

COMPRESSED PREDICTION OF LARGE-SCALE URBAN TRAFFIC

Nikola Mitrovic¹, Muhammad Tayyab Asif¹, Justin Dauwels¹, Patrick Jaillet^{2,3}

¹School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore

²Laboratory for Information and Decision Systems, MIT, Cambridge, MA

³Center for Future Urban Mobility, Singapore-MIT Alliance for Research and Technology, Singapore

ABSTRACT

Traffic prediction lies at the core of many intelligent transport systems (ITS). Commonly deployed prediction methods such as support vector regression and neural networks achieve good performance by explicitly predicting the traffic variables (e.g., traffic speed or volume) at each road segment in the network. For large traffic networks, predicting traffic variable at each road segment may be unwieldy, especially in the setting of real-time prediction. To tackle this problem, we propose an alternative approach in this paper. We first generate low-dimensional representation of the network, leveraging on the column-based (CX) decomposition of matrices. The low-dimensional model represents the large network in terms of a small subset of road segments. The future state of the low-dimensional network is predicted by standard procedures, i.e., support vector regression. The future state of the entire network is then inferred by extrapolating the predictions of the subnetwork, using the CX decomposition. Numerical results for a large-scale road network in Singapore demonstrate the efficiency and accuracy of the proposed algorithm.

Index Terms— Prediction in large networks, low-dimensional models

1. INTRODUCTION

For intelligent transportation systems (ITS), the collection of real-time traffic data was the primary goal in the previous decade. Nowadays, traffic sensors such as loop detectors and probe vehicles collect data from thousands of road segments with high temporal resolution [1]. The availability of large amounts of data has enabled Data Driven Intelligent Transport Systems (D²ITS) in many applications such as traffic management, sensing, route guidance and congestion avoidance [2]. Most of these applications highly rely on fast and accurate traffic prediction. Consequently, real-time traffic prediction has been regarded as critical to the success of ITS, and it has become a foremost goal in the recent years [3–5].

The research described in this project was funded in part by the Singapore National Research Foundation (NRF) through the Singapore MIT Alliance for Research and Technology (SMART) Center for Future Mobility (FM).

Traffic forecasting studies mostly deal with time-series analysis and machine learning techniques [3, 5–7]. These techniques consider past and current states of the individual links (road segments) to predict their future states. To improve the prediction accuracy, some studies incorporate past and current values of neighboring links [8]. It has been shown that traffic network exhibits strong spatial correlations. Indeed, neighboring links may yield very similar predictions, and hence there is inherent redundancy. These relationships, between predicted states, have not yet been explicitly exploited to speed up traffic prediction [3, 7–11]. Instead, in all existing studies, the traffic is predicted at each link individually. Such an approach becomes unwieldy for large-scale traffic networks, especially for real-time applications.

In this paper, we overcome this issue by proposing compressed prediction of traffic. In principle, the method is applicable to any kind of multivariate time-series. However, we will limit ourselves to traffic prediction, as we have access to large real-life datasets for that particular application. Unlike the traditional forecasting models, we predict the speed only for a small subset of the links, and use mapping functions between the compressed state and the whole network to perform network wide prediction. This approach significantly reduces the computational cost, and allows real-time prediction for large traffic networks.

Concretely, we organize the traffic data in a large matrix, where columns of the matrix correspond to road segments (links). The rows represent the time intervals of the observed traffic variables. We apply the column-based (CX) matrix decomposition method to decompose the large data matrix in terms of a small number of columns of the original matrix [12, 13]. In other words, we compress the large traffic network in terms of a small subset of road segments in the network, which we will refer to as the “representative set”. The latter is assumed to be representative of the entire network, and is supposed to allow us to efficiently estimate the behavior of any other link in the network. By applying the CX method, we infer linear relationships between the representative set and the rest of the network. To this end, we represent the network matrix as product of two matrices: (1) data matrix of the representative subset of the road network

and (2) matrix of mapping functions between this subset and the entire network.

We predict future traffic state for the representative subset of road segments using standard regression methods, i.e., support vector regression [14]. Any other regression method could be applied for this purpose. Then, we extrapolate the predicted values at the representative subset to the entire network by applying the mapping functions, yielding predictions of the future traffic condition in the entire network.

To assess our approach, we predict traffic speed for roads in a large urban network in Singapore, comprising more than 6000 links. As comparison, we apply the support vector regression (SVR) method at each individual link. Our numerical results demonstrate that the proposed approach yields predictions that are almost as accurate as the standard approach of predicting each link individually. The computational complexity of the proposed approach is much lower, and scales approximately linearly with the size of the representative subset. In practice, this may speed up the computations by a factor of 10 or more, making the approach practical for real-time applications.

The paper is structured as follows. In section II, we briefly review the column based (CX) matrix decomposition, and explain how we apply it for compressed prediction. We also provide a brief analysis of the error in compressed prediction. In Section III, we explain the traffic data set under consideration. In Section IV, we provide numerical results, and compare the accuracy and computational complexity of the proposed and traditional approach. In Section V, we summarize our contributions in relation to prior work, and suggest topics for future work.

2. COMPRESSED PREDICTION

We consider the test network as a directed graph $G = (N, E)$, where $\{s_i \in E\}_{i=1}^p$ is the set of all the road segments in the network (see Fig.1). The vector $\mathbf{a}_i = [z(s_i, t_1) \dots z(s_i, t_d)]^T$ contains the speed profile for respective road s_i . We consider $z(s_i, t_j)$ as the average speed on that road during the interval $(t_j - \delta, t_j)$. Furthermore, the network profile $\mathbf{A} \in \mathbb{R}^{d \times p}$ for the whole network G is $\mathbf{A} = [\mathbf{a}_1 \dots \mathbf{a}_p]$. We use these definitions to propose the algorithm for compressed prediction in the next sections.

2.1. Column based (CX) decomposition

Let us start by considering that speed profiles $\{\mathbf{a}_1, \dots, \mathbf{a}_p\}$ are inter-related by certain global patterns in the network G . Transportation related studies have shown that global traffic patterns exist even in large and diverse networks [8, 15, 16]. Hence, we can find a suitable low-rank approximation $\hat{\mathbf{A}}$ for the network profile \mathbf{A} , such that:

$$\mathbf{A} \approx \hat{\mathbf{A}} = \mathbf{C}\mathbf{X}, \quad (1)$$

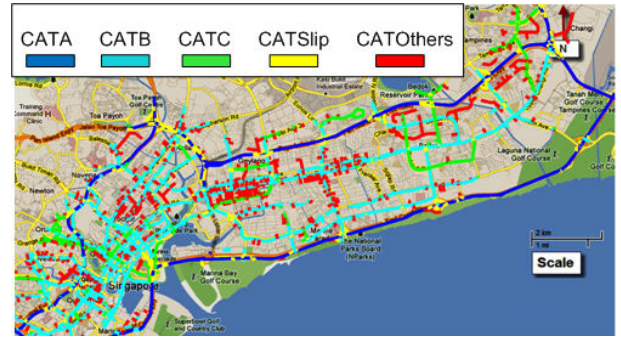


Fig. 1: The study area network.

where $\mathbf{C} \in \mathbb{R}^{d \times r}$ is the compressed network profile. The matrix $\mathbf{C} = [\mathbf{c}_1 \dots \mathbf{c}_r]$ is composed of the columns of network profile \mathbf{A} such that $\{\mathbf{c}_1, \dots, \mathbf{c}_r\} \subseteq \{\mathbf{a}_1, \dots, \mathbf{a}_p\}$. The matrix $\mathbf{X} \in \mathbb{R}^{r \times p}$ is the relationship matrix. The low-dimensional representation shown in (1) is often termed as CX decomposition [12, 17]. To obtain \mathbf{C} , we need to select an appropriate subset of roads $\Theta = \{\theta_1 \dots \theta_r\} : \Theta \subset E$. We can then estimate \mathbf{X} as:

$$\mathbf{X} = (\mathbf{C}^T \mathbf{C})^{-1} \mathbf{C}^T \mathbf{A}, \quad (2)$$

where $\mathbf{C}^+ = (\mathbf{C}^T \mathbf{C})^{-1} \mathbf{C}^T$ can be considered as the Moore-Penrose pseudo-inverse of matrix \mathbf{C} [18]. We utilize historical data to learn the relationship matrix \mathbf{X} .

Let us now explain the construction of \mathbf{C} . To do so, each road $\{s_i\}_{i=1}^p$ in the network G (or column $\{\mathbf{a}_i\}_{i=1}^p$ of \mathbf{A}) is assigned a selection probability $P(s_i)$. We then perform random sampling to obtain the appropriate subset of columns $(\{\mathbf{c}_i\}_{i=1}^r)$ to construct \mathbf{C} . We considered three strategies to obtain $\{P(s_i)\}_{s_i \in E}$. In the first strategy, each link has the same probability of selection i.e. $P_u(s_i) = \frac{1}{p}$.

In the second case, we assign selection probability to the link s_i , based on the energy of the corresponding column \mathbf{a}_i . The probability $P_e(s_i)$ is calculated as:

$$P_e(s_i) = \frac{\|\mathbf{a}_i\|_2^2}{\|\mathbf{A}\|_F^2} \quad \forall i = 1, \dots, p, \quad (3)$$

where $\|\mathbf{A}\|_F$ is the Frobenius norm of the matrix \mathbf{A} .

In the third strategy, we first perform Singular Value Decomposition (SVD) of the network profile $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$, where the matrix $\mathbf{U} \in \mathbb{R}^{d \times r}$ contains the left singular vectors and the matrix $\mathbf{V} \in \mathbb{R}^{p \times r}$ contains the right singular vectors. Then, we assign selection probabilities as follows [12]:

$$P_s(s_i) = \frac{1}{r} \sum_{j=1}^r v_{ij}^2 \quad \forall i = 1, \dots, p, \quad (4)$$

where v_{ij} is i -th coordinate of j -th right singular vector. Since the matrix \mathbf{V} is unitary, the quantities $P_s(s_i)$ sum to one.

2.2. Prediction

In the previous section, we developed a low-dimensional representation to represent a large network G . Suppose that we select a representative subset of roads $\Theta = \{\theta_1 \dots \theta_r\} : \Theta \subset E$ and learn the relationship \mathbf{X} between these roads and the rest of the network. We train SVR predictors for each of these representative links θ_i , for each prediction horizon. Let us represent the compressed predicted state at time t_j for k^{th} prediction horizon by the vector $\{\mathbf{w}_\Theta(t_j + k\delta) \in \mathbb{R}^{r \times 1}\}_{j=1}^f$, such that $\mathbf{w}_\Theta(t_j + k\delta) = [\hat{z}(\theta_1, t_j + k\delta), \dots, \hat{z}(\theta_r, t_j + k\delta)]^T$. The term $\hat{z}(\theta_i, t_j + k\delta)$ represents the k -step ahead predicted value for road θ_i at time t_j , obtained by SVR. We can estimate the future state of the entire network $\{\mathbf{w}_E(t_j + k\delta) \in \mathbb{R}^{p \times 1}\}_{j=1}^f$ as:

$$\mathbf{w}_E(t_j + k\delta) = \mathbf{X}^T \mathbf{w}_\Theta(t_j + k\delta), \quad (5)$$

where $\mathbf{w}_E(t_j + k\delta) = [\hat{z}(s_1, t_j + k\delta), \dots, \hat{z}(s_p, t_j + k\delta)]^T$. We represent the time instances for which we performed prediction by the set $\{t_j\}_{j=1}^f$. The resulting compressed and network profiles would be $\mathbf{C}_{pred}^T = [\mathbf{w}_\Theta(t_1 + k\delta) \dots \mathbf{w}_\Theta(t_f + k\delta)]$ and $\hat{\mathbf{A}}_{pred}^T = [\mathbf{w}_E(t_1 + k\delta) \dots \mathbf{w}_E(t_f + k\delta)]$ respectively, such that $\hat{\mathbf{A}}_{pred} = \mathbf{C}_{pred} \mathbf{X}$.

2.3. Compressed prediction error

The overall performance of the proposed algorithm is sensitive to errors due to prediction (SVR predictors) as well as estimation (CX-approximation). Suppose $\hat{\mathbf{A}} = \mathbf{C}_{pred} \mathbf{X}$ is the compressed prediction for the network profile \mathbf{A} . Let us represent compressed profile as $\mathbf{C}_{pred} = \mathbf{C}_0 + \Delta\mathbf{C}$, where \mathbf{C}_0 contains the actual speed values and $\Delta\mathbf{C}$ represents the variation due to prediction error. Furthermore, let $\mathbf{H} = \mathbf{C}_0 \mathbf{X}$ be the estimated network profile, without any prediction error in \mathbf{C} . The mean squared error (MSE) can be decomposed as:

$$\frac{1}{pf} \|\mathbf{A} - \hat{\mathbf{A}}\|_F^2 = \frac{1}{pf} \|(\mathbf{A} - \mathbf{C}_0 \mathbf{X}) - (\mathbf{C}_{pred} \mathbf{X} - \mathbf{C}_0 \mathbf{X})\|_F^2, \quad (6)$$

$$= \frac{1}{pf} \|(\mathbf{A} - \mathbf{H}) - (\hat{\mathbf{A}} - \mathbf{H})\|_F^2, \quad (7)$$

$$= \frac{1}{pf} \left(\sum_{i=1}^f \sum_{j=1}^p (a_{ij} - h_{ij})^2 + \sum_{i=1}^f \sum_{j=1}^p (\hat{a}_{ij} - h_{ij})^2 - 2 \sum_{i=1}^f \sum_{j=1}^p (a_{ij} - h_{ij})(\hat{a}_{ij} - h_{ij}) \right). \quad (8)$$

We refer to the error factors in (8) as:

$$MSE_{total} = MSE_{cx} + MSE_{svr} - 2\xi_{cor}. \quad (9)$$

We will analyze the behavior of these errors for different compression ratios (total number of links divided by the number of representative links) in Section 4.

3. EXPERIMENTAL SETUP

We deploy the proposed method in large traffic network in Singapore (see Fig. 1), comprising of $n = 6024$ links. The selected area contains different types of roads. The Land Transportation Authority (LTA) in Singapore provided speed data for the months of August, September and October 2011, with sampling interval of 5 minutes. The reported speed represents the average speed of all vehicles which traverse a link during the given sampling interval. There is 3% of missing data in the data set at hand due to sensor's malfunctions and other reasons. We impute missing data values by following the procedure in [19]. We represent the data set as a matrix, where each column represents the average speed data from a particular road segment (link) during the three months of data. Each row of matrix corresponds to a 5-minute time interval. The data set is divided into training and testing subsets. Training part involves first two months of data. Testing subset contains speed data for October 2011.

First, we apply column based (CX) decomposition method on the training data for two purposes: (1) to extract most suitable roads in the network and (2) to learn mapping functions between selected roads and the rest of the network. We test different selection methods (random, energy-based, and SVD-based) in order to choose the best one for further implementation. Then, we predict the speed values for the subset of the selected links, using SVR. We learn SVR predictors using two months of training set and further apply it for October data. For the rest of the network, future state is estimated using learned mapping functions. Consequently, prediction of the entire network is provided. As comparison, we follow the traditional approach which explicitly predicts traffic speed for each link in the network using SVR technique. We refer to it as the uncompressed method. We perform experiments for different sizes of selected subset and various prediction horizons. To evaluate performances of the proposed method, we consider percent root mean distortion (PRD) error [20].

4. RESULTS

In this section, we provide results for the proposed (compressed) and traditional (uncompressed) prediction methods. First, we investigate how to select representative links in the network. We use two months of data (August and September 2011) to test the performance of three selection methods (random, energy-based, and SVD-based). We select a representative subset of the roads and learn relationships between this subset and the rest of the network. The speed at each link in the network is approximated using the selected subset of roads according to the CX approximation (1). Fig. 2 shows the performance of this approximation for the three methods for selecting representative links. As can be seen from that figure, the SVD-based method generates the best

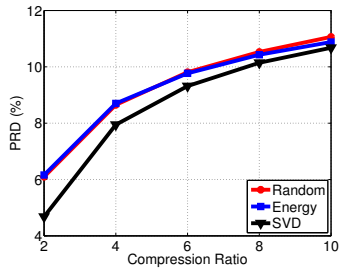


Fig. 2: Performance of CX method for different sampling procedures.

	2	4	6	8	10
MSE_{SVR}	29.23	25.66	22.47	19.84	18.20
MSE_{CX}	11.58	26.80	33.26	37.82	41.42
ξ_{cor}	1.75	2.98	2.94	2.79	2.89
MSE_{Total}	37.32	46.50	49.85	52.08	53.84

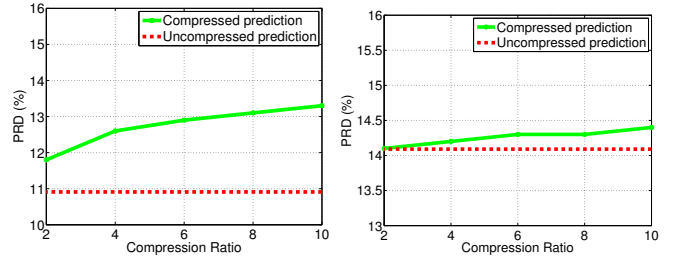
Table 1: The MSE error of proposed method for different compression ratios.

prediction performance (lowest PRD) for all compression ratios (total number of links divided by the number of representative links). This is in agreement with Theorem 1 in [12] which propose SVD method for column selection. Consequently, we will adopt this selection method for the proposed compressed prediction approach.

The prediction error of the proposed method has two components: the error induced by the (imperfect) SVR predictions, and the approximation associated with the column based (CX) decomposition. Table 1 shows the mean squared error (MSE) associated with those two error components, the correlation between those two errors, and the total MSE, for different compression ratios. From the table it can be seen that there is small correlation between the two error components. By increasing the compression ratio, the error associated with the CX approximation increases. In other words, leveraging on smaller subset of representative links will induce larger estimation error, as expected.

Fig. 3 shows the percent root mean distortion (PRD) error of the proposed and the traditional methods for 5 and 30 minutes prediction horizons. These results suggest that the proposed CX-based method has (slightly) larger prediction error than the standard (uncompressed) approach. The increased prediction error is caused by the low-dimensional network representation, where the predicted state for most of the links is computed as a linear combination of predictions for the small set of representative links. Intuitively, this additional error will decrease if the set of representative links is increased. Consequently, the additional prediction error associated with the compressed prediction approach increases gradually with the compression ratio (see Fig. 3).

The main contribution of proposed method is the substantial reduction in computational complexity, making it useful for real-time traffic prediction. Table 2 shows the required computation time (in seconds) for predicting the speed for the entire network, for the compressed and traditional SVR method. We assume that training for



(a) 5 minutes prediction horizon.

(b) 30 minutes prediction horizon.

Fig. 3: Performances of proposed and traditional method.

Compression Ratio (CR)	2	4	6	8	10
Traditional Method (SVR)	206.02	206.02	206.02	206.02	206.02
SVR	103.01	51.51	34.34	25.75	20.60
Matrix multiplication	0.0103	0.0076	0.0071	0.0060	0.0049
Proposed Method	103.02	51.51	34.34	25.76	20.61
Savings	50%	75%	83%	87%	90%

Table 2: Comparison of computation time (in seconds) for proposed and traditional method.

both methods is performed offline, which is a reasonable assumption for practical scenarios. The computation time for the traditional method is the required time for SVR prediction at one link times the total number of links in the network. In the proposed compressed method, two computations are required: (1) predicting the speed for the subset of representative links by means of SVR; (2) extrapolation of the predictions to the entire network, which involves a simple matrix multiplication. As can be seen from Table 2, the computation time for the latter is negligible. Consequently, the computation time for the compressed prediction approach is proportional to the number of representative links, and the speedup is proportional to the compression ratio. Therefore, substantial reduction in the computational complexity can be obtained, at the expense of only minor reduction in prediction accuracy.

5. CONCLUSIONS AND FUTURE WORK

In this paper, we applied the column based (CX) matrix decomposition method for fast yet accurate prediction of large scale traffic networks. We predict the traffic speed at a subset of carefully chosen representative road segments, and disseminate the predictions through the rest of the network. The proposed method significantly reduces the computational cost (at the expense of slightly increased prediction error), and may be useful for real-time applications.

In the future, we will incorporate additional information (e.g., weather data) to boost the prediction performance of low-dimensional network models. Moreover, to enhance the robustness of the proposed method, we will investigate irregular conditions due to, e.g., accidents and special events.

6. REFERENCES

- [1] J. Aslam, S. Lim, X. Pan, and D. Rus, "City-scale traffic estimation from a roving sensor network," in *Proceedings of the 10th ACM Conference on Embedded Network Sensor Systems*. ACM, 2012, pp. 141–154.
- [2] J. Zhang, F. Wang, K. Wang, W. Lin, X. Xu, and C. Chen, "Data-driven intelligent transportation systems: A survey," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 4, pp. 1624–1639, 2011.
- [3] M. Lippi, M. Bertini, and P. Frasconi, "Short-term traffic flow forecasting: An experimental comparison of time-series analysis and supervised learning," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 14, no. 2, pp. 871–882, 2013.
- [4] Y.-J. Wu, F. Chen, C. Lu, B. Smith, and Y. Chen, "Traffic flow prediction for urban network using spatio-temporal random effects model," in *91st Annual Meeting of the Transportation Research Board (TRB)*, 2012.
- [5] M. T. Asif, J. Dauwels, C. Y. Goh, A. Oran, E. Fathi, M. Xu, M. M. Dhanya, N. Mitrovic, and P. Jaillet, "Unsupervised learning based performance analysis of n-support vector regression for speed prediction of a large road network," in *Intelligent Transportation Systems (ITSC), 2012 15th International IEEE Conference on*, sept. 2012, pp. 983–988.
- [6] C. van Hinsbergen, J. Van Lint, and H. Van Zuylen, "Bayesian committee of neural networks to predict travel times with confidence intervals," *Transportation Research Part C: Emerging Technologies*, vol. 17, no. 5, pp. 498–509, 2009.
- [7] C. Quek, M. Pasquier, and B. Lim, "Pop-traffic: A novel fuzzy neural approach to road traffic analysis and prediction," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 7, no. 2, pp. 133–146, 2006.
- [8] W. Min and L. Wynter, "Real-time road traffic prediction with spatio-temporal correlations," *Transportation Research Part C: Emerging Technologies*, vol. 19, no. 4, pp. 606–616, 2011.
- [9] J. Park, D. Li, Y. Murphey, J. Kristinsson, R. McGee, M. Kuang, and T. Phillips, "Real time vehicle speed prediction using a neural network traffic model," in *Neural Networks (IJCNN), The 2011 International Joint Conference on*. IEEE, 2011, pp. 2991–2996.
- [10] C. Wu, J. Ho, and D. Lee, "Travel-time prediction with support vector regression," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 5, no. 4, pp. 276–281, 2004.
- [11] E. Schmitt and H. Jula, "Vehicle route guidance systems: Classification and comparison," in *Intelligent Transportation Systems Conference, 2006. ITSC'06. IEEE*. IEEE, 2006, pp. 242–247.
- [12] P. Drineas, M. W. Mahoney, and S. Muthukrishnan, "Relative-error cur matrix decompositions," *SIAM Journal on Matrix Analysis and Applications*, vol. 30, no. 2, pp. 844–881, 2008.
- [13] N. Mitrovic, M. T. Asif, U. Rasheed, J. Dauwels, and P. Jaillet, "CUR decomposition for compression and compressed sensing of large-scale traffic data," in *Intelligent Transportation Systems (ITSC), 2013 16th International IEEE Conference on*, oct. 2013, pp. 1475–1480.
- [14] A. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statistics and computing*, vol. 14, no. 3, pp. 199–222, 2004.
- [15] M. T. Asif, S. Kannan, J. Dauwels, and P. Jaillet, "Data compression techniques for urban traffic data," in *Computational Intelligence in Vehicles and Transportation Systems (CIVTS), 2013 IEEE Symposium on*, 2013, pp. 44–49.
- [16] T. Djukic, J. W. van Lint, and S. Hoogendoorn, "Application of principal component analysis to predict dynamic origin-destination matrices," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2283, no. 1, pp. 81–89, 2012.
- [17] M. W. Mahoney and P. Drineas, "CUR matrix decompositions for improved data analysis," *Proceedings of the National Academy of Sciences*, vol. 106, no. 3, pp. 697–702, 2009.
- [18] G. Golub and W. Kahan, "Calculating the singular values and pseudo-inverse of a matrix," *Journal of the Society for Industrial & Applied Mathematics, Series B: Numerical Analysis*, vol. 2, no. 2, pp. 205–224, 1965.
- [19] M. T. Asif, N. Mitrovic, L. Garg, J. Dauwels, and P. Jaillet, "Low-dimensional models for missing data imputation in road networks," in *38th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2013.
- [20] J. Dauwels, S. Kannan, R. Ramasubba, and A. Cichocki, "Near-lossless multi-channel EEG compression based on matrix and tensor decompositions," *Biomedical and Health Informatics, IEEE Journal of*, vol. PP, no. 99, p. 1, 2012.