OMP-type Algorithm with Structured Sparsity Patterns for Multipath Radar Signals

Tabea Rebafka, Céline Lévy-Leduc, Maurice Charbit

OMP-type Algorithm with Structured Sparsity Patterns for Multipath Radar Signals

Tabea Rebafka, Céline Lévy-Leduc and Maurice Charbit, *Member, IEEE.*

Abstract

A transmitted, unknown radar signal is observed at the receiver through more than one path in additive noise. The aim is to recover the waveform of the intercepted signal and to simultaneously estimate the direction of arrival (DOA). We propose an approach exploiting the parsimonious time-frequency representation of the signal by applying a new OMP-type algorithm for structured sparsity patterns. An important issue is the scalability of the proposed algorithm since high-dimensional models shall be used for radar signals. Monte-Carlo simulations for modulated signals illustrate the good performance of the method even for low signal-to-noise ratios and a gain of 20 dB for the DOA estimation compared to some elementary method.

Index Terms

DOA estimation, waveform recovery, multipath propagation, structured sparsity, OMP-type algorithm.

I. Introduction

The aim of SIGnals INTelligence (SIGINT) is to intercept as much information as possible on received signals. For radar sources, the parameters of interest are the Direction Of Arrival (DOA) of the direct path (assuming it is observed), but also, if possible, the carrier frequency and the modulation scheme. In practice, due to the presence of multipath propagation, the interception of radar sources remains a difficult problem. Indeed, multipath propagation creates nuisance parameters that have to be taken into

T. Rebafka is with the LPMA at UPMC, University Paris 6, France. e-mail: tabea.rebafka@upmc.fr.

C. Lévy-Leduc and M. Charbit are with the CNRS/LTCI of Telecom ParisTech, France. e-mail: {celine.levy-leduc, maurice.charbit}@telecom-paristech.fr.

This work was financially supported by the French National Agency: Direction Générale de l'Armement (DGA).

account such as DOAs and relative Times of Arrival (TOA) of reflected paths, relative signal power levels on the different paths and additive noise variance.

In this paper, we shall consider a unique narrowband farfield source propagating through several paths, corrupted with additive Gaussian white noise and observed on a multi-sensor array. This issue is actually very involved since the observed signals on the different sensors are highly correlated and the different replicates are likely to overlap. This prevents us from using methods such as MUSIC, ESPRIT or MVDR which do not work very well with coherent signals [1]. These approaches, based on subspace methods or on maximum likelihood estimation, have been extended by [2], [3], [4] and [5] to deal with coherent sources. However, to the best of our knowledge, estimating the waveform, which is an infinite-dimensional parameter, in a nonparametric way has never been considered for multipath signals. To solve this problem, our main idea consists in exploiting the sparsity of the time-frequency representation of the waveform, which means that the selection of possibly useful signal components from a huge collection of candidate signals may be driven by a sparsity constraint. Such kind of approach was initially proposed by [6] but only for the estimation of the time-delays, which is a finite dimensional parameter, and is extended here for estimating the DOAs and TOAs as well as the waveform.

For dealing with the estimation in sparse linear regression models, the Lasso [7] and the greedy orthogonal matching pursuit (OMP) ([8], [9]) have become very popular tools. In our case, an inspection of our model shows that the parameter vector is not sparse in an arbitrary way: its sparsity pattern has a specific structure. In order to include prior information concerning the sparsity structure, different approaches have recently been proposed. On the one hand, there are methods based on composed ℓ_1/ℓ_2-penalties (elastic nets [10]; fused Lasso [11]; group Lasso [12]; composite absolute penalty [13]; overlapping groups [14], [15]), on the other hand, [16] proposed the group-OMP method for non-overlapping groups, which is an extension of the orthogonal matching pursuit to structured solutions.

In this paper we explain how to extend the group-OMP to overlapping groups and how to deal with the scalability of our algorithm which is a crucial issue in the context of intercepted radar signals since the model dimension is usually very high. A simulation study shows that the proposed method performs well for composed and modulated signals even when the signal-to-noise ratio is low. A gain of more than 20 dB for the DOA estimation is achieved in comparison with some elementary method.

The paper is organized as follows. In Section II the mathematical model for radar signals is introduced, together with a reformulation of the problem in the form of a sparse, partly linear model. In Section III an OMP-type algorithm is developed which is well adapted to our model. Section IV deals with the scalability issue of the algorithm, while Section V presents an appropriate stopping criterion for the

OMP-type algorithm. Finally, the experimental results of Section VI illustrate the performance of the new algorithm and a comparison with some elementary method is provided. Conclusions are made in the final Section VII.

II. Model

A. Signal Modelling

Let us denote by $s_0(t)$ the original radar signal emitted by the source, where t is the time. Due to propagation and reflexion on obstacles, the signal arrives via several pathways on the receiver. With each pathway we associate a direction of arrival, say D_u, a time-delay, say T_u and a complex constant A_u representing the attenuation of the signal on the u-th pathway. The signal is detected by an antenna with C sensors whose array response is denoted by $\vec{a}(\cdot)$. Then the signal arriving at the sensor array at time t is a C-dimensional vector given by

$$\vec{s}(t) = \sum_{u=1}^{U} A_u \, s_0(t - T_u) \, \vec{a}(D_u) \ , \tag{1}$$

where U denotes the number of propagation pathways, which is unknown. As the time lags T_u are typically much shorter than the total length of the signal s_0, the arriving signal $\vec{s}$ is the superposition of several delayed replicates s_0 with different amplitudes.

In practice, the signal $\vec{s}$ is observed with some additional noise, say $\vec{\varepsilon}$, and at a sample frequency, say F_s. With $\Delta_s = 1/F_s$ the observations are hence given by

$$\vec{y}_m = \vec{s}(m\Delta_s) + \vec{\varepsilon}_m \ , \qquad m = 1, \ldots, M \ , \tag{2}$$

where $\vec{\varepsilon}_m \in \mathbb{C}^C$ is a noise vector. Generally one considers circular gaussian independent random vectors with zero mean and covariance $\sigma^2 I_C$.

B. Linear Regression Model

In practice, all model parameters are unknown, except the array response function $\vec{a}(\cdot)$, which is determined by the array geometry or by calibration. As the relationship between the parameters is rather involved, we propose to reformulate the estimation problem by linearizing the model. This is obtained, on the one hand, by representing the waveform in an overcomplete basis and, on the other hand, by discretizing the parameter spaces of the other parameters. This leads to a linear model with a high-dimensional parameter vector and nonlinear constraints. As this vector is structured and sparse, *i.e.* most entries are zero, regularization methods with structured sparsity inducing norms can be applied [13] .

1) Decomposition of the signal : To estimate the waveform, the radar signal s_0 is represented in some given (overcomplete) basis or dictionary $\mathcal{D} = \{\varphi_j, j = 1, \dots, J\}$, such that

$$s_0 = \sum_{j=1}^{J} \beta_j \varphi_j \ , \tag{3}$$

with appropriate coefficients $\beta_j \in \mathbb{C}$. Hence, estimating s_0 becomes recovering the coefficients β_j. When using appropriate, large dictionaries, a small number of elements φ_j is sufficient to reconstruct s_0. In other words, it is assumed that the dictionary is such that most coefficients β_j are zero and hence that the representation of s_0 is sparse.

2) Propagation model: To estimate the parameters A_u, D_u and T_u by discretization, we introduce grids $\{\tau_1, \dots, \tau_P\}$ and $\{\theta_1, \dots, \theta_Q\}$ of potential values of the delay times T_u and the angles of arrival D_u respectively, as in [6], [17]. Then $\vec{s}(t)$ can be rewritten as

$$\vec{s}(t) = \sum_{p=1}^{P} \sum_{q=1}^{Q} \alpha_{p,q} \ s_0(t - \tau_p) \, \vec{a}(\theta_q) \ , \tag{4}$$

$$\text{with} \quad \alpha_{p,q} = \sum_{u=1}^{U} A_u \mathbb{1}\{T_u = \tau_p, D_u = \theta_q\} \ ,$$

where $\mathbb{1}$ is the indicator function. Clearly, there are only U coefficients $\alpha_{p,q}$ that are nonzero. Hence, when fine grids are used, *i.e.* P and Q are large, this is another sparse estimation problem. As the number of pathways U is generally unknown, model selection is required, *i.e.* the estimation of the model order U.

By combining (3) and (4), the observation $\vec{y}_m$ reads

$$\vec{y}_m = \sum_{j=1}^{J} \sum_{p=1}^{P} \sum_{q=1}^{Q} \alpha_{p,q} \beta_j \varphi_j(m\Delta_s - \tau_p) \, \vec{a}(\theta_q) + \vec{\varepsilon}_m \ , \tag{5}$$

for $m = 1, \dots, M$. It is convenient to represent the model in matrix notation. Let $\mathbf{X}$ be the matrix of size $CM \times JPQ$, where all known quantities are stored. More precisely, the (j, p, q)-th predictor, *i.e.* the column $x_{j,p,q}$ of $\mathbf{X}$, is given by

$$x_{j,p,q} = \begin{pmatrix} \varphi_j(\Delta_s - \tau_p) \, \vec{a}(\theta_q) \\ \vdots \\ \varphi_j(M\Delta_s - \tau_p) \, \vec{a}(\theta_q) \end{pmatrix} \in \mathbb{C}^{CM} \ .$$

In other words, $x_{j,p,q}$ corresponds to the signal received by the antenna (without noise) when the emitted signal is the element φ_j arriving with delay τ_p and angle θ_q. Denote the observation vector by $Y = (y_1^T, \dots, y_m^T)^T \in \mathbb{C}^{CM}$ and the noise vector by $\varepsilon = (\vec{\varepsilon}_1^T, \dots, \vec{\varepsilon}_m^T)^T \in \mathbb{C}^{CM}$. Finally, let the Kronecker

product $\beta \otimes \alpha$ of the vectors $\beta = (\beta_1, \ldots, \beta_J)^T$ and $\alpha = (\alpha_{1,1}, \ldots, \alpha_{P,Q})^T$ be the JPQ-vector w defined as

$$w = \begin{pmatrix} \vdots \\ w_{j,p,q} \\ \vdots \end{pmatrix} = \begin{pmatrix} \beta_1\alpha_{1,1} \\ \beta_1\alpha_{1,2} \\ \vdots \\ \beta_j\alpha_{p,q} \\ \vdots \\ \beta_J\alpha_{P,Q} \end{pmatrix} . \tag{6}$$

Then model (5) can be written as

$$Y = \mathbf{X}(\beta \otimes \alpha) + \varepsilon \quad \text{ with } \alpha \in \mathbb{C}^{PQ}, \beta \in \mathbb{C}^J \ . \tag{7}$$

Finally, introduce the following set

$$\mathcal{W} = \{w \in \mathbb{C}^{JPQ} : w = \beta \otimes \alpha,\ \alpha \in \mathbb{C}^{PQ}, \beta \in \mathbb{C}^J\} \ ,$$

which is a subspace of $\mathbb{C}^{JPQ}$ that can be parametrized using $J + PQ$ variables. Remark that, given a vector $w \in \mathcal{W}$, one can recover the vectors α and β only up to a multiplicative constant, which does not matter in our application. Thus, model (5) can be written as the following linear regression model

$$Y = \mathbf{X}w + \varepsilon \quad \text{ with } w \in \mathcal{W}, \tag{8}$$

where the design matrix $\mathbf{X}$ is known and the parameter vector w is to be estimated from the observations Y. The particularity of this regression model is that the search space for w is not the entire space $\mathbb{C}^{JPQ}$ but the smaller, nonconvex set $\mathcal{W}$.

III. Estimation procedure

The problem now is to estimate the vectors α and β by fitting model (7), or equivalently, to estimate w in model (8). The advantage of (8) over (7) is the linearity of the model. Since the dimension of the space $\mathcal{W}$ is much higher than that of the observations Y, the minimization problem

$$\min_{w \in \mathcal{W}} \|Y - \mathbf{X}w\|^2 \tag{9}$$

is ill-posed. However, we note that w must be a sparse vector, since α and β are assumed sparse. It is well known that sparsity leads to stable estimation procedures based on a regularized fitting criterion with an ℓ_1-penalty (Lasso [7]; basis pursuit [18]; Lars [19]) or greedy algorithms (as matching pursuit and orthogonal matching pursuit (OMP) [8], [20]). A closer inspection of $\mathcal{W}$ reveals some structure in the

sparsity patterns, that means, there are constraints on the distribution of the zero entries for all vectors w in $\mathcal{W}$. More precisely, the set of indices $\mathcal{S} = \{(j,p,q) : \text{ such that } w_{j,p,q} = 0\}$ is called the sparsity pattern of some given vector w. Indeed, all elements w of $\mathcal{W}$ verify that $\alpha_{p',q'} = 0$ implies that the components $w_{j,p',q'} = 0$ for all j, and likewise, if $\beta_{j'} = 0$ then $w_{j',p,q} = 0$ for all p and q. Thus, w must not be sparse in an arbitrary way, but its sparsity pattern has to respect some structure. Modern estimation procedures force specific structures of the sparse vector. On the one hand, there are methods based on composed ℓ_1/ℓ_2-penalties (elastic nets [10]; fused Lasso [11]; group Lasso [12]; composite absolute penalty [13]; overlapping groups [14], [15]), on the other hand, the orthogonal matching pursuit can be extended to structured solutions (group-OMP for non-overlapping groups [16]).

Our problem can be related to a penalized minimization problem of the form

$$\min_{w \in \mathbb{C}^{JPQ}} \left\{ \|Y - \mathbf{X}w\|^2 + \Omega(w) \right\} , \tag{10}$$

with a structured ℓ_1/ℓ_2-penalty $\Omega(w)$ (detailed below) in the sense that the sparsity pattern of a solution of this problem, say $\tilde{w}$, respects the required structure. However, there is no guarantee that the nonzero components of $\tilde{w}$ satisfy relation (6), that is, that there are coefficients $\alpha_{p,q}$ and β_j such that $\tilde{w}_{j,p,q} = \alpha_{p,q}\beta_j$ and hence $\tilde{w}$ may not belong to $\mathcal{W}$. The problem lies in the nonzero components of $\tilde{w}$. To compute the solution of (10), different solutions have been proposed in the literature as the active set algorithm in in [21] and [15]. In this paper we propose an extension of the group-OMP [16] for overlapping groups. Now, to fit the model given in (7) we propose the following procedure consisting of two steps, that are developed below in more detail.

Step 1. (Model selection) Solve problem (10) via an OMP-type algorithm to obtain a solution $\tilde{w}$ with admissible sparsity structure. Identify the nonzero components $\tilde{w}_{j,p,q}$ and the indices $\mathcal{A}$ of the associated coefficients $\tilde{\alpha}_{p,q}$ and $\tilde{\beta}_j$ that must be nonzero.

Step 2. (Estimation in reduced dimension) Reduce the dimension of the regression matrix $\mathbf{X}$ by keeping the predictor $x_{j,p,q}$ only if $\tilde{w}_{j,p,q} \neq 0$. Then compute the least squares estimator in model (7) by the Nelder-Mead simplex method. This problem is now well-posed because of the largely reduced dimension.

A. Model selection step

To specify the penalty $\Omega(w)$ in (10), we describe the structure of the sparsity pattern $\mathcal{S}$ of $w \in \mathcal{W}$. To this end, for a given couple (p,q), we denote the set of indices $G^{\alpha}_{p,q} = \{(j,p,q),\ j \in \{1,\dots,J\}\}$, and likewise, for a given j, the set of indices $G^{\beta}_{j} = \{(j,p,q),\ p \in \{1,\dots,P\}, q \in \{1,\dots,Q\}\}$. It is

clear that the sparsity pattern $\mathcal{S}$ of any w in $\mathcal{W}$ can be written as the union of all sets $G^{\alpha}_{p,q}$ and G^{β}_{j} with $\alpha_{p,q} = 0$ and $\beta_j = 0$.

Given a set of indices G, let $\|w_G\|_2$ denote the ℓ_2-norm of the subvector of w composed of the elements $w_{j,p,q}$ with $(j,p,q) \in G$. Then one can show that the solution, say $\tilde{w}$, of (10) with the following penalty

$$\Omega(w) = \lambda_1 \sum_{p,q} \|w_{G^{\alpha}_{p,q}}\|_2 + \lambda_2 \sum_{j} \|w_{G^{\beta}_{j}}\|_2 \,, \tag{11}$$

where $\lambda_1, \lambda_2 > 0$ are regularization parameters, has the required sparsity structure, *i.e.* the sparsity pattern of $\tilde{w}$ can be written as the union of sets $G^{\alpha}_{p,q}$ and G^{β}_{j}. Note that these groups are possibly overlapping. Indeed, the penalty can be viewed as a mixture of ℓ_1- and ℓ_2-norms appropriate for overlapping group structures. More precisely, the penalty is the sum (*i.e.* the ℓ_1-norm) of the ℓ_2-norm of subvectors w_G. Likewise to the standard Lasso, the ℓ_1-norm entails that for several groups G the terms $\|w_G\|_2$ are zero, implying in turn that all entries of w_G are exactly zero. By using two regularization parameters $\lambda_1 > 0$ and $\lambda_2 > 0$ one can use different degrees of sparsity for the vectors α and β.

To solve the problem (10) with penalty $\Omega(w)$ given by (11) we propose an OMP-type algorithm. To describe the algorithm we introduce the collection $\mathcal{G} = \{G_g, g \in \mathcal{I}\}$ of all sets $G^{\alpha}_{p,q}$ and G^{β}_{j} indexed by a common index g with index set $\mathcal{I}$. Note that the cardinality of $\mathcal{I}$ is $PQ + J$. For a set of indices A, denote by $\mathbf{X}_A$ the matrix made of the corresponding predictors $x_{j,p,q}$ of $\mathbf{X}$ such that $(j,p,q) \in A$.

Recall that the principle of OMP consists in adding iteratively the predictor to the current solution $w^{(t)}$ which is the most correlated with the current residual $r^{(t)} = Y - \mathbf{X}w^{(t)}$. In [16] this principle is extended to selecting non-overlapping groups of variables, where $\mathcal{G}$ is a set of pairwise disjoint sets G_g and one adds all variables of group G_{g^*} if $\|\mathbf{X}^T_{G_{g^*}} r^{(t)}\|_2^2 = \max_{g \in \mathcal{I}} \|\mathbf{X}^T_{G_g} r^{(t)}\|_2^2$. Proceeding in this way guarantees that at every step of the algorithm, the current solution respects the required sparsity structure, which means that the sparsity pattern of the current solution $w^{(t)}$ equals the union of some groups $G_g \in \mathcal{G}$.

Now this procedure can be extended to our context with overlapping group structures, where the aim remains the same: the sparsity pattern of every intermediate solution $w^{(t)}$ must be the union of some groups $G_g \in \mathcal{G}$. The difference to the group-OMP of [16] is at the level of the selection of variables that may enter the solution. Indeed, due to the overlapping group structure, the sets of variables that may be activated at the next step are not simply the groups $G_g \in \mathcal{G}$. To be more precise on this issue, we introduce the set $\mathcal{Z}^{(t)} \subset \mathcal{I}$ of indices of zero groups G_g associated with the current solution $w^{(t)}$, such that $\mathcal{S}^{(t)} = \cup_{g \in \mathcal{Z}^{(t)}} G_g$ is the sparsity pattern of $w^{(t)}$. Clearly, the indices of the non zero entries of $w^{(t)}$ are given by $\mathcal{A}^{(t)} = (\mathcal{S}^{(t)})^c$, where A^c denotes the complementary set of A with respect to $\mathcal{G}$. Now the candidate groups of components that may be activated at the next step are obtained by deleting

one of the indices from the zero groups $\mathcal{Z}^{(t)}$. That means, every potential group of indices has the form $\mathcal{P}_{g_0} = (\cup_{g\in\mathcal{Z}^{(t)}\setminus\{g_0\}} G_g)^c \setminus \mathcal{A}^{(t)}$ where $g_0 \in \mathcal{Z}^{(t)}$. Then, following the philosophy of OMP, one activates the components given by $\mathcal{P}_{g^*}$ if $\|\mathbf{X}_{\mathcal{P}_{g^*}}^T r^{(t)}\|_2^2 = \max_{g_0\in\mathcal{Z}^{(t)}} \|\mathbf{X}_{\mathcal{P}_{g_0}}^T r^{(t)}\|_2^2$. Once a group $\mathcal{P}_{g^*}$ is selected, one derives the new set of active indices $\mathcal{A}^{(t+1)} = \mathcal{A}^{(t)} \cup \mathcal{P}_{g^*}$ and updates the current solution by solving the least squares problem of reduced dimension $w_{\mathcal{A}^{(t+1)}}^{(t+1)} = \arg\min_w \|Y - \mathbf{X}_{\mathcal{A}^{(t+1)}} w\|_2^2$.

We notice that the algorithm can be viewed as a walk through a directed acyclic graph, which also appears in the algorithm proposed in [15]. Formally, our OMP-type algorithm can be described as follows.

Algorithm 1 (OMP-type algorithm for overlapping groups): Initialization. Set $t = 0$.

`Repeat until` some stopping criterion is satisfied.

`If` $t = 0$ (initial step)

Compute $(j^*, p^*, q^*) = \arg\max_{(j,p,q)} \|x_{(j,p,q)}^T Y\|_2^2$.

Denote by g_α^* and g_β^* the group indices associated with

the groups G_{p^*,q^*}^α and $G_{j^*}^\beta$, respectively

Put the set of zero group indices $\mathcal{Z}^{(1)} = \mathcal{I}\setminus\{g_\alpha^*, g_\beta^*\}$.

Put the set of active indices $\mathcal{A}^{(1)} = \{(j^*, p^*, q^*)\}$. `Else`

For each $g_0 \in \mathcal{Z}^{(t)}$ set

$$\mathcal{P}_{g_0} = \left(\cup_{g\in\mathcal{Z}^{(t)}\setminus\{g_0\}} G_g\right)^c \setminus \mathcal{A}^{(t)} \ .$$

Compute $g^* = \arg\max_{g_0\in\mathcal{Z}^{(t)}} \|(\mathbf{X}_{\mathcal{P}_{g_0}})^T r^{(t)}\|_2^2$.

Update $\mathcal{Z}^{(t+1)} = \mathcal{Z}^{(t)}\setminus\{g^*\}$ and $\mathcal{A}^{(t+1)} = \mathcal{A}^{(t)} \cup \mathcal{P}_{g^*}$.

`End`

Update the current solution

$$\tilde{w}_{\mathcal{A}^{(t+1)}}^{(t+1)} = \arg\min_{w\in\mathbb{C}^{|\mathcal{A}^{(t+1)}|}} \|Y - \mathbf{X}_{\mathcal{A}^{(t+1)}} w\|_2^2 \ . \tag{12}$$

Update the residual $r^{(t+1)} = Y - \mathbf{X}_{\mathcal{A}^{(t+1)}} \tilde{w}^{(t+1)}$.

Set $t = t + 1$.

`End`

Note that in any case the algorithm stops after a finite number of iterations, more exactly after $J+PQ-1$ steps. Then all groups are activated and the solution $\tilde{w}$ is the ordinary least squares estimator. The important issue of an appropriate stopping criterion for the algorithm is addressed in Section V. Clearly, the aim is to stop the algorithm earlier to provide a relevant, sparse solution.

We remark that the proposed algorithm is a natural extension of the orthogonal matching pursuit. Moreover it is similar to the active set algorithm described in [15], which provides an interesting mathematical analysis of the optimization problem given by (10). Nevertheless, our OMP-type algorithm provides an approach which is feasible in practice and computationally faster than the active set algorithm. Indeed, in (12) the solution is updated by the ordinary least squares estimator that is known explicitly, whereas [15] proposes to solve the *penalized* problem of reduced dimension by a more expensive second-order cone programming.

B. Solution of the reduced problem

The solution of (10) is a sparse vector, say $\tilde{w}$, whose sparsity pattern serves to identify the indices of the nonzero coefficients $\alpha_{p,q}$ and β_j. More precisely, if $\tilde{w}_{j,p,q} \neq 0$, this implies that the associated coefficients $\alpha_{p,q}$ and β_j are both non zero. Conversely, $\tilde{w}_{j,p,q} = 0$ implies that at least one of the coefficients is 0.

To obtain estimates of these nonzero coefficients, we reduce the dimension of the regression matrix $\mathbf{X}$ by keeping the predictor $x_{j,p,q}$ only if $\tilde{w}_{j,p,q} \neq 0$. Denote the resulting matrix by $\mathbf{X}_{\text{red}}$. Then compute the least squares estimator by minimizing

$$L(\tilde{\alpha}, \tilde{\beta}) = \|Y - \mathbf{X}_{\text{red}}(\tilde{\beta} \otimes \tilde{\alpha})\|^2 \,, \tag{13}$$

where $\tilde{\alpha}$ and $\tilde{\beta}$ are the associated α- and β-vectors of reduced dimension $\tilde{P}\tilde{Q}$ and $\tilde{J}$, respectively This minimization is a well-posed problem as the dimension of $\mathbf{X}_{\text{red}}$ is much smaller than that of $\mathbf{X}$.

To compute the solution of (13) one can use the fact that the model is linear in $\tilde{\alpha}$ for fixed $\tilde{\beta}$ and conversely. That means for fixed $\tilde{\beta}$ we define the matrix $\mathbf{X}(\tilde{\beta})$ with $\tilde{P}\tilde{Q}$ columns by

$$\mathbf{X}_{\text{red}}(\tilde{\beta}) = \sum_{j=1}^{\tilde{J}} \tilde{\beta}_j (x^{\text{red}}_{j,1,1}, \ldots, x^{\text{red}}_{j,\tilde{P},\tilde{Q}}) \,.$$

Then the minimum of $\tilde{\alpha} \mapsto L(\tilde{\alpha}, \tilde{\beta})$ is the ordinary least squares estimator given by

$$\tilde{\alpha}^{\text{OLS}}(\tilde{\beta}) = (\mathbf{X}_{\text{red}}(\tilde{\beta})^T \mathbf{X}_{\text{red}}(\tilde{\beta}))^{-1} \mathbf{X}_{\text{red}}(\tilde{\beta})^T Y \,.$$

Furthermore, as $\tilde{\alpha}$ and $\tilde{\beta}$ are identifiable only up to some multiplicative constant, one can set $\tilde{\beta}_1 = 1$, for instance. Finally, it remains to minimize

$$(\tilde{\beta}_2, \ldots, \tilde{\beta}_{\tilde{J}}) \mapsto L\left(\tilde{\alpha}^{\text{OLS}}((1, \tilde{\beta}_2, \ldots, \tilde{\beta}_{\tilde{J}})^T), (1, \tilde{\beta}_2, \ldots, \tilde{\beta}_{\tilde{J}})^T\right)$$

which can be done by the Nelder-Mead simplex method. This is feasible since the number of unknown variables, *i.e.* $\tilde{J} - 1$, is quite low.

IV. SCALABILITY

To obtain a good representation of the waveform, a large dictionary shall be used. Likewise, to avoid biased estimators of angles and time delays, we may use fine grids. However, large dictionary sizes J and large grid sizes PQ entail a huge regression matrix $\mathbf{X}$ having JPQ columns (see Section VI where $JPQ \approx 6.2\,10^9$). This raises computational difficulties, namely concerning the storage of the regression matrix $\mathbf{X}$.

A closer look at the OMP-type algorithm shows that there are no computations involving the entire regression matrix. Essentially, the algorithm considers one by one all possible groups of variables $G \in \mathcal{G}$ that may enter the solution in the present iteration. For each group of variables G the correlation of the associated predictors with the current residual $r^{(t)} = Y - \mathbf{X}_G \tilde{w}^{(t)}$ is computed. Hence, instead of the whole matrix $\mathbf{X}$, only the predictors of group G are required.

In short, the storage of $\mathbf{X}$ can be avoided by recomputing the required predictors at each iteration. With such a programming there are almost no limits on the size of the regression matrix, and thus almost arbitrary dictionaries and grids may be used.

To be more precise, let us have a look at the dictionary used in the following simulation study. The dictionary is composed of sinusoids with different frequencies f, different lengths l and different starting points s. The index j is hence a triple index (f, l, s). Now instead of storing the whole vector $x_{j,p,q}$, we just use the triple index (f, l, s) to reconstruct the dictionary element φ_j, whenever it is needed. Then further translation and multiplication with the appropriate steering vector yield the required predictor $x_{j,p,q}$. Finally, we only have to store the steering matrix for all possible angles of arrival, which is a matrix of very moderate size $C \times P$. With this knowledge, the predictor $x_{j,p,q}$ is easily reconstructed at any iteration of the algorithm.

In regression problems it is common to work with a normalized regression matrix, that is all columns have mean zero and standard deviation one. To avoid the computation of the normalization constants for every predictor, they may be computed just once and stored for later usage. Note that the normalization constants do not depend neither on the starting point of the dictionary element nor on the delay on the pathway but only on the frequency f and the length l of the dictionary element and the angle of arrival. Hence, the number of constants to store remains reasonable.

The most time-consuming step in the algorithm is the scan of all predictors (or more precisely, of all relevant groups of predictors) in every iteration, to identify the one that is the most correlated with the current residual. It is noteworthy that this step can be speeded up by parallelization of the algorithm.

Indeed, there is no specific order to visit the predictors and the problem can be split in several independent subtasks.

V. Stopping criterion

The OMP-type algorithm requires an appropriate stopping criterion. A common approach to conceive a good stopping criterion is based on cross-validation [13]. Note that in our context, we do not have to deal with a single, but with two sparse vectors, namely α and β and their degrees of sparsity may not be equal. Hence, cross-validation would be computationally demanding.

Here we propose a stopping criterion that on the one hand takes into account available prior information on the number of propagation pathways (and hence on the number of α-variables) and on the other hand has a data-driven component to determine the total number of non zero $\alpha_{p,q}$- and β_j-coefficients.

A. Sparsity of $\alpha_{p,q}$-coefficients

In the context of intercepted radar signals, one generally expects a small number of relevant propagation pathways. Indeed, we are mainly interested in the recovery of the direct path to determine the principal DOA. It is hence convenient to introduce an upper limit for the number of paths $U_{\max}$ and to modify the OMP-type algorithm such that at every iteration the number of activated $\alpha_{p,q}$-coefficients is checked. If their number has achieved $U_{\max}$, then we stop activating further $\alpha_{p,q}$-coefficients and concentrate on activating only β_j-coefficients.

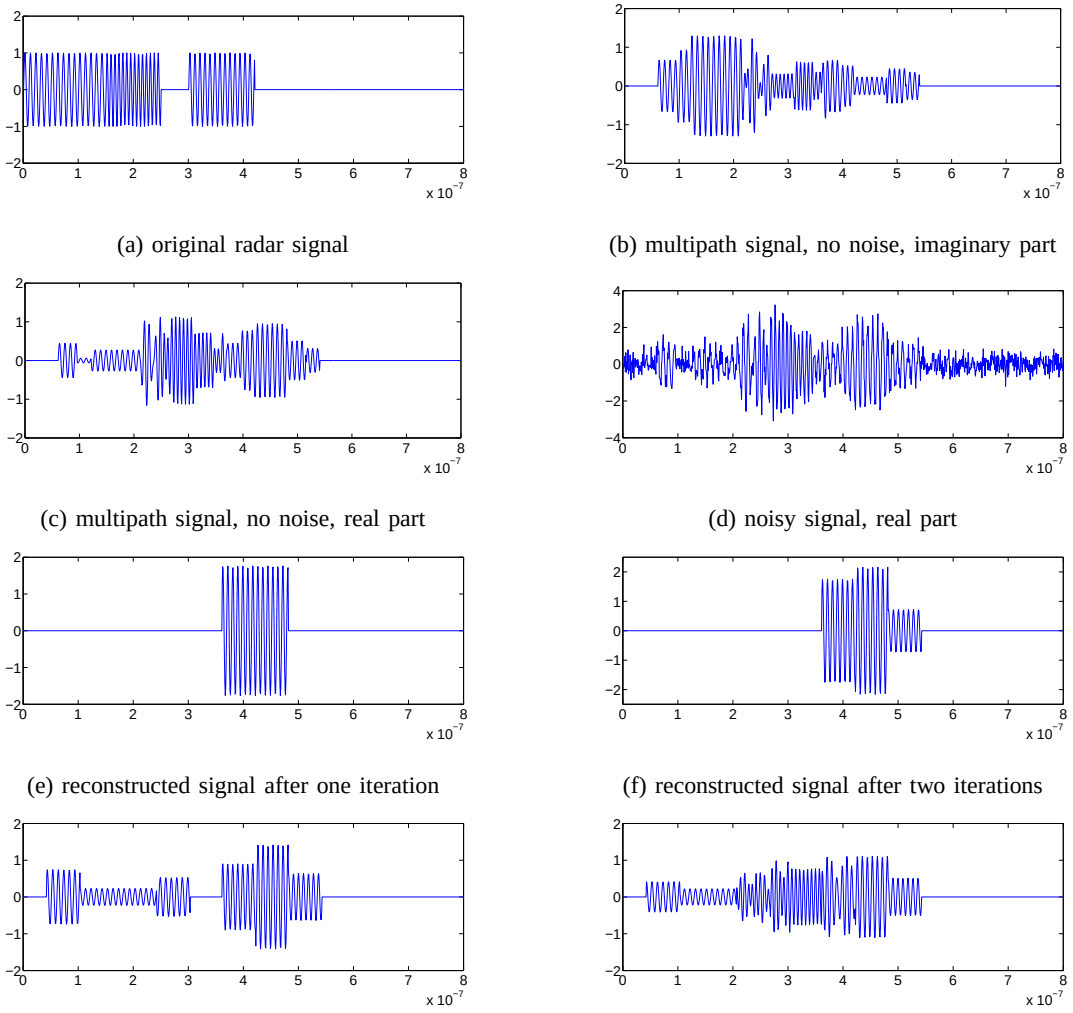
In simulations we observed that this procedure yields satisfactory results. The only important assumption for a good performance is that the direct path is not too weak compared to the indirect pathways, since the algorithm activates the components by their importance in the signal.

Note that, if $U_{\max}$ is smaller than the 'true' number of propagation pathways, then the observed signal Y cannot be completely explained. In other words, the solution obtained by the OMP-type algorithm is not the optimal solution of the minimization problem (10).

B. Control of the squared error

The degree of sparsity in the β-variables depends on the chosen dictionary. In general, very few prior information is available on the number $\tilde{J}$ of basis elements used to decompose the radar signal.

A simple but useful tool for model selection consists in studying the evolution of the squared error, that is the evolution of the square norm of the current residual $L(\tilde{w}^{(t)}) = \|Y - \mathbf{X}\tilde{w}^{(t)}\|_2^2$, where $\tilde{w}^{(t)}$



(a) original radar signal

(b) multipath signal, no noise, imaginary part

(c) multipath signal, no noise, real part

(d) noisy signal, real part

(e) reconstructed signal after one iteration

(f) reconstructed signal after two iterations

(g) reconstructed signal after three iterations

(h) reconstructed signal after four iterations

Fig. 1. Radar signal and its reconstruction. All signals, except (a), are on the first sensor. (e)-(h) represent the real part.

denotes the solution at the t-th iteration. Obviously, the squared error $L(\tilde{w}^{(t)})$ is monotone decreasing over the iterations.

The OMP-type algorithm always adds the group of variables that are the most correlated with the current residual, that is the components that are likely to diminish the squared error the most. This explains why the algorithm tends to first activating all 'true' variables, and afterwards selecting variables that are not in the true model. Hence the problem of selecting the model order becomes the detection of the point when all 'true' variables are activated and the algorithm starts adding pure noise variables.

In fact, activating a 'true' variable generally improves the squared error a lot, since the new variable

explains a considerable part of the observations. Whereas adding a variable which is not in the true model is less beneficial and the squared error hardly decreases. Consequently, first the squared error $L(\tilde{w}^{(t)})$ decreases steeply until the point where all 'true' variables are activated. Then the graph of the squared error becomes very flat. This change of the slope yields a kink in the graph of the squared error $L(\tilde{w}^{(t)})$. Thus, the kink indicates the number of 'true' variables.

An efficient way to determine the kink consists in computing the convex hull of the graph. Then the kink is given by the endpoint of the longest linear piece of the convex hull [22], [23].

We may hence proceed as follows. The OMP-type algorithm is run with some fixed $U_{\max}$ and stopped after a large number of iterations. It is useful to store all intermediate solutions $\tilde{w}^{(t)}$ and the associated square losses. Afterwards we determine the final solution as the intermediate solution $\tilde{w}^{(t)}$ where the squared error presents a kink.

Clearly, the kink is especially pronounced when the signal-to-noise ratio is high. Conversely, when there is too much noise in the data, then it may be difficult to make out a kink. Besides, when the true model contains a relatively 'weak' variable, that is, its contribution to the signal is rather low, then the number of variables may be underestimated as the 'weak' variable can be confounded with noise.

VI. Experimental Results

A. Setting

We illustrate the algorithm by an example. Let us consider a radar signal made of three subsequent sinusoids with different frequencies (100, 140, 110 MHz) and a gap between the two last sinus waves (sinusoids with starting points 0, 150, 300 ns and lengths 150, 100, 120ns), see Figure 1(a). The signal is detected by a uniform linear array (ULA) of 5 sensors separated by half a wavelength of the actual narrowband source signals. The sample frequency is $F_s = 1.28$ GHz and the total duration of observation is 0.8 μs. Thus, we have $M = 1024$ observation points and Y is a vector of length 5120. Furthermore, we consider three pathways ($U = 3$) with unknown angles of arrivals (9°, 13°, 38°) and delays (60, 95, 120ns). The signal (without noise) received at the first sensor of the antenna is presented in Figure 1(b) and (c), where the superposition of the sinusoids results from the multiple propagation pathways. From these figures it is clear that the multiple paths have a strong impact on the signal and hence cannot be neglected in the analysis. The signal is corrupted by additive noise with a signal-to-noise ratio of 5 dB, see Figure 1(d). The signal-to-noise ratio is determined on the direct propagation pathway, defined as the $10 \log$ (energy of the signal on the first path/variance of the noise) as used in [24] or [6].

We choose a Fourier dictionary containing sine and cosine functions with frequencies 90, 95,..., 150 MHz, signal lengths 80, 85,..., 200 ns and starting points 0, 3δ, 6δ,..., 255δ ns, where $\delta = 1/F_s \approx 0.78$ ns. The size of the dictionary is thus $J = 166\,400$. For the angles and the delays we use the grids $\{1^\circ$, 2°,..., $90^\circ\}$ and $\{0, 3\delta, 6\delta,\ldots, 255\delta\}$, respectively Hence, w is a vector of dimension $3.83 \cdot 10^9$.

B. Signal reconstruction by the OMP-type algorithm

In the simulations we set $U_{\text{max}} = 2$, that is, at best we recover two of the three pathways and we stop the algorithm after 9 iterations. Figure 1(e)-(h) illustrate the first four iterations of the OMP-type algorithm. In the first iteration the algorithm selects a single sinusoid. Next, a second pathway is selected leading to two overlapping sinusoids. As now two α-variables are activated and $U_{\text{max}} = 2$, the algorithm can only add further β-variables, that is, further dictionary elements. Thus, in the third iteration another sinusoid is added which appears directly on the two pathways. Finally, another sinusoid is activated.

We note that the graph in (h) is quite similar to the one in (c). Nevertheless, the reconstruction is not perfect, as we allowed the algorithm to find only two pathways, while indeed there are three. In this example the second pathway corresponding to the angle of arrival of 13° is missing.

C. Model selection device

As mentioned above, we stopped the algorithm after 9 iterations. Then we used the graph of the squared error to determine the number of variables in the model by the point where the curve exhibits a kink. The kink is determined by computing the convex hull of the graph. Figure 2 illustrates the evolution of the squared loss for a sample with a signal-to-noise ratio of 5 dB. The curve has a slight kink at 5, which is the appropriate number of variables in this case, as the signal is composed of three sinusoids and we set $U_{\text{max}} = 2$.

TABLE I

EMPIRICAL MEAN AND STANDARD DEVIATION OF THE TOTAL NUMBER OF SELECTED VARIABLES OF α- AND β-COEFFICIENTS WITH $U_{\text{MAX}} = 2$.

SNR	total nb	nb of α's	nb of β's
5 dB	5 (0)	2 (0)	3 (0)
-5 dB	5 (0)	2 (0)	3 (0)
-15 dB	5.1 (0.3)	2 (0)	3.1 (0.3)

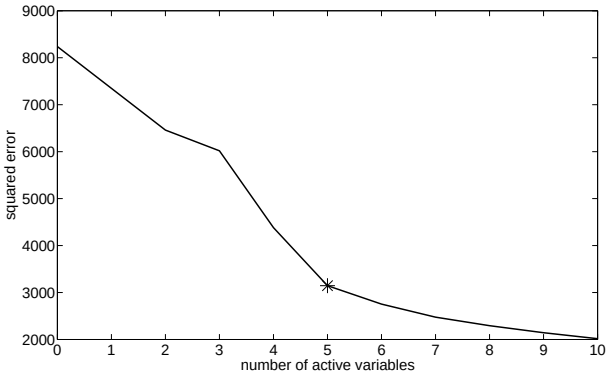


Fig. 2. Graph of the squared error of a dataset with a signal-to-noise ratio of 5 dB. The mark indicates the kink, that is the selected number of model variables.

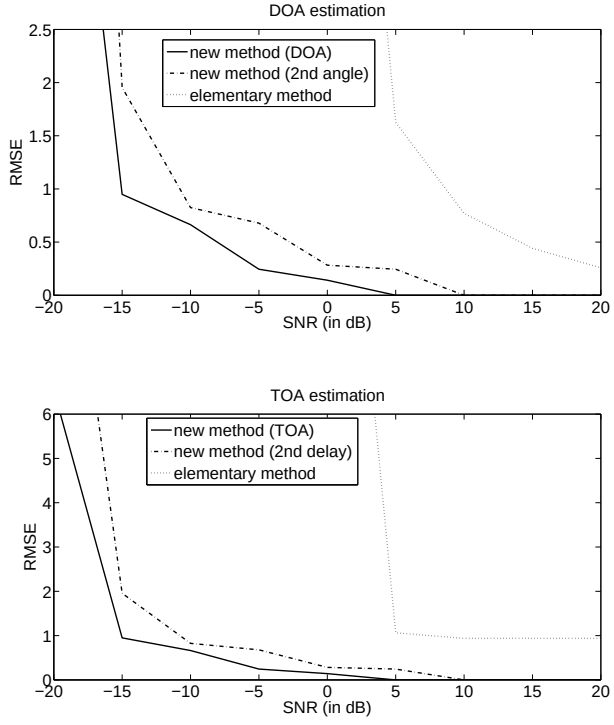


Fig. 3. Comparison of RMSE of the DOA and TOA for the elementary method and the new estimator for different signal-to-noise ratios applied on datasets of size 5 x 1280. The RMSE values are based on 50 repetitions per noise level.

Table I shows the empirical mean and standard deviation of the total number of selected variables, the number of α- and β-coefficients for different signal-to-noise ratios over 50 repetitions. Obviously, in this example the model selection tool works very well, as the model selection device successfully distinguishes 'true' variables from noisy variables, even though the algorithm is prevented from selecting all 'true' variables. Hence, the introduction of an upper bound $U_{\max}$ for the number of α-coefficients does not corrupt the model selection.

D. Comparison with an elementary method

A simple, elementary method for estimating the principal DOA in the radar setting consists in first estimating the TOA by thresholding and then using the point of observation at the TOA to fit the angle of arrival. More precisely, to detect the TOA by the elementary method, the absolute value of the observed signal at some time point t is compared to some threshold, here 3σ, where σ^2 is the variance of the noise distribution that is supposed to be known for the elementary method. The first point t where this happens is considered as the TOA. This detection method fails when the signal-to-noise ratio is too low, since then the signal is too weak compared to the noise such that the observed signal never exceeds the threshold. And when no TOA is detected, the elementary method does not provide an estimate of the DOA.

We compare this elementary method to our OMP-type algorithm in the following setting with an FSK-modulated Barker sequence. We consider the 7-bit Barker sequence [1, 1, -1, -1, 1,-1, -1] with FSK-modulation. Every bit is represented by a sinusoid of length 80 ns with frequency 210 or 230 MHz. As there are four sign changes in the sequence, the transmitted signal is composed of four consecutive sinusoids of varying length. Furthermore, we consider two pathways ($U = 2$) with unknown angles of arrivals (6°and 42°) and delays of 10 and 40 ns.

TABLE II

EMPIRICAL MEAN AND STANDARD DEVIATION (IN PARENTHESES) OF THE PARAMETER ESTIMATES OBTAINED BY THE ELEMENTARY METHOD.

Elementary Method		
	1st angle (DOA)	1st delay (TOA)
true	6°	10 ns
SNR 20 dB	5.99 (0.26)	10.94 (0)
SNR 0 dB	6.81 (8.43)	22.31 (12.11)

Again, the sample frequency is $F_s = 1.28$ GHz and we use a ULA of 5 sensors separated by half a wavelength of the actual narrowband source signals. The observation time is 1 μs, such that Y is a vector of length 6400. The dictionary contains the sine and cosine functions with frequencies 200, 202,..., 240 MHz, lengths 50, 55,..., 170 ns and starting points 0, δ, 2δ,..., 255δ ns, where $\delta = 1/F_s$. The size of the dictionary is thus $J = 268\,800$. We use the grids $\{1^\circ, 2^\circ, \ldots, 90^\circ\}$ and $\{0, \delta, 2\delta, \ldots, 255\delta\}$ for the angles and the delays, respectively Hence, w is a vector of dimension $6.19 \cdot 10^9$. Data are simulated with different levels of the signal-to-noise ratio. Again we choose $U_{\max} = 2$.

Figure 3 compares the root mean square error (RMSE) of the DOA and TOA estimates obtained by both methods based on 50 simulated datasets for each noise level. Compared to the elementary procedure the OMP-type algorithm achieves a gain of around 20 dB. This gain is mainly due to the fact that the algorithm takes into account all observations and not only a single point. For low signal-to-noise ratios the elementary method completely fails.

E. Parameter Estimates

It is important to note that in contrast to the elementary method the OMP-type algorithm provides estimates not only of the DOA and the TOA, but of all the other parameters including the waveform. Indeed, in the setting described above our algorithm identifies several consecutive sinusoids. Table II presents the empirical mean and standard deviation of the estimates obtained by the elementary method and Table III gives the results for the OMP-type method, namely for the estimates of the angles of arrival, the associated delays, the frequencies of the four detected sinusoids, their lengths and their starting points. Estimates are especially accurate for the angles of arrival and the frequencies, whereas the estimates of the lengths of the sinusoids and their starting points are less accurate. There is a clear tendency to overestimate the length, especially for the shortest sinusoid. Indeed, for high noise levels it is not clear if the OMP-type algorithm detects the shortest (and less energetic) sinusoid. The fourth sinusoid generally has not even the correct frequency, so it is possible that the algorithm adds any non relevant predictor, because the shortest sinusoid is too weak for detection.

We observe that the reconstructed radar signal $\hat{s}$ has not exactly the form of a modulated signal due to the overlapping sinusoids. It is hence up to the user to make such an interpretation of the signal. Nevertheless, we emphasize that the algorithm works for any type of modulation (FSK or PSK) without prior knowledge on the modulation and it provides a good idea of the waveform.

VII. Conclusion

The algorithm proposed in this paper is to the best of our knowledge the first method addressing the problem of estimating both the DOA and the unknown waveform of a transmitted radar signal in the presence of multiple propagation pathways. The new method is in the spirit of OMP extended to more sophisticated sparsity structures. The scaling of the OMP-type algorithm has been considered, which is necessary for unbiased estimation. A simulation study illustrates the good performance of the new method and exhibits a considerable improvement with respect to some elementary method.

We would like to emphasize the general character of our method, which is of much interest for the intercepted radar signal setting. First, the method proposed here is not restricted to a specific array geometry. Second and more importantly, the approach can be adapted to a large variety of signals such as sinusoids, chirps, composed signals etc. by using a dictionary where a sparse representation of the signal is possible.

References

[1] H. Krim and M. Viberg, "Two decades of array signal processing research: The parametric approach," *IEEE Signal Processing Magazine*, vol. 13, no. 4, pp. 67–94, 1996.

[2] Y. Zhang, Z. Ye, and C. Liu, "An efficient DOA estimation method in multipath environment," *Signal Processing*, vol. 90, no. 2, pp. 707–713, 2010.

[3] M. C. Vanderveen, C. B. Papadias, and A. Paulraj, "Joint angle and delay estimation (jade) for multipath signals arriving at an antenna array," *IEEE Communications Letters*, vol. 1, no. 1, pp. 12–14, 1997.

[4] C.-W. Ma and C.-C. Teng, "Detection of coherent signals using weighted subspace smoothing," *IEEE Transactions on Antennas and Propagation*, vol. 44, no. 2, pp. 179–187, 1996.

[5] T. J. Shan, M. Wax, and T. Kailath, "On spatial smoothing of estimation of coherent signals," *IEEE on Acoustics, Speech, and Signal Processing*, vol. ASSP-33, pp. 806–811, 1985.

[6] J.-J. Fuchs, "Multipath time-delay detection and estimation," *IEEE Transactions on Signal Processing*, vol. 47, pp. 237–243, 1999.

[7] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society, Series B*, vol. 58, no. 1, pp. 267–288, 1996.

[8] Y. C. Pati, R. Rezaifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition," in *27th Asilomar Conference on Signals, Systems and Computation*, 1993.

[9] S. Mallat, G. Davis, and Z. Zhang, "Adaptive time-frequency decompositions," *SPIE Journal of Optical Engineering*, vol. 33, pp. 2183–2191, 1994.

[10] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society, Series B*, vol. 67, pp. 301–320, 2005.

[11] R. Tibshirani, M. Saunders, S. Rosset, J. Zhu, and K. Knight, "Sparsity and smoothness via the fused lasso," *Journal of the Royal Statistical Society, Series B*, vol. 67, pp. 91–108, 2005.

[12] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *Journal of the Royal Statistical Society, Series B*, vol. 68, no. 1, pp. 49–67, 2006.

[13] P. Zhao, G. Rocha, and B. Yu, "The composite absolute penalties family for grouped and hierarchical variable selection," *The Annals of Statistics*, vol. 6A, pp. 3468–3497, 2009.

[14] L. Jacob, G. Obozinski, and J.-P. Vert, "Group lasso with overlap and graph lasso," in *ICML '09: Proceedings of the 26th Annual International Conference on Machine Learning*. New York, NY, USA: ACM, 2009, pp. 433–440.

[15] R. Jenatton, J.-Y. Audibert, and F. Bach, "Structured variable selection with sparsity-inducing norms," 2009, technical report, arXiv:0908.0050v1.

[16] A. Lozano, G. Swirszcz, and N. Abe, "Grouped orthogonal matching pursuit for variable selection and prediction," in *Advances in Neural Information Processing Systems 22*, 2009.

[17] D. Malioutov, M. Çetin, and A. S. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Transactions on Signal Processing*, vol. 53, no. 8, pp. 3010–3022, 2005.

[18] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, pp. 33–61, 1998.

[19] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," *The Annals of Statistics*, vol. 32, no. 2, pp. 407–499, 2004.

[20] S. S. Chen and D. L. Donoho, "Basis pursuit," Department of Statistics, Stanford University, Tech. Rep., 1994. [Online]. Available: http://stat.stanford.edu/~donoho/Reports/1994/asilomar.pdf

[21] R. Jenatton, J.-Y. Audibert, and F. Bach, "Active set algorithm for structured sparsity-inducing norms," in *OPT 2009: 2nd NIPS Workshop on Optimization for Machine Learning*, 2009.

[22] M. Lavielle, "Using penalized contrasts for the change-point problem," *Signal Processing*, vol. 85, pp. 1501–1510, 2005.

[23] M. Lavielle and C. Lévy-Leduc, "Semiparametric estimation of the frequency of unknown periodic functions and its application to laser vibrometry signals," *IEEE Transactions on Signal Processing*, vol. 53, pp. 2306–2314, 2005.

[24] T. G. Manickam, R. J. Vaccaro, and D. W. Tufts, "A least squares algorithm for multipath time-delay estimation," *IEEE Transactions on Signal Processing*, vol. 42, pp. 3229–3233, 1994.

TABLE III

EMPIRICAL MEAN AND STANDARD DEVIATION (IN PARENTHESES) OF THE PARAMETER ESTIMATES OBTAINED BY THE OMP-TYPE ALGORITHM.

OMP-type algorithm

	1st angle	1st delay	2nd angle	2nd delay
true	6°	10 ns	42°	40 ns
20 dB	6.00 (0)	10.94 (0)	42.0 (0)	40.6 (0)
0 dB	5.94 (0.24)	10.5 (1.34)	41.9 (0.27)	40.2 (1.26)

	1st frequ.	2nd frequ.	3rd frequ.	4th frequ.
true	210 MHz	230 MHz	210 MHz	230 MHz
20 dB	230.0 (0)	210.0 (0)	230.0 (0)	210.0 (0)
0 dB	230.0 (0)	210.0 (0)	230.1 (0.40)	210.0 (0)

	1st length	2nd length	3rd length	4th length
true	160 ns	160 ns	80 ns	160 ns
20 dB	160.0 (0)	160.6 (1.64)	162.4 (12.1)	170.0 (0)
0 dB	163.1 (4.62)	165.5 (4.20)	162.4 (15.1)	170.0 (0)

	1st start	2nd start	3rd start	4th start
true	0 ns	160 ns	320 ns	400 ns
20 dB	0 (0)	158.8 (1.54)	265.3 (12.9)	389.1 (0)
0 dB	0 (0)	153.1 (5.37)	271.0 (19.7)	389.5 (1.34)