[Article] Compressive estimation and imaging based on autoregressive models

Compressive estimation and imaging based on autoregressive models

Matteo Testa and Enrico Magli

***Abstract*—Compressed Sensing (CS) is a fast and efficient way to obtain compact signal representations. Oftentimes, one wishes to extract some information from the available compressed signal. Since CS signal recovery is typically expensive from a computational point of view, it is inconvenient to first recover the signal and then extract the information. A much more effective approach consists in estimating the information directly from the signal's linear measurements. In this paper we propose a novel framework for compressive estimation of autoregressive (AR) process parameters based on ad-hoc sensing matrix construction. More in detail, we introduce a compressive least square estimator for AR(p) parameters and a specific AR(1) compressive Bayesian estimator. We exploit the proposed techniques to address two important practical problems. The first is compressive covariance estimation for Toeplitz structured covariance matrices where we tackle the problem with a novel parametric approach based on the estimated AR parameters. The second is a block-based compressive imaging system, where we introduce an algorithm that adaptively calculates the number of measurements to be acquired for each block from a set of initial measurements based on its degree of compressibility. We show that the proposed techniques outperform state-of-the-art methods for these two problems.**

***Index Terms*—Compressed Sensing, Compressive Estimation, Adaptive Imaging, Compressive Covariance Estimation, Autoregressive Process.**

I. Introduction

Compressed Sensing (CS) [1] [2] is a well-established paradigm in which signal acquisition and compression become a single operation. Being able to sample a signal well below the Nyquist rate made CS a popular approach over the last few years. The key concept which CS relies on, is that the acquired signal must be sparse in some domain. This assumption allows to provide theoretical guarantees under which the signal can be exactly recovered with overwhelming probability. However, due to its nature CS shifts the computational cost at the recovery stage. In many cases, one wishes to infer some information about the signal that has been acquired, e.g. estimate one or more parameters that describe the signal in order to perform some detection or classification task. In CS, because of the complexity of the signal recovery process, it is inconvenient to first recover the signal and then estimate its parameters. Rather, it is much more desirable to perform estimation directly on the compressed measurements (see e.g. [3]).

M.Testa and E. Magli are with the Department of Electronic and Telecommunication Engineering, Politecnico di Torino, Turin, 10129 ITALY e-mail: {name.surname}@polito.it. 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.

Besides analyzing the signal itself, knowledge of such signal parameters may also be useful during the reconstruction stage. For example, in many CS applications, no knowledge about the nature of the compressed signal is available, except for the assumption that it is sparse in some domain. However, natural signals are typically not exactly sparse, but rather approximately sparse or "compressible". In many cases, such compressibility implies that the signal spectrum is decreasing [4]. This kind of information on the signal structure has indeed been used to improve the CS reconstruction [5].

If we assume that the signal to be sensed can be approximated by a certain class of parametric signals, then estimating its parameters has important implications in terms of inferring signal characteristics, and using these to improve signal recovery. In [6] a way to estimate the parameters of a compressively sensed autoregressive process (AR) was introduced. The AR model has been widely used to represent many signals of interest in practical applications, including audio and images [7] [8]. More in general, signals having a shaped spectrum can be represented in a parametric AR fashion.

In this paper we generalize the work in [6] by proposing a compressive estimator for AR(p) coefficients. Moreover, in order to improve the robustness of the estimator when considering highly compressed signals, we specialize the estimation in the AR(1) setting through the use of Bayesian techniques. Remarkably, we show experimentally that the proposed algorithms are able to estimate the AR model parameters even when the number of available linear measurements is not sufficient to reconstruct the original signal. This opens the door to many interesting applications. In particular, we focus on two key applications that employ the proposed AR estimation framework to solve important practical estimation problems. The first one is a parametric compressive covariance estimation algorithm. Interestingly, since it is a parametric technique, the coefficients defining the covariance matrix have a very compact representation, which can be conveniently transmitted to a receiver, allowing the construction of the full covariance matrix to be moved to a different stage. Secondly, we introduce an algorithm that adaptively estimates the degree of compressibility of an image, and hence the number of random projections needed to properly compress it, which does not require to have access to the full image at any stage of the process.

Compressive covariance estimation: The knowledge of the covariance matrix of a signal estimated in the compressed domain has a lot of applications, including compressive power-spectrum estimation [9]–[12], wideband spectrum sensing

[13], incoherent imaging [14] and direction-of-arrival estimation [15] [16]. All these techniques take advantage from compressive covariance estimation since the number of sensors needed for the signal acquisition can be dramatically reduced. In order to estimate the covariance matrix of a process in the compressed domain, the main approaches used in literature are: maximum likelihood estimation (MLE), least squares [17] and convex optimization [18]. However, while the second approach is not able to guarantee the positive semidefiniteness of the covariance matrix, the former usually requires to have samples which show a good statistical significance. This problem has been addressed in [19] where, within the wideband spectrum sensing framework, the authors make use of spectral prior information to reduce the number or required samples. It is also worth noting that, in order to make the compressive covariance estimation possible, the first two approaches require the sensing process to be able to preserve the second order statistics of the signals which can the be used to recover the covariance matrix. In order to provide guarantees for the preservation of the correlation matrix structure, in [20] the authors propose optimal sensing matrix design through the use of the sparse rulers.

Differently from the aforementioned methods, the proposed approach focuses instead on the estimation of the parameters of a covariance matrix defined by the structure of the AR process. This approach is computationally light, and it also ensure the positive semidefiniteness of the covariance matrix.

Adaptive compressive imaging: Compressive image sensing and representation have been extensively addressed in literature. Early works exploited the signal sparsity in standard sparsifying bases such as DCT and wavelets [21]. Later methods tried to achieve improved recovery performance using the Total Variation by assuming that images are gradient-sparse [22] [23]. Other works focused on the arrangement of the signal in order to get higher PSNR for the recovered image, or reduce reconstruction complexity, e.g. block CS [24] and BCS-SPL [25]. However, to the best of our knowledge in literature all works assumed the sampling ratio to be fixed at the beginning and to be the same for every block of the image. The question then arises: is it possible to apply a lower sampling ratio to the blocks which contain less information, $i.e.$, are spatially smooth? Even though this question has been extensively addresses in standard image coding [26] [27] [28], the same can not be said when considering CS schemes. The only work which addressed a similar problem [29], requires to estimate the complexity *given the uncompressed signal* in order to acquire a proper number of random measurements; in many cases, e.g. compressive imaging, this is not a feasible approach. Differently, we consider a fully compressed setting in which we only have access to the random projections of the signal of interest.

The remainder of paper is organized as follows: In Section 1 we introduce the compressed estimator for AR processes by first reviewing AR processes and the uncompressed LS estimator. We also validate the proposed sensing matrix. In Section 2 we propose a compressed domain estimator for Toeplitz structured covariance matrices. Then, in Section 3 a novel compressive AR(1) estimation algorithm which achieves improved performance for highly compressed signals is presented. In Section 4 we propose a novel adaptive compressive imaging scheme which relies on the compressive AR(1) estimator. The results are discussed in Section 5 and the concluding remarks are presented in Section 6.

II. Compressive AR(p) estimation

The literature related to the uncompressed estimation of the AR(p) parameters is vast and has been extensively studied. Some of the most well-known and widely used tecniques are the Yule-Walker, Prony and Burg algorithms [30] due to their efficiency. However, among all the different classes of the AR coefficients estimators available in literature, we focus on the LS estimator. This choice, along with an ad-hoc sensing matrix design, allow us to explicitly estimate the regression coefficients in the compressed domain.

A. Uncompressed LS estimator review

Let us start by introducing some notation and reviewing the LS estimator in the uncompressed domain.

An AR process of order p is a parametric model able to describe the time-varying nature of a process in which the output values linearly depend on their previous values. More formally

$$x_t = \sum_{i=1}^{p} x_{t-i} a_i \; + \epsilon_t, \tag{1}$$

where $\boldsymbol{\epsilon}$ is called driving process and $\mathbf{a}$ is the vector of the regression coefficients. In other words, it can be seen as a filtering operation over a process $\boldsymbol{\epsilon}$ with an all-pole filter with coefficients given by $\mathbf{a} = [a_1 \; \dots \; a_i \; \dots \; a_p]^\top$. Given an AR(p) process $\mathbf{x} \in \mathbb{R}^N$, we define $\mathbf{x}^+$ as a subset of $\mathbf{x}$ composed by its samples with index from $(p+1)$ to N. Let us also define the matrix $\mathbf{X} \in \mathbb{R}^{(N-p)\times p}$ constructed in the following way

$$\mathbf{X} = \begin{bmatrix} x_p & x_{p-1} & \dots & x_1 \\ x_{p+1} & x_p & \dots & x_2 \\ \vdots & \dots & \dots & \vdots \\ x_{N-1} & x_{N-2} & \dots & x_{N-p} \end{bmatrix}. \tag{2}$$

Since from (1) we obtain

$$\mathbf{x}^+ = \mathbf{X}\mathbf{a}, \tag{3}$$

it is straightforward to write the LS estimator of $\mathbf{a}$ as the solution to the following minimization problem

$$\arg\min_{\hat{\mathbf{a}}} \|\mathbf{x}^+ - \mathbf{X}\hat{\mathbf{a}}\|_2^2 \tag{4}$$

or, more concisely, as $\hat{\mathbf{a}} = \mathbf{X}^\dagger \mathbf{x}^+$ where "$\dagger$" denotes the pseudo-inverse.

B. Compressed LS estimator

We start our discussion with the compressive AR(p) estimator introduced in [6] which couples an LS estimator with a novel sensing matrix design. In order to have an analogous LS estimator for the compressed domain we need to employ a sensing matrix able to preserve the structure of the regression.

As we can see from (2), the LS estimator for a process of order p, needs $p+1$ shifted versions of the input signal. Hence, the idea is to build a sensing matrix from which, given the output measurements, it is possible to extract the compressed $p+1$ shifted versions of $\mathbf{x}$. This means that the sensing matrix should be made of $p+1$ sub-blocks $\mathbf{\Phi}'$ where each of them senses a shifted version of the unknown signal $\mathbf{x}$.

Then, if we use (3), multiplying both sides by the sensing block $\mathbf{\Phi}'$ we obtain

$$\mathbf{y}^+ = \mathbf{Ya}, \tag{5}$$

where $\mathbf{y}^+ = \mathbf{\Phi}'\mathbf{x}^+$ and $\mathbf{Y} = \mathbf{\Phi}'\mathbf{X}$. Hence, if the sensing matrix is made up of shifted sensing blocks it is possible to extract the quantities $\mathbf{y}^+$ and $\mathbf{Y}$ directly from the measurement vector $\mathbf{y}$.

More formally, let us assume that the main building block $\mathbf{\Phi}' \in \mathbb{R}^{\mu\times(N-p)}$ with $\mu = M/\ (p+1)$, has entries distributed according to $\phi'_{ij} \sim \mathcal{N}(0, \frac{1}{M})$. Then, the proposed sensing matrix $\mathbf{\Phi} \in \mathbb{R}^{M\times N}$ is made of $p+1$ circulant blocks of $\mathbf{\Phi}'$ as depicted in Fig. 1.

$$\begin{bmatrix} \phi'_{1,1} & \phi'_{1,2} & \phi'_{1,3} & \phi'_{1,4} & \phi'_{1,5} & \phi'_{1,6} & 0 & 0 \\ \phi'_{2,1} & \phi'_{2,2} & \phi'_{2,3} & \phi'_{2,4} & \phi'_{2,5} & \phi'_{2,6} & 0 & 0 \\ 0 & \phi'_{1,1} & \phi'_{1,2} & \phi'_{1,3} & \phi'_{1,4} & \phi'_{1,5} & \phi'_{1,6} & 0 \\ 0 & \phi'_{2,1} & \phi'_{2,2} & \phi'_{2,3} & \phi'_{2,4} & \phi'_{2,5} & \phi'_{2,6} & 0 \\ 0 & 0 & \phi'_{1,1} & \phi'_{1,2} & \phi'_{1,3} & \phi'_{1,4} & \phi'_{1,5} & \phi'_{1,6} \\ 0 & 0 & \phi'_{2,1} & \phi'_{2,2} & \phi'_{2,3} & \phi'_{2,4} & \phi'_{2,5} & \phi'_{2,6} \end{bmatrix}$$

Fig. 1: Circulant blocks structure of $\mathbf{\Phi}$.

In order to obtain the measurements of shifted versions of the original signal (which are needed to obtain $\mathbf{y}^+$ and $\mathbf{Y}$), we exploit the structure of the sensing matrix in Fig. 1. As can be seen, each sub-block acquires a shifted version of the input signal through the sub-sensing matrix $\mathbf{\Phi}'$. Hence, by denoting as $\mathbf{y}_{i\to j}$ a sub-set of $\mathbf{y}$ containing its samples from the i-th until the j-th one, we can write $\mathbf{y}_{1+\mu(k-1)\to k\mu} = \mathbf{\Phi}'\mathbf{x}_{k\to(N-p+k-1)}$. This means that the vector $\mathbf{y}$ is made of $p+1$ blocks of length μ which are the measurements corresponding to different shifts of $\mathbf{x}$.

In particular, using (5) we define the compressed LS estimator for AR(p) coefficients, as:

$$\arg\min_{\hat{\mathbf{a}}} \|\mathbf{y}^+ - \mathbf{Y}\hat{\mathbf{a}}\|, \tag{6}$$

where the chosen M must be an integer multiple of $p+1$.

It is worth noting that the proposed estimator, working directly in the reduced space of the measurements domain, is computationally less demanding with respect to the corresponding LS estimator (4) in the uncompressed domain. The complexity is due to the computation of the pseudo-inverse of $\mathbf{X}$ ($\mathbf{Y}$) in the uncompressed (compressed) domain. When considering the uncompressed case, it strictly depends on the value of p and the length of signal N according to $O(p^2(N-p))$, where the most influential term is N because the order of the process is typically small. On the other hand, in the compressed domain the required computational power for the proposed estimator drastically reduces to $O(p^2\mu)$ with $\mu \ll (N-p)$.

Sensing matrix validation: We now numerically validate the proposed sensing matrix comparing its recovery performance to that of the most used ones in literature for which theoretical results on the recovery performance exists [31] [32]. In this experiment we also included in the comparison another kind of sensing matrix induced by the Generalized Nested Sampling (GNS) technique introduced in [33]. The performance of such sensing matrix is here analyzed since it allows to compressively estimate a Toeplitz structured covariance matrix as will be shown in Sec. II-C. In particular, we fix a sparsity level $s = 100$ and randomly pick the support of the non-zero components. Then, we compare the recovery performance of different sensing matrices by running 1000 different Monte Carlo runs over different M values by compressing sparse signals and then recovering them using LASSO. The recovery error is defined as $\|x - \hat{x}\|_2/\|x\|_2$ where $\hat{x}$ is the recovered signal. For what concerns the GNS scheme, since the number of measurements M is determined by the given N, we employed different values of N in order to obtain different compression ratios. It is also worth noting that since a fixed sparsity level would be disadvantageous when considering smaller values of N, for this specific case we fixed the sparsity to be equal to the 10% of sample size.

The results (Fig. 2) show that recovery error of the proposed matrix is slightly higher than the Gaussian sensing matrix and lower than that of the Bernoulli one. The recovery performance of the proposed matrix is hence comparable to that of a circulant matrix, which is a very popular choice, and has a negligible performance loss with respect to a Gaussian matrix. Similar results can be found for other values of M, N, s and p and are omitted for brevity. In the same figure we also show that the GNS scheme, while is a good sampling technique for covariance estimation, is not able to reach satisfactory recovery performance when used inside the CS framework.

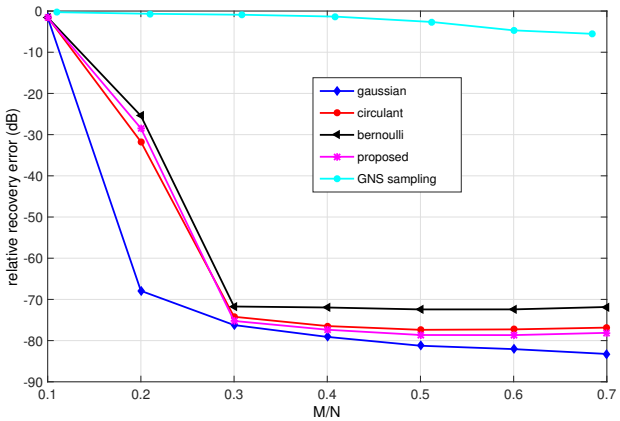


Fig. 2: Comparison of recovery ability of different sensing matrices using a signal of length $N = 1000$, sparsity $s = 100$ and $p = 9$.

C. Compressive Toeplitz covariance estimation

So far we have presented an efficient way to estimate the coefficients of a compressed AR(p) process. In the following

we discuss an important application which arises from the aforementioned technique: the compressive estimation of a structured Toeplitz covariance matrix. This kind of matrices play an important role in different fields such as integral equations, spline functions, mathematics, statistics, and signal processing.

Due to the nature of AR(p) processes, the associated covariance matrix has a Toeplitz structure. Before derivating the compressed covariance matrix estimator, let us introduce some notation: the symbol $\mathrm{tril}(\bullet)$ denotes the operator which extracts the lower triangular part of a matrix, while we will use $\mathrm{toeplitz}(\mathbf{a})$ to denote a matrix built as a Toeplitz matrix constructed from vector $\mathbf{a}$.

Let us start with a simple model of a set of observations of AR(p) processes:

$$\mathbf{AX} = \mathbf{V}, \tag{7}$$

where $\mathbf{X} \in \mathbb{R}^{N\times O}$ are the column-wise AR processes, $\mathbf{V} \in \mathbb{R}^{N\times O}$ are the column-wise driving noise vectors and $\mathbf{A}$ is the regression matrix (common to all the observation vectors) which only depends on the coefficients vector $\mathbf{a} \in \mathbb{R}^{1\times p}$. In particular, we have $\mathbf{A} = \mathrm{tril}\left(\mathrm{toeplitz}(\mathbf{a}^\star)\right)$, having defined $\mathbf{a}^\star = [1 \ -\mathbf{a}\ 0 \ \ldots \ 0]^\intercal \in \mathbb{R}^{1\times N}$.

We recall that, by definition, the driving noise processes $\mathbf{V}$ are distributed according to $\mathcal{N}\left(0, \mathbb{I}\sigma^2_{\mathbf{v}}\right)$. Therefore, it immediately follows that the mean of the autoregressive processes $\mathbf{X}$ is $\boldsymbol{\mu}_{\mathbf{X}} = \underline{0}$. Hence, we can write the covariance matrix of the AR(p) process as

$$\boldsymbol{\Sigma} = \mathbf{A}^{-1}\sigma^2_{\mathbf{V}}\mathbb{I}\,\mathbf{A}^{-1^\intercal}. \tag{8}$$

This means that, given the coefficients $\mathbf{a}$ of the process and the variance of the driving noise $\sigma^2_{\mathbf{V}}$, the covariance matrix of the process is uniquely defined by these two parameters.

Given the compressed measurements $\mathbf{y} = \boldsymbol{\Phi}\mathbf{X}$, we can use (6) to estimate $\mathbf{a}$, without any prior knowledge of $\sigma^2_{\mathbf{V}}$. In order to estimate $\sigma^2_{\mathbf{V}}$ let us write

$$\begin{aligned}\langle \mathbf{y}^\intercal\mathbf{y}\rangle = \langle \mathbf{X}^\intercal\boldsymbol{\Phi}^\intercal\boldsymbol{\Phi}\mathbf{X}\rangle = \left\langle \mathbf{V}^\intercal\mathbf{A}^{-1^\intercal}\boldsymbol{\Phi}^\intercal\boldsymbol{\Phi}\mathbf{A}^{-1}\mathbf{V}\right\rangle = \\ = \sigma^2_{\mathbf{V}}\mathrm{tr}(\mathbf{A}^{-1^\intercal}\boldsymbol{\Phi}^\intercal\boldsymbol{\Phi}\mathbf{A}^{-1}),\end{aligned} \tag{9}$$

where $\mathrm{tr}(\bullet)$ denotes the trace operator and $\langle\bullet\rangle$ denotes the expectation operator computed with respect to the only random variable involved if not differently specified.
Thus, we can estimate $\sigma^2_{\mathbf{V}}$ as

$$\hat{\sigma^2_{\mathbf{v}}} = \frac{\langle \mathbf{y}^\intercal\mathbf{y}\rangle}{\mathrm{tr}(\mathbf{A}^{-1^\intercal}\boldsymbol{\Phi}^\intercal\boldsymbol{\Phi}\mathbf{A}^{-1})}, \tag{10}$$

using the sample mean estimator to compute the term $\langle \mathbf{y}^\intercal\mathbf{y}\rangle$.

To summarize, the covariance $\boldsymbol{\Sigma}$ of a compressed AR(p) process can be estimated according to (8) exploiting the coefficients vector $\mathbf{a}$ computed with (6) and the driving noise variance $\sigma^2_{\mathbf{v}}$ computed with (10).

Results: In the following we show the performance of the proposed technique for compressive covariance estimation and compare it with three different approaches proposed in literature to address the same task. The algorithm proposed by Eldar et al. in [18] employs convex optimization modeling in order to recover Toeplitz structured covariance matrices. For the experiments, we generated synthetic AR processes of order $p = 4$, size $N = 100$ and considered $O = 200$ observations. Then, we compressed the resulting $\mathbf{X} \in \mathbb{R}^{N\times O}$ with different M/N ratios. The results were then averaged by running 1000 different experiments employing random sensing matrices.

To assess the performance we employ the normalized mean squared error (NMSE) defined as $\frac{\|\boldsymbol{\Sigma}-\hat{\boldsymbol{\Sigma}}\|^2_F}{\|\boldsymbol{\Sigma}\|^2_F}$ where $\boldsymbol{\Sigma}$ and $\hat{\boldsymbol{\Sigma}}$ are the true and the estimated covariance matrices respectively. As we can see from Fig. 3, at all undersampling rates the proposed algorithm shows lower NMSE than [18]. We also analyze the effects of higher orders of the regression and model mismatch. In the same figure we show the NMSE of both techniques when the order of the process is $p = 10$. We considered this value since this is typically the largest order considered for natural signals. The experiment shows that the proposed algorithm is able to achieve lower NMSE compared to [18] and the results are comparable with those obtained with a smaller order.

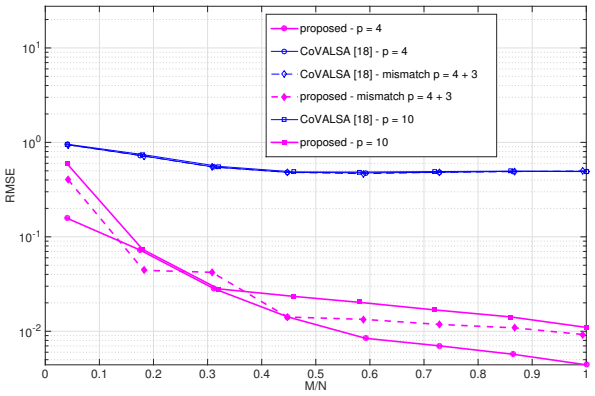


Fig. 3: NMSE on compressive covariance estimation computed for the proposed technique and [18] evaluated at different compression ratios. We consider two different scenarios namely $p = 4, 10$ and also include a comparison with the mismatched case.

Next, we analyze the mismatch scenario. Typically when the order of the process p is unknown, a good practice is to use an order slightly larger than the guessed one. Therefore we show an experiment in which we purposely used an increased order for the proposed AR(p) compressive covariance estimation algorithm. As in the previous experiments, the true order of the process is $p = 4$ while the augmented one is $p+3$. As can be seen in Fig. 3 the proposed methods deals well with model mismatch and achieves results similar to those obtained using the correct value of p.

Since we have shown so far that a Toeplitz structured covariance matrix can be defined by the parameters of an AR(p) process, we also perform a comparison with the method proposed in [34]. In particular we focus on the NMSE related to the AR(p) coefficients. The technique in [34] is a Bayesian algorithm with *Metropolis-Hastings* sampling to estimate the parameters of the regression. In order to evaluate the NMSE of the coefficients of the two methods we performed a comparison using the same setting as described in Sec. IV of [34] and using the results provided in the same paper. The results

TABLE I: AR(2) coefficients estimation comparison

M/N	$\hat{a}_1$ **proposed**	$\hat{a}_1$ **[34]**	$\hat{a}_2$ **proposed**	$\hat{a}_2$ **[34]**	$\hat{\sigma^2}$ **proposed**	$\hat{\sigma^2}$ **[34]**
0.10	4.80E-02	≈ 2.00E-03	1.69E-02	≈2.00E-03	1.40E-02	≈5.00E-04
0.17	2.20E-02	≈7.00E-04	5.00E-03	≈7.00E-04	1.1E-02	≈4.00E-04
0.25	6.00E-03	≈3.00E-04	2.50E-03	≈5.00E-04	5.40E-03	≈2.00E-04
0.33	5.00E-03	≈3.00E-04	2.00E-03	≈4.00E-04	4.80E-03	≈1.50E-04
0.40	5.00E-03	≈1.50E-04	1.00E-03	≈1.80E-04	3.5E-03	≈1.00E-04

depicted in Table I show that a more complex modeling as the one by Kail et al. leads to improved results. However, such a complex modeling comes with an increased computational cost. In [34] the authors claim that the estimation of the parameters of an AR(2) model and noise variance takes approximately 11s. On the contrary, the techinque we propose is computationally lighter requiring only 0.03s to complete the same task using non optimized Matlab® code on an Intel® i5 processor @ 3.0GHz.

Lastly, we compare our compressive covariance estimation technique with another method [35] which uses a different sampling technique to achieve the same goal: recover the covariance matrix. This technique aims to estimate a low-rank Toeplitz structured covariance matrix given the quadratic measurements obtained through the use of Generalized Nested Sampling. To perform this comparison we generated low-rank Toeplitz covariance matrices through the use of the decomposition shown in [33] with ranks $r = \{4, 8\}$. Then, order to obtain different M/N ratios, since GNS directly relates the size of M with N, we employed two different values for the signal sample size namely $N = \{200, 350\}$[1]. The results are shown in Table II. As can be seen, both techniques reach very low NMSE values at low compression ratios. More in detail, the proposed technique shows better results when considering higher ranks, while the method in [35] seems to perform better on lower ranks. However, while the two results are quite close, the GNS sampling method is not a good candidate for CS recovery as shown in Sec. II-B, in contrary to the proposed one which is still able to correctly recover compressed signals.

TABLE II: Low-rank Toeplitz structured compressive covariance estimation

N	**rank**	**M/N [35]**	**M/N proposed**	**NMSE proposed**	**NMSE [35]**
200	8	0.22	0.20	7.73E-02	8.73E-02
	4	0.22	0.21	6.06E-02	5.04E-02
350	8	0.16	0.16	7.51E-02	8.90E-02
	4	0.13	0.12	6.37E-02	5.55E-02

III. Compressive Bayesian AR(1) estimation

As previously discussed, the coefficients of an AR process can be estimated in an efficient way in the compressed domain and can be used to perform compressive covariance estimation. Among other things, they can be used to estimate the compressibility of a signal.

[1]Since the proposed technique requires $M \propto (p+1)$ and the one in [35] strictly depends on N, the compression ratios M/N of the two techniques are approximately rather than exactly the same.

While the compressive AR(p) estimator we introduced in the previous section achieves excellent performance and it is more general since can be applied to autoregressive processes of any order p, it can be further improved and made more robust by employing Bayesian techniques. In fact, if we consider autoregressive processes of the first order $p = 1$, it is possible to explicitly obtain a compressive Bayesian estimator as described below. It is also worth noting that when higher orders $p > 1$ are considered, the stationarity constrains we impose on the autoregressive coefficient can not be exploited due to its recursive dependency with the reflection coefficients. Thus, it is not possible to obtain an explicit compressive Bayesian AR(p) estimator without employing computationally expensive algorithms such as Markov chain Monte Carlo like techniques. Hence, in the following we focus on the particular case when the order reduces $p = 1$: the AR coefficient turns out to be the correlation coefficient among the samples of the signal. This information is closely related to the complexity of a signal and its inference can improve the knowledge of the uncompressed signal. In the following, we introduce a novel Bayesian estimator for compressed AR(1) processes which leads to better performance for highly compressed signals.

A. Modeling

In order to improve the readability, let us denote with ρ the AR(1) coefficient. According to the CS scheme introduced in Section II, the acquired measurements lead to the following observation model

$$\begin{aligned} \mathbf{\Phi'}\mathbf{x}^{+} &= \mathbf{\Phi'}(\mathbf{x}^{-}\rho + \boldsymbol{\epsilon}) \\ \mathbf{y}^{+} &= \mathbf{y}^{-}\rho + \boldsymbol{\zeta}, \end{aligned}$$

where $\mathbf{x}^{-} = \mathbf{x}_{1\rightarrow(N-1)}$.

The set of parameters $\boldsymbol{\Theta} = \{\rho, \sigma^2_{\boldsymbol{\epsilon}}\}$ we wish to estimate includes the AR(1) coefficient and the variance of the Gaussian process of the AR model. Noting that $\boldsymbol{\zeta}$ follows a Gaussian distribution, we can write the probability of the observation model as:

$$p(\mathbf{y}^{+}|\mathbf{y}^{-}, \boldsymbol{\Theta}) = \mathcal{N}(\mathbf{y}^{+}|\mathbf{y}^{-}\rho, \mathbf{\Phi'}\sigma^2_{\boldsymbol{\epsilon}}\mathbb{I}\mathbf{\Phi'^{\intercal}}). \tag{11}$$

Let us discuss the choice of the prior distributions for the set of parameters $\boldsymbol{\Theta}$. When choosing the probability distribution for ρ it is worth noting that, in order to ensure stability, the necessary condition for the stationarity of the process [36] requires the values of ρ to be bounded in the interval $(-1, 1)$. Among the class of bounded probability distributions we choose the Beta distribution since it allows us to shape the signal distribution in a flexible way, which includes the non informative uniform distribution as a special case. The Beta

distribution is a bounded distribution defined on the interval $[0, 1]$. Hence, in order to bound the Beta distribution in the interval of interest $[-1, 1]$, the probability of ρ can be defined as:

$$p(\rho) = \text{Beta}(0.5 + 0.5\rho|\rho_\alpha, \rho_\beta), \tag{12}$$

where ρ_α, ρ_β are hyperparameters controlling the shape of the distribution. Since we expect σ_ϵ^2 to be positive valued, we put on this parameter an inverse gamma distribution which is commonly used for variance modeling and, being conjugate prior with the Gaussian distribution, allows easier calculations. The probability of σ_ϵ^2 is hence defined as:

$$p(\sigma_\epsilon^2) = \text{IG}\left(\sigma_\epsilon^2 \middle| a, b\right), \tag{13}$$

where a and b are two hyperparameters. According to the Bayesian modeling employed we have four hyperparamters $\boldsymbol{\Theta}_h = \{\rho^\alpha, \rho^\beta, a, b\}$ which, since no assumptions can be made, are manually set. In fact, if there is no additional information regarding the shape of the distributions of $\boldsymbol{\rho}$ and σ_ϵ^2 , it is convenient to shape them as flat priors. On the contrary, if some information is provided it can be incorporated in the distributions through the use of the hyperparamters. By analyzing the empirical probability distributions for $\boldsymbol{\rho}$ and σ_ϵ^2, the hyperparameters can be tuned to shape the distributions accordingly. As an example, given that ρ^α, ρ^β control the distribution on ρ, in image processing problems we may want to peak the distribution around 1 since an image patch has higher probability to be a smooth region than a high frequency one. Moreover, when $\rho^\alpha = \rho^\beta = 1$ this will result in a flat (uninformative) prior on ρ.

B. Inference

In order to obtain better results by taking into account the uncertainties of the estimates, we employ Bayesian inference. The goal is to obtain the probability distributions of the parameters in $\boldsymbol{\Theta}$ instead of point-wise estimates like *maximum likelihood* or *maximum-a-posteriori*. If we consider the log joint distribution $\log p(\mathbf{y}^+, \mathbf{y}^-, \boldsymbol{\Theta})$, the presence of the Beta distribution which is not conjugate prior to the Gaussian distribution does not allow a direct Bayesian inference. For this reason we employ the variational Bayesian framework [37]. This approach, in particular using the Mean-Field approximation [37], seeks a set of disjoint set of distributions that approximate the full posterior which minimizes the Kullback-Leiber (KL) divergence. In particular we have:

$$p(\boldsymbol{\Theta}|\mathbf{y}^-, \mathbf{y}^+) \simeq q(\boldsymbol{\Theta}) = q(\rho)q(\sigma_\epsilon^2). \tag{14}$$

The best function $q(\bullet)$ in terms of KL-divergence is obtained as the expectation $q(\bullet) = \langle p(\boldsymbol{\Theta})\rangle_{\boldsymbol{\Theta}\setminus\bullet}$. Thus, given the log-joint distribution, we can write:

$$\begin{aligned}\log q(\rho) &= \langle \log p(\mathbf{y}^+, \mathbf{y}^-, \boldsymbol{\Theta})\rangle_{\boldsymbol{\Theta}\setminus\rho} = \\ &= c_0 - \frac{1}{2\sigma_\epsilon^2}(\mathbf{y}^+ - \mathbf{y}^-\rho)^\intercal(\boldsymbol{\Phi}'\boldsymbol{\Phi}'^\intercal)^{-1}(\mathbf{y}^+ - \mathbf{y}^-\rho) + \\ &+ \underbrace{(\rho^\alpha - 1)\log(1+\rho)}_{\text{e}} + \underbrace{(\rho^\beta - 1)\log(1-\rho)}_{\text{g}},\end{aligned} \tag{15}$$

where c_0 is a constant term in ρ. Since this distribution does not allow an easy form, we propose to use a good approximation for the terms e and g which is very strict around 0. In particular we have:

$$\begin{aligned}e &= (\rho^\alpha - 1)\log(1+\rho) \leq (\rho^\alpha - 1)\rho \\ g &= (\rho^\beta - 1)\log(1-\rho) \leq (\rho^\beta - 1)(-\rho - \frac{\rho^2}{2}).\end{aligned}$$

Using such approximations, we can see that (15) can be rewritten in the standard form of the Gaussian distribution. Thus, we can compute the first and second order statistics by derivating twice with respect to ρ and hence the probability distribution of ρ becomes:

$$q(\rho) \sim \mathcal{N}\left(\rho \middle| -\frac{m}{2n}, \frac{1}{2n}\right) \tag{16}$$

having defined

$$\begin{aligned}m &= -\frac{1}{2}(\rho^\beta - 1) - \frac{1}{2\sigma_\epsilon^2}\mathbf{y}^{-\intercal}(\boldsymbol{\Phi}'\boldsymbol{\Phi}'^\intercal)^{-1}\mathbf{y}^- \\ n &= (\rho^\alpha - 1) - (\rho^\beta - 1) + \frac{1}{\sigma_\epsilon^2}\mathbf{y}^{-\intercal}(\boldsymbol{\Phi}'\boldsymbol{\Phi}'^\intercal)^{-1}\mathbf{y}^+.\end{aligned} \tag{17}$$

The same process applies for the parameter σ_ϵ^2. Since the inverse Gamma distribution is conjugate prior with the Gaussian one, the result is still an inverse Gamma distribution defined by different parameters. In fact, it can be shown that σ_ϵ^2 it is distributed as an inverse gamma defined by:

$$q(\sigma_\epsilon^2) \sim \text{IG}\left(\sigma_\epsilon^2 \middle| \tilde{a}, \tilde{b}\right) \tag{18}$$

where

$$\begin{aligned}\tilde{a} &= a + \frac{M}{2} \\ \tilde{b} &= b + \frac{1}{2}\left(\mathbf{y}^+ - \mathbf{y}^-\rho\right)^\intercal\left(\boldsymbol{\Phi}'\boldsymbol{\Phi}'^\intercal\right)^{-1}\left(\mathbf{y}^+ - \mathbf{y}^-\rho\right).\end{aligned}$$

Then, as can be seen from (17), in order to compute $q(\rho)$ we only need the expectation of the inverse of σ_ϵ^2 which can be computed as:

$$\left\langle \frac{1}{\sigma_\epsilon^2} \right\rangle = \frac{\tilde{a}}{\tilde{b}}. \tag{19}$$

To conclude, the whole inference process of the AR(1) parameters is summarized in Algorithm 1. It is worth noting that in order to define a criterion for evaluating the convergence, a good metric is the difference in the likelihood $p\left(\mathbf{y}^+|\boldsymbol{\Theta}, \mathbf{y}^-\right)$ between two subsequent iterations.

Algorithm 1 Bayesian AR(1) parameter estimation algorithm

INPUT: $\mathbf{y}^+, \mathbf{y}^-, \rho^\alpha, \rho^\beta, a, b$
INITIALIZE: $\hat{\rho} = 1, (1/\hat{\sigma_\epsilon^2}) = 1$
1: **while** not reached convergence **do**
2: Compute $\hat{\rho} = \langle q(\rho)\rangle$ according to (16) and (17)
3: Compute $\hat{\rho}^2 = \text{Var}(q(\rho))$ according to (16) and (17)
4: Compute $\frac{\hat{1}}{\hat{\sigma_\epsilon^2}} = \left\langle \frac{1}{q(\sigma_\epsilon^2)} \right\rangle$ according to (19)
5: **end while**
OUTPUT: $\hat{\rho}, \hat{\sigma_\epsilon^2} = \left\langle q(\sigma_\epsilon^2)\right\rangle$

C. Comparison

This algorithm, despite being specific for the AR(1) model, is able to improve the estimation performance when high compression ratios are employed. To better show this, in Figure 4 we can see a comparison of the two algorithms. For the experiment, a stationary AR(1) process was generated with $\rho \in (-1, 1)$ and the signals were compressed using the sensing matrix design introduced in Section II. For both algorithms we computed the NMSE defined as $\text{NMSE} = \left(\frac{\hat{\rho}-\rho_{\text{true}}}{\rho_{\text{true}}}\right)^2$. The results were averaged over 1000 experiments. It is worth noting that, according to the described setup, the convergence was usually reached in no more than 5 iterations.

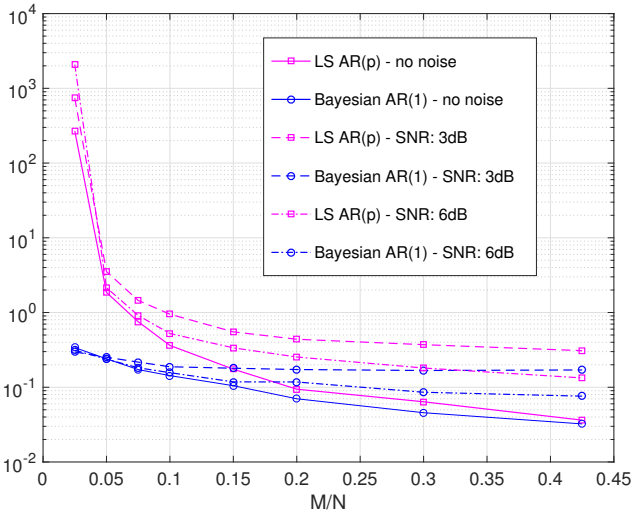


Fig. 4: NMSE comparison of LS method in (6) and the Bayesian method described in Algorithm 1. In this experiment $\rho = 0.8$, $N = 100$. Here, we consider non-noisy signals along with noisy signals having two different SNRs: 3dB and 6dB. Only high compression ratios are shown since the two methods tends to converge as M/N approaches 1.

We can see that, for extremely compressed signals whose compression ratio M/N ranges between 0.01 and 0.3 the Bayesian AR(1) specific algorithm achieves significantly lower estimation errors. Moreover, in the same picture the effects of additive noise are depicted. We repeated the same experiment by introducing an AWGN noise source corrupting the signals. The noise was generated in such a way to obtain two different Signal to Noise (SNR) ratios namely 3dB and 6dB. As we can see, the Bayesian algorithm shows superior robustness to noise by leading to consistently lower errors at different compression ratios. We rely on this important feature to design an adaptive compressive imaging scheme that is explained more in detail in the next section.

IV. Adaptive Compressive Imaging

In this Section we propose a novel algorithm for adaptive compressive imaging based on Bayesian AR(1) inference. To motivate our algorithm, let us start by considering two blocks $\mathbf{b}_1, \mathbf{b}_2 \in \mathbb{R}^{B\times B}$ extracted from an image $\mathbf{I} \in \mathbb{R}^{NB\times NB}$.

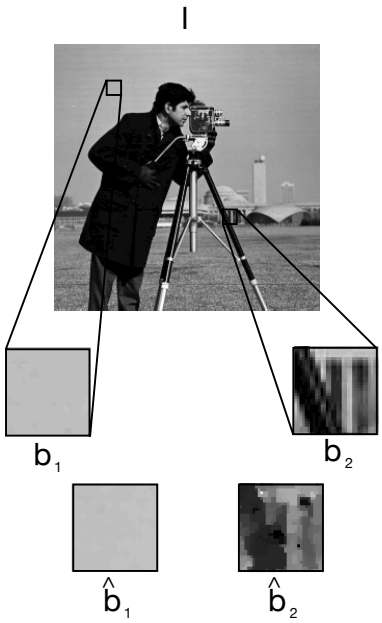


Fig. 5: Blocks in natural images exhibit high ($\mathbf{b}_1$) and low ($\mathbf{b}_2$) spatial frequencies. $\hat{\mathbf{b}}_1$ and $\hat{\mathbf{b}}_2$ are the recovered blocks with $M/N = 0.1$

In Figure 5 two common kinds of block characteristics are depicted: low ($\mathbf{b}_1$) and high ($\mathbf{b}_2$) spatial frequency blocks. Natural images typically involve large smooth regions [38] hence the number of low frequency blocks is significantly higher than the high frequency ones. Therefore, a compressive imaging scheme taking the same number of measurements on all blocks is going to provide significantly suboptimal performance, since smooth blocks are going to be over represented ($\hat{\mathbf{b}}_1$), and high-frequency blocks under represented ($\hat{\mathbf{b}}_2$) in the compressed representation. This may also lead to significant blocky artifacts in the reconstructed image. This can be seen in the recovered blocks $\hat{\mathbf{b}}_1$ and $\hat{\mathbf{b}}_2$ in Fig. 5 where the block with high spatial frequencies has a noticeably less visually satisfying recovery. The algorithm we are going to introduce aims to adapt the sensing process by selecting a suitable number of measurements for each block depending on its statistics. Let us start by considering the block $\mathbf{b}_1$ which shows low spatial frequencies. The correlation coefficient ρ computed using an AR(1) LS estimator (as described in (4)) using $\text{vec}(\mathbf{b}_1)$ is high, as one may expect, *i.e.,* $\rho = 0.99 \simeq 1$. The same coefficient, estimated from compressed measurements using Algorithm 1 (compression ratio of 0.3, with $B = 16$) is $\hat{\rho} = 0.95$.

On the other hand, a block which contains high spatial frequencies $\mathbf{b}_2$ will result in lower correlation coefficient ($\rho = 0.6$ and $\hat{\rho} = 0.55$ respectively computed from uncompressed and compressed block).

This means that by working in the compressed domain, hence without the need of recovering the signal, we can adapt the number of needed measurements depending on the complexity of each block. We remark that, in this specific application, since we are only interested in measuring the compressibility of each block of the image, an AR(1) model

is perfectly adequate to the task at hand.

A. Adapting the compression ratio

In order to correlate the compression ratio $r = M/N$ and the AR(1) coefficient ρ we must define a function that will result in high compression ratio when $\rho \simeq 1$, and will reduce it as the correlation coefficient goes toward zero. In this paper we propose to use a function $f_c(\rho)$ which has been empirically shown to be effective. The function is defined as:

$$f_c(\rho) = \lfloor (M_{\max} - M_{\min}) \, |1 - \rho|^{\gamma} + M_{\min} \rfloor, \tag{20}$$

where $M_{\max}$ ($M_{\min}$) corresponds to the maximum (minimum) value of M which is desired for the problem of interest and γ is a parameter which can be used to tune the recovery quality. For values of γ higher than 0.5, as the signal becomes less complex (ρ approaches 1) the number of measurements slowly decreases. On the other hand, when γ is smaller than 0.5, as long as ρ is not very close 1 the number of required measurements does not decrease steeply. This behavior is highlighted in our experiments where we show that different γ lead to different reconstruction qualities.

B. Adaptive compressive imaging scheme

We now put it all together and introduce a scheme in which the number of measurements is adapted to the complexity of signal itself. The scheme we consider involves a sensor and a reconstruction unit (S-RU). This approach is very general and allows us to include many different subproblems as special cases. In fact, this scheme can be employed for any kind of signal, although in this paper we consider its application to block-based compressive imaging. The basic concept is that the sensor first acquires a small batch of measurements. From those measurements, using the proposed estimator the actual number of measurements needed to achieve the desired quality is calculated, and more measurements are acquired so as to reach this number. It should be noted that the estimation algorithm can run directly on the sensor; alternatively, one could envisage that the first batch is sent to the RU, which runs the estimation and then requests from the sensor the extra measurements needed.

Though not considered in this paper, we could equivalently employ this scheme in a compressive imaging system [39] in which measurements are acquired at subsequent time intervals, instead of block by block.

We assume the sensor acquires the image by means of CS which is performed separately on each block $\mathbf{b}_i$ of size $B \times B$ which the image is composed of. At first, each block $\mathbf{b}_i \; \forall i \leq n_B$ is sensed using the minimum number of measurements $M_{\min}$, which is typically not sufficient for a visually satisfactory reconstruction. From this batch of measurements, the complexity can be efficiently estimated using Algorithm 1. At this point, the sensor is aware of the complexity of each block since it has access to $\hat{\rho}_i \; \forall i \leq n_B$. The number of needed measurements M_R to achieve a satisfactory recovery is then computed using (20) for each block and the missing $M_i = M_R - M_{\min}$ measurements $\mathbf{y}_i^{\star} \in \mathbb{R}^{M_i \times 1} \; \forall i \leq n_B$ are requested from the sensor. The last step consist in the recovery of each block. To solve the problem we employ the Block Compressed Sensing with Smooth Projected Landweber Reconstruction using Directional DWT (BCS-SPL-DDWT) technique introduced in [25] which performs block CS image recovery and showed superior recovery capabilities. In order to suit our problem, this algorithm has been adapted by substituting fixed sensing matrices with sensing matrices which are different in size for each block. More in detail the adaptive sensing problem defined at each block $\mathbf{b}_i$ becomes

$$\begin{bmatrix} \mathbf{y}_i^{\star} \\ \mathbf{y}_i \end{bmatrix} = \begin{bmatrix} \mathbf{\Phi}^{\star} \\ \mathbf{\Phi}_i \end{bmatrix} \, \mathrm{vec}(\mathbf{b}_i), \tag{21}$$

where $\mathbf{y}_i^{\star}$ and $\mathbf{\Phi}^{\star}$ are the measurements and the sensing matrix obtained during the first coarse compression at a fixed compression ratio given $M_{\min}$; $\mathbf{y}_i$ and $\mathbf{\Phi}_i$ are instead the measurements and the sensing matrix resulting after the estimation of the correlation coefficient ρ_i.

Block size: The choice of the block size is crucial to achieve good performance. In fact, if the blocks are not sufficiently small the approximation of the block with an AR(1) process will not hold since the block is more likely to contain subregions which exhibit different types of correlations among the pixels. More formally, we can state that we seek a dimension B for the blocks such that the stationarity assumption on the blocks is satisfied. We ran experiments over different images and we found that one of most common choices for block size *i.e.,* $B = 16$ is the largest block size that fits the stationarity requirements. Hence, this is the block size we have used for the experiments presented in the next section.

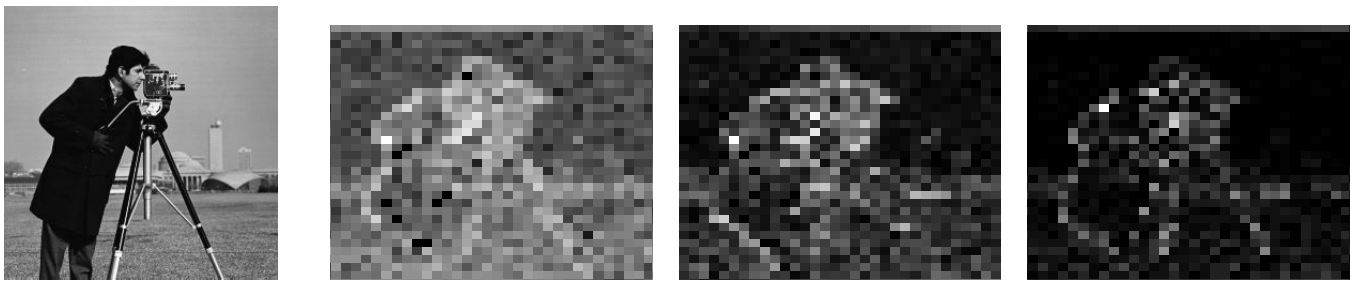
V. Results

In this section we show the performance of the proposed adaptive compressive imaging scheme. The parameters of Algorithm 1 denoted by $\mathbf{\Theta}_h$ were set according the prior knowledge we have on the blocks statistics. We chose $\rho_\alpha = 0.8$ and $\rho_\beta = 0.8$ because, as previously discussed, this values lead to a distribution on the parameter ρ highly peaked around 1 since we expect most of the blocks to be smooth. The Bayesian parameters were set as $a = 2$, $b = 1$. These values were experimentally found to yield better recovery results. Then we set $M_{\min} = 22$ to be large enough to allow good complexity estimation from the first batch of measurements, but still not sufficient image recovery, and $M_{\max} = 250$ to be large but still having $M/N < 1$.

Given the image *Cameraman*, we start by showing in Fig. 6 the actual number of measurements chosen by the algorithm for each block by taking into account three different values of γ.

Higher values of γ favor more complex blocks by allocating most of the measurements to this class of blocks. As the value of γ decreases, more measurements are added to blocks which show *medium* complexity. The result is an increased compression ratio for high γ values and a reduced compression for smaller values.

Next, we assess the end-to-end performance of the adaptive compressive imaging system by evaluating the PSNR and SSIM [40] values of recovered images compared with non-adaptive BCS-SPL-DDWT algorithm. For this experiment the



(a) Original image (b) $\gamma = 0.2$ (c) $\gamma = 0.5$ (d) $\gamma = 0.8$

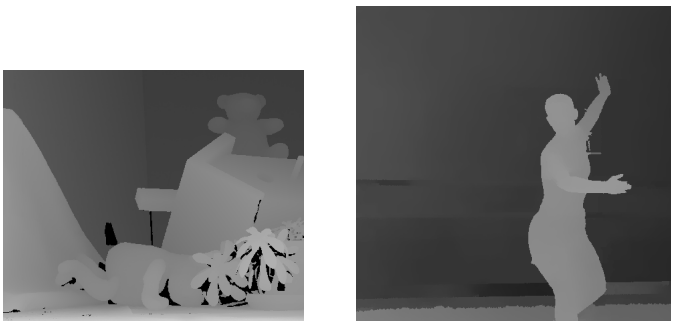
Fig. 6: Number of required measurements for each block for different values of the parameter γ. In (b)—(d) gray intensity indicates the number of measurements needed for the corresponding block. Black corresponds to $M_{\min}$ and white to $M_{\max}$.

(a) Original image (b) Proposed adaptive system (c) Non-adaptive [25]

Fig. 7: Detail of *Lena* recovered from its compressed version given $M/N = 0.17$; (a) original image; (b) image recovered using the proposed algorithm; (c) image obtained using [25].

parameters are set as above; in both the adaptive and non-adaptive case, the same total number of measurements is used; in the non adaptive case, the measurements are equally split among all blocks. The results of these experiments are shown in Table III. We can see that in the vast majority of the cases considered, the ability to adapt the number of measurements according to the complexity of the block leads to superior recovery performance. More in detail, the proposed algorithm reached PSNR gains ranging from 0.4dB up to more than 6dB. We also considered a very sparse image containing well defined edges and smooth regions *i.e.,* the Shepp-Logan phantom image. The results are extremely good for the proposed adaptive algorithm which is able to efficiently allocate the measurements only where needed. For this particular case the gain reached up to 4 dB. Very high gains are also achieved when considering depth-map images (shown in Fig. 8) where allocating more measurements mainly in high complexity regions is crucial.

In Fig. 7 we show a crop of the image $Lena$ acquired with the adaptive and non-adaptive algorithms. The detail shows a high frequency region which is good tesbed for evaluating the visual quality of the recovered images. As can be seen a higher visual quality is achieved when the adaptive algorithm is used. The details are better preserved due to the higher number of measurements allocated in this region.



(a) Teddy (b) Ballet

Fig. 8: Depth-map images taken from [41] and [42].

VI. Conclusion

In this paper we considered the problem of the AR(p) parameters estimation from compressed sensing measurements. We have shown that, through an appropriate design of the sensing matrix, one can obtain efficient AR(p) estimation while retaining the ability to reconstruct the signal, and particularly a robust AR(1) Bayesian estimator. Moreover, we have shown that such AR(p) estimators can be employed in several practical applications. In particular, they enable Toeplitz structured compressive covariance estimation with high accuracy at a very low degree of complexity. Moreover, they also enable to estimate the degree of complexity of a compressively sensed signal block directly from its measurements, which is useful in adaptive compressive sensing/imaging in order to choose the optimal number of measurements for each block; experimental results have shown significant gains with respect to a nonadaptive scheme. Indeed, future work will point towards the investigation of other signal models which can enable further compressive signal processing applications.

TABLE III: Comparison between the proposed adaptive compressive imaging system and BCS-SPL-DDWT [25].

	γ	PSNR (dB)		SSIM		M/N
		proposed	[25]	proposed	[25]	
Lena	0.2	**37.1390**	36.7500	**1.0000**	**1.0000**	0.5360
	0.5	**30.9200**	30.4680	**1.0000**	**1.0000**	0.1930
	0.8	**28.1990**	27.3360	**1.0000**	0.9970	0.1110
Cameraman	0.2	**34.2470**	33.4370	**1.0000**	**1.0000**	0.4760
	0.5	**29.0850**	28.8230	**1.0000**	0.9990	0.2550
	0.8	**25.9430**	25.0910	**1.0000**	0.9940	0.1320
Barbara	0.2	**28.4160**	27.8840	**1.0000**	0.9980	0.4820
	0.5	**24.3780**	23.9040	**1.0000**	0.9970	0.2320
	0.8	22.8490	**22.8590**	**1.0000**	0.9900	0.1430
Monarch	0.2	**36.5910**	35.3400	**1.0000**	0.9980	0.4860
	0.5	**28.1490**	27.6660	**1.0000**	0.9950	0.2040
	0.8	**25.1980**	23.9850	**1.0000**	0.9910	0.1260
Shepp-Logan Phantom	0.2	**32.2890**	31.1050	**1.0000**	0.9980	0.1640
	0.5	**31.7860**	28.1710	**0.9980**	0.9960	0.1010
	0.8	**30.0290**	26.2460	**0.9980**	0.9940	0.0830
Teddy	0.2	**39.2730**	33.4550	**1.0000**	0.9980	0.3750
	0.5	**29.3980**	27.2180	**0.9970**	0.9960	0.1710
	0.8	**26.5270**	24.9280	**0.9950**	0.9940	0.0990
Ballet	0.2	**34.7370**	30.8670	**0.9990**	0.9980	0.3500
	0.5	**30.4340**	26.2080	**0.9950**	0.9940	0.1360
	0.8	**26.5790**	26.1400	**0.9950**	**0.9950**	0.1010

REFERENCES

[1] Emmanuel J Candès, Justin Romberg, and Terence Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *Information Theory, IEEE Transactions on*, vol. 52, no. 2, pp. 489–509, 2006.

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

[3] 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.

[4] Gregory W Wornell, "Wavelet-based representations for the 1/f family of fractal processes," *Proceedings of the IEEE*, vol. 81, no. 10, pp. 1428–1450, 1993.

[5] Richard G Baraniuk, Volkan Cevher, Marco F Duarte, and Chinmay Hegde, "Model-based compressive sensing," *Information Theory, IEEE Transactions on*, vol. 56, no. 4, pp. 1982–2001, 2010.

[6] Matteo Testa and Enrico Magli, "Autoregressive process paramter estimation from compressed sensing measurements," in *Signals, Systems and Computers, 2015 49th Asilomar Conference on*. IEEE, 2015.

[7] Xiaolin Wu and Xiangjun Zhang, "Adaptive structured recovery of compressive sensing via piecewise autoregressive modeling," in *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*. IEEE, 2010, pp. 3906–3909.

[8] Robert Bos, Stijn De Waele, and Piet MT Broersen, "Autoregressive spectral estimation by application of the burg algorithm to irregularly sampled data," *Instrumentation and Measurement, IEEE Transactions on*, vol. 51, no. 6, pp. 1289–1294, 2002.

[9] Michael A Lexa, Mike E Davies, John S Thompson, and Janosch Nikolic, "Compressive power spectral density estimation," in *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*. IEEE, 2011, pp. 3884–3887.

[10] Chia-Pang Yen, Yingming Tsai, and Xiaodong Wang, "Wideband spectrum sensing based on sub-nyquist sampling," *Signal Processing, IEEE Transactions on*, vol. 61, no. 12, pp. 3028–3040, 2013.

[11] Geert Leus and Dyonisius Dony Arianandа, "Power spectrum blind sampling," *IEEE Signal Processing Letters*, vol. 18, no. 8, pp. 443–446, 2011.

[12] Dyonisius Dony Arianandа and Geert Leus, "Compressive wideband power spectrum estimation," *IEEE Transactions on Signal Processing*, vol. 60, no. 9, pp. 4775–4789, 2012.

[13] Zhi Tian and Georgios B Giannakis, "Compressed sensing for wideband cognitive radios," in *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*. IEEE, 2007, vol. 4, pp. IV–1357.

[14] James D Krieger, Yuval Kochman, and Gregory W Wornell, "Design and analysis of multi-coset arrays," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*. IEEE, 2013, pp. 3781–3785.

[15] Piya Pal and PP Vaidyanathan, "Nested arrays: a novel approach to array processing with enhanced degrees of freedom," *Signal Processing, IEEE Transactions on*, vol. 58, no. 8, pp. 4167–4181, 2010.

[16] Yuri Abramovich, Douglas Gray, Alexei Y Gorokhov, Nicholas K Spencer, et al., "Positive-definite toeplitz completion in doa estimation for nonuniform linear antenna arrays. i. fully augmentable arrays," *Signal Processing, IEEE Transactions on*, vol. 46, no. 9, pp. 2458–2471, 1998.

[17] Daniel Romero, Dyonisius Dony Arianandа, Zhi Tian, and Geert Leus, "Compressive covariance sensing," *Signal Processing Magazine, IEEE*, vol. 33, no. 1, pp.78–93, 2016.

[18] José M Bioucas-Dias, David Cohen, and Yonina C Eldar, "Covalsa: Covariance estimation from compressive measurements using alternating minimization," in *Signal Processing Conference (EUSIPCO), 2014 Proceedings of the 22nd European*. IEEE, 2014, pp. 999–1003.

[19] Daniel Romero and Geert Leus, "Wideband spectrum sensing from compressed measurements using spectral prior information," *Signal Processing, IEEE Transactions on*, vol. 61, no. 24, pp. 6232–6246, 2013.

[20] Daniel Romero, Roberto Lopez-Valcarce, and Geert Leus, "Compression limits for random vectors with linearly parameterized second-order statistics," *Information Theory, IEEE Transactions on*, vol. 61, no. 3, pp. 1410–1425, 2015.

[21] Justin Romberg, "Imaging via compressive sampling [introduction to compressive sampling and recovery via convex programming]," *Signal Processing Magazine, IEEE*, vol. 25, no. 2, pp. 14–20, 2008.

[22] Yilun Wang, Junfeng Yang, Wotao Yin, and Yin Zhang, "A new alternating minimization algorithm for total variation image reconstruction," *SIAM Journal on Imaging Sciences*, vol. 1, no. 3, pp. 248–272, 2008.

[23] Jian Zhang, Shaohui Liu, Ruiqin Xiong, Siwei Ma, and Debin Zhao, "Improved total variation based image compressive sensing recovery by nonlocal regularization," in *Circuits and Systems (ISCAS), 2013 IEEE International Symposium on*. IEEE, 2013, pp. 2836–2839.

[24] James E Fowler, Sungkwang Mun, and Eric W Tramel, "Block-based compressed sensing of images and video," *Foundations and Trends in Signal Processing*, vol. 4, no. 4, pp. 297–416, 2012.

[25] Sungkwang Mun and James E Fowler, "Block compressed sensing of images using directional transforms," in *Image Processing (ICIP), 2009 16th IEEE International Conference on*. IEEE, 2009, pp. 3021–3024.

[26] Marius Goldberg, Paul R Boucher, and Seymour Shlien, "Image compression using adaptive vector quantization," *Communications, IEEE Transactions on*, vol. 34, no. 2, pp. 180–187, 1986.

[27] S Grace Chang, Bin Yu, and Martin Vetterli, "Adaptive wavelet thresholding for image denoising and compression," *Image Processing, IEEE Transactions on*, vol. 9, no. 9, pp. 1532–1546, 2000.

[28] Chuo-Ling Chang and Bernd Girod, "Direction-adaptive discrete wavelet transform for image compression," *Image Processing, IEEE Transactions on*, vol. 16, no. 5, pp. 1289–1302, 2007.

[29] Hai-bo ZHENG and Xiu-chang ZHU, "Sampling adaptive block compressed sensing reconstruction algorithm for images based on edge

detection," *The Journal of China Universities of Posts and Telecommunications*, vol. 20, no. 3, pp. 97–103, 2013.
[30] Steven M Kay and Stanley L Marple, "Spectrum analysisa modern perspective," *Proceedings of the IEEE*, vol. 69, no. 11, pp. 1380–1419, 1981.
[31] Holger Rauhut, "Circulant and toeplitz matrices in compressed sensing," *arXiv preprint arXiv:0902.4394*, 2009.
[32] Emmanuel J Candes and Terence Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?," *Information Theory, IEEE Transactions on*, vol. 52, no. 12, pp. 5406–5425, 2006.
[33] Heng Qiao and Piya Pal, "Generalized nested sampling for compression and exact recovery of symmetric toeplitz matrices," in *Signal and Information Processing (GlobalSIP), 2014 IEEE Global Conference on*. IEEE, 2014, pp. 443–447.
[34] Georg Kail and Geert Leus, "Compressive modeling of stationary autoregressive processes," in *Information Theory and Applications Workshop (ITA), 2015*. IEEE, 2015, pp. 108–114.
[35] Heng Qiao and Piya Pal, "Generalized nested sampling for compressing low rank toeplitz matrices," *IEEE Signal Processing Letters*, vol. 22, no. 11, pp. 1844–1848, 2015.
[36] Jonathan D Cryer and Natalie Kellet, *Time series analysis*, vol. 101, Springer, 1986.
[37] Dimitris G Tzikas, Aristidis C Likas, and Nickolaos P Galatsanos, "The variational approximation for bayesian inference," *Signal Processing Magazine, IEEE*, vol. 25, no. 6, pp. 131–146, 2008.
[38] Antonio Torralba and Aude Oliva, "Statistics of natural image categories," *Network: computation in neural systems*, vol. 14, no. 3, pp. 391–412, 2003.
[39] Michael B Wakin, Jason N Laska, Marco F Duarte, Dror Baron, Shriram Sarvotham, Dharmpal Takhar, Kevin F Kelly, and Richard G Baraniuk, "An architecture for compressive imaging," in *Image Processing, 2006 IEEE International Conference on*. IEEE, 2006, pp. 1273–1276.
[40] Zhou Wang, Alan Conrad Bovik, Hamid Rahim Sheikh, and Eero P Simoncelli, "Image quality assessment: from error visibility to structural similarity," *Image Processing, IEEE Transactions on*, vol. 13, no. 4, pp. 600–612, 2004.
[41] Daniel Scharstein and Richard Szeliski, "High-accuracy stereo depth maps using structured light," in *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*. IEEE, 2003, vol. 1, pp. I–195.
[42] C Lawrence Zitnick, Sing Bing Kang, Matthew Uyttendaele, Simon Winder, and Richard Szeliski, "High-quality video view interpolation using a layered representation," in *ACM Transactions on Graphics (TOG)*. ACM, 2004, vol. 23, pp. 600–608.

Matteo Testa received the B.Sc. degree and the M.Sc. degree in telecommunications engineering from the Politecnico di Torino, Turin, Italy, in 2011 and 2012, respectively, and the Ph.D. degree in electronic and communications engineering from the Electronics Department, Politecnico di Torino, in 2016, under the supervision of Prof. E. Magli. He currently holds a post-doctoral position with the IPL lab, Politecnico di Torino, led by Prof. E. Magli in collaboration with SONY EuTEC. His research is focused on compressed sensing with a particular interest on imaging systems, forensic applications and bayesian inference.

Enrico Magli (S'97-M'01-SM'07) received the M.Sc. and Ph.D. degrees from the Politecnico di Torino, Italy, in 1997 and 2001, respectively. He is currently an Associate Professor with the same university. His research interests include compressive sensing, image and video coding, and vision. He is an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, the IEEE TRANSACTIONS ON MULTIMEDIA, and the *EURASIP Journal on Image and Video Processing*, and an IEEE Distinguished Lecturer from 2015 to 2016. He was the recipient of the IEEE Geoscience and Remote Sensing Society 2011 Transactions Prize Paper Award, the IEEE ICIP 2015 Best Student Paper Award (as senior author), and the 2010 and 2014 Best Associate Editor Award of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY.