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 circularly 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 \( \{x_t\}_{t=1}^{\infty} \) to represent a high-dimensional data stream, where \( x_t \in \mathbb{C}^n \) is the data sample generated at the \( t \)th time satisfying \( \mathbb{E}[x_t] = 0 \) and the covariance \( \mathbb{E}[x_t x_t^H] = \Sigma \). Moreover, the covariance matrix \( \Sigma \) 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 \( \{\alpha_i\}_{i=0}^{n-1} \), where each entry of \( \alpha_i \in \mathbb{C}^n \) 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
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 \( \ell_i = \text{mod} (t - 1, m) \), where \( \text{mod} (\cdot) \) is the modulo function, and observe a single quadratic sketch

   \[ s_t = (a_i^H x_t)^2. \]

   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 \( a_i \) are aggregated and stored in a single sketching measurement \( y_i \), which, due to stationarity, as \( t \to \infty \), converges to

   \[ y_i = \mathbb{E}[(a_i^H x_t)^2] = a_i^H \mathbb{E}[x_t x_t^H] a_i = a_i^H \Sigma a_i, \tag{1} \]

   which is linear in \( \Sigma \) and quadratic in \( a_i \), \( i = 0, \ldots, m - 1 \).

   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 \( a_i \) by \( y_i^t \), then it can be updated from \( y_i^{t-1} \) and \( n_i^{t-1} \) as

   \[ y_i^t = \frac{1}{n_i^t} \sum_{r=1}^{t} s_{t} 1_{\ell_r = i} = \frac{n_i^{t-1}}{n_i^t} y_i^{t-1} + \frac{1}{n_i^t} s_t 1_{\ell_t = i}, \tag{2} \]

   where \( 1_{\ell_r = i} \) is the indicator function, \( n_i^t = \sum_{r=1}^{t} 1_{\ell_r = i} = n_i^{t-1} + 1 \) counts the number of sketches employing \( a_i \). 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 \( \Sigma \) 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):

\[ y_i = a_i^H \Sigma a_i + \eta_i = \langle \Sigma, a_i a_i^H \rangle + \eta_i = \langle \Sigma, A_i \rangle + \eta_i, \tag{3} \]

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

\[ y = A(\Sigma) + \eta, \tag{4} \]

where \( y = \{ y_i \}_{i=0}^{m-1} \) is a set of \( m \) measurements, and \( A(\Sigma) : \mathbb{C}^{n \times n} \to \mathbb{C}^m \) is the linear operator mapping \( \Sigma \) to \( \{ \Sigma, A_i \} \\}_{i=0}^{m-1} \).

To motivate the low-rank structure of \( \Sigma \), we resort to the following convex optimization algorithm:

\[ \Sigma = \arg\min_M \text{Tr}(M) \text{ s.t. } \| y - A(M) \|_1 \leq \epsilon, \tag{5} \]

where \( \text{Tr}(M) \) denotes the trace of \( M \), and \( M \succeq 0 \) 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 \( \Sigma \) to (5) satisfies

\[ \| \Sigma - \Sigma \|_F \leq C_1 \frac{\text{Tr}(\Sigma - \Sigma_i)}{\sqrt{r}} + C_2 \frac{\epsilon}{m}, \tag{6} \]

for all covariance matrices \( \Sigma \) with the best rank-\( r \) approximation \( \Sigma_r \), where \( C_0, C_1, C_2 \) 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 \( y_i^t = \{ y_i^t \}_{i=0}^{m-1} \) from \( y_i^{t-1} = \{ y_i^{t-1} \}_{i=0}^{m-1} \) via (2). While it is possible to directly obtain an online estimate of the covariance matrix using \( y_i^t \) 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 in 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 circularly 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 \( \Sigma_{k-1} \) between the measurement constraints and the structural constraints, based on the current sketching measurements \( y_i^t \).

**Algorithm 1** Covariance Tracking via Alternating Projection

**Input:** the data stream \( \{ x_t \}_{t=1}^{\infty} \), the sketching vectors \( \{ a_i \}_{i=0}^{m-1} \), the covariance rank \( r \);

**Initialization:** a random \( n \times n \) rank-\( r \) PSD matrix \( \Sigma^0 \); an all-zero vector \( y_0 = \{ y_i^0 \}_{i=0}^{m-1} \);

1: for \( t = 1, 2, \ldots \) do
2: Update the sketching measurements by the covariance sketching scheme (2) to obtain \( y_i^t \);
3: if \( \text{mod} (t - 1, m) = m - 1 \) then
4: Set \( k = \lfloor t/m \rfloor \);
5: Project the previous covariance estimate \( \Sigma_{k-1} \) onto the affine set \( S_1 : \{ M : y^t = A(M) \} \) determined \( y_i^t \):
6: \[ Q^k = \arg\min_{Q} \left\| Q - \Sigma_{k-1} \right\|_F, \text{ s.t. } Q \in S_1; \tag{7} \]
7: end if
8: end for

where \( \text{Tr}(M) \) denotes the trace of \( M \), and \( M \succeq 0 \) 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 \( \Sigma \) to (5) satisfies

\[ \| \Sigma - \Sigma \|_F \leq C_1 \frac{\text{Tr}(\Sigma - \Sigma_i)}{\sqrt{r}} + C_2 \frac{\epsilon}{m}, \]

for all covariance matrices \( \Sigma \) with the best rank-\( r \) approximation \( \Sigma_r \), where \( C_0, C_1, C_2 \) are universal constants.
First, we project $\Sigma^{k-1}$ to the closest matrix in Frobenius norm satisfying the affine constraint $S_1 = \{ M : y^t = A(M) \}$ by solving (7), whose solution can be written as

$$Q^k = \Sigma^{k-1} - A^* (A A^*)^{-1} (A (\Sigma^{k-1}) - y^t),$$

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

$$Q^k = \sum_{i=1}^{r} \rho_i^k u_i^k (u_i^k)^H,$$

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

$$\Sigma^k = \sum_{i=1}^{m} \max \{ \rho_i^k, 0 \} u_i^k (u_i^k)^H.$$ 

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

$$Q^k = \Sigma^{k-1} - \sum_{i=1}^{m-1} \left( \sum_{t=1}^{r} \max \{ \rho_i^k, 0 \} |a_i^H u_i^k|^2 - y_t^2 \right) a_i a_i^H,$$

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 $y_t^k$ 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 $z_i^{k-1}$ and the current sketch $s_t$, denoted by

$$z_i^t = \begin{cases} \frac{1}{1 - \lambda (t=1)} + \lambda (t=1) \frac{z_i^{t-1}}{\lambda} + s_t 1\{t=1\} & \text{if } \lambda < 1 \\ \frac{1}{\lambda} z_i^{t-1} + s_t, & \text{otherwise} \end{cases},$$

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

$$z_i^t = \sum_{t=1}^{t} \lambda^{t-n_i^t} s_t 1\{t=1\} = a_i^H \left( \sum_{t=1}^{t} \lambda^{t-n_i^t} x_t x_t^H \right) a_i,$$

where $n_i^t = \sum_{s=1}^{t} 1\{s=1\}$. As $t$ tends to infinity, $z_i^t$ converges to

$$z_i = a_i^H \left( \sum_{t=0}^{\infty} \lambda^t E[x_t x_t^H] \right) a_i = \frac{1}{1 - \lambda} a_i^H \Sigma a_i,$$

which corresponds to measuring the covariance matrix $\Sigma$ up to a scaling factor determined by the discounting factor $\lambda$. In practice, $\lambda$ 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 $x_t = U g_t$, where $U \in \mathbb{R}^{n \times r}$ is a fixed matrix composed of i.i.d. standard Gaussian entries, and $g_t \in \mathbb{R}^r$ is generated with i.i.d. standard Gaussian entries. We also generate $m$ sketching vectors $\{a_t\}_{t=1}^m$ 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 $y_t$ after all the $m$ sketching measurements have been updated once. We calculate the normalized mean squared error (NMSE) as $\| \Sigma^k - \Sigma \|_F / \| \Sigma \|_F$, where $\Sigma = UU^T$ denotes the true covariance matrix and $\Sigma^k$ 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.

#### 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 $x_t = U g_t$, where $U$ is composed of i.i.d. standard Gaussian entries and changes abruptly in the middle of the stream twice and $g_t$ 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...
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.

\begin{align*}
\text{NMSE} = \frac{1}{m} \sum_{t=1}^{m} \left( \frac{x_t - \hat{x}_t}{x_t} \right)^2
\end{align*}

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 $\lambda = 0.98$.

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 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.

5. CONCLUSION
6. REFERENCES


