A Missing Data Recovery Method of Sparse Graph Signal in GFT Domain

Chien-Cheng Tseng[1], *Senior Member, IEEE* and Su-Ling Lee[2]
[1]National Kaohsiung University of Science and Technology, Kaohsiung, Taiwan
[2]Chang Jung Christian University, Tainan, Taiwan

***Abstract*-In this paper, a missing data recovery method of sparse graph signal in graph Fourier transform (GFT) domain is presented. First, *K*-sparse graph signal which only has *K* non-zero elements in GFT domain is defined. The conventional *K*-bandlimited constraint is only a special case of *K*-sparse constraint. Then, missing data recovery problem of *K*-sparse graph signal is formulated as an optimization problem such that it can be solved by using basis pursuit, orthogonal matching pursuit or iterative hard thresholding method. Finally, real temperature data are used to demonstrate the effectiveness of the proposed recovery method.**

I. INTRODUCTION

In the conventional digital signal processing theory, the uniformly-sampled signal is a time-domain sequence or a spatial-domain array. Examples are digital speech or digital image signals which are regular data. However, in recent years, there is a massive progress in the researches of sensor network, traffic transportation network, social network and biological brain network. This leads to our great attentions in the researches of processing irregular signals and data from these networks. Generally speaking, the irregular signals are represented by the graphs composed of several vertices, edges and weights. In order to process irregular graph signals, the theory of graph signal processing (GSP) has been proposed [1]. In this new research area, the conventional time-domain or spatial domain operators, theorems and tools are extended to the vertex-domain ones including Fourier transform, frequency-selective filter and vertex-frequency analysis etc. These graph signal processing analysis tools can be used to solve various irregular signal processing problems such as missing data recovery, filtering, prediction, de-noising, compression, modeling, classification and clustering etc [2]. The purpose of this paper is to study the missing data recovery problem of sparse graph signal.

II. MISSING DATA RECOVERY OF SPARSE GRAPH SIGNAL

First, the background of GSP is briefly reviewed. Let graph signal $\boldsymbol{f} = [f(1), f(2), \cdots, f(N)]^T$ be defined on a finite undirected graph $G = (V, \boldsymbol{A})$ where $V = \{v_1, v_2, \cdots, v_N\}$ is the vertex set and $\boldsymbol{A}$ is the adjacent matrix. The elements $a_{n,m}$ of symmetric adjacent matrix $\boldsymbol{A}$ give the weight of the edge connecting vertices v_n to v_m. The weight value denotes the similarity between the corresponding signal values. We assume the graph does not have self loops; that is, $a_{n,n} = 0$. Let d_n be the sum of all edge weights connected to vertex v_n, then $\boldsymbol{D} = diag\,[d_1, d_2, \cdots, d_N]$ is the well-known diagonal degree matrix. The graph Laplacian matrix is defined as $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{A}$ whose eigen-decomposition is given by

$$\boldsymbol{L} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^T \tag{1}$$

The diagonal matrix $\boldsymbol{\Lambda}$ is constructed by the real eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_N$ and unitary matrix $\boldsymbol{U}$ is composed of the eigenvectors $\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_N$. The graph Fourier transform (GFT) of signal $\boldsymbol{f}$ is then defined as $\hat{\boldsymbol{f}} = \boldsymbol{U}^T\boldsymbol{f}$. And, its corresponding inverse graph Fourier transform is given by $\boldsymbol{f} = \boldsymbol{U}\hat{\boldsymbol{f}}$. A graph signal is *K*-sparse in GFT domain if the following condition is satisfied:

$$\|\hat{\boldsymbol{f}}\|_0 = K \tag{2}$$

where ℓ_0 pseudo-norm $\|\hat{\boldsymbol{f}}\|_0$ denotes the number of non-zero elements of GFT $\hat{\boldsymbol{f}} = [\hat{f}(\lambda_1), \hat{f}(\lambda_2), \cdots, \hat{f}(\lambda_N)]^T$. Moreover, the graph signal is called *K*-bandlimited in spectral domain if the condition is met:

$$\hat{f}(\lambda_k) = 0 \quad \text{for all } k > K \tag{3}$$

Obviously, if the non-zero elements of $\hat{\boldsymbol{f}}$ are only the first *K* elements, then *K*-sparse condition is the same as the *K*-bandlimited condition. So, the conventional *K*-bandlimited constraint is only a special case of *K*-sparse constraint studied in this paper.

Next, let us study how to solve the missing data recovery problem of *K*-sparse graph signal in GFT domain. If the index set of measured signal is denoted by $R_1 = \{i_1, i_2, \cdots, i_M\}$, then index set of missing data is $R_2 = \{1, 2, \cdots, N\} \setminus R_1$, where notation \ stands for the set difference operator. Thus, the $f(i)$ with $i \in R_1$ are measured data and the $f(i)$ with $i \in R_2$ are missing data. Given the measured data $f(i)$ with $i \in R_1$, the problem is how to recover the original graph signal $\boldsymbol{f}$ such that the missing data $f(i)$ with $i \in R_2$ can be estimated. In order to solve this problem, let us define the measured signal vector below:

$$\boldsymbol{y} = [f(i_1), f(i_2), \cdots, f(i_M)]^T \tag{4}$$

Based on inverse GFT $\boldsymbol{f} = \boldsymbol{U}\hat{\boldsymbol{f}}$, vector $\boldsymbol{y}$ can be written as

$$\boldsymbol{y} = \begin{bmatrix} \boldsymbol{b}_{i_1}^T \\ \vdots \\ \boldsymbol{b}_{i_M}^T \end{bmatrix} \hat{\boldsymbol{f}} = \boldsymbol{\Phi}\hat{\boldsymbol{f}} \tag{5}$$

where notation $\boldsymbol{b}_{i_n}^T$ denotes the $i_n - th$ row of the matrix $\boldsymbol{U}$. Because $\hat{\boldsymbol{f}}$ is a *K*-sparse signal, the recovery problem can be formulated as the following optimization problem:

$$\text{Minimize } \|\hat{\boldsymbol{f}}\|_0 \quad \text{Subject to } \boldsymbol{y} = \boldsymbol{\Phi}\hat{\boldsymbol{f}} \tag{6}$$

This optimization problem has been well investigated in the research of sparse representation [3][4]. So far, several methods have been presented to solve this problem including basis pursuit (BP), orthogonal matching pursuit (OMP) or iterative hard thresholding (IHT) method. In the BP method, the ℓ_0 norm $\|\hat{\boldsymbol{f}}\|_0$ in (6) is replaced by ℓ_1 norm $\|\hat{\boldsymbol{f}}\|_1$ such that the convex programming method can be used to get solution. In the OMP method, the procedure to solve (6) is listed below:

input: Matrix $\boldsymbol{\Phi}$, vector $\boldsymbol{y}$ and integer K

initialize: $\boldsymbol{r}_0 = \boldsymbol{y}$, $\hat{\boldsymbol{f}}_0 = \boldsymbol{0}$, $\Lambda_0 = empty\ set$, $\ell = 0$

while $\ell \le K$ **do**

  match: $\boldsymbol{s}_\ell = \boldsymbol{\Phi}^T \boldsymbol{r}_\ell$

  identify: $\Lambda_{\ell+1} = \Lambda_\ell \cup \{\arg\max_k |\boldsymbol{s}_\ell(k)|\}$

  update: $\hat{\boldsymbol{f}}_{\ell+1} = \arg \min_{z:\Gamma(z)\subseteq\Lambda_{\ell+1}} \|\boldsymbol{y} - \boldsymbol{\Phi z}\|_2$

  $\boldsymbol{r}_{\ell+1} = \boldsymbol{y} - \boldsymbol{\Phi}\hat{\boldsymbol{f}}_{\ell+1}$

  $\ell = \ell + 1$

end while

output $\hat{\boldsymbol{f}} = \hat{\boldsymbol{f}}_\ell$

The above notation $\Gamma(\boldsymbol{z})$ stands for the support set of vector $\boldsymbol{z}$, i.e., the index set of non-zero elements of $\boldsymbol{z}$. In the IHT method, the procedure to solve (6) is listed below:

input: Matrix $\boldsymbol{\Phi}$, vector $\boldsymbol{y}$, integer K, and step size μ

initialize: $\hat{\boldsymbol{f}}_0 = \boldsymbol{0}$, $n = 0$

while $(n < Itr)$

  $\hat{\boldsymbol{f}}_{n+1} = H_K\left(\hat{\boldsymbol{f}}_n + \mu\boldsymbol{\Phi}^T(\boldsymbol{y} - \boldsymbol{\Phi}\hat{\boldsymbol{f}}_n)\right)$

  $n = n + 1$

end while

output $\hat{\boldsymbol{f}} = \hat{\boldsymbol{f}}_n$

where $H_K(\cdot)$ is a nonlinear operator which sets all but the largest K elements in a vector to zero and the Itr is the prescribed maximum number of iteration. Finally, given the index set $R_1 = \{i_1, i_2, \cdots, i_M\}$, the GFT matrix $\boldsymbol{U}$, measured data $\boldsymbol{y}$ and integer K, the procedure of the proposed missing data recovery method is listed below:

Step 1. Use GFT matrix $\boldsymbol{U}$ and set R_1 to get the matrix $\boldsymbol{\Phi}$.

Step 2: Use BP, OMP or IHT method to solve the problem in (6) to get the optimal solution $\hat{\boldsymbol{f}}$.

Step 3: Use the inverse graph Fourier transform $\boldsymbol{f} = \boldsymbol{U}\hat{\boldsymbol{f}}$ to compute the recovered graph signal $\boldsymbol{f}$.

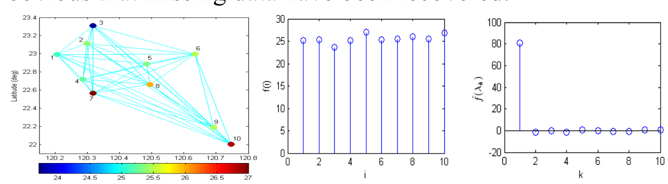
Until now, the proposed method has been described. In next section, one example is used to evaluate the performance.

III. NUMERICAL EXAMPLE

In this section, the missing data recovery problem of temperature signal is studied. The temperature data measured in 10 south Taiwan cities shown in Fig.1(a) are adopted to make experiment. Ten cities include Nantz, Kaohsiung, Chishan, Lioukei, Pingdong, Fongong, Hengchun, Tainan, Shanhua, and Shinying. The elements of adjacency matrix $\boldsymbol{A}$ are chosen as

$$a_{n,m} = \begin{cases} e^{-\frac{dist(m,n)^2}{2\sigma^2}} & m \ne n \\ 0 & m = n \end{cases} \tag{7}$$

where $dist(m,n)$ is the distance between m-th and n-th cities. The $a_{n,n} = 0$ are chosen to assume that undirected graph does not have self loops. Here, parameters are chosen as $N = 10$ and $\sigma = 80$. Fig.1(b)(c) show the graph signal $\boldsymbol{f}$ and its GFT $\hat{\boldsymbol{f}}$. It is clear that graph signal in GFT domain is sparse, so the proposed method can be used. Fig.2(a) shows a signal with measured index set $R_1 = \{1,2,4,5,6,7,9,10\}$, that is, $f(3)$ and $f(8)$ are missed. Fig.2(c)-(d) show the recovered graph signals by using BP, OMP and IHT methods with $K = 4$. Compared these results with original signal in Fig.1(a), it is obvious that missing data have been recovered.



(a) (b) (c)

Fig.1 (a) Ten cities. (b) The graph signal $\boldsymbol{f}$. (b) The GFT $\hat{\boldsymbol{f}}$ of graph signal.

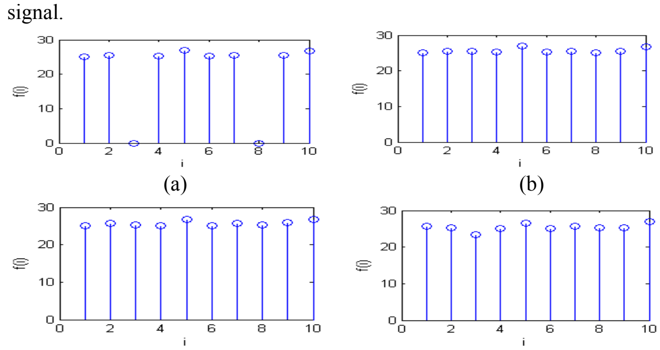


Fig.2 (a) The measured graph signal $\boldsymbol{f}$ with missing data. (b) Recovered signal of BP. (c) Recovered signal of OMP. (d) Recovered signal of IHT.

IV. CONCLUSIONS

In this paper, a missing data recovery method of sparse graph signal in GFT domain has been presented. However, missing data recovery problem is only solved by using sparse representation method. Thus, it is interesting to use other optimization methods to solve recovery problem in the future.

REFERENCES

[1] O. Teke and P.P. Vaidyanathan, "Uncertainty principles and sparse eigenvectors of graphs," *IEEE Trans. on Signal Processing*, vol.65, no.20, pp.5406-5420, Oct. 2017.

[2] C.C. Tseng and S.L. Lee, "A missing temperature data estimation method using graph Fourier transform," *IEEE ICCE-TW*, pp.87-88, June 2017.

[3] M. Elad, *Sparse and Redundant Representation*, Springer, 2013.

[4] Y.C. Eldar and G. Kutyniok, *Compressed Sensing: Theory and Applications*, Cambridge University Press, 2012.