is $\sqrt{p^3}$ times our recovery error. Thus, our approach outperforms the existing one by recovering and clustering data simultaneously even in the special case of no corruptions.

When there is no missing data, the recovery error by [27] is $\mathcal{O}\left(\sqrt{\frac{d}{m}}\right)$, which is slightly tighter than our error bound in (10). This is due to our techniques to handle the missing data.

3.2 Fundamental limit of any recovery method

The following theorem establishes the minimum possible error by any method from unquantized measurements. We consider the case that the number of corruptions is at most a constant fraction of the measurements. To simplify the analysis, we assume

$$s \leq \min\left(C_0 mn, mn - \frac{64m}{d}\right) \tag{11}$$

where C_0 is a constant smaller than $1/2$. Let

$$\mathcal{S}_{fX} = \left\{X : X = L + E, (L, E, C) \in \mathcal{S}_f\right\}. \tag{12}$$

Theorem 2 *Let $N \in \mathbb{R}^{m\times n}$ contain i.i.d. entries from $\mathcal{N}\left(0, \sigma^2\right)$. Assume (11) holds. Consider any algorithm that, for any $X \in \mathcal{S}_{fX}$, takes $M_{ij} = X_{ij} + N_{ij}, (i,j) \in \Omega$ as the input and returns an estimate $\hat{X}$ of X. Then, there always exists some $X \in \mathcal{S}_{fX}$ such that with probability at least $\frac{3}{4}$,*

$$\frac{\left\|\hat{X} - X\right\|_F}{\sqrt{mn}} \geq \min\left(C_3, C_4\sigma\sqrt{\frac{d - \frac{d}{n}\left\lfloor\frac{s}{m}\right\rfloor - \frac{64}{n}}{fm - \frac{s_\Omega}{n}}}\right) \tag{13}$$

holds for some fixed constants C_3 and C_4, where $C_3 = \sqrt{\frac{1-2C_0}{8}}\min(\alpha_1, \alpha_2)$ and $C_4 < \sqrt{\frac{1-2C_0}{256}}$. s_Ω is the number of errors in X_Ω.

Note that C_3 is a constant. When f is a constant, (13) indicates that

$$\|\hat{X} - X\|_F/\sqrt{mn} \geq \Theta(\sqrt{d/m}). \tag{14}$$

The recovery error from unquantized measurements is at least $\Theta\left(\sqrt{\frac{d}{m}}\right)$. Comparing it with our error bound $\sqrt{\frac{d^3}{m}}$ in (10), one can see that our method is close to optimal. If the corrupted entries are randomly distributed, s_Ω is approximately $\Theta(fs)$. Then, the second term inside the minimization of (13) scales as $\Theta\left(\frac{1}{\sqrt{f}}\sqrt{\frac{d}{m}}\right)$.

3.3 Privacy from the recovery perspective

3.3.1 *Recovery of a single user from its own data only*

An intruder is often interested in the data of a certain user. If the adversary only has access to one user's data, then problems (3)–(5) are reduced to

$$\begin{aligned}&\min_{L,E\in\mathbb{R}^m} F(L,E)\\ &\text{s.t.}\|L\|_\infty \leq \alpha_1, \|E\|_\infty \leq \alpha_2, \|E\|_0 \leq s.\end{aligned} \tag{15}$$

Note that since $n = 1$, there is no constraint on C. (15) maximizes the log-likelihood of one user given the information about the quantized measurements. It can be viewed as a special case of the low-rank matrix recovery from quantized measurements considered in [37]. One can check that the average recovery error is upper bounded by $\mathcal{O}\left(\sqrt{d^3}\right)$ by setting $n = 1$ in Theorem 5 of [37]. Similarly, the relative recovery by any method is at least in the order of $\Theta(\sqrt{d})$ by setting $n = 1$ in Theorem 4 of [37]. This error bound does not depend on m, the number of measurements of this user. Therefore, if an intruder only has one user's data, even if m is very large, the average recovery error is nonzero and does not diminish as m increases. Then, the privacy of the energy consumption behavior of this user is protected.

3.3.2 Recovery of a single user by leveraging other users in the same group

One can exploit the measurements from other users to increase the estimation accuracy of one target user. Suppose one can access n users' data in m time steps, and these users all share similar load patterns as the target user, then from either Theorem 1 of this paper or Theorem 5 of [37], the average recovery error is at most $\mathcal{O}\left(\sqrt{\frac{d^3}{\min(m,n)}}\right)$. Compared with the previous case of accessing the data of one single user only, the recovery error is significantly reduced. We emphasize that the decrease of the recovery error results from exploiting correlations among users.

The number of quantization levels K also affects privacy. Intuitively, a smaller value of K corresponds to a higher level of privacy. However, the privacy level also depends on the selection of bin boundaries, and decreasing K does not necessarily increase privacy. For instance, if a pair of boundaries are chosen very close to each other so that no measurements located within the interval, then $K = 3$ could reach the same privacy and recovery error as $K = 2$. Therefore, K does not directly appear in Theorem 1 but rather affects the privacy indirectly through γ_α and L_α. The bin boundaries usually tend to be closer in the region where the measurements concentrate.

For smart meter data, the bin boundaries can be selected in the range of a typical household consumption level. If a certain house has some electrical appliances with an energy consumption level significantly higher than normal households, this abnormal pattern of high energy consumption can in fact be masked in the noisy and quantized measurements due to the way how bin boundaries are selected. However, since this house has a different load pattern from other households, one cannot exploit other users' data to enhance the recovery accuracy of this user. The recovered data of this user will have a nonzero error as discussed in the first paragraph of Section 3.3.

3.4 Clustering guarantee

The clustering performance is evaluated through the subspace-preserving property of $\hat{C}$. A sufficient condition for $\hat{C}$ to be subspace-preserving is stated as follows.

Proposition 1 *Suppose columns of $\hat{L}$ are i.i.d. drawn from certain unknown continuous distribution supported on $\hat{p}$ distinct d-dimensional subspaces, then the global minimizer $\hat{C}$ of (3) has the subspace-preserving property for $\hat{L}$.*

Ref. [27] also provides a sufficient condition for $\hat{C}$ to be subspace-preserving. The subspaces are required to be independent with each other in [27]. Two independent

subspaces intersect only at zero. Here, we require subspaces to be distinct from each other. Two subspaces are distinct if for each subspace, there exists one point that belongs to this subspace but not the other. The data points are generated based on some continuous distribution supported on these distinct subspaces.

4 Distributed sparse alternative proximal algorithm for data recovery and clustering

We next propose a distributed algorithm to solve (3) by W nodes collaboratively such that node i can estimate $L^*_{\Phi_i}$ from its acquired measurements Y_{Φ_i}, while it does not know Y_{Φ_j} or $L^*_{\Phi_j}$ for all other j's nodes. This further enhances user privacy.

We first follow [27] and move some constraints to the objective function to simplify the algorithm design. Since the rank of L is at most r, we factorize L as $L = UV^T$, where $V \in \mathbb{R}^{n\times r}$ and $U \in \mathbb{R}^{m\times r}$. We replace the equality constraints $L = LC$ and $L = UV^T$ by adding $\frac{\lambda_1}{2}\left\|V^T - V^TC\right\|_F^2$ and $\frac{\lambda_2}{2}\left\|UV^T - L\right\|_F^2$ to the objective function. The parameters λ_1 and λ_2 affect the tightness of the original constraints. Note that $V^T = V^TC$ is a sufficient but not necessary condition for $L = LC$. Then, (3) is changed into

$$\begin{aligned}&\left(\hat{U},\hat{V},\hat{L},\hat{E},\hat{C}\right)\\ &= \underset{\substack{U\in\mathbb{R}^{m\times r},V\in\mathbb{R}^{n\times r}\\ L,E,C\in\mathbb{R}^{m\times n}}}{\arg\min} H(U,V,L,E,C)\ \text{s.t.}(L,E,C)\in\mathcal{SF},\end{aligned} \tag{16}$$

where

$$\begin{aligned}H(U,V,L,E,C) =& F(L,E) + \frac{\lambda_1}{2}\left\|V^T - V^TC\right\|_F^2\\ &+ \frac{\lambda_2}{2}\left\|UV^T - L\right\|_F^2,\end{aligned} \tag{17}$$

$$\begin{aligned}\mathcal{SF} =& \{(L,E,C) : \|L\|_\infty \le \alpha_1, \|E\|_\infty \le \alpha_2,\\ &\|E\|_0 \le s, \|C_{\star i}\|_0 \le d, C_{i,i} = 0, \forall i \in [n]\},\end{aligned} \tag{18}$$

The solution of (16) is the same as that of (3) when λ_1 and λ_2 approach the infinity.

We next decompose V into W parts, and let $V_{\Phi_i\star} \in \mathbb{R}^{q\times r}, i \in [W]$ denote the rows of V with row indices Φ_i. Then, the objective in (17) can be decomposed as follows:

$$H(U,V,L,E,C) = \sum_{i=1}^{W} \mathcal{H}\left(U,V,L_{\Phi_i},E_{\Phi_i},C_{\Phi_i}\right) \tag{19}$$

where

$$\begin{aligned}&\mathcal{H}\left(U,V,L_{\Phi_i},E_{\Phi_i},C_{\Phi_i}\right)\\ &= F\left(L_{\Phi_i},E_{\Phi_i}\right) + \frac{\lambda_1}{2}\left\|V^T_{\Phi_i\star} - V^TC_{\Phi_i}\right\|_F^2\\ &\quad + \frac{\lambda_2}{2}\left\|UV^T_{\Phi_i\star} - L_{\Phi_i}\right\|_F^2,\end{aligned} \tag{20}$$

$$\begin{aligned}F\left(L_{\Phi_i},E_{\Phi_i}\right) = -\sum_{\substack{(k,j+iq-q)\\ \in\Omega,\\ \forall k\in[m],j\in[q]}}\sum_{l=1}^{K}&\\ \mathbf{1}_{[(Y_{\Phi_i})_{k,j}=l]}\log\left(\varphi_l\left((L_{\Phi_i})_{k,j} + \left(E_{\Phi_i}\right)_{k,j}\right)\right).&\end{aligned} \tag{21}$$

and V contains $V_{\Phi_1\star}$ to $V_{\Phi_W\star}$, i.e., $V = \begin{bmatrix} V_{\Phi_1\star} \\ V_{\Phi_2\star} \\ \vdots \\ V_{\Phi_W\star} \end{bmatrix}$. The constraint set $\mathcal{SF}$ in (18) is equivalent to the intersection of $\mathcal{SF}_i$'s ($\forall j \in [q]$), with[4]

$$\mathcal{SF}_i = \Big\{ \left(L_{\Phi_i}, E_{\Phi_i}, C_{\Phi_i}\right) : \left\|L_{\Phi_i}\right\|_\infty \le \alpha_1, \left\|E_{\Phi_i}\right\|_\infty \le \alpha_2, \left\|E_{\Phi_i}\right\|_0 \le \frac{s}{W}, \left\|(C_{\Phi_i})_{\star j}\right\|_0 \le d, (C_{\Phi_i})_{iq-q+j,j} = 0 \Big\}. \tag{22}$$

Then, (16) can be equivalently written as

$$\begin{aligned} &\left(\hat{U}, \hat{V}_{\Phi_i\star}, \hat{L}_{\Phi_i}, \hat{E}_{\Phi_i}, \hat{C}_{\Phi_i}\right) \\ &= \underset{\substack{C_{\Phi_i}\in\mathbb{R}^{n\times q}, U\in\mathbb{R}^{m\times r} \\ V_{\Phi_i\star}\in\mathbb{R}^{q\times r} \\ L_{\Phi_i}, E_{\Phi_i}\in\mathbb{R}^{m\times q}, \forall i\in[W]}}{\arg\min} \sum_{i=1}^{W} \mathcal{H}\left(U, V, L_{\Phi_i}, E_{\Phi_i}, C_{\Phi_i}\right) \\ &\qquad \text{s.t.} \left(L_{\Phi_i}, E_{\Phi_i}, C_{\Phi_i}\right) \in \mathcal{SF}_i, \forall i \in [W]. \end{aligned} \tag{23}$$

where the estimated variables are U and W components of L, E, C, and V.

The constraints in (23) can be decomposed for W nodes, while the objective function cannot, due to the coupling of U and V. Here, we develop a synchronized Distributed Sparse Alternative Proximal Algorithm (DSAPA) to solve (23) with the convergence guarantee. The node i owns Y_{Φ_i} and estimates $V_{\Phi_i\star}$, L_{Φ_i}, E_{Φ_i}, C_{Φ_i}, and U. Since all nodes have the estimates of U, and $L_{\Phi_i} = UV_{\Phi_i\star}$, the key to protect user privacy of node i is not to share the estimate of $V_{\Phi_i\star}$, as well as Y_{Φ_i}, to any other nodes.

In the $(t+1)$th iteration, node i sequentially updates $C_{\Phi_i}^{t+1}$, $V_{\Phi_i\star}^{t+1}$, $L_{\Phi_i}^{t+1}$, $E_{\Phi_i}^{t+1}$, U^{t+1} in Subroutines 1–5. Each subroutine essentially follows the projected gradient. The gradient of H with respect to $V_{\Phi_i\star}$, L_{Φ_i}, E_{Φ_i}, U, and C_{Φ_i} are

$$\begin{aligned} \nabla_{C_{\Phi_i}} H &= -\lambda_1 V\left(V_{\Phi_i\star}^T - V^T C_{\Phi_i}\right) \\ &= -\lambda_1 V\left(V_{\Phi_i\star}^T - V_{\Phi_i\star}^T \left(C_{\Phi_i}\right)_{\Phi_i\star} - \sum_{j=1, j\neq i}^{W} V_{\Phi_j\star}^T \left(C_{\Phi_i}\right)_{\Phi_j\star}\right) \\ &:= -\lambda_1 V M_{\Phi_i}, \end{aligned} \tag{24}$$

$$\begin{aligned} \nabla_{V_{\Phi_i\star}} H &= \lambda_2\left(V_{\Phi_i\star}U^T - L_{\Phi_i}^T\right)U \\ &\quad + \lambda_1\left[\left(V_{\Phi_i\star} - C_{\Phi_i}^T V\right) - \sum_{j=1}^{W}\left(C_{\Phi_j}\right)_{\Phi_i\star}\left(V_{\Phi_j\star} - C_{\Phi_j}^T V\right)\right] \\ &= \lambda_1\left(M_{\Phi_i}^T - \sum_{j=1}^{W}\left(C_{\Phi_j}\right)_{\Phi_i\star} M_{\Phi_j}^T\right) + \lambda_2\left(V_{\Phi_i\star}U^T - L_{\Phi_i}^T\right)U, \end{aligned} \tag{25}$$

$$\nabla_{L_{\Phi_i}} H = \nabla F\left(L_{\Phi_i}, E_{\Phi_i}\right) - \lambda_2\left(UV_{\Phi_i\star}^T - L_{\Phi_i}\right), \tag{26}$$

$$\nabla_{E_{\Phi_i}} H = \nabla F\left(L_{\Phi_i}, E_{\Phi_i}\right), \tag{27}$$

[4] We assume for simplicity that the corruptions are distributed evenly such that the number of nonzero entries in E_{Φ_i} is at most $\frac{s}{W}$. The algorithm can be easily extended to cases that the numbers of corruptions are different as long as a reasonable accurate upper bound of the number of corruptions is available.

$$\nabla_U H = \lambda_2 \left(UV^T - L \right) V := \lambda_2 \left(U \sum_{i=1}^{W} \iota_i - \sum_{i=1}^{W} \zeta_i \right), \tag{28}$$

where $M = V^T - V^T C$, $\iota_i = V_{\Phi_i \star}^T V_{\Phi_i \star}$, $\zeta_i = L_{\Phi_i} V_{\Phi_i \star}$, and

$$[\nabla F(L_{\Phi_i}, E_{\Phi_i})]_{k,j} = \frac{\dot{\Psi}\left(\omega_{(Y_{\Phi_i})_{k,j}} - (X_{\Phi_i})_{k,j}\right) - \dot{\Psi}\left(\omega_{(Y_{\Phi_i})_{k,j}-1} - (X_{\Phi_i})_{k,j}\right)}{\Psi\left(\omega_{(Y_{\Phi_i})_{k,j}} - (X_{\Phi_i})_{k,j}\right) - \Psi\left(\omega_{(Y_{\Phi_i})_{k,j}-1} - (X_{\Phi_i})_{k,j}\right)}, \tag{29}$$

$\forall k \in [m], j \in [q]$.

The step sizes in the $(t+1)$th iteration are selected as

$$\tau_C = \frac{1}{\lambda_1 \left\| V^t (V^t)^T \right\|_F} = \frac{1}{\lambda_1 \left\| \sum_{i=1}^{W} \iota_{\Phi_i} \right\|_F}, \tag{30}$$

$$\tau_{V_{\Phi_i \star}} = \frac{1}{e_U^t + \lambda_1 \max_{i \in [W]} F_i^t}, \tag{31}$$

$$\tau_{L_{\Phi_i}} = \frac{1}{\frac{1}{\sigma^2 \beta^2} + \lambda_2}, \tag{32}$$

$$\tau_{E_{\Phi_i}} = \sigma^2 \beta^2, \tag{33}$$

and

$$\tau_U = \frac{1}{\lambda_2 \left\| \left(V^{t+1}\right)^T V^{t+1} \right\|_F} = \frac{1}{\lambda_2 \left\| \sum_{i=1}^{W} \iota_{\Phi_i} \right\|_F}, \tag{34}$$

where $e_U = \lambda_2 \left\| \left(U^t\right)^T U^t \right\|_F$, $F_i^t = \left\| I_{q \times q} + \left(C_{\Phi_i \star}^{t+1}\right) \cdot \left(C_{\Phi_i \star}^{t+1}\right)^T - \left(C_{\Phi_i}\right)_{\Phi_i \star}^{t+1} - \left((C_{\Phi_i})_{\Phi_i \star}^{t+1} \right)^T \right\|_F$.

These step sizes are no greater than the reciprocals of the smallest Lipschitz constants of $\nabla_{C_{\Phi_i}} H$, $\nabla_{V_{\Phi_i \star}} H$, $\nabla_{L_{\Phi_i}} H$, $\nabla_{E_{\Phi_i}} H$, and $\nabla_U H$ in the tth iteration, respectively. Details of the calculations are shown in Appendix 6. This property is useful for the convergence analysis of the DSAPA.

The constraints in (22) are met by projecting the updated estimates to $\mathcal{SF}_i$. For the constraints on C_{Φ_i}, in steps 10–15 of Subroutine 1, we first set diagonal entries of $(C_{\Phi_i})_{\Phi_i \star}^{t+1}$ to zero. Then, we keep the d entries with the largest absolute value of $(C_{\Phi_i})_{\star j}^{t+1}$ and set all other entries to zero for any $j \in [q]$. The infinity norm on L_{Φ_i} is met by setting all entries larger than α_1 to be α_1 and setting all entries smaller than $-\alpha_1$ to be $-\alpha_1$ (step 4 in Subroutine 3). A similar approach applies to E_{Φ_i}. We also keep $\frac{s}{W}$ entries with the largest absolute values and set other nonzero entries to zero (steps 3–6 in Subroutine 4).

Note that L_{Φ_i} and E_{Φ_i} can be updated by node i independently and are not shared with other nodes. Updating C_{Φ_i}, V_{Φ_i}, and U needs communication from other nodes due to the coupling in the objective function. V_{Φ_i} cannot be shared with other nodes, since otherwise other nodes can estimate L_{Φ_i} by multiplying U and V_{Φ_i}. Thus, node i computes the intermediate terms that depend on V_{Φ_i} and send to other nodes instead of sending V_{Φ_i}, as illustrated in Fig. 3.

The algorithm is initialized as follows. $L_{\Phi_i}^0$ in node i is defined as,

$$(L_{\Phi_i}^0)_{k,j} = \begin{cases} \frac{\omega_l - \omega_{l-1}}{2} & \text{if } (Y_{\Phi_i})_{k,j} = l, 0 < l < K \\ \frac{\alpha_1 - \omega_{K-1}}{2} & \text{if } (Y_{\Phi_i})_{k,j} = K \\ \frac{\alpha_1 - \omega_1}{2} & \text{if } (Y_{\Phi_i})_{k,j} = 0 \end{cases} \tag{35}$$

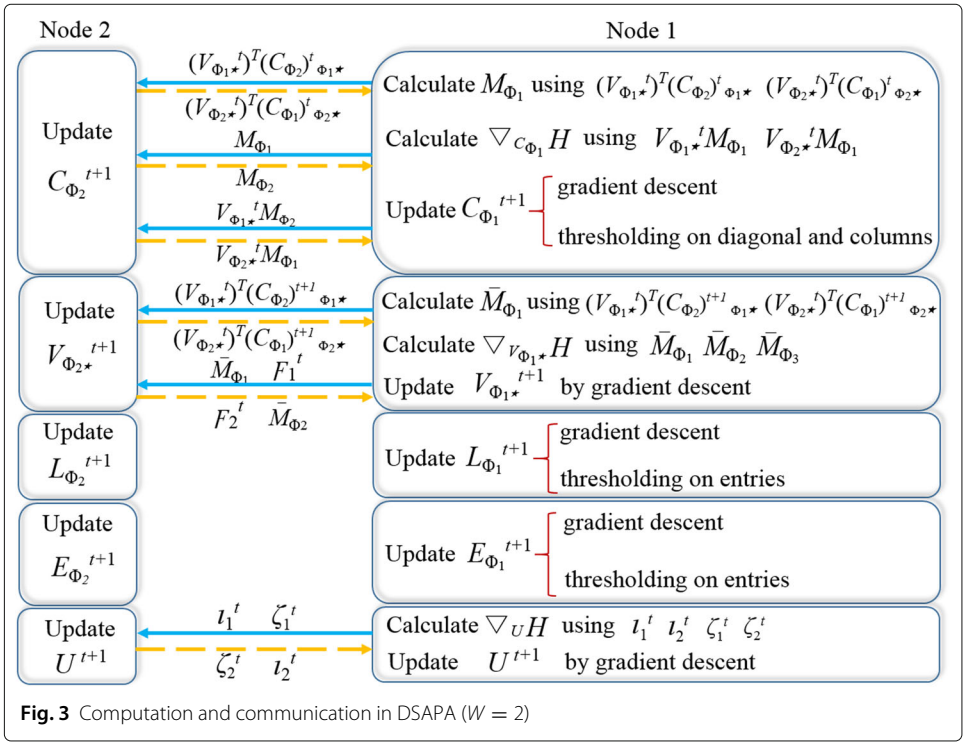

Fig. 3 Computation and communication in DSAPA ($W = 2$)

Then, node i performs the truncated singular value decomposition on $L_{\Phi_i}^0$ and let $U_i^{(r)}\Sigma_i^{(r)}(V_{\Phi_i\star}^{(r)})^T$ denote the rank-r approximation to $L_{\Phi_i}^0$. Then, node i transmits $U_i^{(r)}$ to all other nodes. Each node initializes at

$$U^0 = \frac{1}{W}\sum_{i=1}^{W} U_i^{(r)}\left(\Sigma_i^{(r)}\right)^{1/2}, \tag{36}$$

$$V_{\Phi_i\star}^0 = \left(L_{\Phi_i}^0\right)^T U^0\left(\left(U^0\right)^T U^0\right)^{-1}, \text{ and} \tag{37}$$

$$E_{\Phi_i}^0 = C_{\Phi_i}^0 = 0. \tag{38}$$

The convergence of DSAPA is summarized as follows.

Theorem 3 *From any initial point, DSAPA always converges to a critical point of (23).*

The computational complexities of Subroutines 1–5 are $\mathcal{O}(nqr)$, $\mathcal{O}(mqr)$, $\mathcal{O}(mq)$, $\mathcal{O}(mq)$, and $\mathcal{O}(mqr)$, respectively. The per-node per-iteration complexity of DSAPA is $\mathcal{O}(nqr)$. In contrast, the complexity of the centralized algorithm in [27] is $\mathcal{O}(nmr)$. The communication cost of Subroutines 1, 2, and 5 are $\mathcal{O}(n^2)$, $\mathcal{O}(nWr)$, and $\mathcal{O}(mWr)$, respectively.

For data clustering, a central node collects $\hat{C}_{\Phi_i}$ from all the nodes and applies spectral clustering [41] to obtain the clustering results.

When λ_1 and λ_2 are large enough, (23) approximates (3), but the step sizes in (30)–(32) and (34) are small and that reduces the convergence rate. One practical solution is to dynamically increase λ_1 and λ_2 [55]. We suggest the following practical selection. Initialize with small λ_1 and λ_2, and replace λ_2 with $\rho\lambda_2$ ($\rho > 1$) for the first T_0 iterations. Then, reset λ_2 to the initial value and update them with $\rho\lambda_1$ and $\rho\lambda_2$ simultaneously in each iteration. The algorithm terminates after T iterations.

Subroutine 1 Iterate C_{Φ_i} in the i-th distributed node

1: Compute $\left(V_{\Phi_i\star}^t\right)^T \left(C_{\Phi_j}\right)_{\Phi_i\star}^t, \forall j \in [\,W]$.
2: Send $\left(V_{\Phi_i\star}^t\right)^T \left(C_{\Phi_j}\right)_{\Phi_i\star}^t$ to the j-th node $\forall j \in [\,W]\,, j \neq i$.
3: Compute $M_{\Phi_i} = \left(V_{\Phi_i\star}^t\right)^T - \left(V_{\Phi_i\star}^t\right)^T \left(C_{\Phi_i}\right)_{\Phi_i\star}^t - \sum_{j=1,j\neq i}^{W} \left(V_{\Phi_j\star}^t\right)^T (C_{\Phi_i})_{\Phi_j\star}^t$.
4: Send M_{Φ_i} to the j-th node $\forall j \in [\,W]\,, j \neq i$.
5: Compute $V_{\Phi_i\star}^t M_{\Phi_j}, \forall j \in [\,W]$.
6: Send $V_{\Phi_i\star}^t M_{\Phi_j}$ to the j-th node, $\forall j \in [\,W]\,, j \neq i$.
7: Compute $\nabla_{C_{\Phi_i}} H$ according to (24).
8: Compute $\tau_{C_{\Phi_i}^t}$ according to (30).
9: Compute $C_{\Phi_i}^{t+1} = C_{\Phi_i}^t - \tau_{C_{\Phi_i}^t} \nabla_{C_{\Phi_i}} H$.
10: Set $\left(C_{\Phi_i}\right)_{iq-q+j,j}^{t+1} = 0, \forall j \in [\,q]$
11: **for** every $j = 1, 2, ..., q$ **do**
12: **if** $\sum_k \mathbf{1}_{\left[\left(C_{\Phi_i}\right)_{k,j}^{t+1} \neq 0\right]} > d$, **then**
13: $\left(C_{\Phi_i}\right)_{\star j}^{t+1}$ only keeps d entries with the largest absolute values. Other nonzero entries are set to be zero.
14: **end if**
15: **end for**
16: Send $\left(C_{\Phi_i}\right)_{\Phi_j\star}^{t+1}$ to the j-th node, $\forall j \in [\,W]\,, j \neq i$.

Subroutine 2 Iterate $V_{\Phi_i\star}$ in the i-th distributed node

1: Compute $\left(V_{\Phi_i\star}^t\right)^T \left(C_{\Phi_j}\right)_{\Phi_i\star}^{t+1}, \forall j \in [\,W]$.
2: Send $\left(V_{\Phi_i\star}^t\right)^T \left(C_{\Phi_j}\right)_{\Phi_i\star}^{t+1}$ to the j-th node $\forall j \in [\,W]\,, j \neq i$.
3: Compute $\bar{M}_{\Phi_i} = \left(V_{\Phi_i\star}^t\right)^T - \left(V_{\Phi_i\star}^t\right)^T \left(C_{\Phi_i}\right)_{\Phi_i\star}^{t+1} - \sum_{j=1,j\neq i}^{W} \left(V_{\Phi_j\star}^t\right)^T \left(C_{\Phi_i}\right)_{\Phi_j\star}^{t+1}$, and F_i^t.
4: Send $\bar{M}_{\Phi_i}, F_i^t$ to the j-th node $\forall j \in [\,W]\,, j \neq i$.
5: Compute $\nabla_{V_{\Phi_i\star}} H$ according to (25).
6: Compute $\tau_{V_{\Phi_i\star}^t}$ by (31).
7: Compute $V_{\Phi_i\star}^{t+1} = V_{\Phi_i\star}^t - \tau_{V_{\Phi_i\star}^t} \nabla_{V_{\Phi_i\star}} H$.

Subroutine 3 Iterate L_{Φ_i} in the i-th distributed node

1: Compute $\tau_{L_{\Phi_i}}$ by (32).
2: Compute $\nabla_{L_{\Phi_i}} H$ according to (26).
3: Compute $L_{\Phi_i}^{t+1} = L_{\Phi_i}^t - \tau_{L_{\Phi_i}} \nabla_{L_{\Phi_i}} H$.
4: If $\left(L_{\Phi_i}\right)_{k,j}^{t+1} > \alpha_1$, set $\left(L_{\Phi_i}\right)_{k,j}^{t+1} = \alpha_1$. If $\left(L_{\Phi_i}\right)_{k,j}^{t+1} < -\alpha_1$, set $\left(L_{\Phi_i}\right)_{k,j}^{t+1} = -\alpha_1$. $\forall k \in [\,m]\,, j \in [\,q]$.

5 Results: numerical experiments

We evaluate the performance on the Irish smart meter dataset (ISMD) [56] and the UMass smart* microgrid dataset (USMD) [57]. The ISMD consists of more than 5000 residential customers. The measurements are obtained every 30 min and have a unit of kilowatt

Subroutine 4 Iterate E_{Φ_i} in the i-th distributed node

1: Compute $\nabla_{E_{\Phi_i}} H$ according to (27).
2: Compute $E_{\Phi_i}^{t+1} = E_{\Phi_i}^{t} - \tau_{E_{\Phi_i}} \nabla_{E_{\Phi_i}} H$.
3: If $\left(E_{\Phi_i}\right)_{k,j}^{t+1} > \alpha_2$, set $\left(E_{\Phi_i}\right)_{k,j}^{t+1} = \alpha_2$. If $\left(E_{\Phi_i}\right)_{k,j}^{t+1} < -\alpha_2$, set $\left(E_{\Phi_i}\right)_{k,j}^{t+1} = -\alpha_2$. $\forall k \in [m], j \in [q]$.
4: **if** $\sum_j \sum_k \mathbf{1}_{[(E_{\Phi_i})_{k,j}^{t+1} \neq 0]} > s/W, \forall k \in [m], \forall j \in [q]$ **then**
5: $E_{\Phi_i}^{t+1}$ only keeps s/W entries with the largest absolute values. Other nonzero entries are set to be zero.
6: **end if**

Subroutine 5 Iterate U in the i-th distributed node

1: Compute $\iota_i^t = \left(V_{\Phi_i\star}^{t+1}\right)^T V_{\Phi_i\star}^{t+1}$ and send to all other nodes.
2: Compute $\zeta_i^t = L_{\Phi_i}^{t+1} V_{\Phi_i\star}^{t+1}$ and send to all other nodes.
3: Compute $\nabla_U H$ according to (28).
4: Compute τ_{U^t} by (34).
5: $U^{t+1} = U^t - \tau_{U^t} \nabla_U H$.

(kW). The UMSD contains 443 users in 24 h, and the power consumption is measured every minute. Some users have long sequences of zero power consumption, and some users have significantly high power consumption occasionally. We suspect these measurements have data quality issues resulting from devices or communication and remove these users from the datasets. We use 4780 customers in 30 days for ISMD and 438 customers in 6 h for USMD. Thus, the size of the data matrix L is 1440×4780 for ISND and 360×438 for USMD. The power consumption is at most 6 kW and 99 kW, respectively. Since the raw measurements are noisy, L is approximated by a rank-r matrix $L^*_{\text{rank-}r}$ by keeping only the largest r singular values. The recovery error is measured by $\|L^*_{\text{rank-}r} - \tilde{L}\|_F^2 / \|L^*_{\text{rank-}r}\|_F^2$, where $\tilde{L}$ is the recovered matrix. We choose r to be about 10% of the total number of the singular values. Then, r is set to 150 for ISMD and 40 for USMD. The following experiments are tested on ISMD, if not otherwise specified.

As described in Section 2.3, normalized mutual information is used to measure the data privacy. We now calculate the average normalized mutual information of 4780 users $\hat{NI} = \frac{1}{4780} \sum_{i=1}^{4780} NI(L_{\star i}, Y_{\star i})$. As a comparison, we also calculate the normalized mutual information between the noisy data (before quantization) and the actual data. The quantization level K is chosen as 2 or 5. The quantization boundaries and quantized values are summarized in Table 2 ($K = 2, 5$). We place more boundaries in the region where data concentrate. Selecting the optimal quantized boundaries is beyond the scope of this paper and will be left for the future work. We believe these parameters can be optimized if a small portion of ground-truth data are available for training. The noise level σ varies from 0.1 to 0.4 with a step size of 0.02. To compute the probabilistic distribution of $L_{\star i}$, we divide the range 0–6 kW into 100 or 300 equal intervals and compute the empirical distributions. As shown in Fig. 4, the normalized mutual information between the power after quantization and the actual power consumption is always smaller than that between the noisy value before quantization and the actual power consumption. This indicates the proposed quantization process enhances the data privacy. In addition, the norma-

Table 2 Quantization boundaries and quantized values

$\omega(K=2)$:	$\omega_0^*=-\infty, \omega_1^*=1$ kW, $\omega_2^*=\infty$
$Q(K=2)$:	$Q_1=0.5$ kW, $Q_2=3$ kW
$\omega(K=5)$:	$\omega_0^*=-\infty, \omega_1^*=0.25$ kW, $\omega_2^*=0.5$ kW, $\omega_3^*=1$ kW, $\omega_4^*=3$ kW, $\omega_5^*=\infty$
$Q(K=5)$:	$Q_1=0.2$ kW, $Q_2=0.4$ kW, $Q_3=0.85$ kW, $Q_4=2.5$ kW, $Q_5=4.5$ kW
$\omega(K=7)$:	$\omega_0^*=-\infty, \omega_1^*=0.5$ kW, $\omega_2^*=1$ kW, $\omega_3^*=3$ kW, $\omega_4^*=5$ kW, $\omega_5^*=10$ kW, $\omega_6^*=20$ kW, $\omega_7^*=\infty$
$Q(K=7)$:	$Q_1=0.2$ kW, $Q_2=0.7$ kW, $Q_3=2$ kW, $Q_4=4$ kW, $Q_5=7$ kW, $Q_6=15$ kW, $Q_7=35$ kW

lized mutual information $\hat{NI}$ decreases when either K decreases or σ increases. That is consistent with the intuition.

Since no ground-truth clustering result exists for this dataset, we define an index CI to evaluate the clustering performance. Let a_j denote the maximum angle of all the data points in group j to the estimated subspace of this group. Let b_j denote the minimum angle of any point in group j to the other subspaces. The clustering index CI measures the clustering accuracy and is defined as

$$CI = \frac{1}{N}\sum_{j=1}^{N}\frac{b_j - a_j}{\max\{a_j, b_j\}}. \tag{39}$$

CI is large if a_j's are small and b_j's are large, which means that points in the same group are close to the subspace of that group and away from other groups. A larger CI corresponds to a better clustering result. We apply Sparse Subspace Clustering (SSC) [21] to this dataset with different cluster numbers and compare the resulting CI's. We use the Alternating Direction Method of Multipliers (ADMM) [58] to solve SSC. When the number of clusters is $p=4$, we obtain the maximum $CI=0.085$. Thus, we set the number of clusters to be 4 in the following experiments.

We generate corruptions E^* and noise N randomly. The nonzero entries of E^* are selected from $[-4,-0.5]$ and $[0.5,4]$ uniformly. Every entry of N is drawn from the $\mathcal{N}(0,0.3^2)$. The quantization level K is set to 5. The locations of the missing data are selected randomly. The simulations run in MATLAB on a computer with 3.4 GHz Intel Core i7.

We evaluate DSAPA on the quantized measurements. We choose $W=5$ agents. We assume the upper bound of the magnitudes of the sparse error and the power consumption are known. For simplicity, we use the largest value of the given error and set $\alpha_2=4$.

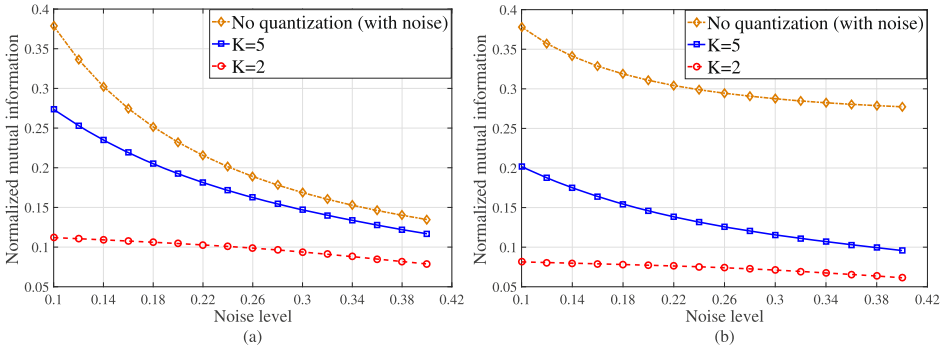

Fig. 4 Normalized mutual information between the original power usage and the quantized measurements. **a** 0.06 kW per interval. **b** 0.02 kW per interval

Similarly, we set $\alpha_1 = 6$. We set $d = 50$. λ_1 and λ_2 are initialized to be 0.5, and $\rho = 1.05$. The maximum iteration number T is set to be 200. T_0 is set to be 40.

Here, d is selected to be approximately $r/(p-1)$. We use $p-1$ considering the overlap between subspaces. We remark that varying d around the selected value does not affect the result. λ_1 and λ_2 are self-adjusted in our algorithm as discussed in the last paragraph of Section 4.

Figure 5 shows the energy consumption of a single user in 24 h. It compares the actual data, the rank-150 approximation of the actual data, the quantized observations, the recovered data by DSAPA, and the average quantized data of the users in the same group. One can see that the rank-150 approximation of the actual data has a similar pattern to the actual data. Clearly, the details of power consumption are hidden in the quantized measurements. For instance, the two peak consumptions are no longer visible in quantized measurements. Thus, an intruder does not know the user pattern if only accessing the quantized measurements of that user only. On the other hand, DSAPA recovers the power consumption trend accurately from the quantized data. The two peak loads are accurately identified in the recovered data as shown in Fig. 5. The recovered data can be used for grid planning.

After obtaining $\hat{C}$ using DSAPA, we implemented spectral clustering [41] to cluster the data points. To visualize the recovered consumption pattern of users in each group, we normalize the power consumptions and compute the average of users in the same group. Figure 6 shows the average profile obtained by our method in 1 day (no missing data and with 15% missing data). For comparison, the mean daily profile of the ground-truth data clustered by SSC is also shown in Fig. 6. One can see that the data losses do not affect the recovery performance of DSAPA. The recovered patterns are close to the actual patterns obtained by SSC, considering that the measurements are highly noisy and quantized. Now we pick some users in the same group and average the quantized value ($K = 5$) of these users. We calculate the normalized mutual information between one user and the averaged quantized value of the selected users. Figure 7 shows the normalized mutual information when the number of selected users varies. The value does not decrease

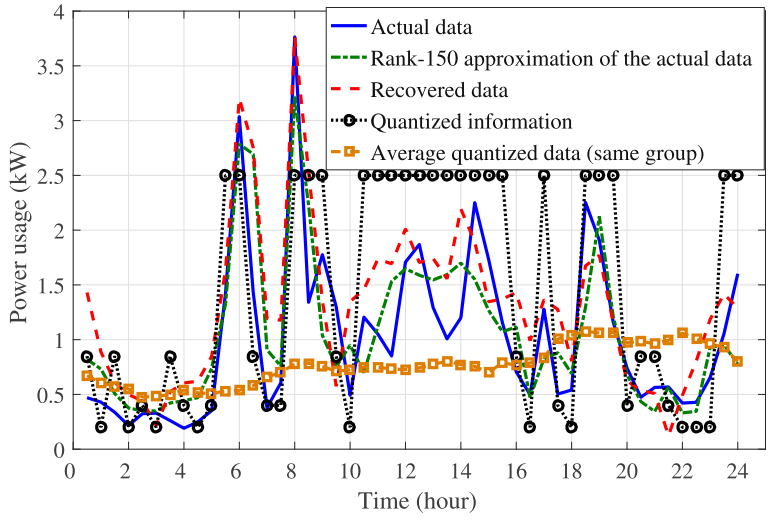

Fig. 5 Comparison of actual data, rank-150 approximation, recovered data, quantized data, and average quantized data in the same group

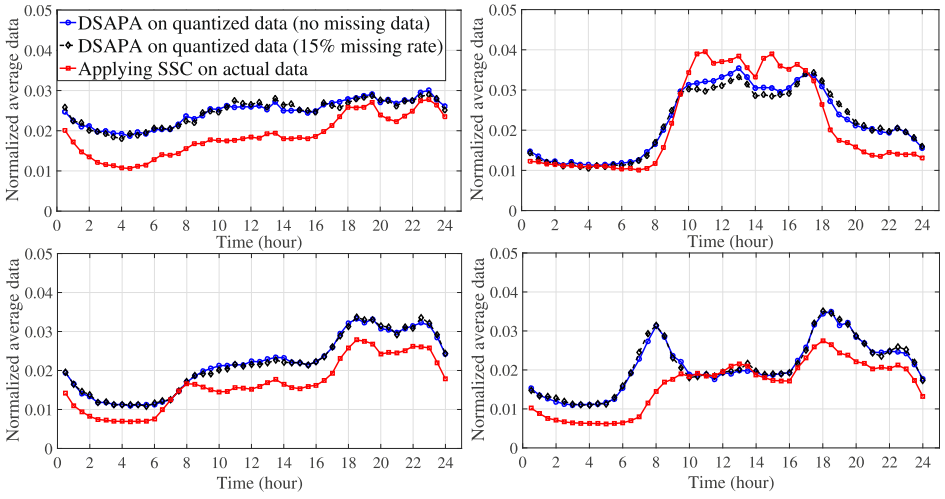

Fig. 6 Average consumptions of four different groups in 1 day (actual and recovered by DSAPA)

much when the number of users increases. Compared with Fig. 4b, one can see that the averaged quantized value of the same group does not provide much information to the single user.

We compare DSAPA with Approximate Projected Gradient Method (APGM) [28] and Quantized Robust Principal Component Analysis (QRPCA) [35] for data recovery in Fig. 8a. We apply SSC on the recovered data by APGM (or QRPCA) to obtain the clustering result, labeled by "APGM + SSC" ("QRPCA + SSC") in Fig. 8b. If we simply use the quantized value $Q_1, Q_2, \cdots, Q_5$ to estimate the actual power consumption, the relative recovery error is 0.869, which is much larger than the results in Fig. 8a. When the missing data rate changes from 0 to 0.4, our method always outperform the other methods both in data recovery and data clustering. For comparison, $CI = 0.085$ for SSC on the ground-truth data, and $CI = 0.05$ for a random clustering. Our method achieves $CI = 0.08$ using quantized measurements with 5% corruptions and no data losses.

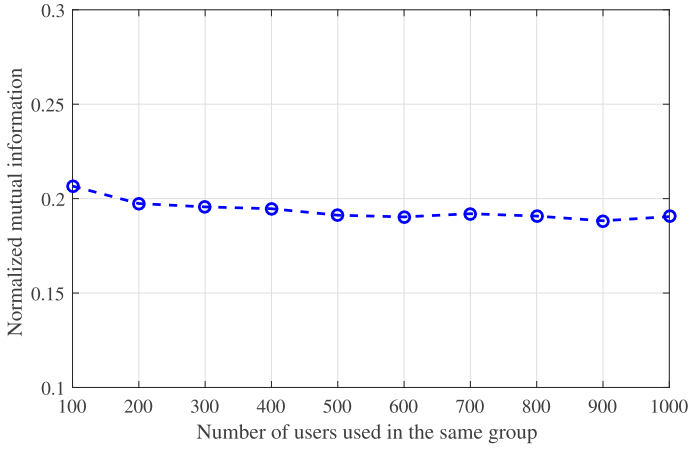

Fig. 7 Normalized mutual information between one user and the averaged quantized value of users in the same group (0.02 kW per interval)

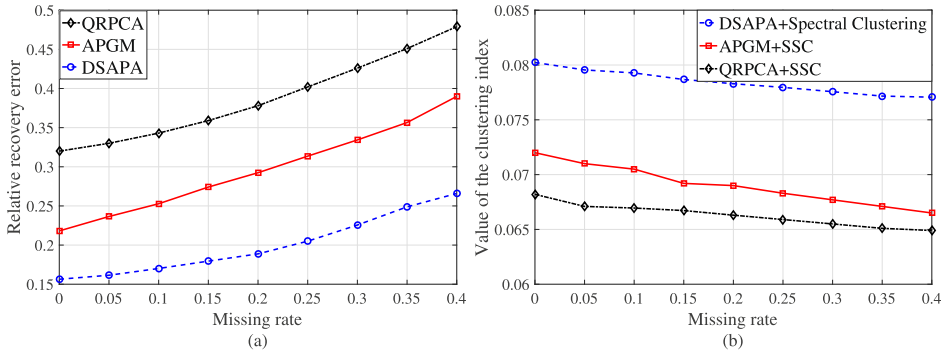

Fig. 8 Relative recovery error and clustering accuracy when the missing rate changes ($s/mn = 5\%$)

We vary the number of users by randomly selecting a subset of the 4780 users. Under the 15% missing rate and no corruption, Fig. 9 shows the recovery error when the number of users varies. The recovery error is 0.35 when the user number is to 500 and decreases to 0.2 when there are 2500 users.

We test the case when no additional noise is added before quantization. We vary the estimated noise level when implementing DSAPA since the measurements usually contain observation noise. As shown in Fig. 10, DSAPA can recover the data with no additional noise. However, adding no noise can lead to a low privacy level. The normalized mutual information when $K = 2$ and $K = 5$ are 0.2862 and 0.9579, respectively (0.02 kW per interval). These values are much higher than those shown in Fig. 4, indicating a lower level of privacy when no noise is added.

In Fig. 11, we compare the relative recovery error and the clustering index CI of DSAPA and the centralized algorithm Sparse-APA in [27]. Since Sparse-APA does not consider missing data, we study the case with full observations. The corruption rate is set as

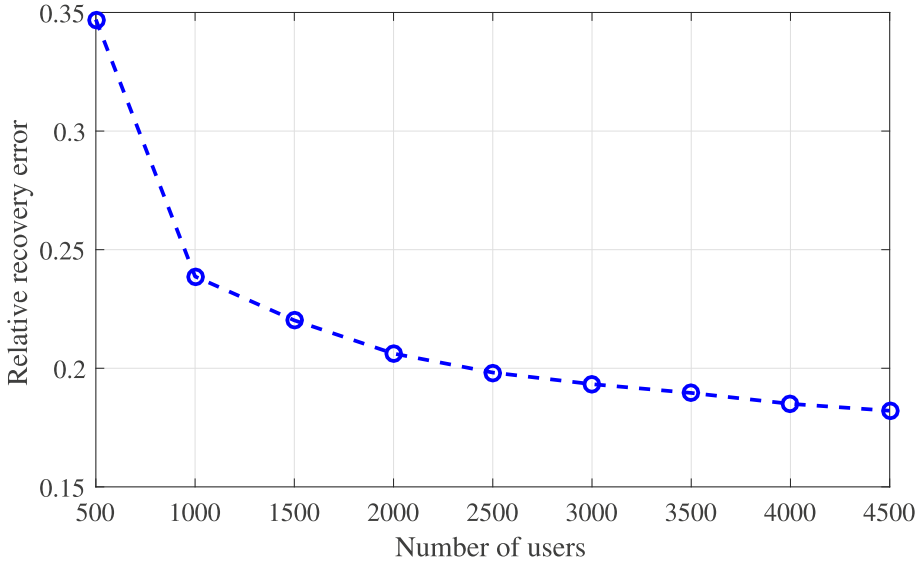

Fig. 9 Relative recovery error when the number of users increases

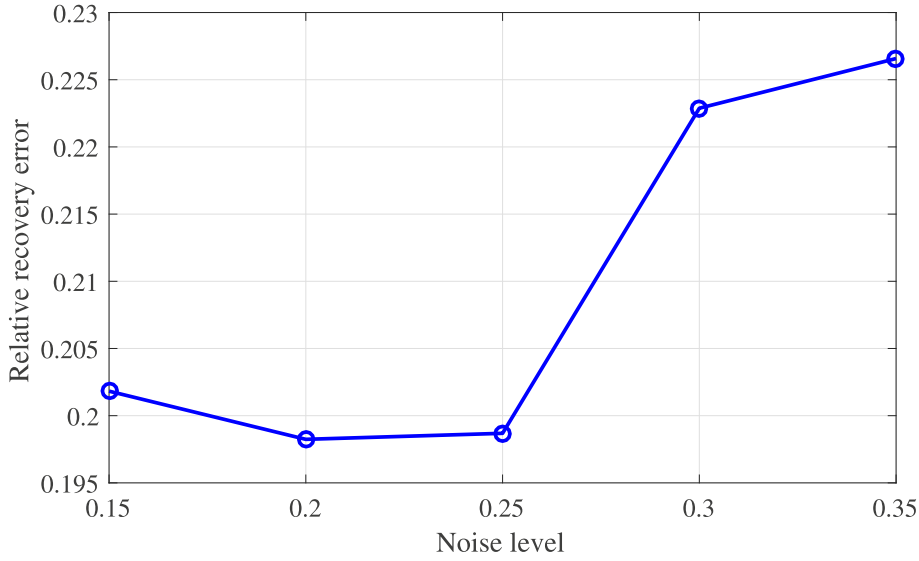

Fig. 10 Relative recovery error when noise level in the recovery process increases (without adding additional noise)

$s/mn = 5\%$. The recovery error of Sparse-APA is small than our method when the algorithm initializes, because Sparse-APA can compute a better initialization in a centralized fashion. However, the difference decreases as the iteration number increases. After 200 iterations, both algorithms perform similarly.

We next show the performance of DSAPA on USMD. Since the measurements vary from 0 to 100 kW, we set $K = 7$, and $\alpha_1 = 50$. The quantization boundaries and quantized values are in Table 2 ($K = 7$). p and d are set to be 4 and 15, respectively, using the same technique as discussed in the previous experiments. We generate the corruptions E^* and the noise N randomly. The nonzero entries of E^* are selected from $[-10, 10]$ uniformly, and the corruption rate is 5%. Every entry of N is drawn from the $\mathcal{N}(0, 0.3^2)$. Similar to Figs. 5 and 8a, we show the results on USMD in Fig. 12.

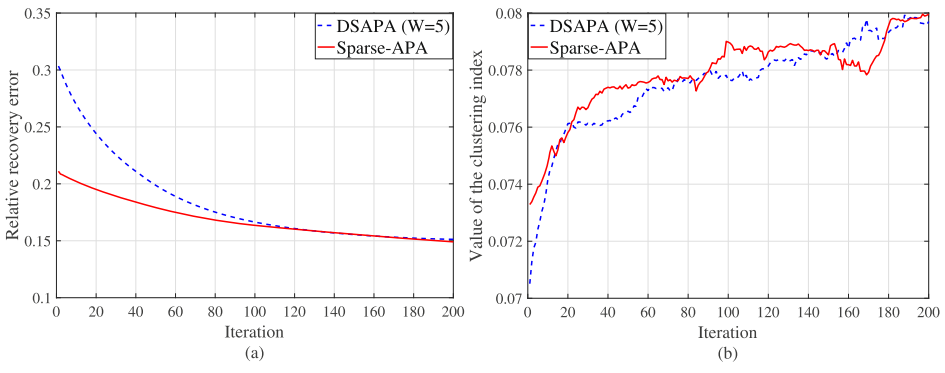

Fig. 11 Comparisons between DSAPA and Sparse-APA (centralized) [27]

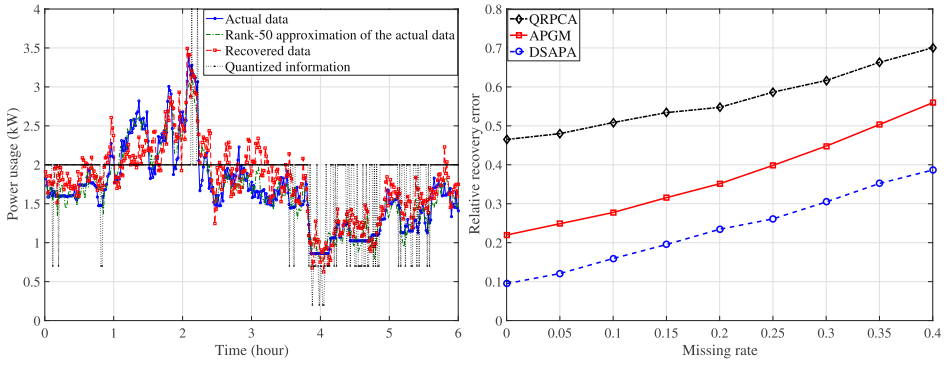

Fig. 12 **a** Comparison of actual data, rank-50 approximation, recovered data, and quantized data (USMD). **b** Relative recovery error when the missing rate changes

6 Conclusion and discussions

This paper for the first time shows that the two seemingly contradicting objectives of data privacy and information accuracy of smart meter data can be achieved simultaneously. The central technical contribution is the development of a decentralized data recovery and clustering method from highly quantized, partially lost, and partially corrupted measurements. Distributed nodes do not share raw data with each other and cannot estimate the actual data of other nodes. We propose a Distributed Sparse Alternative Proximal Algorithm (DSAPA) with a convergence guarantee to solve the nonconvex problem. The recovery error of our method is nearly optimal. The method is evaluated on actual smart meter datasets. Future works include leveraging the time correlation within each user to further improve the method and developing unsynchronized decentralized data recovery algorithms.

Appendix 1

Supporting lemmas used in the Proof of Theorem 1

Lemma 1 *Under Assumptions 1 and 2, the following inequalities hold*

$$\left\|\hat{L}_i - L_i^*\right\|_F \le \sqrt{2gd}a \left\|\left(\hat{L}_i - L_i^*\right)_{\Omega_i}\right\|_F + 2\sqrt{2gd}\alpha_1 b, \tag{40}$$

$$\left\|\hat{L} - L^*\right\|_F \le \sqrt{2gd}a \left\|\left(\hat{L} - L^*\right)_{\Omega}\right\|_F + 2\sqrt{2gdp}\alpha_1 b. \tag{41}$$

where $a = \frac{\sqrt{\xi mn}}{h\sqrt{p}}$, $b = \frac{\sqrt{\xi gdmn}\mathcal{C}}{\sqrt{hp}}$. Ω_i *includes the indices of the observed entries.*

Proof From Assumptions 1 and 2,

$$\begin{aligned}\left\|\hat{L}_i - L_i^*\right\|_F &\overset{(a)}{\le} \sqrt{2gd}\frac{\sqrt{\xi mn}}{\sigma_1(G_i)\sqrt{p}}\left\|\left(\hat{L}_i - L_i^*\right)_{\Omega_i}\right\|_F \\ &+ 2\sqrt{2gd}\alpha_1\frac{\sqrt{g\xi dmn}\sigma_2(G_i)}{\sigma_1(G_i)\sqrt{p}} \\ &\overset{(b)}{\le} \sqrt{2gd}\frac{\sqrt{\xi mn}}{h\sqrt{p}}\left\|\left(\hat{L}_i - L_i^*\right)_{\Omega_i}\right\|_F \\ &+ 2\sqrt{2gd}\alpha_1\frac{\sqrt{g\xi dmn}\mathcal{C}}{\sqrt{hp}}\end{aligned} \tag{42}$$

where (a) holds from Lemma 8 and Lemma 9 in [28], and the assumption $n_i \leq \xi n/p$. (b) holds because of $\sigma_1(G_i) \geq h$ and $\sigma_2(G_i) \leq \mathcal{C}\sqrt{h}$.

Then

$$\begin{aligned}
&\left\|\hat{L} - L^*\right\|_F^2 = \sum_{i=1}^{p} \left\|\hat{L}_i - L_i^*\right\|_F^2 \\
&\overset{(c)}{\leq} \sum_{i=1}^{p} \left(2gda^2 \left\| \left(\hat{L}_i - L_i^*\right)_{\Omega_i} \right\|_F^2 \right. \\
&\left. + 8gd\alpha_1 ab \|(\hat{L}_i - L_i^*)_{\Omega_i}\|_F + 8gd{\alpha_1}^2 b^2 \right) \\
&\overset{(d)}{\leq} 2gda^2 \left\| \left(\hat{L} - L^*\right)_{\Omega} \right\|_F^2 + 8gd\alpha_1 ab\sqrt{p} \left\| \left(\hat{L} - L^*\right)_{\Omega} \right\|_F \\
&+ 8gd{\alpha_1}^2 b^2 p \\
&= \left(\sqrt{2gd}a \left\| \left(\hat{L} - L^*\right)_{\Omega} \right\|_F + 2\sqrt{2gdp}\alpha_1 b \right)^2 .
\end{aligned} \tag{43}$$

where (c) follows from (40) (or (42)). (d) holds from $\sum_{i=1}^{p} \left\| \left(\hat{L}_i - L_i^*\right)_{\Omega_i} \right\|_F \leq \sqrt{p} \left\| \left(\hat{L} - L^*\right)_{\Omega} \right\|_F$. Then, we have the desired result. □

Lemma 2 *Let $\hat{\theta} = vec(\hat{X})$, $\theta^* = vec(X^*)$, $\mathcal{F}(\theta^*) = F(\theta^*)$, and $\hat{X}, X^* \in \mathcal{S}_{fX}$. Follow the same assumptions as those of Theorem 1. Then, with probability at least $1 - pC_1 e^{-C_2 \xi n/p}$,*

$$\begin{aligned}
&\left| \left\langle \nabla_\theta \mathcal{F}(\theta^*), \hat{\theta} - \theta^* \right\rangle \right| \\
&\leq 4.02 L_\alpha gda\sqrt{\xi n} \left\| \left(\hat{X} - X^*\right)_{\Omega} \right\|_F \\
&+ 8.04 L_\alpha gda\sqrt{\xi n}\alpha_2\sqrt{s} + 8.04 L_\alpha gd\sqrt{\xi np}\alpha_1 b \\
&+ 2\alpha_2 s L_\alpha ,
\end{aligned} \tag{44}$$

holds for the positive constants C_1 and C_2. $\langle ., . \rangle$ denotes the inner product of two matrices, i.e., the sum of entry-wise products.

Proof The proof is generalized from the proof of Lemma 2 in [27] which does not consider missing data. Here we extend the analysis to handle missing data. According to the definition, there exists a permutation matrix Γ^* such that L^* can be written as $L^* = \left[L_1^*, L_2^*, ..., L_p^*\right]\Gamma^*$. By Assumption 2, $\hat{L}$ can be written as $\hat{L} = \left[\hat{L}_1, \hat{L}_2, ..., \hat{L}_p\right]\Gamma^*$, where the dimension of $\hat{L}_i$ is smaller or equal to $(g-1)d$.

Note that $\sum_{l=1}^{K} \varphi_l\left(\left(X_i^*\right)_{k,j}\right) = 1$ and $\left[L_\alpha^{-1}\nabla_X F\left(X_i^*\right)\right]_{k,j} = -L_\alpha^{-1}\sum_{l=1}^{K} \frac{\dot{\varphi}_l\left(\left(X_i^*\right)_{k,j}\right)}{\varphi_l\left(\left(X_i^*\right)_{k,j}\right)} \cdot \mathbf{1}_{[(Y_i)_{k,j}=l]}$. Combining them with (7), one can conclude that the elements of $L_\alpha^{-1}\nabla_X F\left(X_i^*\right)$ have zero mean, and the variances are bounded by one. Using the result of Lemma 1 in [27], we have

$$\left\| L_\alpha^{-1}\nabla_X F\left(X_i^*\right) \right\|_2 \leq 2.01\sqrt{\xi n/p} \tag{45}$$

holds with probability at least $1 - C_1 e^{-C_2 \xi n/p}$. X_i^* is the same ith group as L_i^* under the permutation Γ^*. Then

$$
\begin{aligned}
&\left|\left\langle \nabla_\theta \mathcal{F}(\theta^*), \hat{\theta} - \theta^* \right\rangle\right| = \left|\left\langle \nabla_X F(X^*), \hat{X} - X^* \right\rangle\right| \\
&\leq \left|\left\langle \nabla_X F(X^*), \hat{L} - L^* \right\rangle\right| + \left|\left\langle \nabla_X F(X^*), \hat{E} - E^* \right\rangle\right| \\
&\overset{(a)}{=} \left|\sum_{i=1}^{p} \left\langle \nabla_X F(X_i^*), \hat{L}_i - L_i^* \right\rangle\right| \\
&\quad + \left|\left\langle \nabla_X F(X^*), \hat{E} - E^* \right\rangle\right| \\
&\overset{(b)}{\leq} \sum_{i=1}^{p} \left\| \nabla_X F\left(X_i^*\right) \right\|_2 \left\| \hat{L}_i - L_i^* \right\|_* + 2\alpha_2 s L_\alpha \\
&\overset{(c)}{\leq} 2.01 L_\alpha \sqrt{\xi n/p} \sum_{i=1}^{p} \sqrt{2gd} \left\| \hat{L}_i - L_i^* \right\|_F + 2\alpha_2 s L_\alpha \\
&\overset{(d)}{\leq} 2.01 L_\alpha \sqrt{\xi n/p} \sum_{i=1}^{p} (\sqrt{2gd}(\sqrt{2gd}a) \left\| \left(\hat{L}_i - L_i^*\right)_{\Omega_i} \right\|_F \\
&\quad + \sqrt{2gd}(2\sqrt{2gd}\alpha_1 b)) + 2\alpha_2 s L_\alpha \\
&\overset{(e)}{\leq} 2.01 L_\alpha (\sqrt{2gd}a) \sqrt{2\xi gdn} \left\| \left(\hat{L} - L^*\right)_{\Omega} \right\|_F \\
&\quad + 8.04 L_\alpha gd \sqrt{\xi np} \alpha_1 b + 2\alpha_2 s L_\alpha \\
&\leq 4.02 L_\alpha (\sqrt{gd}a) \sqrt{\xi gdn} \left\| \left(\hat{X} - X^*\right)_{\Omega} \right\|_F \\
&\quad + 4.02 L_\alpha (\sqrt{gd}a) \sqrt{\xi gdn} \left\| \left(\hat{E} - E^*\right)_{\Omega} \right\|_F \\
&\quad + 8.04 L_\alpha gd \sqrt{\xi np} \alpha_1 b + 2\alpha_2 s L_\alpha \\
&\overset{(f)}{\leq} 4.02 L_\alpha gda \sqrt{\xi n} \left\| \left(\hat{X} - X^*\right)_{\Omega} \right\|_F \\
&\quad + 8.04 L_\alpha gda \sqrt{\xi n} \alpha_2 \sqrt{s} \\
&\quad + 8.04 L_\alpha gd \sqrt{\xi np} \alpha_1 b + 2\alpha_2 s L_\alpha
\end{aligned}
$$

holds with probability at least $1 - pC_1 e^{-C_2 \xi n/p}$.

(a) holds from the linearity of the inner product. The first term of (b) holds from $|\langle A, B\rangle| \leq \|A\|_2 \|B\|_*$. The second term of (b) holds from the fact that both $\hat{E}, E^*$ have at most s nonzero entries and $|\nabla_X F(X^*)_{i,j}| \leq 1$. (c) holds from (45) and the fact $\left\| \hat{L}_i - L_i^* \right\|_* \leq \sqrt{2gd} \left\| \hat{L}_i - L_i^* \right\|_F$. (d) holds from Lemma (40). (e) holds from $\sum_{i=1}^{p} \left\| \left(\hat{L}_i - L_i^*\right)_{\Omega_i} \right\|_F \leq \sqrt{p} \|(\hat{L} - L^*)_\Omega\|_F$. (f) holds because $\|(\hat{E} - E^*)_\Omega\|_F \leq 2\alpha_2 \sqrt{s}$, which results from the fact that $|\hat{E}_{i,j} - E_{i,j}^*|$ is bounded by $2\alpha_2$. The probability $1 - pC_1 e^{-C_2 \xi n/p}$ comes from the union bound for $P(\max_{i \in [p]} \|\nabla_X F(X_i^*)\|_2 \leq 2.01 L_\alpha \sqrt{\xi n/p})$. □

Appendix 2

Proof of Theorem 1

Proof The proof follows and extends the proofs of Theorem 1 in [28] and Theorem 5 in [37]. We extend from the low-rank matrices in [28, 37] to matrices with columns in p low-dimensional subspaces. Moreover, ref. [28] does not consider corruptions, and ref. [37] does not consider missing data. Here we consider both missing data and corruptions.

The first bound $2\alpha_1 + 2\alpha_2\sqrt{\frac{s}{mn}}$ in (8) follows from the fact that $\hat{L}, L^*, \hat{E}, E^* \in \mathcal{S}_f$. We discuss the second bound in (8) as follows. We denote (4) to be $F(X)$ when we treat X to be the variable. Note that $\mathcal{S}_{fX}$ is a compact set, and the objective function is continuous in X. $F(X)$ then achieves a minimum in $\mathcal{S}_{fX}$. Suppose that $\hat{X} \in \mathcal{S}_{fX}$ minimizes $F(X)$.

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

$$\begin{aligned}\mathcal{F}_{\Omega,Y}(\theta) &= \mathcal{F}_{\Omega,Y}(\theta^*) + \left\langle \nabla_\theta \mathcal{F}_{\Omega,Y}(\theta^*), \theta - \theta^* \right\rangle \\ &+ \frac{1}{2}\left\langle \theta - \theta^*, (\nabla^2_{\theta\theta}\mathcal{F}_{\Omega,Y}(\tilde{\theta}))(\theta - \theta^*) \right\rangle,\end{aligned} \tag{46}$$

where $\tilde{\theta} = \theta^* + \bar{\eta}(\theta - \theta^*)$ for some $\bar{\eta} \in [0, 1]$, with corresponding matrices $\tilde{X} = X^* + \bar{\eta}(X - X^*)$.

From (46), Lemma 2, and Lemma A.3 in [38], we have

$$\begin{aligned}0 &\geq F(\hat{X}) - F(X^*) \\ &\geq -c_f \|(\hat{X} - X^*)_\Omega\|_F + \frac{\gamma_\alpha}{2}\|(\hat{X} - X^*)_\Omega\|_F^2 - \eta.\end{aligned} \tag{47}$$

holds with probability at least $1 - pC_1 e^{-C_2 \xi n/p}$ where $c_f = 4.02 L_\alpha g d a \sqrt{\xi n}$, $\eta = 8.04 L_\alpha g d a \sqrt{\xi n}\alpha_2\sqrt{s} + 8.04 L_\alpha g d \sqrt{\xi np}\alpha_1 b + 2\alpha_2 s L_\alpha$.

By solving (47), we then have

$$\|(\hat{X} - X^*)_\Omega\|_F \leq (c_f + \sqrt{c_f^2 + 2\gamma_\alpha \eta})/\gamma_\alpha. \tag{48}$$

Thus,

$$\begin{aligned}&\|\hat{L} - L^*\|_F/\sqrt{mn} \\ &\overset{(a)}{\leq} (\sqrt{2gd}a\|(\hat{L} - L^*)_\Omega\|_F + 2\sqrt{2gdp}\alpha_1 b)/\sqrt{mn} \\ &\leq (\sqrt{2gd}a)(\|(\hat{X} - X^*)_\Omega\|_F + \|(\hat{E} - E^*)_\Omega\|_F)/\sqrt{mn} \\ &\quad + 2\sqrt{2gdp}\alpha_1 b/\sqrt{mn} \\ &\overset{(b)}{\leq} (\sqrt{2gd}a)((c_f + \sqrt{c_f^2 + 2\gamma_\alpha\eta})/\gamma_\alpha + 2\alpha_2\sqrt{s})/\sqrt{mn} \\ &\quad + 2\sqrt{2gdp}\alpha_1 b/\sqrt{mn} \\ &\overset{(c)}{\leq} \frac{M_1 d^{\frac{3}{2}} n m^{\frac{1}{2}}}{h^2 p} + \frac{M_2 d^{\frac{5}{4}} n^{\frac{1}{2}} m^{\frac{1}{4}}}{h^{\frac{5}{4}} p^{\frac{1}{2}}} + \frac{M_3 d n^{\frac{1}{2}} m^{\frac{1}{4}} s^{\frac{1}{4}}}{h^{\frac{3}{2}} p^{\frac{3}{4}}} \\ &\quad + \frac{M_4 d^{\frac{1}{2}} s^{\frac{1}{2}}}{h p^{\frac{1}{2}}} + \frac{M_5 d}{h^{\frac{1}{2}}} \\ &\overset{(d)}{\leq} C_1' \frac{\kappa d\sqrt{d}}{f^2\sqrt{m}} + C_2' \frac{d\kappa^{3/4}}{f^{3/2} m^{1/4}} \left(\frac{s}{mn}\right)^{1/4} \\ &\quad + C_3' \frac{\sqrt{\kappa d}}{f} \left(\frac{s}{mn}\right)^{1/2},\end{aligned} \tag{49}$$

where M_1–M_5 are constants. (a) holds because of (41). (b) holds according to (48). (c) holds because of the Cauchy-Schwarz inequality. (d) holds because $f = h/m$, $\frac{M_2 d^{\frac{5}{4}} n^{\frac{1}{2}} m^{\frac{1}{4}}}{h^{\frac{5}{4}} p^{\frac{1}{2}}} = \frac{M_2 \kappa d^{\frac{5}{4}}}{f^{\frac{5}{4}} m^{\frac{1}{2}}}$, and $\frac{M_5 d}{h^{\frac{1}{2}}} = \frac{M_5 d}{f^{\frac{1}{2}} m^{\frac{1}{2}}}$. The order of both terms are smaller than $O\left(\frac{\kappa d\sqrt{d}}{f^2\sqrt{m}}\right)$. □

Appendix 3

Supporting lemmas for Theorem 2

Lemma 3 *There exists a set* $\mathcal{X} \subset \mathcal{S}_{fX}$ *with*

$$|\mathcal{X}| \geq \exp\left(\frac{dn - d\lfloor \frac{s}{m} \rfloor}{16}\right) \tag{50}$$

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

1. *For all* $X \in \mathcal{X}$, $X_{i,j} = \pm\alpha\gamma$ *or 0,* $\forall(i, j)$, *where* $\alpha = \min(\alpha_1, \alpha_2)$.
2. *For all* $X^{(i)}, X^{(j)} \in \mathcal{X}$, $i \neq j$,

$$\|X^{(i)} - X^{(j)}\|_F^2 > \alpha^2\gamma^2\left(\frac{mn}{2} - s\right). \tag{51}$$

Proof Now we independently generate a set $\mathcal{X}$ of $\left\lceil \exp\left(\frac{dn - d\lfloor \frac{s}{m} \rfloor}{16}\right)\right\rceil$ random matrices from the following distribution. According to columns' indices, X is first been divided into $X_1, X_2, \cdots, X_p$, which correspond to indices $\{1, ..., \lfloor \frac{n}{p} \rfloor\}, \{\lfloor \frac{n}{p} \rfloor + 1, ..., 2\lfloor \frac{n}{p} \rfloor\}, \{2\lfloor \frac{n}{p} \rfloor + 1, ..., 3\lfloor \frac{n}{p} \rfloor\}, \cdots, \{(p-1)\lfloor \frac{n}{p} \rfloor + 1, ..., n\}$, respectively. For the first d rows of X_1, fix the locations of $\lfloor \frac{s}{pm} \rfloor$ entries in each row and set the values to zero. The remaining $d\lfloor \frac{n}{p} \rfloor - d\lfloor \frac{s}{pm} \rfloor$ entries take values $\pm\alpha\gamma$ with equal probabilities. For all $i \in \{d + 1, ..., m\}, j \in [\lfloor \frac{n}{p} \rfloor]$,

$$X_{i,j} := X_{k,j}, \text{ where} k = i(\text{mod}d) + 1. \tag{52}$$

The same process is applied to $X_2, X_3, \cdots, X_p$. Then, one can see that X can be written as $X = L + E$, where L can span subspaces with dimension smaller or equal to d, and E is a sparse matrix. We further have

$$\|L\|_\infty = \alpha\gamma \leq \alpha_1, \quad \|E\|_\infty = \alpha\gamma \leq \alpha_2, \text{ and} \|E\|_0 \leq s. \tag{53}$$

Each column of L can be represented by at most d other columns. Thus, $\mathcal{X} \in \mathcal{S}_{fX}$.

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

$$\begin{aligned}
&\|X - \hat{X}\|_F^2 = \sum_{i,j}(X_{ij} - \hat{X}_{ij})^2 \\
&\geq \lfloor \frac{m}{d} \rfloor \sum_{i=1}^{d} \left(\sum_{j=1}^{\lfloor \frac{n}{p} \rfloor}(X_{ij} - \hat{X}_{ij})^2 + \sum_{j=\lfloor \frac{n}{p} \rfloor + 1}^{2\lfloor \frac{n}{p} \rfloor}(X_{ij} - \hat{X}_{ij})^2 \right. \\
&\left. + \cdots + \sum_{j=(p-1)\lfloor \frac{n}{p} \rfloor + 1}^{n}(X_{ij} - \hat{X}_{ij})^2 \right) \\
&\geq 4\alpha^2\gamma^2\lfloor \frac{m}{d} \rfloor \sum_{i=1}^{dn - d\lfloor \frac{s}{m} \rfloor} \delta_i,
\end{aligned} \tag{54}$$

where δ_i's are independent 0/1 Bernoulli random variables and the means are all $\frac{1}{2}$. Following the same proof technique of Lemma 4 in [37], one can show that $\mathcal{X}$ satisfies the property 2. □

Let $Y = X + N$, where the entries in matrix N are i.i.d. and generated from Gaussian distribution $\mathcal{N}(0, \sigma^2)$. Suppose that $X \in \mathcal{X}$ is chosen uniformly at random. Lemma 4 bounds the mutual information $I(X_\Omega, Y_\Omega)$.

Lemma 4

$$I(X_\Omega, Y_\Omega) \leq \frac{|\Omega| - s_\Omega}{2} \log\left(1 + \left(\frac{\alpha\gamma}{\sigma}\right)^2\right) \tag{55}$$

Proof The proof is similar to the proof of Lemma 5 in [31], but [31] does not consider corruptions. We modify the proof to handle corruptions. From Lemma 5 in [31], one can obtain

$$I(X_\Omega, Y_\Omega) \leq H(\tilde{X}_\Omega + N_\Omega) - H(N_\Omega). \tag{56}$$

where $\aleph$ denotes a matrix with all entries are i.i.d. generated from $\{+1, -1\}$. $\tilde{X} = X \cdot \aleph$ denotes the entry-wise product of X and $\aleph$.

The vectorization of $\tilde{X}_\Omega + N_\Omega$ is denoted by $\mathrm{vec}(\tilde{X}_\Omega + N_\Omega) \in \mathbb{R}^{|\Omega|}$. We compute the covariance matrix as

$$\Sigma := \mathbb{E}[\,\mathrm{vec}(\tilde{X}_\Omega + N_\Omega)\mathrm{vec}(\tilde{X}_\Omega + N_\Omega)^T\,]. \tag{57}$$

Then, by Theorem 8.6.5 in [59], we have

$$\begin{aligned} H(\tilde{X}_\Omega + N_\Omega) &\leq \frac{1}{2}\log((2\pi e)^{|\Omega|} det(\Sigma)) \\ &= \frac{1}{2}\log((2\pi e)^{|\Omega|}(\alpha^2\gamma^2 + \sigma^2)^{|\Omega|-s_\Omega}\sigma^{2s_\Omega}), \end{aligned} \tag{58}$$

The equality holds since $\tilde{X}$ has s_Ω zero entries.

We have $H(N_\Omega) = \frac{1}{2}\log((2\pi e)^{|\Omega|}\sigma^{2|\Omega|})$ and thus

$$I(X_\Omega, Y_\Omega) \leq \frac{1}{2}\log\left(\frac{(\alpha^2\gamma^2 + \sigma^2)^{|\Omega|-s_\Omega}\sigma^{2s_\Omega}}{\sigma^{2|\Omega|}}\right), \tag{59}$$

which establishes the lemma. □

Appendix 4

Proof of Theorem 2

Proof The proof follows Theorem 4 in [31] which does not consider the corruptions. Our proof is more involved due to the corruptions. Choose ϵ so that

$$\epsilon^2 = \min\{\frac{(1-2C_0)\alpha^2}{8}, C_4^2\sigma^2\frac{dn - d\lfloor\frac{s}{m}\rfloor - 64}{|\Omega| - s_\Omega}\} \tag{60}$$

where C_4 is a constant to be determined later. The set $\mathcal{X}$ is defined in Lemma 3. γ is set to be

$$\frac{2\epsilon}{\alpha}\sqrt{\frac{2mn}{mn-2s}} \leq \gamma \leq \frac{2\epsilon}{\alpha}\sqrt{\frac{2}{1-2C_0}} \leq 1. \tag{61}$$

Suppose for the sake of a contradiction that there exists an efficient algorithm such that for any $X \in \mathcal{S}_{fX}$, given the measurements Y, returns an $\hat{X}$, and

$$\|X - \hat{X}\|_F^2/mn \leq \epsilon^2 \tag{62}$$

holds with probability at least 1/4. Let

$$X^* = \arg\min_{X' \in \mathcal{X}} \|X' - \hat{X}\|_F^2. \tag{63}$$

Following the proof of Theorem 4 in [31], one can find that if (62) holds, then $X^* = X$. By the assumption of (62),

$$P(X \neq X^*) \leq 3/4. \tag{64}$$

Let X be a matrix chosen uniformly at random from $\mathcal{X}$. Considering running the algorithm on X, then by Fano's inequality, the probability that $X \neq X^*$ is at least

$$\begin{aligned} P(X \neq X^*) &\geq \frac{H(X|Y_\Omega) - 1}{\log|\mathcal{X}|} \\ &= \frac{H(X) - I(X, Y_\Omega) - 1}{\log|\mathcal{X}|} \geq 1 - \frac{I(X_\Omega, Y_\Omega) + 1}{\log|\mathcal{X}|}. \end{aligned} \tag{65}$$

We have obtained $|\mathcal{X}|$ from Lemma 3 and $I(X_\Omega, Y_\Omega)$ from Lemma 4. Then, using the inequality $\log(1+z) \leq z$, we obtain

$$P(X \neq \hat{X}) \geq 1 - \frac{16}{dn - d\lfloor \frac{s}{m} \rfloor} \left(\frac{|\Omega| - s_\Omega}{2} \left(\frac{\alpha\gamma}{\sigma} \right)^2 + 1 \right). \tag{66}$$

Combining (66) with (61) and (64), we obtain

$$\frac{16}{dn - d\lfloor \frac{s}{m} \rfloor} \left((|\Omega| - s_\Omega) \frac{4}{1 - 2C_0} \left(\frac{\epsilon}{\sigma} \right)^2 + 1 \right) \geq \frac{1}{4}, \tag{67}$$

which implies that

$$\epsilon^2 \geq \frac{(1 - 2C_0)\sigma^2}{256} \frac{dn - d\lfloor \frac{s}{m} \rfloor - 64}{|\Omega| - s_\Omega}. \tag{68}$$

Setting $C_4^2 < \frac{1-2C_0}{256}$ leads to a contradiction, hence (62) must fail to hold with probability at least 3/4. Using the definition $f = \frac{|\Omega|}{mn}$, we obtain the desired result. □

Appendix 5

Proof of Proposition 1

Proof Given any i, from (5), we know that $\hat{L}_{\star i} = \hat{L}\hat{C}_{\star i}$. Without loss of generality, we assume $\hat{L}_{\star i} \in \hat{S}_1$, where the $\hat{p}$ subspaces are denoted by $\hat{S}_i$ ($i \in [\hat{p}]$). Then, from the constraint $\hat{C}_{i,i} = 0, \forall i \in [n]$, we have $\hat{L}_{\star i} = [\hat{L}_{1\setminus\star i} \ \hat{L}_{-1}] \begin{bmatrix} \hat{C}_{\star i}^{(1\setminus\star i)} \\ \hat{C}_{\star i}^{(-1)} \end{bmatrix}$, where $\hat{L}_{1\setminus\star i}$ denotes all data points belonging to $\hat{S}_1$ except $\hat{L}_{\star i}$. $\hat{L}_{-1}$ denotes all data points belonging to $\{\hat{S}_j\}_{j=2}^{\hat{p}}$. $\hat{C}_{\star i}^{(1\setminus\star i)}$ and $\hat{C}_{\star i}^{(-1)}$ are sparse coefficients corresponding to $\hat{L}_{1\setminus\star i}$ and $\hat{L}_{-1}$, respectively. Now we only need to prove that $\hat{C}_{\star i}^{(-1)} = 0$.

If $\hat{C}_{\star i}^{(-1)} \neq 0$, then $\hat{L}_{\star i}$ belongs to a subspace $\hat{S}_1'$ which is different from $\hat{S}_1$, and spanned by data points corresponding to nonzero entries of $\begin{bmatrix} \hat{C}_{\star i}^{(1\setminus\star i)} \\ \hat{C}_{\star i}^{(-1)} \end{bmatrix}$. Moreover, the dimension of $\hat{S}_1'$ must be smaller or equal to d since $\| \begin{bmatrix} \hat{C}_{\star i}^{(1\setminus\star i)} \\ \hat{C}_{\star i}^{(-1)} \end{bmatrix} \|_0 \leq d$. Therefore, $\hat{L}_{\star i} \in \hat{S}_1'' = \hat{S}_1' \bigcap \hat{S}_1$, where $\bigcap$ denotes the intersection of two subspaces. We first consider the case when the dimension of $\hat{S}_1''$ is smaller than d. Since the data points of $\hat{L}_\star$ are sampled from a continuous distribution of $\hat{p}$ subspaces, the probability that the data point $\hat{L}_{\star i}$ lying in a data-point-spanned hyperplane in $\hat{S}_1$ that has dimension smaller than d is 0 (to see this, consider the probability of a data point lying in a pre-fix line within a plane). Next we show that the number of such hyperplanes is finite. Because the data points are fixed beforehand, there is only a finite number of combinations of data points that can span $\hat{S}_1'$ and further intersect with $\hat{S}_1$ to form $\hat{S}_1''$. Then, the probability of the union of a finite of combinations is still zero. Therefore, the dimension of $\hat{S}_1''$ equals to d, which indicates that the dimensions of $\hat{S}_1'$ and $\hat{S}_1$ are both d. This leads to $\hat{S}_1'' = \hat{S}_1' = \hat{S}_1$. This results in

a contradiction, since the data points corresponding to $\hat{C}_{\star i}^{(-1)} \neq 0$ do not belong to $\hat{S}_1$. Thus, $\hat{C}_{\star i}^{(-1)} = 0$, and the claim holds. □

Appendix 6

DSAPA: proof of the Lipschitz differential property and calculation of Lipschitz constants

A function is Lipschitz differentiable if and only if all its partial gradients are Lipschitz continuous. The definition is shown in Definition 3.

Definition 3 [60] *For any fixed matrices $z_1, z_2, .., z_n$, matrix variable y, and a function $y \to \Upsilon(y, z_1, z_2, ..., z_n)$, the partial gradient $\nabla_y \Upsilon(y, z_1, z_2, ..., z_n)$ is said to be Lipschitz continuous with Lipschitz constant $L_p(z_1, z_2, ..., z_n)$, if the following holds*

$$\begin{aligned}&\|\nabla_y \Upsilon(y, z_1, z_2, ..., z_n) - \nabla_y \Upsilon(y', z_1, z_2, ..., z_n)\|_F \\ &\leq L_p(z_1, z_2, ..., z_n)\|y - y'\|_F, \ \ \forall y, y'.\end{aligned}$$

We provide the Lipschitz differential property of H and compute the corresponding Lipschitz constants of its partial gradients with respect to $C_{\Phi_i}, V_{\Phi_i\star}, L_{\Phi_i}, E_{\Phi_i}, \forall i \in [W]$. Let L_{p1}^{t+1}, L_{p2}^{t+1}, L_{p3}^{t+1}, L_{p4}^{t+1}, and L_{p5}^{t+1} denote the smallest Lipschitz constants of $\nabla_{C_{\Phi_i}} H$, $\nabla_{V_{\Phi_i\star}} H$, $\nabla_{L_{\Phi_i}} H$, $\nabla_{E_{\Phi_i}} H$, and $\nabla_U H$ in the $(t+1)$th iteration. We have

$$\begin{aligned}&\|\nabla_{C_{\Phi_i}} H(C_{\Phi_i}) - \nabla_{C_{\Phi_i}} H(C'_{\Phi_i})\|_F \\ &= \|\lambda_1 V^t (V^t)^T (C_{\Phi_i} - C'_{\Phi_i})\|_F \\ &\leq \|\lambda_1 V^t (V^t)^T\|_F \|C_{\Phi_i} - C'_{\Phi_i}\|_F \\ &= \|\lambda_1 \sum_{i=1}^{W} \iota_i^t\|_F \|C_{\Phi_i} - C'_{\Phi_i}\|_F \\ &\overset{(a)}{=} \frac{1}{\tau_C(V^t)} \|C_{\Phi_i} - C'_{\Phi_i}\|_F,\end{aligned} \tag{69}$$

where (a) follows from (30). Equation (69) implies that

$$L_{p1}^{t+1} \leq \|\lambda_1 \sum_{i=1}^{W} \iota_i^t\|_F, \text{ and } \tau_C(V^t) \leq 1/L_{p1}^{t+1}. \tag{70}$$

$$\begin{aligned}&\|\nabla_{V_{\Phi_i\star}} H(V_{\Phi_i\star}) - \nabla_{V_{\Phi_i\star}} H(V'_{\Phi_i\star})\|_F \\ &= \|\lambda_2 (V_{\Phi_i\star} - V'_{\Phi_i\star})(U^t)^T U^t + \lambda_1 (V_{\Phi_i\star} - V'_{\Phi_i\star}) \cdot \\ &(I_{q\times q} - (C_{\Phi_i})_{\Phi_i\star}^{t+1} - ((C_{\Phi_i})_{\Phi_i\star}^{t+1})^T + (C_{\Phi_i\star}^{t+1})(C_{\Phi_i\star}^{t+1})^T)\|_F \\ &\overset{(b)}{\leq} \|V_{\Phi_i\star} - V'_{\Phi_i\star}\|_F \cdot (\|\lambda_2 (U^t)^T U^t\|_F + \lambda_1 \cdot \\ &\|I_{q\times q} + (C_{\Phi_i\star}^{t+1})(C_{\Phi_i\star}^{t+1})^T - (C_{\Phi_i})_{\Phi_i\star}^{t+1} - ((C_{\Phi_i})_{\Phi_i\star}^{t+1})^T\|_F) \\ &\overset{(c)}{\leq} \frac{1}{\tau_V(U^t, C^{t+1})} \|V_{\Phi_i\star} - V'_{\Phi_i\star}\|_F,\end{aligned} \tag{71}$$

where (b) follows from the triangle inequality, and (c) follows from (31). Equation (71) implies that

$$\begin{aligned}&L_{p2}^{t+1} \leq \max_{i\in[W]} \lambda_1 \|I_{q\times q} + (C_{\Phi_i\star}^{t+1})(C_{\Phi_i\star}^{t+1})^T - (C_{\Phi_i})_{\Phi_i\star}^{t+1} - \\ &((C_{\Phi_i})_{\Phi_i\star}^{t+1})^T\|_F + e_U^t, \text{ and } \tau_V(U^t, C^{t+1}) \leq 1/L_{p2}^{t+1}.\end{aligned} \tag{72}$$

$$
\begin{aligned}
&\|\nabla_{L_{\Phi_i}} H(L_{\Phi_i}) - \nabla_{L_{\Phi_i}} H(L'_{\Phi_i})\|_F = \\
&\|\nabla F(L_{\Phi_i}, E^t_{\Phi_i}) - \nabla F(L'_{\Phi_i}, E^t_{\Phi_i}) + \lambda_2 (L_{\Phi_i} - L'_{\Phi_i})\|_F \\
&\overset{(d)}{=} \|\mathrm{diag}(\nabla^2 F(\bar{L}_{\Phi_i}))\mathrm{vec}(L_{\Phi_i} - L'_{\Phi_i})\|_2 \\
&\quad + \lambda_2 \|L_{\Phi_i} - L'_{\Phi_i}\|_F \\
&\leq (\|\mathrm{diag}(\nabla^2 F(\bar{L}_{\Phi_i}))\|_2 + \lambda_2)\|L_{\Phi_i} - L'_{\Phi_i}\|_F \\
&\overset{(e)}{=} (\|\nabla^2 F(\bar{L}_{\Phi_i})\|_\infty + \lambda_2)\|L_{\Phi_i} - L'_{\Phi_i}\|_F \\
&\overset{(f)}{\leq} (\frac{1}{\sigma^2 \beta^2} + \lambda_2)\|L_{\Phi_i} - L'_{\Phi_i}\|_F \\
&\overset{(g)}{=} \frac{1}{\tau_L(E^t_{\Phi_i})}\|L_{\Phi_i} - L'_{\Phi_i}\|_F,
\end{aligned} \tag{73}
$$

where (d) comes from the differential mean value theorem. $\nabla^2 F(\bar{L}_{\Phi_i}) \in \mathbb{R}^{m \times q}$ has the (k,j)th entry equaling to $\frac{\partial^2 F}{\partial^2 (L_{\Phi_i})_{k,j}}|_{(\bar{L}_{\Phi_i})_{k,j}}$, and $\mathrm{diag}(\nabla^2 F(\bar{L}_{\Phi_i})) \in \mathbb{R}^{mq \times mq}$ is a diagonal matrix with the diagonal vector equaling to $\mathrm{vec}(\nabla^2 F(\bar{L}_{\Phi_i}))$. (e) follows from the fact that the l_2 norm of a diagonal matrix is equal to its entry-wise infinity norm. Note that (1) is lower bounded by β, and the probability density function of the normal distribution and its derivative are upper bounded by $\frac{1}{\sqrt{2\pi}\sigma}$ and $\frac{e^{-1/2}}{\sqrt{2\pi}\sigma^2}$, respectively. Then, one can easily check that $\|\nabla^2 F(\bar{L}_{\Phi_i})\|_\infty$ is bounded by $\frac{1}{\sigma^2\beta^2}$. (f) is thus obtained by upper bounding $\|\nabla^2 F(\bar{L}_{\Phi_i})\|_\infty$. (g) follows from (32). Thus, $\tau_L(E^t_{\Phi_i}) \leq \frac{1}{L^{t+1}_{p3}}$.

$$
\begin{aligned}
&\|\nabla_{E_{\Phi_i}} H\left(E_{\Phi_i}\right) - \nabla_{E_{\Phi_i}} H(E'_{\Phi_i})\|_F \\
&= \|\nabla F(L^{t+1}_{\Phi_i}, E_{\Phi_i}) - \nabla F(L^{t+1}_{\Phi_i}, E'_{\Phi_i})\|_F \\
&\overset{(h)}{=} \|\mathrm{diag}(\nabla^2 F(\bar{E}_{\Phi_i}))\mathrm{vec}(E_{\Phi_i} - E'_{\Phi_i})\|_F \\
&\leq \|\nabla^2 F(\bar{E}_{\Phi_i})\|_\infty \|E_{\Phi_i} - E'_{\Phi_i}\|_F \\
&\overset{(i)}{\leq} \frac{1}{\sigma^2\beta^2}\|E_{\Phi_i} - E'_{\Phi_i}\|_F \\
&\overset{(j)}{=} \frac{1}{\tau_E(L^{t+1}_{\Phi_i})}\|E_{\Phi_i} - E'_{\Phi_i}\|_F,
\end{aligned} \tag{74}
$$

where (h) follows from the differential mean value theorem. (i) is obtained by upper bounding $\|\nabla^2 F(\bar{E}_{\Phi_i})\|_\infty$ by $\frac{1}{\sigma^2\beta^2}$. (j) follows from (33). (74) implies that $\tau_E(L^{t+1}_{\Phi_i}) = \sigma^2\beta^2 \leq \frac{1}{L^{t+1}_{p4}}$.

$$
\begin{aligned}
&\|\nabla_U H(U) - \nabla_U H(U')\|_F \\
&= \|\lambda_2 (U - U')(V^t)^T V^{t+1}\|_F \\
&\leq \|\lambda_2 (V^{t+1})^T V^{t+1}\|_2 \|U - U'\|_F \\
&\overset{(k)}{\leq} \|\lambda_2 (V^{t+1})^T V^{t+1}\|_F \|U - U'\|_F \\
&\overset{(l)}{=} \|\lambda_2 \sum_{i=1}^{W} \iota_i^{t+1}\|_F \|U - U'\|_F \\
&\overset{(m)}{=} \frac{1}{\tau_U(V^{t+1})}\|U - U'\|_F,
\end{aligned} \tag{75}
$$

where (k) follows from the inequality $\|\cdot\|_2 \leq \|\cdot\|_F$. (l) follows from $(V^{t+1})^T V^{t+1} = \sum_{i=1}^{W} \iota_{\Phi_i}^{t+1}$. Since $\|\lambda_2 \sum_{i=1}^{W} \iota_{\Phi_i}^{t+1}\|_F \geq L_{p5}^{t+1}$, (m) follows from (34). (75) implies that $L_{p5}^{t+1} \leq \|\lambda_2 \sum_{i=1}^{W} \iota_i^{t+1}\|_F$, and $\tau_U(V^{t+1}) \leq 1/L_{p5}^{t+1}$.

Based on Definition 3, (69)–(75) guarantee the Lipschitz differentiability of H and provide the Lipschitz constants and the step sizes of the DSAPA.

Appendix 7

Proof of Theorem 3

Proof The constraints in (22) can be transferred to the following indicator functions.

$$K_1(C_{\Phi_i}) = \begin{cases} \infty & \text{if there exists a} \\ & (C_{\Phi_i})_{iq-q+j,j} \neq 0, \forall j \in [q] \\ 0 & \text{otherwise} \end{cases} \tag{76}$$

$$K_2(C_{\Phi_i}) = \begin{cases} \infty & \text{if there exists a} \\ & (C_{\Phi_i})_{\star j} \text{ s.t. } \|(C_{\Phi_i})_{\star j}\|_0 > d, \\ & j \in [q] \\ 0 & \text{otherwise} \end{cases} \tag{77}$$

$$B(L_{\Phi_i}) = \begin{cases} \infty & \text{if } \|L_{\Phi_i}\|_\infty > \alpha_1 \\ 0 & \text{otherwise} \end{cases} \tag{78}$$

$$J_1\left(E_{\Phi_i}\right) = \begin{cases} \infty & \text{if } \|E_{\Phi_i}\|_\infty > \alpha_2 \\ 0 & \text{otherwise} \end{cases} \tag{79}$$

$$J_2\left(E_{\Phi_i}\right) = \begin{cases} \infty & \text{if } \|E_{\Phi_i}\|_0 > s/W \\ 0 & \text{otherwise} \end{cases} \tag{80}$$

(76)–(80) correspond to the operations of projection in DSAPA.

Similar to the proof of Theorem 3 in [27], DSAPA globally converges to a critical point of (16) from any initial point, provided that H is Lipschitz differentiable, and

$$H + \sum_{i=1}^{W} (K_1(C_{\Phi_i}) + K_2(C_{\Phi_i}) + B(L_{\Phi_i}) + J_1\left(E_{\Phi_i}\right) + J_2\left(E_{\Phi_i}\right)) \tag{81}$$

satisfies the Kurdyka-Lojasiewicz (KL) property.

The proof of the Lipschitz differentiable property of H is shown in Appendix 6. $B(L_{\Phi_i})$, $J_1\left(E_{\Phi_i}\right)$, $J_2\left(E_{\Phi_i}\right)$, $K_1(C_{\Phi_i})$, and $K_2(C_{\Phi_i})$ are indicator functions of semi-algebraic sets. Therefore, they are KL functions according to [60]. Since H is differentiable everywhere, or equivalently, real analytic, H also has the KL property according to the examples in session 2.2 of [61]. Thus, (81) satisfies the KL property. □

Abbreviations

UoS: Union of Subspaces; DSAPA: Distributed Sparse Alternative Proximal Algorithm; c.d.f: Cumulative distribution function; SSC: Sparse Subspace Clustering; PMU: Phasor measurement unit; KL: Kullback-Leibler; NI: Normalized mutual information; APGM: Approximate projected gradient method; QRPCA: Quantized Robust Principal Component Analysis; NILM: Non-intrusive load monitoring

Acknowledgements

This research is supported in part by ARO W911NF-17-1-0407 and the Rensselaer-IBM AI Research Collaboration (http://airc.rpi.edu), part of the IBM AI Horizons Network (http://ibm.biz/AIHorizons).

Authors' contributions
Ren and Meng conceived and designed the method and the experiments. Ren performed the experiments and drafted the manuscript. Meng revised the manuscript. Jinjun provided many helpful suggestions. The authors read and approved the final manuscript.

Availability of data and materials
The Irish smart meter datasets that support the findings of this study are available from the Irish Social Science Data Archive (ISSDA) but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of the Irish Social Science Data Archive (ISSDA). The UMass smart* microgrid dataset analyzed during the current study is available in http://traces.cs.umass.edu/index.php/Smart/Smart.

Consent for publication
Informed consent was obtained from all authors included in the study.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Department of Electrical, Computer, and Systems Engineering, Rensselaer Polytechnic Institute, Troy, NY, USA. [2]IBM Thomas J. Watson Research Center, Yorktown Heights, NY, USA.

Received: 15 October 2019 Accepted: 17 April 2020
Published online: 07 May 2020

References

1. G. W. Hart, Nonintrusive appliance load monitoring. Proc. IEEE. **80**(12), 1870–1891 (1992)
2. E. J. Aladesanmi, K. A. Folly, Overview of non-intrusive load monitoring and identification techniques. IFAC-PapersOnLine. **48**(30), 415–420 (2015)
3. Z. Erkin, G. Tsudik, in *International Conference on Applied Cryptography and Network Security*, Private computation of spatial and temporal power consumption with smart meters (Springer, Singapore, 2012), pp. 561–577
4. P. Barbosa, A. Brito, H. Almeida, S. Clauß, in *Proceedings of the 29th Annual ACM Symposium on Applied Computing, SAC '14*, Lightweight privacy for smart metering data by adding noise (ACM, Gyeongju, 2014), pp. 531–538
5. J. M. Bohli, C. Sorge, O. Ugus, in *2010 IEEE International Conference on Communications Workshops*, A privacy model for smart metering (IEEE, Cape Town, 2010), pp. 1–5
6. M. Backes, S. Meiser, in *Data Privacy Management and Autonomous Spontaneous Security*, Differentially private smart metering with battery recharging (Springer, Berlin, 2014), pp. 194–212
7. D. Varodayan, A. Khisti, in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Smart meter privacy using a rechargeable battery: minimizing the rate of information leakage (IEEE, Prague, 2011), pp. 1932–1935
8. D. Egarter, C. Prokop, W. Elmenreich, in *2014 IEEE International Conference on Smart Grid Communications (SmartGridComm)*, Load hiding of household's power demand (IEEE, Venice, 2014), pp. 854–859
9. S. McLaughlin, P. McDaniel, W. Aiello, in *Proceedings of the 18th ACM Conference on Computer and Communications Security*, Protecting consumer privacy from electric load monitoring (ACM, Chicago, 2011), pp. 87–98
10. X. He, X. Zhang, C. C. J. Kuo, A distortion-based approach to privacy-preserving metering in smart grids. IEEE Access. **1**, 67–78 (2013)
11. M. Savi, C. Rottondi, G. Verticale, Evaluation of the precision-privacy tradeoff of data perturbation for smart metering. IEEE Trans. Smart Grid. **6**(5), 2409–2416 (2015)
12. O. Tan, D. Gunduz, H. V. Poor, Increasing smart meter privacy through energy harvesting and storage devices. IEEE J. Sel. Areas Commun. **31**(7), 1331–1341 (2013)
13. F. L. Quilumba, W.-J. Lee, H. Huang, D. Y. Wang, R. L. Szabados, Using smart meter data to improve the accuracy of intraday load forecasting considering customer behavior similarities. IEEE Trans. Smart Grid. **6**(2), 911–918 (2015)
14. A. Albert, R. Ram, Smart meter driven segmentation: what your consumption says about you. IEEE Trans Power Syst. **28**(4), 4019–4030 (2013)
15. N. Mahmoudi-Kohan, M. P. Moghaddam, M. K. Sheikh-El-Eslami, E. Shayesteh, A three-stage strategy for optimal price offering by a retailer based on clustering techniques. Int. J. Electr. Power Energy Syst. **32**(10), 1135–1142 (2010)
16. C. Dwork, in *International Conference on Theory and Applications of Models of Computation, Differential privacy: A survey of results*, Differential privacy (Springer, Xi'an, 2008), pp. 1–19
17. L. Sankar, S. Kar, R. Tandon, H. V. Poor, in *Proc. IEEE International Conference on Smart Grid Communications (SmartGridComm)*, Competitive privacy in the smart grid: an information-theoretic approach (IEEE, Brussels, 2011), pp. 220–225
18. C. Y. Ma, D. K. Yau, in *Proceedings of the 10th ACM Symposium on Information, Computer and Communications Security*, On information-theoretic measures for quantifying privacy protection of time-series data (ACM, Singapore, 2015), pp. 427–438
19. S. Li, A. Khisti, A. Mahajan, Information-theoretic privacy for smart metering systems with a rechargeable battery. IEEE Trans. Inf. Theory. **64**(5), 3679–3695 (2018)
20. A. Reinhardt, F. Englert, D. Christin, Averting the privacy risks of smart metering by local data preprocessing. Pervasive Mob. Comput. **16**, 171–183 (2015)
21. E. Elhamifar, R. Vidal, Sparse subspace clustering: algorithm, theory, and applications. IEEE Trans. Pattern Anal. Mach. Intell. **35**(11), 2765–2781 (2013)

22. B. Eriksson, L. Balzano, R. Nowak, in *Proc. Int. Conf. Artif. Intell. Stat*, High-rank matrix completion (JMLR, La Palma, 2012), pp. 373–381
23. G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, Y. Ma, Robust recovery of subspace structures by low-rank representation. IEEE Trans. Pattern Anal. Mach. Intell. **35**(1), 171–184 (2013)
24. V. M. Patel, H. Van Nguyen, R. Vidal, Latent space sparse and low-rank subspace clustering. IEEE J. Sel. Topics Signal Process. **9**(4), 691–701 (2015)
25. M. Soltanolkotabi, E. J. Candès, A geometric analysis of subspace clustering with outliers. Ann. Stat. **40**(4), 2195–2238 (2012)
26. M. Soltanolkotabi, E. Elhamifar, E. J. Candés, Robust subspace clustering. Ann. Stat. **42**(2), 669–699 (2014)
27. R. Wang, M. Wang, J. Xiong, Data recovery and subspace clustering from quantized and corrupted measurements. IEEE J. Sel. Topics Signal Process., Spec Issue Robust Subspace Learn. Tracking Theory Algoritm Appl. **12**(6), 1547–1560 (2018)
28. S. A. Bhaskar, Probabilistic low-rank matrix completion from quantized measurements. J. Mach. Learn. Res. **17**(60), 1–34 (2016)
29. Y. Cao, Y. Xie, in *Proc. IEEE Int. Workshop Comput. Adv. Multi-Sensor Adapt. Process*, Categorical matrix completion (IEEE, Cancun, 2015)
30. T. Cai, W.-X. Zhou, A max-norm constrained minimization approach to 1-bit matrix completion. J. Mach. Learn. Res. **14**(1), 3619–3647 (2013)
31. M. A. Davenport, Y. Plan, E. van den Berg, M. Wootters, 1-bit matrix completion. Inf. Infer. **3**(3), 189–223 (2014)
32. P. Gao, M. Wang, J. H. Chow, M. Berger, L. M. Seversky, Missing data recovery for high-dimensional signals with nonlinear low-dimensional structures. IEEE Trans. Signal Process. **65**(20), 5421–5436 (2017)
33. O. Klopp, J. Lafond, É. Moulines, J. Salmon, Adaptive multinomial matrix completion. Electron. J. Stat. **9**(2), 2950–2975 (2015)
34. J. Lafond, O. Klopp, E. Moulines, J. Salmon, in *Adv. Neural Inf. Process. Syst*, Probabilistic low-rank matrix completion on finite alphabets (Curran Associates, Montreal, 2014), pp. 1727–1735
35. A. S. Lan, C. Studer, R. G. Baraniuk, in *Proc. IEEE Int. Conf. Acoust Speech Signal Process*, Matrix recovery from quantized and corrupted measurements (IEEE, Florence, 2014), pp. 4973–4977
36. A. S. Lan, A. E. Waters, C. Studer, R. G. Baraniuk, Sparse factor analysis for learning and content analytics. J. Mach. Learn. Res. **15**(1), 1959–2008 (2014)
37. P. Gao, R. Wang, M. Wang, J. H. Chow, Low-rank matrix recovery from noisy, quantized and erroneous measurements. IEEE Trans. Signal Process. **66**(11), 2918–2932 (2018)
38. S. A. Bhaskar, in *Proc. Asilomar Conf. Signals Syst. Comput*, Probabilistic low-rank matrix recovery from quantized measurements: application to image denoising, (2015), pp. 541–545
39. S. A. Bhaskar, Localization from connectivity: a 1-bit maximum likelihood approach. IEEE/ACM Trans. Netw. **24**(5), 2939–2953 (2016)
40. Y. Yang, J. Feng, N. Jojic, J. Yang, T. S. Huang, in *European Conference on Computer Vision*, l0-sparse subspace clustering (Springer, Amsterdam, 2016), pp. 731–747
41. A. Y. Ng, M. I. Jordan, Y. Weiss, in *Adv. Neural Inf. Process. Syst*, On spectral clustering: analysis and an algorithm (Morgan Kaufmann Publishers, Vancouver, 2002), pp. 849–856
42. J. Lin, E. Keogh, L. Wei, S. Lonardi, Experiencing sax: a novel symbolic representation of time series. Data Min. Knowl. Discov. **15**(2), 107–144 (2007)
43. E. Keogh, K. Chakrabarti, M. Pazzani, S. Mehrotra, in *Proceedings of the 2001 ACM SIGMOD International Conference on Management of Data*, Locally adaptive dimensionality reduction for indexing large time series databases (ACM, Santa Barbara, 2001), pp. 151–162
44. R. Basri, D. W. Jacobs, Lambertian reflectance and linear subspaces. IEEE Trans. Pattern Anal. Mach. Intell. **25**(2), 218–233 (2003)
45. P. Gao, M. Wang, S. G. Ghiocel, J. H. Chow, B. Fardanesh, G. Stefopoulos, Missing data recovery by exploiting low-dimensionality in power system synchrophasor measurements. IEEE Trans. Power Syst. **31**(2), 1006–1013 (2016)
46. M. B. Hossain, I. Natgunanathan, Y. Xiang, L.-X. Yang, G. Huang, Enhanced smart meter privacy protection using rechargeable batteries. IEEE Internet Things J. **6**(4), 7079–7092 (2019)
47. A. Reinhardt, D. Egarter, G. Konstantinou, D. Christin, in *2015 IEEE International Conference on Smart Grid Communications (SmartGridComm)*, Worried about privacy? Let your PV converter cover your electricity consumption fingerprints (IEEE, Miami, 2015), pp. 25–30
48. L. Sweeney, k-anonymity: a model for protecting privacy. Int. J. Uncertain. Fuzziness Knowl-Based Syst. **10**(05), 557–570 (2002)
49. R. L. Lagendijk, Z. Erkin, M. Barni, Encrypted signal processing for privacy protection: conveying the utility of homomorphic encryption and multiparty computation. IEEE Signal Proc. Mag. **30**(1), 82–105 (2012)
50. T. Baumeister, in *2011 IEEE International Conference on Smart Grid Communications (SmartGridComm)*, Adapting PKI for the smart grid (IEEE, Brussels, 2011), pp. 249–254
51. G. Giaconi, D. Gündüz, H. V. Poor, in *2015 IEEE International Conference on Communications (ICC)*, Smart meter privacy with an energy harvesting device and instantaneous power constraints (IEEE, Miami, 2015), pp. 7216–7221
52. G. Kalogridis, C. Efthymiou, S. Z. Denic, T. A. Lewis, R. Cepeda, in *2010 First IEEE International Conference on Smart Grid Communications*, Privacy for smart meters: towards undetectable appliance load signatures (IEEE, Gaithersburg, 2010), pp. 232–237
53. J. Gomez-Vilardebo, D. Gündüz, Smart meter privacy for multiple users in the presence of an alternative energy source. IEEE Trans. Inf. Forensic Secur. **10**(1), 132–141 (2014)
54. Y. Hong, W. M. Liu, L. Wang, Privacy preserving smart meter streaming against information leakage of appliance status. IEEE Trans. Inf. Forensic Secur. **12**(9), 2227–2241 (2017)
55. J. A. Snyman, N. Stander, W. J. Roux, A dynamic penalty function method for the solution of structural optimization problems. Appl. Math. Model. **18**(8), 453–460 (1994)

56. Commission for Energy Regulation Smart Metering Project. http://www.ucd.ie/issda/data/commissionforenergyregulationcer. Accessed 5 July 2018
57. S. Barker, A. Mishra, D. Irwin, E. Cecchet, P. Shenoy, J. Albrecht, *et al*, Smart*: an open data set and tools for enabling research in sustainable homes. SustKDD, August. **111**(112), 108 (2012)
58. S. P. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, Distributed optimization and statistical learning via the alternating direction method of multipliers. Found. Trends® Mach. Learn. **3**(1), 1–122 (2011)
59. T. M. Cover, J. A. Thomas, *Elements of Information Theory*. (Wiley, Hoboken, 2012)
60. J. Bolte, S. Sabach, M. Teboulle, Proximal alternating linearized minimization for nonconvex and nonsmooth problems. Math. Program. **146**(1-2), 459–494 (2014)
61. Y. Xu, W. Yin, A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion. SIAM J. Imag. Sci. **6**(3), 1758–1789 (2013)

Publisher's Note