Projection Matrix Optimization for Sparse Signals in Structured Noise

Sebastian Pazos, *Member, IEEE*, Martin Hurtado, *Member, IEEE*, Carlos H. Muravchik, *Senior Member, IEEE*, and Arye Nehorai, *Fellow, IEEE*

***Abstract*—We consider the problem of estimating a signal which has been corrupted with structured noise. When the signal of interest accepts a sparse representation, only a small number of measurements are required to retain all the information. The measurements are mapped to a lower dimensional space through a projection matrix. We propose a method to optimize the design of this matrix where the objective is not only to reduce the amount of data to be processed but also to reject the undesired signal components. As a result, we reduce the computation time and the error on the estimation of the unknown parameters of the sparse model, with respect to the uncompressed data. The proposed method has tunable parameters that can affect its performance. Optimal tuning would require a comprehensive study of parameter variations and options. To avoid this learning burden, we also introduce a variant of the algorithm that is free from tuning, without significant loss of performance. Using synthetic data, we analyze the performance of the proposed algorithms and their robustness against errors in the model parameters. Additionally, we illustrate the performance of the method through a radar application using real clutter data with a still target and with a synthetic moving target.**

***Index Terms*— Projection matrix optimization, sparse models, compressive sensing, radar.**

I. Introduction

In several signal processing applications, such as communications [1], radar [2] and brain source localization [3], the observed data $\boldsymbol{y}$ can be represented as the contribution of three sources: the signal of interest $\boldsymbol{x}$, the interference $\boldsymbol{z}$ and the noise $\boldsymbol{w}$, i.e. $\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{z} + \boldsymbol{w}$. A sparse representation of the signal of interest $\boldsymbol{x}$ results when it is expressed as a linear combination of a set of predetermined signals that forms an over-complete dictionary, $\boldsymbol{x} = X\boldsymbol{b}$. Sparse signals are particularly suited to be compressed [4]; hence, we only need a reduced number of corrupted measurements in order to recover the useful information. The set of compressed measurements can be obtained by projecting the signal on a lower dimensional space through a projection matrix A.

Finding the sparse solution $\boldsymbol{b}$ is a non-convex optimization problem because the solution is constrained to minimum ℓ_0 norm. A common procedure to circumvent this difficulty consists in using the ℓ_1 norm as a surrogate measure of sparsity, e.g. basis pursuit denoising (BPDN) [5], Orthogonal Matching Pursuit (OMP) [6] and Dantzig selector (DS) [7]. These methods only ensure a sparse solution if the joint effect of the dictionary and the projection matrix AX satisfies the restricted isometry property (RIP) [4]. It has been shown that random Gaussian matrices satisfy this condition [8]. However, for general matrices this condition is not easily attained [9] and is expensive to compute. Additionally, there exists another class of methods for solving the sparse inverse problem which applies statistical models and tools, such as automatic relevance determination (ARD) [10], automatic double over-relaxation (ADORE) [11], Fast Bayesian Matching Pursuit (FBMP) [12], Sparse Bayesian Learning (SBL) [13]. In a previous work, we developed a method, casted in the latter class, that combines the expectation maximization (EM) algorithm with a decision test [14]; herein referred to as Enhanced Sparse Bayesian Learning (ESBL). This algorithm exploits the structured representation of the interference for better denoising of the sparse signal.

In this paper, we devise a method for designing an optimal projection matrix. In this context, optimality refers to the fact that the resulting matrix not only compresses the measurements but also improves the estimation results by rejecting the undesired signals.

Previous works addressed the construction of the projection matrix through different strategies and in different scenarios. In the case of a noise-free data, [15] introduced an iterative algorithm based on shrinkage, named optimized projection, that minimizes an average measure of the mutual coherence of the effective dictionary AX. Still in the noiseless case, [16] constructed the projection matrix based on *Multidimensional Scaling* which enforces pairwise distances between new atoms in the effective dictionary AX to match as well as possible the original atoms in original dictionary X. Similarly, in [17] the target of its algorithm is to make the Gram matrix $(AX)^H(XA)$

Manuscript received September 08, 2014; revised December 26, 2014 and April 23, 2015; accepted May 05, 2015. Date of publication May 18, 2015; date of current version June 17, 2015. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Yue Rong. The work of S. Pazos, M. Hurtado, and C. Muravchik was supported by ANPCyT PICT-PRH 2009-0097, PICT 2011-0909, and UNLP 11-I-166. The work of A. Nehorai was supported by the NSF under Grants CCF-1014908 and CCF-0963742, and by ONR Grant N000140810849.

S. Pazos is with the Research Institute of Electronics, Control, and Signal Processing (LEICI), La Plata 1900, Argentina (e-mail: sebastianpazos@gmail.com).

M. Hurtado and C. H. Muravchik are with the Department of Electrical Engineering, National University of La Plata, La Plata 1900, Argentina, and also with the Research Institute of Electronics, Control, and Signal Processing (LEICI), La Plata 1900, Argentina (e-mail: martin.hurtado@ing.unlp.edu.ar; carlosm@ing.unlp.edu.ar).

A. Nehorai is with the Department of Electrical and Systems Engineering, Washington University in St. Louis, St. Louis, MO 63130 USA (e-mail: nehorai@ese.wustl.edu).

This paper has supplementary downloadable multimedia material available at http://ieeexplore.ieee.org provided by the authors. This includes the supporting Matlab code for the paper. This material is 44 KB in size.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2015.2434328

as close as possible to the identity matrix. More specific constructions are introduced in [18] based on graph coding theory which guarantees recovery; and in [19] based on BCH codes which offers a construction of binary sampling matrices.

Extending the scenario to include interference sources and noise, [20] analyzed the MIMO radar application and proposed two construction methods, the first minimized a linear combination of the coherence of the sensing matrix and the inverse of the signal-to-interference ratio and the second enforced a structure on the projection matrix to enhance SIR while maintaining a coherence similar to Gaussian matrices. Other works which targeted specific applications, imposed additional structure on the projection matrix resulting in block sparsity. [21] used an equiangular tight frame instead of the identity matrix to approximate the Gram matrix of the equivalent dictionary; and [22] minimized a weighted sum of inter-block coherence and sub-block coherence. Furthermore [23] extended [17] for signals with a Gaussian mixture model.

More related to our work is [24], which used a similar model, encompassing signal of interest, interference and noise, but used a two step process. First, it applied a projection on the orthogonal space of the interference and then solved the inverse problem with a Gaussian random matrix using Compressed Sensing. Similarly [25] minimized the mutual information between the signal and the measurements but applied the projection matrix only to the signal, not the interference.

In our work, we propose a Bayesian approach for the optimal design of the projection matrix. We select a matrix that minimizes the estimation error of the parameters of interest given the compressed measurements. This is achieved through the minimization of the trace of the posterior covariance, resulting in a highly concentrated posterior distribution. To avoid the excessive burden of optimizing entry by entry of the projection matrix, we propose a structured compression matrix to reduce the number of degree of freedom. We show that the proposed design of the projection matrix combined with the ESBL algorithm [14] reduces both computation time and estimation error. To the best of our knowledge, no other algorithm for solving the inverse sparse problem separates the source of contamination into structured interference and unstructured noise. Hence, conventional sparse methods do not take full advantage of our design procedure.

In the next section we present the data model used in this paper and formulate the optimization problem. In Section III we analyze the cost function in order to make the problem solvable, find its optimal solution when an energy restriction is used, and propose a parameter-free suboptimal solution. In Section IV we provide several simulations to analyze the performance of the proposed algorithm. We show results for a fixed value of the energy restriction and an automatic selection of its value, we analyze the robustness of our method when the statistics of the problem are not precisely or accurately known and we analyze the behavior of the method with varying sparsity of the signal. In Section V we show an application of the method to real radar data with a static target and then we apply the method to a moving synthetic target over real radar clutter. We conclude in Section VI.

Notation: We adopt the following notation. A scalar variable is denoted x, a vector is $\boldsymbol{x}$ and a matrix is X. The vectors $\boldsymbol{x}_{\cdot j}$ and $\boldsymbol{x}_{i\cdot}$ represent the jth column and ith row of matrix X, respectively. The scalar x_{ij} is the ijth entry of matrix X, and x_i is ith the element of vector $\boldsymbol{x}$. The identity matrix of size p is I_p. The conjugate transpose of a complex vector and matrix is $\boldsymbol{x}^H$ and X^H. The trace of a square matrix is $\mathrm{tr}(X)$. The operator $\otimes$ denotes the Kronecker product. $\|X\|_F$ is the Frobenius norm of matrix X.

The random vector $\boldsymbol{x} \sim \mathcal{CN}_P(\boldsymbol{\mu}, \Sigma_u)$ is a circularly symmetric complex Gaussian vector of dimension P with mean $\boldsymbol{\mu}$ and covariance Σ_u.

II. Problem Formulation

A. Model Assumptions

We consider the following noise-free linear model,

$$\boldsymbol{y}_{\mathrm{nf}} = X\boldsymbol{b} + \sum_{p=1}^{P} Z_p \boldsymbol{\mu}_p = X\boldsymbol{b} + Z\boldsymbol{\mu}, \tag{1}$$

where $\boldsymbol{y}_{\mathrm{nf}}$ is the $M \times 1$ vector of noise-free observed data, $\boldsymbol{b}$ is the $N \times 1$ vector representing the signal of interest, with $M < N$. Each $\boldsymbol{\mu}_p$ is a $Q \times 1$ vector representing one of the P interference sources. X and Z_p are the $M \times N$ and $M \times Q$ matrices of linear regressors, which are fixed for the applications treated in this paper. These variables can be compacted into $\boldsymbol{\mu} = [\boldsymbol{\mu}_1', \ldots, \boldsymbol{\mu}_P']'$ and $Z = [Z_1, \ldots, Z_P]$. In this work, we assume that the sparsity $s = \|\boldsymbol{b}\|_0$ of the vector $\boldsymbol{b}$ is lower than M and thus, it can be further compressed and fewer measurements could be taken to estimate $\boldsymbol{b}$. This operation is represented by a $K \times M$ projection matrix A, where K is chosen such that $s < K \leq M$. If there is no prior information on the sparsity of the signal then we should set $K = M$. Then the model for the compressed measurements can be expressed as

$$\boldsymbol{y} = AX\boldsymbol{b} + AZ\boldsymbol{\mu} + \boldsymbol{w}, \tag{2}$$

where $\boldsymbol{y}$ is the $K \times 1$ vector of compressed measurements and $\boldsymbol{w}$ represents not only the noise introduced by the system but also any modeling errors that may be present. A classic approach to compressed acquisition is to simultaneously optimize the projection matrix A and the dictionary X [17]. However, in several applications the dictionary X is constrained by physical considerations and must be kept fixed. Hence, we will optimize the projection matrix A alone.

The motivation for our procedure are applications to radar and communications systems, such as that described in Section V where the information involved is typically represented by complex data. However, the proposed method to build the projection matrix is not limited to radar or spatial applications. The problem is general enough to be used in a larger range of applications which share the model (1), as seen in [1] and [3] for example. Additionally, the use of Gaussian priors is also common in Bayesian frameworks of sparse problems [13], [26]. Hence, we assume that all random variables involved in the signal model (2) follow zero-mean circularly symmetric

complex Gaussian distributions. The case of real random variables can be enclosed within the more general complex ones. Furthermore there exists statistical independence among interferences, among the noise and between them. The covariance matrices for $\boldsymbol{b}$, $\boldsymbol{w}$ and $\boldsymbol{\mu}_p$ are Σ_b, σI_K and Σ_u, respectively. Thus the probability density distributions are

$$\begin{aligned}\boldsymbol{w} &\sim \mathcal{CN}_K(0, \sigma I_K),\\ \boldsymbol{\mu} &\sim \mathcal{CN}_{PQ}(0, I_P \otimes \Sigma_u),\\ \boldsymbol{b} &\sim \mathcal{CN}_M(0, \Sigma_b).\end{aligned}$$

A generalization of the previous problem arises when several measurements or snapshots are available. This new problem is known as sparse representation of multiple measurements vectors (MMV) [27]. Thus, considering a sequence of D snapshots, the data output becomes

$$Y = AXB + AZU + W, \tag{3}$$

where matrices Y, B, U, and W are built using the respective vectors from (2) as their columns. For instance $Y = [\boldsymbol{y}_{\cdot 1}, \ldots, \boldsymbol{y}_{\cdot D}]$. Assuming the data can be represented by the same elements of the dictionary, the matrix B has a small number of nonzero rows. The goal is to estimate the columns of B under the condition of common sparsity profile and statistical stationarity of the interference and noise across different snapshots. All matrices in (3) are complex.

We remark that the optimization procedure to construct the projection matrix will be based on the statistics of one snapshot and then applied to the whole sequence to solve the inverse problem.

B. ESBL Algorithm

We provide a brief description of the ESBL algorithm whose details can be found in [14]. Similar to the SBL method, this is an iterative algorithm that solves the sparse inverse problem depicted by (3). It not only computes a point estimate of B through the posterior mean but also estimates the nuisance parameters Σ_b, Σ_u and σ. Each iteration of the algorithm consists of two steps. In the first step the EM algorithm is used to find estimates of the model parameters and B, which results in a non-sparse matrix. The second step of the algorithm implements a decision test, with a threshold determined by the false alarm probability P_{FA}, to prune the statistically insignificant components of this point estimate.

C. Problem Statement

The goal is to find the projection matrix A that minimizes the reconstruction error for $\boldsymbol{b}$, assuming that the statistical parameters of the signal, interference and noise are known. In order to minimize this error we want to find a scalar cost function that measures the spread of the posterior distribution. Under the Gaussian assumption, the covariance matrix is directly related to the so-called confidence ellipsoid that encloses the values whose probability exceeds a threshold. The ellipsoid axis lengths are proportional to the eigenvalues of the covariance matrix [28, Ch. 7.5]. Thus, a classic option is to use the determinant of the posterior covariance $\Sigma_{b/y}$, which minimizes the volume of the confidence ellipsoid. Another option is to use the the trace of that matrix, that minimizes the sum of the length of the axes. With the latter measure we obtain a closed form solution to the problem and provide a significant gain on the performance, as shown through simulations in Section IV and in a real application in Section V. Therefore, we adopt the trace of the posterior covariance $\Sigma_{b/y}$, leading to a highly concentrated posterior distribution. The problem can be stated as

$$\min_A \operatorname{tr}\left(\Sigma_{b/y}\right) \quad \text{s.t.} \quad \|A\|_F^2 = E, \tag{4}$$

where the constraint is introduced to bound the energy of A.

III. Proposed Solution

A. Cost Function

In order to find the optimal solution to (4), we first analyze the cost function and find the right expression to work with. From (2) we obtain the covariance of the compressed measurements $\Sigma_y = A\Sigma A^H + \sigma I_K$ where

$$\Sigma = X\Sigma_b X^H + Z(I_P \otimes \Sigma_u)Z^H, \tag{5}$$

and then the posterior covariance is [29, Thm 10.3]

$$\Sigma_{b/y} = \Sigma_b - \Sigma_b X^H A^H \left(A\Sigma A^H + \sigma I_K\right)^{-1} AX\Sigma_b. \tag{6}$$

Applying the Woodbury identity [30, eq. 145]

$$\Sigma_{b/y} = \Sigma_b - \\ -\frac{\Sigma_b}{\sigma} X^H A^H \left[I_K - A\left(I_M + \frac{\Sigma}{\sigma} A^H A\right)^{-1} \frac{\Sigma}{\sigma} A^H\right] AX\Sigma_b. \tag{7}$$

Defining $G = A^H A$, a $M \times M$ matrix of rank K, and applying one of the Searle identities [30, eq. 154] we obtain

$$\Sigma_{b/y} = \Sigma_b - \frac{\Sigma_b}{\sigma} X^H \left(I_M + G\frac{\Sigma}{\sigma}\right)^{-1} GX\Sigma_b. \tag{8}$$

The minimization problem (4) cannot be easily solved for an arbitrary matrix A. Inspired by the procedure introduced in [31], we restrict our solution space $\mathcal{A}$ to matrices A with a defined structure. This restriction space is spanned by the eigenvectors of matrix G, which we set equal to eigenvectors of the covariance matrix Σ. Ideally, the eigenvectors of G should come from the signal space by means of Σ_b however, this makes the problem untractable and we compromise by projecting into the signal-plus-interference space. As a result the solution is no longer optimal for problem (4) and we obtain a sub-optimal A. In coming Sections, it is shown by means of simulations and radar real data, that this procedure, despite its sub-optimality, provides great performance gains over a random projection. Thus we take the eigendecomposition of Σ and G,

$$\Sigma = VDV^H \qquad G = \Gamma\Lambda\Gamma^H, \tag{9}$$

where V is the $M \times M$ matrix of eigenvectors of Σ and $D = \operatorname{diag}\{d_1, \ldots, d_M\}$ is the $M \times M$ diagonal matrix of eigenvalues of Σ; $\Lambda = \operatorname{diag}\{\lambda_1, \ldots, \lambda_K\}$ is the $K \times K$ matrix of eigenvalues of G and Γ is the $M \times K$ matrix of eigenvectors of G. From the previous discussion, the eigenvectors of G will

be a subset of the columns of V, which will be defined by the sub-optimal solution to the problem in the next section. From here on, for simplicity and without loss of generality, they will be assumed to be the first K columns of V. With these considerations we can rewrite (8) as

$$\begin{aligned}\Sigma_{b/y} &= \Sigma_b - \frac{\Sigma_b}{\sigma} X^H \left(I_M + \Gamma\Lambda\Gamma^H \frac{\Sigma}{\sigma} \right)^{-1} \Gamma\Lambda\Gamma^H X\Sigma_b \\ &= \Sigma_b - \frac{\Sigma_b}{\sigma} X^H\Gamma \left[I_M - \left(I_K + \Lambda\Gamma^H \frac{\Sigma}{\sigma}\Gamma \right)^{-1} \Lambda\Gamma^H \frac{\Sigma}{\sigma}\Gamma \right] \\ &\quad \times \Lambda\Gamma^H X\Sigma_b \end{aligned} \tag{10}$$

Since

$$\begin{aligned}\Gamma^H\Sigma\Gamma &= \Gamma^H V D V^H \Gamma = (I_K \quad \mathbf{0}) D \begin{pmatrix} I_K \\ \mathbf{0} \end{pmatrix} \\ &= \mathrm{diag}\{d_1, \cdots, d_K\} \triangleq \tilde{D}, \end{aligned} \tag{11}$$

then using, once again, a Searle identity [30, eq. 154]

$$\Sigma_{b/y} = \Sigma_b - \frac{\Sigma_b}{\sigma} X^H \Gamma G \Gamma^H X \Sigma_b, \tag{12}$$

where

$$Q = \left[I_K - \left(I_K + \Lambda\frac{\tilde{D}}{\sigma} \right)^{-1} \Lambda\frac{\tilde{D}}{\sigma} \right] \Lambda \tag{13}$$

is a diagonal matrix with elements

$$[Q]_{kk} = \frac{\lambda_k}{1 + \frac{d_k\lambda_k}{\sigma}}. \tag{14}$$

With this definition, we can rewrite the cost function as

$$tr\left(\Sigma_{b/y}\right) = tr(\Sigma_b) - \sum_{k=1}^{K} \frac{\lambda_k}{\sigma + d_k\lambda_k} t_k, \tag{15}$$

where $t_k = \|v_k\|_2^2$ and $v_k = [\Sigma_b X^H V]_{\cdot k}$ is the kth column of the matrix $\Sigma_b X^H \Gamma$. Notice that we have transformed a minimization problem over $K \times M$ matrices to one over a vector of size K, containing the eigenvalues of G.

B. Restricted Optimal Solution

Taking into account the restriction on the solution space mentioned in the previous section, and the fact that $\|A\|_F^2 = \mathrm{tr}(A^H A) = \mathrm{tr}(G) = \sum_{k=1}^{K} \lambda_k$, we define a modified problem described by

$$\min_{A\in\mathcal{A}} \mathrm{tr}\left(\Sigma_{b/y}\right) \quad \text{s.t.} \quad \sum_{k=1}^{K} \lambda_k = E, \quad \lambda_k \geq 0 \ \forall k. \tag{16}$$

We prove in the Appendix that the optimal eigenvalues of G for the previous problem are

$$\lambda_k = \begin{cases} \dfrac{t_k}{d_k} \dfrac{\sum_{i=1}^{L} \sigma/d_i + E}{\sum_{i=1}^{L} t_i/d_i} - \dfrac{\sigma}{d_k}, & 1 \leq k \leq L \\ 0, & k > L \end{cases}. \tag{17}$$

where L is the number of nonzero eigenvalues defined by the optimization process described in the Appendix. Finally, the optimal $L \times M$ projection matrix A can be obtained from G using the L optimal eigenvalues and their corresponding eigenvectors from Σ,

$$A = \mathrm{diag}\left\{ \sqrt{\lambda_1}, \ldots, \sqrt{\lambda_L} \right\} [V]_{1:L}^H. \tag{18}$$

We remark that the energy restriction E impacts the value of L and may be used as a tuning parameter affecting the behavior of the algorithm.

Notice that the optimization procedure needs the value of Σ_b, Σ_u and σ as input parameters. The assumption of knowing these parameters is not uncommon in the literature [23], [25], [32], [33]. In real applications their true values are not known. However, they can be roughly determined from previous snapshots with suitable estimation algorithms. In Section IV-C we analyze the sensitivity of the optimization with respect to the parameters.

C. Automatic selection of E

If the application does not require an energy constraint E on the rows of the matrix, then we have a new degree of freedom. In this section we introduce a heuristic methodology to choose a quasi-optimal value for the constraint E and consequently L, the number of eigenvalues used.

Fig. 1 shows the optimal values of L, obtained as solution of (16), as the constraint E is varied in a typical scenario for the problem. The optimal value for this constraint should be infinite but as E increases, this translates to larger numerical errors in the algorithm and therefore poor performance. It can be appreciated that there are threshold values of E for which the size L of the optimal matrix increases. These threshold values of E, noted as E_L, can be calculated as follows. For a fixed value of L, set the eigenvalues $\lambda_{L+1}, \lambda_{L+2}, \ldots, \lambda_K$ to zero and compute the required value of E needed to nullify λ_{L+1}, from (17)

$$E_L = \sum_{1}^{L} \frac{\sigma}{d_i} \left(\sqrt{\frac{t_i}{t_{L+1}}} - 1 \right), \tag{19}$$

and consequently, for each L using (19) in (17), the required eigenvalues λ_i are

$$\lambda_i = \max\left\{ 0, \frac{\sigma}{d_i} \left(\sqrt{\frac{t_i}{t_{L+1}}} - 1 \right) \right\} \tag{20}$$

The analytical expression of the cost function results

$$tr\left(\Sigma_{b/y}\right) = tr(\Sigma_b) - \sum_{k=1}^{L} \frac{t_k}{d_k} + \frac{\sum_{k=1}^{L} \frac{\sigma\sqrt{t_k}}{d_k} \sum_{l=1}^{L} \frac{\sqrt{t_l}}{d_l}}{\sum_{l=1}^{L} \frac{\sigma}{d_k} + E}, \tag{21}$$

whose value when $E \to 0$ is $tr(\Sigma_b)$ and when $E \to \infty$ is $tr(\Sigma_b) - \sum_{k=1}^{L} \frac{t_k}{d_k}$.

Additionally, we also illustrate in Fig. 1 the behavior of the cost function (21) as E increases, leading to an infinite optimal value for E, as mentioned before. Therefore, we have a tradeoff between choosing a lower L, in order to compress the signal,

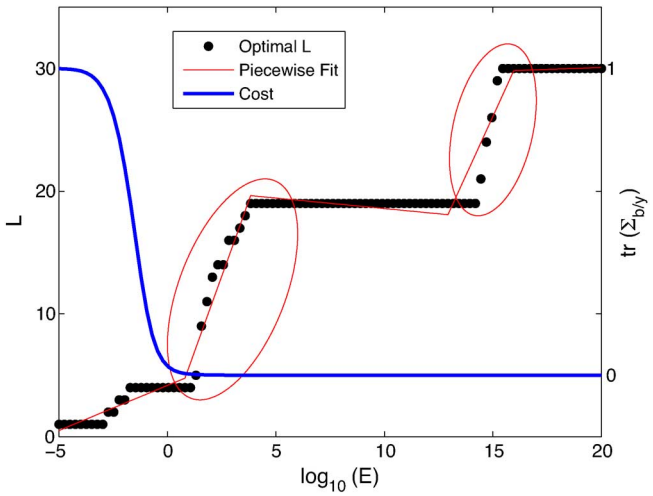

Fig. 1. Optimal L value for different values of the energy constraint E, and its corresponding cost.

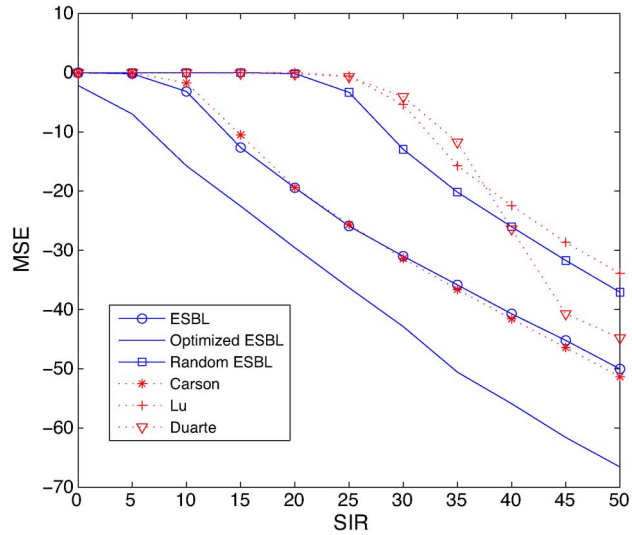

Fig. 2. Estimation relative error as a function of the SIR using a fixed value of E using no projection, optimized projection and a random projection.

and a higher L, in order to decrease the value of the cost function. The latter results in larger values for E which introduces numerical errors in the procedure.

We introduce a heuristic method to resolve that tradeoff. In Fig. 1 there are two groups of points, the first one has low values of E and L and the other has extremely high values of E, which we would like to avoid. Thus, we should choose the largest E_L of the first batch. In order to achieve this automatically, we propose the use of Multivariate Adaptive Regression Splines (MARS) [34] to fit a linear piecewise function to these points, obtain the breaking point and choose that value of E. The details of this procedure exceed the scope of this paper but they can be found in the previous reference. The results are shown in the next section.

IV. Performance Analysis

A. Fixed Value of E

To compute the empirical performance, we propose the following example of the model (2). The lengths of the vectors $\boldsymbol{y}$, $\boldsymbol{b}$ and $\boldsymbol{u}_p$ are $M = 30$, $N = 60$ and $Q = 5$, respectively. There are $P = 3$ sources of interference and the data $\boldsymbol{y}$ is observed $D = 10$ independent times.

The use of random dictionaries is a frequent approach in sparse signal processing [13], [35], [36]. Since many real applications measurements can usually be modeled as random, these random dictionaries represent a wide range of phenomena. Therefore, we use random dictionaries as fair indicators of the procedure's performance. Matrices X and Z are randomly generated only once, at the beginning of the simulation, with entries distributed as $\mathcal{CN}(0,1)$ and normalized with their Frobenius norm. The sparse vector $\boldsymbol{b}$ has a single nonzero component, thus $s = 1$. The signal power is $P_b = \|\boldsymbol{b}\|_2^2$. The interference components covariance is $\Sigma_u = P_u I_Q/\sqrt{Q}$. Then the interference power is $P_u = \|\Sigma_u\|_{\mathrm{F}}$. The measurement noise power is $\sigma = P_u/10$. For evaluating the performance empirically, we computed 500 runs of Monte Carlo simulations in which the support of $\boldsymbol{b}$ is randomly selected for each run. We estimate $\boldsymbol{b}$ using the ESBL algorithm at different signal-to-interference ratios, which we define as $\mathrm{SIR} = 10\log_{10}(P_b/P_u)$ for a false alarm rate of 10^{-3}.

For the compression, we start with a K value of 30 and an energy restriction $E = 500$. The parameters Σ_b, Σ_u and σ are assumed to be known in the simulations but they could be approximated using training data or previous estimations, as shown in the next section. Additionally, we include the use of a random matrix A to compare its performance to that of the optimal one proposed. In this random matrix each entry has a complex Gaussian distribution $A_{ij} \sim \mathcal{CN}(0,1)$ normalized to satisfy the energy restriction and with the same dimensions as the optimized one to have a fair comparison.

Using the ESBL algorithm with the optimal projection we obtain a gain of 12 dB for an estimation error of 10^{-4} over the uncompressed case, as shown in Fig. 2, even though we used fewer observations, between 10 and 15, than the 30 original ones. Additionally, we can appreciate that a random choice of the projection matrix aggravates the corresponding estimation error.

Fig. 2 also includes the performances for algorithms by Lu [24], Carson [25] and Duarte[17]. The orthogonal projection algorithm by Lu behaves practically as if using a random matrix since the number of interferers is high in relation to the measurements. As explained in their paper, the orthogonal projection suppresses the signal of interest as well as the interference in this scenario. While the model (1) considered in this paper applies the projection matrix to the signal of interest and the interference terms, Carson's algorithm designs the projection matrix based on the model $\boldsymbol{y} = AX\boldsymbol{b} + \boldsymbol{w}$, i.e. applying the projection only to the signal of interest. Not projecting the interference results in a behavior similar to uncompressed case. Finally, Duarte's algorithm does not take into account the interference, and behaves poorly at low SIR and improves as the interference is no longer significant, although not as much as our algorithm.

B. Automatic Selection of E

For this modification we use the same parameters as in the previous section but in this case the energy restriction E is au-

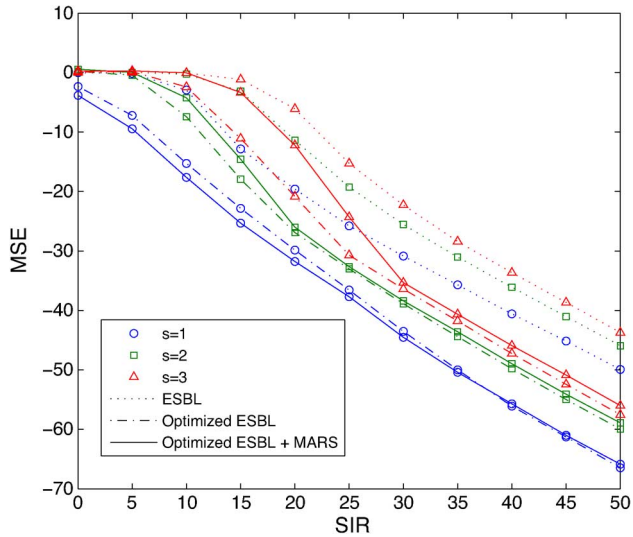

Fig. 3. Estimation relative error as a function of the SIR for different sparsity values s, with a fixed value and automatic selection of E.

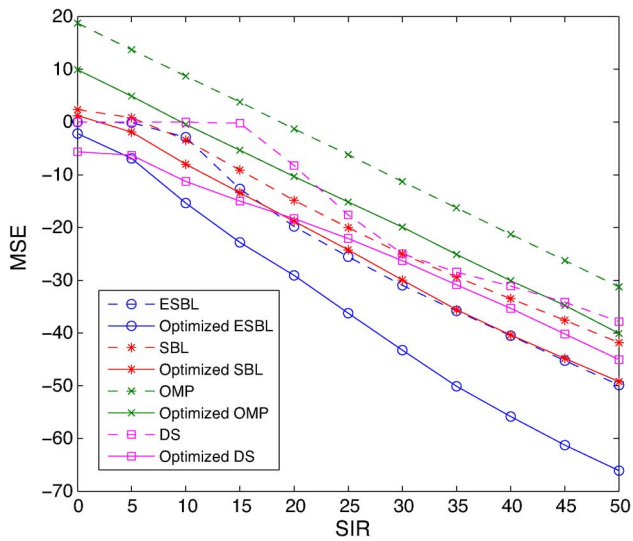

Fig. 4. Estimation relative error, $\|\hat{\boldsymbol{b}}-\boldsymbol{b}\|_2^2/\|\hat{\boldsymbol{b}}\|_2^2$, as a function of the SIR when using automatic selection of E.

tomatically determined by the procedure explained before. The results are shown in Fig. 3, where we analyze three levels of sparsity s for the signal from 1 to 3. For $s = 1$ we obtain a better performance than fixing $E = 500$. For higher sparsity levels $s = 2$ and $s = 3$ we lose a little performance but we gain independence from choosing the wrong parameter E, which is the goal of the modification. On the other hand, the cost of this independence translates into a higher number of measurements but still smaller than the original.

We also analyzed the behavior of the optimal projection matrix in combination to other sparse algorithms rather than ESBL. Fig. 4 shows that the proposed projection procedure improves the behavior of all algorithms and that the best performance is achieved by applying the optimal projection matrix to the ESBL algorithm. Moreover, as shown in Fig. 5, this combination obtains the largest gain over the uncompressed case. This is the

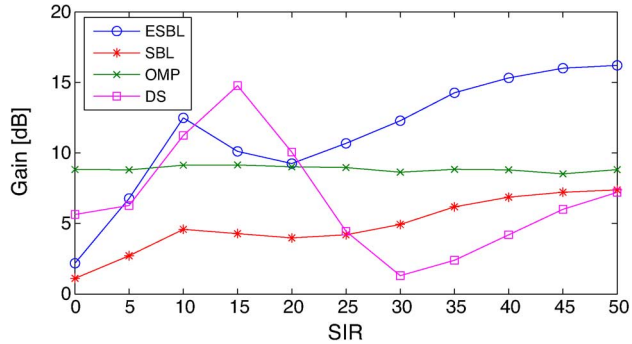

Fig. 5. Gain using the optimal matrix for different algorithms as a function of the SIR when using automatic selection of E.

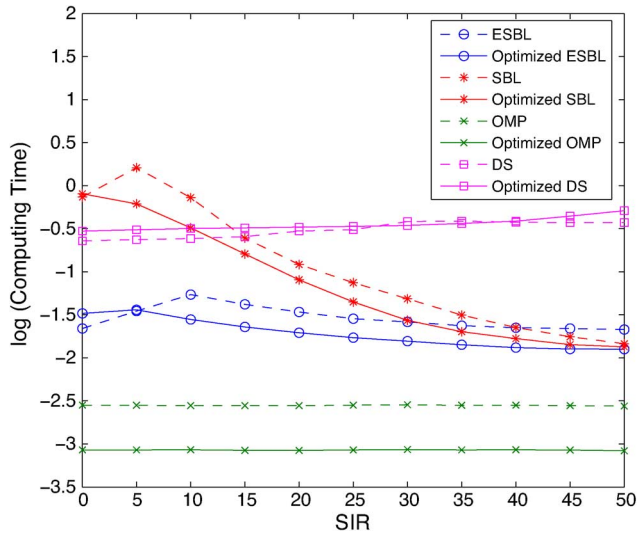

Fig. 6. Average running time for raw and optimized data.

result of being the only algorithm that exploits the interference information in the estimation process.

In Fig. 6 we illustrate the time required by the method, with and without compression, to solve the problem. Since the dimension of the input to the estimation algorithm is reduced, it operates faster resulting in a shorter runtime. Even tough OMP is the fastest algorithm, its performance, shown in Fig. 4, is not acceptable. The overhead of computing the eigenvalues to design the optimal matrix A is negligible.

Fig. 7 shows the estimation error for different values of the signal's sparsity level s. As expected, when the sparsity increases the estimation error is worse. For $s = 1$ we obtain a considerable gain over the uncompressed case but for $s = 9$ the performance is lost.

C. Robustness

In most practical cases the statistical parameters Σ_b, Σ_u, σ, will be unknown and will need to be estimated. This will result in estimation errors affecting the optimality of the method and potentially degrading its performance. We analyze several escenarios of possible sources of discrepancies in the estimation.

First, we analyze the lack of precision in the estimation of the support of Σ_b; that is, it is not as concentrated on the true support as it should be. This imprecise estimation is based on

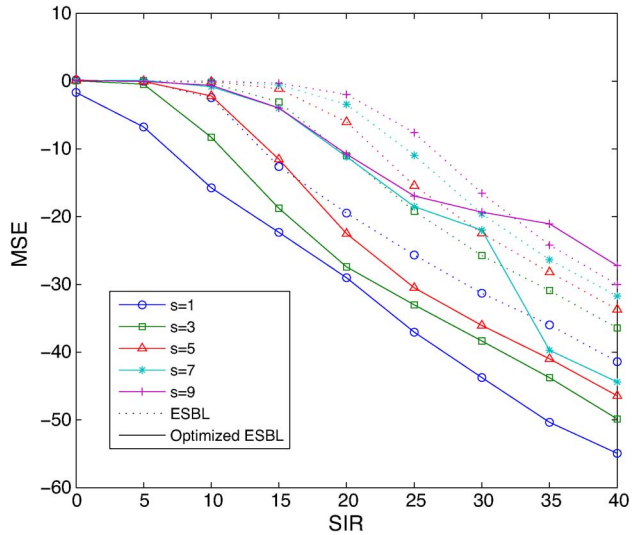

Fig. 7. Estimation relative error, $||\hat{\boldsymbol{b}} - \boldsymbol{b}||_2^2 / ||\hat{\boldsymbol{b}}||_2^2$, as a function of the sparsity s of the original signal.

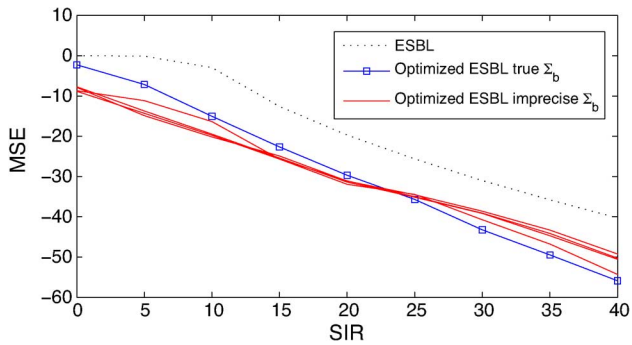

Fig. 8. Estimation relative error as a function of the SIR when the estimation of Σ_b is not precise, with a range of 0.5 to 1.5 for the standard deviation.

a Gaussian distribution with a standard deviation from the true value in the range of 0.5 to 1.5 support elements. The results are shown in Fig. 8, where each solid line corresponds to a different value of the standard deviation. It can be appreciated that we still have a noticeable gain over the uncompressed case for this range of values. There is an unexpected behavior for low SIR, where there seems to be a gain due to the imprecision but we suspect that it should be mostly due to the near optimality of the procedure.

Second, we analyze the case when Σ_b is not accurately estimated, that is the support of Σ_b is not correct. Using a bias from 1 to 6 support elements, the results shown in Fig. 9 indicate that we have a small performance degradation if the support is not correct but it does not strongly depend on the amount of bias. However, we can still obtain a much better performance than that obtained by using no compression.

Third, we explore the lack of precision in the estimation of Σ_u, by adding a scaled random Gaussian matrix to its true value. For each scale factor, ranging from 0.1 to 1, we obtain a different solid line in Fig. 10. The degradation is large for small values of SIR, indicating that a precise estimation of the interference properties is critical in this regime but we still have an acceptable performance at mid-level to high SIR.

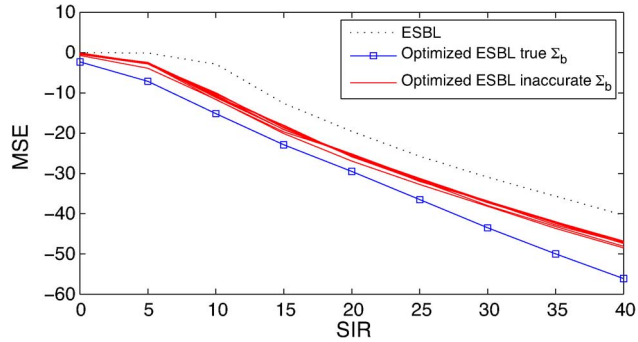

Fig. 9. Estimation relative error as a function of the SIR when the estimation of Σ_b is not accurate, with a range of 1 to 6 for the bias.

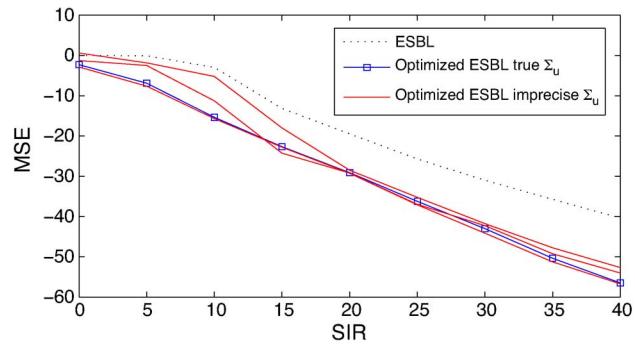

Fig. 10. Estimation relative error as a function of the SIR when the estimation of Σ_u is not precise, with a range of 0.1 to 1 for the standard deviation.

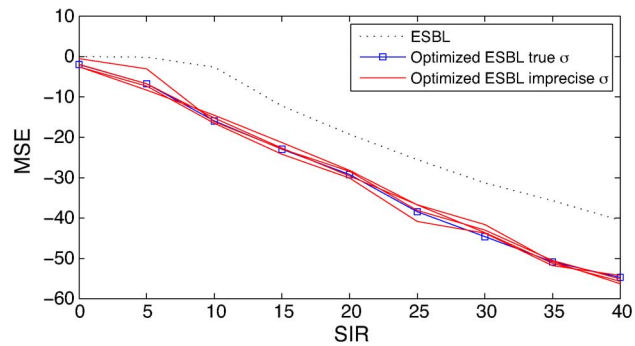

Fig. 11. Estimation relative error as a function of the SIR when the estimation of σ is not precise, with a range of 0.5 to 5 for the standard deviation.

Finally, we show in Fig. 11 that errors in the estimation of σ have practically no effect on the performance since the noise level is insignificant against the interference level in the proposed scenarios.

V. Real Radar Data

A. Fixed Target

We evaluate the optimized method using radar data collected by the McMaster University IPIX polarimetric radar [37]. We processed the dataset stare1 recorded on Nov. 11, 1993. The data corresponds to a beachball wrapped with aluminum foil floating on the sea surface in mild weather (wave height was 0.67 m and wind speed was 21 km/h). A detailed account of casting the

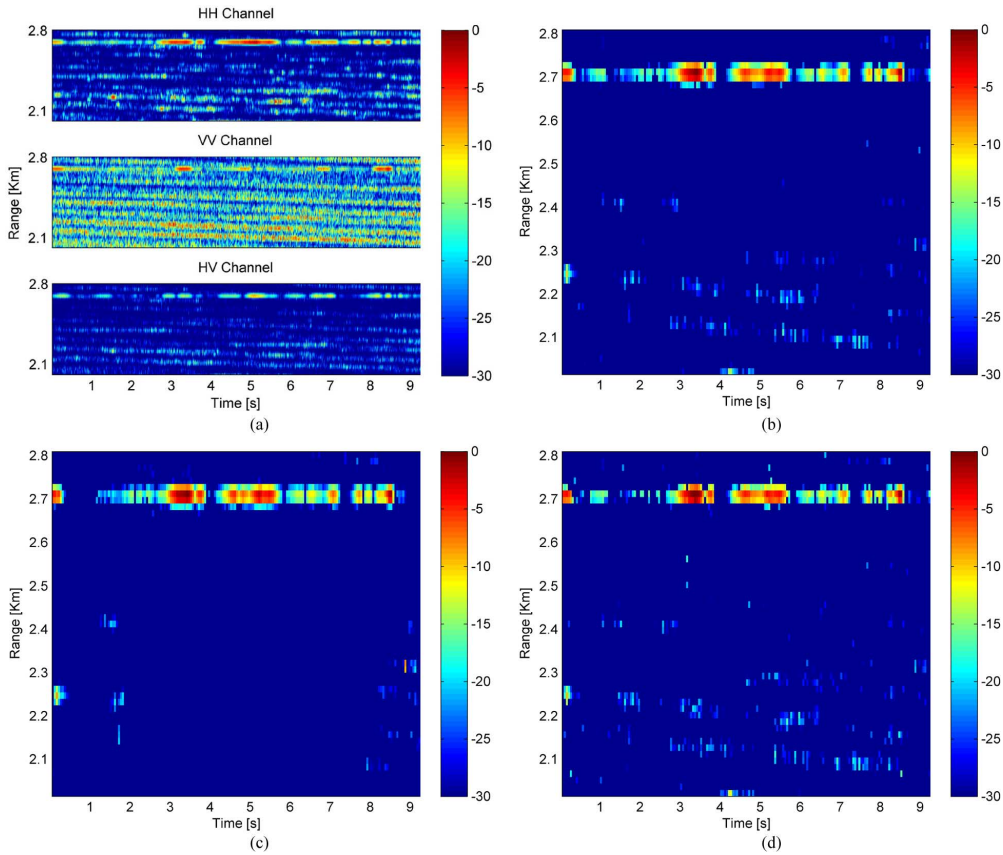

Fig. 12. Radar images in range and time domain for (a) Original raw data from different polarimetric channels, (b) reconstructed data with no compression, (c) reconstructed data using optimal compression and (d) reconstructed data using Duarte's Algorithm, for a still target.

radar model into (1), including the design of the matrices X and Z, is contained in [38].

We generated the overcomplete dictionary by allowing the presence of a target in each of the 54 range cells which form the radar footprint. Each target is represented using Krogager's decomposition of the scattering matrix [39]. For each range cell, we consider nine components: a sphere, left and right helix, and six diplanes with different orientations. This configuration results in a signal vector of length $N = 486$. On the other hand, we have four polarizations for each range cell resulting in $M = 216$ measurements. The initial value of the clutter covariance Σ_u and noise variance σ are estimated from the first snapshot using the concentrated MLE [40] and the signal covariance Σ_b is set to the identity matrix. After this first set, the previous outputs of the ESBL algorithm are used as estimates in the optimization procedure for the next snapshots.

We first show in Fig. 12(a) the energy for different polarimetric channels of the raw radar data in a logarithmic scale. It is possible to locate the target at approximately 2.7 km, which corresponds to range cell 47; additionally, this figure shows that the maritime clutter generates reflections almost as strong as the target, specially in the VV channel. In Fig. 12(b) we show the radar image reconstructed from the same dataset using the ESBL algorithm with a probability of false alarm of $P_{FA} = 10^{-5}$ and no projection matrix applied. The target response remains unchanged; the clutter amplitude is now 20 dB weaker than the target. However, residual clutter is still present after the processing. In Fig. 12(c) we apply the optimal projection proposed, resulting in an even better rejection of the clutter, most of it is now at least 30 dB weaker than the target, as suggested by the previous performance simulations. We also processed the real data using the algorithms previously mentioned in the simulations by Lu [24], Carson [25] and Duarte [17]. The first two failed to correctly estimate the target in most snapshots, despite the use of interference information by Carson's algorithm, and are not included. The best performance of the alternatives is achieved by Duarte's algorithm, shown in Fig. 12(d). Since this algorithm does not exploit prior interference information, it cannot efficiently reject clutter and it also degrades the target signature when the signal is weak.

As mentioned before, we start with $M = 216$ measurements. The procedure takes a set of 10 snapshots, computes the optimal projection matrix and applies it to these measurements. For this specific dataset, the size of the inverse problem is then reduced from 216 to somewhere in the range of 20 to 120 compressed measurements as shown in Fig. 13. Notice that we obtain a better compression and clutter rejection when the target signal is stronger. This reduction translates into a much smaller computing time of the ESBL algorithm. Fig. 14 shows that applying the optimal projection matrix reduces to 1/3 the computing time in this case.

B. *Moving Synthetic Target*

We analyze the performance of the algorithm when dynamics are present in the scenario contrary to the previous case of a static target. Due to the unavailability of real data of moving

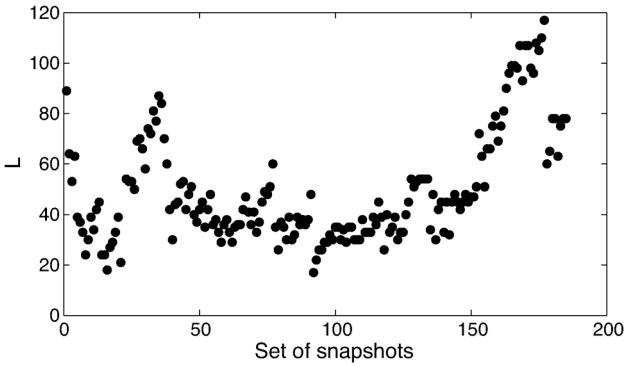

Fig. 13. Mean number of measurements after projection for each set of snapshots.

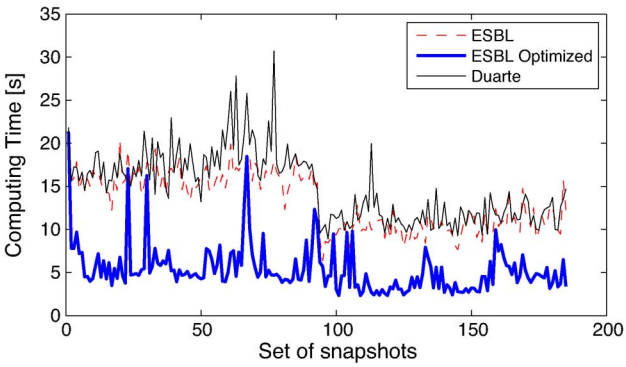

Fig. 14. Comparison of the Computing Time for real radar data processing.

targets we simulate their behavior and add a moving synthetic target over real data of sea clutter from the same IPIX polarimetric radar. The dataset used was stare4 recorded on Nov. 12, 1993, consisting of only sea clutter data and no real target as in the previous section. The target added has a scattering matrix corresponding to a diplane on the VV polarimetric channel; moves at 22.5 m/s away from the radar during the 10 seconds of observation time; and has an approximate power of 10 dB over the clutter. Fig. 15(a) shows the polarimetric channels for the raw data in this scenario. Fig. 15(a) shows the reconstructed image after applying the ESBL algorithm and no compression revealing noticeable clutter in some regions. The proposed algorithm provides better rejection of the undesired clutter in the scene, as shown in Fig. 15(c), even though the target is moving and therefore the estimations of the covariances are not precise. Duarte's algorithm performs well but still presents high clutter levels in some regions, as shown in Fig. 15(d) and uses larger computing times.

VI. Conclusions

In this paper, we introduced a method to optimally compress the measurements from a sparse linear model with structured noise. Compression is achieved by projecting the contaminated signal on a lower dimensional space. In order to design the projection matrix, we proposed to minimize the trace of the conditional covariance of the sparse signal conditioned to the compressed data. This method resulted in an inverse water-filling procedure for building the projection matrix. The algorithm presented outperforms in simulations other current methods for designing the projection matrix.

To demonstrate the performance of the algorithm in a real scenario we processed real polarimetric radar data that correspond to a target in the presence of strong sea clutter. In this case the algorithm proved to have a better rejection of clutter compared to the non optimized case, despite using fewer measurements.

Appendix
Proof of Optimal λ_i

To solve the problem (16), we introduce the Lagrangian

$$\mathcal{L} = \operatorname{tr}\left(\Sigma_{b/y}\right) + \alpha\left(E - \sum_{k=1}^{K}\lambda_k\right) + \sum_{k=1}^{K}\nu_k\lambda_k. \tag{A.1}$$

Using the Karush-Kuhn-Tucker conditions [28], at the optimal point $(\boldsymbol{\lambda}^*, \alpha^*, \boldsymbol{\nu}^*)$ we have, for $i = 1, \ldots, K$,

$$\begin{aligned} \nabla\mathcal{L} &= \mathbf{0}, \\ \nu_i &\geq 0, \\ \lambda_i &\geq 0, \\ \nu_i\lambda_i &= 0, \\ E - \sum_{k=1}^{M}\lambda_k &= 0. \end{aligned} \tag{A.2}$$

We remove $\boldsymbol{\nu}$ using the first condition on the Lagrangian

$$\nu_i = \alpha - \frac{\partial \operatorname{tr}\left(\Sigma_{b/y}\right)}{\partial\lambda_i}, \tag{A.3}$$

then the optimal solution conditions reduce to

$$\nabla\left(\operatorname{tr}\left(\Sigma_{b/y}\right) + \alpha\left(E - \sum_{k=1}^{K}\lambda_k\right) + \sum_{k=1}^{K}\nu_k\lambda_k\right) = \mathbf{0}, \tag{A.4}$$

$$\frac{\partial \operatorname{tr}\left(\Sigma_{b/y}\right)}{\partial\lambda_i} + \alpha \leq 0, \tag{A.5}$$

$$\left(\frac{\partial \operatorname{tr}\left(\Sigma_{b/y}\right)}{\partial\lambda_i} + \alpha\right)\lambda_i = 0, \tag{A.6}$$

$$\lambda_i \geq 0, \tag{A.7}$$

$$\sum_{k=1}^{K}\lambda_k - E = 0. \tag{A.8}$$

Then we need to find the partial derivative of the conditional covariance

$$\begin{aligned} \frac{\partial \operatorname{tr}\left(\Sigma_{b/y}\right)}{\partial\lambda_i} &= \frac{\partial}{\partial\lambda_i}\left[\operatorname{tr}(\Sigma_b) - \sum_{k=1}^{K}\frac{\lambda_k}{\sigma + d_k\lambda_k}t_k\right] \\ &= -\frac{t_i\sigma}{(\sigma + d_i\lambda_i)^2}. \end{aligned} \tag{A.9}$$

Now we have two conditions for α. If $\alpha < \frac{t_i}{\sigma}$ then by (A.5) $\lambda_i > 0$ and by (A.6), $\alpha = \frac{t_i\sigma}{(\sigma + d_i\lambda_i)^2}$. Thus

$$\lambda_i = \frac{\sigma}{d_i}\left(\sqrt{\frac{t_i}{\sigma\alpha}} - 1\right). \tag{A.10}$$

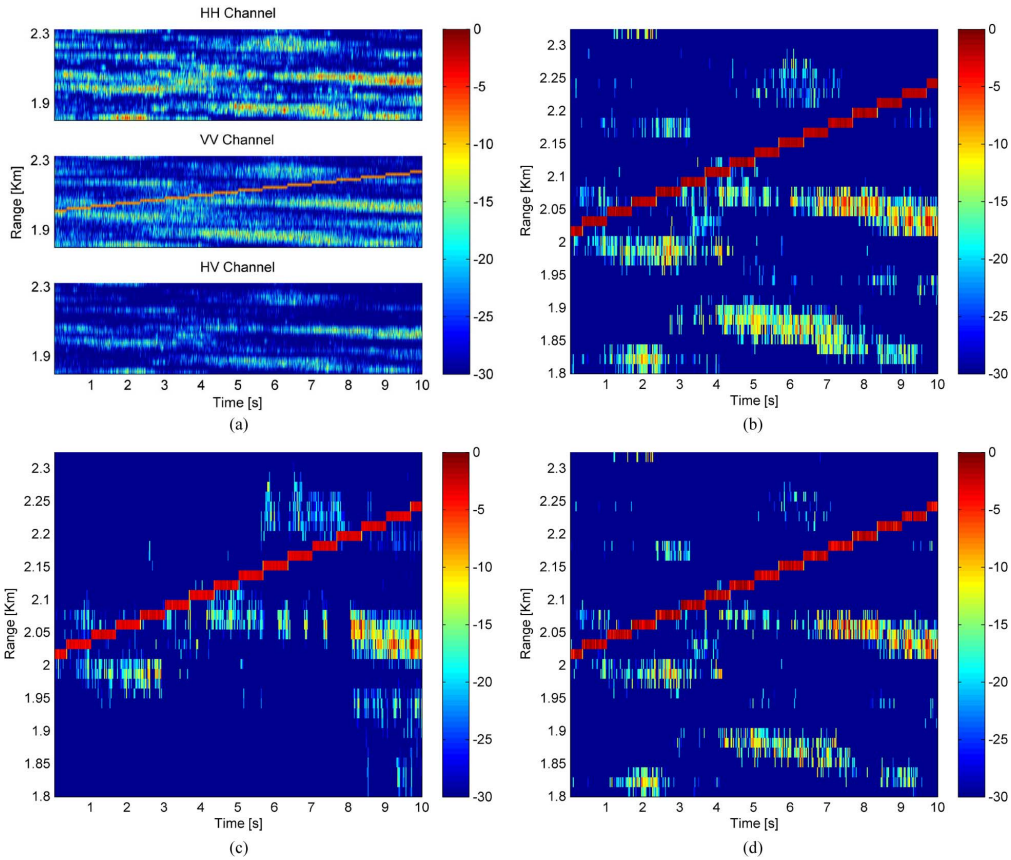

Fig. 15. Radar images in range and time domain for (a) Original raw data from different polarimetric channels, (b) reconstructed data with no compression, (c) reconstructed data using optimal compression and (d) reconstructed data using Duarte's Algorithm, for a moving target.

On the other hand, if $\alpha \geq \frac{t_i}{\sigma}$ then λ_i must be zero. To prove this, assume that λ_i is positive, then

$$\alpha \geq \frac{t_i}{\sigma} > \frac{t_i}{\sigma} \frac{1}{\left(1 + \frac{d_i \lambda_i}{\sigma}\right)^2}, \tag{A.11}$$

and

$$\left(\frac{t_i \sigma}{(\sigma + d_i \lambda_i)^2} - \alpha \right) \lambda_i \neq 0, \tag{A.12}$$

which violates the optimality of the solution in (A.6). Thus, λ_i must be zero. In summary, the optimal eigenvalues for G are

$$\lambda_i = \max \left\{ 0, \frac{\sigma}{d_i} \left(\sqrt{\frac{t_i}{\sigma \alpha}} - 1 \right) \right\}. \tag{A.13}$$

This is an inverse water-filling type of solution where the energy constraint is distributed among the eigenvalues whose corresponding values of t_i are larger than $\sigma\alpha$, as shown in Fig. 16. Assuming that the first L optimal eigenvalues λ_i are nonzero, we use (A.13) in (A.8) and find the Lagrange multiplier

$$\alpha = \frac{\left(\sum_{i=1}^{L} \sqrt{\frac{t_i}{\sigma}} \frac{\sigma}{d_i} \right)^2}{\left(\sum_{i=1}^{L} \frac{\sigma}{d_i} + E \right)^2}. \tag{A.14}$$

In order to find L, first we assume that the t_i are in decreasing order, then we start with $L = K$ and we compute α and λ_L. If

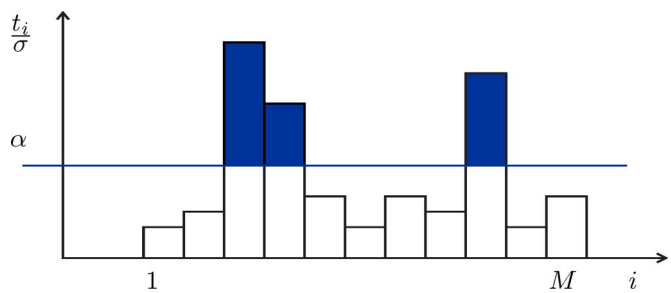

Fig. 16. Inverse water-filling solution for the eigenvalues of H based on the relationship of t_i and α.

λ_L is negative then the optimal value of L was lower and we try with $L = K - 1$. This process is repeated until λ_L is positive. Then the rest of the eigenvalues λ_i can be computed with the right value of α and L, according to (A.13).

REFERENCES

[1] R. Behrens and L. Scharf, "Signal processing applications of oblique projection operators," *IEEE Trans. Signal Process.*, vol. 42, no. 6, pp. 1413–1424, Jun. 1994.

[2] M. Hurtado and A. Nehorai, "Polarimetric detection of targets in heavy inhomogeneous clutter," *IEEE Trans. Signal Process.*, vol. 56, no. 4, pp. 1349–1361, Apr. 2008.

[3] N. von Ellenrieder, M. Hurtado, and C. Muravchik, "Electromagnetic source imaging for sparse cortical activation patterns," in *Proc. IEEE Int. Conf. Eng. Med. Biol. Soc. (EMBC)*, Sep. 2010, pp. 4316–4319.

[4] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[5] S. S. Chen, D. L. Donoho, Michael, and A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 20, pp. 33–61, 1998.
[6] J. Tropp and A. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.
[7] E. Candes and T. Tao, "The Dantzig selector: Statistical estimation when p is much larger than n," *Ann. Statist.*, vol. 35, p. 2313, 2007.
[8] E. J. Candès *et al.*, "Compressive sampling," *Proc. Intern. Congr. Math.*, vol. 3, pp. 1433–1452, 2006.
[9] G. Tang and A. Nehorai, "Performance analysis of sparse recovery based on constrained minimal singular values," *IEEE Trans. Signal Process.*, vol. 59, no. 12, pp. 5734–5745, Dec. 2011.
[10] D. Wipf and S. Nagarajan, "A new view of automatic relevance determination," presented at the Adv. Neural Inf. Process. Syst. 20, 2008.
[11] K. Qiu and A. Dogandžić, "Double overrelaxation thresholding methods for sparse signal reconstruction," presented at the 44rd Annu. Conf. Inform. Sci. Syst., Princeton, NJ, USA, 2010.
[12] P. Schniter, L. Potter, and J. Ziniel, "Fast bayesian matching pursuit," in *Proc. Inf. Theory Appl. Workshop*, Feb. 2008, pp. 326–333.
[13] D. Wipf and B. Rao, "Sparse bayesian learning for basis selection," *IEEE Trans. Signal Process.*, vol. 52, no. 8, pp. 2153–2164, Aug. 2004.
[14] M. Hurtado, C. Muravchik, and A. Nehorai, "Enhanced sparse bayesian learning via statistical thresholding for signals in structured noise," *IEEE Trans. Signal Process.*, vol. 61, no. 21, pp. 5430–5443, Nov. 2013.
[15] M. Elad, "Optimized projections for compressed sensing," *IEEE Trans. Signal Process.*, vol. 55, no. 12, pp. 5695–5702, Dec. 2007.
[16] V. D. M. Nhat, D. Vo, S. Challa, and S. Lee, "Efficient projection for compressed sensing," in *Proc. 7th IEEE/ACIS Int. Conf. Comput. Inf. Sci.*, May 2008, pp. 322–327.
[17] J. Duarte-Carvajalino and G. Sapiro, "Learning to sense sparse signals: Simultaneous sensing matrix and sparsifying dictionary optimization," *IEEE Trans. Image Process.*, vol. 18, no. 7, pp. 1395–1408, Jul. 2009.
[18] A. Tehrani, A. Dimakis, and G. Caire, "Optimal deterministic compressed sensing matrices," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, May 2013, pp. 5895–5899.
[19] A. Amini and F. Marvasti, "Deterministic construction of binary, bipolar, ternary compressed sensing matrices," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 2360–2370, Apr. 2011.
[20] Y. Yu, A. Petropulu, and H. Poor, "Measurement matrix design for compressive sensing-based mimo radar," *IEEE Trans. Signal Process.*, vol. 59, no. 11, pp. 5338–5352, Nov. 2011.
[21] S. Li, Z. Zhu, G. Li, L. Chang, and Q. Li, "Projection matrix optimization for block-sparse compressive sensing," in *Proc. Int. Conf. Signal Process., Commun. Comput. (ICSPCC)*, Aug. 2013, pp. 1–4.
[22] L. Zelnik-Manor, K. Rosenblum, and Y. Eldar, "Sensing matrix optimization for block-sparse decoding," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4300–4312, Sep. 2011.
[23] J. Duarte-Carvajalino, G. Yu, L. Carin, and G. Sapiro, "Adapted statistical compressive sensing: Learning to sense gaussian mixture models," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Mar. 2012, pp. 3653–3656.
[24] Y. Lu, S. Hegler, C. Statz, A. Finger, and D. Plettemeier, "Improving detection performance of compressed sensing by orthogonal projection," in *Proc. Int. Radar Symp.*, Jun. 2013, vol. 1, pp. 107–112.
[25] W. Carson, M. Chen, M. Rodrigues, R. Calderbank, and L. Carin, "Communications-inspired projection design with application to compressive sensing," *SIAM J. Imaging Sci.* vol. 5, no. 4, pp. 1185–1212, 2012.
[26] M. E. Tipping, "Sparse bayesian learning and the relevance vector machine," *J. Mach. Learn. Res.* vol. 1, pp. 211–244, Sep. 2001 [Online]. Available: http://dx.doi.org/10.1162/15324430152748236
[27] J. Chen and X. Huo, "Theoretical results on sparse representations of multiple-measurement vectors," *IEEE Trans. Signal Process.*, vol. 54, no. 12, pp. 4634–4643, Dec. 2006.
[28] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[29] S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*. Englewood Cliffs, NJ, USA: Prentice Hall, 1993.
[30] K. B. Petersen and M. S. Pedersen, *The Matrix Cookbook*. Lyngby, Denmark: Technical University of Denmark, Oct. 2008, version 20081110.
[31] D. Fuhrmann, "One-step optimal measurement selection for linear gaussian estimation problems," in *Proc. Int. Waveform Diversity Design Conf.*, Jun. 2007, pp. 224–227.
[32] K. L. Bell, Y. Ephraim, and H. L. Van Trees, "A Bayesian approach to robust adaptive beamforming," *IEEE Trans. Signal Process.*, vol. 48, no. 2, pp. 386–398, Feb. 2000.
[33] S. Kay, "Optimal signal design for detection of Gaussian point targets in stationary gaussian clutter/reverberation," *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 1, pp. 31–41, Jun. 2007.
[34] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, ser. Springer Series in Statistics. New York, NY, USA: Springer-Verlag, 2001.
[35] B. Rao and K. Kreutz-Delgado, "An affine scaling methodology for best basis selection," *IEEE Trans. Signal Process.*, vol. 47, no. 1, pp. 187–200, Jan. 1999.
[36] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ_1 minimization," *Proc. Nat. Acad. Sci.*, vol. 100, no. 5, pp. 2197–2202, Mar. 2003.
[37] S. Haykin, C. Krasnor, T. J. Nohara, B. W. Currie, and D. Hamburger, "A coherent dual-polarized radar for studying the ocean environment," *IEEE Trans. Geosci. Remote Sens.*, vol. 29, no. 1, pp. 189–191, Jan. 1991.
[38] M. Hurtado, N. von Ellenrieder, C. Muravchik, and A. Nehorai, "A sparse modeling for polarimetric radar," presented at the IEEE Workshop Stat. Signal Process., 2011.
[39] E. Krogager, "New decomposition of the radar target scattering matrix," *Electron. Lett.*, vol. 26, pp. 1525–1527, 1990.
[40] P. Stoica and A. Nehorai, "On the concentrated stochastic likelihood function in array signal processing," *Circuits, Syst., Signal Process.*, vol. 14, no. 5, pp. 669–674, 1995.

Sebastian Pazos (S'00–A'01–M'08) graduated as an Electrical Engineer from the Buenos Aires Institute of Technology (ITBA), Argentina, in 2001. He received his Ph.D. degree in electrical engineering at National University of La Plata and the M.Sc. degree in electrical engineering (2006) and the M.Sc. degree in Mathematics (2008) from the University of Illinois at Urbana-Champaign, IL.

He is a Professor at the Department of Basic Sciences of the National University of La Plata and a member of its Industrial Electronics, Control and Instrumentation Laboratory (LEICI). He is the recipient of a Fulbright scholarship. His research interests are in the area of statistical and array signal processing applied to radar.

Martin Hurtado (S'95–A'96–M'05) received the B.Eng. and M.Sc. degrees in electrical engineering from the National University of La Plata, Argentina, in 1996 and 2001, respectively. He received his Ph.D. degree in electrical engineering from Washington University in St. Louis in 2007.

Currently, he is a research associate of the National Council of Scientific and Technical Research of Argentina and an adjunct professor in the Department of Electrical Engineering at National University of La Plata. His research interests are in the area of statistical signal processing, detection and estimation theory, and their applications in sensor arrays, communications, and remote sensing systems.

Carlos H. Muravchik (S'81–M'83–SM'99) graduated as an Electronics Engineer from the National University of La Plata, Argentina, in 1973. He received the M.Sc. in Statistics (1983) and the M.Sc. (1980) and Ph.D. (1983) degrees in Electrical Engineering, all from Stanford University, Stanford, CA.

He is a Professor at the Department of Electrical Engineering in the National University of La Plata (Argentina); the past director of its Industrial Electronics, Control and Instrumentation Laboratory (LEICI) and a member of the Comision de Investigaciones Cientificas de la Pcia. de Buenos Aires. He was a Visiting Professor to Yale University in 1983 and 1994, to the University of Illinois at Chicago in

1996, 1997, 1999 and 2003 and to Washington University in St Louis in 2006 and 2010. Since 1999 he is a member of the Advisory Board of the journal Latin American Applied Research and was an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING (2003–2006). His research interests are in the area of statistical and array signal processing with biomedical, communications and control applications, and in nonlinear systems.

Arye Nehorai (S'80–M'83–SM'90–F'94) is the Eugene and Martha Lohman Professor and Chair of the Preston M. Green Department of Electrical and Systems Engineering (ESE) at Washington University in St. Louis (WUSTL). He is also Professor in the Division of Biology and Biomedical Studies (DBBS) and Director of the Center for Sensor Signal and Information Processing at WUSTL. Earlier, he was a faculty member at Yale University and the University of Illinois at Chicago. He received the B.Sc. and M.Sc. degrees from the Technion, Israel and the Ph.D. from Stanford University, California. Under his leadership as department chair, the undergraduate enrollment has more than tripled in the last four years.

Dr. Nehorai served as Editor-in-Chief of the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 2000 to 2002. From 2003 to 2005 he was the Vice President (Publications) of the IEEE Signal Processing Society (SPS), the Chair of the Publications Board, and a member of the Executive Committee of this Society. He was the founding editor of the special columns on Leadership Reflections in *IEEE Signal Processing Magazine* from 2003 to 2006.

Dr. Nehorai received the 2006 IEEE SPS Technical Achievement Award and the 2010 IEEE SPS Meritorious Service Award. He was elected Distinguished Lecturer of the IEEE SPS for a term lasting from 2004 to 2005. He received best paper awards in IEEE journals and conferences. In 2001 he was named University Scholar of the University of Illinois. Dr. Nehorai was the Principal Investigator of the Multidisciplinary University Research Initiative (MURI) project titled Adaptive Waveform Diversity for Full Spectral Dominance from 2005 to 2010. He is a Fellow of the IEEE since 1994, Fellow of the Royal Statistical Society since 1996, and Fellow of AAAS since 2012.