Robust Kronecker Product PCA for Spatio-Temporal Covariance Estimation

Kristjan Greenewald, Student Member, IEEE, and Alfred O. Hero III, Fellow, IEEE

Abstract—Kronecker PCA involves the use of a space vs. time Kronecker product decomposition to estimate spatio-temporal covariances. It was shown in [1], [2] that a diagonally loaded covariance matrix is not well modeled by such a decomposition and that using a diagonal correction factor in the decomposition significantly reduces the required separation rank of the KronPCA estimate. In this work the addition of a sparse correction factor is considered, which corresponds to a model of the covariance as a sum of KronKronecker products and a sparse matrix. This sparse correction includes diagonal correction as a special case but allows for sparse unstructured “outliers” anywhere in the covariance matrix. This paper introduces a robust PCA-based algorithm to estimate the covariance under this model, extending the nuclear norm penalized LS KronKronecker PCA approaches of [2], [3]. This paper also provides an extension to Toeplitz temporal factors, producing a parameter reduction for temporally stationary measurement modeling. High dimensional MSE performance bounds are given for these extensions. Finally, the proposed extension of KronPCA is evaluated and compared on both simulated and real data coming from yeast cell cycle experiments. This establishes the practical utility of the sparse correction in biological and other applications.

I. INTRODUCTION

In this paper, we develop a method for the estimation of spatio-temporal covariances and apply it to multivariate time series modeling and parameter estimation. The covariance for spatio-temporal processes manifests itself as a multiframe covariance, i.e. the covariance not only between variables or features in a single frame (time point), but also between variables in a set of nearby frames. In streaming applications, at each time \( t \) the covariance may be estimated over a sliding time window of \( p_t \) frames. If each frame contains \( p_s \) spatial variables, then the covariance is described by a \( p_t p_s \) by \( p_t p_s \) matrix:

\[
\Sigma_t = \text{Cov} \left( \{ I_n \}_{n=1}^{t-1} \right)
\]

where \( I_n \) denotes the \( p_s \) variables or features of interest in the \( n \)th video frame. We make the standard piecewise stationarity assumption that \( \Sigma_t \) can be approximated as unchanging over each consecutive set of \( k_t \) frames.

As \( p_s p_t \) can be very large, even for moderately large \( p_s \) and \( p_t \) the number of degrees of freedom \( (p_s p_t (p_s p_t + 1)/2) \) in the covariance matrix can greatly exceed the number \( n \) of training samples available to estimate the covariance matrix. One way to handle this problem is to introduce structure and/or sparsity into the covariance matrix, thus reducing the number of parameters to be estimated.

A natural non-sparse option is to introduce structure by modeling the covariance matrix \( \Sigma \) as the Kronecker product of two smaller matrices, i.e.

\[
\Sigma \approx T \otimes S.
\]

When the measurements are Gaussian with covariance of this form they are said to follow a matrix-normal distribution [4]. This model lends itself to coordinate decompositions [3]. For spatio-temporal data, we consider the natural decomposition of space (variables) vs. time (frames) [3], [1]. In this setting, the \( S \) matrix is the “spatial covariance” and \( T \) is the “time covariance,” both determined up to a multiplicative constant.

An extension to the representation (2) discussed in [3] approximates the covariance matrix using a sum of Kronecker product factors

\[
\Sigma \approx \sum_{i=1}^{r} T_i \otimes S_i
\]

where \( r \) is the separation rank. We call this the Kronecker PCA (KronPCA) covariance representation. This model with \( r > 1 \) has been used in various applications, including video modeling and classification [1], [5], network anomaly detection [2], synthetic aperture radar, and MEG/EEG covariance modeling (see [3] for references). In [6] it was shown that any covariance matrix can be represented in this form.

This allows for more accurate approximation of the covariance when it is not in KronKronecker product form but most of its energy is in the first few KronKronecker components. An algorithm (Permuted Rank-penalized Least Squares (PRLS)) for fitting the model (3) to a measured sample covariance matrix was introduced in [3] and was shown to have strong high dimensional guarantees in MSE performance. This KronKronecker PCA model does not naturally accommodate additive noise since the diagonal elements (variances) must conform to the KronKronecker structure.

In [1] we extended this KronPCA model, and the PRLS algorithm of [3], by adding a structured diagonal matrix to (3). This model is called Diagonally Loaded KronKronecker PCA (DL-KronPCA) and, although it has an additional \( p_s \) parameters, it was shown that for fixed \( r \) it performs significantly better for both inverse covariance estimation and inference since it better captures the effect of additive uncorrelated (white) noise.

The DL-KronPCA model [1] is the \( r + 1 \)-Kronecker model (where \( U \) are diagonal)

\[
\Sigma \approx \left( \sum_{i=1}^{r} T_i \otimes S_i \right) + U.
\]
where the diagonal matrix $U$ is called a “diagonal loading matrix.” Following Pitsianis-VanLoan rearrangement of the square $p_t p_s \times p_t p_s$ matrix $\Sigma$ to an equivalent rectangular $p_t^2 \times p_s^2$ matrix $B$ as in [3], [7], this becomes the approximation problem of finding a low rank plus diagonal approximation [1], [3]. A weighted least squares solution to this problem is given in [1], [2].

This paper extends DL-KronPCA to the case of sparse loading that is not necessarily diagonal. In other words, we model the covariance as the sum of a low separation rank matrix and a sparse matrix $\Sigma_{\text{sparse}}$:

$$\Sigma \approx \left( \sum_{i=1}^{r} T_i \otimes S_i \right) + \Sigma_{\text{sparse}}. \quad (5)$$

DL-KronPCA is trivially a special case of this model. The motivation behind this extension is that while the KronPCA model may work well for most variables in the problem, there are sometimes a few variables (or correlations) that cannot be well modeled using KronPCA, due to complex non-Kronecker structured covariance patterns, e.g., covariance outliers. Thus, putting these variables and correlations in their own sparse term allows the separation rank of the remainder to decrease, thus reducing the number of model parameters. We call this model Robust Kronecker PCA (Robust KronPCA), and propose regularized least squares based estimation algorithms closely related to those proposed for Robust PCA. In particular, we consider a singular value thresholding approach (arising from a nuclear norm). We derive high dimensional consistency results for the SVT-based algorithm and apply the method to modeling a cell cycle dataset. Following [2], we also allow for the enforcement of a temporal block Toeplitz constraint, which corresponds to a temporally stationary covariance and results in a further reduction in the number of parameters when the process under consideration is temporally stationary and the time samples are uniformly spaced.

The rest of the paper is organized as follows: in Section II we introduce and motivate our Robust KronPCA model. Section III introduces an algorithm for estimating Robust KronPCA covariances, and Section IV provides high dimensional consistency convergence theorems. Simulations and the application to cell cycle data are presented in Section V and our conclusions are given in Section VI.

II. Robust KronPCA

Let $X$ be a $p_t \times p_s$ matrix with entries $X(m,t)$ denoting samples of a space-time random process defined over a $p_t$-grid of space samples $m \in \{1, \ldots, p_t\}$ and a $p_s$-grid of time samples $t \in \{1, \ldots, p_t\}$. Let $x = \text{vec}(X)$ denote the $p_t p_s$ column vector obtained by lexicographical reordering. Define the $p_t p_s \times p_t p_s$ spatiotemporal covariance matrix $\Sigma = \text{Cov}[x]$.

We consider extending DL-KronPCA to the case of sparse loading. As noted above, we model the covariance as the sum of a low separation rank matrix $\Theta$ and a sparse matrix $\Gamma$:

$$\Sigma \approx \left( \sum_{i=1}^{r} T_i \otimes S_i \right) + \Theta + \Gamma. \quad (6)$$

The Pitsianis-VanLoan rearrangement operator $R(\cdot)$ maps $p_t p_s \times p_t p_s$ matrices to $p_t^2 \times p_s^2$ matrices and is defined in [3] as setting the $(i - 1)p_t + j$th row of $R(M)$ equal to $\text{vec}(M(i,j))^T$. $M(i,j)$ is defined as the $i,j$th $p_s \times p_s$ subblock of $M$, i.e. $M(i,j) = [M(i(-1)p_t + 1:p_t, j(-1)p_s + 1:p_s)]$. Following the approach of [1], [3], we propose to fit the model (5) to the sample covariance matrix $\Sigma_{\text{SCM}} = n^{-1} \sum_{i=1}^{n} (X_i - \bar{X})(X_i - \bar{X})^T$, where $\bar{X}$ is the sample mean, and $n$ is the number of samples of the space time process $X$. The estimation of the parameters $T_i$, $S_i$ and $\Gamma$ in (5) is performed by minimizing the following objective function

$$\min_{L,S} ||R - \tilde{L} - \check{S}||^2_F + \lambda_0 ||L||_s + \lambda_1 ||\check{S}||_1, \quad (8)$$

where $R = R(\Sigma_{\text{SCM}})$. The objective function is minimized over all $p_t^2 \times p_s^2$ matrices $R = \check{L} + \check{S}$. The matrices $\check{L}$ and $\check{S}$ correspond to the estimates of $R(\Theta)$ and $R(\Gamma)$ respectively. Hence, $\check{\Sigma} = R^{-1}(\check{L} + \check{S})$ when $\check{L}, \check{S}$ minimize the objective.

As shown in [1], the left and right singular vectors of $\check{L}$ correspond to the vectorized $T_i$ and $S_i$ respectively. Following the method of [3], the optimization can then be shown to be equivalent to

$$\min_{L,S} ||B - \check{L} - \check{S}||^2_F + \lambda_0 ||L||_s + \lambda_1 \sum_{j \in \mathcal{I}} c_j ||\check{S}_{j+p_t}||_1, \quad (9)$$

where $A_j$ denotes the $j$th row of $A$, $\check{R} = \check{P} \check{R}$, $\check{S} = \check{P} \check{S}$, and $B = \check{P} \check{R}$. The matrix $P$ is the $p_t^2 \times p_t^2$ identity matrix, the summation is over $\mathcal{I} = \{-p_t + 1, \ldots, p_t^2 - p_t\}$, and the constants $c_j = 1$. We maintain this level of generality in the expression for the objective as different $P$, $c_j$, and $\mathcal{I}$ will arise when, for example, the block Toeplitz constraint is introduced in the next section. This well-studied optimization problem (nuclear norm penalized low rank matrix approximation) is considered in [9], where it is shown to be strictly convex. Several fast solution methods are available, including the iterative SVD-based method shown in Algorithm [1]–[10]. If the sparse correction were to be omitted, the resulting optimization problem could be solved directly using the SVD [3]. The covariance estimate is given by

$$\check{\Sigma} = R^{-1}(\check{P}^T (\check{L} + \check{S})), \quad (10)$$

where $\check{L}, \check{S}$ are the minimizers of (9). We thus have the following:

**Theorem III.1** (Robust KronPCA). The objective function in Equation (9) is strictly convex and is equivalent to the Robust...
PCA objective function of [10] etc. Following [10] and others, convergence to the unique global minimum can be achieved using Algorithm [7].

Algorithm 1 Proximal Gradient Robust KronPCA
1: $B = PR(\Sigma_{SCM})$
2: Initialize $M, S, L$
3: while Not converged do
4: $L^k = SVT\tau_k\lambda^k_0(M^{k-1} - S^{k-1})$
5: for $j \in I$ do
6: $S^k_{j+p_t} = \text{soft}_{\tau_k\lambda^k_{j_p}}(M^k_{j+p_t} - L^k_{j+p_t})$
7: end for
8: $M^k = L^k + S^k - \tau_k(L^k + S^k - B)$
9: end while
10: $\Sigma = R^{-1}(P^T(L + S))$
11: return $\Sigma$

A. Toeplitz Case

In this section, we propose an estimation algorithm for Robust KronPCA estimation with a block Toeplitz constraint. For simplicity we consider the case that the covariance is Toeplitz with respect to time, however, extensions to the cases of Toeplitz spatial structure and Toeplitz structure simultaneously in both dimensions are straightforward. The objective function remains [8], subject to the constraint that both $\Theta$ and $\Gamma$ are temporally block Toeplitz.

For $\Theta$, the block Toeplitz constraint is equivalent to a Toeplitz constraint on the $T_i$. Following the results of [8], we note that the Toeplitz constraint on $T_i$ is equivalent to

$$[a^i_k] = a^i_{j+p_t}, \forall k \in K(j), j \in [-p_t + 1, p_t - 1].$$

for some vector $a^i$ where

$$K(j) = \{k : (k-1)p_t + k + j \in [-p_t + 1, p_t - 1]\}.$$  

Let

$$i^r_{j+p_t} = a^i_{j+p_t}/\sqrt{p_t - |j|}, \quad j \in [-p_t + 1, p_t - 1].$$

In the objective function [9], this implies [2] that for Toeplitz estimation the summation set $I$, the $\ell_1$ weighting constants $c_j$, and the $(2p_t - 1) \times p_t^2$ matrix $P$ are all defined as

$$I = \{-p_t + 1, \ldots, p_t - 1\},$$

$$c_j = 1/\sqrt{p_t - |j|},$$

$$P_{i+p_t,j+i} = \begin{cases} \frac{1}{\sqrt{p_t - |j|}} & i \in K(j), \\ 0 & \text{otherwise}. \end{cases}$$

where the last line holds for all $j = -p_t + 1, \ldots, p_t - 1, i = 1, \ldots, p_t^2$. Note that this corresponds to an unbiased Toeplitz estimate. It is of interest to investigate the extension of biased Toeplitz estimators [11] to this case. Note that this imposition of Toeplitz structure also results in a significant reduction in computational cost due to a reduction in the size of the optimization problem [2]. As the problem is of the form [9], the block Toeplitz estimate can also be found using [15] and computed via Algorithm [1].

The block Toeplitz estimate is given by

$$\Sigma = R^{-1}(P^T(L + S)),$$

where $L, S$ are the minimizers of [9].

The non Toeplitz and Toeplitz objectives are both invariant with respect to replacing $\Theta$ with $\Theta^2$ and $\Gamma$ with $\Gamma^T$. Furthermore, $R_\alpha((A + A^T)/2) \leq R_\alpha(a)$ by the triangle inequality. Hence we have the following corollary by the uniqueness of the global optimum (Theorem III.1):

Corollary III.2. For both the Toeplitz and non Toeplitz cases, the Robust KronPCA covariance estimates $\hat{\Theta}$ and $\hat{\Gamma}$ are symmetric with probability one.

IV. HIGH DIMENSIONAL CONSISTENCY

In this section, we assume that the data obeys the additive noise model $x_i = z_i + w_i$, where $z_i$ and $w_i$ are statistically independent and are zero mean, $z_i$ has true (population) covariance $\Theta$ and $w_i$ has a sparse true (population) covariance $\Gamma$. Under this model, the sample covariance is given by:

$$\frac{1}{n} \sum_{i=1}^n x_i x_i^T = \Theta + \Gamma + W$$

$W$ is the Wishart distributed error (noise) between the sample covariance and the population covariance respectively.

We first define the subspace pairs $M_{\gamma}, M^\perp_{\gamma}$ [12] associated with the set of either low separation rank ($\gamma_1 = \Theta$) or sparse ($\gamma_2 = \Gamma$) matrices. For the sparse case, let $S$ be the set of indices on which $\text{vec}(\Theta)$ is nonzero. Then $M_{\gamma} = M_F$ is the subspace of vectors in $\mathbb{R}^{p_t^2}$ that are have support contained in $S$, and $M^\perp_{\gamma}$ is the subspace of vectors orthogonal to $M_{\gamma}$, i.e. the subspace of vectors with support contained in $S^c$. For the case of $\Theta$, note that by [6] a decomposition of $\Theta$ exists of the form

$$\Theta = \min_{i=1}^{p_t^2} \sigma_i T^{(i)} \otimes S^{(i)}$$

where for all $i$, $\|T^{(i)}\|_F = \|S^{(i)}\|_F = 1$, $\sigma_i \geq 0$ and nonincreasing, the $p_t \times p_t$ matrices $(T^{(i)})_i$ are all linearly independent, and the $p_t \times p_t$ matrices $(S^{(i)})_i$ are all linearly independent. It is easy to show that this decomposition can be computed by extracting and rearranging the singular value decomposition of $R(\Theta)$ [6], [8], [7] and thus the $\sigma_i$ are uniquely determined. Let $r$ be such that $\sigma_r = 0$ for all $i > r$. Define the matrices $U_{\gamma}$ and $U^\perp_{\gamma} = \text{span}(\text{vec}(T^{(i)})) \otimes \text{span}(\text{vec}(S^{(i)}))$ respectively, where $e_i$ is the $i$th unit vector and $S_i$ is the $i$th entry of $S$. Then the subspaces $M_{\gamma} = \text{span}(U_{\gamma})$ and $M^\perp_{\gamma} = \text{span}(U^\perp_{\gamma})$.

Consider the covariance estimator given by solving the optimization problem in Equation [8]. As is typical in Robust PCA, an incoherence assumption is required to ensure that $\Theta$ and $\Gamma$ can be distinguished. Our incoherence assumption is as follows:

$$\max \left\{ \sigma_{\max}(P_{\bar{M}_{\Theta}}P_{M_F}), \sigma_{\max}(P_{M_{\Theta}}P_{\bar{M}_F}) \right\} \leq \frac{16}{L^2}$$

$$\max \left\{ \sigma_{\max}(P_{\bar{M}_{\Theta}}P_{\bar{M}_F}), \sigma_{\max}(P_{M_{\Theta}}P_{\bar{M}_F}) \right\} \leq \frac{16}{L^2} \text{ Corollary III.2. For both the Toeplitz and non Toeplitz cases, the Robust KronPCA covariance estimates } \Theta \text{ and } \Gamma \text{ are symmetric with probability one.}$$
where

$$\Lambda = 2 + \max \left\{ \frac{3\sqrt{2}\eta}{\lambda_s}, \frac{3\sqrt{\beta}}{\beta \lambda_s} \right\}$$  \hspace{1cm} (19)$$

$P_{\mathcal{M}_y}$ is the matrix corresponding to the projection operator onto the subspace $\mathcal{M}_y$.

By way of interpretation, note that the maximum singular value of the product of projection matrices is essentially measuring the “angle” between the subspaces. Hence, the incoherence condition is essentially requiring that the subspaces based on $\Theta$ and $\Gamma$ live are sufficiently “orthogonal” to each other i.e. “incoherent.” This ensures identifiability in the sense that a portion of $\Gamma$ (a portion of $\Theta$) cannot be well approximated by adding a small number of additional terms to the Kronecker factors of $\Theta$ ($\Gamma$). Thus $\Theta$ cannot be sparse and $\Gamma$ cannot have low separation rank. In [12] it was noted that this incoherency condition is significantly weaker than many typically imposed.

Further suppose that in the robust KronPCA model, that the $n$ training samples are multivariate Gaussian distributed and IID. Suppose further that $L$ is at most rank $r$ and $S$ has $s$ nonzero entries. Choose the regularization parameters to be

$$\lambda_\Theta = k\|\Sigma\| \max(\alpha^2, \alpha), \quad \lambda_\Gamma = 32\rho(\Sigma) \sqrt{\frac{\log p_t p_s}{n}} \hspace{1cm} (20)$$

where $\rho(\Sigma) = \max_{ij} \Sigma_{ij}$, $M = \max(p_t, p_s, n)$, and $k$ is smaller than an increasing function of $t_0$ given in the proof (35).

Given these assumptions, we have the following bound on the estimation error.

**Theorem IV.1 (Robust KronPCA).** Assume that the incoherence assumption (18) and the assumptions in the previous paragraph hold, and the regularization parameters are given by Equation (20) with $\alpha = \sqrt{t_0(p_t^2 + p_s^2 + \log M)/n}$ for any chosen $t_0 > 1$. Then the Frobenius norm error of the solution to the optimization problem (8) is bounded as:

$$\|\hat{L} - L^*\|_F \leq \frac{3}{\kappa} \max \left\{ k\|\Sigma\| \sqrt{r \max(\alpha^2, \alpha), 32\rho(\Sigma)} \sqrt{\frac{\log p_t p_s}{n}} \right\} \hspace{1cm} (21)$$

with probability at least $1 - c \exp(-c_0 \log p_t p_s)$, $c, c_0$ constants, $c_0$ dependent on $t_0$ but bounded from above by an absolute constant.

Next, we derive a similar bound for the Toeplitz Robust KronPCA estimator.

It is easy to show that a decomposition of $\Theta$ of the form (17) exists where all the $T_i$ are Toeplitz. Hence the definitions of the relevant subspaces are apparent. In the Gaussian robust Toeplitz KronPCA model (Equation (9) using the parameters in (14)), further suppose $L$ is at most rank $r$ and $S$ has $s$ nonzero entries.

**Theorem IV.2 (Toeplitz Robust KronPCA).** Assume that the incoherence assumption (18) and the assumptions of the previous paragraph hold, and the regularization parameters are as in (20) with $\alpha = \sqrt{t_0(2p_t + p_s^2 + \log M)/n}$ for any $t_0 > 1$. Then the Frobenius norm error of the solution to the Toeplitz Robust KronPCA optimization problem (8) with coefficients given in (14) is bounded as:

$$\|\hat{L} - L^*\|_F \leq \frac{3}{\kappa} \max \left\{ k\|\Sigma\| \sqrt{r \max(\alpha^2, \alpha), 32\rho(\Sigma)} \sqrt{\frac{\log p_t p_s}{n}} \right\} \hspace{1cm} (22)$$

and since $L = P^T \tilde{L}$

$$\|\hat{L} - L^*\|_F \leq \frac{3}{\kappa} \max \left\{ k\|\Sigma\| \sqrt{r \max(\alpha^2, \alpha), 32\rho(\Sigma)} \sqrt{\frac{\log p_t p_s}{n}} \right\} \hspace{1cm} (23)$$

with probability at least $1 - c \exp(-c_0 \log p_t p_s)$, $c, c_0$ constants.

Hence the resulting improvement (as expected from the reduction in the number of temporal parameters) is essentially that $p_t^2$ is replaced by $2p_t$.

The proof of these theorems is deferred to Appendix A.

V. Results

A. Simulations

In this section we evaluate the performance of the proposed algorithms by conducting mean squared covariance estimation error simulations. For the first simulation, we consider a covariance that is a sum of 3 kronecker products ($p_t = 50, p_s = 10$), with each term being a Kronecker product of two autoregressive (AR) covariances. To create a covariance matrix following the “KronPCA plus sparse” model, we create a new “corrupted” covariance by taking the 3 term AR covariance and deleting a random set of row/column pairs (in the non Toeplitz case), adding a diagonal term, and sparsely adding high correlations (whose magnitude depends on the distance to the diagonal) at random locations. Figure 1 shows the uncorrupted covariance and an example corrupted version. For our experiments, we have the sparse term to be on the same mean-square order as the original to better illustrate the balance between accuracy estimating the sparse term and the low separation rank term.

Figures 2 and 3 show results for estimating the Toeplitz corrupted covariance and non Toeplitz corrupted covariance respectively, each with the appropriate form of Robust KronPCA. For both simulations, the average MSE of the covariance that is a sum of 3 kronecker products ($p_t = 50, p_s = 10$), with each term being a Kronecker product of two autoregressive (AR) covariances. To create a covariance matrix following the “KronPCA plus sparse” model, we create a new “corrupted” covariance by taking the 3 term AR covariance and deleting a random set of row/column pairs (in the non Toeplitz case), adding a diagonal term, and sparsely adding high correlations (whose magnitude depends on the distance to the diagonal) at random locations. Figure 1 shows the uncorrupted covariance and an example corrupted version. For our experiments, we have the sparse term to be on the same mean-square order as the original to better illustrate the balance between accuracy estimating the sparse term and the low separation rank term.

Figures 2 and 3 show results for estimating the Toeplitz corrupted covariance and non Toeplitz corrupted covariance respectively, each with the appropriate form of Robust KronPCA. For both simulations, the average MSE of the covariance estimate is computed for a range of Gaussian training sample sizes. Average 3-ahead prediction MSE using the learned covariance to form the predictor coefficients ($\Sigma_{yx} \hat{\Sigma}_{xx}$) is shown in Figure 4. The regularization parameters are set at $10^5$ samples and the curves are formed by adjusting them according to the relevant theorems. The choices of regularization parameters shown were selected via cross validation for best performance in the appropriate region. Note the significant gains achieved using regularization, and the effectiveness of using the regularization parameter formulas derived in the theorems, particularly for achieving consistency. In addition, note that separation rank penalization alone does not maintain
dominance over the unregularized (SCM) estimate in the high sample regime, whereas the full Robust KronPCA method maintains a consistent advantage (as predicted by the theorems).

Fig. 1. Covariance used for the MSE simulations, along with the corrupted version used in the first experiment. Subfigures (top half, clockwise starting from upper left): Original three term AR covariance; its KronPCA and PCA spectra; and the first spatial and temporal factors of the original covariance. (Bottom half, clockwise from upper left): covariance with sparse corruptions; its KronPCA, Robust KronPCA, and PCA spectra; and the first (non robust) spatial and temporal factors of the corrupted covariance. Note how the corruption spreads the KronPCA spectrum and the significant corruption of the first Kronecker factor in the standard KronPCA estimate.

B. Cell Cycle Modeling

As a real data application, we consider the yeast (S. cerevisiae) metabolic cell cycle dataset used in [13]. The dataset consists of 9335 gene probes sampled approximately every 24 minutes for a total of 36 time points, and about 3 complete cell cycles [13].

In [13], it was found using a variety of algorithms that many of the gene expression intensities exhibit periodic behavior in the dataset due to the periodic cell cycle. Our goal is to establish that periodicity can also be detected in the temporal component of the KronPCA spatio-temporal correlation model for the dataset. We use $p_T = 36$ so only one spatio-temporal training sample is available. Due to their high dimensionality, the spatial factor estimates have very low accuracy, but the first few temporal $T_i$ factors ($36 \times 36$) can be effectively estimated (verified low variance via bootstrapping) due to the large number of spatial variables and non excessive spatial correlations. We learn the spatiotemporal covariance (both space and time factors) using block toeplitz robust KronPCA and then analyze the estimated Toeplitz time factors ($T_i$) to discover periodicity. The Toeplitz (stationary) assumption can be made due to the regular sampling of the time series. This allows us to consider the overall periodicity of the gene set, taking into account relationships between the genes, as opposed to the univariate analysis in [13]. The sparse correction to the covariance allows for the partial or complete removal of genes and correlations that are outliers in terms of their temporal behavior from the estimation of the overall temporal behavior.

Figure 5 shows the quantiles of the empirical distribution of the entries of the sample covariance versus those of the normal distribution. The extremely heavy tails motivate the use of a sparse correction term as opposed to the purely $l_2$ approach of standard KronPCA. The results showing plots of the first row of each Toeplitz temporal factor estimates are shown in Figure 6. On the left, the first three factors are shown when the entire 9335 gene dataset is used to create the sample covariance. Note the strong temporal periodicity
were derived. Finally, simulations and experiments with yeast cell cycle data were performed that demonstrate advantages of our methods, both relative to the sample covariance and relative to the standard (without sparse correction) KronPCA.

APPENDIX A

DERIVATION OF HIGH dimensional consistency

In this section we first prove Theorem VI.1 i.e. the bound for the non Toeplitz case.

A general theorem for decomposable regularization of this type was proven in [12]. In [12] the theorem was applied to Robust PCA directly on the sample covariance, hence we follow a similar approach in our proof for the Robust KronPCA case.

Consider the more general M-estimation problem

$$
\min_{(\theta_\alpha)_{\alpha \in I}} \mathcal{L}(\sum_{\alpha \in I} \theta_\alpha) + \sum_{\alpha \in I} \lambda_\alpha \mathcal{R}_\alpha(\theta_\alpha),
$$

where $\mathcal{R}_\alpha$ are regularizers, with regularization parameters $\lambda_\alpha \geq 2\mathcal{R}_\alpha^*(\nabla_{\theta_\alpha} \mathcal{L}(\theta_\alpha; X))$. $\mathcal{R}_\alpha^*$ is the dual norm of the norm $\mathcal{R}_\alpha$. Let $\mathcal{M}_\alpha$ be the model subspace associated with the constraints enforced by $\mathcal{R}_\alpha$ [12]. Assume the following conditions are satisfied:

1) The loss function $\mathcal{L}$ is convex and differentiable.
2) Each norm $\mathcal{R}_\alpha$ ($\alpha \in I$) is decomposable with respect to the subspace pairs $(\mathcal{M}_\alpha, \mathcal{M}_\alpha^+)$, where $\mathcal{M}_\alpha \subseteq \mathcal{M}_\alpha^+$.
3) (Restricted Strong Convexity) For all $\Delta \in \Omega_\alpha$, where $\Omega_\alpha$ is the parameter space for parameter component $\alpha$,

$$
\delta \mathcal{L}(\Delta_\alpha; \theta^*) := \mathcal{L}(\theta^* + \Delta_\alpha) - \mathcal{L}(\theta^*) - \langle \Delta_\alpha, \mathcal{L}(\theta^*) \rangle \geq & \mathcal{R}_\alpha^* \|\Delta_\alpha\|^2 - g_\alpha \mathcal{R}_\alpha^2(\Delta_\alpha),
$$

where $\mathcal{R}_\alpha$ is a “curvature” parameter, and $g_\alpha \mathcal{R}_\alpha^2(\Delta_\alpha)$ is a “tolerance” parameter.

4) (Structural Incoherence) For all $\Delta_\alpha \in \Omega_\alpha$,

$$
|\mathcal{L}(\theta^* + \sum_{\alpha \in I} \Delta_\alpha) + \sum_{\alpha \in I} \mathcal{L}(\theta^* + \Delta_\alpha)| \leq \frac{\mathcal{R}_\alpha^*}{2} \sum_{\alpha \in I} \|\Delta_\alpha\|^2 + \sum_{\alpha \in I} h_\alpha \mathcal{R}_\alpha^2(\Delta_\alpha).
$$

Define the subspace compatibility constant as $\Psi_\alpha(\mathcal{M}, \|\cdot\|) := \sup_{u \in \mathcal{M} \setminus \{0\}} \mathcal{R}_\alpha(u) \|u\|$. Given these assumptions, the following theorem holds (Corollary 1 in [12]):
Theorem A.1. Suppose that the subspace-pairs are chosen such that $\theta^* \in \mathcal{M}_\Theta$. Then the parameter error bounds are given as:
\[
\|\hat{\theta} - \theta^*\| \leq \left(\frac{3I}{2\kappa}\right) \max_{\alpha \in \ell} \lambda_\alpha \Psi_\alpha(\mathcal{M}_\alpha),
\]
where
\[
\kappa := \frac{\kappa_{\mathcal{E}}}{2} - 32\gamma |I| \left(\max_{\alpha \in \ell} \lambda_\alpha \Psi_\alpha(\mathcal{M}_\alpha)\right)^2
\]
\[
\gamma := \max_{\alpha \in \ell} \frac{1}{\lambda_\alpha} \sqrt{g_\alpha + h_\alpha}.
\]

We can now prove Theorem IV.1.

Proof of Theorem IV.1: To apply Theorem A.1 to the KronPCA estimation problem, we first check the conditions. In our objective function (9), we have a loss $L(\Sigma) = \|B - \Sigma\|_F^2$, which of course satisfies condition 1. It was shown in [12] that the nuclear norm and the $\ell_1$ norm both satisfy Condition 2. The restricted strong convexity condition (Condition 3) holds trivially with $\kappa_{\mathcal{E}} = 1$ and $g_\alpha = 0$ [12].

It was shown in [12] that for the linear L2 loss function in [9], the following simpler structural incoherence condition implies Condition 4:
\[
\max \left\{ \sigma_{\max}(P \mathcal{M}_\Theta, P \mathcal{M}_\Gamma), \sigma_{\max}(P \mathcal{M}^\perp_\Theta, P \mathcal{M}^\perp_\Gamma) \right\} \leq \frac{1}{16\Lambda^2}
\]
where $\Lambda = \max_{\gamma, \gamma^2} \left\{ 2 + \frac{3\gamma_1 \gamma_2 \Psi_{\gamma, \gamma^2}(\mathcal{M}_{\gamma, \gamma^2})}{\gamma_1 \gamma_2} \right\}$.

The subspace compatibility constant remains the same for the nuclear norm, and for the sparse case we have for all $\Delta$
\[
\mathcal{R}_\Gamma(\Delta) \leq \|\Delta\|_1
\]
hence, the supremum under the L1 norm is greater than the supremum under the row weighted norm. Thus, the subspace compatibility constant is still less than or equal to $\sqrt{s}$, where $s$ is now the number of nonzero entries in $\mathcal{M}_\Gamma$. The first follows from the fact that for all $\Theta \in \mathcal{M}_\Theta$, $\text{rank}(\Theta) \leq 2r$ since both the row and column spaces of $\Theta$ must be of rank $r$ [12].

Finally, we need to show that the regularization parameters satisfy
\[
\lambda_\Theta \geq 2\mathcal{R}_{\Theta}(\nabla_{\Theta, \alpha} L(\theta^*; X))
\]
with high probability. Since the L1 norm is invariant under rearrangement, the argument from [12] still holds and we have that
\[
\lambda_\Theta = 32\rho(\Sigma) \sqrt{\frac{\log p_t p_s}{n}}
\]
satisfies (30) with probability at least $1 - 2 \exp(-c_2 \log p_s p_t)$.

For the low rank portion, (30) will hold if [12]
\[
\lambda_\theta \geq 4\|\mathcal{R}((\Sigma_{\mathcal{SCM}} - \Sigma))\|.
\]

From [3] we have that for $t_0 \geq f(\epsilon) = 4C \log(1 + \frac{\epsilon}{2})$ ($C$ absolute constant given in [3]), $C$ an absolute constant, and $\alpha^2 \geq \alpha$
\[
\|\mathcal{R}((\Sigma_{\mathcal{SCM}} - \Sigma))\| \leq \frac{\|\Sigma\| p_t^2 + p_s^2 + \log M}{1 - 2\epsilon}
\]
with probability at least $1 - 2M^{-t_0/4C}$ and otherwise
\[
\|\mathcal{R}((\Sigma_{\mathcal{SCM}} - \Sigma))\| \leq \frac{\|\Sigma\| \sqrt{t_0} \sqrt{p_t^2 + p_s^2 + \log M}}{1 - 2\epsilon}
\]
with probability at least $1 - 2M^{-t_0/4C}$. Thus our choice of $\lambda_\theta$ satisfies (32) with high probability. To satisfy the constraints on $t$, we need $t_0 > f^2(\epsilon)$. Clearly, $\epsilon$ can be adjusted to satisfy the constraint and
\[
k = 4/(1 - 2\epsilon).
\]

Recalling the sparsity probability $1 - 2 \exp(-c_2 \log p_t p_s)$, by the union bound (30) is satisfied for both regularization parameters with probability at least $1 - 2 \exp(-c_2 \log p_t p_s) - 2 \exp(-t_0/4C) \log M \geq 1 - c \exp(-c_0 \log p_t p_s)$ and the proof of Theorem IV.1 is complete.

Next, we present the proof for Theorem IV.2, emphasizing only the parts that differ from the non-Toeplitz proof of Theorem IV.1 for brevity, since much of the proof for the previous theorem carries over to the Toeplitz case.

Proof of Theorem IV.2: Condition 1 still holds as in the general non-Toeplitz case, and Condition 2 holds because $\mathcal{R}_\Gamma$ is a positively weighted sum of norms, forming a norm on the product space (which is clearly the entire space). $\mathcal{R}_\Gamma$ is decomposable because the L1 norm is decomposable and the overall model subspace is the product of the model subspaces for each row. The remaining two conditions trivially remain the same from the non Toeplitz case.

The subspace compatibility constant remains the same for the nuclear norm, and for the sparse case we have for all $\Delta$
\[
\mathcal{R}_\Gamma(\Delta) \leq \|\Delta\|_1
\]
with probability at least $1 - 2M^{-t_0/4C}$ and otherwise
\[
\|\mathcal{R}((\Sigma_{\mathcal{SCM}} - \Sigma))\| \leq \frac{\|\Sigma\| \sqrt{t_0} \sqrt{p_t^2 + p_s^2 + \log M}}{1 - 2\epsilon}
\]
with probability at least $1 - 2M^{-t_0/4C}$. Thus our choice of $\lambda_\theta$ satisfies (32) with high probability. To satisfy the constraints on $t$, we need $t_0 > f^2(\epsilon)$. Clearly, $\epsilon$ can be adjusted to satisfy the constraint and
\[
k = 4/(1 - 2\epsilon).
\]

Recalling the sparsity probability $1 - 2 \exp(-c_2 \log p_t p_s)$, by the union bound (30) is satisfied for both regularization parameters with probability at least $1 - 2 \exp(-c_2 \log p_t p_s) - 2 \exp(-t_0/4C) \log M \geq 1 - c \exp(-c_0 \log p_t p_s)$ and the proof of Theorem IV.1 is complete.

We now show that the regularization parameters chosen satisfy (30) with high probability. For the sparse portion, we need to find the dual of $\mathcal{R}_\Gamma$, defined as
\[
\mathcal{R}_\Gamma^*(Z) = \sup \{ \langle (Z, X) \rangle | \mathcal{R}_\Gamma(X) \leq 1 \}
\]
where $\langle (Z, X) \rangle = \text{trace} \{Z^T X\}$. Let the matrix $P_1 = \text{diag}\{ \sqrt{p_t - |j|} \}_{j=-p_t+1}^{p_t-1}$. Define the matrices $X'$ such that $X' = P_1^{-1} X$. Then $\mathcal{R}_\Gamma(X) = \|X\|_1$ and
\[
\mathcal{R}_\Gamma^*(Z) = \sup \{ \langle P_1 Z, X' \rangle | \|X'\|_1 \leq 1 \}
\]

since the dual of the $\ell_1$ norm is the $\ell_\infty$ norm. From [14], (30) now takes the form
\[
\lambda_\Gamma \geq 4\|P_1 P W\|_\infty = \|\tilde{W}\|_\infty
\]
where
\[
\tilde{W}_{j+p_t, i} = \sum_{\ell \in K(j)} W_{\ell, i}.
\]
Hence
\[ |\tilde{W}_{i+p,t} - p_t| \leq (p_t - n)|W| \] (41)

\[ \|W\|_\infty \leq p_t |W|_\infty \]

From (41) (via the union bound), we have
\[ \Pr \left( \|W\|_\infty > \rho(\Sigma) \sqrt{\frac{\log(p_t p_s)}{n}} \right) \leq 2 \exp(-c_2 \log(p_t p_s)) \] (42)
giving
\[ \Pr \left( \|\hat{W}\|_\infty > \rho(\Sigma) p_t \sqrt{\frac{\log(p_t p_s)}{n}} \right) \leq 2 \exp(-c_2 \log(p_t p_s)) \] (43)
which demonstrates that our choice for \( \lambda_f \) is satisfactory with high probability.

For the low rank portion, we use the following corollary based on an extension of a theorem in [3] to the Toeplitz case:

**Corollary A.2.** Suppose \( \Sigma_n \) is a \( p_t p_s \) \times \( p_t p_s \) covariance matrix, \( \|\Sigma_0\|_2 \) is finite for all \( p_t, p_s \), and let \( M = \max(p_t, p_s, n) \).

Let \( \epsilon' < 0.5 \) be fixed and assume that \( t_0 \geq f(\epsilon) \) and \( C = \max(C_1, C_2) > 0 \). We have that
\[ \|\Delta_n\|_2 \leq \frac{\|\Sigma_0\|_2}{1 - 2\epsilon'} \max \left\{ t_0 \alpha^2, \sqrt{t_0} \alpha \right\} \] (44)
with probability at least \( 1 - 2M^{-\frac{t_0}{4\alpha^2}} \), where
\[ \alpha = \frac{2p_t + p_s^2 + \log M}{n} \] (45)
We prove this in Appendix B.
As before, for the low rank portion (30) will hold if
\[ \lambda_\theta \geq 4\|\mathcal{R}(\Sigma_{SCM} - \Sigma)\| \] (46)
From Corollary A.2 we have that for \( t \geq f(\epsilon) \), \( C \) an absolute constant, and \( \alpha^2 \geq \alpha \)
\[ \|\mathcal{R}(\Sigma_{SCM} - \Sigma)\| \leq \frac{\|\Sigma\|_2 t_0}{1 - 2\epsilon'} \] (47)
with probability at least \( 1 - 2M^{-t_0/4C} \) and otherwise
\[ \|\mathcal{R}(\Sigma_{SCM} - \Sigma)\| \leq \frac{\|\Sigma\|_2 \sqrt{t_0}}{1 - 2\epsilon'} \] (48)
with probability at least \( 1 - 2M^{-t_0/4C} \). Hence, the theorem follows in the same way as in the non Toeplitz case.

**APPENDIX B**

**GAUSSIAN CHAOS OPERATOR NORM BOUND**

We first note the following corollary from [3]:

**Corollary B.1.** Let \( \mathbf{x} \in S_{p_t-1} \) and \( \mathbf{y} \in S_{p_s-1} \). Let \( \mathbf{z}_i \sim \mathcal{N}(0, \Sigma_0) \) \( i = 1, \ldots, n \) be \( p_t p_s \) dimensional iid training samples. Let \( \Delta_n = \mathcal{R}(\frac{1}{n} \sum \mathbf{z}_i \mathbf{z}_i^T - \Sigma_0) \). Then for all \( \tau > 0 \),
\[ \Pr(|\mathbf{x}^T \Delta_n \mathbf{y}| \geq \tau) \leq \exp \left( \frac{-n\tau^2}{2} \frac{1}{C_1 \|\Sigma_0\|_2^2 + C_2 \|\Sigma_0\|_2 \|\Sigma_0\|_2} \right) \] (49)
where \( C_1, C_2 \) are absolute constants.

The proof (appropriately modified from that of a similar theorem in [3]) of Corollary A.2 then proceeds as follows:

**Proof:** Define \( \mathcal{N}(S_{p_t-1}, \epsilon') \) as an \( \epsilon' \) net on \( S_{p_t-1} \). Choose \( \mathbf{x}_1 \in S_{2p_t-2}, \mathbf{y}_1 \in S_{p_s-1} \) such that \( |\mathbf{x}_1^T \mathbf{P} \Delta_n \mathbf{y}_1| = \|\mathbf{P} \Delta_n\|_2 \). By definition, there exists \( \mathbf{x}_2 \in \mathcal{N}(S_{2p_t-2}, \epsilon'), \mathbf{y}_2 \in \mathcal{N}(S_{p_s-1}, \epsilon') \) such that \( |\mathbf{x}_1 - \mathbf{x}_2|_2 \leq \epsilon', |\mathbf{y}_1 - \mathbf{y}_2|_2 \leq \epsilon' \). Then
\[ |\mathbf{x}_1^T \mathbf{P} \Delta_n \mathbf{y}_1 - |\mathbf{x}_2^T \mathbf{P} \Delta_n \mathbf{y}_2| \leq |\mathbf{x}_1^T \mathbf{P} \Delta_n \mathbf{y}_1 - \mathbf{x}_1^T \mathbf{P} \Delta_n \mathbf{y}_2| \] (50)
\[ \leq 2 \epsilon' \|\mathbf{P} \Delta_n\|_2 \]

We then have
\[ \|\mathbf{P} \Delta_n\|_2 (1 - 2\epsilon') \]
\[ \leq \max \left\{ |\mathbf{x}_1^T \mathbf{P} \Delta_n \mathbf{y}_2| : \mathbf{x}_2 \in \mathcal{N}(S_{2p_t-2}, \epsilon'), \mathbf{y}_2 \in \mathcal{N}(S_{p_s-1}, \epsilon') \right\} \] (51)
\[ \leq \max \left\{ |\mathbf{x}_1^T \mathbf{P} \Delta_n \mathbf{y}_2| : \mathbf{x}_2 \in \mathcal{N}(S_{2p_t-2}, \epsilon'), \mathbf{y}_2 \in \mathcal{N}(S_{p_s-1}, \epsilon') \right\} \] (52)

\[ \Pr(\|\mathbf{P} \Delta_n\|_2 > \epsilon') \]
\[ \leq \frac{1}{1 - 2\epsilon'} \max \left\{ x \in \mathcal{N}(S_{2p_t-2} - \epsilon'), y \in \mathcal{N}(S_{p_s-1} - \epsilon') \right\} |\mathbf{x}^T \mathbf{P} \Delta_n \mathbf{y}| \] (53)
\[ \Pr(\|\mathbf{P} \Delta_n\|_2 > \epsilon') \]
\[ \leq \frac{1}{1 - 2\epsilon'} \max \left\{ x \in \mathcal{N}(S_{2p_t-2} - \epsilon'), y \in \mathcal{N}(S_{p_s-1} - \epsilon') \right\} |\mathbf{x}^T \mathbf{P} \Delta_n \mathbf{y}| \] (54)
\[ \leq \frac{1}{1 - 2\epsilon'} \max \left\{ x \in \mathcal{N}(S_{2p_t-2} - \epsilon'), y \in \mathcal{N}(S_{p_s-1} - \epsilon') \right\} |\mathbf{x}^T \mathbf{P} \Delta_n \mathbf{y}| \] (55)
\[ \leq \frac{1}{1 - 2\epsilon'} \max \left\{ x \in \mathcal{N}(S_{2p_t-2} - \epsilon'), y \in \mathcal{N}(S_{p_s-1} - \epsilon') \right\} |\mathbf{x}^T \mathbf{P} \Delta_n \mathbf{y}| \] (56)
\[ \leq 2 \left( 1 + \frac{2}{\epsilon'} \right)^2 \exp \left( \frac{-n\epsilon^2 (1 - 2\epsilon')^2}{2} \right) \]

Note that
\[ |\mathbf{x}^T \mathbf{P}|_2^2 = \sum \frac{x_j^2 + p_t}{p_t - |j|} \] (55)
\[ = \sum x_j^2 = |\mathbf{x}|_2^2 = 1 \]
so \( \mathbf{x}^T \mathbf{P} \in S_{p_t-1} \). We can thus use Corollary B.1 giving
\[ \Pr(\|\mathbf{P} \Delta_n\|_2 > \epsilon') \]
\[ \leq 2 \left( 1 + \frac{2}{\epsilon'} \right)^2 \exp \left( \frac{-n\epsilon^2 (1 - 2\epsilon')^2}{2} \right) \]
Two regimes emerge from this expression. The first is where 
\[ \epsilon \leq \frac{C_1 \| \Sigma_0 \|_2}{C_2 (1 - 2 \epsilon^2)} \] which allows
\[
\Pr \left( \| P \Delta_n \|_2 > \epsilon \right) \leq 2 \left( 1 + \frac{2}{\epsilon} \right)^{2p_t + p_s^2} \exp \left( -n \epsilon^2 (1 - 2 \epsilon^2)^2 / 2C_1 \| \Sigma_0 \|_2^2 \right)
\] (57)
Choose
\[ \epsilon = \sqrt{t_0} \| \Sigma_0 \|_2 \sqrt{\frac{2p_t + p_s^2 + \log M}{n}} \] (58)
This gives:
\[
\Pr \left( \| P \Delta_n \|_2 > \sqrt{t_0} \| \Sigma_0 \|_2 \sqrt{\frac{2p_t + p_s^2 + \log M}{n}} \right) \leq 2 \left( 1 + \frac{2}{\epsilon} \right)^{2p_t + p_s^2} \exp \left( -t^2 (2p_t + p_s^2 + \log M) / 4C_1 \right)
\] (59)
\[
\leq 2 \left( 1 + \frac{2}{\epsilon} \right)^{2p_t + p_s^2} e^{-t_0} M^{-t_0/(4C_1)} \leq 2M^{-t_0/(4C_1)}
\]
The second regime \( \epsilon > \frac{C_1 \| \Sigma_0 \|_2}{C_2 (1 - 2 \epsilon^2)} \) allows us to set \( \epsilon \) to
\[
\epsilon = \frac{t_0 \| \Sigma_0 \|_2}{1 - 2 \epsilon^2} \frac{2p_t + p_s^2 + \log M}{n} \] (60)
which gives
\[
\Pr \left( \| P \Delta_n \|_2 > \frac{t_0 \| \Sigma_0 \|_2}{1 - 2 \epsilon^2} \frac{2p_t + p_s^2 + \log M}{n} \right) \leq 2 \left( 1 + \frac{2}{\epsilon} \right)^{2p_t + p_s^2} \exp \left( -t^2 (2p_t + p_s^2 + \log M) / 4C_2 \right)
\] (61)
\[
\leq 2 \left( 1 + \frac{2}{\epsilon} \right)^{2p_t + p_s^2} e^{-t_0} M^{-t_0/(4C_2)} \leq 2M^{-t_0/(4C_2)}
\]
Combining both regimes (noting that \( t_0 > 1 \) and \( \sqrt{t_0} C_1 / C_2 > 1 \)) completes the proof.

REFERENCES

[1] K. Greenewald, T. Tsiligkaridis, and A. Hero, “Kronecker sum decompositions of space-time data,” in Proceedings of IEEE CAMSAP, 2013.
[2] K. Greenewald and A. Hero, “Regularized block toeplitz covariance matrix estimation via kronecker product expansions,” in Proceedings of IEEE SSP, 2014.
[3] T. Tsiligkaridis and A. Hero, “Covariance estimation in high dimensions via kronecker product expansions,” IEEE Trans. on Sig. Proc. 61(21), pp. 5347–5360, 2013.
[4] T. Tsiligkaridis, A. Hero, and S. Zhou, “On convergence of kronecker graphical lasso algorithms.” IEEE Trans. Signal Proc. 61(7), pp. 1743–1755, 2013.
[5] K. Greenewald and A. Hero, “Kronecker pca based spatio-temporal modeling of video for dismount classification,” in Proceedings of SPIE, 2014.
[6] C. V. Loan and N. Pitsianis, “Approximation with kronecker products,” in Linear Algebra for Large Scale and Real Time Applications, pp. 293–314, Kluwer Publications, 1993.
[7] K. Werner, M. Jansson, and P. Stoica, “On estimation of cov. matrices with kronecker product structure,” IEEE Trans. on Sig. Proc. 56(2), pp. 478–491, 2008.
[8] J. Karrn and J. Nagy, “Opt. kronecker product approx. of block toeplitz matrices,” SIAM Journal on Matrix Analysis and App. 22(1), pp. 155–172, 2000.