# Low-Dimensional Models for Compressed Sensing and Prediction of Large-Scale Traffic Data

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

<table>
<thead>
<tr>
<th><strong>Citation</strong></th>
<th>Mitrovic, Nikola, Muhammad Tayyab Asif, Justin Dauwels, and Patrick Jaillet. “Low-Dimensional Models for Compressed Sensing and Prediction of Large-Scale Traffic Data.” IEEE Transactions on Intelligent Transportation Systems 16, no. 5 (October 2015): 2949–2954.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>As Published</strong></td>
<td><a href="http://dx.doi.org/10.1109/tits.2015.2411675">http://dx.doi.org/10.1109/tits.2015.2411675</a></td>
</tr>
<tr>
<td><strong>Publisher</strong></td>
<td>Institute of Electrical and Electronics Engineers (IEEE)</td>
</tr>
<tr>
<td><strong>Version</strong></td>
<td>Author's final manuscript</td>
</tr>
<tr>
<td><strong>Accessed</strong></td>
<td>Fri Feb 08 08:22:01 EST 2019</td>
</tr>
<tr>
<td><strong>Citable Link</strong></td>
<td><a href="http://hdl.handle.net/1721.1/100719">http://hdl.handle.net/1721.1/100719</a></td>
</tr>
<tr>
<td><strong>Terms of Use</strong></td>
<td>Creative Commons Attribution-Noncommercial-Share Alike</td>
</tr>
<tr>
<td><strong>Detailed Terms</strong></td>
<td><a href="http://creativecommons.org/licenses/by-nc-sa/4.0/">http://creativecommons.org/licenses/by-nc-sa/4.0/</a></td>
</tr>
</tbody>
</table>
Low-dimensional Models for Compressed Sensing and Prediction of Large-Scale Traffic Data

Nikola Mitrovic, Student Member, IEEE, Muhammad Tayyab Asif, Student Member, IEEE, Justin Dauwels, Senior Member, IEEE and Patrick Jaillet

Abstract—Advanced sensing and surveillance technologies often collect traffic information with high temporal and spatial resolutions. The volume of the collected data severely limits the scalability of online traffic operations. To overcome this issue, we propose a low-dimensional network representation where only a subset of road segments is explicitly monitored. Traffic information for the subset of roads is then used to estimate and predict conditions of the entire network. Numerical results show that such approach provides 10 times faster prediction at a loss of performance of 3% and 1% for 5 and 30 minutes prediction horizons, respectively.

Index Terms—Low-dimensional models, traffic prediction.

I. INTRODUCTION

Intelligent Transportation Systems (ITS) collect real-time traffic information from various sources such as probe vehicles, smartphone devices and infrastructure based traffic sensors. With advancements in sensor technology, traffic data (e.g., volume and speed) can be recorded on a large scale and with high temporal resolution. Recorded data is frequently used for historical analysis and traffic management operations such as network monitoring, transportation planning and congestion avoidance applications [1]. These applications heavily rely on fast and accurate assessment of current (estimation) and future (prediction) network states.

To model the road network, existing studies explicitly address each road segment in that network. For large traffic networks and online applications such an approach may not be feasible. To overcome this problem we focus on low-dimensional network models where only a subset of road segments needs to be explicitly monitored.

In this study, we use column-based (CX) matrix decomposition to express the original network in terms of a small subnetwork. We refer to the small subnetwork as the compressed state of the original network. We learn the relationship between compressed and original (uncompressed) network by analyzing the recorded data in offline manner. In this way, we can represent the traffic network as a product of two low-rank matrices: (i) the subnetwork data and (ii) the corresponding relationship matrix. We refer to this as CX compression scheme. The CX compression scheme is a stepping stone to compressed sensing and compressed prediction applications.

In the case of compressed sensing we aim to infer the present state of the entire network from the current traffic state of the subnetwork [2]. We use the training data (collected offline) to infer the relationship matrix. To assess the network state we multiply the data from the subnetwork, obtained from the testing set, with the relationship matrix, inferred from the training data set. Our underlying assumption is that traffic variables often vary smoothly across the traffic network [3].

In the matter of compressed prediction we apply the CX-based method to infer the future state of the network. First, we explicitly predict traffic state for the subnetwork using traditional prediction algorithms. Then we multiply the predicted data of the subnetwork with the relationship matrix, inferred from the training data set. Similarly to compressed sensing we rely on the observation that traffic conditions tend to follow distinct patterns and traffic parameters often vary smoothly [3], [4].

For our analysis, we consider the city-scale traffic network in Singapore, comprising 17,967 road segments. The numerical results show that the proposed methods can infer the current and future states of the network, while substantially improving the processing speed of the underlying modeling algorithm. The reduction in computational time is proportional to the compression ratio, i.e., the ratio of the number of links in the subnetwork and the total number of links.

The paper is structured as follows. In Section II we briefly review relevant literature. In Section III we introduce the column based (CX) matrix decomposition method. In Section IV we present three applications of the CX matrix decomposition methods in the realm of traffic modeling: compression, compressed sensing, and compressed prediction. In Section V we describe the traffic data set analyzed in this paper. In Section VI we provide and discuss results for our experiments. In Section VII we summarize our contributions and suggest topics for future research.

II. RELATED WORK

In this paper we propose a novel low-dimensional network model to improve the scalability of estimation and prediction operations in ITS. Low-dimensional representation of large traffic data sets is traditionally obtained by Principal Component Analysis (PCA) [3], [5], [6]. PCA provides an
effective low-dimensional representation in terms of latent variables and corresponding basis vectors. However, these latent variables are hard to interpret. Moreover, even if we obtain the basis vectors from historical data, we still need to collect data from all sensors during online operation. Due to this reason, PCA is often used for offline operations such as compression and data preprocessing [6].

B. Compressed sensing

So far, we have assumed that the matrix $X_h$ is stored together with $C_h$ leading to the compression of matrix $A_p$. In this scenario, the matrix $X_h$ is computed for a given data matrix $A_p$ and a column matrix $C_h$. Alternatively, one may precompute a matrix $X_h$ and re-use the same matrix to infer $A$ for any given $C$. Although we still need data from all the links to precompute $X_h$, this operation can easily be performed offline. Hence, during online operations, the system would only require data from a small number of sensors. We refer to this scenario as compressed sensing. It is noteworthy that low-dimensional PCA models can not be used for compressed sensing since PCA requires data from all sensors for both offline and online operations.

The underlying assumption of the proposed method is that the traffic conditions are stationary, so that a fixed matrix $X$ allows us to accurately reconstruct the original data matrix $A$ from $C$ [3], [4]. Therefore, we can estimate the present network state $(\hat{\alpha}_p^i)$ as $\hat{\alpha}_p^i = c_{p,i}^s X_h \forall i = k, \ldots, m$, where $\hat{\alpha}_p^i \in \mathbb{R}^{1 \times n_p}$ is a row vector which represents the current state of the entire network for test data $(i = k, \ldots, m)$. Row vector $c_{p,i}^s \in \mathbb{R}^{1 \times c}$ contains the information about current traffic conditions at $c$ specific locations in the network. Matrix $X_h$ is the relationship matrix, learned from a training data set. We define the compression ratio for compressed sensing as $\rho$.

Large traffic networks contain a diverse set of road segments. We want to explore whether homogeneous subnetworks can improve the overall performances of compressed sensing. We divide the traffic network into $s$ mutually exclusive subnetworks such that $\alpha_p = [\alpha_1 \ldots \alpha_s]$ where $\alpha_i \in \mathbb{R}^{1 \times n_i} \forall i = 1, \ldots, s$. Then, we perform compressed sensing for each subnetwork separately. At last, we merge the results of the clustered subnetworks to infer the traffic state of the entire network. Although different choices of temporal and/or spatial clustering can be applied, we consider simple clustering based on different road categories in this study.

The overall performance of the proposed compressed sensing method is sensitive to the “compressibility” of the network and “non-stationarity” in the traffic data. For compression, we represent the traffic data as a product of two low-rank matrices, i.e., the subnetwork data matrix and the most appropriate relationship matrix. As the compression is lossy, we expect the reconstructed matrix $\hat{A}_p$ to be different from the original matrix $A_p$. The issue of non-stationarity is due to the fact that matrix $X_h$ is inferred from training (historical) data instead of the current data. The matrix $A_p = [\alpha_1^k \ldots \alpha_s^k]^T$ is not available, and the goal is to infer that matrix by extrapolating the matrix $C_p = [c_{p,1}^k \ldots c_{p,m}^k]^T$ according to the CX decomposition. Obviously, the matrix $X_p$ cannot be extracted from the current data $A_p$, since the matrix $A_p$ is not available. Instead we determine $X_h$ from training data set. Since traffic is not perfectly stationary, this approximation will induce an additional reconstruction error. We refer to it as the error due to non-stationarity of traffic spatial relationships. To quantify this error, let us call $B = C_p X_h$ the reconstruction of the data matrix $A_p$, assuming the latter is available to compute the CX decomposition. The reconstruction $\hat{A}_p = [\hat{\alpha}_1^k \ldots \hat{\alpha}_s^k]^T$ in the scenario of compressed sensing is less accurate, since we need to replace $X_p$ (determined from the test data matrix $A_p$) by $X_h$ (determined from training data matrix $A_h$). The mean squared error (MSE) incurred for compressed sensing can be written as:

$$\frac{1}{rn} \|A_p - \hat{A}_p\|^2_F = \frac{1}{rn} \|A_p - (C_p X_h - C_p X_h)\|^2_F,$$

$$= \frac{1}{rn} \|A_p - B - (\hat{A}_p - B)\|^2_F,$$

$$= \frac{1}{rn} \left( \sum_{i=1}^{m} \sum_{j=1}^{n} (a_{ij} - b_{ij})^2 + \sum_{i=1}^{m} \sum_{j=1}^{n} (\hat{a}_{ij} - b_{ij})^2 
- 2 \sum_{i=1}^{m} \sum_{j=1}^{n} (a_{ij} - b_{ij})(\hat{a}_{ij} - b_{ij}) \right),$$

where $r = (m - k + 1)$ represents the number of time instances in test data matrix $A_p$. The first component of the error corresponds to the compressibility of the network and the second component is due to the non-stationarity of spatial patterns within the network (see (5)). The third component of the error refers to the correlations between aforementioned error components (see (5)). To make this interpretation more explicit, we rewrite (5) as:

$$\text{MSE}_{\text{est}} = \text{MSE}_{\text{com}} + \text{MSE}_{\text{ns}} - 2\xi_{\text{est}},$$

where $\xi_{\text{est}}$ is correlation coefficient between compressibility and non-stationarity. We will analyze the behavior of these errors for different compression ratios in Section VI.

C. Compressed prediction

In the previous section, we inferred the condition of the entire traffic network by observing traffic conditions at a small subset of links. Here we will extend this approach to prediction; we aim to predict the state of the entire traffic network from the predicted state of a small subset of links. We recall that low-dimensional models generated by PCA can not be utilized for this task since PCA requires information for all links in the network. Instead, we use state-of-the-art algorithm to predict the traffic speed only for a selected subset of locations. Then, we utilize the proposed method to extrapolate the predictions to the rest of the network using the precomputed relationship matrix. This can be written as $\hat{\alpha}_f^i = \hat{c}_f^s X_h \forall i = k, \ldots, m$, where $\hat{c}_f^s \in \mathbb{R}^{1 \times c}$ is the row vector containing the predicted values of the traffic variable at the selected locations and $f$th time instance, $\hat{\alpha}_f^i \in \mathbb{R}^{1 \times n_i}$ contains the predictions for all locations at $f$th time instance, and $X_h$ is the relationship matrix. If the predictions $\hat{c}_f^s$ would be identical to the true values $c_f^s$, then the problem boils down to compressed sensing, which we discussed in the previous section. In practice, however, the predictions have some inaccuracies. Therefore, we can write $\hat{c}_f^s = \hat{c}_f^s + \Delta c_f$ where $\Delta c_f$ represents the prediction error for the subnetwork at time $f$. Furthermore, let $D = C_f X_h$ be the estimated network profile, during the entire observational period, without any prediction error in $C_f$ ($C_f = [c_f^s \ldots c_f^m]^T$). Then, the MSE between predicted $\hat{A}_f$ ($\hat{A}_f = [\hat{\alpha}_f^1 \ldots \hat{\alpha}_f^m]^T$) and true future values $A_f$ ($A_f = [\alpha_f^1 \ldots \alpha_f^m]^T$) can be written as:

$$\frac{1}{rn} \|A_f - \hat{A}_f\|^2_F = \frac{1}{rn} \|A_f - C_f X_h\|^2_F,$$

$$= \frac{1}{rn} \|A_f - D + (\hat{A}_f - D)\|^2_F,$$

$$= \frac{1}{rn} \left( \sum_{i=1}^{m} \sum_{j=1}^{n} (a_{ij} - d_{ij})^2 + \sum_{i=1}^{m} \sum_{j=1}^{n} (\hat{a}_{ij} - d_{ij})^2 
- 2 \sum_{i=1}^{m} \sum_{j=1}^{n} (a_{ij} - d_{ij})(\hat{a}_{ij} - d_{ij}) \right),$$

where the first component of the error corresponds to the non-stationarity of spatial patterns within the network and the
second component is due to inaccurate predictions. The third component of the error shows the correlations between these two error components. We refer to (9) as MSE for compressed prediction. We rewrite (9) in more explicit form:

\[
\text{MSE}_{\text{total}} = \text{MSE}_{\text{est}} + \text{MSE}_{\text{pred}} - 2 \xi_{\text{pred}},
\]

(10)

where \( \xi_{\text{pred}} \) refers to correlation between non-stationarity and predictability. By substituting (6) in (10), we obtain:

\[
\text{MSE}_{\text{total}} = \text{MSE}_{\text{com}} + \text{MSE}_{\text{ns}} + \text{MSE}_{\text{pred}} - 2 \xi_{\text{est}} - 2 \xi_{\text{pred}}.
\]

(11)

Hence, the total error of compressed prediction can be decomposed into four error components: (i) error due to compression; (ii) error due to changes in spatial relationships (non-stationarity); (iii) error due to inaccurate predictions; (iv) correlations among the previous error components.

Compressed prediction provides significant reduction in computational complexity by explicitly predicting the traffic variables for only a small subset of road segments in the network. Compressed prediction involves two computations: (i) prediction of the traffic conditions at representative locations in the network and (ii) extrapolation of the predicted values to the entire network. In the former, the computational complexity depends on the underlying prediction algorithms, and is proportional to the number of locations \( c \) in the subnetwork. The second step (extrapolation) requires a single matrix-vector multiplication with complexity \( O(cn) \). In practice, the predictions at each link in the subnetwork are computationally complex. By contrast, the extrapolation can be executed much faster. Therefore, by performing prediction only for a small subnetwork, the computational complexity can be drastically reduced.

V. EXPERIMENTAL SETUP

We consider the nationwide traffic network in Singapore which contains diverse types of roads (see Fig.1). The variable of interest is the average traffic speed, i.e., the average speed of all vehicles which traverse a link during the given sampling interval of 5 minutes. The data set contains the average speed at each link of the transportation network for a period of three months (August - October 2011). We selected 17967 links which had less than 5% of missing values. We performed imputation by applying the Low Dimensional CP Weighted OPTimization (LDCP-WOPT) imputation method as it is able to deal with the large data set [20], [21].

We represent the data set in the form of a matrix as explained in Section IV. For compressed sensing and compressed prediction, we need training data to: (i) determine the subnetwork of \( c \) links, corresponding to the matrix \( C \) (see

\[
\text{PRD} = \frac{\| A - CX \|_F}{\| A \|_F}.
\]

(12)

We use the mean squared error (MSE) to assess the impact of several error components on the overall error. MSE is defined as:

\[
\text{MSE} = \frac{1}{mn} \| A - CX \|_F^2.
\]

(13)

VI. RESULTS

First we investigate the CX-based method for the compression of traffic data. We apply SVD sampling strategy to find the appropriate set of columns. We repeat sampling five times and report the average reconstruction accuracy. Our benchmark is PCA, as it is considered as the optimal linear transformation. Fig. 2a shows the compression performance of the proposed and baseline methods. As expected, PCA outperforms the proposed method in terms of compression error. However, as we pointed out earlier, the low-dimensional model obtained by PCA cannot be applied for compressed estimation and prediction, since it requires data from all links. By contrast, the low-dimensional models generated by the CX method are perfectly suitable for both applications, which is the main advantage of our approach compared to PCA.

An important question about the proposed CX-based method is whether the sampling scheme leads to subnetworks that are stable over time. To assess the stability of the subnetwork generated by the SVD sampling method, we applied this method to each of the three months (August, September, October, 2011) of traffic data separately. For
each month, we sort the road segments in descending order according to the assigned probability by the SVD sampling method. Hence, the most representative roads are at the top of these lists. Next we select the first \( k \) links of each list, with \( k \) corresponding to 10\%, 25\%, and 50\% of the links in the network. If the subnetwork is stable across time, the three short lists of top-\( k \) links should have many links in common. The results of this analysis are summarized in Table I, where the percentage of common links is provided. As it can be seen from this Table, most links in the subnetwork are consistently selected for all three months, suggesting that the SVD sampling method results in a subnetwork that is stable over time.

We now investigate the case of compressed sensing. We aim to reconstruct the average speed at each link in the entire network by collecting data from a small subset of roads. The relationship matrix \( X_k \) is determined from the training set (data from Aug-Sep, 2011), and the reconstruction error is assessed on the test set (data from October, 2011). Fig. 2b shows the reconstruction accuracy of the proposed method for three different approaches: In the first approach we select the subset of road segments according to SVD sampling scheme (see solid line in Fig. 2b). In the second approach we cluster the network according to the category of the road. For each cluster, we select the subset of the road segments using SVD sampling scheme. Then, we perform compressed sensing for each cluster separately (see dashed line in Fig. 2b). In the third approach we use the identical set of roads as defined in the second approach to perform network estimation. Unlike in the second approach, we do not perform any clustering here (see dotted line in Fig. 2b). Intuitively, the difference between the second and the third approach shows the gain obtained by network clustering. Fig. 2b indicates that applying the compressed sensing method to different road categories leads to better estimation performance for the entire network. As expected, the reconstruction accuracy of all three approaches increases with the size of subnetwork. Let us now investigate the error of compressed sensing in more details. In our analysis, we consider the subset of links as defined in SVD scheme without clustering.

The overall compressed sensing (estimation) error is caused by information loss due to compression of traffic data and changes in traffic behavior between training and testing periods. Table II shows the MSE of the individual error components, the correlation between the two errors components and the total MSE, for different compression ratios. As it can be seen from Table II the non-stationarity of the traffic data is the main contributor to the estimation error. As expected, the error associated with the compressibility of traffic data increases with the compression ratio. Furthermore, compression ratio.

We decompose the MSE of compressed prediction into estimation and prediction components. Table III shows the contribution of these two error components as well as the correlation between them for 5 minute prediction horizon. As it can be seen from Table III, the estimation error increases with the compression ratio. This increase in estimation error is mainly due to non-stationarity of the error component (see Table II). Table III shows that the prediction error tends to be dominant for smaller compression ratio, i.e., when significant portion of the network is explicitly predicted. From Table III, we can also see that there is some correlation between the two error components.

The proposed approach of compressed prediction provides substantial reduction in computational complexity by explicitly predicting the variables at a small representative set, followed
Compression ratio (see Table IV).

Computational complexity is approximately proportional to the discussions.

A compressed prediction can be applied in conjunction with selection techniques can lead to better performance of a negligible increase in prediction error.

Our numerical results show that the proposed method decomposed the compressed prediction error into several compressed sensing and compressed prediction. We low-dimensional models to enhance the scalability of network wide extrapolation. As Table IV shows, the latter can be neglected. Consequently, the required computation time for compressed prediction is proportional to the number of the road segments in C. Consequently, the reduction in computational complexity is approximately proportional to the compression ratio (see Table IV).

### VII. CONCLUSIONS

In this paper we utilized column based (CX) low-dimensional models to enhance the scalability of compressed sensing and compressed prediction. We decomposed the compressed prediction error into several components and investigated the relationship between them. Our numerical results show that the proposed method significantly reduces the computational cost at the expense of a negligible increase in prediction error.

In future work, we will explore whether other column selection techniques can lead to better performance of compressed prediction. Also, we will investigate how compressed prediction can be applied in conjunction with routing, in order to optimize routes taking future traffic conditions into account.

### VIII. ACKNOWLEDGMENT

Justin Dauwels wishes to thank Andrzej Cichocki for helpful discussions.

### REFERENCES


**Table III:** The MSE error of the proposed method for application of compressed prediction and for different compression ratios.

<table>
<thead>
<tr>
<th>Compression Ratio</th>
<th>2</th>
<th>4</th>
<th>6</th>
<th>8</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE (EST)</td>
<td>19.78</td>
<td>31.25</td>
<td>38.47</td>
<td>41.18</td>
<td>43.28</td>
</tr>
<tr>
<td>MSE (SVR)</td>
<td>33.73</td>
<td>28.23</td>
<td>24.15</td>
<td>20.95</td>
<td>18.61</td>
</tr>
<tr>
<td>ξ (corr)</td>
<td>5.80</td>
<td>6.01</td>
<td>5.52</td>
<td>4.83</td>
<td>4.42</td>
</tr>
<tr>
<td>Total MSE</td>
<td>41.91</td>
<td>47.46</td>
<td>51.58</td>
<td>52.47</td>
<td>53.05</td>
</tr>
</tbody>
</table>

**Table IV:** Computation time (in seconds) of the compressed method for 5 minute prediction horizon. The traditional approach requires 91.63 sec to perform prediction for the whole network.

by (fast and efficient) extrapolation to the entire network. This reduction in computation time is obtained at the expense of a small increase in the prediction error (see Fig 3). The computation times for the compressed and traditional methods are reported in Table IV. For the purpose of benchmarking, we tested the compressed and uncompressed prediction algorithms on 2.67 GHz MacPro server on a single core with 32GB of random-access memory (RAM). We assume that training phases are performed offline for both methods. Prediction time for compressed method involves the time to predict traffic variable for a subset of links and time required to perform network wide extrapolation. As Table IV shows, the latter can be neglected. Consequently, the required computation time for compressed prediction is proportional to the number of the road segments in C. Consequently, the reduction in computational complexity is approximately proportional to the compression ratio (see Table IV).