COVARIANCE TRACKING FROM SKETCHES OF RAPID DATA STREAMS

Yiran Jiang[†] *and Yuejie Chi*[†,‡]

[†]Department of Electrical and Computer Engineering, The Ohio State University
[‡]Department of Biomedical Informatics, The Ohio State University

ABSTRACT

Estimating and tracking the covariance matrix of high-dimensional data streams with low complexities in acquisition, storage and computation are of great interest in modern data-intensive applications. This paper develops an online covariance estimation and tracking algorithm for a recently developed covariance sketching framework that requires a single sketch per sample [1], by leveraging the low-rank structure of the covariance matrix. In particular, we devise a discounting mechanism in the aggregation procedure to enable faster tracking when the covariance structure changes over time. The performance of the proposed algorithm is validated through numerical examples.

***Index Terms*—** streaming data, covariance estimation and tracking, alternating projection, sketching

1. INTRODUCTION

Modern data-intensive applications have generated an explosive amount of high-dimensional and high-rate data samples that have overwhelmed traditional sensor suites to be fully observed and stored. In many cases, the data has to be processed on-the-fly [2] to respect time and resource constraints, making it challenging to extract useful information from the data stream in real time. What saves the day is that many real-world data streams can be in fact described by a number of parameters much smaller than the ambient dimension, such that each data sample lies approximately in a low-dimensional subspace possibly varying over time. Therefore, it is of great interest to estimate and track the underlying subspace for signal processing tasks, such as anomaly detection, target tracking and video surveillance [3].

Subspace estimation and tracking is a central topic in signal processing with many well-known algorithms such as Oja's rule and its many variations [4], Yang's PAST algorithm [5], and etc. However, they all require fully observed data samples which become ineffective or break down completely when the data samples are contaminated with missing entries. Until recently, efficient online algorithms have been developed to track partially observed data streams with low computational costs [6, 7, 8, 9, 10, 11] motivated by the advance in matrix completion [12]. However, these algorithms are unsatisfactory because the number of observed entries per sample has to be at least greater than the subspace rank, which may still be too high for a storage-limited sensor.

This limitation has been partially addressed by combining effective random sketching schemes to directly recover the covariance matrix rather than the data stream itself [1, 13, 14]. In particular, a low-complexity covariance sketching scheme has been developed in [1, 13, 14] that only takes a single energy sketch per sample by projecting it to a random rank-one subspace spanned by some sketching vector. By leveraging the ergodicity of the data stream, these sketches are aggregated into a set of sketching measurements that are linear with respect to the covariance matrix, and quadratic with respect to the sketching vectors. Universal reconstruction performance guarantees of the covariance matrix are established from a near-optimal number of sketching vectors via convex optimization by promoting low-dimensional covariance structures such as sparsity and low rank.

While the covariance sketching scheme in [1, 13, 14] is appealing for its low acquisition and storage complexities, the convex algorithm used to recover the covariance matrix is computationally expensive to be implemented in a streaming setting. The main contribution of this paper is to propose an online algorithm to update the estimate of a low-rank covariance matrix, by performing one round of alternating projection between the observation constraint and the low-rank constraint at each time the sketching measurements are circulantly updated once. Moreover, to allow faster tracking when the low-rank covariance matrix changes over time, we incorporate a discounting mechanism in the aggregation procedure by geometrically reweighting the historic sketches in a similar fashion to [6]. We validate the performance of the proposed algorithm on direction-of-arrival estimation for a unitary linear array, which achieves superior performance from only a single energy sketch per sample.

The rest of this paper is organized as follows. Section 2 introduces the covariance sketching scheme, and Section 3 describes the proposed online algorithm. Numerical examples are demonstrated in Section 4 and finally we conclude in Section 5.

2. COVARIANCE SKETCHING

In this section, we review the covariance sketching scheme proposed in [1, 13, 14] for estimating low-rank covariance matrices of ergodic data streams. Specifically, we use $\{\boldsymbol{x}_t\}_{t=1}^{\infty}$ to represent a high-dimensional data stream, where $\boldsymbol{x}_t \in \mathbb{C}^n$ is the data sample generated at the tth time satisfying $\mathbb{E}[\boldsymbol{x}_t] = \mathbf{0}$ and the covariance $\mathbb{E}[\boldsymbol{x}_t\boldsymbol{x}_t^H] = \boldsymbol{\Sigma}$. Moreover, the covariance matrix $\boldsymbol{\Sigma}$ is assumed low-rank or approximately low-rank. The prohibitively high rate at which data are generated and the severely limited resources at the sensing platforms force inference methods to function with as small memory and computational costs as possible [2].

2.1. Covariance Sketching

We start by introducing a set of non-adaptive sketching vectors $\{\boldsymbol{a}_i\}_{i=0}^{m-1}$, where each entry of $\boldsymbol{a}_i \in \mathbb{C}^n$ is generated from i.i.d. zero-mean sub-Gaussian distributions, e.g. Gaussian or Bernoulli. The covariance sketching scheme [1, 13, 14] requires only a single

This work is supported in part by National Science Foundation under the grant CCF-1422966 and a Google Faculty Research Award.

sketch per sample of the data stream from which the covariance matrix can be efficiently estimated. At each time $t \geq 1$, the covariance sketching scheme consists of the following key steps:

1. **Sketching:** We choose a sketching vector indexed by $\ell_t = \mod(t-1, m)$, where $\mod(\cdot)$ is the modulo function, and observe a single quadratic sketch
$$s_t = (\boldsymbol{a}_{\ell_t}^H \boldsymbol{x}_t)^2.$$
Note that only *one pass* of each data sample is required with linear complexity to compute the sketch.

2. **Aggregation:** All sketches employing the same sketching vector $\boldsymbol{a}_i$ are aggregated and stored in a single sketching measurement y_i, which, due to stationarity, as $t \to \infty$, converges to
$$y_i = \mathbb{E}[(\boldsymbol{a}_i^H \boldsymbol{x}_t)^2] = \boldsymbol{a}_i^H \mathbb{E}[\boldsymbol{x}_t \boldsymbol{x}_t^H] \boldsymbol{a}_i = \boldsymbol{a}_i^H \boldsymbol{\Sigma} \boldsymbol{a}_i, \quad (1)$$
which is linear in $\boldsymbol{\Sigma}$ and quadratic in $\boldsymbol{a}_i, i = 0, \ldots, m-1$.

In the above covariance sketching scheme, the data acquisition stage in the sketching and aggregation steps makes no assumption about the covariance structures and can be implemented in a fully distributed and online manner without storing the entire data stream. To be specific, denote the measurement at time t corresponding to the sketching vector $\boldsymbol{a}_i$ by y_i^t, then it can be updated from y_i^{t-1} and n_i^{t-1} as

$$y_i^t = \frac{1}{n_i^t} \sum_{\tau=1}^{t} s_t 1_{\{\ell_\tau = i\}} = \frac{n_i^{t-1}}{n_i^t} y_i^{t-1} + \frac{1_{\{\ell_t = i\}}}{n_i^t} s_t, \quad (2)$$

where $1_{\{\cdot\}}$ is the indicator function, $n_i^t = \sum_{\tau=1}^{t} 1_{\{\ell_\tau = i\}} = n_i^{t-1} + 1_{\{\ell_t = i\}}$ counts the number of sketches employing $\boldsymbol{a}_i$. Also, y_i^t converges to y_i as t tends to infinity yielding (1). The storage requirement is only m which can be made much smaller than the dimensionality of $\boldsymbol{\Sigma}$ and doesn't grow with time. Finally, it is worth noting that the sketches are energy measurements which, using energy detectors, are easier to measure and often more accurate than the phase measurements for high-frequency signals in optical systems [15] and wideband spectrum sensing [16]. This offers additional benefits of the covariance sketching scheme for potential applications.

2.2. Covariance Estimation

In order to account for the noise introduced in the sketching process, consider the noisy version of (1):

$$y_i = \boldsymbol{a}_i^H \boldsymbol{\Sigma} \boldsymbol{a}_i + \eta_i = \langle \boldsymbol{\Sigma}, \boldsymbol{a}_i \boldsymbol{a}_i^H \rangle + \eta_i = \langle \boldsymbol{\Sigma}, \boldsymbol{A}_i \rangle + \eta_i, \quad (3)$$

where $\boldsymbol{A}_i \triangleq \boldsymbol{a}_i \boldsymbol{a}_i^H$ is the corresponding rank-one measurement matrix with respect to $\boldsymbol{\Sigma}$ which is quadratic in $\boldsymbol{a}_i$, and $\boldsymbol{\eta} \triangleq \{\eta_i\}_{i=0}^{m-1}$ denotes the bounded (adversarial) noise with $\|\boldsymbol{\eta}\|_1 \leq \epsilon$. The measurements can be expressed succinctly as

$$\boldsymbol{y} = \mathcal{A}(\boldsymbol{\Sigma}) + \boldsymbol{\eta}, \quad (4)$$

where $\boldsymbol{y} = \{y_i\}_{i=0}^{m-1}$ is a set of m measurements, and $\mathcal{A}(\boldsymbol{\Sigma}) : \mathbb{C}^{n \times n} \to \mathbb{C}^m$ is the linear operator mapping $\boldsymbol{\Sigma}$ to $\{\langle \boldsymbol{\Sigma}, \boldsymbol{A}_i \rangle\}_{i=0}^{m-1}$. To motivate the low-rank structure of $\boldsymbol{\Sigma}$, we resort to the following convex optimization algorithm:

$$\hat{\boldsymbol{\Sigma}} = \underset{\boldsymbol{M} \succeq 0}{\operatorname{argmin}} \operatorname{Tr}(\boldsymbol{M}) \quad \text{s.t.} \quad \|\boldsymbol{y} - \mathcal{A}(\boldsymbol{M})\|_1 \leq \epsilon, \quad (5)$$

Algorithm 1 Covariance Tracking via Alternating Projection

Input: the data stream $\{\boldsymbol{x}_t\}_{t=1}^{\infty}$, the sketching vectors $\{\boldsymbol{a}_i\}_{i=0}^{m-1}$, the covariance rank r;

Initialization: a random $n \times n$ rank-r PSD matrix $\boldsymbol{\Sigma}^0$; an all-zero vector $\boldsymbol{y}_0 = \{y_i^0\}_{i=0}^{m-1}$;

for $t = 1, 2, \ldots$ **do**

  Update the sketching measurements by the covariance sketching scheme (2) to obtain $\boldsymbol{y}^t$;

  if $\mod(t-1, m) = m-1$ **then**

    Set $k = \lceil t/m \rceil$;

    Project the previous covariance estimate $\boldsymbol{\Sigma}^{k-1}$ onto the affine set $\mathcal{S}_1 = \{\boldsymbol{M} : \boldsymbol{y}^t = \mathcal{A}(\boldsymbol{M})\}$ determined $\boldsymbol{y}^t$:

$$\boldsymbol{Q}^k = \underset{\boldsymbol{Q}}{\operatorname{argmin}} \left\| \boldsymbol{Q} - \boldsymbol{\Sigma}^{k-1} \right\|_F, \quad \text{s.t.} \quad \boldsymbol{Q} \in \mathcal{S}_1; \quad (7)$$

    Project $\boldsymbol{Q}^k$ to the nearest rank-r PSD matrix to obtain the current covariance estimate $\boldsymbol{\Sigma}^k$:

$$\boldsymbol{\Sigma}^k = \underset{\operatorname{rank}(\boldsymbol{M}) \leq r}{\operatorname{argmin}} \left\| \boldsymbol{M} - \boldsymbol{Q}^k \right\|_F, \quad \text{s.t.} \quad \boldsymbol{M} \succeq 0. \quad (8)$$

  end if

end for

where $\operatorname{Tr}(\boldsymbol{M})$ denotes the trace of $\boldsymbol{M}$, and $\boldsymbol{M} \succeq 0$ denotes the positive-semidefinite (PSD) constraint. It is shown in [1] that, as soon as the number of measurements m exceeds the order of nr, with high probability, the solution $\hat{\boldsymbol{\Sigma}}$ to (5) satisfies

$$\left\| \hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma} \right\|_{\mathrm{F}} \leq C_1 \frac{\operatorname{Tr}(\boldsymbol{\Sigma} - \boldsymbol{\Sigma}_r)}{\sqrt{r}} + C_2 \frac{\epsilon}{m}, \quad (6)$$

for all covariance matrices $\boldsymbol{\Sigma}$ with the best rank-r approximation $\boldsymbol{\Sigma}_r$, where c_0, c_1, C_1 and C_2 are universal constants.

3. COVARIANCE TRACKING

In a streaming data environment, it is highly desirable to maintain an online estimate of the covariance matrix. Notice that the sketches (2) can be updated in a fully online fashion, by updating $\boldsymbol{y}^t = \{y_i^t\}_{i=0}^{m-1}$ from $\boldsymbol{y}^{t-1} = \{y_i^{t-1}\}_{i=0}^{m-1}$ via (2). While it is possible to directly obtain an online estimate of the covariance matrix using $\boldsymbol{y}^t$ via the algorithm in (5), it is computationally expensive. However, given that the measurements at consecutive times only differ slightly, it is natural to consider an alternative procedure, which, incorporates the previous covariance estimate as a *warm-start* to minimize computational costs in obtaining the current covariance estimate. In this section, we devise an alternating projection scheme to update the covariance estimate in a computation-efficient fashion.

3.1. Covariance Tracking via Alternating Projection

We focus on estimating and tracking a rank-r covariance matrix after all m sketching measurements are updated circulantly once using the covariance sketching scheme in Section 2.1. Let $k = \lceil t/m \rceil$, which counts the number of updates for each sketching measurements until time t. Summarized in Algorithm 1, at each k, our algorithm performs one round of alternating projection of the previous covariance estimate $\boldsymbol{\Sigma}^{k-1}$ between the measurement constraints and the structural constraints, based on the current sketching measurements $\boldsymbol{y}^t$.

First, we project $\mathbf{\Sigma}^{k-1}$ to the closest matrix in Frobenius norm satisfying the affine constraint $\mathcal{S}_1 = \{\boldsymbol{M} : \boldsymbol{y}^t = \mathcal{A}(\boldsymbol{M})\}$ by solving (7), whose solution can be written as

$$\boldsymbol{Q}^k = \mathbf{\Sigma}^{k-1} - \mathcal{A}^*(\mathcal{A}\mathcal{A}^*)^{-1}(\mathcal{A}(\mathbf{\Sigma}^{k-1}) - \boldsymbol{y}^t), \tag{9}$$

where $\mathcal{A}^*(\boldsymbol{y}) = \sum_{i=0}^{m-1} y_i \boldsymbol{a}_i \boldsymbol{a}_i^H : \mathbb{C}^m \mapsto \mathbb{C}^{n\times n}$ is the conjugate operator of $\mathcal{A}$. Note that $\boldsymbol{Q}^k$ computed from (9) is a Hermitian matrix. We then project $\boldsymbol{Q}^k$ to the nearest rank-r PSD matrix via (8), whose solution can be found by computing the eigenvalue decomposition (EVD) of $\boldsymbol{Q}^k$. Denote the EVD of $\boldsymbol{Q}^k$ as

$$\boldsymbol{Q}^k = \sum_{i=1}^{n} \rho_i^k \boldsymbol{u}_i^k (\boldsymbol{u}_i^k)^H, \tag{10}$$

where $\boldsymbol{u}_i^k$ and ρ_i^k are the corresponding eigenvectors and eigenvalues in the descending order, for $i = 1, \ldots, n$. Then the solution to (8) can be described as

$$\mathbf{\Sigma}^k = \sum_{i=1}^{r} \max\{\rho_i^k, 0\} \boldsymbol{u}_i^k (\boldsymbol{u}_i^k)^H. \tag{11}$$

The computational complexity of Algorithm 1 per iteration mainly comes from computing the top r eigenvectors and eigenvalues of $\boldsymbol{Q}^k$, which is of much lower complexity than running (5). Furthermore, if the operator $\mathcal{A}$ can be chosen to satisfy $\mathcal{A}\mathcal{A}^*(\boldsymbol{y}) = \boldsymbol{y}$ for any $\boldsymbol{y}$, (9) can be further simplified to

$$\boldsymbol{Q}^k = \mathbf{\Sigma}^{k-1} - \sum_{i=0}^{m-1} \left(\sum_{i=1}^{r} \max\{\rho_i^{k-1}, 0\} |\boldsymbol{a}_i^H \boldsymbol{u}_i^{k-1}|^2 - y_i^t \right) \boldsymbol{a}_i \boldsymbol{a}_i^H,$$

by plugging into (11). Therefore it is possible to exploit the incremental approach in [17] to compute the top-r eigenvectors and eigenvalues as the entries of $\boldsymbol{y}_i^t$ are updated.

3.2. Discounted Aggregation in Covariance Sketching

When the covariance structure of the data stream evolves over time, it is necessary to track these changes as agile as possible. To enable faster tracking, we modify the aggregation step (2) by reweighting the previous aggregate z_i^{t-1} and the current sketch s_t, denoted by

$$\begin{aligned} z_i^t &= \left((1 - 1_{\{\ell_t = i\}}) + \lambda 1_{\{\ell_t = i\}}\right) z_i^{t-1} + s_t 1_{\{\ell_t = i\}} \\ &= \begin{cases} \lambda z_i^{t-1} + s_t, & \ell_t = i \\ z_i^{t-1}, & \text{otherwise} \end{cases}, \end{aligned} \tag{12}$$

where λ is a discounting factor $0 \ll \lambda \leq 1$ that discounts the previous data samples [6]. To see this, expand (12) over t and obtain

$$z_i^t = \sum_{\tau=1}^{t} \lambda^{n_i^t - n_i^\tau} s_\tau 1_{\{\ell_\tau = i\}} = \boldsymbol{a}_i^H \left(\sum_{\tau=1}^{t} \lambda^{n_i^t - n_i^\tau} \boldsymbol{x}_\tau \boldsymbol{x}_\tau^H \right) \boldsymbol{a}_i,$$

where $n_i^t = \sum_{\tau=1}^{t} 1_{\{\ell_\tau = i\}}$. As t tends to infinity, z_i^t converges to

$$z_i = \boldsymbol{a}_i^H \left(\sum_{p=0}^{\infty} \lambda^p \mathbb{E}[\boldsymbol{x}_p \boldsymbol{x}_p^H] \right) \boldsymbol{a}_i = \frac{1}{1-\lambda} \boldsymbol{a}_i^H \mathbf{\Sigma} \boldsymbol{a}_i, \tag{13}$$

which corresponds to measuring the covariance matrix $\mathbf{\Sigma}$ up to a scaling factor determined by the discounting factor λ. In practice, λ is usually selected as a constant close to 1, therefore the bias introduced by the scaling is small. We apply the covariance tracking algorithm similarly to the discounted aggregations by replacing step 2 in Algorithm 1 with (12), and demonstrate its performance in the direction-of-arrival estimation in Section 4.3.

4. NUMERICAL EXPERIMENTS

We examine the performance of the proposed covariance tracking algorithm in this section on both synthetic examples with fixed and time-varying covariance matrices, and tracking direction-of-arrivals in array signal processing.

4.1. Tracking a fixed covariance matrix

Let $n = 20$, $r = 1$ and $m = 80$. We generate the data samples in the stream by $\boldsymbol{x}_t = \boldsymbol{U}\boldsymbol{g}_t$, where $\boldsymbol{U} \in \mathbb{R}^{n\times r}$ is a fix matrix composed of i.i.d. standard Gaussian entries, and $\boldsymbol{g}_t \in \mathbb{R}^r$ is generated with i.i.d. standard Gaussian entries. We also generate m sketching vectors $\{\boldsymbol{a}_i\}_{i=1}^m$ composed of i.i.d. standard Gaussian entries. We first compare the performance of Algorithm 1 with running the batch algorithm (5) using the online obtained sketching measurements $\boldsymbol{y}^t$ after all the m sketching measurements have been updated once. We calculate the normalized mean squared error (NMSE) as $\|\mathbf{\Sigma}^k - \mathbf{\Sigma}\|_F / \|\mathbf{\Sigma}\|_F$, where $\mathbf{\Sigma} = \boldsymbol{U}\boldsymbol{U}^T$ denotes the true covariance matrix and $\mathbf{\Sigma}^k$ denotes the estimated covariance matrix using the respective algorithms. Fig. 1 shows the average NMSE over 20 Monte Carlo runs with respect to the data stream index normalized by m. The proposed covariance sketching algorithm has a larger error at the beginning of the data stream, due to the poor approximation during the aggregation with insufficient samples, it approaches the performance of the batch algorithm with the increase of time.

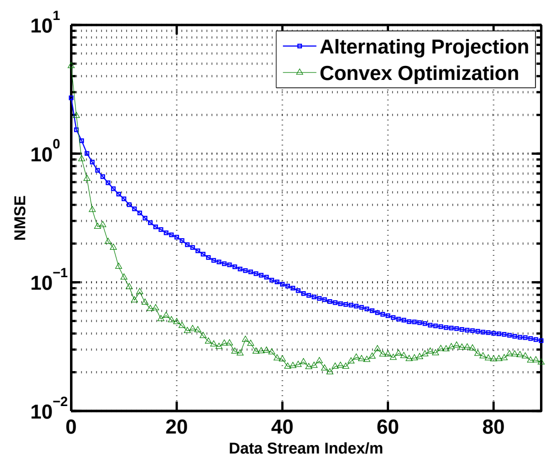
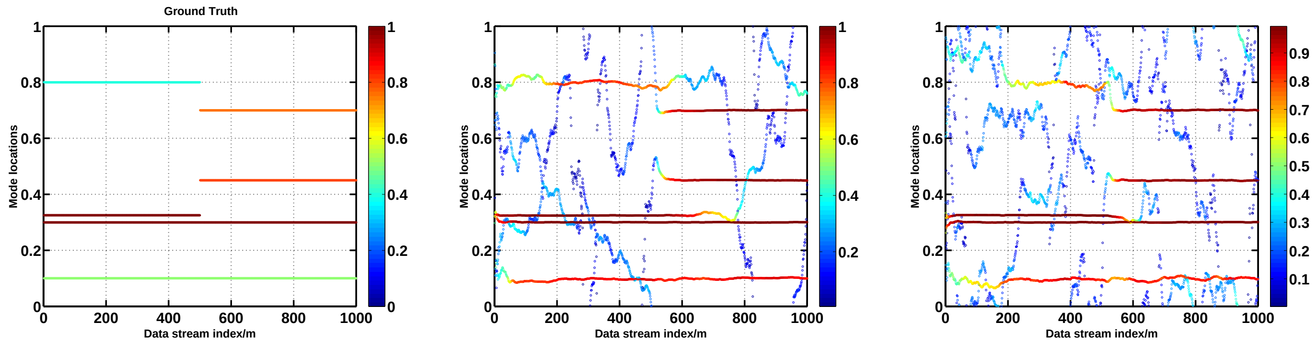


Fig. 1. NMSE of covariance estimation with respect to the data stream index using the proposed covariance sketching algorithm in Algorithm 1 and the batch algorithm (5).

4.2. Tracking a time-varying covariance matrix

We evaluate the performance of Algorithm 1 for tracking a time-varying covariance matrix. In this section, we show the performances of our proposed covariance tracking algorithm when the covariance of the data stream changes abruptly over time.

Let $n = 40$, $r = 3$ and $m = 600$. Let each data sample be $\boldsymbol{x}_t = \boldsymbol{U}\boldsymbol{g}_t$, where $\boldsymbol{U}$ is composed of i.i.d. standard Gaussian entries and changes abruptly in the middle of the stream twice and $\boldsymbol{g}_t$ is randomly generated with i.i.d. Gaussian entries. We compare the original aggregation scheme and the discounted aggregation scheme with a discounting factor $\lambda = 0.98$. Fig. 3 (a) shows the average



(a) Ground Truth (b) $\lambda = 0.99$ (c) $\lambda = 0.98$

Fig. 2. Direction-of-arrival estimation with respect to the data stream index using the proposed covariance sketching and tracking scheme. (a) Ground truth of mode locations; (b) estimated mode locations with a discounting factor $\lambda = 0.99$; and (c) estimated mode locations with discounting factor $\lambda = 0.98$.

NMSE over 20 Monte Carlo runs with respect to the data stream index normalized by m. The NMSE of the covariance estimate decays faster for the original covariance sketching scheme initially, since the sketching measurements converges faster to the desired measurements. However, it responds slower to changes in the covariance structure. On the contrary, the discounting mechanism allows us to track the covariance changes in a timely fashion.

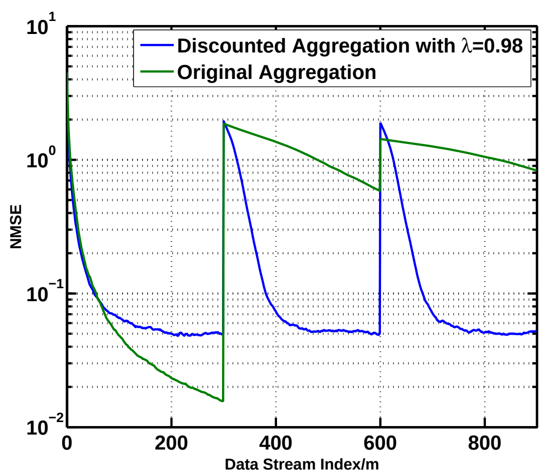


Fig. 3. NMSE of covariance estimation with respect to the data stream index when the covariance changes abruptly.

4.3. Tracking direction-of-arrivals

We evaluate the performance of the proposed covariance tracking algorithm for tracking direction-of-arrivals in a unitary linear array (ULA). Specifically, consider a ULA with $n = 40$ sensors, each data sample can be represented as

$$\boldsymbol{x}_t = \boldsymbol{V}\boldsymbol{D}\boldsymbol{g}_t, \tag{14}$$

where $\boldsymbol{V} = [\boldsymbol{\alpha}(\omega_1), \ldots, \boldsymbol{\alpha}(\omega_r)] \in \mathbb{C}^{n\times r}$ is a Vandermonde matrix composed of $r = 4$ columns where each column $\boldsymbol{\alpha}(\omega) = [1, e^{j2\pi\omega}, ..., e^{j2\pi\omega(n-1)}]^T$ corresponds to a mode with normalized frequency $\omega \in [0, 1)$, $\boldsymbol{D} = \mathrm{diag}\{\boldsymbol{d}\} = \mathrm{diag}\{d_1, ..., d_r\} \in \mathbb{R}^{r\times r}$ is a diagonal matrix characterizing the strength of each mode, and $\boldsymbol{g}_t \in \mathbb{C}^r$ is generated with i.i.d. Gaussian entries. We start the experiment by setting $\boldsymbol{\omega} = [0.1, 0.3, 0.325, 0.8]$ and $\boldsymbol{d} = [0.5, 1, 1, 0.4]$ in the beginning of the scene, which was altered abruptly to $\boldsymbol{\omega} = [0.1, 0.3, 0.45, 0.7]$ and $\boldsymbol{d} = [0.5, 1, 0.8, 0.75]$ at the middle of the scene. The ground truth is depicted in Fig. 2 (a), with color indicating the strength of the modes.

Our goal is to examine the proposed covariance sketching scheme for estimating and tracking the set of modes $\boldsymbol{\omega} = \{\omega_j\}_{j=1}^4$. We implement the sketching scheme with a set of $m = 600$ sketching vectors generated with i.i.d. standard Gaussian entries. We choose discounting factors $\lambda = 0.99$ and $\lambda = 0.98$ for aggregation, and run Algorithm 1 with an estimated rank $\hat{r} = 6$, which is a slightly over estimation of the true rank $r = 4$. The mode locations are then estimated by running ESPRIT [18] on the principal subspace of the estimated covariance matrix at each time.

Fig. 2 (b) and (c) show the estimated mode locations with respect to the data stream index. It can be observed that the proposed algorithm is capable of tracking the changes in the dynamic scene from only a single energy sketch per sample, in particular, the two close-located modes are separated clearly from the estimation. As we decrease λ, the algorithm drops the old modes that are exiting the scene (e.g. the mode at $\omega = 0.325$) faster; however, it also takes a longer time to track the weaker mode (e.g. the mode at $\omega = 0.8$), demonstrating an interesting performance trade-off.

5. CONCLUSION

We developed an efficient covariance tracking algorithm for the recently proposed covariance sketching scheme to obtain online estimates of the covariance matrix of a high-dimensional data stream. Our algorithm uses the previous estimate as a warm start, and projects it first to the affine subspace determined by the updated sketching measurements, and then projects it to the nearest low-rank PSD matrix as the covariance estimate. Moreover, a discounting mechanism is introduced in the aggregation procedure to improve the tracking performance. Numerical examples are provided to empirically validate the performance of the proposed algorithm. Future work includes analysis of the theoretical performance of the proposed tracking algorithm and applications in real-world data.

6. REFERENCES

[1] Y. Chen, Y. Chi, and A. Goldsmith, "Exact and stable covariance estimation from quadratic sampling via convex programming," *arXiv preprint arXiv:1310.0807*, 2013.

[2] S. Muthukrishnan, *Data streams: Algorithms and applications*. Now Publishers Inc, 2005.

[3] F. Porikli, O. Tuzel, and P. Meer, "Covariance tracking using model update based on lie algebra," in *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, vol. 1. IEEE, 2006, pp. 728–735.

[4] E. Oja, *Subspace methods of pattern recognition*. Research Studies Press England, 1983, vol. 4.

[5] B. Yang, "Projection approximation subspace tracking," *IEEE Transactions on Signal Processing*, vol. 43, no. 1, pp. 95–107, 1995.

[6] Y. Chi, Y. C. Eldar, and R. Calderbank, "PETRELS: Parallel Estimation and Tracking of Subspace by Recursive Least Squares from Partial Observations," *IEEE Trans. on Signal Processing*, vol. 61, pp. 5947–5959, 2013.

[7] C. Qiu, N. Vaswani, and L. Hogben, "Recursive robust PCA or recursive sparse recovery in large but structured noise," *arXiv preprint arXiv:1211.3754*, 2012.

[8] L. Balzano, R. Nowak, and B. Recht, "Online identification and tracking of subspaces from highly incomplete information," *Proc. Allerton 2010*, 2010.

[9] M. Mardani, G. Mateos, and G. B. Giannakis, "Subspace learning and imputation for streaming big data matrices and tensors," *arXiv preprint arXiv:1404.4667*, 2014.

[10] J. Feng, H. Xu, and S. Yan, "Online robust pca via stochastic optimization," in *Advances in Neural Information Processing Systems*, 2013, pp. 404–412.

[11] Y. Xie, J. Huang, and R. Willett, "Change-point detection for high-dimensional time series with missing data," *IEEE Journal of Selected Topics in Signal Processing*, vol. 7, no. 1, pp. 12–27, 2013.

[12] E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," *Foundations of Computational Mathematics*, vol. 9, no. 6, pp. 717–772, April 2009.

[13] 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*, May 2014, pp. 7669–7673.

[14] ——, "Robust and universal covariance estimation from quadratic measurements via convex programming," in *IEEE International Symposium on Information Theory*, July 2014.

[15] J. R. Fienup, "Reconstruction of an object from the modulus of its fourier transform," *Optics letters*, vol. 3, no. 1, pp. 27–29, 1978.

[16] A. Ghasemi and E. Sousa, "Collaborative spectrum sensing for opportunistic access in fading environments," in *First IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks, DySPAN*, 2005, pp. 131–136.

[17] M. Brand, "Incremental singular value decomposition of uncertain data with missing values," in *Computer VisionECCV 2002*. Springer, 2002, pp. 707–720.

[18] R. Roy and T. Kailath, "ESPRIT-estimation of signal parameters via rotational invariance techniques," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 37, no. 7, pp. 984 –995, Jul 1989.