HIGH-DIMENSIONAL SPARSE COVARIANCE ESTIMATION FOR RANDOM SIGNALS

Ahmed O. Nasif, *Zhi Tian*

Dept. of Electrical & Computer Engineering
Michigan Technological University
Houghton, Michigan 49931 USA

Qing Ling

Department of Automation
Univ. of Science and Technology of China
Hefei, Anhui 230027 China

ABSTRACT

This paper considers the problem of covariance matrix estimation from the viewpoint of statistical signal processing for high-dimensional or wideband random processes. Due to limited sensing resources, it is often desired to accurately estimate the covariance matrix from a small number of sample observations. To make up for the lack of observations, this paper leverages the structural characteristics of the random processes by considering the interplay of three widely-available signal structures: stationarity, sparsity and the underlying probability distribution of the observed random signal. New problem formulations are developed that incorporate both compressive sampling and sparse covariance estimation strategies. Tradeoff study is provided to illustrate the design choices when estimating the covariance matrices using a handful of sample observations.

Index Terms— sparse covariance estimation, compressed sensing, high-dimensional data, convex optimization

1. INTRODUCTION

Considerable research efforts have been invested on statistical modeling and model selection for different problems involving a large amount of data, for example, in fields such as machine learning, data mining, computational biology, econometrics, astronomy and so on. A class of such problems deal with identifying the covariance matrix of the observed data. Significant advances have been made in statistical analysis to develop fast and robust algorithms to estimate the covariance matrices of high-dimensional data. For example, sparse inverse covariance estimation has been discussed in [2, 8], and sparse covariance estimation has been investigated in [3, 4, 5, 6, 7]. In these works, sparsity is typically enforced on the (inverse) covariance for robust model selection and interpretation. They also exploit the underlying probabilistic distribution of the data set; for example, covariance estimation under multivariate Gaussian distribution is a canonical problem.

This paper considers the problem of high-dimensional covariance matrix estimation from the perspective of statistical signal processing, where the data collection process and signal characteristics of interest may differ from other fields. Particularly for wideband signals or random processes, the number of available signal observations can be quite limited due to constrained sensing resources in real-time sampling. In such a scenario, it is desired to accurately estimate the covariance matrix from a small number of signal observations. However, when the number of samples is smaller than the signal dimension, the problem of covariance estimation becomes ill-posed. Fortunately, the lack of observations can be made up by leveraging the known structures of the signal of interest. Here, we consider the interplay of three widely-available statistical structures, namely, signal stationarity, sparsity and knowledge of the underlying probability distribution of the observed random signals. For example, the data collection process for real-time signal acquisition may allow for compressive sampling [1], which can be incorporated into the statistical approach for sparsity-enforcing covariance estimation for Gaussian data [2, 3, 4, 5]. Accordingly, we formulate new optimization problems for covariance matrix estimation, and discuss their implementation algorithms. Our study shows that accurate estimation of the covariance matrix is indeed possible by appropriately exploiting the signal structures, even with a handful of samples observations.

This paper deals with statistical signal processing where real-time signal acquisition can be performed using compressive sampling [1]. We aim to accurately estimate the covariance matrix of a high-dimensional signal from a limited number of compressed data, which differ from the aforementioned prior work dealing with uncompressed data [2]-[8]. Given the same number of total samples, the tradeoff between data compression and sample size is considered in this paper. Moreover, for wide-sense stationary signals, we consider the special Toeplitz structure of the covariance matrix to further improve the estimation accuracy and reduce the sampling costs.

A large body of research has been done on compressive sampling (CS), but it mostly deals with deterministic signals where the goal is to perfectly reconstruct the original signals by exploiting sparsity [1]. Departing from this deterministic approach, we do not necessarily impose sparsity constraints on the signal itself; instead, we find sparsity in the signal covariance, which may arise because elements of a (non-sparse) random signal can be marginally independent. In fact, it is recently recognized that a deterministic approach to CS can be wasteful of sampling resources, when dealing with random signals characterized by their probabilistic statistics [9, 10, 11]. This line of work on CS for random processes originates from digital communications applications, where the goal is to estimate power spectrum or cyclic spectrum. However, it does not consider the underlying probabilistic distribution of the signals, such as Gaussianity. In contrast, this paper considers the basic covariance estimation problem, which can be generalized to subsume the estimation of other second-order statistics such as (cyclic) power spectrum. We consider the interplay of all three widely-available signal structures, including stationarity, sparsity and Gaussianity.

2. PROBLEM STATEMENT

Consider an N-variate random vector $\mathbf{x}$ with covariance matrix $\mathbf{R}_x = E\{\mathbf{x}\mathbf{x}^T\}$. For clear exposition, $\mathbf{x}$ is assumed to be real-valued, and of zero mean.

Suppose that there are L sample realizations $[\mathbf{x}_0, \ldots, \mathbf{x}_{L-1}]$ from the distribution of $\mathbf{x}$. Then, $\mathbf{R}_x$ can be estimated as its sample covariance matrix $\mathbf{S}_x$, given by

$$\mathbf{S}_x = \tfrac{1}{L} \textstyle\sum_{l=0}^{L-1} \mathbf{x}_l \mathbf{x}_l^T. \tag{1}$$

The data dimension N can be much larger than the sample size L, which renders $\mathbf{S}_x$ to be rank-deficient. In this case, it is difficult to accurately estimate $\mathbf{R}_x$. On the other hand, in many applications the covariance matrix of high-dimensional data exhibits some form of sparsity, because variables can be marginally independent. It is hence prudent to exploit the sparsity property of covariance matrices in order to improve the estimation accuracy given a limited sample size. It is true that the covariance matrix may not be sparse in many scenarios; in that case, it is reasonable to assume that sparsity can be attained in a suitable transform domain [11]. Therefore, it is well motivated to study the covariance matrix, which can be extended to estimate other second-order statistics as well. The goal here is to estimate $\mathbf{R}_x$ given limited sensing resources.

3. SPARSE COVARIANCE ESTIMATION FOR GAUSSIAN DATA

This section briefly reviews recent results on sparse estimation of covariance matrices [3, 4, 5, 6]. This line of work in statistical analysis focuses on model selection and high-dimensional data fitting for multivariate Gaussian data. Depending on the application domains, data compression is often irrelevant. Our description is primarily based on the results in [5].

Consider the case in which the sample size L is smaller than the data dimension N. To reach an accurate estimate of $\mathbf{R}_x$, the probabilistic (Gaussian) distribution of the data is utilized to construct a maximum likelihood (ML) estimator, and an ℓ_1-norm penalty term on $\mathbf{R}_x$ can be imposed to induce sparse estimates. Following [4, 5, 6], the objective function for sparsity-enforcing ML estimation of $\mathbf{R}_x$ is given by

$$\min_{\mathbf{R}_x} \log\det \mathbf{R}_x + \mathrm{tr}(\mathbf{R}_x^{-1}\mathbf{S}_x) + \lambda \|\mathbf{W} * \mathbf{R}_x\|_1. \tag{2}$$

We choose to impose a component-wise penalization on $\mathbf{R}_x$ via the matrix $\mathbf{W}$ containing only nonnegative elements, i.e., $||\mathbf{W} * \mathbf{R}_x||_1$, where $*$ denotes elementwise multiplication. Depending on the a priori knowledge available regarding the sparsity structure of $\mathbf{R}_x$, we can choose $\mathbf{W}$ differently. For example [5], (i) $\mathbf{W}_{ij} = 1, \forall i, j$, (ii) $\mathbf{W}_{ii} = 1, \forall i \neq j$ and $\mathbf{W}_{ii} = 0, \forall i$, (iii) $\mathbf{W}_{ij} = (|\mathbf{S}_x(i,j)| + \Delta)^{-1}, \forall i \neq j, \mathbf{W}_{ii} = 0, \forall i$, where Δ is a small positive scalar introduced for numerical stability. Essentially, (ii) enforces sparsity on off-diagonal elements of $\mathbf{R}_x$ only, while (iii) approaches an ℓ_0-norm penalty on off-diagonal terms when $\mathbf{S}_x$ is close to the true $\mathbf{R}_x$. Throughout this paper, the ℓ_1-norm for a matrix $\mathbf{B}$ is defined as its entry-wise norm, $\|\mathbf{B}\|_1 := \sum_{i,j} |\mathbf{B}_{ij}|$.

Note in (2) that the available parameter is the uncompressed sample covariance $\mathbf{S}_x$ obtained from (1), while the desired output is a (sparse) $\mathbf{R}_x$ reflecting the prior knowledge of the data distribution. The objective function is non-convex, posing numerical challenges. For solving (2), several fast algorithms have been developed, e.g., [5, 6]. We adopt the algorithm presented in [5]. It uses a majorization-minimization approach to approximate the nonconvex objective (2) by iteratively solving the following convex problem:

$$\begin{aligned} \hat{\mathbf{R}}_x^{(t)} = \arg\min_{\mathbf{R}_x} \quad & \mathrm{tr}\Big\{(\hat{\mathbf{R}}_x^{(t-1)})^{-1}\mathbf{R}_x\Big\} + \mathrm{tr}(\mathbf{R}_x^{-1}\mathbf{S}_x) \\ & + \lambda||\mathbf{W} * \mathbf{R}_x||_1 && \text{(3a)} \\ s.t. \quad & \mathbf{R}_x \text{ is p. s. d.} && \text{(3b)} \end{aligned}$$

Here, $\hat{\mathbf{R}}_x^{(t-1)}$ is the estimate of $\mathbf{R}_x$ obtained from the previous iteration step, and p.s.d. refers to positive semi-definite. The steps for solving this problem is given in Section 3.2 and Appendix 3 of [5]. Although (3) is not guaranteed to give a global minimum, it has been shown that limit points of such an algorithm are critical points of the objective (2) [12].

Besides the Gaussianity and sparsity properties of $\mathbf{R}_x$ considered in (2), we will develop new algorithms that further exploit the stationarity property of many signals and take advantage of compressed sensing for analog signals, as elaborated in ensuing sections.

4. THE ROLE OF STATIONARITY

Consider wide-sense stationary processes $\mathbf{x}$. Let $r_{xx}(i) = E\{x(j)x(j+i)\}, \forall i, j$. Due to stationarity, the $N \times N$ covariance matrix $\mathbf{R}_x$ becomes a structured Toeplitz matrix:

$$\mathbf{R}_x = \begin{bmatrix} r_{xx}(0) & r_{xx}(1) & \cdots & r_{xx}(N-1) \\ r_{xx}(1) & r_{xx}(0) & r_{xx}(1) & \vdots \\ \vdots & & \ddots & \vdots \\ r_{xx}(N-1) & \cdots & \cdots & r_{xx}(0) \end{bmatrix} \tag{4}$$

There are N free parameters in $\mathbf{R}_x$, which we can organize into an $N \times 1$ vector $\mathbf{r}_x$ in the form:

$$\mathbf{r}_x = [r_{xx}(0), r_{xx}(1), \cdots, r_{xx}(N-1)]^T. \quad (5)$$

By mapping elements in $\mathbf{r}_x$ to those in $\mathbf{R}_x$, it can be shown that

$$\text{vec}\{\mathbf{R}_x\} = \bar{\mathbf{P}}_N \mathbf{r}_x \quad (6)$$

where vec$\{\cdot\}$ is the vectorization operator that stacks a matrix column-by-column into a vector, and $\bar{\mathbf{P}}_N \in \{0,1\}^{N^2 \times N}$ is the specific mapping matrix; c.f., Section IV.A in [11].

In light of (6), the problem of estimating the covariance matrix $\mathbf{R}_x$ can be equivalently solved by finding $\mathbf{r}_x$, with a reduced number of unknowns. Accordingly, we refine the formulation in (3) by incorporating the stationarity property, as follows:

$$\hat{\mathbf{r}}_x^{(t)} = \arg\min_{\mathbf{r}_x} \left(\text{vec}\left\{(\hat{\mathbf{R}}_x^{(t-1)})^{-T}\right\}\right)^T \bar{\mathbf{P}}_N \mathbf{r}_x + \text{tr}(\mathbf{H}_x^{-1}\mathbf{S}_x)$$
$$+ \lambda \|\mathbf{w} * \mathbf{r}_x\|_1 \quad (7a)$$
$$s.t. \quad \text{vec}\{\mathbf{R}_x\} = \bar{\mathbf{P}}_N \mathbf{r}_x \quad (7b)$$
$$\mathbf{R}_x \text{ is p. s. d.} \quad (7c)$$

Here $\mathbf{H}_x := [\mathbf{H}_1\mathbf{r}_x \; \mathbf{H}_2\mathbf{r}_x \; \cdots \; \mathbf{H}_N\mathbf{r}_x]$, where $\mathbf{H}_i := \bar{\mathbf{P}}_N((i-1)N+1 : iN, :)$ is the $i-$th block of N rows in $\bar{\mathbf{P}}_N$. Also, $\mathbf{w} := \mathbf{Q}_N \text{vec}\{\mathbf{W}\}$, where $\mathbf{Q}_N$ is a known mapping matrix (cf. (15)). Note that the objective function in (7a) is convex in $\mathbf{r}_x$, and (7b) enforces the Toeplitz symmetry.

5. COMPRESSED SENSING FOR COVARIANCE ESTIMATION

We now consider the use of compressive sampling of random signals to save the sensing resources. Compressive sample vectors are collected via a $K \times N$ sampling matrix $\mathbf{A}$, in the form

$$\mathbf{z}_l = \mathbf{A}\mathbf{x}_l, \quad l = 0, \ldots, L-1. \quad (8)$$

Let $\mathbf{R}_z = E\{\mathbf{z}\mathbf{z}^T\}$ denote the covariance matrix of $\mathbf{z} = \mathbf{A}\mathbf{x}$. The uncompressed covariance matrix $\mathbf{R}_x$ is related to the compressed covariance matrix $\mathbf{R}_z$ by

$$\mathbf{A}\mathbf{R}_x\mathbf{A}^T = \mathbf{R}_z. \quad (9)$$

The resulting sample covariance of the compressed data is

$$\mathbf{S}_z = \tfrac{1}{L}\textstyle\sum_{l=0}^{L-1} \mathbf{z}_l \mathbf{z}_l^T. \quad (10)$$

Evidently, $\mathbf{S}_z$ and $\mathbf{S}_x$ are related via the equality

$$\mathbf{A}\mathbf{S}_x\mathbf{A}^T = \mathbf{S}_z. \quad (11)$$

Given $\{\mathbf{z}_l\}_{l=0}^{L-1}$, or $\mathbf{S}_z$, the goal is to estimate $\mathbf{R}_x$ by estimating $\mathbf{S}_x$ first. Because the dimension of $\mathbf{S}_x$ is larger than that of $\mathbf{S}_z$, we cannot find accurate estimate unless we utilize certain structural knowledge of $\mathbf{R}_x$. By imposing sparsity on $\mathbf{R}_x$, the following formulation arises:

$$\min_{\mathbf{R}_x} \quad \|\mathbf{W} * \mathbf{R}_x\|_1 \quad (12a)$$
$$s.t. \quad \mathbf{A}\mathbf{R}_x\mathbf{A}^T = \mathbf{S}_z. \quad (12b)$$
$$\mathbf{R}_x \text{ is p. s. d.} \quad (12c)$$

Here we choose $\mathbf{W}$ according to choice (ii), since the sample covariancce $\mathbf{S}_x$ is unavailable when compressive sensing is employed.

When $\mathbf{x}$ is stationary, $\mathbf{R}_x$ becomes Toeplitz, and (12) can be further simplified. To do so, we vectorize both sides of (9), and utilize the equality vec$\{\mathbf{U}\mathbf{V}\mathbf{W}\} = (\mathbf{W}^T \otimes \mathbf{U})vec\{\mathbf{V}\}$, to reach

$$(\mathbf{A} \otimes \mathbf{A})\text{vec}\{\mathbf{R}_x\} = \text{vec}\{\mathbf{R}_z\}. \quad (13)$$

Note that $\mathbf{R}_z$ is symmetric, but does not possess the special Toeplitz structure of $\mathbf{R}_x$. Hence, there are $K(K+1)/2$ unique elements in $\mathbf{R}_z$ due to symmetry. These elements, say, those in the upper triangle, can be organized into a $\frac{K(K+1)}{2} \times 1$ vector $\mathbf{r}_z$:

$$\mathbf{r}_z = [r_{zz}(0,0), r_{zz}(1,0), \ldots, r_{zz}(K-1,0), \ldots...]^T \quad (14)$$

where $r_{zz}(i,j) = E\{z(i)z(i+j)\}$.

By mapping the elements between $\mathbf{r}_z$ and vec$\{\mathbf{R}_z\}$, it can be shown that [11, Appendix A]

$$\text{vec}\{\mathbf{R}_z\} = \mathbf{P}_K \mathbf{r}_z, \quad \mathbf{r}_z = \mathbf{Q}_K \text{vec}\{\mathbf{R}_z\} \quad (15)$$

where $\mathbf{P}_K \in \{0,1\}^{K^2 \times \frac{K(K+1)}{2}}$ and $\mathbf{Q}_K \in \{0, 1/2, 1\}^{\frac{K(K+1)}{2} \times K^2}$ are the mapping matrices.

Putting together (6), (13) and (15), one has

$$\mathbf{r}_z = \mathbf{Q}_K \text{vec}\{\mathbf{R}_z\} = \underbrace{\mathbf{Q}_K(\mathbf{A} \otimes \mathbf{A})\bar{\mathbf{P}}_N}_{:=\mathbf{\Phi}} \mathbf{r}_x, \quad (16)$$

where $\mathbf{\Phi}$ is of size $\frac{K(K+1)}{2} \times N$. When $\frac{K(K+1)}{2} \geq N$ and $\mathbf{A}$ is properly chosen to ensure that $\mathbf{\Phi}$ is full rank, $\mathbf{r}_x$ can be estimated via the least-squares solution. Further, if $\mathbf{r}_x$ is sparse, ℓ_1-norm penalty on $\mathbf{r}_x$ can be imposed to ensure a sparse solution to $\mathbf{r}_x$. Once $\mathbf{r}_x$ is estimated, $\mathbf{R}_x$ can be constructed from (6) or by element-by-element mapping.

Summarizing, the sparse covariance of a stationary signal can be estimated from compressive samples as follows:

$$\min_{\mathbf{r}_x} \quad \|\mathbf{s}_z - \mathbf{\Phi}\mathbf{r}_x\|_2^2 + \lambda\|\mathbf{w} * \mathbf{r}_x\|_1 \quad (17a)$$
$$s.t. \quad \text{vec}\{\mathbf{R}_x\} = \bar{\mathbf{P}}_N \mathbf{r}_x \quad (17b)$$
$$\mathbf{R}_x \text{ is p. s. d.} \quad (17c)$$

Here, the known input $\mathbf{s}_z$ is obtained from the sample covariance $\mathbf{S}_z$ according to (15), and $\mathbf{A}$ is assumed to have been chosen properly to ensure full-rankness of $\mathbf{\Phi}$. This is a convex problem that can be readily solved. Because $\mathbf{R}_x$ is reconstructed from the estimate of $\mathbf{r}_x$, the special Toeplitz structure of $\mathbf{R}_x$ has been explicitly incorporated. Different from (12), we choose $\mathbf{w}$ according to choice (iii) by first estimating $\mathbf{s}_x$ using least squares as $\hat{\mathbf{s}}_x = (\mathbf{\Phi}^T\mathbf{\Phi})^{-1}\mathbf{\Phi}^T\mathbf{s}_z$, and then reconstructing $\mathbf{S}_x$.

6. GAUSSIANITY, STATIONARITY, SPARSITY AND COMPRESSIVE SAMPLING

When all probabilistic structures of Gaussianity, Stationarity and Sparsity are present in the random signals of interest, we can combine these structural knowledge to improve the estimation accuracy or reduce the sample sizes. This combination has been reflected in the developed formula (7). We are further motivated to utilize compressed sensing, which can reduce the sampling rate or the number of samples per block. By doing so, even when the sample size L is smaller than the signal dimension N (thus causing the sample covariance $\mathbf{S}_x$ to be rank-deficient), the compressed sample covariance $\mathbf{S}_z$ of size $K \times K$ may still be full rank, provided that the compression ratio K/N is chosen such that $K \leq L$.

By jointly considering the design objectives and constraints in (7) and (17), we reach a sparsity-enforcing maximum likelihood estimator from the compressive samples of a Gaussian stationary signal, as follows:

$$(\hat{\mathbf{r}}_x^{(t)}, \hat{\mathbf{s}}_x^{(t)}) = \arg \min_{(\mathbf{r}_x, \mathbf{s}_x)} \left(\text{vec}\left\{ (\hat{\mathbf{R}}_x^{(t-1)})^{-T} \right\} \right)^T \bar{\mathbf{P}}_N \mathbf{r}_x$$

$$+\text{tr}(\mathbf{H}_x^{-1}\mathbf{G}_x) + \lambda ||\mathbf{w} * \mathbf{r}_x||_1 + \rho \|\mathbf{s}_z - \mathbf{\Phi}\mathbf{r}_x\|_2^2 \quad (18a)$$

$$s.t. \quad \text{vec}\{\mathbf{R}_x\} = \bar{\mathbf{P}}_N \mathbf{r}_x \quad (18b)$$

$$\mathbf{R}_x, \mathbf{G}_x \text{ are p. s. d.} \quad (18c)$$

Here, $\mathbf{G}_x := [\mathbf{H}_1\mathbf{s}_x \ \mathbf{H}_2\mathbf{s}_x \ \cdots \ \mathbf{H}_N\mathbf{s}_x]$ and $\mathbf{H}_x := [\mathbf{H}_1\mathbf{r}_x \ \mathbf{H}_2\mathbf{r}_x \ \cdots \ \mathbf{H}_N\mathbf{r}_x]$, where $\mathbf{H}_i := \bar{\mathbf{P}}_N((i-1)N+1 : iN, 1 : N)$.

In (18), the uncompressed sample covariance quantity $\mathbf{s}_x$ is a nuisance parameter that is used to find the desired parameter $\mathbf{r}_x$ from the available compressed sample covariance vector $\mathbf{s}_z$. An alternating direction method of multipliers can be devised to solve for $\mathbf{s}_x$ and $\mathbf{r}_x$ via an iterative procedure.

7. COMPARATIVE STUDY

To compare the different techniques, we generate observation data samples from a zero-mean multivariate Gaussian distribution with covariance matrix $\mathbf{R}_x$, where $\mathbf{R}_x$ is of size $N \times N$. $\mathbf{R}_x$ is chosen such that it is symmetric, positive definite, Toeplitz and sparse. According to our model, the locations of the unique non-zero entries in the first row of $\mathbf{R}_x$ (i.e, $r_{xx}(i)$ where $i = 0, 1, 2, \cdots, N-1$) are chosen for a given sparsity level. Their values are chosen randomly, according to $\mathcal{N}(0, 1)$. We choose to find suitable values of λ via cross validation [13] to optimize our performance metric, which is chosen to be the root mean-square error (RMSE), $||\hat{\mathbf{R}}_x - \mathbf{R}_x||_F/N$ averaged over 100 realizations. We set $N = 32$, the sparsity of $\mathbf{R}_x$ to be 25%, and $N_L = 512$, where N_L is the total number of samples. It is given by $N_L = L \times K$ in the compressive sampling case, and $N_L = L \times N$ in the uncompressed case. All the convex formulations are solved using the CVX Matlab package [14], except (3) which is solved using the majorization-minimization algorithm of [5].

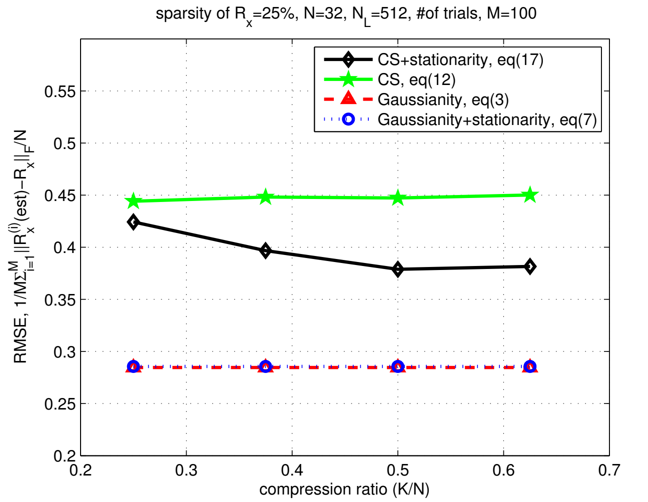

Fig. 1. RMSE as a function of compression ratio, K/N, for $N = 32$ and $N_L = KL = 512$.

Compressive Sampling: In Fig. 1, (17) performs better compared to (12) for all compression levels, showing the advantage of incorporating stationarity in the estimator. The best performance of (17) is achieved at 50% compression, which indicates the best tradeoff between improvement due to increased sample size, and degradation due to increased compression. Interestingly, the performance of (12) remains almost constant for the different compression levels.

Gaussianity: The sample size is fixed at $L = 16$ when solving (3), corresponding to the data-starved case. On the other hand, the sample size $L = N_L/K$ varies when solving (17), depending on the varying compression ratio K/N. The Gaussianity case incorporates the knowledge of the data distribution, while the compressed sensing case allows for a larger sample size. For the chosen parameters, Gaussianity yields better accuracy. Almost identical result is achieved by (7). This is not surprising, since both (3) and (7) is based on the same convex objective function, and (3) implicitly incorporates the Toeplitz structure in its penalty matrix $\mathbf{W}$.

8. SUMMARY

We have developed several techniques to estimate the covariance matrix of random signals based on the exploitation of three different structures in the data: Gaussianity, stationarity, and compression. Relative performance evaluation shows that joint use of Gaussianity and Toeplitz symmetry is crucial to achieving accurate estimates.

In our future work, we would like to study estimators that exploit all of the four structures in the data mentioned above. We are also interested in generalizing our study to nonstationary and cyclostationary data models. In particular, we are interested to explore other statistical structures and properties that can help enhance estimation performance.

9. REFERENCES

[1] E. J. Candes, and M. B. Wakin, "An Introduction To Compressive Sampling," *IEEE Signal Processing Magazine*, Vol. 25, no. 2, pp. 21-30, March 2008.

[2] J. Friedman, T. Hastie, and R. Tibshirani, "Sparse inverse covariance estimation with the graphical lasso," *Biostatistics*, Vol. 9, pp. 432-441, 2008.

[3] P. Bickel, E. Levina, "Covariance regularization by thresholding," *The Annuals of Statistics*, vol. 36, pp. 2577-2604, 2008.

[4] O. Banerjee, L. El Ghaoui, and A. d'Aspremont, "Model Selection Through Sparse Maximum Likelihood Estimation for Multivariate Gaussian or Binary Data," *Journal of Machine Learning Research*, Vol. 9, pp. 485-516, March 2008.

[5] J. Bien, R. Tibshirani, "Sparse Estimation of a Covariance Matrix," *Biometrika*, Vol. 98(4), pp. 807-820, 2011.

[6] P. Ravikumar, M. J. Wainwright, G. Raskutti and B. Yu, "High-dimensional covariance estimation by minimizing ℓ_1-penalized log-determinant divergence," *Electronic Journal of Statistics*, Vol. 5, pp. 935-980, 2011.

[7] X. Ma, "Alternating Direction Method of Multipliers for Covariance Selection Models," *Journal of Scientific Computing*, Vol. 51, pp. 261-273, 2012.

[8] K. Scheinberg, S. Ma and D. Goldfarb, "Sparse Inverse Covariance Selection via Alternating Linearization Mathods," *Proceedings of NIPS Conf.*, 2010.

[9] G. Leus and D. D. Ariananda, "Power spectrum blind sampling," *IEEE Signal Processing Letter*, vol. 18, no. 8, pp. 443-446, Aug. 2011.

[10] G. Leus, Z. Tian, "Recovering Second-Order Statistics from Compressive Measurements," *IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing* (CAMSAP), pp. 337-340, Dec. 2011.

[11] Z. Tian, Y. Tafesse, and B. M. Sadler, "Cyclic Feature Detection from Sub-Nyquist Samples for Wideband Spectrum Sensing," *IEEE Journal of Selected Topics in Signal Processing, Special Issue on Robust Measures and Tests Using Sparse Data for Detection and Estimation*, vol. 6, no. 1, pp. 58-69, February 2012.

[12] L. An and P. Tao, "The dc (difference of convex functions) programming and dca revisited with dc models of real world nonconvex optimization problems," *Ann. Oper. Res.* Vol. 133, pp. 23-46, 2005.

[13] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer, 2th edition, 2009.

[14] M. Grant and S. Boyd, CVX: Matlab Software for Disciplined Convex Programming, [Online]. Available: http://cvxr.com/cvx/