COMPRESSIVE PARAMETER ESTIMATION WITH MULTIPLE MEASUREMENT VECTORS VIA STRUCTURED LOW-RANK COVARIANCE ESTIMATION

Yuanxin Li† and Yuejie Chi†‡

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

ABSTRACT

In this paper, we study the problem of frequency estimation from partial observations of multiple measurement vectors under a sparsity constraint. We develop a two-step approach which first estimates a low-rank Hermitian Toeplitz covariance matrix from the partially observed sample covariance matrix via convex optimization, then recovers the set of frequencies via conventional spectrum estimation methods such as MUSIC. Our method doesn’t assume discretization of the underlying frequencies, therefore overcomes the basis mismatch problem in conventional compressed sensing [1], and can possibly recover a higher number of frequencies than the number of samples per measurement vector. Numerical examples are provided to validate the performance of the proposed algorithm, with comparisons against several existing approaches.

Index Terms— low rank, Toeplitz covariance estimation, multiple measurement vectors, partial observations

1. INTRODUCTION

Many signal processing applications encounter a signal ensemble where each signal in the ensemble can be represented as a sparse superposition of complex sinusoids sharing the same frequencies, for example in remote sensing and super-resolution imaging, and the goal is to recover the set of frequencies from a small number of linear measurements of the signal ensemble. Compressed sensing (CS) [2, 3] has been proposed as an efficient way to reduce the number of measurements while still guarantees stable reconstruction by exploiting sparsity in the reconstruction algorithms. When Multiple Measurement Vectors (MMV) present, a common approach is to exploit the group sparsity [4–6], which further improves the performance. However, conventional CS approaches usually assume a discretized grid for the continuous-valued frequencies, resulting in an undesired performance degeneration when the mismatch between the grid and the actual frequencies is severe [1].

It has been well recognized that covariance structures can be explored to improve spectrum estimation when multiple observations of a stochastic signal are available [7]. With a mild second-order statistical assumption on the sparse coefficients, a correlation-aware approach is proposed in [8, 9] to improve the size of recoverable support by exploring the sparse representation of the covariance matrix in the Khatri-Rao product of the signal sparsity basis. However, due to the above-mentioned basis mismatch issue, the correlation-aware approach can not estimate frequencies off the grid.

In this paper, we propose an algorithm to estimate the set of continuous-valued frequencies from partial observations of multiple spectrally sparse signals. Under the statistical assumption that the frequencies are uncorrelated, the full covariance matrix is a Hermitian Toeplitz matrix whose rank is the number of distinct frequencies. This holds in a variety of applications in array signal processing [10]. We first calculate the partial sample covariance matrix from partial observations of the measurement vectors. A convex optimization formulation is formulated to estimate the full Hermitian Toeplitz covariance matrix whose submatrix on the set of observed entries is close to the partial sample covariance matrix, with an additional trace regularization that promotes the low-rank structure. Trace regularization for positive semidefinite matrices is a widely adopted convex relaxation of the normal rank constraint. Finally, the set of frequencies can be estimated from the estimated full covariance matrix using conventional methods such as MUSIC [11]. Compared with directly applying MUSIC to the partial sample covariance matrix, the proposed algorithm has the potential to recover a higher number of frequencies than the number of samples per measurement vector by taking advantages of the array geometry [12]. We provide numerical examples to validate the proposed algorithm, and demonstrate the performance improves when the number of measurement vectors increases. We also compare against CS, atomic norm minimization [13], and the correlation-aware approach. As our algorithm only requires the partially observed sample covariance matrix rather than the observed signals, the computational complexity does not grow with the number of measurement vectors, in contrast to CS and atomic norm minimization approaches that aim to recover the signals.

The rest of this paper is organized as follows. We formulate the problem in Section 2, and present our algorithm in Section 3. Numerical experiments are provided in Section 4 to validate the proposed algorithm. Finally, conclusions and future work are discussed in Section 5. Throughout the paper, matrices are denoted by bold capitals and vectors by bold lowercase. We use $T(u)$ to represent a Hermitian Toeplitz matrix with vector $u$ as its first column, and $Tr(X)$ to represent the trace of $X$.

2. PROBLEM FORMULATION

We consider spectrally sparse signals that are composed of a small set of distinct frequency components. Specifically, suppose that $y = [y_0, y_1, \ldots, y_{n-1}]^T \in \mathbb{C}^n$ is a spectrally sparse signal that can be written as

$$y = \sum_{r=1}^{k} \theta_r \nu(f_r) \triangleq V \theta,$$

(1)

where $k$ is the spectral sparsity level, each atom $\nu(f_r)$ is defined as

$$\nu(f_r) = \left[ 1, e^{j2\pi f_r}, \ldots, e^{j2\pi f_r(n-1)} \right]^T,$$

(2)
with \( f_r \in [0, 1], r = 1, 2, \ldots, k \) and the coefficient vector \( \theta = [\theta_1, \theta_2, \ldots, \theta_k]^T \in \mathbb{C}^k \). The matrix \( V = [v(f_1), \ldots, v(f_k)] \) is a Vandermonde matrix. We denote \( F = \{f_r\}_{r=1}^k \) the set of frequencies, which are assumed to be distinct with each other.

Let \( Y = [y_1, y_2, \ldots, y_L] \in \mathbb{C}^{n \times L} \) be composed of \( L \) spectrally sparse signals with the same frequencies, where each vector \( y_l \) can be given as

\[
y_l = \sum_{r=1}^k \theta_{r,l} v(f_r) \triangleq V \theta_l,
\]

where \( y_l = [y_{0,l}, y_{1,l}, \ldots, y_{n-1,l}]^T \) and \( \theta_l = [\theta_{1,l}, \theta_{2,l}, \ldots, \theta_{k,l}]^T \). Therefore the signal matrix \( Y \) can be written as

\[
Y = V \Theta,
\]

where \( \Theta = [\theta_1, \theta_2, \ldots, \theta_L] \in \mathbb{C}^{k \times L} \).

In particular, we assume that the coefficients \( \theta_{r,l} \)'s for different signals satisfy \( \mathbb{E}[\theta_{r,l}] = 0 \) and the following second-order statistical property:

\[
\mathbb{E}[\theta_{r,l}\theta_{r',l'}] = \begin{cases} \sigma_r^2, & \text{if } r = r', \quad l = l', \\ 0, & \text{otherwise}. \end{cases}
\]

To put it differently, the coefficients from different signals are uncorrelated, and the coefficients for different frequencies in the same signal are also uncorrelated. As an example, (5) is satisfied if \( \theta_{r,l} \)'s are generated i.i.d. from \( \mathcal{CN}(0, \sigma_r^2) \).

We are interested in the scenario when only partial observations of \( Y \) are available. We assume that a random or deterministic subset of entries of each vector in \( Y \) is observed, and the observation pattern is denoted by \( \Omega \subset \{0, \ldots, n-1\} \) with cardinality \( |\Omega| = m \). In the absence of noise, the partially observed signal matrix is given as

\[
Y_{\Omega} = P_{\Omega}(Y) = [y_{0,\Omega}, y_{1,\Omega}, \ldots, y_{n-1,\Omega}] \in \mathbb{C}^{m \times L},
\]

where \( P_{\Omega} = I_{\Omega} \) is a partial identity matrix with rows specified in the set of \( \Omega \).

### 3. Proposed Algorithm

We develop an algorithm to estimate the set of frequencies \( F \) using the observation \( Y_{\Omega} \). Instead of focusing on reconstructing the complete signal matrix \( Y \), our algorithm explores the low-dimensional structure of its covariance matrix. Given (5), it is straightforward that the covariance matrix of the signal \( y_l \) in (3) can be written as

\[
R = \mathbb{E}[y_l y_l^*] = \sum_{r=1}^k \sigma_r^2 v(f_r) v(f_r)^*,
\]

which is a positive semidefinite (PSD) Hermitian Toeplitz matrix. This matrix is also low-rank with rank \( (R) = k \ll n \). In other words, the spectral sparsity translates into the rank of the covariance matrix. Let the first column of \( R \) be \( u = \sum_{r=1}^k \sigma_r^2 v(f_r) \in \mathbb{C}^n \), then \( R \) can be rewritten as

\[
R = T(u).
\]

The covariance matrix of the partially observed samples \( y_{\Omega,l} \) can then be given as

\[
R_{\Omega} = \mathbb{E}[y_{\Omega,l} y_{\Omega,l}^*] = I_{\Omega} R I_{\Omega} \in \mathbb{C}^{m \times m},
\]

corresponding to the submatrix of \( R \) with the rows and columns indexed by \( \Omega \).

If \( R_{\Omega} \) can be perfectly estimated, e.g. using an infinite number of measurement vectors, one might directly seek a low-rank Hermitian Toeplitz matrix \( T(u) \) which agrees with \( R_{\Omega} \) restricted on the submatrix indexed by \( \Omega \). Unfortunately, the ideal covariance matrix in (9) cannot be perfectly obtained; rather, we will first construct the sample covariance matrix of the partially observed samples as

\[
S_{\Omega} = \frac{1}{L} \sum_{l=1}^L y_{\Omega,l} y_{\Omega,l}^* = \frac{1}{L} Y_{\Omega} Y_{\Omega}^* \in \mathbb{C}^{m \times m}.
\]

We then seek a low-rank PSD Hermitian Toeplitz matrix whose restriction on the submatrix indexed by \( \Omega \) is close to the sample covariance matrix \( S_{\Omega} \) in (10). A natural algorithm would be

\[
\hat{u} = \arg\min_{u \in \mathbb{C}^n} \| I_{\Omega} T(u) I_{\Omega} - S_{\Omega} \|^2_F + \lambda \text{rank}(T(u))
\]

s.t. \( T(u) \succeq 0 \),

\[
\hat{u} = \arg\min_{u \in \mathbb{C}^n} \| I_{\Omega} T(u) I_{\Omega} - S_{\Omega} \|^2_F + \lambda \text{Tr}(T(u))
\]

s.t. \( T(u) \succeq 0 \),

where \( \lambda \) is a regularization parameter balancing the data fitting term and the rank regularization term. However, as directly minimizing the rank is NP-hard, we consider a convex relaxation for rank minimization over the PSD cone, which replaces the rank minimization by trace minimization, resulting in

\[
\hat{u} = \arg\min_{u \in \mathbb{C}^n} \| I_{\Omega} T(u) I_{\Omega} - S_{\Omega} \|^2_F + \lambda \text{Tr}(T(u))
\]

s.t. \( T(u) \succeq 0 \),

where the second term can be simplified as \( \text{Tr}(T(u)) = n u_0 \), where \( u_0 \) is the first entry of \( u \). The algorithm (12) can be solved efficiently using off-the-shelf semidefinite program solvers.

The proposed algorithm works with the sample covariance matrix \( S_{\Omega} \) of \( Y_{\Omega} \) rather than \( Y_{\Omega} \) directly. Therefore our algorithm does not require storing \( Y_{\Omega} \) of size \( mL \), but only \( S_{\Omega} \) of size \( m^2 \), which greatly reduces the necessary memory space when \( m \ll L \).

It is also worthwhile to connect the proposed algorithm with the correlation-aware method proposed in [8, 9]. The method in [8, 9], when specialized to a unitary linear array, can be regarded as a discretized version of our algorithm (12), where the atoms \( \{f_r\} \)’s in the covariance matrix (9) are discretized over a discrete grid.

### 4. Numerical Examples

We conduct several numerical experiments to validate the proposed algorithm. In particular, we examine the influence of the number of measurement vectors on the performance of covariance estimation and frequency estimation, and compare the proposed algorithm against several existing approaches. In all the experiments, we set \( \lambda = 5 \times 10^{-3} \frac{1}{(\log L)^2 \log m} \) which gives good performance empirically. A theoretical investigation of the choice of the regularization parameter is left for future work.

#### 4.1. Influence of \( L \) on structured covariance estimation

We first examine the influence of \( L \) on estimating the structured covariance matrix. We fix \( n = 64 \), and select \( m = 15 \) entries uniformly at random from each measurement vector. The frequencies
are selected uniformly from [0, 1), and the coefficients for each frequency are randomly drawn from $\mathcal{CN}(0, 1)$. For various number of measurement vectors $L$ and sparsity level $k$, we conduct the algorithm (12) and record the normalized mean squared error (NMSE), defined as $\frac{||\hat{\mathbf{R}} - \mathbf{R}||_F}{||\mathbf{R}||_F}$, where $\hat{\mathbf{R}}$ is the estimate obtained from (12) while $\mathbf{R}$ is the first column of the true covariance matrix. Each experiment is repeated 50 times, and the average NMSE is calculated. Fig. 1 shows the average NMSE with respect to the sparsity level $k$ for $L = 20, 100, 500, 1000$ and 5000. It can be seen that as $L$ increases, the average NMSE decreases for a fixed sparsity level.

**Fig. 1.** The NMSE with respect to the sparsity level $k$ for various $L$ when $n = 64$ and $m = 15$.

4.2. Influence of $L$ on frequency estimation

In this subsection we examine the influence of $L$ on frequency estimation using the obtained Toeplitz covariance matrix. This is done in MATLAB via applying the “rootsmusic” function with the true model order (i.e. the sparsity level $k$). We fix $n = 64$, and pick $m = 8$ entries uniformly at random from each measurement vector. Fig. 2 (a) shows the ground truth of the set of frequencies, where the amplitude of each frequency is given as the variance in (5). Fig. 2 (b)–(d) demonstrate the set of estimated frequencies when $L = 50$, 200, and 400. As we increase $L$, the estimates of the frequencies get more accurate, especially at separating close-located frequencies. It is also worth noticing that the amplitudes of the frequencies are not as well estimated, due to the small value of $m$.

**Fig. 2.** Frequency estimation of the proposed algorithm for different $L$’s when $n = 64$, $m = 8$ and $k = 6$. (a) Ground truth; (b) $L = 50$; (c) $L = 200$; and (d) $L = 400$.

4.3. Comparison with existing approaches on covariance estimation and frequency estimation

We compare the performance of the proposed algorithm for covariance estimation with existing approaches including CS using group sparsity [4], atomic norm minimization [13], and the correlation-aware approach [9] for MMV models. We will calculate the NMSE as $\frac{||\hat{\mathbf{R}} - \mathbf{R}||_F}{||\mathbf{R}||_F}$, where $\hat{\mathbf{R}}$ is the reconstructed covariance matrix, and $\mathbf{R}$ is the truth covariance matrix. Since the first two methods aim at recovering the signal matrix $\mathbf{Y}$, we will construct the estimated covariance matrix as the sample covariance matrix of the reconstructed signal matrix $\hat{\mathbf{Y}}$ as $\hat{\mathbf{R}} = \hat{\mathbf{Y}}\hat{\mathbf{Y}}^T / L$.

Let $n = 64$, $m = 15$, $k = 8$, and the signals are generated in the same method as in Section 4.1. For CS and correlation-aware method, we assume a sparsity dictionary of a DFT frame with an oversampling factor 4. For the correlation-aware method, we empirically set its regularization parameter to be $h = 2 \times 10^{-4} \frac{1}{(\log L)^2 (\log m)^2}$ which gives good performance [9]. Fig. 3 shows the NMSE with respect to the number of measurement vectors $L$ for different algorithms. Both our method and the correlation-aware method outperform the CS and atomic norm minimization approach, and ours is slightly better than the correlation-aware approach.

**Fig. 3.** The NMSE with respect to the number of measurement vectors $L$ for different algorithms when $n = 64$, $m = 15$ and $k = 8$.

In the next example, we compare the performance of frequency estimation using CS with a DFT frame, atomic norm minimization, correlation-aware with a DFT frame, and the proposed method. Let $n = 64$, $L = 400$ and $k = 6$. We generate a spectrally sparse ground truth scene in Fig. 4 (a) in the same way as Fig. 2 (a). Fig. 4 (b)–(e) respectively show the estimated frequencies in a unit circle for different methods, with $m = 8$ and $m = 5$ respectively in the
first row and the second row. Notice when \( m = 5 \) we assume \( \Omega = \{0, 32, 39, 47, 57\} \). The proposed algorithm works well to locate all the frequencies accurately in both cases. Due to the off-the-grid mismatch, CS and correlation-aware techniques predict frequencies on the lattice of the DFT frame, and result in a larger number of estimated frequencies. On the other hand, atomic norm minimization fails to distinguish the two close frequencies and misses one frequency due to insufficient number of measurements per vector.

5. CONCLUSION

We proposed an algorithm to estimate off-the-grid frequencies from partially observed samples of multiple spectrally sparse signals by first estimating a low-rank Toeplitz covariance matrix via convex optimization, and then applying conventional spectrum estimation to obtain the set of frequencies. The effectiveness of the proposed method is demonstrated through numerical examples via comparison with existing CS approaches. Future work includes the theoretical analysis of the proposed algorithm, as well as its performance in noisy environments. It will also be interesting to extend the proposed approach to coherent source models and higher dimensional frequencies.

6. REFERENCES


