Distributed covariance estimation for compressive signal processing

Distributed covariance estimation for compressive signal processing

Matteo Testa and Enrico Magli
Department of Electronics and Telecommunications – Politecnico di Torino (Italy)

***Abstract*—In this paper we present a novel technique for the distributed estimation of the covariance matrix of an additive colored noise process affecting Compressed Sensing (CS) measurements. The main application is in wireless sensor networks, where nodes sense signals in CS format in order to save energy in the computation and transmission stages. The proposed technique enables a variety of compressive signal processing operations to be performed at each node directly on the linear measurements, such as detection, exploiting the knowledge of the noise statistics, thereby achieving improved performance. The parametric approach we introduce promises to yield good results while keeping the communication cost low. Hence, we validate our technique by evaluating the error on the estimated covariance matrix, and by including it in a compressive detection task.**

***Keywords*—compressed sensing, autoregressive model, distributed estimation, detection.**

I. Introduction

In recent years, wireless sensor networks (WSN) emerged as an inexpensive way to collect spatially distributed data such as temperature or gas concentrations. The aim of WSN is not only that of collecting distributed data, but also that of collaborating to accomplish a task (*e.g.,* estimating the covariance matrix of a process) given that each node has only a portion of the whole data required for the task to be accomplished.

Since signal transmission among nodes in a WSN is limited due to energy consumption of the radio interface, compressed sensing (CS) which allows to *compress while acquiring* a signal, emerged as a viable solution to reduce the communication load in WSN. Moreover, since the recovery of a compressed signal is computationally expensive and often is not required (*e.g.,* only some parameters of the signals are needed), detection and estimation problems from CS measurements [1] are extremely suitable to be performed directly at nodes. A common assumption is to model the noise affecting the signal with additive white gaussian noise (AWGN), in which case the noise variance is a sufficient statistic to perform signal processing operations. However, when the noise affecting the signal is not AWGN, estimating the covariance matrix of the noise process is fundamental to improve the detection/estimation performance in a variety of inference techniques [2].

The problem of distributed covariance estimation, which has received a lot of attention recently in the non-compressed case, is mainly addressed by the distributed estimation of the principal eigenvectors or eigenvalues of the covariance matrix. Moreover, a large number of techniques proposed in literature tackle this problem using a fusion center to which the nodes send the measurements for processing, which however, it introduces a single point of failure. Differently, the setting we consider in this paper is a fully distributed one with no fusion center and local computations at the nodes. The distributed eigenvector problem is considered in [3] via distributed estimation of the sample covariance matrix followed by local eigendecomposition. Other authors [4] propose instead algorithms able to directly estimate the smallest (largest) eigenvectors. Ad hoc algorithms for the distributed estimation of the largest eigenvalues have also been proposed as in [5]. Moreover, very few papers in literature consider the distributed architecture when the covariance matrix is that of a noise process having observation matrix Y, whose columns, distributed among the nodes, contains different realizations of the process. This setting is important for detection and estimation tasks because the resulting covariance matrix captures the statistics of the process common to the nodes. This model is considered in [6] where the authors compare the centralized estimation with two different fully distributed approaches based on the average consensus protocol to estimate the largest eigenvector of the covariance matrix of the process.

The drawback of such approaches is the potential increase of the exchanged data among nodes and hence energy consumption. In particular, the aforementioned techniques requires the nodes to exchange (with all the other nodes or with a smaller subset) signals whose length is the same as the one of the acquired signal. In this paper we also consider the problem of distributed covariance estimation. Specifically, we aim at developing estimation techniques that require a small communication load to perform the distributed estimation task, thereby requiring much less energy than existing techniques. Consequently, the proposed method differs markedly from existing methods for two main reasons: instead of estimating the eigenvectors of the covariance matrix we estimate the whole covariance matrix but in a parametric fashion in order to reduce the communication load; moreover, we estimate the covariance matrix of a noise process corrupting CS measurements instead of the original signal samples. This latter difference, besides complicating the estimation process, has interesting practical implications. In fact, once the covariance has been estimated, each node can perform signal processing operations on the compressed data, *e.g.,* inference and detection tasks [1] exploiting the knowledge of the noise statistics to obtain improved accuracy. In particular, in this paper we propose a new distributed algorithm for the estimation of the covariance matrix of a colored noise process affecting CS measurements using a parametric approach. In fact, we propose to model the colored noise with an autoregressive process (AR) of order p since this model is able to characterize the colored noise process using only few parameters. Hence, given that we model the colored noise with only few parameters, the approach we propose is extremely parsimonious on the communication cost among the nodes. Moreover, the noise covariance matrix estimate obtained with the proposed algorithm can be used

This work has received funding from the European Research Council under the European Communitys Seventh Framework Programme (FP7/2007-2013) ERC Grant agreement no. 279848.

to improve the performance of compressive detection and estimation tasks in non-AWGN noise. In order to show this, we derive a compressive detector and assess its performance, showing that with the distributed estimation of the covariance matrix indeed improves the detection accuracy.

II. Preliminaries

A. Model and assumptions

The topology of the sensor network we are considering throughout this paper is represented by a graph $G = (V, E)$. For each node of the network $v \in V$ a signal is acquired according to the model

$$y^{(v)} = \Phi^{(v)} x^{(v)} + n^{(v)}, \tag{1}$$

where $x^{(v)} \in \mathbb{R}^n$ is a sparse or compressible signal in some domain (*e.g.,* Fourier or DCT) and $\Phi^{(v)} \in \mathbb{R}^{m \times n}$ with $m < n$ is the sensing matrix: the operator which performs the dimensionality reduction. According to CS theory [7] Φ is chosen to be a random sensing matrix whose entries are distributed according to $N \sim (0, \frac{1}{m})$. Then, $\Phi^{(v)} x^{(v)}$ can be approximated as white noise [8]. The vector $n^{(v)} \in \mathbb{R}^m$ is assumed to be a colored noise process that corrupts the linear measurements. In this paper, the colored noise is approximated with a parametric model, *i.e.,* an autoregressive process of a order p, denoted as AR(p). The goal of this paper is to perform distributed estimation of the covariance matrix of the noise given that each node is corrupted by a different realization of the same noise process. Having the colored noise n approximated with an AR(p) process, we can formally define the time-varying nature of such a process as $n_t = \sum_{i=1}^{p} n_{t-i} a_i \ + w_t$ being $a = [a_1 \ \dots a_i \ a_p]^\top$ the coefficients of the regression and w the driving noise process. Then the covariance matrix of the process can be written as $C = A^{-1} A^{-\intercal}$, where $(\cdot)^{-\intercal}$ denotes the transpose of the inverse operator, $A = \text{tril}(\text{toeplitz}(u))$ and $u = [1 \ a \ 0 \ \dots \ 0]^\intercal \in \mathbb{R}^m$. In order to estimate C we need to estimate the parameter vector a. Although many estimators have been proposed in literature [9], our work relies on the least squares estimator as it allows us to obtain closed form expression for the covariance estimator.

B. Unbiased least-squares estimate

For notation simplicity let us drop the superscript (v) and consider a single node in this subsection. Given a realization $n \in \mathbb{R}^m$ of an AR(p) process and the associated regression vector $\mathbf{n}_t^\intercal = [n_{t-1} \dots n_{t-p}]$, the t-th sample can be written as $n_t = \mathbf{n}_t^\intercal a + w_t$, where $w_t \sim N(0, \sigma_v^2)$ is the t-th sample of the driving noise process vector w. The least-squares estimator of the regression coefficient vector a is then given by

$$a_{LS} = \min_a \|n - N a_{LS}\|_2^2 \text{ where } N = \begin{bmatrix} \mathbf{n}_{p+1}^\intercal \\ \vdots \\ \mathbf{n}_m^\intercal \end{bmatrix}. \tag{2}$$

By means of the bias compensation principle [10], the least-squares estimate can be decomposed in the unbiased estimate term a and the bias term as

$$a = a_{LS} + \sigma_w^2 R^{-1} a, \tag{3}$$

where $R = \sum_{t=1}^{m} \mathbf{n}_t \mathbf{n}_t^\top$ and $\sigma_w^2 = \text{var}(\Phi x)$. As we can see from (3), there are two unknowns: a and σ_w^2. Therefore we have developed an iterative algorithm based on alternating the estimation of the unknowns. The distributed algorithm we introduce in Section III-A unbiases the least-squares estimate relying on the technique developed in [10], where the authors obtain the unbiased estimate $\hat{a}_{ILS}$ according to Algorithm 1.

Algorithm 1 Iterative AR(p) LS estimate unbiasing

$k^2 = \frac{\sigma_y^2}{\sigma_w^2}$
$\hat{J}(a_{LS}) = \sigma_w^2 (k^2 + 1 + a_{LS}^\intercal a)$
while not stopIter **do**
  $\hat{\sigma}_w^2(i) \leftarrow \frac{\hat{J}(\hat{a}_{LS})}{k^2 + 1 + \hat{a}_{LS}^\intercal \hat{a}_{ILS}(i-1)}$
  $\hat{a}_{ILS}(i) \leftarrow \hat{a}_{LS} + \hat{\sigma}_w^2(i) R^{-1} \hat{a}_{ILS}(i-1)$
  $\hat{\sigma}_n^2(i) \leftarrow k^2 \hat{\sigma}_w^2(i)$
end while

III. Distributed covariance estimation algorithm

A. Proposed method

As introduced in section II-A, the estimation of the covariance matrix can be reduced to the estimation of the common parameter vector a of the AR(p) process realizations among the nodes. The idea of using a parametric representation for the covariance matrix of the process allow us to distribute the estimation task keeping the communication cost low. Namely, at each iteration the nodes only need to exchange with their neighbors a small parameter vector of length $p \ll m$. To do so, we split the problem into sub-problems to be iteratively solved at each node, in such a way that at the end of the iterations each node v has a consistent estimate of the covariance matrix $\hat{C}$ across the whole network.

Let us start by defining a global functional in which we have a term related to the node-dependent least-squares estimation of the AR parameters $\hat{a}_{LS}^{(v)}$ and a second term, derived from (3), for bias removal which contains the vector variable a that is common to all nodes. In fact, since the noise component affecting the nodes corresponds to a different realization of the same noise process, the consensus is on the parameters $a^{(v)} \forall v \in V$ that are the local *unbiased* coefficients vector estimates. Since the local least-squares estimates at each node are biased according to the compressed measurements the node has acquired, this parameter is updated and kept local. Hence, the functional $\mathcal{F}$ is defined as:

$$\begin{aligned} &\mathcal{F}(a, a_{LS}^{(1)}, \dots, a_{LS}^{(v)}, \dots, a_{LS}^{(|V|)}) = \min \sum_{v \in V} f^{(v)}(a, a_{LS}^{(v)}) = \quad (4) \\ &= \min \sum_{v \in V} \underbrace{\|y_+^{(v)} - [\mathbf{0} \ Y^{(v)}] \bar{a}^{(v)}\|^2}_{\text{least squares term}} \\ &+ \lambda \underbrace{\|[(\mathbb{I} \ - \sigma_w^{2\,(v)} R^{-1\,(v)}) \ - \mathbb{I}] \ \bar{a}^{(v)}\|^2}_{\text{unbiasing term}}, \end{aligned}$$

given $\bar{a}^{(v)} = [a_{LS}^{(v)} \ a]^\intercal$, $\mathbf{y}_+^{(v)} = [y_{p+1}^{(v)} \ \dots \ y_m^{(v)}]$ and $Y^{(v)} = [\mathbf{y}_{p+1 \to 1}^{(v)\intercal} \ \dots \ \mathbf{y}_{m \to m-p+1}^{(v)\intercal}]^\intercal$, where $\mathbf{y}_{a \to b} \triangleq [\mathbf{y}_a \ \mathbf{y}_{a-1} \ \dots \ \mathbf{y}_{b+1} \ \mathbf{y}_b]$. In particular, the first term of the functional, which only updates the least-squares component $a_{LS}^{(v)}$ of the vector $\bar{a}^{(v)}$, is equivalent to (2) as its solution corresponds to the AR(p) least-square estimate. The second term instead links the local least-squares estimates with the unbiased estimate a (that is common to all nodes) thought $\sigma_w^{2\,(v)}$ and $R^{-1\,(v)}$. Hence, this latter term is defined as the least-squares solution of the bias equation (3). Moreover, a regularization parameter λ is used to weight the second term.

To distribute the functional $\mathcal{F}$ we use the subgradient consensus method proposed in [11], where each node v alternates between a gradient descent step towards the minimum of the function $f^{(v)}$ and a consensus step. Since, as already pointed out, by nature of the noise process all the nodes share the same unbiased coefficients, the consensus step is performed on the coefficients $a^{(v)}$. Therefore, in order to perform the consensus, each node v only needs to exchange a vector made of p parameters with its N_v neighbors corresponding to the coefficients vector $a^{(v)}$. The resulting step to be performed at each node at each iteration is then:

$$a^{(v)}(i+1) = \sum_{j=1}^{|V|} [W]^h_{vj}(a^{(v)}(i) - \tau g^{(v)}(a^{(v)}(i))), \quad (5)$$

where $g^{(v)}(a^{(v)}) = \partial f^{(v)}(a^{(v)}(i))$, τ is the gradient descent step size and the notation $[W]^h$ indicates that h consensus iterations are performed according to the adjacency matrix W. However, computing the step in (5) requires the knowledge of ${\sigma_w^2}^{(v)}$ due to its dependency on $g^{(v)}$. Since this variable is assumed to be unknown, it is iteratively estimated according to step 5 of the unbiasing algorithm described in Algorithm 1. The distributed algorithm procedure is summarized in Algorithm 2.

Once the algorithm has reached the convergence, namely $\|\hat{a}^{(v)}(i) - \hat{a}^{(v)}(i-1)\|_2 < \alpha$ given the arbitrary parameter $\alpha \ll 1$, each node can hence locally build the estimate of the covariance matrix $\hat{C}^{(v)}$.

The parameter h specifies the number of the consensus iterations of the algorithm and it is a trade-off parameter between consistency of the estimates and communication cost across the network. More formally, the communication cost (expressed in exchanged data samples) can be written as

$$D = h(|V||N_v|Rp), \quad (6)$$

where $|N_v|$ is constant and R is the number of iterations of the algorithm. As we experimentally show in Sec. IV, $h = 2$ is a good compromise between estimation quality and communication cost.

Algorithm 2 Distributed parametric covariance matrix estimation

Intialize:
- $\bar{a}^{(v)} \leftarrow \mathbf{0}$
- ${\sigma_w}^{2(v)} \leftarrow var(y^{(v)})$
- ${k}^{2(v)} = \frac{\sigma_v^2}{{\sigma_w^2}^{(v)}}$

while not stopIter **do**
- For each node $v \in V$
- $\bar{a}^{(v)}(i+1) \leftarrow \bar{a}(i) - \tau g(\bar{a}^{(v)}(i), {\sigma_w^2}^{(v)}(i))$
- $\hat{\sigma_w}^{2(v)}(i+1) \leftarrow \frac{\hat{J}(\bar{a}^{(v)}(i))}{{k^{(v)}}^2 + 1 + a_{LS}^{\intercal}{}^{(v)}(i) a(i)^{(v)}}$
- $a^{(v)}(i+1) = \sum_{v \in N_v} W a^{(v)}(i)$
- $\bar{a}^{(v)}(i+1) = [a^{(v)}(i+1)\ a_{LS}^{(v)}(i+1)]^{\intercal}$

end while

$\hat{A}^{(v)} \leftarrow$ tril(toeplitz($[1\ a^{(v)}\ 0\ \dots\ 0]$)), $\hat{C}^{(v)} \leftarrow \hat{A}^{(v)-1}\hat{A}^{(v)-\intercal}$

B. Convergence

Although in this paper we are not presenting a proof of convergence of the proposed algorithm, we discuss its convergence properties by separately analyzing the two directions in which the algorithm moves that are: the subgradient consensus step and the variance estimation. At first, let us assume that $\hat{\sigma}_w^2$ is known at each node. Then the following theorem can be stated:

Theorem 1. *Having the sequence* $\{\bar{a}(i)^{(1)}, \dots, \bar{a}(i)^{(|V|)}\}_{i=0}^{\infty}$ *generated by algorithm 2, with* $h \geq (\log(\beta) - \log(4NM(\beta + \alpha C)))/\log(\gamma)$ *and* $f^* > -\infty$ *we have that:*

$$\lim_{i \to \infty} \inf f(a(i)^{(v)}) \leq f^* + \alpha N C^2/2 + 3nc\beta\ , \forall v \in V$$

Proof: This theorem comes from Th. 1 in [11], we hence need to prove that in our case all the related assumptions are satisfied. Assumption 1 is satisfied since we have that $\|g^{(v)}(a)\| \leq\ C\ =\ (2\|Y^{\intercal}Y\| + \|\lambda S^{\intercal}S\|)\|a\| + \|2Y^{\intercal}y^{+}\|$ with $S = [(\mathbb{I} - {\sigma_w^2}^{(v)} R^{-1(v)})\ - \mathbb{I}]$, then if $\|a\| \leq \infty$ we have that $C \leq \infty$. Assumption 2 is satisfied for the topology we are mainly considering throughout this paper: the ring topology. Lastly, Assumption 3 is satisfied since the problem we are considering is unconstrained. ■

In the proposed algorithm we are assuming that σ_w^2 is not known at each node and hence must be estimated. The unbiasing estimator given by the Algorithm 1 is proven to converge in [10] to the unbiased estimate and hence σ_w^2 converges to ${\sigma_w^2}^*$. Differently from this estimator, the proposed algorithm does not have the knowledge of the true least squares estimate $a_{LS}^{(v)}$. It has instead, at each step, a local estimate of $a_{LS}^{(v)}$ given by the subgradient step which is going towards the optimum $a_{LS}^{(v)*}$. As we numerically show in the next section, the proposed algorithm shows empirical convergence. Intuitively we can hence say that, even though the convergence is not proved when σ_w^2 is not known in advance, the alternated estimation of least squares term and the noise variance, tends to go towards the optimum of the functional defined in (4).

IV. Experimental results

A. Estimation of AR coefficients

For the purpose of numerical evaluation of the proposed algorithm, at first we consider the estimation error and the consensus reached on the parameters of the AR process in a given network after a suitable number of iterations. The considered network is arranged in a ring topology and is given by the graph $\mathcal{G}$ with $|V| = 10$, and $|N_v| = 2$. Each node acquires a signal $y^{(v)} = \Phi^{(v)}x^{(v)}$ where $\Phi^{(v)}$ are gaussian random sensing matrices whose entries are drawn from $N \sim (0, \frac{1}{m})$ and $x^{(v)}$ are sinusoidal signals (sparse in frequency domain) for which different signal to noise ratios are taken into account. The length of $x^{(v)}$ is $n = 800$, and that of $y^{(v)}$ is $m = 200$. For the experiments, along with AR noise, we also considered pink noise. This kind of noise, which is a non-white gaussian noise, exhibits a decreasing spectrum and is present in many physical, biological and economical systems. Hence, for this experiment we considered both synthetic AR noise of order $p = 3$ and pink noise which is approximated with an AR process. Since from our experiments the pink noise is well approximated with an AR process of third order, we choose $p = 3$. When considering pink noise, since the true AR coefficients are not available, the relative error we consider is given by $\frac{\|a^* - \hat{a}^{(v)}\|}{\|a^*\|}$ where a^* is the estimate obtained in a centralized fashion. The results in Fig. 1 show that a relatively small number of exchanged data samples (160000 samples) are sufficient to reach very

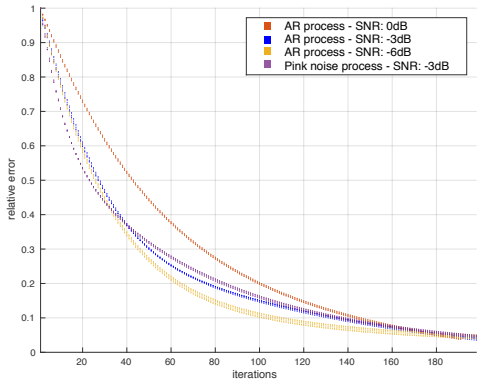

Fig. 1. Relative error $\frac{\|a-\hat{a}^{(v)}\|_2}{\|a\|_2}$ $\forall v \in V$. The vertical bars at each iteration corresponds to the range of errors of all the nodes in the network at the given iteration.

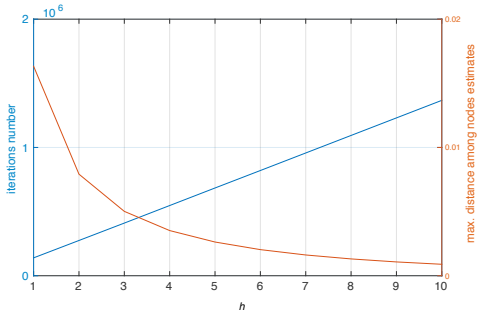

Fig. 2. Trade-off curve for different values of consensus parameter h.

low errors on the estimates $\hat{a}^{(v)}$. Along with a low estimation error, we also show that the local estimates at each node reach consensus. More in detail, for the AR process with SNR = -3dB, the maximum distance among the nodes reduces down to $9e{-}3$ as the number of iterations are bigger than 80.

Since the number of consensus iterations h is a trade-off between closer estimates and higher communication cost, we experimentally evaluate the role of the parameter h. We can see from Fig. 2 that, given a fixed error on the AR coefficients estimates to be reached, increasing this consensus parameter leads to closer estimates but at a higher communication cost. For the experiments we show in the following sections, we chose $h = 2$ as it is a good trade-off for keeping the communication cost low while maintaining accurate estimates across the network.

B. Estimation of covariance matrix

Starting from the good results obtained for the AR parameters estimates, in this section we present some results on the estimated covariance matrix in order to assess the performance of the proposed technique. The distance metric we use to compare the estimate is the Forstner distance [12], a widely used method for evaluating the distance of two positive semi-definite covariance matrices. It is defined as $d(A,B) = \text{tr}(\ln^2(\sqrt{A^{-1}}B\sqrt{A^{-1}}))$, where in our setting A is C, the sample covariance matrix of the process (averaged over 5000 realizations), and B is the local estimate of the covariance $\hat{C}^{(v)}$. As shown in Fig. 3, the distance between the matrices C and $\hat{C}^{(v)}$ decreases as the proposed algorithm iterates, moreover the local distances (*i.e.,* computed using the

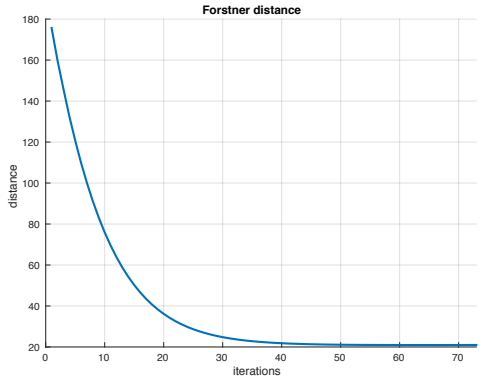

Fig. 3. Forstner distance between $\hat{C}^v$ and C computed for each estimate $\hat{C}^{(v)} \forall v \in V$. This figure only shows the Forstner distance up to 70 iterations since in the remaining iterations the distance does not change having reached convergence.

nodes local estimates) tend to converge. Then, we compare our technique with the one proposed in [6] for the estimation of the eigenvector associated with the largest eigenvalue of the covariance matrix in a distributed setting. More in detail, among the techniques the authors propose, we use the technique *adapt then combine* that, according to their experiments, leads to better results. The setting we use for the comparison is the same as described above, and the number of iterations for the compared algorithm is fixed to 200. The metric used in this experiment is the angle between the principal axes of the true and estimated covariance matrices.

In Table I we summarize the results of the comparison. As can be seen, the proposed technique leads to lower angles between the principal axes while keeping the communication cost extremely low. Moreover, the proposed technique estimates the whole covariance matrix of the noise process instead of just the principal eigenvector and hence it allows a wider range of applications.

TABLE I.

	exchanged data (samples)	angle (rad)
Proposed algorihtm	160000	0.24
Ghadban *et al.* [6]	2000000	0.45

C. Compressive detection in distributed setting

The proposed algorithm allows us to estimate in a distributed fashion the covariance matrix of a non-white noise process keeping the communication cost very low. Hence, we further validate the proposed technique by including it in a compressive signal processing task. More in detail we consider the detection of a compressed signal given its CS measurements corrupted by additive colored noise as introduced in [1]. Similarly to Sec. IV-A, let us assume to have a wireless sensor network represented by the graph $\mathcal{G}$ arranged in a ring topology with $|V| = 10$. Then, we assume that each node v acquiring a CS signal is then connected to its neighbors given $|N_v| = 2$. We are now interested in the signal detection problem at node $v^\star$, more in detail we want to distinguish between the hypotheses:

$$\begin{cases} \mathcal{H}_0 & : y^{(v^\star)} = n^{(v^\star)} \\ \mathcal{H}_1 & : y^{(v^\star)} = \Phi^{(v^\star)} x^{(v^\star)} + n^{(v^\star)}, \end{cases}$$

where $\Phi^{(v^\star)}$, $x^{(v^\star)}$ are known, and $n^{(v^\star)} \sim \mathcal{N}(0, C)$ is the colored noise with unknown covariance matrix C.

In order to improve the performance of the detector, the knowledge of the noise statistics is required. In particular we need the covariance matrix of the noise process. Hence, assuming that the node $v^\star$ needs an estimate of the covariance matrix of the noise process to accurately solve the detection problem, it runs together with its neighbors the distributed covariance estimation algorithm described in Algorithm 2. From now on, assuming that the node $v^\star$ has obtained the estimate $\hat{C}^\star$, we will drop the superscript "$\star$" to improve readability.

Relying on standard detection theory [2] we can show that, in our setting, the *Neyman-Pearson* (NP) optimal detector, namely the likelihood ratio test, can be written as $t = y^\intercal \hat{C}^\intercal \Phi x$. Then, denoting the probability of false alarm as P_F and that of detection as P_D, it can be shown that the P_D in function $P_F = \alpha$ is given by:

$$P_D = Q\left(\frac{\sqrt{V_0}Q^{-1}(\alpha) + \mu^\intercal \hat{C}^{-1}\Phi x - E_1}{\sqrt{V_1}}\right), \qquad (7)$$

where

$$\mu = \mathbb{E}[n], V_0 = \sigma^2 x^\intercal \Phi^\intercal (\hat{C}\hat{C}^\intercal)^{-1}\Phi x - (\mu^\intercal \hat{C}^{-1}\Phi x)^2,$$
$$E_1 = x^\intercal \Phi^\intercal \hat{C}^{-\intercal}(\Phi x + \mu^\intercal),$$
$$V_1 = x^\intercal \Phi^\intercal \hat{C}^{-\intercal}\Phi x - x^\intercal \Phi^\intercal \hat{C}^{-\intercal}\mu^\intercal \mu C^{-\intercal}\Phi x.$$

We hence show the receiver operating characteristics (ROC) curves for the detection problem we introduced. We compare the covariance matrix estimate $\hat{C}$ with those obtained by assuming the noise to be white and hence having no knowledge of noise statistics averaging the results over 50 different runs of the algorithm. Using the estimated covariance $\hat{C}$, we show both theoretical results in (7) and experimental ROC curves obtained by running 500 Monte Carlo (MC) tests. Then, we compare them with those obtained by a standard detector assuming the noise to be white, hence being unaware of the noise statistics.

The results we present are for two different compression ratio values, namely $\frac{m}{n} = \{0.11, 0.22\}$. Moreover, the SNR we consider is extremely low SNR = -19dB, this choice is made to show the detection performances for extremely noisy signals, using the proposed setup. It is worth noting that when higher SNRs are considered, the ROC curve reaches the optimality. As we can see from Fig. 4, the compressive signal detection using the estimated covariance matrix according to the proposed method, outperforms a compressive signal detection algorithm unaware of the noise covariance matrix. Moreover it can be seen that even though the number of measurements m and the SNR value are very low, the detection task is efficiently solved. Lastly, we can see that the experimental results confirm the theoretical performance.

V. Conclusions

In this paper we proposed a new technique for distributed estimation of the covariance matrix of a colored noise process corrupting CS measurements. The main advantage of the proposed method relies in its parametric structure, removing the dependence of the communication cost from the size of the data samples at each node. Moreover, when compared to other techniques, the proposed algorithm is able to achieve both lower errors and data samples to be transmitted in the network.

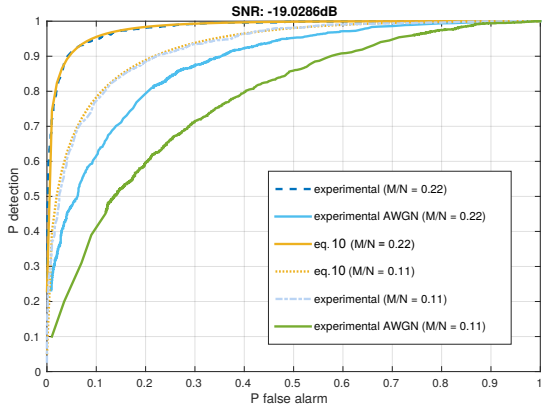

Fig. 4. ROC performance comparison of noise unaware signal detection (AWGN) and using the proposed covariance matrix estimate.

In fact, our results showed that the number of parameters to be transmitted in order to get an accurate estimate of the noise covariance matrix is much smaller than the number of samples acquired by each node. Lastly, since the numerical results we presented showed that we are able to accurately estimate the covariance matrix of the noise process (with low communication cost) we were also able to show that our estimate of the covariance matrix performs well when used in compressive signal processing tasks.

References

[1] Mark A Davenport, Petros T Boufounos, Michael B Wakin, and Richard G Baraniuk, "Signal processing with compressive measurements," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 4, no. 2, pp. 445–460, 2010.

[2] Steven M Kay, "Fundamentals of statistical signal processing, vol. ii: Detection theory," *Signal Processing. Upper Saddle River, NJ: Prentice Hall*, 1998.

[3] Sergio Valcarcel Macua, Pavle Belanovic, and Santiago Zazo, "Consensus-based distributed principal component analysis in wireless sensor networks," in *Signal Processing Advances in Wireless Communications (SPAWC), 2010 IEEE Eleventh International Workshop on*. IEEE, 2010, pp. 1–5.

[4] Alexander Bertrand and Marc Moonen, "Low-complexity distributed total least squares estimation in ad hoc sensor networks," *Signal Processing, IEEE Transactions on*, vol. 60, no. 8, pp. 4321–4333, 2012.

[5] Federico Penna and Slawomir Stanczak, "Decentralized Eigenvalue Algorithms for Distributed Signal Detection in Wireless Networks," *IEEE Transactions on Signal Processing*, , no. c, pp. 1–1, 2014.

[6] Nisrine Ghadban, Paul Honeine, Clovis Francis, Farah Mourad-Chehade, and Joumana Farah, "Strategies for principal component analysis in wireless sensor networks," in *Proc. eighth IEEE Sensor Array and Multichannel Signal Processing Workshop*, 2014.

[7] David L Donoho, "Compressed sensing," *Information Theory, IEEE Transactions on*, vol. 52, no. 4, pp. 1289–1306, 2006.

[8] Giulio Coluccia, Aline Roumy, and Enrico Magli, "Operational rate–distortion performance of single-source and distributed compressed sensing," 2013.

[9] James Douglas Hamilton, *Time series analysis*, vol. 2, Princeton university press Princeton, 1994.

[10] WX Zheng, "Study of a least-squares-based algorithm for autoregressive signals subject to white noise," *Mathematical Problems in Engineering*, vol. 3, no. October 2002, pp. 93–101, 2003.

[11] Björn Johansson, Tamás Keviczky, Mikael Johansson, and Karl Henrik Johansson, "Subgradient methods and consensus algorithms for solving convex optimization problems," in *Decision and Control, 2008. CDC 2008. 47th IEEE Conference on*. IEEE, 2008, pp. 4185–4190.

[12] Wolfgang Förstner and Boudewijn Moonen, "A Metric for Covariance Matrices," *Quo vadis geodesia*, vol. 66, pp. 113–128, 1999.