IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received February 21, 2018, accepted April 19, 2018, date of publication May 1, 2018, date of current version May 24, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2832043

Nonconvex l_p-Norm Regularized Sparse Self-Representation for Traffic Sensor Data Recovery

XIAOBO CHEN[1,2,3], YINGFENG CAI[1,2], QINGCHAO LIU[1,2], AND LEI CHEN[3]

[1]Automotive Engineering Research Institute, Jiangsu University, Zhenjiang 212013, China
[2]School of Automotive and Traffic Engineering, Jiangsu University, Zhenjiang 212013, China
[3]Jiangsu Key Laboratory of Big Data Security and Intelligent Processing, Nanjing University of Posts and Telecommunications, Nanjing 210003, China

Corresponding authors: Xiaobo Chen (xbchen82@gmail.com) and Yingfeng Cai (caicaixiao0304@126.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61773184, Grant U1564201, Grant U1664258, Grant 61601203, and Grant 61203244, in part by the Key Research and Development Program of Jiangsu Province under Grant BE2016149, in part by the Key Project for the Development of Strategic Emerging Industries of Jiangsu Province under Grant 2016-1094 and Grant 2015-1084, in part by the Natural Science Foundation of Jiangsu Province under Grant BK20161516, in part by the Six Talent Peaks Project of Jiangsu Province under Grant 2017-JXQC-007, in part by the Jiangsu Key Laboratory of Big Data Security and Intelligent Processing, NJUPT, under Grant BDSIP1802, in part by the China Postdoctoral Science Foundation under Grant 2017M611729, in part by the Natural Science Foundation of the Jiangsu Higher Education Institutions of China under Grant 17KJB580003, and in part by the Talent Foundation of Jiangsu University, China, under Grant 14JDG066.

ABSTRACT Recovering missing values from incomplete traffic sensor data is an important task for intelligent transportation system because most algorithms require data with complete entries as input. Self-representation-based matrix completion attempts to optimally represent each sample by linearly combining other samples when conducting missing values recovery. Typically, it implements sparse or dense combination through imposing either l_1-norm or l_2-norm regularization over the representation coefficients, which is not always optimal in practice. To permit more flexibility, we propose in this paper a novel approach termed as l_p-norm regularized sparse self-representation (SSR-l_p) by incorporating nonconvex l_p-norm with $0 < p < 1$ as regularization. In such a way, it is able to produce more sparsity than l_1-norm and in turn facilitates the accurate recovery of missing data. We further develop an efficient iterative algorithm for solving SSR-l_p. The performance of this method is evaluated on a real-world road network traffic flow data set. The experimental results verify the advantage of our method over other competing algorithms in recovering missing values.

INDEX TERMS Traffic sensor data, missing values, l_p-norm regularization, sparse self-representation.

I. INTRODUCTION

Traffic flow data collected by geographically distributed sensors has come to play an important role in advanced intelligent transportation system (ITS) because most traffic services provided by ITS depend on the accuracy and completeness of data. For example, short-term traffic flow forecasting [1], [2], which is of paramount importance for realizing proactive traffic control and effective route planning, requires data fed into specific predictive models, e.g. support vector machine [2], [3], neural networks [4], etc., to be complete without missing entries. Despite the fast growing reliability of sensing equipment and transmission network [5], [6], missing sensor data is still prevalent and inevitable in current ITS. For example, it was reported that for a dense road network in the city of Melbourne, about 8% of sensor can reach up to 56% missing data. Similarly, about 10% of daily traffic flow in Beijing is missing. There are many reasons that lead to missing data, such as sensor malfunction, transmission distortion, and other unexpected exogenous factors. Without proper preprocessing, datum with missing entries cannot be directly utilized by most machine learning algorithms.

To address the above mentioned missing sensor data problem, many imputation methods have been proposed in the literature during the past decades. Here, imputation means the procedure that generates plausible estimations for the missing values (MVs) in a given incomplete data [7]. By means of MV imputation, the incomplete data can be converted into complete one and then used in traditional machine learning algorithms. Due to the connectivity of road network and the regularity of human travel activity, traffic sensor data collected at different time intervals and different road segments is essentially correlated with each other. Consequently, such a kind of intrinsic correlation between traffic sensor data makes the recovery of missing values feasible and reliable in practice. Nowadays, some typical imputation methods have been developed in the literature, including K-nearest neighbors (KNN) [8], singularity value decomposition (SVD) [9], local least squares regression (LLS) [10], [11], probabilistic principle component analysis (PPCA) [12], [13], low-rank matrix completion (LRMC) [14]–[17], etc.

Recently, self-representation based matrix completion [18] was developed and shows competitive performance in comparison with other imputation methods. Essentially, self-representation, as a general concept, refers to that each sample can be well represented as a linear combination of other samples with representation coefficients (weights) characterizing the contribution of other samples. Self-representation has already been widely exploited in some pattern recognition tasks, such as subspace clustering [19], [20], feature selection [21], etc., because of its simplicity and effectiveness. However, far less work exists on missing data recovery via self-representation. Different from self-representation with complete data, the representation of each sample cannot be obtained directly in the context of missing data. As a result, the recovery of missing data and the search for representation have to be implemented simultaneously. A key factor in self-representation based matrix completion is the selection of suitable regularization on representation coefficients. After comparing l_2-norm, l_1-norm, and nuclear-norm [17], the previous work [18] verified from experiments that l_1-norm regularization performs much better than other forms of regularization in terms of recovery accuracy. The advantages of l_1-norm regularization are two-fold. First, the representation vector of a target sample with respect to all other samples is sparse, implying that only a few of samples have nonzero coefficients. This is because l_1-norm is the tightest convex relaxation of l_0-norm [22]. Second, l_1-norm minimization generally leads to convex optimization problem [23] with efficient implementation. Nevertheless, a potential issue concerning l_1-norm is that it may fail to find desired solution [22]. From this view of point, l_1-norm could be too restrictive and not sufficiently flexible, which in turn influences the imputation accuracy of missing data.

Inspired by the above discussion, in this paper, we propose a novel self-representation based matrix completion approach for missing data recovery by incorporating l_p-norm regularization with $0<p<1$ [24]. It has been observed that as a nonconvex surrogate of l_0-norm, l_p-norm minimization can often achieve more sparsity than l_1-norm minimization [25], since it is closer to l_0-norm when p is smaller than 1. Theoretically, l_p-norm requires weaker conditions than l_1-norm to guarantee successful recovery of sparse signal [26]. We also take into account the nonnegative property of traffic sensor data through optimization, which is an inherent requirement for many real-world physical systems. The proposed method is termed as l_p-norm regularized sparse self-representation (abbreviated as SSR-l_p for brevity). Despite more flexibility with l_p-norm, it has difficulty in solving the resultant model because l_p-norm is typically nonconvex. To address this issue, we further develop an efficient alternating algorithm which combines iterative reweighted least squares (IRLS) [22] as well as classic gradient projection (GP) method.

We summarize the main contributions of this paper as follows: (1) SSR-l_p is a general framework benefiting from both self-representation and l_p-norm regularization, providing flexible framework for MV imputation. (2) An efficient alternating optimization algorithm is proposed to solve SSR-l_p model. (3) Extensive experiments on real-world traffic sensor data verify the effectiveness of our method in comparison with other related algorithms.

The rest of this paper is organized as follows. In Section II, we briefly review and analyze some popular methods for missing data recovery. In Section III, we present the proposed SSR-l_p model and its optimization algorithm. Section IV reports the experimental results on real-world traffic sensor data. Finally, Section V gives the conclusions and discusses future work.

II. RELATED WORKS

So far, many MV imputation or recovery approaches have been proposed in the literature. These methods can be roughly classified into the following categories.

A. REGRESSION BASED METHODS

The methods in this category attempt to characterize the relationship between missing values and observed values by regression models built based on training data. Regression models can be divided into parametric and nonparametric regression, thus further refining the division of imputation methods in this category. Some typical regression models include K-nearest neighbor (KNN) regression, least squares regression [10], support vector regression [27], neural networks [28], [29], [43], etc. For example, KNN imputation first select K nearest samples for the sample with MVs, followed which the MVs can be estimated as the weighted average of those selected samples. Following similar idea, local least squares (LLS) imputation [10], [30] also selects K nearest samples for the sample with MVs, but different from KNN imputation, it describes the relation between MVs and observed values by virtue of least squares regression, allowing more flexibility than simple weighted average. LLS has been proved to yield promising performance in traffic sensor data recovery [31] and other domains [11].

B. PROBABILISTIC MODEL BASED METHODS

In this type of methods, the complete data is supposed to follow a probabilistic distribution with specific form but unknown model parameters. Based on the observed values, both the model parameters and the missing data can be simultaneously estimated following maximum likelihood estimation (MLE) or full Bayesian framework [32], [33]. A popular algorithm to achieve such joint estimation is based on expectation-maximization (EM) [34]. A typical method belonging to this category is the so-called probabilistic principal component analysis (PPCA) [13] assuming that the data follows multivariate Gaussian model. The missing data and model parameters are alternatively estimated by EM algorithm. This method has shown promising results in the imputation of traffic sensor data. This type of methods [13], [34], [35] imposes a global distribution assumption about data, thus being effective when data is consistent with the assumed distribution.

C. MATRIX COMPLETION BASED METHODS

This type of methods organizes all samples into a matrix and achieve MVs recovery based on certain property of the matrix. One of the most well-known approach belonging to this class is low-rank matrix completion (LRMC) [14] which assumes the matrix is of low-rank structure. For traffic matrix, this assumption is reasonable to a certain extent, because traffic flows within the same road network are spatially and temporally correlated with each other. Owing to such inherent correlation, it was reported that LRMC is able to produce accurate recovery of missing data [36], [37], [44]. Recently, LRMC has attracted considerable attention and researchers have developed many optimization algorithms dedicated to solving LRMC model, such as SVT [14], FPCA [38], ADMM [39], etc. However, LRMC takes a global view on the data matrix, without sufficiently accounting for the difference between samples [17]. It may produce suboptimal recovery for samples with complex inherent structure, e.g., multiple subspaces. To this end, self-representation based matrix completion [18] was presented recently, aiming to characterize the relations between samples through linear combination. The missing data recovery and effective representation are solved jointly, indicating the two tasks can facilitate each other.

III. THE PROPOSED METHOD

In this section, we first present the formulation of l_p-norm regularized sparse self-representation. Then, an efficient optimization algorithm for solving this model is proposed.

Formally, let $X = [x_1, x_2, \ldots, x_N] \in \mathbb{R}^{d\times N}$ be the given data matrix, where $x_i \in \mathbb{R}^d$ denotes the i-th sample with d features, N is the total number of samples. Notice that in our problems, not all of the elements in X are observable. Due to diverse causes, many elements in X are missing. Here, we use Ω to denote the indexes of missing values in X. Thereby, the central task of MV imputation is to estimate X_Ω as accurately as possible.

A. l_p-NORM REGULARIZED SPARSE SELF-REPRESENTATION

In the spirit of self-representation [18], [19], each data sample can be represented as a linear combination of other samples. In particular, we have

$$x_i \approx \sum_{j=1, j\neq i}^{N} w_i(j)x_j \tag{1}$$

where $w_i(j)$ denotes the combinatorial coefficient or weight of x_j in the resulting linear combination. Further, we introduce the following l_p-norm ($0<p<1$) regularization [22], [24] on the weight vector $w_i = [w_i(1), w_i(2), \ldots, w_i(N)]^T$

$$\|w_i\|_p^p = \sum_{j=1}^{N} |w_i(j)|^p \tag{2}$$

with hope that most elements in w_i should be zero or close to zero such that the corresponding samples can be eliminated from the representation of x_i.

Besides the above task, another difficulty comes from the fact that many elements in X, i.e. X_Ω, are unknown while only the rest part of X are observed. As a result, it is infeasible to find weight vector w_i directly given a set of incomplete data samples. In other words, we need to discover the above sparse linear representation structure among data and meanwhile estimate the involved missing values. In fact, it is expected that the reliable estimation of MVs and the discovery of sparse linear relations among data are related and thus would benefit from each other, indicating the two tasks can be solved in a uniformed framework. In addition, for many applications, the data that real physical system records is usually nonnegative [6]. As a result, nonnegativity should also be taken into account. Based on the above discussion, we present l_p-norm regularized sparse self-representation (SSR-l_p) for MV imputation as follows

$$\min_{X_\Omega, W} \frac{1}{2}\sum_{i=1}^{N} \|x_i - \sum_{j=1,j\neq i}^{N} w_i(j)x_j\|^2 + \lambda \sum_{i=1}^{N} \|w_i\|_p^p \quad \text{s.t.} \quad X_\Omega \geq 0 \tag{3}$$

where $\lambda > 0$ is a prameter controlling the strength of l_p-norm regularization.

Let $W = [w_1, w_2, \ldots, w_N]$ and $\mathrm{diag}(W)$ stand for the diagonal elements of W. The above problem (3) can be expressed in matrix form as follows

$$\min_{X_\Omega, W} \frac{1}{2}\|X - XW\|^2 + \lambda \sum_{i=1}^{N} \|w_i\|_p^p \quad \text{s.t.} \quad X_\Omega \geq 0, \mathrm{diag}(W) = 0 \tag{4}$$

Notice that in model (4), both X_Ω and W are the decision variables, differentiating it from traditional subspace clustering [19], [20] where only W is the variable need to solve. It should be pointed out that problem (4) naturally reduces to the models developed in [18] given $p = 1$ or $p = 2$.

B. OPTIMIZATION ALGORITHM

The problem (4) does not allow a closed-form solution because of the coupling between decision variables X_Ω and W, which leads the problem difficult to solve. However, we observe that the problem can be simplified if only one variable is optimized each time while fixing the other one. To this end, we develop an iterative algorithm to solve (4) by alternatively optimizing over X_Ω and W while holding the other variable fix.

Algorithm 1 Solve X_Ω When W Is Fixed

Input current estimation of W
1: Initialize X_Ω, let $c = 10^{-4}$
2: **while** not converged **do**
3: Compute gradient $\nabla g(X_\Omega)$ of $g(X_\Omega)$
4: Find step-size l with Armijo rule, i.e., choose $l = \min\{1, \frac{1}{2}, \frac{1}{2^2}, \ldots\}$ such that $g(X_\Omega^*) \leq g(X_\Omega) + c \cdot trace((X_\Omega^* - X_\Omega)^T \nabla g(X_\Omega))$ where $X_\Omega^* = \max\{X_\Omega - l\nabla g(X_\Omega), 0\}$
5: Update MV estimation as $X_\Omega \leftarrow X_\Omega^*$
6: **end while**
Output the estimated missing data X_Ω

Concretely, we first fix W and seek for the optimal solution of X_Ω. In such a case, the problem (4) is equivalent to the following constrained problem

$$\min_{X_\Omega} \frac{1}{2}\|X - XW\|^2 \quad \text{s.t.} \quad X_\Omega \geq 0 \tag{5}$$

In order to solve (5), we develop an iterative algorithm based on gradient projection [23]. Let the derivative of $g(X_\Omega) = \frac{1}{2}\|X - XW\|^2$ with respect to X_Ω be denoted by $\nabla g(X_\Omega) = (X(I - W)(I - W)^T)_\Omega$, then the algorithm for solving (5) can be summarized in **Algorithm 1**.

Secondly, we attempt to optimize W while holding X_Ω. In such a case, the original problem (4) becomes

$$\min_{W} \frac{1}{2}\|X - XW\|^2 + \lambda \sum_{i=1}^{N} \|w_i\|_p^p \quad \text{s.t.} \quad \text{diag}(W) = 0 \tag{6}$$

Notice that problem (6) is separable with respect to the columns of W, therefore, we can solve each w_i by

$$\min_{w_i} \frac{1}{2}\|x_i - Xw_i\|^2 + \lambda\|w_i\|_p^p \quad \text{s.t.} \quad w_i(i) = 0 \tag{7}$$

In order to handle the constraint in (7), we denote $\overline{X}_i = [x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_N] \in \mathbb{R}^{d\times(N-1)}$, $\overline{w}_i = [w_i(1), \ldots, w_i(i-1), w_i(i+1), \ldots, w_i(N)]^T$, then, the constrained problem (7) can be converted into the unconstrained problem as follows

$$\min_{\overline{w}_i} \frac{1}{2}\|x_i - \overline{X}_i\overline{w}_i\|^2 + \lambda\|\overline{w}_i\|_p^p \tag{8}$$

As we can see, a major difficulty is that problem (8) is nonconvex when $0<p<1$. In this work, in order to take the advantage of the problem structure, we present an algorithm targeting at (8) by means of iteratively reweighted least squares (IRLS) [22] which is a popular technique in optimization. Specifically, by computing the derivative of the objective (8) with respect to $\overline{w}_i$ and setting to zero, we have

$$\overline{X}_i^T(\overline{X}_i\overline{w}_i - x_i) + \frac{\lambda p \overline{w}_i(j)}{(\overline{w}_i(j)^2 + \epsilon)^{1-\frac{p}{2}}} = 0 \tag{9}$$

where ϵ is a small number to avoid division by zeros.

Algorithm 2 Iterative Algorithm for Solving SSR-l_p

Initialization Give an initial W, parameter λ, p
1: **while** not converged **do**
2: Use **Algorithm 1** to update current MVs X_Ω
3: **for** $i = 1, 2, \ldots, N$ **do**
4: Compute D_t by using (11)
5: Update w_i by using (10) or (13)
6: **end for**
7: **end while**
8: **Output** the estimated missing data X_Ω

Then, the iterative procedure for solving $\overline{w}_i$ is given by

$$\overline{w}_i = (\overline{X}_i^T\overline{X}_i + \lambda D_t)^{-1}\overline{X}_i^T x_i \tag{10}$$

where D_t is a diagonal matrix defined as

$$D_t = \text{diag}\left(\frac{p}{(\overline{w}_i^t(j)^2 + \epsilon)^{1-\frac{p}{2}}}\right), \quad j = 1, \ldots, i-1, i+1, \ldots, N \tag{11}$$

and $\overline{w}_i^t$ is the solution at the t-th iteration.

Note that when the total number of samples is large, e.g. $N \gg d$, the inversion of matrix $\overline{X}_i^T\overline{X}_i + \lambda D_t$ in (10) is computationally expensive, i.e., $O(N^3)$ in time complexity. To address this problem and make our algorithm feasible given large number of samples, we apply the well-known Sherman-Morrison-Woodbury formula [40] shown below

$$(UCV + A)^{-1} = A^{-1} - A^{-1}U(C^{-1} + VA^{-1}U)^{-1}VA^{-1} \tag{12}$$

By combing (10) and (12), the optimal solution of $\overline{w}_i$ can be rewritten as

$$\overline{w}_i = \frac{1}{\lambda}D_t^{-1} - \frac{1}{\lambda}D_t^{-1}\overline{X}_i^T(\lambda I + \overline{X}_i D_t^{-1}\overline{X}_i^T)^{-1}\overline{X}_i D_t^{-1} \tag{13}$$

where I is a $d \times d$ identity matrix. In formula (13), we only need to solve the inversion of a matrix with size $d \times d$. To this end, when $N \gg d$, the time complexity for solving w_i can be significantly reduced from $O(N^3)$ to $O(d^3)$. The time complexity for solving all of w_i can be estimated as $O(Nd^3)$, which is linear with respect to the number of samples.

Finally, the whole iterative algorithm for solving SSR-l_p model (4) is summarized in **Algorithm 2**.

IV. EXPERIMENTS AND ANALYSIS

A. DATA DESCRIPTION

In this study, we evaluate the proposed SSR-l_p algorithm on a real-world traffic flow dataset. The data was collected from Interstate 205 (I205) highways, serving the Portland-Vancouver metropolitan area in Oregon and Washington states, USA. The selected sub-area road network is shown in Fig. 1. Thirty inductive loop detectors which records the vehicle volume counts are chosen. The aggregation period is 15 minutes, thus yielding 96 sampling points in each day. In other words, each data sample can be viewed a point in a 96 dimensional space. The collection time period used in this study was from Mar. 1st to Aug. 31st in

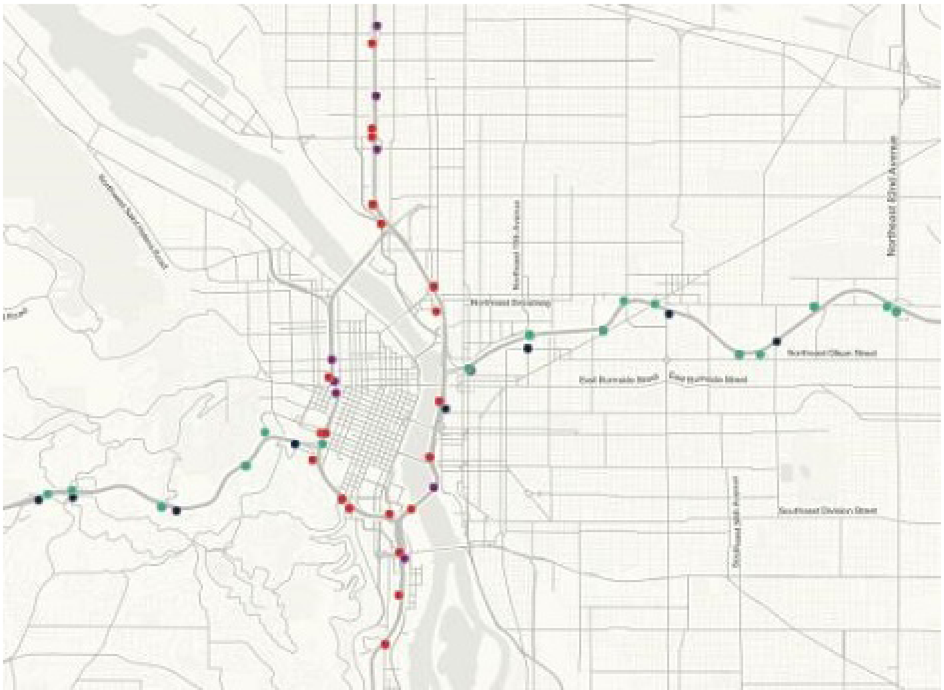


FIGURE 1. The selected sub-area road network of Portland, OR, USA.

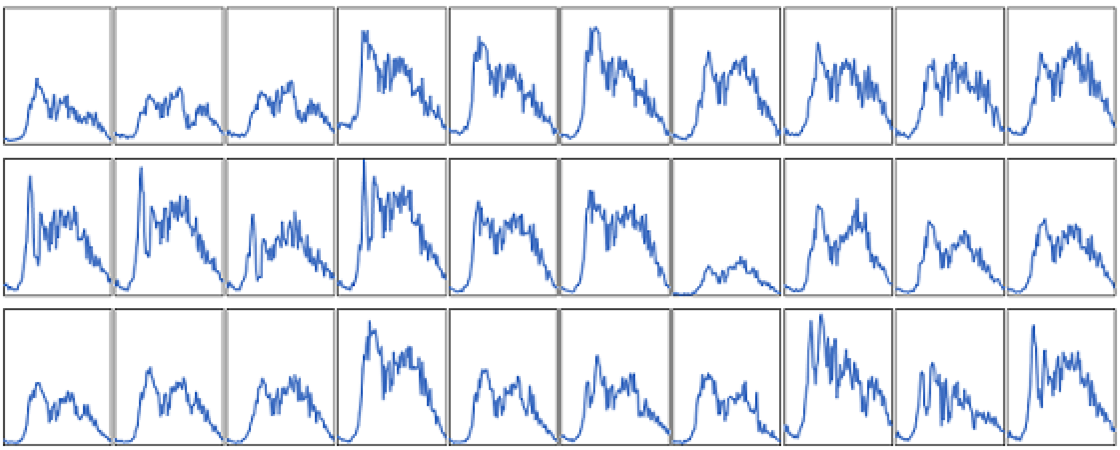


FIGURE 2. Illustration of traffic flow profiles from 30 detectors in the same day.

the year 2015. The data is publicly available at website (http://portal.its.pdx.edu/). After excluding weekends as well as holidays, we finally get volume data of 97 days. Finally, the total number of volume counts reaches $96 \times 30 \times 97 = 279360$. The whole traffic sensor data is organized as a 96×2910 matrix with each column representing a data sample. In Fig. 2, we illustrate 30 data samples, each of which was captured by a distinct detector in the same day. Note that the horizontal axis and the vertical axis in Fig. 2 represent time and traffic volume, respectively. These samples intuitively reflect the traffic flow profiles at different road segments. As can be seen, despite of overall similarity among traffic flow profiles, it does exhibit some distinctions with respect to the variation patterns of different detectors. For instance, the maximum flow of some detectors is significantly larger than that of other detectors. In addition, the traffic flow at certain detectors clearly shows two peaks at rush hours while it is not very notable for other detectors. These slight yet important differences pose great challenge for MV imputation problem and render us to develop more flexible model such that the homogeneity as well as heterogeneity can be stimutanlously taken into accout.

B. CONFIGURATION

To comprehensively compare different methods, beside the proposed SSR-l_p, we also include some closely-related

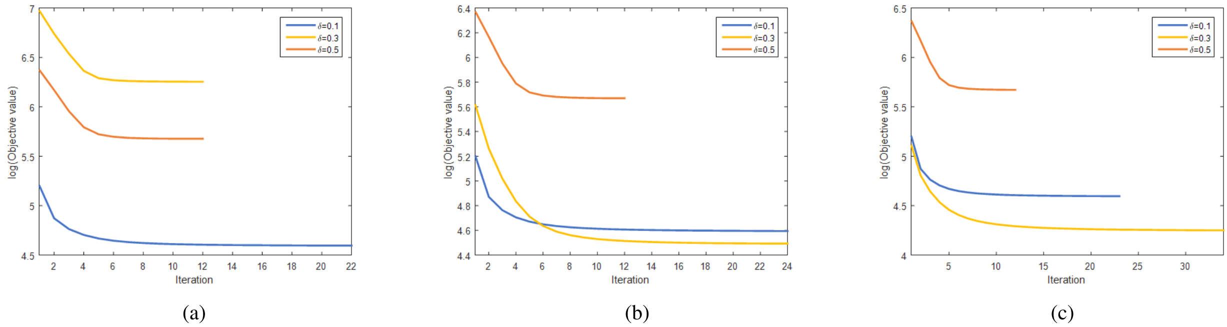


FIGURE 3. **Convergence curve under different missing patterns. (a) MAR. (b) MIXED.**

TABLE 1. **Imputation error ($\times 10^{-2}$) under MCAR missing pattern.**

Method		KNN	LLS	PPCA	LRMC	SR-l_2	SRSp	SSR-l_p
0.1	Mean	97.40	73.40	81.28	86.15	81.35	**55.72**	57.03(p=0.8)
	Std	0.85	0.87	0.73	0.68	0.75	0.49	0.60
0.2	Mean	109.34	79.92	85.50	88.52	83.87	61.21	**60.91(p=0.8)**
	Std	0.56	0.39	0.38	0.37	0.41	1.12	0.55
0.3	Mean	121.96	87.89	88.11	91.05	86.70	71.99	**65.21(p=0.2)**
	Std	0.58	0.45	0.23	0.22	0.21	0.29	0.42
0.4	Mean	134.50	97.46	91.75	94.12	89.93	78.99	**71.79(p=0.2)**
	Std	0.12	0.39	0.22	0.27	0.21	0.43	0.18
0.5	Mean	147.16	110.86	96.04	97.82	93.69	87.56	**81.44(p=0.2)**
	Std	0.22	0.77	0.25	0.21	0.25	0.52	0.41

algorithms, including KNN, LLS [10], PPCA [13], LRMC [41], SR-l_2 [18] and SRSp [18]. Note that SRSp can be viewed as a special case of SSR-l_p when p equals to 1. These methods covers the mainstream techniques for traffic data imputation, such as regression model, probabilistic model, etc. All of these algorithms were implemented in MATLAB environment on a PC with Intel(R) Core(TM) i7-4712MQ CPU and 12GB DDR4 RAM. There are some parameters involved in each method, such as the number of nearest neighbors for KNN and LLS, the subspace dimensionality for PPCA, etc. Following previous stuides [18], we adjust the parameters in each method such that best performance is achieved.

In order to simulate MVs and evaluate the imputation performance, we artificially generate missing entries for the data. Specifically, three missing patterns [13], [17] are considered in the experiment. (i) missing completely at random (MCAR) where the data points to be missing are completely independent of each other and occur as a set of isolated points randomly distributed, (ii) missing at random (MAR) where the appearance of missing points depends on their neighboring points. Therefore, this type of missing pattern looks like a group of successive MVs, (iii) a mixture of MCAR and MAR (MIXED), where half of MVs obey MCAR and the other half are from MAR.

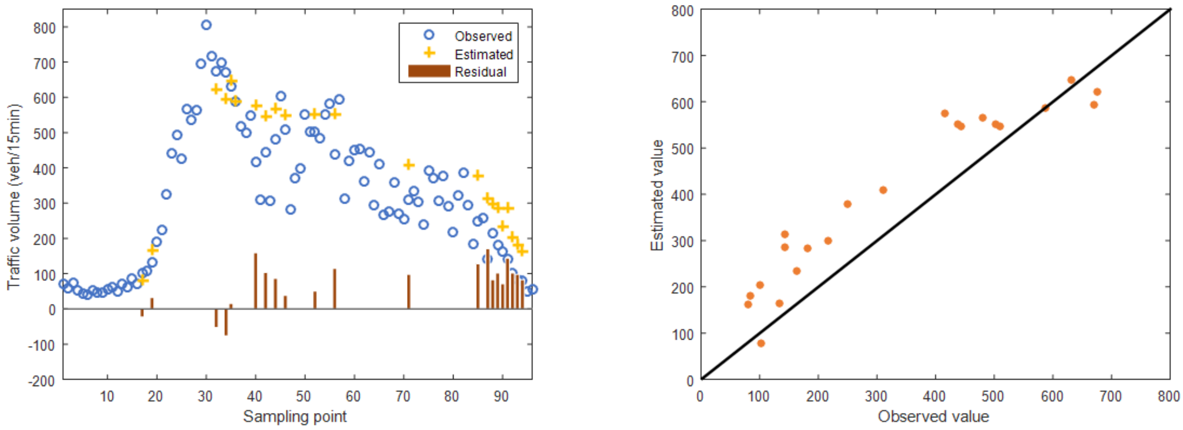
We measure the recovery performance of each method by root mean squared error (RMSE) between the estimated values and the real values for those missing entries. Clearly, smaller RMSE indicates better recovery performance. We also define the missing ratio δ as the ratio of the number of missing entries to the total number of entries. Moreover, δ is changed from 0.1 to 0.5 with step 0.1 in order to investigate the variation of recovery performance against different missing ratios.

C. CONVERGENCE ANALYSIS

In this work, an iterative algorithm alternatively recovering the missing data and optimizing sparse representation coefficients is developed to solve the proposed SSR-l_p model. Next, we empirically investigate the convergence behavior of this algorithm under varying missing ratios and different missing patterns. Some convergence curves obtained in the experiments are shown in Fig.3 where the x-axis denotes the number of iterations and the y-axis denotes the logarithm of objective function. From these results, we can observe that our algorithm reduces the objective (13) in each iteration, regardless of specific missing ratio and missing pattern. Moreover, the iterative algorithm we develop is able to converge quickly, usually requiring about 10-40 iterations in most cases.

D. IMPUTATION PERFORMANCE COMPARISON

Considering the randomness when artificially simulating missing entries, we repeat each experiment five times and calculate the average imputation error (Mean) as well as the associated standard deviation (Std). The experimental results under MCAR, MAR, and MIXED missing patterns are reported in Tables 1, 2 and 3, respectively. Note that the number in parenthesis of SSR-l_p column indicates the value



(a) (b)

FIGURE 4. Imputation results obtained by KNN.

TABLE 2. Imputation error ($\times 10^{-2}$) under MAR missing pattern.

Method		KNN	LLS	PPCA	LRMC	SR-l_2	SRSp	SSR-l_p
0.1	Mean	127.69	82.84	92.87	105.28	98.27	**65.62**	66.20(p=0.8)
	Std	1.30	0.58	1.03	1.19	1.09	0.91	0.49
0.2	Mean	135.73	90.40	96.21	105.37	97.97	74.08	**71.03(p=0.8)**
	Std	1.98	1.05	0.98	0.93	1.00	1.60	0.99
0.3	Mean	143.45	98.82	98.75	106.70	99.11	85.67	**78.36(p=0.6)**
	Std	0.79	0.79	0.41	0.55	0.40	1.54	0.78
0.4	Mean	150.82	110.23	101.66	108.44	101.14	93.38	**85.54(p=0.6)**
	Std	0.66	1.12	0.37	0.40	0.45	0.82	0.72
0.5	Mean	158.09	126.44	105.00	110.79	104.14	100.37	**94.22(p=0.2)**
	Std	0.38	0.88	0.57	0.41	0.54	0.54	0.83

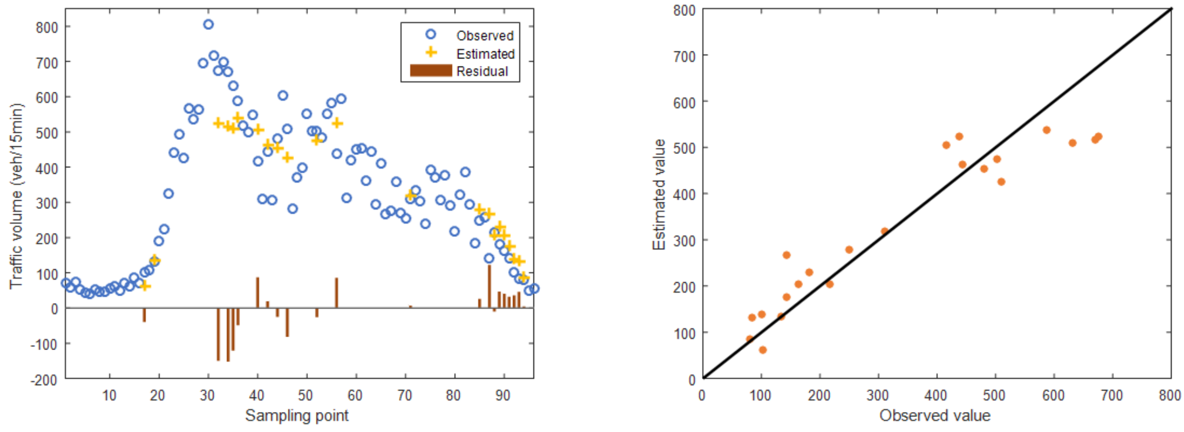
TABLE 3. Imputation error ($\times 10^{-2}$) under MIXED missing pattern.

Method		KNN	LLS	PPCA	LRMC	SR-l_2	SRSp	SSR-l_p
0.1	Mean	113.51	77.72	87.12	96.91	91.03	**60.25**	61.31(p=0.8)
	Std	0.51	0.67	0.93	1.44	1.37	0.63	0.47
0.2	Mean	124.36	84.39	90.67	97.83	91.65	67.33	**65.85(p=0.8)**
	Std	1.01	0.22	0.55	0.70	0.65	1.48	0.32
0.3	Mean	134.29	92.33	93.48	99.59	93.29	79.30	**71.90(p=0.8)**
	Std	1.53	1.05	0.89	0.69	0.66	2.12	0.63
0.4	Mean	144.22	101.58	96.86	101.91	96.13	86.54	**79.27(p=0.2)**
	Std	1.05	0.60	0.59	0.51	0.58	0.72	0.68
0.5	Mean	154.01	115.49	100.86	105.08	99.60	94.51	**88.34(p=0.2)**
	Std	0.89	0.95	0.08	0.37	0.01	0.29	0.42

of p used in this experiment. As we can see, overall, MCAR missing pattern is the easiest situation while MAR is the most difficult case in terms of imputation. It is reasonable because successive missing will lose much valuable information about correlation, thus increasing the difficulty of accurate recovery. With respect to recovery performance, we find that KNN performs worst among these algorithms. LLS, PPCA, LRMC, SR-l_2, and SRSp all significantly outperform KNN. In particular, LLS works well in low missing ratio, however, rapidly degrades when missing ratio increases. Self-representation based methods, including SR-l_2, SRSp, and our proposed SSR-l_p, obtain superior imputation performance than other competing methods. Comparing these three methods, we find that SRSp obtains better performance than SR-l_2, indicating that sparsity is a crucial factor for self-representation based imputation. This conclusion is consistent with that drawn in [18]. At last, SSR-l_p achieves best performance in most cases. In fact, it is interesting to notice that as the missing ratio increases, a smaller p is preferred which implies fewer samples should be selected for MV recovery in such situations. Some examples under MIXED missiont pattern and $\delta = 0.3$ are shown in Fig.4-Fig.10. As we can see, the proposed SSR-l_p achieves small residual in the recovery of missing data.

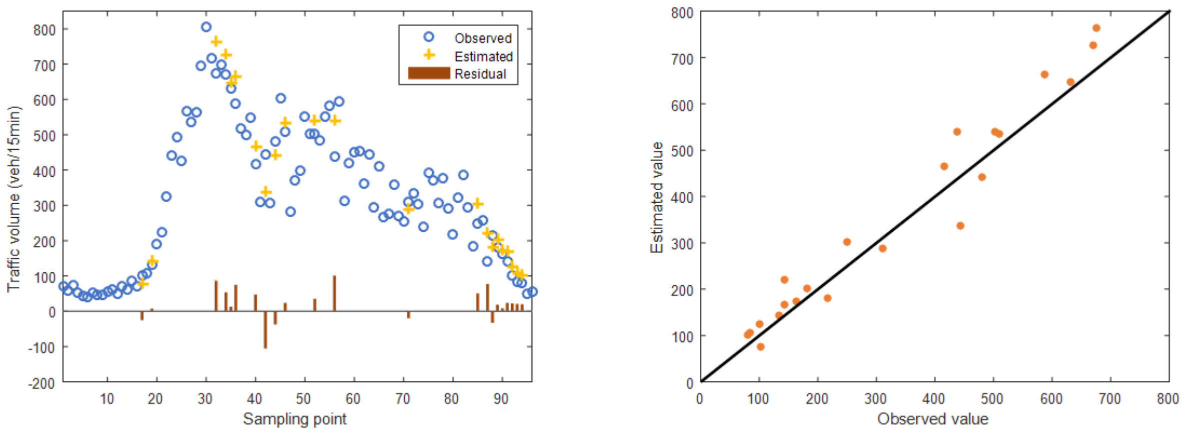
E. INFLUENCE OF l_p ON PERFORMANCE

In what follows, we investigate the recovery performance when varying the value of p in order to confirm that it is an



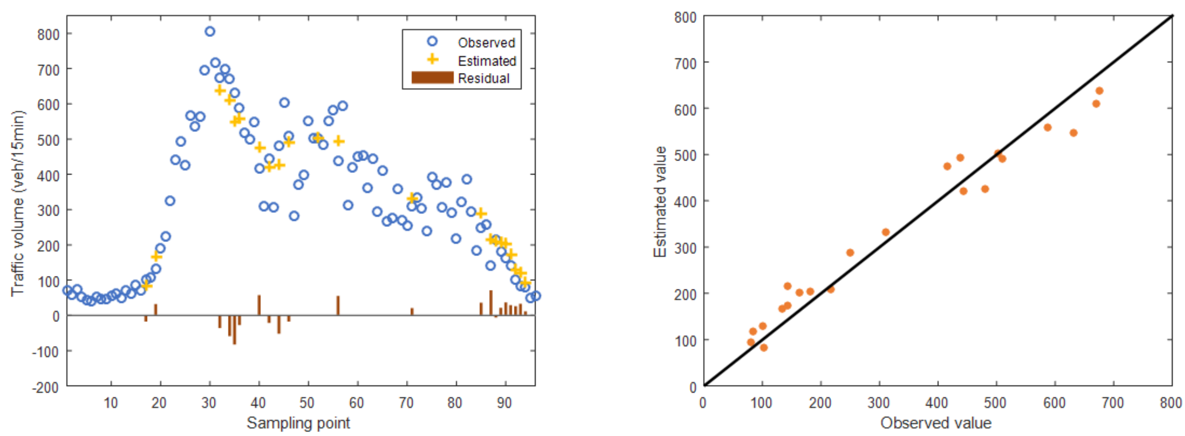
(a) (b)

FIGURE 5. **Imputation results obtained by LLS.**



(a) (b)

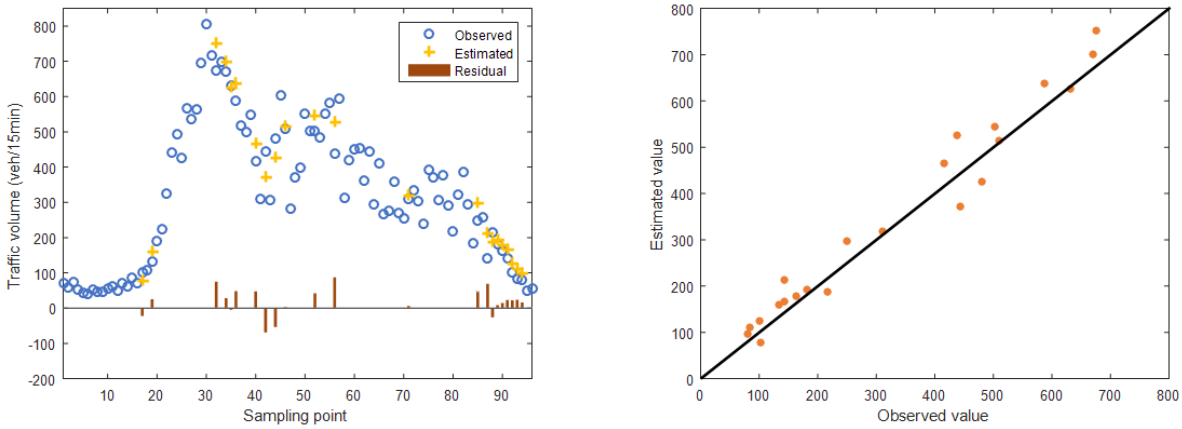
FIGURE 6. **Imputation results obtained by PPCA.**



(a) (b)

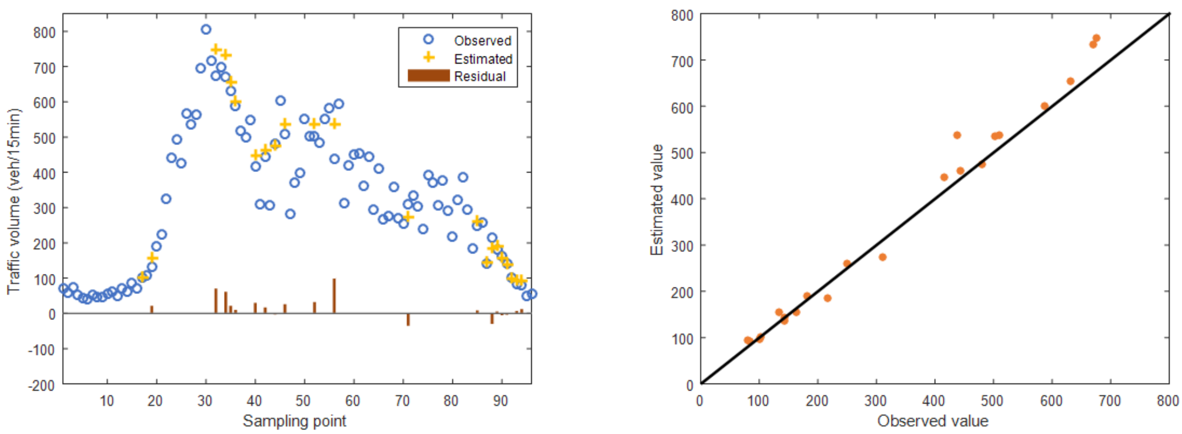
FIGURE 7. **Imputation results obtained by LRMC.**

important factor for self-representation based matrix completion. In particular, we change the parameter p in the range of $\{0.2, 0.4, 0.6, 0.8, 1.0\}$ and record the best performance for each candidate value. Some experimental results under MCAR, MAR and MIXED missing patterns are presented in Fig.11. As we can see, the variation trends of performance



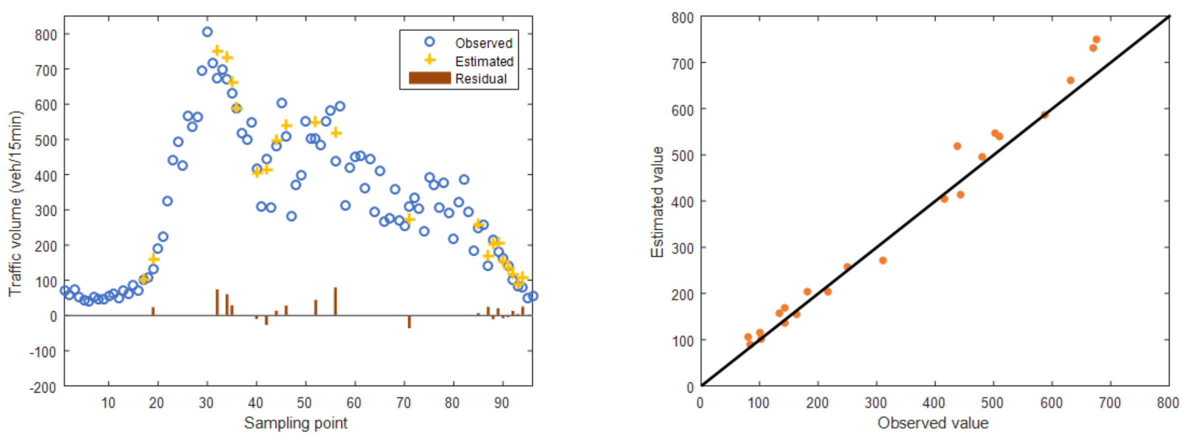
(a) (b)

FIGURE 8. Imputation results obtained by SR-l_2.



(a) (b)

FIGURE 9. Imputation results obtained by SRSp.



(a) (b)

FIGURE 10. Imputation results obtained by SSR-l_p.

when varying p is different under different missing ratio δ. When δ is small, the performance can be improved by increasing the value of p. In contrast, when δ is large, smaller p is preferred. The results indicate that given more observed entries, our method is apt to use more samples for accurate imputation. On the contrary, when there are too many missing

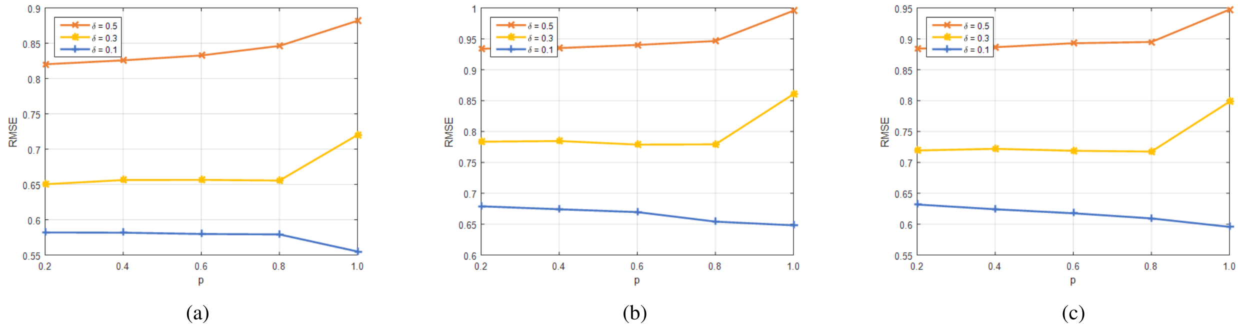


FIGURE 11. **Imputation performance variation with respect to different value of the parameter p. (a) MCAR, $\delta = 0.1$. (b) MCAR, $\delta = 0.3$. (c) MCAR, $\delta = 0.5$.**

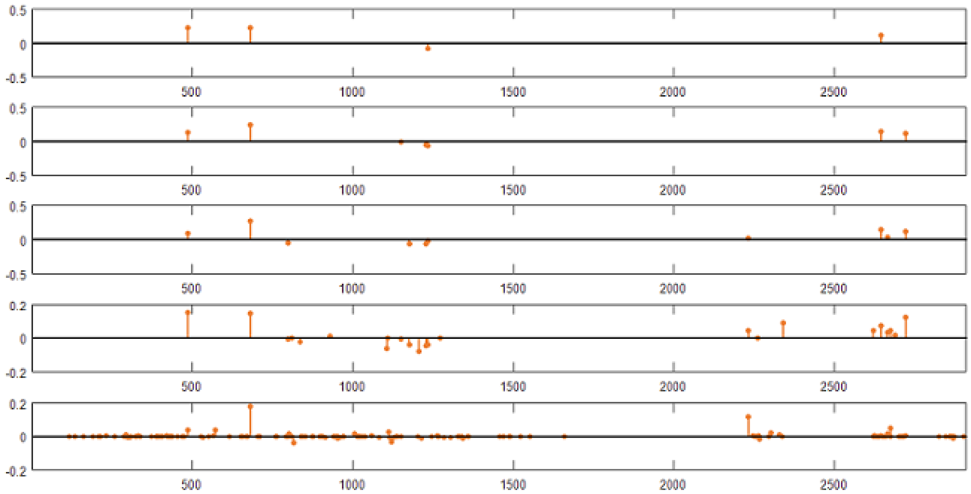


FIGURE 12. **Variation of weights when increasing p from 0.2 (top row) to 1.0 (bottom row) with step 0.2.**

values, our method adjusts itself to utilize samples as fewer as possible. These results are essentially consistent with the above quantitative comparison.

Next, we investigate how the parameter p influences the sparsity of the resulting model. In this experiments, we take one data sample as instance and fix λ to a constant and change p in the same range above. Fig. 6 demonstrates the resulting $w_1 \in \mathbb{R}^{2910}$ after optimization. We show the weights with absolute value larger than 10^{-4} following [42], [22]. As we can see, when p equals to 0.2, only four samples are selected. When p increases to 0.4, three extra samples are selected. In a similar way, with the increase of p, more and more samples are selected. These results empirically verify that, the proposed SSR-l_p model is able to produce the solution with more sparsity given a small value of p.

V. CONCLUSIONS

In this paper, we develop a novel MV imputation algorithm based on self-representation and l_p-norm minimization. With the introduction of l_p-norm, our method is able to find sparser representation for each sample, which in turn facilitate the accurate recovery of missing data. To solve the resulting model, we further develop an algorithm which optimizes the missing data and the representation coefficients alternatively. The experimental results confirm the effectiveness of our method. An interesting extension of our work is investigating the nonlinear formulation of the proposed model, which is more powerful in modeling nonlinear structure of data. Another future work is to extend our proposal to large-scale problems by applying some techniques, such as parallelization, etc.

REFERENCES

[1] X. Chen, Z. Wei, X. Liu, Y. Cai, Z. Li, and F. Zhao, "Spatiotemporal variable and parameter selection using sparse hybrid genetic algorithm for traffic flow forecasting," *Int. J. Distrib. Sens. Netw.*, vol. 13, no. 6, pp. 1–14, 2017.

[2] X. Chen, X. Cai, J. Liang, and Q. Liu, "Ensemble learning multiple LSSVR with improved harmony search algorithm for short-term traffic flow forecasting," *IEEE Access*, vol. 6, pp. 9347–9357, 2018.

[3] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.

[4] H. Wang, L. Dai, Y. Cai, X. Sun, and L. Chen, "Salient object detection based on multi-scale contrast," *Neural Netw.*, vol. 101, pp. 47–56, May 2018.

[5] D. Tian, J. Zhou, Z. Sheng, M. Chen, Q. Ni, and V. C. M. Leung, "Self-organized relay selection for cooperative transmission in vehicular ad-hoc networks," *IEEE Trans. Veh. Technol.*, vol. 66, no. 10, pp. 9534–9549, Oct. 2017.

[6] D. Tian, J. Zhou, and Z. Sheng, "An adaptive fusion strategy for distributed information estimation over cooperative multi-agent networks," *IEEE Trans. Inf. Theory*, vol. 63, no. 5, pp. 3076–3091, May 2017.

[7] J. Luengo, S. García, and F. Herrera, "On the choice of the best imputation methods for missing values considering three groups of classification methods," *Knowl. Inf. Syst.* vol. 32, no. 1, pp. 77–108, Jul. 2012.
[8] G. E. Batista and M. C. Monard, "An analysis of four missing data treatment methods for supervised learning," *Appl. Artif. Intell.*, vol. 17, nos. 5–6, pp. 519–533, 2003.
[9] O. Troyanskaya *et al.*, "Missing value estimation methods for DNA microarrays," *Bioinformatics*, vol. 17, no. 6, pp. 520–525, 2001.
[10] H. Kim, G. H. Golub, and H. Park, "Missing value estimation for DNA microarray gene expression data: Local least squares imputation," *Bioinformatics*, vol. 21, no. 2, pp. 187–198, Jan. 2005.
[11] Z. Yu, T. Li, S.-J. Horng, Y. Pan, H. Wang, and Y. Jing, "An iterative locally auto-weighted least squares method for microarray missing value estimation," *IEEE Trans. Nanobiosci.*, vol. 16, no. 1, pp. 21–33, Jan. 2017.
[12] M. E. Tipping and C. M. Bishop, "Probabilistic principal component analysis," *J. Roy. Statist. Soc. B*, vol. 61, no. 3, pp. 611–622, 1999.
[13] L. Qu, L. Li, Y. Zhang, and J. Hu, "PPCA-based missing data imputation for traffic flow volume: A systematical approach," *IEEE Trans. Intell. Transp. Syst.*, vol. 10, no. 3, pp. 512–522, Sep. 2009.
[14] E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," *Found. Comput. Math.*, vol. 9, no. 6, pp. 717–772, 2009.
[15] J.-F. Cai, E. J. Candèls, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM J. Optim.*, vol. 20, no. 4, pp. 1956–1982, 2010.
[16] H. Tan, "Traffic missing data completion with spatial-temporal correlations," in *Proc. TRB Annu. Meeting*, 2014, pp. 1–16.
[17] X. Chen, Z. Wei, Z. Li, J. Liang, Y. Cai, and B. Zhang, "Ensemble correlation-based low-rank matrix completion with applications to traffic data imputation," *Knowl.-Based Syst.*, vol. 132, pp. 249–262, Sep. 2017.
[18] J. Fan and T. W. S. Chow, "Matrix completion by least-square, low-rank, and sparse self-representations," *Pattern Recognit.*, vol. 71, pp. 290–305, Nov. 2017.
[19] E. Elhamifar and R. Vidal, "Sparse subspace clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 2790–2797.
[20] E. Elhamifar and R. Vidal, "Sparse subspace clustering: Algorithm, theory, and applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 11, pp. 2765–2781, Nov. 2013.
[21] P. Zhu, W. Zuo, L. Zhang, Q. Hu, and S. C. K. Shiu, "Unsupervised feature selection by regularized self-representation," *Pattern Recognit.*, vol. 48, no. 2, pp. 438–446, 2015.
[22] Q. Lyu, Z. Lin, Y. She, and C. Zhang, "A comparison of typical ℓ_p minimization algorithms," *Neurocomputing*, vol. 119, pp. 413–424, Nov. 2013.
[23] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[24] M. Zhao, H. Zhang, W. Cheng, and Z. Zhang, "Joint ℓ_p- and ℓ_2,p-norm minimization for subspace clustering with outlier pursuit," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2016, pp. 3658–3665.
[25] R. Chartrand, "Nonconvex compressed sensing and error correction," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, vol. 3. Apr. 2007, pp. 889–892.
[26] R. Saab, R. Chartrand, and O. Yilmaz, "Stable sparse approximations via nonconvex optimization," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Mar./Apr. 2008, pp. 3885–3888.
[27] X. Chen, J. Yang, and L. Chen, "An improved robust and sparse twin support vector regression via linear programming," *Soft Comput.*, vol. 18, no. 12, pp. 2335–2348, Dec. 2014.
[28] E.-L. Silva-Ramírez, R. Pino-Mejías, M. López-Coello, and M.-D. Cubiles-de-la-Vega, "Missing value imputation on missing completely at random data using multilayer perceptrons," *Neural Netw.*, vol. 24, no. 1, pp. 121–129, Jan. 2011.
[29] X. Sun, L. Chen, Z. Yang, and H. Zhu, "Speed-sensorless vector control of a bearingless induction motor with artificial neural network inverse speed observer," *IEEE/ASME Trans. Mechatronics*, vol. 18, no. 4, pp. 1357–1366, Aug. 2013.
[30] Z. Cai, M. Heydari, and G. Lin, "Iterated local least squares microarray missing value imputation," *J. Bioinform. Comput. Biol.*, vol. 4, no. 5, pp. 935–957, 2006.
[31] G. Chang, Y. Zhang, and D. Yao, "Missing data imputation for traffic flow based on improved local least squares," *Tsinghua Sci. Technol.*, vol. 17, no. 3, pp. 304–309, Jun. 2012.
[32] C. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer 2006.
[33] Y. Zhang, Y. Wang, J. Jin, and X. Wang, "Sparse Bayesian Learning for obtaining sparsity of EEG frequency bands based feature vectors in motor imagery classification," *Int. J. Neural Syst.*, vol. 27, no. 2, p. 1650032, 2016.
[34] F. Shi, D. Zhang, J. Chen, and H. R. Karimi, "Missing value estimation for microarray data by Bayesian principal component analysis and iterative local least squares," *Math. Problems Eng.*, vol. 2013, Mar. 2013, Art. no. 162938.
[35] L. Qu, Y. Zhang, J. Hu, L. Jia, and L. Li, "A BPCA based missing value imputing method for traffic flow volume data, " in *Proc. IEEE Intell. Vehicles Symp.*, Jun. 2008, pp. 985–990.
[36] M. T. Asif, N. Mitrovic, L. Garg, J. Dauwels, and P. Jaillet, "Low-dimensional models for missing data imputation in road networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2013, pp. 3527–3531.
[37] B. Ran, H. Tan, J. Feng, Y. Liu, and W. Wang, "Traffic speed data imputation method based on tensor completion," *Comput. Intell. Neurosci.*, vol. 2015, Jan. 2015, Art. no. 364089.
[38] S. Ma, D. Goldfarb, and L. Chen, "Fixed point and Bregman iterative methods for matrix rank minimization," *Math. Program.*, vol. 128, nos. 1–2, pp. 321–353, 2011.
[39] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Jan. 2011.
[40] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 2012.
[41] M. T. Asif, N. Mitrovic, J. Dauwels, and P. Jaillet, "Matrix and tensor based methods for missing data estimation in large traffic networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 7, pp. 1816–1825, Jul. 2016.
[42] K. Huang, D. Zheng, J. Sun, Y. Hotta, K. Fujimoto, and S. Naoi, "Sparse learning for support vector classification," *Pattern Recognit. Lett.*, vol. 31, no. 13, pp. 1944–1951, Oct. 2010.
[43] X. Sun, B. Su, L. Chen, Z. Yang, X. Xu, and Z. Shi, "Precise control of a four degree-of-freedom permanent magnet biased active magnetic bearing system in a magnetically suspended direct-driven spindle using neural network inverse scheme," *Mech. Syst. Signal Process.*, vol. 88, pp. 36–48, May 2017.
[44] Y. Zhang, Q. Zhao, G. Zhou, J. Jin, X. Wang, and A. Cichocki, "Removal of EEG artifacts for BCI applications using fully Bayesian tensor completion," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Mar. 2016, pp. 819–823.

XIAOBO CHEN received the Ph.D. degree in pattern recognition and intelligent systems from Nanjing University of Science and Technology, China, in 2013. From to 2011, he served as a Research Assistant with the Hong Kong Polytechnic University, Hong Kong. From 2015 to 2017, he served as a Post-Doctoral Research Associate with the University of North Carolina at Chapel Hill, USA. He is currently an Associate Professor with the Automotive Engineering Research Institute, Jiangsu University, China.

His major research interests concentrate on pattern recognition and its applications. He has published over 50 papers in numerous international journals and conference proceedings, such as pattern recognition, human brain mapping, neurocomputing, knowledge-based systems, neuroinformatics, AAAI, ICPR, and MICCAI.

He received the Best Paper Award from the Asian Conference on Pattern Recognition in 2017, the Excellent Doctoral Thesis prize from NJUST in 2013, and the Best Student Paper Award from the Chinese Conference on Pattern Recognition in 2010.

YINGFENG CAI received the B.S., M.S., and Ph.D. degrees from the School of Instrument Science and Engineering, Southeast University, Nanjing, China, respectively. In 2013, she joined the Automotive Engineering Research Institute, Jiangsu University, as an Assistant Professor.

Her research interests include computer vision, intelligent transportation systems, and intelligent automobiles.

QINGCHAO LIU received the B.S. and M.S. degrees from the Jiangsu University, Zhenjiang, China, and the Ph.D. degree from Southeast University, Nanjing, China. In 2015, he joined the Automotive Engineering Research Institute, Jiangsu University, as a Lecturer. His research interests include traffic safety, intelligent transportation systems, and intelligent automobiles.

LEI CHEN received the Ph.D. degree in communication and information system from the Nanjing University of Posts and Telecommunications, Nanjing, China, in 2014. He is currently an Associate Professor with the Nanjing University of Posts and Telecommunications. His research interests include pattern recognition, machine learning, and neural computing.