Kronecker Covariance Sketching for Spatial-Temporal Data

Yuejie Chi
Department of Electrical and Computer Engineering
The Ohio State University, Columbus, Ohio 43210
Email: chi.97@osu.edu

***Abstract*—Covariance sketching has been recently introduced as an effective strategy to reduce the data dimensionality without sacrificing the ability to reconstruct second-order statistics of the data. In this paper, we propose a novel covariance sketching scheme with reduced complexity for spatial-temporal data, whose covariance matrices satisfy the Kronecker product expansion model recently introduced by Tsiligkaridis and Hero. Our scheme is based on quadratic sampling that only requires magnitude measurements, hence is appealing for applications when phase information is difficult to obtain, such as wideband spectrum sensing and optical imaging. We propose to estimate the covariance matrix based on convex relaxation when the separation rank is small, and when the temporal covariance is additionally Toeplitz structured. Numerical examples are provided to demonstrate the effectiveness of the proposed scheme.**

***Index Terms*—spatial-temporal data modeling, covariance sketching, kronecker product, convex optimization**

I. Introduction

Effective statistical inference of high-dimensional data sets plays a critical role in modern scientific and knowledge discovery. However, the volume and velocity of the data generated often overwhelms the storage, bandwidth and computational capabilities of the sensing platforms, making it challenging to perform inference within the prescribed sensing budget.

Recently, covariance sketching has been proposed [1]–[7] as an effective strategy to reduce the data dimensionality without sacrificing the ability to recover covariance information. This is motivated by the observation that, the covariance matrix often leads to a sufficient statistic to perform the inference task of interest, e.g. signal detection and parameter estimation. These schemes often first apply a reduced-dimension linear operator to each of the temporal data vectors, also known as *sketching*, and then estimate the covariance matrix using the sketches that can be stored in much smaller dimension than that of the covariance matrix, whose complexity does not scale with the increase of the number of data snapshots. The effectiveness of covariance sketching lies upon the availability of parsimonious representations of the covariance matrices, such that they can be recovered from a much smaller number of measurements than the ambient dimension. Example structures that have been exploited include positive semidefiniteness, sparsity, low-rankness, and Toeplitz. Notably, no parsimonious assumptions such as sparsity are necessary for the data sets themselves.

This work is supported in part by NSF under the grants CCF-1422966, ECCS-1462191 and by AFOSR under the grant FA9550-15-1-0205.

High-dimensional spatial-temporal data is ubiquitous and can be used to model climate data, gene expression data, network traffic, wideband power spectrum [8], and so on. Due to the spatial-temporal correlations, it is beneficial to consider the covariance matrices of *spatial-temporal data blocks*, which can be modeled as a *Kronecker product* of two smaller positive-semidefinite (PSD) matrices [9], [10] – one corresponding to the spatial correlations, and the other corresponding to the temporal correlations. This greatly reduces the number of parameters to describe the covariance matrix. Recently, this model is extended to a *Kronecker product expansion model* by Tsiligkaridis and Hero [11], [12], where the covariance matrix is a sum of Kronecker products of smaller PSD matrices, to better account for the variability of the data sets. The number of distinct Kronecker products in the expansion is known as the *separation rank* of the covariance matrix, which typically is very small.

In this paper, we propose novel covariance sketching schemes for high-dimensional spatial-temporal data whose covariance matrices satisfy the Kronecker product expansion model with a small separation rank. The existence of Kronecker product structures in the covariance allows for a more efficient design of the sketching schemes. One prominent benefit is to apply sketching simultaneously to the spatial dimension and the temporal dimension, i.e. on a spatial-temporal data block, with a bilinear rank-one operator, to reduce its computational cost. Specifically, for a $p \times q$ data block, the sketching cost is reduced from $O(pq)$ to $O(p+q)$. We only record the magnitudes of the sketches, which are nonnegative, and can be interpreted as energy projections of the data onto certain one-dimensional linear subspaces, which are easy to obtain using an energy detector in high-frequency applications. After aggregations, the sketches can be written as linear measurements of the covariance matrix with respect to *Kronecker-structured* rank-one measurement ensembles. One then wishes to recover the covariance matrix from a small number of sketches by exploiting the fact that it has a small separation rank. Recall that the covariance matrix satisfies the Kronecker product expansion model, using properties of the Kronecker product, we propose a convex optimization algorithm based on nuclear norm minimization as a surrogate for minimizing the separation rank [11]. In particular, we further tailor the reconstruction algorithm to the case when the temporal covariance is a Toeplitz PSD matrix

978-0-9928-6265-7/16/$31.00 ©2016 IEEE

[13] that accounts for stationarity. Finally, we demonstrate the effectiveness of the proposed sketching schemes with numerical simulations.

Our work is closely related to the covariance sketching scheme proposed by Chen et al. [1], but is tailored to the Kronecker structures of the covariance matrices to further reduce the sketching complexity. The utility of Kronecker products has also been studied in the compressive sensing literature [14], by assuming the sparsifying basis admits a Kronecker product representation. Another line of work considers covariance sketching using quantized measurements [15]–[18].

The rest of this paper is organized as follows. Section II introduces the spatial-temporal data and the covariance models with Kronecker products. Section III describes the proposed covariance sketching schemes and the corresponding reconstruction algorithms. Section IV provides numerical examples. Finally, we provide some additional discussions in Section V and conclude in Section VI.

II. Spatial-Temporal Data Model

We first describe the spatial-temporal data model, and then introduce the Kronecker product expansion model for its covariance structure that reduces the number of parameters.

A. Spatial-temporal data model

Consider a vector-valued random process $\{\boldsymbol{x}[t]\}_{t=0}^{\infty}$, where the entries of $\boldsymbol{x}[t] \in \mathbb{R}^p$ correspond to different spatial coordinates. Start by concatenating q consecutive temporal samples of $\boldsymbol{x}[t]$ together and denote

$$\boldsymbol{u}[t] = \left[\boldsymbol{x}[tq]^T, \boldsymbol{x}[tq+1]^T, \cdots, \boldsymbol{x}[tq+q-1]^T\right]^T \in \mathbb{R}^{pq}.$$

On the other hand, reshaping $\boldsymbol{u}[t]$ into a $p \times q$ matrix gives a spatial-temporal data block

$$\boldsymbol{U}[t] = \left[\boldsymbol{x}[tq], \boldsymbol{x}[tq+1], \cdots, \boldsymbol{x}[tq+q-1]\right] \in \mathbb{R}^{p\times q}, \quad (1)$$

where the columns correspond to the temporal dimension, and the rows correspond to spatial dimension. We have $\boldsymbol{u}[t] = \mathrm{vec}(\boldsymbol{U}[t])$, where $\mathrm{vec}(\cdot)$ denotes the vectorization operator. For spatial-temporal data, we're interested in the covariance matrix of $\{\boldsymbol{u}[t]\}_{t=0}^{\infty}$, which captures both spatial and temporal correlations. Define the covariance matrix of $\{\boldsymbol{u}[t]\}$ as

$$\boldsymbol{R}_{\boldsymbol{u}} = \mathbb{E}\left\{\boldsymbol{u}[t]\boldsymbol{u}[t]^T\right\} \in \mathbb{R}^{pq\times pq}. \quad (2)$$

Note that the parameter q shall be picked wisely to make sense of the data. For example, if $\boldsymbol{x}[t]$ is a cyclostationary random process, picking q as the period of $\boldsymbol{x}[t]$ will make the new process $\boldsymbol{u}[t]$ wide-sense stationary [19].

B. Kronecker Product (Expansion) Model

To describe $\boldsymbol{R}_{\boldsymbol{u}}$, we need $\Theta(p^2q^2)$ parameters which may be too large for high-dimensional data. Therefore, we consider models of $\boldsymbol{R}_{\boldsymbol{u}}$ that reduces its degree of freedom. The first is a *Kronecker product model* [20], where we have

$$\boldsymbol{R}_{\boldsymbol{u}} = \boldsymbol{D} \otimes \boldsymbol{E}, \quad (3)$$

where $\otimes$ denotes the Kronecker product, $\boldsymbol{D} \in \mathbb{R}^{q\times q}$ and $\boldsymbol{E} \in \mathbb{R}^{p\times p}$ are PSD matrices that correspond to the temporal and spatial correlations, respectively. This structure arises, and finds applications in many signal processing problems. The number of parameters to describe $\boldsymbol{R}_{\boldsymbol{u}}$ is now $\Theta(p^2+q^2)$, which is much smaller than $\Theta(p^2q^2)$.

An important extension of the Kronecker product model is recently proposed in [11], [12] with improved modeling powers. Specifically, $\boldsymbol{R}_{\boldsymbol{u}}$ is said to satisfy the *Kronecker product expansion model* if it can be written as a sum of Kronecker products:

$$\boldsymbol{R}_{\boldsymbol{u}} = \sum_{i=1}^{r} \boldsymbol{D}_i \otimes \boldsymbol{E}_i, \quad (4)$$

where r is called the separation rank [11] which is typically a small number, $\boldsymbol{D}_i \in \mathbb{R}^{q\times q}$ and $\boldsymbol{E}_i \in \mathbb{R}^{p\times p}$ are PSD matrices corresponding to the spatial and temporal correlation in the ith factor, $1 \le i \le r$. The number of parameters is now given as $\Theta(r(p^2+q^2))$. When $r = 1$, this model reduces to the Kronecker product model discussed above.

Define the permutation rearrangement operator $\mathcal{P}(\cdot) : \mathbb{R}^{pq\times pq} \mapsto \mathbb{R}^{p^2\times q^2}$:

$$[\mathcal{P}(\boldsymbol{\Phi})]_{pi_1+i_2, qj_1+j_2} = [\boldsymbol{\Phi}]_{pj_1+i_1, pj_2+i_2}, \quad (5)$$

where the subscripts denote the coordinate of the entries, for $0 \le i_1, i_2 \le p-1$, $0 \le j_1, j_2 \le q-1$. Using the definition of the Kronecker product, one can see that

$$\mathcal{P}(\boldsymbol{R}_{\boldsymbol{u}}) = \sum_{i=1}^{r} \mathrm{vec}(\boldsymbol{E}_i)\,\mathrm{vec}(\boldsymbol{D}_i)^T. \quad (6)$$

Therefore, if $\boldsymbol{R}_{\boldsymbol{u}}$ has a small separation rank, it can be transformed linearly through permutation and reshaping into a matrix with low rank in the usual sense.

Additionally, the factors may possess structures that are worth exploring and further reduce the degrees of freedom. Of particular interest is to incorporate the temporal stationarity [13], where we consider $\boldsymbol{D}_i$'s are Toeplitz matrices. We refer to this as *Kronecker product expansion model with Toeplitz enhancement*. Then, the number of parameters to describe $\boldsymbol{R}_{\boldsymbol{u}}$ is further reduced to $\Theta(r(p^2+q))$.

C. Motivating example

To demonstrate the applicability of the above models, consider the space-time data block in (1) as generated from the following factor model:

$$\boldsymbol{U}[t] = \sum_{i=1}^{r} \boldsymbol{h}_i[t]\boldsymbol{s}_i[t]^H, \quad (7)$$

where each factor $\boldsymbol{h}_i[t]$, $\boldsymbol{s}_i[t]$ are zero-mean and statistically uncorrelated. Then we have $\boldsymbol{u}[t] = \sum_{i=1}^{r} \boldsymbol{s}_i[t] \otimes \boldsymbol{h}_i[t]$, and the

covariance matrix $\boldsymbol{R_u}$ can be written as

$$\begin{aligned}\boldsymbol{R_u} &= \mathbb{E}\left\{\left(\sum_{i=1}^{r}\boldsymbol{s}_i[t]\otimes\boldsymbol{h}_i[t]\right)\left(\sum_{i=1}^{r}\boldsymbol{s}_i[t]\otimes\boldsymbol{h}_i[t]\right)^T\right\}\\ &= \sum_{i=1}^{r}\mathbb{E}\left\{(\boldsymbol{s}_i[t]\boldsymbol{s}_i[t]^T)\otimes(\boldsymbol{h}_i[t]\boldsymbol{h}_i[t]^T)\right\}\\ &= \sum_{i=1}^{r}\mathbb{E}\left\{\boldsymbol{s}_i[t]\boldsymbol{s}_i[t]^T\right\}\otimes\mathbb{E}\left\{\boldsymbol{h}_i[t]\boldsymbol{h}_i[t]^T\right\}\\ &= \sum_{i=1}^{r}\boldsymbol{D}_i\otimes\boldsymbol{E}_i,\end{aligned}$$

where $\boldsymbol{D}_i = \mathbb{E}\left\{\boldsymbol{s}_i[t]\boldsymbol{s}_i[t]^T\right\}$ and $\boldsymbol{E}_i = \mathbb{E}\left\{\boldsymbol{h}_i[t]\boldsymbol{h}_i[t]^T\right\}$. The second equality follows from the statistical uncorrelation between different factors, and the third equality follows from the statistical uncorrelation between $\boldsymbol{s}_i[t]$ and $\boldsymbol{h}_i[t]$.

III. Kronecker Covariance Sketching

We first propose a covariance sketching scheme for the Kronecker product expansion model, where the sketching vectors are applied simultaneously on the spatial and temporal dimensions to reduce computational costs. We then propose covariance recovery algorithms based on convex relaxation when the separation rank is small.

A. Kronecker covariance sketching

For high-dimensional spatial-temporal data, it may not be possible to acquire the full data due to sensing budget limitations in either time, storage or power. In what follows, we describe a one-pass sampling strategy that only takes a single nonnegative energy sketch for each spatial-temporal block $\boldsymbol{U}[t]$ (i.e. $\boldsymbol{u}[t]$). It is straightforward to extend the scheme when multiple sketches per $\boldsymbol{U}[t]$ can be acquired.

Define a set of d sketching vectors $\{\boldsymbol{a}_i\in\mathbb{R}^p, \boldsymbol{b}_i\in\mathbb{R}^q\}_{i=1}^d$ composed of i.i.d. standard Gaussian/Bernoulli entries. The sketching scheme is similar to that in [1], and constitutes of two steps. Given Tq samples of $\{\boldsymbol{x}[t]\}_{t=0}^{Tq-1}$, or equivalently T samples of $\{\boldsymbol{u}[t]\}_{t=0}^{T-1}$, we perform the following at each $t=0,\dots,T-1$:

1) *Sketching:* for each $\boldsymbol{U}[t]$, we first select a pair of sketching vectors $\{\boldsymbol{a}_{\ell_t},\boldsymbol{b}_{\ell_t}\}$, $\ell_t\in\{1,\dots,d\}$, in the cyclic order[1], take a quadratic sketch that measures only the magnitude of the inner product of $\boldsymbol{U}[t]$ with a bilinear rank-one matrix $\boldsymbol{a}_{\ell_t}\boldsymbol{b}_{\ell_t}^T$:
$$\begin{aligned}\tilde{y}_t = \left|\boldsymbol{a}_{\ell_t}^T\boldsymbol{U}[t]\boldsymbol{b}_{\ell_t}\right|^2 &= \left|(\boldsymbol{b}_{\ell_t}\otimes\boldsymbol{a}_{\ell_t})^T\boldsymbol{u}[t]\right|^2\\ &:= \left|\boldsymbol{z}_{\ell_t}^T\boldsymbol{u}[t]\right|^2,\end{aligned}\tag{8}$$
where $\boldsymbol{z}_i=\boldsymbol{b}_i\otimes\boldsymbol{a}_i$, $i=1,\dots,d$.
2) *Aggregation:* obtain the average of the sketches for each pair of sketching vectors, i.e.
$$\bar{y}_{i,T} = \frac{\sum_{t=0}^{T-1}\tilde{y}_t 1_{\{\ell_t=i\}}}{\sum_{t=0}^{T-1}1_{\{\ell_t=i\}}},$$
where $1_{\{\cdot\}}$ is the indicator function. The aggregation can be performed in an online fashion.

As T approaches infinity, the aggregates converge to the expectation of quadratic sketches, i.e.

$$\begin{aligned}y_i &= \mathbb{E}\left[\left|\boldsymbol{z}_i^T\boldsymbol{u}[t]\right|^2\right] = \boldsymbol{z}_i^T\mathbb{E}\left\{\boldsymbol{u}[t]\boldsymbol{u}[t]^T\right\}\boldsymbol{z}_i\\ &= \boldsymbol{z}_i^T\boldsymbol{R_u}\boldsymbol{z}_i := \langle\boldsymbol{R_u},\boldsymbol{Z}_i\rangle,\quad i=1,\dots,d.\end{aligned}\tag{9}$$

where $\boldsymbol{Z}_i=\boldsymbol{z}_i\boldsymbol{z}_i^T$ is again rank-one. Denote $\boldsymbol{y}=\{y_i\}_{i=1}^d$, we can represent (9) in a vector form as

$$\boldsymbol{y}=\mathcal{Q}(\boldsymbol{R_u}),\tag{10}$$

where $\mathcal{Q}(\cdot):\boldsymbol{R_u}\mapsto\{\langle\boldsymbol{R_u},\boldsymbol{Z}_i\rangle\}_{i=1}^d$. Importantly, the aggregated sketches $\boldsymbol{y}$ produces a set of *linear* measurements of the covariance matrix $\boldsymbol{R_u}$, even when the sketches are quadratic with respect to the data $\boldsymbol{u}[t]$. When T is finite, we can regard $\boldsymbol{y}$ as noisy linear measurements of $\boldsymbol{R_u}$, by adding a noise term into (9). In this paper, we only consider the noise-free case, and leave the study for finite T as future work.

Compared with the sketching scheme in [1], the new scheme applies sketching onto spatial-temporal data blocks using a bilinear rank-one matrix as in (8), while the sketching operator in [1] is unstructured. This allows one to reduce the sketching cost to $O(p+q)$, compared with using an unstructured sketching vector, which is $O(pq)$.

B. Kronecker covariance estimation

Given the linear model in (10), one may attempt to recover the covariance matrix $\boldsymbol{R_u}$ via least-squares estimation, as long as d is large enough, e.g. $d\geq p^2q^2/2$. However, as we will show, by leveraging the small separation rank of $\boldsymbol{R_u}$ in the Kronecker product expansion model, we can use a much smaller number of measurements to recover $\boldsymbol{R_u}$ than its ambient dimension based on convex relaxations. Note that we can rewrite y_i as a linear measurement of the rank-r matrix $\mathcal{P}(\boldsymbol{R_u})$ in (6),

$$y_i=\langle\mathcal{P}(\boldsymbol{R_u}),\mathcal{P}(\boldsymbol{Z}_i)\rangle,$$

where

$$\begin{aligned}\mathcal{P}(\boldsymbol{Z}_i)=\mathcal{P}(\boldsymbol{z}_i\boldsymbol{z}_i^T) &= \mathcal{P}((\boldsymbol{b}_i\otimes\boldsymbol{a}_i)(\boldsymbol{b}_i\otimes\boldsymbol{a}_i)^T)\\ &= \mathcal{P}((\boldsymbol{b}_i\boldsymbol{b}_i^T)\otimes(\boldsymbol{a}_i\boldsymbol{a}_i^T))\\ &= (\boldsymbol{b}_i\otimes\boldsymbol{b}_i)(\boldsymbol{a}_i\otimes\boldsymbol{a}_i)^T\end{aligned}\tag{11}$$

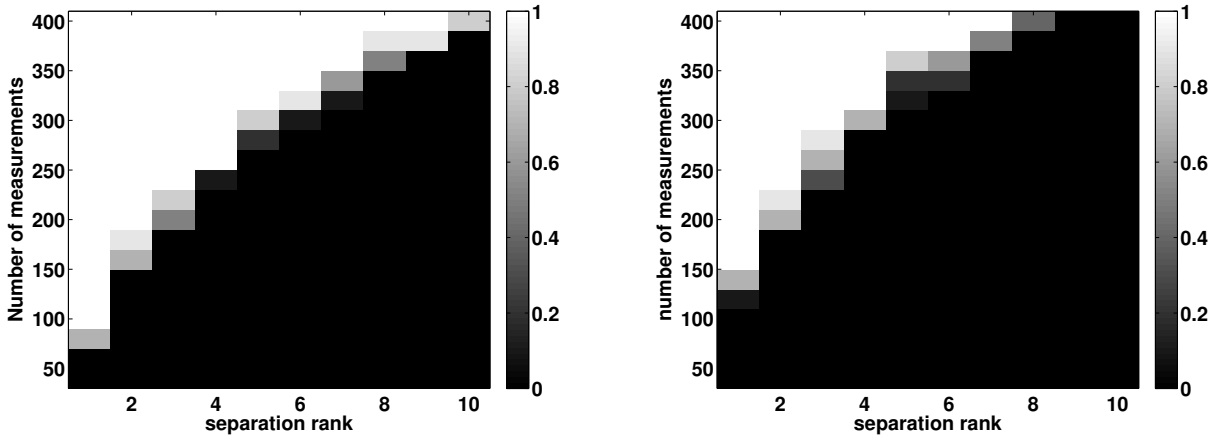
is also a rank-one matrix. Hence, we propose to seek the matrix with the minimum rank that satisfies the observation:

$$\min_{\boldsymbol{W}}\ \text{rank}(\boldsymbol{W})\quad\text{s.t.}\quad y_i=\langle\mathcal{P}(\boldsymbol{Z}_i),\boldsymbol{W}\rangle,\quad i=1,\dots,d,\quad \mathcal{P}^{-1}(\boldsymbol{W})\succeq 0,$$

where $\mathcal{P}^{-1}$ denotes the inverse permutation of $\mathcal{P}$ defined in (5) which is a linear operation.

Since minimizing the rank constraint is NP-hard, we look to the nuclear norm relaxation of the rank constraint [21],

[1] the exact selection ordering is not critical as long as each pair of sketching vectors sees enough data samples for reliable aggregation.



(a) Gaussian sketching vectors (b) Bernoulli sketching vectors

Fig. 1. Phase transition of covariance estimation under the Kronecker product expansion model. The success rate is plotted with respect to the number of measurements and the separation rank when $p = q = 6$, where the sketching vectors are generated with i.i.d. (a) Gaussian entries; (b) Bernoulli entries.

yielding the following algorithm:

$$\min_{\boldsymbol{W}\in\mathbb{R}^{p^2\times q^2}} \|\boldsymbol{W}\|_* \quad \text{s.t.} \quad y_i = \langle \mathcal{P}(\boldsymbol{Z}_i), \boldsymbol{W}\rangle, \quad i = 1,\dots,d, \\ \mathcal{P}^{-1}(\boldsymbol{W}) \succeq 0, \tag{12}$$

where $\|\cdot\|_*$ denotes the nuclear norm. In Section IV, our numerical simulations suggest that indeed the above algorithm allows exact recovery of the covariance matrix even when $d \ll p^2q^2$ when the separation rank r is small.

As a useful extension, we further consider the case when the temporal covariance $\boldsymbol{D}_i = \mathcal{T}(\boldsymbol{t}_i)$'s are Toeplitz matrices, and $\boldsymbol{E}_i$'s are unconstrained, where $\mathcal{T}(\boldsymbol{t}_i)$ denotes the symmetric Toeplitz matrix with $\boldsymbol{t}_i$ as the first column. Further define the matrix $\boldsymbol{T} \in \mathbb{R}^{q^2\times q}$ which corresponds to the linear mapping $\mathrm{vec}(\mathcal{T}(\boldsymbol{t})) = \boldsymbol{T}\boldsymbol{t}$. In this case, we have that

$$\begin{aligned}\mathcal{P}(\boldsymbol{R_u}) &= \sum_{i=1}^{r} \mathrm{vec}(\boldsymbol{E}_i)\,\mathrm{vec}(\boldsymbol{D}_i)^T \\ &= \sum_{i=1}^{r} \mathrm{vec}(\boldsymbol{E}_i)\,\mathrm{vec}(\mathcal{T}(\boldsymbol{t}_i))^T \\ &= \sum_{i=1}^{r} \mathrm{vec}(\boldsymbol{E}_i)\boldsymbol{t}_i^T\boldsymbol{T}^T := \boldsymbol{V}\boldsymbol{T}^T.\end{aligned}$$

where $\boldsymbol{V} = \sum_{i=1}^{r} \mathrm{vec}(\boldsymbol{E}_i)\boldsymbol{t}_i^T \in \mathbb{R}^{p^2\times q}$ is a rank-r matrix. We then wish to recover $\boldsymbol{V}$ using the following nuclear norm relaxation algorithm:

$$\min_{\boldsymbol{V}\in\mathbb{R}^{p^2\times q^2}} \|\boldsymbol{V}\|_* \quad \text{s.t.} \quad y_i = \langle \mathcal{P}(\boldsymbol{Z}_i), \boldsymbol{V}\boldsymbol{T}^T\rangle, \quad i = 1,\dots,d, \\ \mathcal{P}^{-1}(\boldsymbol{V}\boldsymbol{T}^T) \succeq 0. \tag{13}$$

Again, we will demonstrate the effectiveness of the proposed approach in the numerical simulations in Section IV.

IV. Numerical Examples

Let $p = q = 6$. We generate a synthetic covariance matrix satisfying the Kronecker product expansion model by randomly generating PSD matrices $\boldsymbol{E}_i \in \mathbb{R}^{p\times p}$ and $\boldsymbol{D}_i \in \mathbb{R}^{q\times q}$, $i = 1,\cdots,r$. We also randomly generate the measurement vectors with standard Gaussian or Bernoulli entries. For each (r,d) pair, we apply the sketching scheme and calculate the normalized mean squared error (NMSE) for reconstruction as $\|\hat{\boldsymbol{R}}_{\boldsymbol{u}} - \boldsymbol{R_u}\|_{\mathrm{F}}^2/\|\boldsymbol{R_u}\|_{\mathrm{F}}^2$, where $\hat{\boldsymbol{R}}_{\boldsymbol{u}}$ is the reconstructed covariance matrix from (12). For each Monte Carlo simulation, the reconstruction is claimed successful if the NMSE is below 10^{-5}. Fig. 1 shows the success rate of covariance reconstruction with respect to the number of measurements d and the separation rank r over 10 Monte Carlo experiments for each (r,d) pair, when the sketching vectors are generated i.i.d. using (a) standard Gaussian entries $\mathcal{N}(0,1)$, and (b) Bernoulli entries with probability $1/2$. It can be seen that successful recovery is possible even when d is much smaller than the ambient dimension of $\boldsymbol{R_u}$, especially when the separation rank is small.

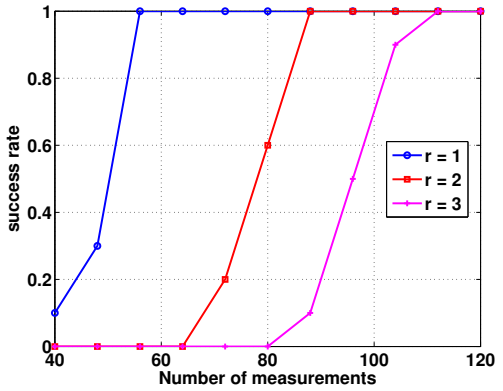


Fig. 2. Success rate of covariance estimation under the Kronecker product expansion model where the temporal covariance is a Toeplitz matrix. The success rate is plotted with respect to the number of measurements for separation rank $r = 1, 2, 3$, where the sketching vectors are generated with i.i.d. standard Gaussian entries.

Fig. 2 plots the success rate of covariance estimation when $\boldsymbol{D}_i$'s are additionally Toeplitz with respect to the number of measurements for separation rank $r = 1, 2, 3$, where the sketching vectors are generated with i.i.d. standard Gaussian entries, for $p = q = 6$. It can be seen that the additional Toeplitz structure further helps reducing the number of required measurements to achieve exact recovery.

V. Discussions

Here, we briefly discuss the possibility of using bilinear compressive measurements of each spatial-temporal data block $\boldsymbol{U}[t]$ for covariance sketching, i.e.

$$\boldsymbol{Y}[t] = \boldsymbol{A}\boldsymbol{U}[t]\boldsymbol{B}^T,$$

where $\boldsymbol{A} \in \mathbb{R}^{m_1 \times p}$ and $\boldsymbol{B} \in \mathbb{R}^{m_2 \times q}$, where $m_1 < p$ and $m_2 < q$ since we're interested in compression. This can be regarded as an extension of the sketching scheme in [3]–[5] to spatial-temporal data. Then $\mathrm{vec}(\boldsymbol{Y}[t]) = (\boldsymbol{B} \otimes \boldsymbol{A})\boldsymbol{u}[t]$ and

$$\begin{aligned}
\boldsymbol{R}_{\boldsymbol{Y}} &= \mathbb{E}\{\mathrm{vec}(\boldsymbol{Y}[t])\,\mathrm{vec}(\boldsymbol{Y}[t])^T\} \\
&= (\boldsymbol{B} \otimes \boldsymbol{A})\boldsymbol{R}_{\boldsymbol{u}}(\boldsymbol{B} \otimes \boldsymbol{A})^T \\
&= \sum_{i=1}^{r} (\boldsymbol{B} \otimes \boldsymbol{A})(\boldsymbol{D}_i \otimes \boldsymbol{E}_i)(\boldsymbol{B} \otimes \boldsymbol{A})^T \\
&= \sum_{i=1}^{r} (\boldsymbol{B}\boldsymbol{D}_i\boldsymbol{B}^T) \otimes (\boldsymbol{A}\boldsymbol{E}_i\boldsymbol{A}^T) \in \mathbb{R}^{m_1 m_2 \times m_1 m_2}.
\end{aligned}$$

From the covariance matrix of the compressed data $\boldsymbol{R}_{\boldsymbol{Y}}$, we wish to recover the covariance matrix of the uncompressed data $\boldsymbol{R}_{\boldsymbol{u}}$. Applying a permutation operator $\tilde{\mathcal{P}}(\cdot) : \mathbb{R}^{m_1 m_2 \times m_1 m_2} \mapsto \mathbb{R}^{m_1^2 \times m_2^2}$ similar to (5) to the above equation, we have

$$\begin{aligned}
\tilde{\mathcal{P}}(\boldsymbol{R}_{\boldsymbol{Y}}) &= \sum_{i=1}^{r} \mathrm{vec}(\boldsymbol{A}\boldsymbol{E}_i\boldsymbol{A}^T)\,\mathrm{vec}(\boldsymbol{B}\boldsymbol{D}_i\boldsymbol{B}^T)^T \\
&= \sum_{i=1}^{r} [(\boldsymbol{A} \otimes \boldsymbol{A})\,\mathrm{vec}(\boldsymbol{E}_i)][(\boldsymbol{B} \otimes \boldsymbol{B})\,\mathrm{vec}(\boldsymbol{D}_i)]^T \\
&= (\boldsymbol{A} \otimes \boldsymbol{A})\mathcal{P}(\boldsymbol{R}_{\boldsymbol{u}})(\boldsymbol{B} \otimes \boldsymbol{B})^T.
\end{aligned}$$

Due to the non-empty null space of $\boldsymbol{A}$ and $\boldsymbol{B}$, it is not possible to uniquely identify $\mathcal{P}(\boldsymbol{R}_{\boldsymbol{u}})$, hence $\boldsymbol{R}_{\boldsymbol{u}}$ from $\tilde{\mathcal{P}}(\boldsymbol{R}_{\boldsymbol{Y}})$. However, we may still be able to identify $\boldsymbol{R}_{\boldsymbol{u}}$ for Kronecker product expansion models with enhancements, particularly when $\boldsymbol{D}_i$'s and $\boldsymbol{E}_i's$ have additional Toeplitz or sparsity constraints.

VI. Conclusion

In this paper, we have proposed a novel covariance sketching scheme for high-dimensional spatial-temporal data, whose covariance matrices exhibit structures involving Kronecker products that can be harnessed to further reduce the sketching complexity. Efficient algorithms for covariance estimation are also proposed based on convex relaxations using properties of the Kronecker product. Numerical examples are provided to demonstrate the effectiveness of the proposed schemes. In future works, we plan to study the theoretical guarantees of the proposed algorithms, and applications of the proposed schemes for spectrum sensing for cognitive radios, data compression, and more.

References

[1] Y. Chen, Y. Chi, and A. Goldsmith, "Exact and stable covariance estimation from quadratic sampling via convex programming," *Information Theory, IEEE Transactions on*, vol. 61, no. 7, pp. 4034–4059, July 2015.

[2] Y. Chen, Y. Chi, and A. J. Goldsmith, "Estimation of simultaneously structured covariance matrices from quadratic measurements," in *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*. IEEE, 2014, pp. 7669–7673.

[3] G. Dasarathy, P. Shah, B. N. Bhaskar, and R. D. Nowak, "Sketching sparse matrices, covariances, and graphs via tensor products," *Information Theory, IEEE Transactions on*, vol. 61, no. 3, pp. 1373–1388, 2015.

[4] G. Leus and Z. Tian, "Recovering second-order statistics from compressive measurements," in *Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP),*, 2011, pp. 337–340.

[5] D. Romero, D. D. Ariananda, Z. Tian, and G. Leus, "Compressive covariance sensing: Structure-based compressive sensing beyond sparsity," *Signal Processing Magazine, IEEE*, vol. 33, no. 1, pp. 78–93, 2016.

[6] D. D. Arianand and G. Leus, "Compressive joint angular-frequency power spectrum estimation," in *Signal Processing Conference, Proceedings of the 21st European*. IEEE, 2013, pp. 1–5.

[7] J. M. Bioucas-Dias, D. Cohen, and Y. C. Eldar, "Covalsa: Covariance estimation from compressive measurements using alternating minimization," in *Signal Processing Conference (EUSIPCO), 2014 Proceedings of the 22nd European*. IEEE, 2014, pp. 999–1003.

[8] J. Lundén, V. Koivunen, A. Huttunen, and H. V. Poor, "Collaborative cyclostationary spectrum sensing for cognitive radio systems," *Signal Processing, IEEE Transactions on*, vol. 57, no. 11, pp. 4182–4195, 2009.

[9] K. Werner, M. Jansson, and P. Stoica, "On estimation of covariance matrices with kronecker product structure," *Signal Processing, IEEE Transactions on*, vol. 56, no. 2, pp. 478–491, 2008.

[10] J. C. De Munck, H. M. Huizenga, L. J. Waldorp, and R. M. Heethaar, "Estimating stationary dipoles from meg/eeg data contaminated with spatially and temporally correlated background noise," *Signal Processing, IEEE Transactions on*, vol. 50, no. 7, pp. 1565–1572, 2002.

[11] T. Tsiligkaridis and A. Hero, "Covariance estimation in high dimensions via kronecker product expansions," *Signal Processing, IEEE Transactions on*, vol. 61, no. 21, pp. 5347–5360, 2013.

[12] K. Greenewald, T. Tsiligkaridis, and A. O. Hero III, "Kronecker sum decompositions of space-time data," *arXiv preprint arXiv:1307.7306*, 2013.

[13] K. Greenewald and A. O. Hero, "Regularized block toeplitz covariance matrix estimation via kronecker product expansions," in *Statistical Signal Processing (SSP), 2014 IEEE Workshop on*. IEEE, 2014, pp. 9–12.

[14] M. F. Duarte and R. G. Baraniuk, "Kronecker compressive sensing," *Image Processing, IEEE Transactions on*, vol. 21, no. 2, pp. 494–504, 2012.

[15] O. Mehanna and N. Sidiropoulos, "Frugal sensing: Wideband power spectrum sensing from few bits," *Signal Processing, IEEE Transactions on*, vol. 61, no. 10, pp. 2693–2703, 2013.

[16] Y. Chi, "One-bit principal subspace estimation," in *Signal and Information Processing (GlobalSIP), 2014 IEEE Global Conference on*. IEEE, 2014, pp. 419–423.

[17] D. Romero, S.-J. Kim, R. López-Valcarce, and G. B. Giannakis, "Spectrum cartography using quantized observations," in *Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on*. IEEE, 2015, pp. 3252–3256.

[18] H. Fu and Y. Chi, "Principal subspace estimation for low-rank toeplitz covariance matrices with binary sensing," in *Proceeding of the Asilomar Conference on Signals, Systems, and Computers*, 2016.

[19] W. Gardner, "Exploitation of spectral redundancy in cyclostationary signals," *Signal Processing Magazine, IEEE*, vol. 8, no. 2, pp. 14–36, 1991.

[20] P. Dutilleul, "The mle algorithm for the matrix normal distribution," *Journal of statistical computation and simulation*, vol. 64, no. 2, pp. 105–123, 1999.

[21] B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM Review*, vol. 52, no. 3, pp. 471–501, 2010.