

COVALSA: COVARIANCE ESTIMATION FROM COMPRESSIVE MEASUREMENTS USING ALTERNATING MINIMIZATION

José M. Bioucas-Dias⁽¹⁾, Deborah Cohen⁽²⁾, and Yonina C. Eldar⁽²⁾

⁽¹⁾ Telecomunicações, Instituto Superior Técnico, Universidade de Lisboa, Portugal

⁽²⁾ Technion - Israel Institute of Technology, Haifa, Israel

ABSTRACT

The estimation of covariance matrices from compressive measurements has recently attracted considerable research efforts in various fields of science and engineering. Owing to the small number of observations, the estimation of the covariance matrices is a severely ill-posed problem. This can be overcome by exploiting prior information about the structure of the covariance matrix. This paper presents a class of convex formulations and respective solutions to the high-dimensional covariance matrix estimation problem under compressive measurements, imposing either Toeplitz, sparseness, null-pattern, low rank, or low permuted rank structure on the solution, in addition to positive semi-definiteness. To solve the optimization problems, we introduce the CoVariance by Augmented Lagrangian Shrinkage Algorithm (CoVALSA), which is an instance of the Split Augmented Lagrangian Shrinkage Algorithm (SALSA). We illustrate the effectiveness of our approach in comparison with state-of-the-art algorithms.

Index Terms— Covariance matrix estimation, compressive acquisition, alternating optimization, SALSA

1. INTRODUCTION

Numerous applications in machine learning, economics and financial time series analysis, optics, communications, require estimation of the signal statistics, especially its covariance, rather than the signal itself. In certain circumstances, when the number of measurements is too small, the signal cannot even be recovered. Oftentimes, covariance matrices have structure that can be exploited for their recovery, which is not necessarily the case for the signals themselves. Furthermore, the averaging performed in evaluating the statistics can lead to improved performance in noisy environments over sparse recovery techniques applied directly to the signals.

In this paper, we are concerned with the estimation of covariance matrices under compressive measurements [1, 2]. This problem is severely ill-posed and calls for some form of prior information imposing constraints on the estimated covariance matrices and thereby improving the conditioning of

the inverse problem. Representative forms of prior information are a) stationarity, which leads to Toeplitz covariance matrices [3], b) diagonal covariance [4, 5], c) sparsity either on the covariance [6–8], its inverse [9–11], or its Fourier transform [3–5] d) null-pattern (*e.g.*, banded matrices) [12], e) low rank [13, 14], and d) Kronecker product [15, 16].

A popular approach to perform covariance matrix estimation under prior constraints is to turn it into a vector recovery estimation problem, by use of Kronecker products and vec operators [3–5]. The vector containing the unknown correlation values can then be recovered from the resulting set of linear measurements using constrained linear estimation methods. Although popular, this approach has two notable drawbacks. The first is that in general the use of the Kronecker product enlarges the dimensionality of the problem, increasing the computational complexity. The second, and more important, is that in vector form, some of the structure of the problem is lost. In particular, it is difficult to enforce positive semidefiniteness of the covariance matrix in this form. Thus, the vector approach typically ignores this constraint, and the resulting estimated covariance values do not necessarily lead to a valid covariance matrix. By taking into account the fact that any covariance matrix must be positive semidefinite, we can improve recovery ability and obtain satisfactory performance even from a small number of compressive measurements.

In order to properly exploit prior information in covariance estimation we introduce a class of convex formulations and respective solutions to the high-dimensional covariance matrix estimation problem under compressive measurements. We consider the following types of covariance structure: a) Toeplitz, b) sparseness, c) null-pattern, d) low rank, and e) low permuted rank. We also constrain the covariance matrix to be positive semidefinite. All formulations yield semidefinite programs which we solve effectively with an instance of the SALSA algorithm [17] herein termed CoVariance by Augmented Lagrangian Shrinkage Algorithm (CoVALSA). We note that the work [25] addresses conditions under which unknown sparse matrices observed in the compressed regime may be recovered exactly and efficiently using a convex program. In this paper, we are also concerned with matrix recovering from compressive measurements. We consider, however, a different problem scenario concerning 1) the optimization variables, which are covariance matrices to be inferred from sample covariance counterparts, and 2) the classes of regularizers/constraints considered.

⁽²⁾This work is supported in part by the Israel Science Foundation under Grant no. 170/10, in part by the SRC, in part by the Intel Collaborative Research Institute for Computational Intelligence (ICRI-CI), and in part by the Portuguese Science and Technology Foundation under ProjectsPEst-OE/EEI/LA0008/2013 and PTDC/EEI-PRO/1470/2012

Our approach is based on the alternating method of multipliers (ADMM), adapted to our setting. CoVALSA has two distinctive features: a) it allows to solve a number of disparate problems in a highly flexible and unified framework, and b) it has complexity of $O(n^3)$ per iteration for covariance matrices of size n , which is much smaller than $O(n^6 \log(1/\varepsilon))$ of the interior point methods (e.g., SeDuMi and SDPT3) widely used to solve small scale semidefinite programs. In comparison to the popular Kronecker-based vector formulation, it results in problems of smaller dimension and yields improved performance particularly in low signal to noise ratio (SNR) regimes.

The paper is organized as follows. Section 2 formulates the constrained covariance estimation problem. The general template of the CoVALSA algorithm and the instances related to the different types of covariance structure mentioned above are discussed in Section 3. Section 4 presents experimental results illustrating the effectiveness of the proposed algorithm and comparing it to the traditional vector approach.

2. COMPRESSIVE COVARIANCE ESTIMATION

2.1. Problem Formulation

Let $\mathbf{x} \in \mathbb{R}^n$ denote a zero-mean random vector with covariance matrix $\mathbf{X} \equiv \mathbb{E}[\mathbf{x}\mathbf{x}^T]$, where $(\cdot)^T$ denotes the transpose operator. Our goal is to recover \mathbf{X} from compressive measurements

$$\mathbf{Y} = \mathbf{A}\mathbf{X}\mathbf{A}^T, \quad (1)$$

where $\mathbf{A} \in \mathbb{R}^{m \times n}$ is a sampling matrix with $m < n$.

One way in which the model (1) may arise is when we are given a sequence of compressed measurements $\mathbf{y}_t = \mathbf{A}\mathbf{x}_t$, where \mathbf{x}_t , for $t = 1, \dots, L$, are sampled from independent random vectors distributed as \mathbf{x} . In general, we cannot recover \mathbf{x}_t from \mathbf{y}_t since there are fewer measurements than unknowns. However, if the covariance matrix \mathbf{X} has structure, then we can attempt to recover it from the empirical covariance of the vectors \mathbf{y}_t .

Specifically, define \mathbf{Y} as the sample covariance matrix of \mathbf{y}_t , for $t = 1, \dots, L$. We then have that

$$\begin{aligned} \mathbf{Y} &\equiv \frac{1}{L} \sum_{i=1}^L \mathbf{y}_i \mathbf{y}_i^T = \mathbf{A} \left(\frac{1}{L} \sum_{i=1}^L \mathbf{x}_i \mathbf{x}_i^T \right) \mathbf{A}^T \\ &= \mathbf{A}\mathbf{X}\mathbf{A}^T + \underbrace{\mathbf{A} \left(\frac{1}{L} \sum_{i=1}^L \mathbf{x}_i \mathbf{x}_i^T \right) \mathbf{A}^T - \mathbf{A}\mathbf{X}\mathbf{A}^T}_{\mathbf{N}}. \end{aligned} \quad (2)$$

In this work, we only assume that \mathbf{x} is zero-mean, and has finite covariance \mathbf{X} . Under these conditions, the perturbation \mathbf{N} defined in (2) satisfies $\mathbb{E}[\mathbf{N}] = \mathbf{0}$ and $\mathbb{E}\|\mathbf{N}\|_F^2 \propto (1/L)$, where $\|\mathbf{N}\|_F^2 \equiv \text{tr}(\mathbf{N}\mathbf{N}^T)$ is the Frobenius norm of \mathbf{N} . Motivated by these properties of \mathbf{N} , we may formulate the covariance matrix estimation as the optimization problem

$$\begin{aligned} \min_{\mathbf{X}} \quad & (1/2)\|\mathbf{Y} - \mathbf{A}\mathbf{X}\mathbf{A}^T\|_F^2 + \lambda\phi(\mathbf{X}) \\ \text{s.t.} \quad & \mathbf{X} \in S_+. \end{aligned} \quad (3)$$

Here $\phi : \mathbb{R}^{n \times n} \mapsto [-\infty, \infty]$ is a closed, proper, and convex function, termed the regularizer, promoting solutions with *a priori* known characteristics, $\lambda > 0$ is a regularization parameter setting the relative weight between the data term and the regularizer, and S_+ denotes the convex set of symmetric and positive semi-definite matrices of a given fixed dimension.

In the formulation (3) we implicitly assume that the covariance matrix \mathbf{X} has structure which can be incorporated in order to promote stable recovery. In describing the structure it will be useful to define the indicator function $\iota_C : \mathbb{R}^{n \times n} \mapsto \mathbb{R}$ of the set C , by

$$\iota_C(\mathbf{X}) = \begin{cases} 0, & \mathbf{X} \in C \\ \infty, & \mathbf{X} \notin C. \end{cases} \quad (4)$$

Herein, we consider the following instances of ϕ , which correspond to different types of structure:

- R1) **Toeplitz:** $\phi(\mathbf{X}) = \iota_{T_k}(\mathbf{X})$, where T_k is the set of k banded Toeplitz matrices (i.e., $\mathbf{X} \in T_k$ iff $X_{ij} = X_{j-i}$ and $X_{i,j} = 0$ for $|j-i| > k$). The Toeplitz structure arises when the random vector is sampled from a second-order stationary process with finite-length correlation.
- R2) **sparseness:** $\phi(\mathbf{X}) = \|\mathbf{X}\|_1 \equiv \sum_{i,j} |X_{ij}|$.
- R3) **zero-pattern:** $\phi(\mathbf{X}) = \iota_M(\mathbf{X})$, where M is a mask (i.e., $M_{ij} \in \{0, 1\}$) such that $M \equiv \{\mathbf{X} \in \mathbb{R}^{n \times n} : \mathbf{X} = M \circ \mathbf{X}\}$, and \circ denotes the Hadamard product.
- R4) **low rank:** $\phi(\mathbf{X}) = \|\mathbf{X}\|_*$ is the nuclear norm of \mathbf{X} given by $\|\mathbf{X}\|_* \equiv \sum_{i=1}^n \sigma_i$, where $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n \geq 0$ are the singular values of \mathbf{X} .
- R5) **low permuted rank:** $\phi(\mathbf{X}) = \|\mathcal{R}_q(\mathbf{X})\|_*$, where $\mathcal{R}_q : \mathbb{R}^{n \times n} \mapsto \mathbb{R}^{p^2 \times q^2}$ is a permutation rearrangement operator such that given $\mathbf{A} \in \mathbb{R}^{p \times p}$ and $\mathbf{B} \in \mathbb{R}^{q \times q}$ with $n = pq$, $\mathcal{R}_q(\mathbf{A} \otimes \mathbf{B}) = \text{vec}(\mathbf{A})\text{vec}(\mathbf{B})^T$. Here $\text{vec}(\mathbf{A})$ denotes the column vector formed by staking the columns of \mathbf{A} . The regularizer $\|\mathcal{R}_q(\mathbf{X})\|_*$ promotes covariance matrices given by low rank Kronecker product expansions [18].

2.2. Vector Formulation

A popular approach to solve (3) is to turn the problem into vector form. Specifically, by using properties of the Kronecker operation, we can rewrite (1) as

$$\text{vec}(\mathbf{Y}) = (\mathbf{A} \otimes \mathbf{A}) \text{vec}(\mathbf{X}), \quad (5)$$

where $\text{vec}(\mathbf{Y})$ stacks the columns of \mathbf{Y} into a column vector. Denoting $\mathbf{y} = \text{vec}(\mathbf{Y})$, $\mathbf{x} = \text{vec}(\mathbf{X})$, and $\mathbf{C} = \mathbf{A} \otimes \mathbf{A}$, we have that

$$\|\mathbf{Y} - \mathbf{A}\mathbf{X}\mathbf{A}^T\|_F^2 = \|\mathbf{y} - \mathbf{C}\mathbf{x}\|_2^2. \quad (6)$$

The regularizers $\phi(\mathbf{X})$ can also be described directly in terms of \mathbf{x} . This often necessitates changing slightly the regularizer

function to $\psi(\mathbf{x})$ such that $\psi(\mathbf{x}) = \phi(\mathbf{X})$. Problem (3) can then be written as

$$\min_{\mathbf{x}} (1/2)\|\mathbf{y} - \mathbf{C}\mathbf{x}\|_2^2 + \lambda\psi(\mathbf{x}). \quad (7)$$

However, in this approach, the positive semi-definite constraint on \mathbf{X} is not enforced. Furthermore, the length of \mathbf{y} is equal to m^2 and the length of \mathbf{x} is n^2 so that the problem dimensions have increased considerably. More specifically, the overall dimensions of the data to be reconstructed do not change. However the Kronecker product increases the measurement matrix dimensions since \mathbf{C} is a $m^2 \times n^2$ matrix.

Several examples of this approach have been previously considered in the literature for some of the regularizers mentioned above. In [3], the authors treat the case in which \mathbf{x} is stationary and therefore its covariance matrix \mathbf{X} is Toeplitz. Diagonal covariances matrices are considered in [4, 5]. A sparsity constraint can be further added to the above structures [3–5]. Before discussing our matrix approach to covariance estimation, we note that in [19], the authors solve the optimization problem (3) directly in matrix form. In [20] the authors consider an extension of OMP and FISTA to treat the matrix formulation of (3) for arbitrary matrices \mathbf{X} subject to a sparsity prior. However, both works do not include the positive semidefinite constraint.

3. COVALSA ALGORITHM

In this section, we develop the CoVALSA algorithm which is an instance of the SALSAs methodology introduced in [17]. We start by converting the constrained optimization problem (3) into the equivalent version

$$\begin{aligned} \min_{\mathbf{X}} f_1(\mathbf{X}) + f_2(\mathbf{V}) \\ \text{subject to } \mathbf{V} = \mathcal{G}(\mathbf{X}). \end{aligned} \quad (8)$$

Here

$$f_1(\mathbf{X}) \equiv 0 \quad (9)$$

$$f_2(\mathbf{V}) \equiv g_1(\mathbf{V}_1) + g_2(\mathbf{V}_2) + g_3(\mathbf{V}_3) \quad (10)$$

with

$$g_1(\mathbf{V}_1) \equiv (1/2)\|\mathbf{Y} - \mathbf{A}\mathbf{V}_1\mathbf{A}^T\|_F^2 \quad (11)$$

$$g_2(\mathbf{V}_2) \equiv \iota_{S_+}(\mathbf{V}_2) \quad (12)$$

$$g_3(\mathbf{V}_3) \equiv \lambda\phi(\mathcal{H}(\mathbf{V}_3)) \quad (13)$$

and

$$\underbrace{\begin{bmatrix} \mathbf{V}_1 \\ \mathbf{V}_2 \\ \mathbf{V}_3 \end{bmatrix}}_{\mathbf{V} \in \mathbb{R}^{n_1 \times n_2}} = \underbrace{\begin{bmatrix} \mathcal{I} \\ \mathcal{I} \\ \mathcal{H} \end{bmatrix}}_{\mathcal{G}}(\mathbf{X})$$

where \mathcal{I} denotes the identity operator. It holds that $\mathcal{H} \equiv \mathcal{I}$ for the regularizers R1), R2), R3), and R4), and $\mathcal{H} \equiv \mathcal{R}_q$ for the regularizer R5) (*i.e.*, the permutation rearrangement operator). Finally, n_1, n_2 are, respectively, the number of rows and of columns of \mathbf{V} .

Algorithm CoVALSA

1. Set $k = 0$, choose $\mu > 0$, $\mathbf{V}_0 = (\mathbf{V}_{1,0}, \mathbf{V}_{2,0}, \mathbf{V}_{3,0})$, and
2. $\mathbf{D}_0 = (\mathbf{D}_{1,0}, \mathbf{D}_{2,0}, \mathbf{D}_{3,0})$
3. **repeat**
4. $\mathbf{X}_{k+1} := \arg \min_{\mathbf{X}} \|\mathcal{G}(\mathbf{X}) - \mathbf{V}_k - \mathbf{D}_k\|_F^2$
5. $\boldsymbol{\nu}_1 := \mathbf{X}_{k+1} - \mathbf{D}_{1,k}$
6. $\mathbf{V}_{1,k+1} := \arg \min_{\mathbf{V}_1} g_1(\mathbf{V}_1) + \frac{\mu}{2}\|\mathbf{V}_1 - \boldsymbol{\nu}_1\|_F^2$
7. $\boldsymbol{\nu}_2 := \mathbf{X}_{k+1} - \mathbf{D}_{2,k}$
8. $\mathbf{V}_{2,k+1} := \arg \min_{\mathbf{V}_2} g_2(\mathbf{V}_2) + \frac{\mu}{2}\|\mathbf{V}_2 - \boldsymbol{\nu}_2\|_F^2$
9. $\boldsymbol{\nu}_3 := \mathcal{G}(\mathbf{X}_{k+1}) - \mathbf{D}_{3,k}$
10. $\mathbf{V}_{3,k+1} := \arg \min_{\mathbf{V}_3} g_3(\mathbf{V}_3) + \frac{\mu}{2}\|\mathbf{V}_3 - \boldsymbol{\nu}_3\|_F^2$
11. $\mathbf{D}_{1,k+1} := -\boldsymbol{\nu}_1 + \mathbf{V}_{1,k+1}$
12. $\mathbf{D}_{2,k+1} := -\boldsymbol{\nu}_2 + \mathbf{V}_{2,k+1}$
13. $\mathbf{D}_{3,k+1} := -\boldsymbol{\nu}_3 + \mathbf{V}_{3,k+1}$
14. $k \leftarrow k + 1$
15. **until** stopping criterion is satisfied.

Fig. 1. CoVariance by Augmented Lagrangian Shrinkage Algorithm (CoVALSA).

The next step consists in applying ADMM [21, 22] to (8). The following is a simplified version of a theorem of Eckstein and Bertsekas, adapted to our setting, stating convergence conditions of ADMM.

Theorem 1 ([21]) *Let $\text{kernel}(\mathcal{G}) = \{0\}$ and f_2 be closed, proper, and convex. Consider arbitrary $\mu > 0$ and $\mathbf{V}_0, \mathbf{D}_0 \in \mathbb{R}^{n_1 \times n_2}$. Consider three sequences $\{\mathbf{X}_k \in \mathbb{R}^{n_1 \times n_2}, k = 0, 1, \dots\}$, $\{\mathbf{V}_k \in \mathbb{R}^{n_1 \times n_2}, k = 0, 1, \dots\}$, and $\{\mathbf{D}_k \in \mathbb{R}^{n_1 \times n_2}, k = 0, 1, \dots\}$ that satisfy*

$$\mathbf{X}_{k+1} = \arg \min_{\mathbf{X}} \|\mathcal{G}(\mathbf{X}) - \mathbf{V}_k - \mathbf{D}_k\|_F^2 \quad (14)$$

$$\mathbf{V}_{k+1} = \arg \min_{\mathbf{V}} f_2(\mathbf{V}) + \frac{\mu}{2}\|\mathcal{G}(\mathbf{X}_{k+1}) - \mathbf{V} - \mathbf{D}_k\|_F^2 \quad (15)$$

$$\mathbf{D}_{k+1} = \mathbf{D}_k - [\mathcal{G}(\mathbf{X}_{k+1}) - \mathbf{V}_{k+1}]. \quad (16)$$

Then, if (8) has a solution, the sequence $\{\mathbf{X}_k\}$ converges to it; otherwise, at least one of the sequences $\{\mathbf{V}_k\}$ or $\{\mathbf{D}_k\}$ diverges.

The proof of Theorem 1 uses the equivalence between ADMM and the Douglas-Rachford splitting (DRS) method applied to the dual of (8).

Fig. 1 shows the pseudocode of the derived algorithm, which we name CoVariance by Augmented Lagrangian Shrinkage Algorithm (CoVALSA). In order to implement it, we need to solve the optimizations shown in lines 4, 6, 8, and 10. We note that the operator \mathcal{R}_q is a permutation so that $\mathcal{R}_q^* \mathcal{R}_q = \mathcal{I}$ [$(\cdot)^*$ denotes adjoint operator]. It follows then $\mathcal{H}^* \mathcal{H} = \mathcal{I}$ for all five regularizers and then $\mathcal{G}^* \mathcal{G} = 3\mathcal{I}$. The solution of quadratic minimization stated in line 4 is then

$$\begin{aligned} \mathbf{X}_{k+1} &:= (1/3)\mathcal{G}^*(\mathbf{V}_k + \mathbf{D}_k) \\ &:= (1/3)((\mathbf{V}_{1,k} + \mathbf{D}_{1,k}) + (\mathbf{V}_{2,k} + \mathbf{D}_{2,k}) \\ &\quad + \mathcal{H}^*(\mathbf{V}_{3,k} + \mathbf{D}_{3,k})). \end{aligned} \quad (17)$$

The remaining optimizations are addressed in the next section.

3.1. Moreau proximity operators

The optimizations shown in lines 6, 8, and 10 are the Moreau proximity operators [23] for the convex functions g_1 , g_2 , and g_3 , respectively, which are interpretable as generalizations of projections onto convex sets.

3.1.1. Moreau proximity operator for g_1

From line 6 of CoVALSA, we have

$$\psi_{g_1/\mu}(\boldsymbol{\nu}) = \arg \min_{\mathbf{V}_1} \frac{1}{2} \|\mathbf{Y} - \mathbf{A}\mathbf{V}_1\mathbf{A}^T\|_F^2 + \frac{\mu}{2} \|\mathbf{V}_1 - \boldsymbol{\nu}\|_F^2,$$

which can be computed by solving the linear system of equations

$$\mathbf{A}^T \mathbf{A} \mathbf{V}_1 \mathbf{A}^T \mathbf{A} + \mu \mathbf{X} = \mathbf{A}^T \mathbf{Y} \mathbf{A} + \mu \boldsymbol{\nu}.$$

We then use the vec properties and the eigendecomposition $\mathbf{A}^T \mathbf{A} = \mathbf{E} \boldsymbol{\Lambda} \mathbf{E}^T$ where $\mathbf{E} \in \mathbb{R}^{n \times m}$ holds the m eigenvectors corresponding to the m largest eigenvalues (the remaining $n - m$ are zero). After some manipulation, we have

$$\psi_{g_1/\mu}(\boldsymbol{\nu}) = \frac{1}{\mu} \mathbf{Z} - \mathbf{E}[(\mathbf{E}^T \mathbf{Z} \mathbf{E}) \circ \mathbf{D}] \mathbf{E}^T, \quad (18)$$

where $\mathbf{Z} \equiv \mathbf{A}^T \mathbf{Y} \mathbf{A} + \mu \boldsymbol{\nu}$ and $\mathbf{D} \equiv (\mathbf{a} \mathbf{a}^T) \circ [(\mathbf{a} \mathbf{a}^T + \mu) \mu]$, with $\mathbf{a} \equiv \text{diag}(\boldsymbol{\Lambda})$ and \circ standing for Hadamard division.

3.1.2. Moreau proximity operator for g_2

From line 8 of CoVALSA, we have

$$\begin{aligned} \psi_{g_2/\mu}(\boldsymbol{\nu}) &= \arg \min_{\mathbf{V}_2} \iota_{S_+}(\mathbf{V}_2) + \frac{\mu}{2} \|\mathbf{V}_2 - \boldsymbol{\nu}\|_F^2 \quad (19) \\ &= \mathbf{E}(\boldsymbol{\Lambda})_+ \mathbf{E}^T, \quad (20) \end{aligned}$$

where $(\mathbf{E}, \boldsymbol{\Lambda})$ are the eigenvalues and eigenvectors of the $(\boldsymbol{\nu} + \boldsymbol{\nu}^T)/2$ (the symmetric part of $\boldsymbol{\nu}$) and $(\boldsymbol{\Lambda})_+$ is the non-negative part of $\boldsymbol{\Lambda}$ [24].

3.1.3. Moreau proximity operator for g_3

From line 10 of CoVALSA, we have

$$\psi_{g_3/\mu}(\boldsymbol{\nu}) = \arg \min_{\mathbf{V}_3} \lambda \phi(\mathbf{V}_3) + \frac{\mu}{2} \|\mathbf{V}_3 - \boldsymbol{\nu}\|_F^2. \quad (21)$$

The solution of (21), for the different regularizers, is the following:

$$\begin{aligned} \text{R1)} \quad & \psi_{g_3/\mu}(\boldsymbol{\nu}) = \mathcal{T}_k(\boldsymbol{\nu}) \\ \text{R2)} \quad & \psi_{g_3/\mu}(\boldsymbol{\nu}) = \max\{0, \boldsymbol{\nu} - \lambda/\mu\} \text{sign}(\boldsymbol{\nu}) \\ \text{R3)} \quad & \psi_{g_3/\mu}(\boldsymbol{\nu}) = \mathbf{M} \circ \boldsymbol{\nu} \\ \text{R4)} \quad & \psi_{g_3/\mu}(\boldsymbol{\nu}) = \mathbf{E}(\boldsymbol{\Lambda})_+ \mathbf{F}^T \\ \text{R5)} \quad & \psi_{g_3/\mu}(\boldsymbol{\nu}) = \mathbf{E}(\boldsymbol{\Lambda})_+ \mathbf{F}^T, \end{aligned} \quad (22)$$

where $\mathcal{T}_k(\boldsymbol{\nu})$ replaces the diagonals $-k, \dots, k$ of $\boldsymbol{\nu}$ with its mean value and with zero elsewhere, $\psi_{g_3/\mu}$ for R2) is the componentwise soft threshold and $(\mathbf{E} \boldsymbol{\Lambda} \mathbf{F}^T)$ is the singular value decomposition of $\boldsymbol{\nu}$ [24].

The computational complexity of CoVALSA per iteration is $O(n^3)$ and is dominated by the eigendecomposition and the singular value decomposition necessary to compute $\psi_{g_2/\mu}$ and $\psi_{g_3/\mu}$, respectively.

4. EXPERIMENTAL RESULTS

To illustrate the effectiveness of CoVALSA, we show results of experiments only for the Toeplitz constraint, due to lack of space. The measurement matrices are sampled from a δ -random bipartite ensemble: $\mathbf{A} \in \{0, 1\}^{m \times n}$ are drawn independently (without replacement) and uniformly from the δ -random bipartite ensemble (see Definition 4 [25] for details).

In this section, we consider two sources of noise. The first one is the perturbation due to the finite number of samples, as described in Section 2.1. The equivalent SNR corresponding to L measurement vectors is given by

$$\text{SNR}_L = 10 \log_{10} \left(L \frac{\|\mathbf{Y}\|_F^2}{\|\mathbf{Y}\|_F^2 + (\sum_{i=1}^m \mathbf{Y}_{ii})^2} \right). \quad (23)$$

We then consider additive noise in the form of an independent identically distributed (iid) Gaussian matrix \mathbf{N}_g . In that context, the SNR is defined as

$$\text{SNR} = 10 \log_{10} \frac{\|\mathbf{A} \mathbf{X} \mathbf{A}^T\|_F^2}{\|\mathbf{N}_g\|_F^2}. \quad (24)$$

Figures 2 and 3 shows the Toeplitz covariance matrix $\mathbf{X} = [X_{ij}]$, with $X_{i,j} = 0.7^{|i-j|}$ and its estimation, with $n = 100$ and $m = 50$, without and with noise \mathbf{N} due to the finite number of samples, respectively.

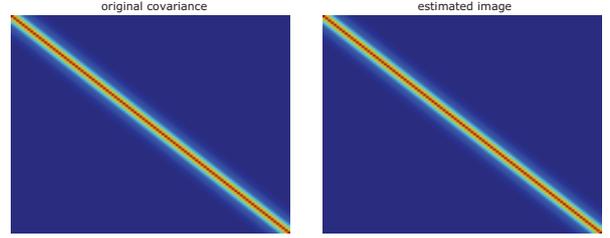


Fig. 2. Toeplitz covariance matrix. Left: original. Right: estimated with $\mathbf{N} = \mathbf{0}$.

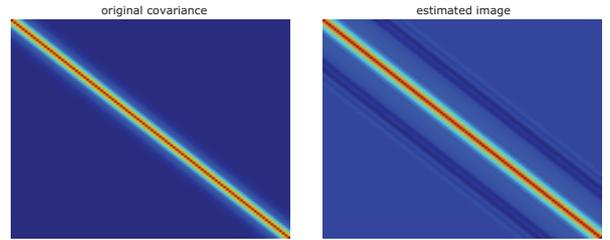


Fig. 3. Toeplitz covariance matrix. Left: original. Right: estimated with $L = 200$.

We then compare CoVALSA to the traditional vectorization approach mentioned in Section 2.2 for the Toeplitz regularizer. This last approach does not enforce the semi-positive definite constraint. We consider both measurement matrices drawn from a δ -random bipartite ensemble, as described above, and normalized iid Gaussian matrices. Figures 4 shows both the performances of CoVALSA and of the standard vectorization approach described in Section 2.2, for different values of the SNR and L , respectively. Each experiment is repeated over 20 realizations. We observe that

our approach outperforms the traditional vector approach, especially in low SNR regimes.

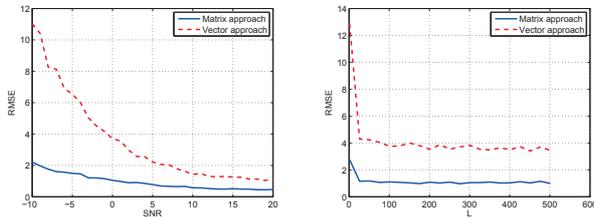


Fig. 4. Relative RMSE of the matrix and vector approaches as a function of the SNR with $L = 200$ (left) and as a function of L with SNR= 0dB (right).

5. CONCLUDING REMARKS

The paper presented a class of convex formulations and respective solutions to the high-dimensional covariance matrix estimation problem under compressive measurements imposing either Toeplitz, sparseness, null-pattern, low rank, or low permuted rank structure on the solution, in addition to positive semi-definiteness. To solve the optimization problems, we introduced the CoVariance by Augmented Lagrangian Shrinkage Algorithm (CoVALSA), which is an instance of the Split Augmented Lagrangian Shrinkage Algorithm (SALSA). Compared with the popular formulations based on Kronecker products and vec operators, and by enforcing positive semi-definiteness on the estimated covariance matrices, the proposed formulation yields improved recovery, even from a small number of compressive measurements. We observe that our approach outperforms the traditional vector approach, especially in low SNR regimes, where the prior information enforced by the symmetric and positive semi-definite constraint has the highest impact.

REFERENCES

- [1] D. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [2] Y.C. Eldar and G. Kutyniok, *Compressed Sensing: Theory and Applications*, Cambridge University Press, 2012.
- [3] D. Ariananda and G. Leus, "Compressive wideband power spectrum estimation," *IEEE Trans. on Sig. Proc.*, vol. 60, no. 9, pp. 4775–4789, 2012.
- [4] A. Lexa, M. Davies, J. Thompson, and J. Nikolic, "Compressive power spectral density estimation," in *ICASSP*, 2011, pp. 3884–3887.
- [5] D. Cohen and Y. C. Eldar, "Sub-Nyquist sampling for power spectrum sensing in cognitive radios: A unified approach," *CoRR*, vol. abs/1308.5149, 2013.
- [6] A. Nasif, Z. Tian, and Q. Ling, "High-dimensional sparse covariance estimation for random signals," in *ICASSP*, 2013, pp. 4658–4662.
- [7] J. Bickel and E. Levina, "Covariance regularization by thresholding," *The Annals of Statistics*, vol. 36, no. 6, pp. 2577–2604, 2008.
- [8] R. Tibshirani J. Bien, "Sparse estimation of a covariance matrix," *Biometrika*, vol. 98, no. 4, pp. 807–820, 2011.
- [9] A. Rothman, P. Bickel, E. Levina, and J. Zhu, "Sparse permutation invariant covariance estimation," *Electronic Journal of Statistics*, vol. 2, pp. 494–515, 2008.
- [10] C. Johnson, A. Jalali, and P. Ravikumar, "High-dimensional sparse inverse covariance estimation using greedy methods," in *AISTATS*, Neil D. Lawrence and Mark A. Girolami, Eds., 2012, vol. 22, pp. 574–582.
- [11] P. Ravikumar, M. 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.
- [12] Peter J Bickel and Yulia R Gel, "Banded regularization of autocovariance matrices in application to parameter estimation and forecasting of time series," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 73, no. 5, pp. 711–728, 2011.
- [13] K. Lounici, "High-dimensional covariance matrix estimation with missing observations," *arXiv preprint arXiv:1201.2577*, 2012.
- [14] M. Johnstone and A. Lu, "On consistency and sparsity for principal components analysis in high dimensions," *Journal of the American Statistical Association*, vol. 104, no. 486, pp. 682–693, 2009.
- [15] T. Tsiligkaridis, A. Hero III, and S. Zhou, "On convergence of kronecker graphical lasso algorithms," *IEEE Trans. on Sig. Proc.*, vol. 61, no. 7, pp. 1743–175, 2013.
- [16] K. Werner, M. Jansson, and P. Stoica, "On estimation of covariance matrices with kronecker product structure," *IEEE Trans. on Sig.l Proc.*, vol. 56, no. 2, pp. 478–491, 2008.
- [17] M. Afonso, J. Bioucas-Dias, M., and Figueiredo, "An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems," *IEEE Transactions on Image Processing*, vol. 20, no. 3, pp. 681–695, 2011.
- [18] T. Tsiligkaridis and A. Hero III, "Covariance estimation in high dimensions via kronecker product expansions," *arXiv preprint arXiv:1302.2686*, 2013.
- [19] S. A. Razavi, M. Valkama, and D. Cabric, "High-resolution cyclic spectrum reconstruction from sub-Nyquist samples," *IEEE SPAWC*, pp. 205–254, 2013.
- [20] T. Wimalajeewa, Y. C. Eldar, and P. K. Varshney, "Recovery of sparse matrices via matrix sketching," *CoRR*, vol. abs/1311.2448, 2013.
- [21] D. Bertsekas J. Eckstein, "On the douglas-rachford splitting method and the proximal point algorithm for maximal monotone operators," *Mathematical Programming*, vol. 55, no. 1-3, pp. 293–318, 1992.
- [22] D. Gabay and B. Mercier, "A dual algorithm for the solution of nonlinear variational problems via finite element approximation," *Computers & Mathematics with Applications*, vol. 2, no. 1, pp. 17–40, 1976.
- [23] P. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, pp. 185–212. 2011.
- [24] J.-B. Urruty and C. Lemaréchal, *Fundamentals of convex analysis*, Springer, 2001.
- [25] G. Dasarathy, P. Shah, B. Bhaskar, and R. Nowak, "Sketching sparse matrices," *arXiv preprint arXiv:1303.6544*, 2013.