SUPPLEMENTARY MATERIAL FOR
"ONLINE MATRIX COMPLETION WITH GAUSSIAN MIXTURE MODEL"

Chunsheng Liu[1], *Chunlei Chen*[2], *Hong Shan*[1], *Bin Wang*[1]

[1] National University of Defense Technology, China. [2] Weifang University, China

ABSTRACT

In this supplementary material, we provide some additional experimental results to further demonstrate the superiority of our algorithm, compared with other state-of-the-art methods. Moreover, we also give the derivation of the proposed algorithm in detail. Additionally, we present the details of the real-world public traffic data sets. Finally, we illustrate the time complexity and memory complexity of our algorithm.

***Index Terms*—** additional experiments, inference details, time and memory complexity.

1. DETAILS FOR DYNAMIC TRAFFIC DATA SETS

We use the public dynamic traffic data sets[1] (Abilene and Geant) as the real-world data sets to demonstrate the superiority of our ONCE method. The data sets record the traffic matrix information of the network topology in a certain period of time. Abilene's topology consists of 12 nodes, and the duration of the collected traffic data is 10 days and the collecting interval is 5 minutes. Geant's topology consists of 22 nodes, and the duration of the collected traffic data is 10 days and the collecting interval is 15 minutes. We summarize the details of the two data sets in Table.1.

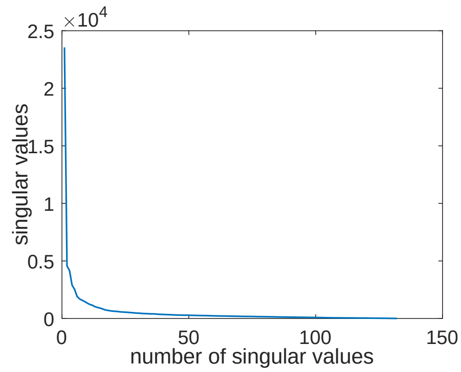
Table 1. Details of dynamic traffic data sets.

	Duration (days)	Collect Gap (min)	size
Abilene	10	5	132×2880
Geant	10	15	462×960

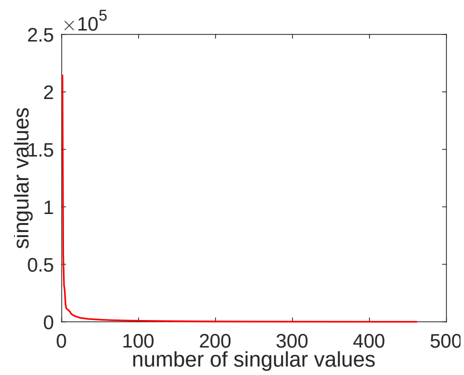
Moreover, Fig.1 shows the singular values of these two data sets. It can be seen that there are only a small number of large singular values that contain most of the information about the data. Therefore, we can approximate these data sets as low-rank matrices.

2. ADDITIONAL EXPERIMENTS

Fig.2 and Fig.3 present the comparison of MAEs for each method on synthetic and real traffic data sets, respectively.



(a) Abilene

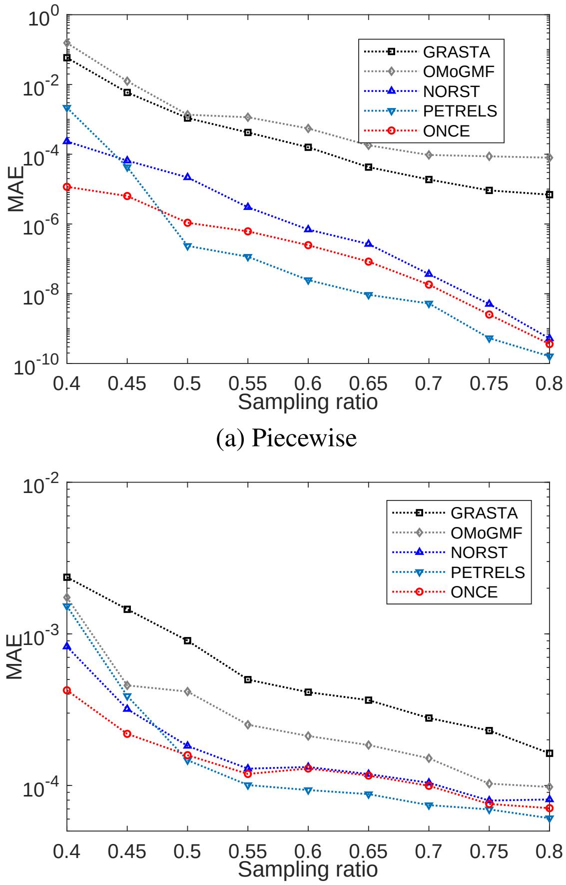


(b) Geant

Fig. 1. The singular values of the two dynamic traffic data sets. (a) Singular values of Abilene. (b) Singular values of Geant.

From Fig.2, it can be seen that MAE decreases monotonously with an increasing sampling ratio, which is reasonable – the larger the number of observations is, the more information of the underlying data can be obtained. Moreover, we can see that GRASTA and OMoGMF have poor performance compared to the other three algorithms, which is consistent with the situation of ER. In particular, when the sampling ratio is small (i.e., < 0.5), ONCE achieves highest performance while PETRELS performs poorly. Fig.3 shows the MAE versus number of samples on the traffic data sets. As can be seen, it is obvious that ONCE achieves the least mean MAE.

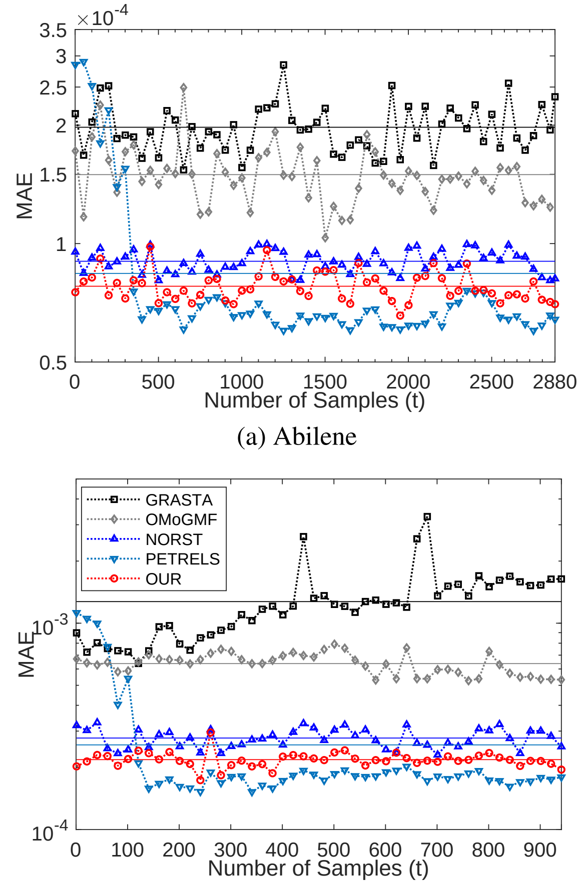
[1]http://sndlib.zib.de/home.action.



(a) Piecewise

(b) Changing

Fig. 2. Comparison of MAEs with different sampling ratio on two kinds of synthetic data sets. The sampling ratio varies successively from 0.4 to 0.8. (a) MAE versus sampling ratio on synthetic data with piecewise constant subspace change; (b) MAE versus sampling ratio on synthetic data with subspace changing at each time.



(a) Abilene

(b) Geant

Fig. 3. Comparison of MAEs on two traffic data sets. The sampling ratio is fixed at 0.5. (a) MAE versus number of samples on Abilene, (b) MAE versus number of samples on Geant, where the solid line indicates the mean MAE of each method.

3. INFERENCE DETAILS ON THE GMM PARAMETERS UPDATING FORMULATIONS

In this section, we present the inference details about the updating formulations for GMM parameters $\{\Pi_t, \Sigma_t\}$ in the M-step of ONCE.

According to Eq.(10) of the main text, we can get the following objective function:

$$\mathcal{L}_t(\Psi) = -\sum_{i=1}^{n}\sum_{k=1}^{K} z_t^{ik}\left(\ln \pi_t^k - \ln \sigma_t^k - \frac{\left(y_t^i - \mathbf{U}_t^i v_t\right)^2}{\left(\sigma_t^k\right)^2}\right) + \left(\sum_{k=1}^{K} N_{t-1}^k\left(\frac{{\sigma_{t-1}^k}^2}{2{\sigma_t^k}^2} + \ln\sigma_t^k\right) - N_{t-1}\sum_{k=1}^{K}\pi_{t-1}^k \ln \pi_t^k\right) + \mu\left|\sum_{k=1}^{K}\pi_t^k - 1\right|. \tag{1}$$

Then we can calculate the partial derivative of σ_t^k, and have:

$$\frac{\partial\mathcal{L}}{\partial\sigma_t^k} = \sum_{i=1}^{n} z_t^{ik}\left(\frac{1}{\sigma_t^k} - \frac{\left(y_t^i - \mathbf{U}_t^i v_t\right)^2}{2\left(\sigma_t^k\right)^3}\right) + N_{t-1}^k\left(\frac{1}{\sigma_t^k} - \frac{\left(\sigma_{t-1}^k\right)^2}{\left(\sigma_t^k\right)^3}\right). \tag{2}$$

Let Eq.(2) be equal to 0, and we can get

$$\left(\sigma_t^k\right)^2 = \frac{N_{t-1}^k\left(\sigma_{t-1}^k\right)^2 + \sum_{i=1}^{n} z_t^{ik}\left(y_t^i - \mathbf{U}_t^i v_t\right)^2}{N_{t-1}^k + \sum_{i=1}^{n} z_t^{ik}}. \tag{3}$$

Then we can calculate the partial derivative of π_k, and have:

$$\frac{\partial\mathcal{L}}{\partial\pi_t^k} = \sum_{i=1}^{n} z_t^{ik} + N_{t-1}^k - \mu\pi_t^k. \tag{4}$$

And let Eq.(4) be equal to 0, we can get

$$\sum_{k=1}^{K}\left(\sum_{i=1}^{n} z_t^{ik} + N_{t-1}^{k} - \mu\pi_t^k\right) = 0. \tag{5}$$

Incorporating $\sum_{k=1}^{K}\pi_t^k = 1$, we have:

$$\pi_t^k = \frac{\sum_{i=1}^{n} z_t^{ik} + N_{t-1}^{k}}{\sum_{k=1}^{K}\left(\sum_{i=1}^{n} z_t^{ik} + N_{t-1}^{k}\right)}. \tag{6}$$

where $N_{t-1}^{k} = \sum_{\tau=1}^{\mathrm{t}-1}\sum_{i=1}^{n} z_\tau^{ik}$.

Then we set

$$\begin{aligned}
&\bar{n} = n + \sum_{\tau=1}^{t-1}\sum_{i=1}^{n}\sum_{k=1}^{K} z_\tau^{ik}, n^k = \sum_{i=1}^{n} z_t^{ik},\\
&\bar{\pi}^k = n^k/n, \bar{n}^k = n^k + \sum_{\tau=1}^{t-1}\sum_{i=1}^{n} z_\tau^{ik},\\
&\left(\bar{\sigma}^k\right)^2 = \tfrac{1}{n^k}\sum_{i=1}^{n} z_t^{ik}\left(y_t^i - \mathbf{U}_t^i v_t\right)^2
\end{aligned} \tag{7}$$

Based on the Eq.(3), Eq.(6) can then be rewritten as the following forms for Π_t and Σ_t:

$$\pi_t^k = \pi_{t-1}^k - \frac{n}{\bar{n}}\left(\pi_{t-1}^k - \bar{\pi}^k\right). \tag{8}$$

$$\left(\sigma_t^k\right)^2 = \left(\sigma_{t-1}^k\right)^2 - \frac{n^k}{\bar{n}^k}\left(\left(\sigma_{t-1}^k\right)^2 - \left(\bar{\sigma}^k\right)^2\right). \tag{9}$$

From Eq.(8) and Eq.(9), we can see that they are the closed-form updating formulations for the GMM parameters.

4. PROOF OF THE COMPUTATIONAL FORMULA OF GAUSSIAN MIXTURE MODEL REGULARIZER

In this section, we present the proof of deduction formula of GMM regularizer in Definition 1. The derivation process of $\|x_t\|_{\mathrm{M}}$ is as follows:

$$\begin{aligned}
&\|x_t\|_{\mathrm{M}} = -\ln p\left(y_t\,|\Psi\right)\\
&= -\ln\left(\prod_{i=1}^{n}\prod_{k=1}^{K}\left(\pi_t^k\right)^{z_t^{ik}}\mathcal{N}\left(y_t^i\left|\mathcal{P}_{\Omega_t}\left(\mathbf{U}_t^i v_t\right), {\sigma_t^k}^2\right)^{z_t^{ik}}\right.\right)\\
&= -\sum_{i=1}^{n}\ln\left(\prod_{k=1}^{K}\left(\pi_t^k\right)^{z_t^{ik}}\mathcal{N}\left(y_t^i\left|\mathcal{P}_{\Omega_t}\left(\mathbf{U}_t^i v_t\right), {\sigma_t^k}^2\right)^{z_t^{ik}}\right.\right)\\
&= -\sum_{i=1}^{n}\sum_{k=1}^{K}\left(\ln\left(\pi_t^k\right)^{z_t^{ik}} + \ln\mathcal{N}\left(y_t^i\left|\mathcal{P}_{\Omega_t}\left(\mathbf{U}_t^i v_t\right), {\sigma_t^k}^2\right)^{z_t^{ik}}\right.\right).
\end{aligned} \tag{10}$$

Then for a Gaussian distribution $x \sim \mathcal{N}\left(\mu, \sigma^2\right)$, the probability density function of it can be formulated as $f(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$, and the Eq.(10) can be further deduced as following:

$$\begin{aligned}
&\|x_t\|_{\mathrm{M}} =\\
&-\sum_{i=1}^{n}\sum_{k=1}^{K} z_t^{ik}\left(\ln\left(\pi_t^k\right) + \ln\left(\frac{1}{\sqrt{2\pi}\sigma_t^k}\exp\left(-\frac{\left(y_t^i - \mathcal{P}_{\Omega_t}\left(\mathbf{U}_t^i v_t\right)\right)^2}{2\left(\sigma_t^k\right)^2}\right)\right)\right)\\
&= -\sum_{i=1}^{n}\sum_{k=1}^{K} z_t^{ik}\left(\ln\frac{\pi_t^k}{\sigma_t^k} - \frac{\left(y_t^i - \mathcal{P}_{\Omega_t}\left(\mathbf{U}_t^i v_t\right)\right)^2}{2\left(\sigma_t^k\right)^2}\right) + C\\
&= -\sum_{i=1}^{n}\sum_{k=1}^{K} z_t^{ik}\left(\ln\frac{\pi_t^k}{\sigma_t^k} - \frac{\left(y_t^i - \mathcal{P}_{\Omega_t}\left(\mathbf{U}_t^i v_t\right)\right)^2}{2\left(\sigma_t^k\right)^2}\right) + C\\
&= \sum_{i=1}^{n}\sum_{k=1}^{K} z_t^{ik}\ln\frac{\sigma_t^k}{\pi_t^k} + \sum_{i=1}^{n}\sum_{k=1}^{K} z_t^{ik}\left(\frac{\left(y_t^i - \mathcal{P}_{\Omega_t}\left(\mathbf{U}_t^i v_t\right)\right)^2}{2\left(\sigma_t^k\right)^2}\right) + C
\end{aligned} \tag{11}$$

Incorporating the equation $\sum_{k=1}^{K} z_t^{ik} = 1$, and Eq.(11) can be simplified as follows:

$$\begin{aligned}
&\|x_t\|_{\mathrm{M}} =\\
&\sum_{i=1}^{n}\sum_{k=1}^{K} z_t^{ik}\ln\frac{\sigma_t^k}{\pi_t^k} + n\sum_{k=1}^{K}\frac{1}{2\left(\sigma_t^k\right)^2}\left\|y_t - \mathcal{P}_{\Omega_t}\left(\mathbf{U}_t v_t\right)\right\|_2^2 + C
\end{aligned} \tag{12}$$

For the convenience of description, if we set $\alpha_\sigma = n\sum_{k=1}^{K}\frac{1}{2\left(\sigma_t^k\right)^2}$ and $\beta_{\pi,\sigma} = \sum_{i=1}^{n}\sum_{k=1}^{K} z_t^{ik}\ln\frac{\sigma_t^k}{\pi_t^k}$. Therefore, Eq.(12) can be rewritten as the following forms:

$$\|x_t\|_{\mathrm{M}} = \alpha_\sigma\left\|y_t - \mathcal{P}_{\Omega_t}\left(\mathbf{U}_t v_t\right)\right\|_2^2 + \beta_{\pi,\sigma} + C. \tag{13}$$

Incorporating the fact (Eq.(2) in main text) that $y_t = \mathcal{P}_{\Omega_t}\left(\mathbf{U}_t v_t\right) + x_t$, then we reformulate Eq.(13) as:

$$\|x_t\|_{\mathrm{M}} = \alpha_\sigma\|x_t\|_2^2 + \beta_{\pi,\sigma} + C. \tag{14}$$

where C is a constant.

The proof is then completed. ■

5. TIME AND MEMORY COMPLEXITY OF ONCE

In this section, we explain the time complexity and memory complexity of our proposed method.

Our work differs the work of [1] by modeling the noise term using GMM distribution, other than adopting a fixed and simple Laplacian distribution (the Laplacian distribution can be equivalently expressed as a scaled GMM [2]). Therefore, ONCE need update the GMM parameters before the noise data x_t updating, while NORST updates x_t directly. In particular, we use EM procedure to update the GMM parameters $\left(\pi_t^k, \sigma_t^k\right)$, and we assume that the required number of iterations is T, the time complexity of this part is $\mathcal{O}(nKT)$, where n is the dimension of data, and K is the number of components in GMM. From [1], the time complexity of NORST is $\mathcal{O}(nr\log 1/\varepsilon)$, where r is the rank of the underlying subspace, and ε is the recovery error. Consequently, the time complexity of ONCE is $\mathcal{O}(n(r\log 1/\varepsilon + KT))$. In addition, ONCE has the same memory complexity with NORST, i.e., $\mathcal{O}(nr\log n\log 1/\varepsilon)$.

6. REFERENCES

[1] Praneeth Narayanamurthy, Vahid Daneshpajooh, and Namrata Vaswani, "Provable subspace tracking from missing data and matrix completion," *IEEE Trans. Signal Processing*, vol. 67, no. 16, pp. 4245–4260, 2019.

[2] D. F. Andrews and C. L. Mallows, "Scale mixtures of normal distributions," *Journal of the Royal Statistical Society*, vol. 36, no. 1, pp. 99–102, 1974.