

LETTER

An Improved Traffic Matrix Decomposition Method with Frequency-Domain Regularization*

Zhe WANG^{†a)}, Nonmember, Kai HU[†], Member, and Baolin YIN[†], Nonmember

SUMMARY We propose a novel network traffic matrix decomposition method named *Stable Principal Component Pursuit with Frequency-Domain Regularization* (SPCP-FDR), which improves the Stable Principal Component Pursuit (SPCP) method by using a frequency-domain noise regularization function. An experiment demonstrates the feasibility of this new decomposition method.

key words: traffic matrix, stable principal component pursuit, frequency-domain regularization, iterative algorithm

1. Introduction

The network traffic matrix has been applied to many significant application problems such as capacity planning, traffic engineering and anomaly detection. A traffic matrix combines diverse traffic components with distinct temporal properties. Therefore, it is necessary to decompose them efficiently, and this problem is named traffic matrix structural analysis [1]. We presented the traffic matrix decomposition model in [2] and decomposed a traffic matrix into three sub-matrices, which is equivalent to the generalized *Robust Principal Component Analysis* (RPCA) problem [3]. The results in [2] were achieved by applying the *Stable Principal Component Pursuit* (SPCP) method in [3].

In this study, we improve the traffic matrix decomposition method by using frequency-domain regularization. This method is a variation of SPCP, and is named *Stable Principal Component Pursuit with Frequency-Domain Regularization* (SPCP-FDR). We design the numerical algorithm for SPCP-FDR, evaluate its decomposition results on the Abilene dataset [7], and show that SPCP-FDR achieves more rational traffic decompositions compared with SPCP.

2. Background

The standard RPCA problem is formally defined in [4]:

Problem 1 (Standard RPCA) *Suppose that a known matrix $X \in \mathbb{R}^{m \times n}$ is of the form $A + E$, where A and E are unknown matrices. It is assumed that A has a low rank and E*

is sparse. The problem is to recover A and E .

In many applications, data matrices satisfying the assumptions in Problem 1 are polluted by dense noise with small magnitude. This leads to the generalized RPCA [3]:

Problem 2 (Generalized RPCA) *Suppose that a known matrix $X \in \mathbb{R}^{m \times n}$ is of the form $A + E + N$, where A , E and N are unknown matrices. It is assumed that A has a low-rank, E is sparse, and N is an i.i.d. entry-wide noise matrix with small magnitude. The problem is to recover A , E and N .*

We refer RPCA as the generalized version as follows. Zhou et al. [3] proved that under surprising board conditions, for "almost all" data matrix X which is the sum of a low-rank matrix and a sparse matrix, and is corrupted by a dense noise matrix N whose Frobenius norm $\|N\|_F \leq \delta$ (δ is a positive constant), one could stably estimate A and E with high probability using this convex program:

$$\min_{A, E} \|A\|_* + \lambda \|E\|_1 \quad s.t. \|X - A + E\|_F \leq \delta, \quad (1)$$

where λ is a positive parameter, $\|\cdot\|_*$ and $\|\cdot\|_1$ denote the matrix nuclear norm and the l_1 norm, respectively. They named this method as *Stable Principal Component Pursuit* (SPCP), which needs to solve a time-consuming constrained program. Thus they turned to solve another similar unconstrained program instead (τ is another positive parameter):

$$\min_{A, E} \tau (\|A\|_* + \lambda \|E\|_1) + \frac{1}{2} \|X - A - E\|_F^2. \quad (2)$$

Suppose $X \in \mathbb{R}^{T \times P}$ is a traffic matrix, and each column $X_j \in \mathbb{R}^T$ ($1 \leq j \leq P$) is the traffic time-series of an Original-Destination (OD) flow. Following the traffic matrix decomposition model [2], X is the sum of three sub-matrices: (1) The deterministic traffic matrix A is a low-rank matrix, contributed by the periodic traffic changes in each OD flow; (2) The anomaly traffic matrix E is a sparse matrix, but its nonzero entries may have large magnitudes; (3) The noise matrix N is constituted of independent random variables with relatively small magnitudes, the entries in one column constitute a white noise vector, but those in different columns have distinct variances. In [2], we estimated the noise traffic variances $\{\sigma_j\}_{j=1}^P$ of all the OD flows in X , and divided X_j by σ_j ($1 \leq j \leq P$). This preprocessing normalizes random variables in N , and preserves the rank of A , as well as the sparsity of E . Thus traffic matrix decomposition is equivalent to RPCA, and we only consider traffic matrices with unit noise traffic variance.

Manuscript received September 12, 2012.

Manuscript revised November 2, 2012.

[†]The authors are with the State Key Laboratory of Software Development Environment, the School of Computer Science and Engineering, Beihang University, Beijing, China.

*This work is supported by the National Natural Science Foundation of China (Nos. 61073013 and 90818024) and the Open Project of State Key Laboratory of Software Development Environment (No. SKLSDE-2012ZX-15).

a) E-mail: wangzhe@cse.buaa.edu.cn

DOI: 10.1587/transinf.E96.D.731

3. Stable Principal Component Pursuit with Frequency-Domain Regularization

Suppose " \otimes " is a traffic matrix decomposition method, and $N^\otimes \in \mathbb{R}^{T \times P}$ is the decomposed noise traffic matrix of X by \otimes . In this letter, SPCP and SPCP-FDR are denoted as " \odot " and " \oplus ", respectively. We consider the frequency-domain property of the decomposed noise traffic matrices. For each noise traffic time-series N_j^\otimes (the j -th column of N^\otimes), $\alpha_j \in \mathbb{C}^T$ denotes its Discrete Fourier Transform (DFT):

$$\alpha_j = W^T N_j^\otimes = [W_1 \ W_2 \ \cdots \ W_T]^T N_j^\otimes, \quad (3)$$

where $W \in \mathbb{C}^{T \times T}$ is the discrete Fourier basis matrix. The t -th column vector $W_t \in \mathbb{C}^T$ ($1 \leq t \leq T$) is defined as:

$$W_t(k) = \frac{1}{\sqrt{T}} e^{-\frac{2\pi i}{T}(t-1)(k-1)} \quad k = 1, 2, \dots, T. \quad (4)$$

N_j^\otimes 's spectral density $\varphi_j \in \mathbb{R}^T$ is defined as:

$$\varphi_j(t) = |\alpha_j(t)|^2 \quad t = 1, 2, \dots, T. \quad (5)$$

As N_j^\otimes is a real signal, $\varphi_j(t) = \varphi_j(T - t + 2)$ for each position $t \in [2, T] \cap \mathbb{N}$, and $(T - t + 2)$ is named the dual position of t . The spectra $\varphi_j(t)$ at positions close to 2 or T describe N_j^\otimes 's power distributed in low-frequency domain; conversely, the spectra at positions close to $(\frac{T}{2} + 1)$ indicate the high-frequency power. Furthermore, the spectral density $\Phi_{N^\otimes} \in \mathbb{R}^T$ of N^\otimes is defined as the entry-wise sum of all the noise traffic time-series' spectral density:

$$\Phi_{N^\otimes}(t) = \sum_{j=1}^P \varphi_j(t) \quad t = 1, 2, \dots, T. \quad (6)$$

We compute the spectral density of eight noise traffic matrices $N01^\odot \sim N08^\odot$ and display them in Fig. 1, which are independently decomposed from the eight Abilene [7] weekly traffic matrices $X01 \sim X08$ by SPCP we adopted in [2]. In this dataset, $P = 121$, $T = 2016$, and the minimal time interval is 5 minutes. Thus for each noise traffic matrix $N0x^\odot$ ($x \in \{1, 2, \dots, 8\}$), $\Phi_{N0x^\odot}(t)$ represents the power captured by periodic traffic with period $\frac{2016}{\min(t-1, 2016-t+2)} \cdot \frac{5}{60}$ hour(s). We discover these two properties for the noise traffic matrices decomposed by SPCP: (1) The low-frequency spectra are generally much larger than the high-frequency spectra. This means that most energy of these noise traffic matrices is contributed by the low-frequency traffic patterns; (2) Quite a few low-frequency spectra have dramatically larger values than their neighbors (No. 8, 15, 29, 57, 113 and 169 spectra, as well as the spectra at dual positions, and their corresponding periods are 24, 12, 6, 3, 1.5 and 1 hour(s), respectively).

These properties match our assumptions of the noise traffic poorly. As each column of the noise traffic matrix is assumed as a white noise vector, if the noise traffic is exactly

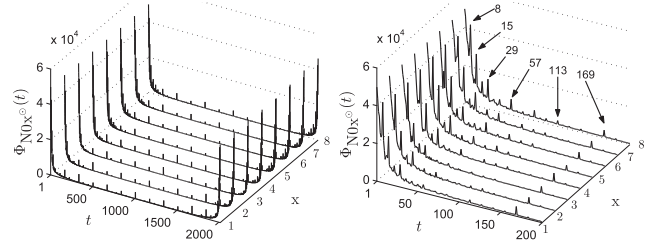


Fig. 1 The spectral density of $N01^\odot \sim N08^\odot$ (left); For each noise traffic matrix, the first 200 positions of its spectral density are specially magnified (right), which describe the low-frequency power.

recovered by SPCP, the spectral density should show a flat distribution. We briefly explain the cause of this limitation. In most cases, the volume of a traffic matrix is mainly contributed by the deterministic traffic. The deterministic traffic changes slowly and shows typical diurnal pattern. Therefore, the spectral density of each OD flow presents unbalanced distribution: The low-frequency spectra usually have much larger values, and the spectra whose periods are the factors of 24 hours show strong peaks. The optimization problem (2) can be written in this equivalent form:

$$\min_{A, E, N} \|A\|_* + \lambda \|E\|_1 + \gamma \|N\|_F^2 \quad s.t. A + E + N = X, \quad (7)$$

where $\gamma = 1/2\tau$, and the objective function for N is the Frobenius norm. Thus problem (7) has no frequency-domain consideration on the noise traffic. With high probability, the noise traffic decomposed by SPCP inherits the unbalanced spectral density distribution of the OD flow traffic.

These discussions motivate us to replace $\|N\|_F^2$ by a new objective function which focuses on N 's frequency-domain property, and the new decomposed noise traffic should have more flat spectral density. We define the frequency-domain weight matrix $C = \text{diag}(c_1, \dots, c_T)$ satisfying:

$$c_t > 0, \quad t = 1, 2, \dots, T; \quad \sum_{t=1}^T c_t^2 = T. \quad (8)$$

We propose the new traffic decomposition method called *Stable Principal Component Pursuit with Frequency-Domain Regularization* (SPCP-FDR). Compared with SPCP, SPCP-FDR adopts the same objective functions for the deterministic and the anomaly traffic, respectively; while $\|CW^T N\|_F$ is chosen as the new objective function for the noise traffic (C 's design is presented in Sect. 4):

$$\min_{A, E, N} \|A\|_* + \lambda \|E\|_1 + \gamma \|CW^T N\|_F^2 \quad s.t. A + E + N = X, \quad (9)$$

where λ and γ are positive parameters balancing the three objective functions. We choose $\lambda = 1/\sqrt{\max(T, P)}$ because it is demonstrated as a proper choice in [4]. For the choice of γ , consider the simplest case: each column in N is a standard Gaussian white noise. The DFT of one Gaussian white noise is also a Gaussian white noise, and it can be proved that $\mathbb{E}[\|CW^T N\|_F^2] = \mathbb{E}[\|N\|_F^2]$ (\mathbb{E} denotes expectation). Therefore, SPCP-FDR can be seen as an approximation of SPCP under the Gaussian white noise assumption. In

[2], we choose $\tau = \sqrt{2 \log(TP) \max(T, P)}$ for problem (2). Thus in this study $\gamma = 1/2\tau = 1/(2\sqrt{2 \log(TP) \max(T, P)})$.

4. Implementation Details

4.1 The Design of Frequency-Domain Weight Matrix

We design C for decomposing the Abilene weekly traffic matrices ($T = 2016$), and we believe that the key technologies are adaptable to other datasets with small modifications. Instinctively, it punishes the low-frequency spectra, especially for the spectra whose corresponding periods are the factors of 24 hours. Define the position set S_1 as:

$$S_1 = S_1^a \cup S_1^b; \\ S_1^a = \{8, 15, 29, 57, 113, 169\}; S_1^b = \{t \mid (T - t + 2) \in S_1^a\}.$$

S_1^b represents the dual positions of S_1^a . We design $\{c_t\}_{t=1}^T$ as:

$$c_t = \begin{cases} \beta[v(t) + \rho] & t \in S_1 \\ \beta v(t) & \text{otherwise} \end{cases}, \quad (10)$$

where $v(x)$ is a positive function in $[1, T]$ satisfying $v(x) = v(T - x + 2)$ in $[2, T]$. Meanwhile, it decreases monotonically in $[1, \frac{T}{2} + 1]$. $\rho > 0$ is an additional penal parameter for the spectra correlated to S_1 . $\beta > 0$ is a scaling parameter. In this study, we choose

$$v(x) = \begin{cases} 4e^{-\frac{(x-1)}{200}} + 1 & x \in [1, \frac{T}{2} + 1] \\ 4e^{-\frac{(T-x+1)}{200}} + 1 & x \in [\frac{T}{2} + 1, T] \end{cases} \quad (11)$$

and $\rho = 2$. At last, we get $\beta = 0.4832$ from assumption (8) for this special choice of $v(x)$ and ρ .

4.2 The APG Algorithm for SPCP-FDR

The *Accelerated Proximal Gradient* (APG) algorithm for SPCP-FDR solves a relaxed approximation problem of (9):

$$\min_{A, E, N} F(A, E, N) = \min_{A, E, N} \mu g(A, E, N) + f(A, E, N) \\ g(A, E, N) \triangleq \|A\|_* + \lambda \|E\|_1 + \gamma \|CW^T N\|_F^2, \quad (12) \\ f(A, E, N) \triangleq \frac{1}{2} \|A + E + N - X\|_F^2,$$

where $\mu > 0$ is a parameter. As $\mu \rightarrow 0$, the solution to (12) approaches to the solution to (9). Not directly minimizing $F(A, E, N)$, this algorithm minimizes a sequence of quadratic approximations $Q(A, E, N, Y^A, Y^E, Y^N)$ to $F(A, E, N)$ at point (Y^A, Y^E, Y^N) (renewed in each step):

$$Q(A, E, N, Y^A, Y^E, Y^N) \\ \triangleq \mu g(A, E, N) + f(Y^A, Y^E, Y^N) + \\ \langle \nabla f(Y^A, Y^E, Y^N), (A, E, N) - (Y^A, Y^E, Y^N) \rangle + \\ \frac{L_f}{2} \|(A, E, N) - (Y^A, Y^E, Y^N)\|_F^2, \quad (13)$$

where the Lipschitz constant $L_f = 3$. It can be derived that:

$$\min_{A, E, N} Q(A, E, N, Y^A, Y^E, Y^N) \\ = \min_A \frac{L_f}{2} \|A - G^A\|_F^2 + \mu \|A\|_* + \\ \min_E \frac{L_f}{2} \|E - G^E\|_F^2 + \mu \lambda \|E\|_1 + \\ \min_N \frac{L_f}{2} \|N - G^N\|_F^2 + \mu \gamma \|CW^T N\|_F^2 + \text{constant}, \quad (14)$$

where

$$G^\square = Y^\square - \frac{1}{L_f} (Y^A + Y^E + Y^N - X), \quad \square \in \{A, E, N\}. \quad (15)$$

Problem (14) splits into three independent optimization problems, and the first two (for A and E) are well studied [5]. We give the solution to the third problem by derivation:

$$N = L_f [L_f I_{T \times T} + 2\mu \gamma W C^2 W^T]^{-1} G^N. \quad (16)$$

Following the main idea in [5], we present the APG algorithm for the SPCP-FDR method in Algorithm 1. This algorithm has the $O(1/k^2)$ convergence rate. Because the proof is very close to Theorem 2.1 in [5] and Theorem 4.4 in [6], we omit it and directly summarize this result:

Theorem 1 Let $F(A, E, N) = \bar{\mu} g(A, E, N) + f(A, E, N)$. Then for all $k > k_0 \triangleq \lceil \log(\frac{\mu_0}{\bar{\mu}}) / \log(\frac{1}{\eta}) \rceil$, we have

$$F(X_k) - F(X^*) \leq 6 \|X_{k_0} - X^*\|_F^2 / (k - k_0 + 1)^2, \quad (17)$$

where $X_k = (A_k, E_k, N_k)$ is defined in Algorithm 1, and $X^* = (A^*, E^*, N^*)$ is a solution to (12) when $\mu = \bar{\mu}$. Notice that in this study $L_f = 3$, while in [5] it equals to 2.

Algorithm 1 The APG Algorithm for SPCP-FDR

Input: traffic matrix $X \in \mathbb{R}^{T \times P}$ with unit noise variance.

Initialization: $A_0 = A_{-1} = E_0 = E_{-1} = N_0 = N_{-1} = \mathbf{0}^{T \times P}$; $t_0 = t_{-1} = 1$; $\mu_0 = 0.99 \|X\|_2$; $\bar{\mu} = 10^{-5} \mu_0$; $k = 0$.

$\eta = 0.9$; $L_f = 3$; $\lambda = 1 / \sqrt{\max(T, P)}$; $\gamma = 1 / (2 \sqrt{2 \log(TP) \max(T, P)})$.

While not converged **do**

$Y_k^A = A_k + \frac{t_{k-1}-1}{t_k} (A_k - A_{k-1})$; $Y_k^E = E_k + \frac{t_{k-1}-1}{t_k} (E_k - E_{k-1})$;

$Y_k^N = N_k + \frac{t_{k-1}-1}{t_k} (N_k - N_{k-1})$;

$G_k^\square = Y_k^\square - \frac{1}{L_f} (Y_k^A + Y_k^E + Y_k^N - X)$, $\square \in \{A, E, N\}$;

$(U, S, V) = \text{svd}[G_k^A]$; $\text{//svd}[\cdot]$ denotes singular value decomposition.

$A_{k+1} = U S_{\frac{\mu_k}{L_f}} [S] V^T$; $E_{k+1} = S_{\frac{\mu_k}{L_f}} [G_k^E]$;

$\text{//} S_\epsilon[\cdot]$ denotes soft-thresholding operator with threshold $\epsilon > 0$.

$N_{k+1} = L_f [L_f I_{T \times T} + 2\mu_k \gamma W C^2 W^T]^{-1} G_k^N$;

$t_{k+1} = (1 + \sqrt{4t_k^2 + 1})/2$; $\mu_{k+1} = \max(\eta \mu_k, \bar{\mu})$; $k = k + 1$.

End while

Output: $A = A_k$; $E = E_k$; $N = E_k$.

5. Experiment Results

We evaluate the SPCP-FDR method by using the Abilene traffic matrices X01 ~ X08. For each $x \in \{1, \dots, 8\}$, suppose X0x is decomposed as $\{A_0 x^\oplus, E_0 x^\oplus, N_0 x^\oplus\}$ and

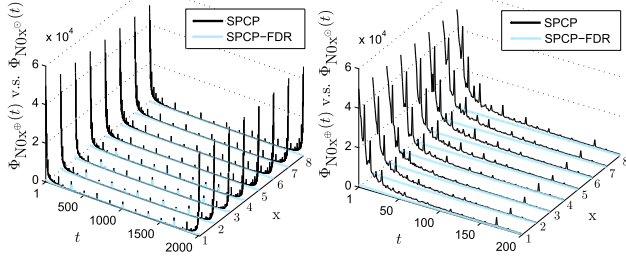


Fig. 2 The spectral density of $N01^{\oplus} \sim N08^{\oplus}$ decomposed by SPCP-FDR, compared with the spectral density of $N01^{\odot} \sim N08^{\odot}$ decomposed by SPCP (left); For each matrix, the first 200 positions of its spectral density are specially magnified (right), which describe the low-frequency power.

Table 1 A comparison between SPCP and SPCP-FDR on the decomposed deterministic traffic matrices.

Traffic matrix	rank($X0x$)	rank($A0x^{\odot}$)	rank($A0x^{\oplus}$)	$\frac{\ A0x^{\oplus}\ _F}{\ A0x^{\odot}\ _F}$
X01 ($x = 1$)	121	10	19	1.0158
X02 ($x = 2$)	121	11	22	1.0171
X03 ($x = 3$)	121	12	20	1.0283
X04 ($x = 4$)	121	11	18	1.0145
X05 ($x = 5$)	121	10	19	1.0148
X06 ($x = 6$)	121	10	22	1.0162
X07 ($x = 7$)	121	13	24	1.0210
X08 ($x = 8$)	121	12	23	1.0182

$\{A0x^{\odot}, E0x^{\odot}, N0x^{\odot}\}$ by SPCP-FDR and SPCP, respectively.

Figure 2 compares the spectral density of the noise traffic matrices $N01^{\oplus} \sim N08^{\oplus}$ (blue) and $N01^{\odot} \sim N08^{\odot}$ (black). For each $x \in \{1, \dots, 8\}$, $N0x^{\oplus}$ has more flat spectral density distribution than $N0x^{\odot}$. As we further regularize noise traffic's spectra whose positions lie in S_1 , their magnitudes dramatically decline. Therefore, for the new decompositions of $X01 \sim X08$ by SPCP-FDR, most low-frequency traffic pattern is efficiently eliminated from the resulting noise traffic matrices $N01^{\oplus} \sim N08^{\oplus}$. Table 1 compares the deterministic traffic matrices $A01^{\oplus} \sim A08^{\oplus}$ and $A01^{\odot} \sim A08^{\odot}$. The ranks of $A01^{\oplus} \sim A08^{\oplus}$ decomposed by SPCP-FDR are nearly twice as those decomposed by SPCP. However, as all these values do not exceed 24 and the original traffic matrices' ranks are 121, $A01^{\oplus} \sim A08^{\oplus}$ still satisfy the low-rank characteristic. For each $x \in \{1, \dots, 8\}$, the Frobenius norm of $A0x^{\oplus}$ is a little larger than that of $A0x^{\odot}$, and it can be explained that most low-frequency traffic pattern eliminated from the noise traffic is added to the deterministic traffic.

As a case study, we analyze the decomposition result of $X01$ in detail. For each OD flow $(X01)_j$ ($1 \leq j \leq 121$), compute the absolute Pearson correlation coefficient between $(A01^{\oplus})_j$ and $(N01^{\oplus})_j$, and the coefficient between $(A01^{\odot})_j$ and $(N01^{\odot})_j$. Figure 3 displays these coefficients arranged by flow ID. This is a proper metric of the cross-correlation and it should be as small as possible. For most OD flows in $X01$, compared with SPCP, SPCP-FDR's absolute Pearson correlation coefficients show significant decline. Actually, $X02 \sim X08$ show quite similar results. Thus SPCP-FDR could efficiently reduce this cross-correlation. Figure 4 compares SPCP-FDR with SPCP on decomposition result of the No. 121 OD flow, which produces the largest volume in $X01$. Instinctively, the noise traffic time-series $(N01^{\odot})_{121}$

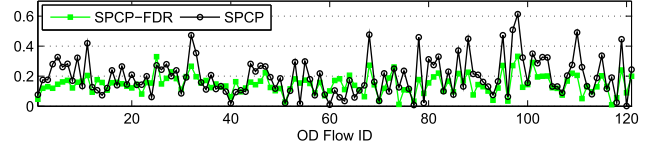


Fig. 3 A comparison between SPCP and SPCP-FDR on the absolute Pearson correlation coefficients (between the deterministic traffic time-series and the noise traffic time-series of each OD flow) of $X01$.

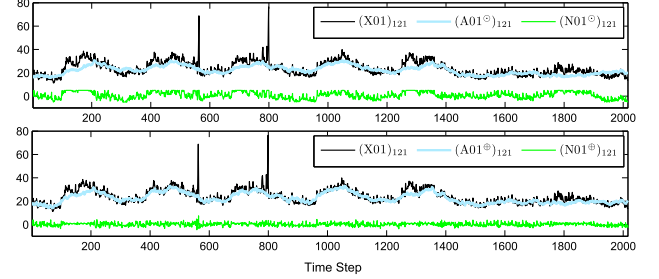


Fig. 4 A comparison between SPCP and SPCP-FDR on the decomposition result of OD flow $(X01)_{121}$. Upper panel: the deterministic and the noise traffic time-series decomposed by SPCP; Bottom panel: the deterministic and the noise traffic time-series decomposed by SPCP-FDR.

contains a distinct diurnal trend, which should be classified as the deterministic traffic better. By contrast, $(N01^{\oplus})_{121}$ presents a more stable temporal pattern. Thus SPCP-FDR gives a more rational decomposition for this flow.

6. Conclusions

This letter presents a novel traffic matrix decomposition method named SPCP-FDR, which improves SPCP by using frequency-domain regularization. We propose an APG algorithm for SPCP-FDR, its convergence rate is $O(1/k^2)$, and demonstrate it on a real-world dataset. SPCP-FDR efficiently eliminates the low-frequency periodic traffic from the noise traffic, maintains deterministic traffic's low-rank property, and shows lower cross-correlation between these two kinds of traffic than SPCP. Therefore, SPCP-FDR achieves more rational traffic decompositions.

References

- [1] A. Lakhina, K. Papagiannaki, M. Crovella, C. Diot, E.D. Kolaczyk, and N. Taft, "Structural analysis of network traffic flows," SIGMETRICS Perform. Eval. Rev., vol.32, no.1, pp.61–72, June 2004.
- [2] Z. Wang, K. Hu, K. Xu, B. Yin, and X. Dong, "Structural analysis of network traffic matrix via relaxed principal component pursuit," Comput. Netw., vol.56, no.7, pp.2049–2067, 2012.
- [3] Z. Zhou, X. Li, J. Wright, E. Candes, and Y. Ma, "Stable principal component pursuit," Proc. IEEE ISIT 2010, 2010.
- [4] E. Candes, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" J. ACM., vol.58, no.3, pp.1–37, 2011.
- [5] Z. Lin, A. Ganesh, J. Wright, L. Wu, M. Chen, and Y. Ma, "Fast convex optimization algorithms for exact recovery of a corrupted low-rank matrix," Proc. IEEE CAMSAP 2009, 2009.
- [6] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," SIAM J. Imaging Sciences, vol.2, no.1, pp.183–202, 2009.
- [7] Abilene dataset:
<http://www.cs.utexas.edu/~yzhang/research/AbileneTM/>