TRANSFORMED SPIKED COVARIANCE COMPLETION FOR TIME SERIES ESTIMATION

Benjamin Eng, Zhizhen Zhao, Farzad Kamalabadi, and Lara Waldrop

Department of Electrical and Computer Engineering and Coordinated Science Laboratory
University of Illinois Urbana-Champaign

ABSTRACT

In this paper, we address the problem of estimating a noisy, incomplete time series of a dynamical system with an unknown state evolution. The technique that we will present is transformed spiked covariance completion (TSCC), a matrix completion method for signal estimation. This method exploits the spiked covariance model of the underlying signal to develop a linear estimator that is resilient to noise. We discuss the conditions in the signal model for which this technique is applicable and compare this method against other state-of-the-art time series estimation techniques with a numerical example. Our algorithm gives estimates that are more robust to noise in comparison to the current state-of-the-art techniques that address this same estimation problem.

Index Terms— Matrix Completion, Time Series Estimation, Singular Spectrum Analysis, Denoising

1. INTRODUCTION

The problem of estimating a noisy, incomplete time series of an unknown dynamical system is pervasive in applications ranging from predicting weather patterns [1] to understanding trends in the global economy [2]. In these application domains, we wish to estimate a true signal (perhaps temperature or GDP) based off some measurement data that may be partially missing in addition to contamination by noise. In the field of space remote sensing, radiance data collected by space-borne sensors often contains incomplete data due to the sampling periodicity of the spacecraft and sensor malfunctions. In addition, the total signal counts collected at a detector usually is small due to the short integration time of the sensor. In this domain, it is quite typical to have SNRs < 10 dB. The observed data represents a dynamical system, where the current state is characterized by some state evolution that is unknown. This paper is concerned with the problem of recovering the state of a system from noisy, incomplete measurements with no prior information of the state evolution.

Traditional methods for estimating time series usually involve the use of Dynamic Linear Models. These methods often use a Kalman Filter (KF) to estimate a signal from noisy measurements. For high dimensional cases, the Ensemble Kalman Filter in [3] is used. The KF provides optimal state estimates in the sense that there is minimum mean square error between measurements and estimated measurements. The classical KF is only applicable in cases where the state transition matrix of the system is known.

Another class of time series estimation techniques assume that the dynamics of the system follow a parametric model. In these methods, given a sequence of time series measurements, a linear model for the true signal is learned and is used in the estimation process. In [4], the estimation technique, temporal regularized matrix factorization (TRMF), performs matrix factorization via methods in [5] on a data matrix (vectorized columns of measurements) to find feature and temporal matrices, where the temporal matrix is modeled as a autoregressive process. By approximating the original data matrix as the multiplication of two matrices, this method effectively denoises the columns of the original data matrix in addition to imputing any missing values. This technique effectively converts a linear estimation problem with a noisy, incomplete signal into a matrix completion problem.

Matrix completion is a class of problems where measurements are put into a matrix and the missing entries of the matrix are filled in according to some desired structure of the matrix. For example, one may desire the estimated matrix to have a low rank structure or some minimum spectral norm. Typically, these algorithms work by estimating the singular values and the left/right singular vectors of a partially observed matrix and then minimizing the Frobenius norm error between the estimated and the observed entries. In general, matrix completion techniques assume that the columns of the matrix are independently and identically distributed (iid) according to some distribution. State-of-the-art techniques such as ones used in [6, 7, 8] impute values into the matrix with these assumptions. Other methods, e.g. [9], set the estimation of a partially observed matrix as the minimization of the nuclear norm of its estimate. These techniques have high reconstruction accuracy but do not incorporate any temporal dependencies in the columns of matrices. In addition, these techniques are often employed in recommendation systems, where the measurement noise is either very low or nonexistent. In this paper, we will introduce Transformed Spiked Covariance Completion (TSCC), an estimation technique that will exploit the spiked covariance model of a time-lagged embedding of observations. The method is a non-parametric estimation approach that initially converts the time series observations into a trajectory matrix via time-lagged embedding in [10] to capture the underlying dynamics. Subsequently, the assumption that the underlying signal lies in a low dimensional linear subspace is exploited to determine the empirical Best Linear Estimator [11]. As a result, we can form a linear estimator to truncate and shrink the singular values of the trajectory matrix similar to methods in [12, 11] to estimate the entries. In this technique, we do not perform estimation by assuming any state evolution model. We show that this technique provides accurate recovery under noise and missing entries and is computationally more efficient than other existing methods. In section 2 and 3 , we discuss the signal model and the details of the proposed method. In section 4, we will present numerical experiments that demonstrate the superior reconstruction accuracy of the TSCC method over the other mentioned estimation techniques.

2. THE SIGNAL MODEL

The model for this signal estimation problem considered for this paper is:

$$X_{i+1} = F_i X_i + w_i \quad Y_i = A_i X_i + \epsilon_i \tag{1}$$

We consider a linear dynamical system whose true state is represented by feature vector X_i in $\mathbb{R}^N$ and whose dynamics are captured by the state transition matrix F_i. Y_i is a measurement vector in $\mathbb{R}^N$, containing data gathered by some sensor. A_i is a diagonal matrix in $\mathbb{R}^{N\times N}$ with entries of 0 or 1 on the diagonal. A_i is referred to as an observation matrix as it maps the true state X_i to the observation Y_i. In the case with missing entries, we model A_i as a diagonal matrix, with values of 0 or 1 on the diagonal. Both observation noise ϵ_i and process noise w_i are assumed to be additive white Gaussian noise. In our estimation technique , we do not assume the knowledge dynamics F_i.The multivariate time series $[Y_0, Y_1, ..Y_{T-1}]$ forms a matrix of size $N \times T$ and its true state is $X = [X_0, X_1, ..X_{T-1}]$. Our goal is to estimate X from Y.

3. PROPOSED METHOD

We develop a non-parametric and non-iterative method to find a linear estimator $\hat{X}$ based on Y under some assumptions of the structure of X.

3.1. Trajectory Matrix Formation

In our approach, we form a trajectory matrix Z from the columns of Y and obtain the following matrix:

$$Z = \begin{bmatrix} Y_L & Y_{L+1} & ... & Y_J \\ Y_{L-1} & Y_L & ... & Y_{J-1} \\ ... & ... & ... & ... \\ Y_1 & Y_2 & ... & ... \end{bmatrix}. \tag{2}$$

The trajectory matrix, Z, is a block Toeplitz matrix of dimension $(NL) \times (T-L+1)$ where N is the dimension of a single observation and L is the number of lagged observation vectors Y_i. We can also declare matrix Q with columns Q_i to be the corresponding trajectory matrix of X. These trajectory matrices are one class of delay embeddings that are derived from dynamical system theory [13, 14, 15, 16]. By representing the set of observations as the delayed versions of itself, the embedding can capture the predictable modes of the dynamical system as seen in Taken's embedding theorem [10]
.

For the case of Singular Spectrum Analysis (SSA) where A_i is an identity matrix with the assumption that Q is low rank, Q is determined by a rank r approximation of Z with proper shrinkage of singular values.

$$\hat{Q} = \sum_{k=1}^{r} \eta(\sigma_k) u_k v_k^\top \tag{3}$$

We see that $\hat{Q}$ follows a block Toeplitz structure. Diagonal averaging of $\hat{Q}$ is applied to obtain the final estimate of X.

$$\hat{X}_i = \frac{1}{L} \sum_{j,k} \hat{Q}_{j,k} \quad \text{where} \quad k - j = i \tag{4}$$

Here, we sum over the j^{th} lagged vector at time k for $0 \leq j \leq L-1$ and $0 \leq k \leq T-L$

In the problem with missing entries, A_i is not an identity matrix. In addition A_i is not full rank therefore we cannot take its inverse. If we perform the procedure above, it will not provide a good estimator of X due to the missing entries and noise. (see Fig. 1). In the following subsection, we will describe a statistical method to estimate X.

3.2. Linear Estimation with Spiked Covariance Model

In Eq (1), we assume in the signal model that our observation is a linear transformation of the true signal with additive white Gaussian noise ϵ_i. To impute and denoise the entries of Z, we can write an estimator similar to the empirical Best Linear Predictor (EBLP) in [11] that can leverage this assumption in estimation. In the case that the noise is not Gaussian, this estimator can be modified by simple whitening techniques shown in [11].

We denote by $\tilde{A}_i$ the truncation matrix for each column Z. This matrix is of size $NL \times NL$ and is a diagonal matrix with values of 1 or 0. The additive white Gaussian noise after time-lagged embedding is denoted by $\tilde{\epsilon}_i$. Following the procedure in [11], we compute the diagonal normalization matrix,

$$M = \frac{1}{T-L+1} \sum_{i=1}^{T-L+1} \tilde{A}_i \tilde{A}_i^\top. \tag{5}$$

We have $\tilde{A}_i \tilde{A}_i^\top = M + E_i$ where E_i is the deviation of $\tilde{A}_i \tilde{A}_i^\top$ from the ensemble mean. In the high dimensional limit, i.e.,

$NL \to \infty, T - L + 1 \to \infty$, and $\frac{NL}{T-L+1} \to \gamma$, the operator norm of the matrix with rows $\frac{E_i Q_i}{\sqrt{T-L+1}}$ vanishes. Therefore, we can write B_i as:

$$B_i = \tilde{A}_i^\top Z_i = MQ_i + E_i Q_i + \tilde{A}_i^\top \tilde{\epsilon}_i \sim MQ_i + \tilde{A}_i^\top \tilde{\epsilon}_i \quad (6)$$

Because M is full rank, we can compute

$$\tilde{B}_i = M^{-1} B_i \sim Q_i + M^{-1} \tilde{A}_i^\top \tilde{\epsilon}_i \quad (7)$$

The additive white Gaussian noise $\tilde{\epsilon}_i$ is linearly transformed to $M^{-1}\tilde{A}_i^\top \tilde{\epsilon}_i$, therefore, on the right hand side of Eq. (7), the true signal is contaminated by colored noise. We use an empirical best linear estimator in [11] to recover Q_i from $\tilde{B}_i$. In signal processing, the best linear predictor is also known as Wiener Filter. We obtain the singular values σ_k and the singular vectors u_k, v_k of $\tilde{B}$. We truncate and shrink the singular values of $\tilde{B}$ using random matrix theory and generalized Marčenko Pastur distribution [17]. The singular values after shrinkage are denoted by λ_k and the estimated Q is,

$$\hat{Q} = \sum_{k=1}^{r} \lambda_k u_k v_k^\top . \quad (8)$$

We obtain an estimate of X_i by diagonal averaging $\hat{Q}$ according to Eq. (4).

3.3. Computational Considerations

In addition to higher accuracy, the TSCC technique can be a faster algorithm when compared to TRMF. The complexity for TSCC is $O(\min(LN, T-L+1)^2 \times \max(LN, T-L+1))$. Like in the signal model, N is the dimension of the state vector, T is the number of measurements, and L is lag parameter for the trajectory matrix. This complexity comes mainly from the fact that EBLP performs an SVD with many other $O(NT)$ calculations in generating the shrinkage coefficients.

TRMF is an iterative algorithm whose single update complexity for each iteration is $O(NTk^2 + L(T - L + 1)k^2 + (L^3 + TL^2))$. Here the variables k and L respectively represent the latent dimension and the lag parameter (of the autoregressive model). Note that the complexity is the sum of three terms. TRMF decomposes a data matrix to two matrices that are constrained by an optimization problem that minimizes the Frobenius norm error between the observations and the approximation and the model error [4]. It is important to note that as the dimensions of the matrix increases, the number of iterations increase.

Since TRMF assumes a model for the dynamical system, to properly find its factorization, one would need to search through various model parameters, namely k and L to best fit the observations. TSCC, a non-iterative algorithm, can potentially have a better complexity compared to TRMF. The TSCC complexity is dominated by a SVD operation; new algorithms like the one presented in [18] show that a randomized SVD for the low rank approximation is linear with $O(lMN)$ where l is slightly larger than rank r of the matrix. Potentially, the incorporation of this algorithm into the current TSCC framework can lead to a complexity that is faster than TRMF.

4. NUMERICAL EXPERIMENTS

4.1. Synthetic Data Generation

To compare our technique against the current, state-of-the-art, we generate synthetic data to control noise variation. We first generate two matrices H in $\mathbb{R}^{N\times k}$ and G in $\mathbb{R}^{k\times L}$. These matrices are drawn from a zero mean, unit variance normal distribution. We then generate a state transition matrix W whose size is the same as G; this matrix is also drawn from the same zero mean, unit variance normal distribution. G can be thought of as a collection of L k-length vectors. The $L + 1^{th}$ vector in the collection can be calculated by

$$G_i = \sum_{t=0}^{L-1} W_t G_{i-t} \quad (9)$$

Here W_t is the diagonal matrix formed from the t^{th} column of W and the G_{i-t} is the t^{th} lagged k-length vector in G. Using this model, we can generate an arbitrary size sequence of vectors $\hat{G}$.
The resulting data matrix that we will use in these experiments would be generated by the simple multiplication $X = F\hat{G}$. The clean signal is stored in $\bar{X}$. To simulate noise, we simply add a normal noise vector of zero mean and some deviation σ to the clean signal. We vary σ throughout our experiments.

4.2. Comparison of Methods

In these experiments, we compare four methods. The first method we consider is TRMF as seen in [4]. TRMF performs time series estimation via matrix completion under the autoregressive assumption. The second method we examine is direct estimation of matrix M with EBLP. The third technique is our TSCC technique detailed in section 3. The fourth technique is the standard SSA technique in [19]. It should be noted that EBLP is a general algorithm that estimates M on the basis of a spiked covariance model. In TSCC, estimation is done on a time-lagged embedding, which is restricted to Toeplitz block matrices.

As input to each technique, we feed in a noisy data matrix Y generated from the methods presented in section 4.1. For TRMF, we set the number of iterations to 10, the point where the algorithm converges. The simulated data has 50 state features and 250 time samples for N and T respectively. We first generate 100 clean data matrices (no noise). For each of the clean data matrices, we form a new noisy data matrix by adding Gaussian Noise vectors with deviations of 0.1,0.3,0.5,0.8 and 1.0. In total, we have 600 matrices exhibiting a unique level of noise. To simulate missing data, we replace 20% of the entries of Y from each column in the generated data matrix with zeros. We apply the mentioned estimation techniques on the data matrices and evaluate the error, via the relative Frobenius Norm Error defined as:

$$Err = \frac{||X - \hat{X}||_F}{||X||_F} \tag{10}$$

In fig. 1, we display the clean data matrix, which contains the noise-less entries of the matrix before the addition of Gaussian noise and the removal of data. The reconstructed matrices from TRMF, SSA, and TSCC are displayed as well in the four panel plot. In this particular figure, the noise deviation was 0.8.

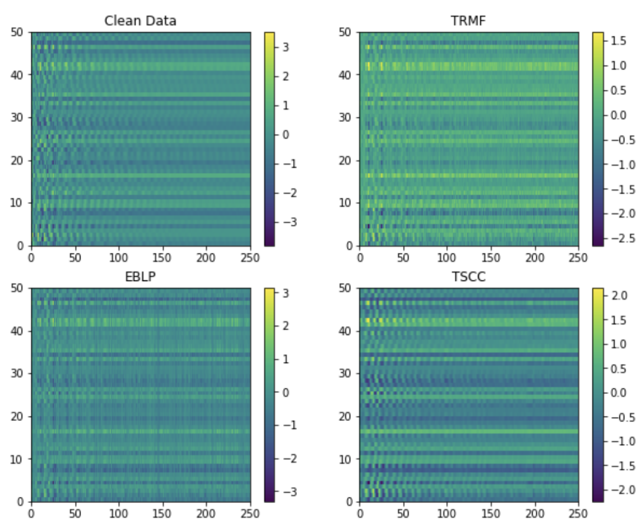

Fig. 1. Estimations for Noise Deviation = 0.8

Visually, the TRMF and the TSCC methods capture the structure of the dynamics when compared to the clean data. The EBLP method by itself is not robust to missing data for this example; this is seen with the striation through temporal axis of the matrix where data is missing. These striations are also seen in the vertical direction of the TRMF figure. TSCC performs better than EBLP by reforming Y as a trajectory matrix. TSCC then has increased estimation accuracy due the mode capture capability of the embedding.

In our experiments, the average compute time was calculated by dividing the time taken to estimate the matrices over the number of data matrices. The average compute time was 294 ms for TRMF, 56 ms for EBLP, and 64 ms for TSCC on a 3.3 GHz Intel i7 processor. Experimentally, TSCC performs faster than the state-of-the-art TRMF.

Comparing these methods across various levels of noise, we see that TSCC outperforms TRMF in all cases of noise levels. This is seen in fig. 2 where reconstruction accuracy versus the noise deviation is plotted for all three techniques.

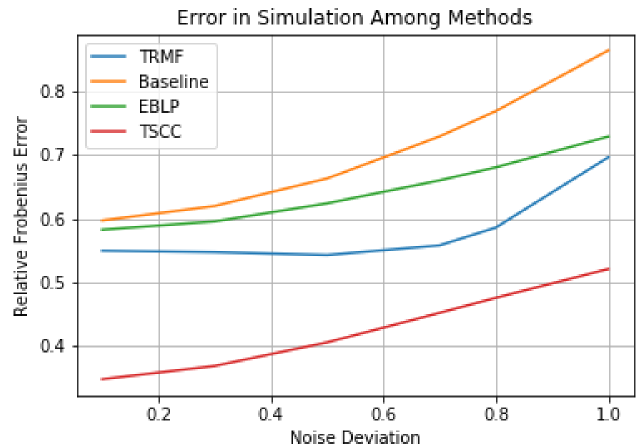

Fig. 2. Reconstruction Accuracy vs. Noise Level

4.3. Analysis of Results

This experiment clearly demonstrates the estimation capability of TSCC for the signal model in (1). We see that TSCC has superior denoising ability as the linear estimator used can find optimal shrinkage coefficients to separate noise from signal. TSCC also proves effective at matrix completion with a time series dataset. In addition, by using a delayed embedding like the trajectory matrix, we can more easily generalize the types of dynamical systems that we want to estimate.

It is surprising to note that TRMF fails to properly estimates the parameters of the time series even though the synthetic data generated was modeled in an autoregressive fashion. Using a non-parametric method like TSCC allows flexibility in estimating the modes of the dynamical system.

5. CONCLUSION

For the problem of estimating the true, underlying signal from observations of a dynamical system, we present TSCC, a method to reconstruct the time series from the partial measurements. We evaluate this technique against the state-of-the-art technique TRMF and demonstrate that our proposed method is more robust to noise in addition to having better computational complexity with a numerical example. The main advantage of our technique is that we can represent a multi-variate time series as an embedding whose noise covariance follows a spiked covariance model, allowing us to develop a linear estimator that is effective at denoising and matrix completion.

6. REFERENCES

[1] D. Comeau, Z. Zhao, D. Giannakis, and A. Majda, "Data-driven prediction strategies for low-frequency patterns of north pacific climate variability," *Climate Dynamics*, vol. 48, no. 5, pp. 1855–1872, Mar 2017.

[2] L. Bauwens, M. Lubrano, and J. Richard, *Bayesian Inference in Dynamic Econometric Models*, Oxford University Press, 2000.

[3] M. D. Butala, R. A. Frazin, Y. Chen, and F. Kamalabadi, "Tomographic imaging of dynamic objects with the ensemble kalman filter," *IEEE Transactions on Image Processing*, vol. 18, no. 7, pp. 1573–1587, July 2009.

[4] H.-F. Yu, N. Rao, and I. S. Dhillon, "High-dimensional Time Series Prediction with Missing Values," *ArXiv e-prints*, Sept. 2015.

[5] N. Rao, H. Yu, and I. Ravikumar, P.and Dhillon, "Collaborative filtering with graph information: Consistency and scalable methods," in *Advances in Neural Information Processing Systems 28*, C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, Eds., pp. 2107–2115. Curran Associates, Inc., 2015.

[6] N. B. Shah, S. Balakrishnan, and M. J. Wainwright, "Low Permutation-rank Matrices: Structural Properties and Noisy Completion," *ArXiv e-prints*, Aug. 2017.

[7] R. H. Keshavan, A. Montanari, and S. Oh, "Matrix Completion from a Few Entries," *ArXiv e-prints*, Jan. 2009.

[8] J. Andén, E. Katsevich, and A. Singer, "Covariance estimation using conjugate gradient for 3D classification in Cryo-EM," *ArXiv e-prints*, Dec. 2014.

[9] E. J. Candes and Y. Plan, "Matrix Completion With Noise," *ArXiv e-prints*, Mar. 2009.

[10] F. Takens, "Detecting strange attractors in turbulence," *Lecture Notes in Mathematics, Berlin Springer Verlag*, vol. 898, pp. 366, 1981.

[11] E. Dobriban, W. Leeb, and A. Singer, "Optimal prediction in the linearly transformed spiked model," *ArXiv e-prints*, Sept. 2017.

[12] D. Donoho, M. Gavish, and I. Johnstone, "Optimal shrinkage of eigenvalues in the spiked covariance model," *arXiv preprint arXiv:1311.0851*, 2013.

[13] N. Packard, J. Crutchfield, J. Farmer, and R. Shaw, "Geometry from a time series," *Physical review letters*, vol. 45, no. 9, pp. 712, 1980.

[14] E. Deyle and G. Sugihara, "Generalized theorems for nonlinear state space reconstruction," *PLoS One*, vol. 6, no. 3, pp. e18295, 2011.

[15] T. Sauer, J. Yorke, and M. Casdagli, "Embedology," *Journal of Statistical Physics*, vol. 65, no. 3, pp. 579–616, Nov 1991.

[16] D. Giannakis and A. Majda, "Nonlinear laplacian spectral analysis for time series with intermittency and low-frequency variability," *Proceedings of the National Academy of Sciences*, vol. 109, no. 7, pp. 2222–2227, 2012.

[17] F. Götze and A. Tikhomirov, "On the Rate of Convergence to the Marchenko–Pastur Distribution," *ArXiv e-prints*, Oct. 2011.

[18] N. Halko, P. Martinsson, Y. Shkolnisky, and M. Tygert, "An algorithm for the principal component analysis of large data sets," *SIAM Journal on Scientific computing*, vol. 33, no. 5, pp. 2580–2594, 2011.

[19] G. Tsagkatakis, B. Beferull-Lozano, and P. Tsakalides, "Singular spectrum-based matrix completion for time series recovery and prediction," *EURASIP Journal on Advances in Signal Processing*, vol. 2016, no. 1, pp. 66, May 2016.