



HAL
open science

A simultaneous sparse approximation method for multidimensional harmonic retrieval

Souleymen Sahnoun, El-Hadi Djermoune, David Brie, Pierre Comon

► **To cite this version:**

Souleymen Sahnoun, El-Hadi Djermoune, David Brie, Pierre Comon. A simultaneous sparse approximation method for multidimensional harmonic retrieval. [Research Report] GIPSA-lab. 2016. hal-01174780v1

HAL Id: hal-01174780

<https://hal.science/hal-01174780v1>

Submitted on 25 May 2016 (v1), last revised 30 Aug 2016 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A Simultaneous Sparse Approximation Method for Multidimensional Harmonic Retrieval

Souleymen Sahnoun^{a,*}, El-Hadi Djermoune^b, David Brie^b and Pierre Comon^a

^a Gipsa-Lab, CNRS, University of Grenoble Alpes
^b CRAN, Université de Lorraine, CNRS

Abstract

In this paper, a new method for the estimation of the parameters of multidimensional (R -D) harmonic and damped complex signals in noise is presented. The problem is formulated as R simultaneous sparse approximations of multiple 1-D signals. To get a method able to handle large size signals while maintaining a sufficient resolution, a multigrid dictionary refinement technique is associated to the simultaneous sparse approximation. The refinement procedure is proved to converge in the single R -D mode case. Then, for the general multiple modes case, the signal tensor model is decomposed in order to handle each mode separately in an iterative scheme. The proposed method does not require an association step since the estimated modes are automatically “paired”. We also derive the Cramér-Rao lower bounds of the parameters of modal R -D signals. The expressions are given in compact form in the single tone case. Finally, numerical simulations are conducted to demonstrate the effectiveness of the proposed method.

Index Terms—Multidimensional harmonic retrieval, frequency estimation, simultaneous sparse approximation, multigrid dictionary refinement, Cramér-Rao lower bound

1 Introduction

The problem of estimating the parameters of sinusoidal signals from noisy measurements is an important topic in signal processing and several parametric and nonparametric approaches have been developed for one-dimensional (1-D) signals [1]. Recently, this problem has received a renewed interest thanks to the emergence of multidimensional (R -D) applications. Indeed, parameter estimation from R -D signals is required in numerous applications in signal processing and communications such as nuclear magnetic resonance (NMR) spectroscopy, wireless communication channel estimation [2] and MIMO radar imaging [3]. In all these applications, signals are assumed to be a superposition of R -D sinusoids or, more generally, of exponentially decaying R -D complex exponen-

tials (modal signals). As for the 1-D case, the crucial step is the estimation of the R -D modes (including frequencies and damping factors) because they are nonlinear functions of the data.

In order to achieve high resolution estimates, parametric approaches are often preferred to nonparametric ones. Several parametric R -D methods ($R \geq 2$) have been proposed. They include linear prediction-based methods such as 2-D TLS-Prony [4], and subspace approaches such as matrix enhancement and matrix pencil (MEMP) [5], 2-D ESPRIT [6], multidimensional folding (MDF) [7], improved multidimensional folding (IMDF) [8, 9], Tensor-ESPRIT [10], principal-singular-vector utilization for modal analysis (PUMA) [11, 12] and the methods proposed in [13, 14]. All these methods perform at various degrees but it is generally admitted that they yield accurate estimates at high SNR scenarios and/or when the frequencies are well separated. This is obtained at the expense of computational effort. In [11], tensor PUMA was proposed as an accurate and computationally efficient multidimensional harmonic retrieval method, which attains the Cramér-Rao lower bound (CRLB) and does not require to build large size matrix or tensor. However its performance degrades rapidly with the increase of the number of components in the R -D signal.

Recently, methods based on sparse approximations have been proposed to address the harmonic or modal retrieval problem [15, 16, 17, 18, 19, 20, 21, 22]. For time-data spectral estimation, the dictionary is formed from a set of (normalized) complex exponentials potentially embedded in the data, which allows one to easily include some prior knowledge about the position of some known modes. More generally, the usual choice is a uniform spectral grid obtained by sampling the frequency and damping factor lines. Clearly, a fine grid is required to get a good resolution but, on the other hand, it will result in a huge dictionary [15]. This complexity is further increased in the case of R -D signals in which we are confronted with $2R$ -D grids. In order to reduce the computational burden, a multigrid scheme for sparse approximation was proposed in [19] to iteratively refine the dictionary starting from a coarse one. At each iteration, a sparse approximation is performed and then new grid points (called “atoms”) are inserted in the vicinity of active ones leading to

This work is funded by the European Research Council under the Seventh Framework Programme FP7/2007–2013 Grant Agreement no. 320594.
e-mail (Gipsa-lab): firstname.lastname@gipsa-lab.grenoble-inp.fr
e-mail (CRAN): firstname.lastname@univ-lorraine.fr

a multiresolution-like scheme. This algorithm, which refines jointly R 2-D grids, is efficient but has mainly two drawbacks: 1) it does not have convergence guarantees, 2) the dictionary becomes intractable for large signals when $R \geq 2$. Recently, several studies have also focused on gridless sparse recovery methods based on continuous dictionaries [23, 24]. However, the proposed algorithms demand a large computational burden even for 1-D signals.

The goal of the present paper is to propose a fast multi-dimensional modal estimation technique able to handle large signals and yielding a good estimation accuracy.

1. First, the proposed approach, as for some parametric methods for modal retrieval, is based on the idea of estimating the parameters independently along each dimension $r = 1, \dots, R$. It will be shown that the *simultaneous* sparse approximation concept [25, 20] is well-suited for R -D modal retrieval ($R \geq 2$).
2. The second contribution consists in the proposition of a new multigrid scheme which amounts to consider a two-step refinement of 1-D grids, the first step for frequencies and the second one for damping factors. One advantage of this procedure is that it reduces the computational time. The convergence of the proposed multigrid strategy is analyzed in the single tone case ($F = 1$), and convergence conditions are expressed in terms of atom positions in the initial dictionaries.
3. The extension of this result to the multiple tones case ($F > 1$) is not trivial because, not only it depends on the selected sparse approximation algorithm, but also on the coherence of the dictionary [25]. Indeed, due to the multigrid strategy, the columns of the refined dictionary are increasingly correlated, which may prevent convergence even in the noiseless case. Consequently, for $F > 1$, we exploit an alternative representation of the data model enabling the extraction of the R -D signal tones separately. Therefore, the third contribution of this paper is the derivation of a new algorithm for estimating parameters of R -D damped signals in which the results of the previous contribution apply. The effectiveness of the new algorithm for multiple R -D tones is also analyzed. One very interesting by-product of this approach is that the pairing of R -D parameters is achieved for free, without any further association stage.

The usual way to assess the performances of an estimation method is to compare the variance of the estimates to the CRLB. In [5] Y. Hua derived the CRLB for 2-D frequencies, i.e., undamped 2-D exponentials; no damped signals are considered. Closed-form expressions of the CRLB for the general undamped R -D case are derived in [26]. CRLB for 2-D damped signals are derived in [27]. Therefore, to the best of our knowledge, no compact expressions of the CRLB's are available for the general R -D damped model. Thus, another contribution of the paper is the derivation of the CRLB's for the frequency, damping factor, amplitude and phase of this model.

The remainder of this paper is organized as follows. In section 2, we introduce notation and present the R -D modal retrieval problem. In section 3, we formulate the R -D modal estimation problem as R simultaneous sparse estimation problems, show how to construct a modal dictionary on a uniform grid and then describe the new fast multigrid strategy. In section 4, we give sufficient conditions for convergence of the multigrid dictionary refinement in the case of single tone R -D signals. In light of these new results, we propose in section 5 a new efficient algorithm for multiple tones R -D signals. In section 6, we derive the expressions of the CRLB's for the parameters of R -D damped exponentials in Gaussian white noise. We then give the CRLB in the cases of single damped and undamped R -D cisoids. The effectiveness of the proposed method is demonstrated using simulation signals in section 7. Finally, conclusions are drawn in section 8.

2 Notation and Problem Statement

2.1 Notation

Scalars are denoted as lower-case letters (a, b, α), column vectors as lower-case bold-face letters (\mathbf{a}, \mathbf{b}), matrices as bold-face capitals (\mathbf{A}, \mathbf{B}), and tensors as calligraphic bold-face letters (\mathcal{A}, \mathcal{B}). Notations $(\cdot)^T, (\cdot)^H$ and $(\cdot)^\dagger$ stand for the transpose, the Hermitian transpose and the pseudo-inverse, respectively. The symbols “ \odot ” and “ \otimes ” will denote the Khatri-Rao product (column-wise Kronecker) and the Kronecker product, respectively. Both words “mode” and “tone” are used to refer to a component of the multidimensional signal. The tensor operations used here are consistent with [28]:

- the outer product of two tensors $\mathcal{A} \in \mathbb{C}^{M_1 \times \dots \times M_R}$ and $\mathcal{B} \in \mathbb{C}^{K_1 \times \dots \times K_N}$ is given by:
$$\mathcal{C} = \mathcal{A} \otimes \mathcal{B} \in \mathbb{C}^{M_1 \times \dots \times M_R \times K_1 \times \dots \times K_N},$$

$$c(m_1, \dots, m_R, k_1, \dots, k_N) = a(m_1, \dots, m_R)b(k_1, \dots, k_N)$$
(1)

- the contraction product acting on the r -th index of a tensor $\mathcal{A} \in \mathbb{C}^{M_1 \times \dots \times M_R}$ and the second index of a matrix $\mathbf{U} \in \mathbb{C}^{K \times M_r}$ is:

$$\mathcal{B} = \mathcal{A} \underset{r}{\bullet} \mathbf{U} \in \mathbb{C}^{M_1 \times \dots \times M_{r-1} \times K \times M_{r+1} \times \dots \times M_R},$$

$$b(m_1, m_2, \dots, m_{r-1}, k_r, m_{r+1}, \dots, m_R) =$$

$$\sum_{m_r=1}^{M_r} a(m_1, m_2, \dots, m_R)u(k_r, m_r)$$
(2)

- the matrix $\mathbf{A}_{(r)} \in \mathbb{C}^{M_r \times (M_1 \dots M_{r-1} M_{r+1} \dots M_R)}$ represents the unfolding (dimension- r matricization) of the tensor \mathcal{A} and corresponds to the arrangement of the dimension- r fibers of \mathcal{A} in the columns of the resulting matrix.
- $\|\mathcal{A}\|$ denotes the the Frobenius norm for tensors.

Finally, throughout this paper, the tilde symbol ($\tilde{\cdot}$) denotes a noisy signal; e.g. $\tilde{y}(\cdot) = y(\cdot) + e(\cdot)$.

2.2 Problem Formulation

An R -D modal signal is modeled as the superposition of F multidimensional damped complex sinusoids:

$$\tilde{y}(m_1, \dots, m_R) = \sum_{f=1}^F c_f \prod_{r=1}^R a_{f,r}^{m_r-1} + e(m_1, \dots, m_R) \quad (3)$$

where $m_r = 1, \dots, M_r$ for $r = 1, \dots, R$. M_r denotes the sample support of the r -th dimension, $a_{f,r} = \exp(\alpha_{f,r} + j\omega_{f,r}) \in \mathbb{C}$ is the f -th mode in the r -th dimension, $\{\alpha_{f,r}\}_{f=1, r=1}^{F,R}$, $\alpha_{f,r} \in \mathbb{R}^-$, are the damping factors, $\{\omega_{f,r} = 2\pi\nu_{f,r}\}_{f=1, r=1}^{F,R}$ are the angular frequencies, and $c_f = \lambda_f \exp(j\phi_f)$ is the complex amplitude of the f -th mode where $\lambda_f = |c_f|$ denotes the magnitude and ϕ_f the phase. $e(m_1, m_2, \dots, m_R)$ is a zero-mean complex Gaussian white noise with variance σ^2 and mutually independent in all dimensions.

In a tensor form, the R -D signal in (3) may be written as

$$\tilde{\mathcal{Y}} = \mathcal{Y} + \mathcal{E} \quad (4)$$

where $\{\tilde{\mathcal{Y}}, \mathcal{Y}, \mathcal{E}\} \in \mathbb{C}^{M_1 \times M_2 \times \dots \times M_R}$. The problem consists in estimating the set of parameters $\{a_{f,r}\}_{f=1, r=1}^{F,R}$ and $\{c_f\}_{f=1}^F$ from the R -D signal samples.

3 Simultaneous Sparse Approximation for R -D Modal Signals

3.1 Tensor Formulation of the Data Model

The noise-free data tensor \mathcal{Y} in (4) can be written in the following form:

$$\mathcal{Y} = \sum_{f=1}^F c_f \mathbf{a}_{f,1} \otimes \mathbf{a}_{f,2} \otimes \dots \otimes \mathbf{a}_{f,R} \quad (5)$$

where $\mathbf{a}_{f,r} = [1, a_{f,r}, \dots, a_{f,r}^{M_r-1}]^T$, $r = 1, \dots, R$. Equation (5) is called the Canonical Polyadic (CP) decomposition form, or the Candecomp/Parafac decomposition of the tensor \mathcal{Y} [28, 29]. The CP model (5) can be concisely denoted by $\mathcal{Y} = \llbracket \mathbf{c}; \mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_R \rrbracket$ where $\mathbf{A}_r = [\mathbf{a}_{1,r}, \mathbf{a}_{2,r}, \dots, \mathbf{a}_{F,r}]$, $r = 1, \dots, R$, and $\mathbf{c} = [c_1, c_2, \dots, c_F]^T$ is the vector of complex amplitudes. Using these definitions, the matricized form of \mathcal{Y} along the r -th dimension is given by

$$\mathbf{Y}_{(r)} = \mathbf{A}_r \mathbf{\Delta}_c (\mathbf{A}_R \odot \dots \odot \mathbf{A}_{r+1} \odot \mathbf{A}_{r-1} \odot \dots \odot \mathbf{A}_1)^T \quad (6)$$

where $\mathbf{\Delta}_c = \text{diag}(\mathbf{c})$. Then, we can write

$$\tilde{\mathbf{Y}}_{(r)} = \mathbf{A}_r \mathbf{H}_r + \mathbf{E}_{(r)} \quad (7)$$

where $\mathbf{H}_r \in \mathbb{C}^{F \times M'_r}$ is

$$\mathbf{H}_r \stackrel{\text{def}}{=} \mathbf{\Delta}_c (\mathbf{A}_R \odot \dots \odot \mathbf{A}_{r+1} \odot \mathbf{A}_{r-1} \odot \dots \odot \mathbf{A}_1)^T \quad (8)$$

and $M'_r = \prod_{\substack{k=1 \\ k \neq r}}^R M_k$. Therefore

$$\mathbf{Y}_{(r)} \stackrel{\text{def}}{=} [\mathbf{y}_{(r),1}, \dots, \mathbf{y}_{(r),M'_r}] = \left[\sum_{f=1}^F h_r(f, 1) \mathbf{a}_{f,r}, \dots, \sum_{f=1}^F h_r(f, M'_r) \mathbf{a}_{f,r} \right] \quad (9)$$

where $h_r(f, m'_r)$ is the (f, m'_r) entry of the matrix \mathbf{H}_r , for $f = 1, \dots, F$ and $m'_r = 1, \dots, M'_r$.

3.2 Simultaneous Sparse Approximation

Assuming¹ that $M_r > F, \forall r$, it is easy to see from (9) that the mode coordinates $\{a_{f,r}\}_{f=1}^{F_r}$ ($F_r \leq F$) in the r -th dimension are identifiable from any column of $\mathbf{Y}_{(r)}$. This process may be repeated on each dimension $r = 1, \dots, R$ to get all the modes coordinates. In practice, we have to replace the matrix $\mathbf{Y}_{(r)}$ by its noisy counterpart $\tilde{\mathbf{Y}}_{(r)}$ accounting for the additive white noise. In this case, (9) holds only approximately. Consequently, for each column $\tilde{\mathbf{y}}_{(r),m'_r}, m'_r = 1, \dots, M'_r$, the modal estimation problem can be formulated as a sparse approximation problem corresponding to the following constrained optimization:

$$\mathbf{x}_{m'_r} = \arg \min_{\mathbf{x}} \|\mathbf{x}\|_0 \quad \text{subject to} \quad \|\tilde{\mathbf{y}}_{(r),m'_r} - \mathbf{Q}_r \mathbf{x}\|_2^2 \leq \epsilon \quad (10)$$

where $\mathbf{Q}_r \in \mathbb{C}^{M_r \times N}$, $N \gg M_r$, is a (known) modal dictionary, $\mathbf{x} \in \mathbb{C}^N$ is a (sparse) vector containing the coefficients of the activated columns in \mathbf{Q}_r , and ϵ is a small reconstruction error related to the noise variance. The pseudo-norm $\|\mathbf{x}\|_0$ counts the number of nonzero elements in \mathbf{x} . The design of \mathbf{Q}_r is discussed in section 3.3. The fact that each vector $\tilde{\mathbf{y}}_{(r),m'_r}$ corresponds to a 1-D signal generated by the same modes implies that the position of nonzero entries in $\mathbf{x}_{m'_r}$ should be the same for $m'_r = 1, 2, \dots, M'_R$. Let \mathbf{X} be the matrix defined by

$$\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{M'_r}], \quad (11)$$

then the sparsity of \mathbf{X} may be measured by computing the Euclidian norms of the rows: those providing a nonzero norm define the rows of active atoms (which are estimations of modes $a_{f,r}$ in the dimension r) in the dictionary \mathbf{Q}_r . Therefore, we are facing a simultaneous sparse approximation problem:

$$\mathbf{X}_r = \arg \min_{\mathbf{X}} \|\mathbf{X}\|_{0,2} \quad \text{subject to} \quad \|\tilde{\mathbf{Y}}_{(r)} - \mathbf{Q}_r \mathbf{X}\|_F^2 \leq \epsilon \quad (12)$$

where $\|\mathbf{X}\|_{0,2}$ is the mixed ℓ_0/ℓ_2 -norm of \mathbf{X} (the number of rows with nonzero ℓ_2 -norm). The simultaneous sparse representation models, called also "multiple measurement vectors" (MMV), have been studied from several angles of view, and different approaches have been proposed (see [30] and references therein). As the goal of the present paper is to develop a fast approach well adapted to large signals, we restrict our attention to the SOMP algorithm [25]. However,

¹Note that this assumption is considered only in this section.

it is worth mentioning that, in more intricate cases and/or small size signals, much more efficient simultaneous sparse algorithms may be used at the price of an increased computational burden. A straightforward way to get the R -tuples $\{(a_{f,1}, \dots, a_{f,R})\}_{f=1}^F$ consists in estimating the modes $a_{f,r}$ in the R dimensions using matrices $\tilde{\mathbf{Y}}_{(r)}, r = 1, \dots, R$, which requires a further pairing step to form the R -D modes. To get accurate estimates using the described scheme, two conditions have to be satisfied, 1) the dictionary should contain all modes in the signal, 2) the sparse approximation method should have sufficient guarantees for selecting the true atoms from the dictionary: this is known as “exact recovery guarantees”. These problems are discussed in the following sections and an alternative representation of the data is used to avoid the pairing stage in the multiple tones case.

3.3 Modal Dictionary Design and Multigrid Strategy

3.3.1 Uniform Modal Dictionary

The dictionary $\mathbf{Q}_r \in \mathbb{C}^{M_r \times N}$ can be defined from a discretization of the (ν, α) plane. Each point of the grid corresponds to a hypothetical mode. Let N_μ be the number of points of a uniform grid covering the frequency interval $[0, 1)$. Similarly, let N_β be the number of points of a uniform grid covering the damping factor interval $(\beta_{\min}, 0]$, where β_{\min} is a lower bound on $\{\alpha_{f,r}\}_{f=1}^F$. Then \mathbf{Q}_r is given by

$$\mathbf{Q}_r = [\mathbf{q}_r(0, 0), \dots, \mathbf{q}_r((N_\mu - 1)\delta_\mu, 0), \mathbf{q}_r(0, \delta_\beta), \dots, \mathbf{q}_r((N_\mu - 1)\delta_\mu, \delta_\beta), \dots, \mathbf{q}_r((N_\mu - 1)\delta_\mu, (N_\beta - 1)\delta_\beta)] \quad (13)$$

where $\mathbf{q}_r(\mu, \beta) = \frac{\mathbf{a}_r(\mu, \beta)}{\|\mathbf{a}_r(\mu, \beta)\|_2}$, $\mathbf{a}_r(\mu, \beta) = [1, e^{(\beta+j2\pi\mu)}, \dots, e^{(\beta+j2\pi\mu)(M_r-1)}]^\top$, $\delta_\beta = \beta_{\min}/N_\beta$, and $\delta_\mu = 1/N_\mu$. The total number of columns in \mathbf{Q}_r is $N = N_\mu N_\beta \gg F$, each of them is called atom. In the aim of reducing the computational complexity, we propose to estimate frequencies and then damping factors by calling twice the sparse approximation method. At the first step, the frequencies are estimated using a harmonic dictionary. In the second step, the damping factors are estimated using a modal dictionary formed by the already estimated frequencies and a damping factor grid. These two steps are explained in section 4.

3.3.2 Multi-Grid Dictionary Refinement

To achieve a high-resolution modal estimation, a possible way is to define uniform grids as before and selecting very small values for δ_μ and δ_β . However, the resulting dictionaries will lead to prohibitive calculation cost and memory capacities requested. Rather, we propose to start with a coarse grid (N_μ and N_β low) and to adaptively refine it through a multigrid scheme as sketched on Figure 1. Let ℓ be the current grid level ($\ell = 0, \dots, L - 1$). At level ℓ , we first restore the signal $\mathbf{X}_r(\ell)$ related to the dictionary $\mathbf{Q}_r(\ell)$ by applying the SOMP

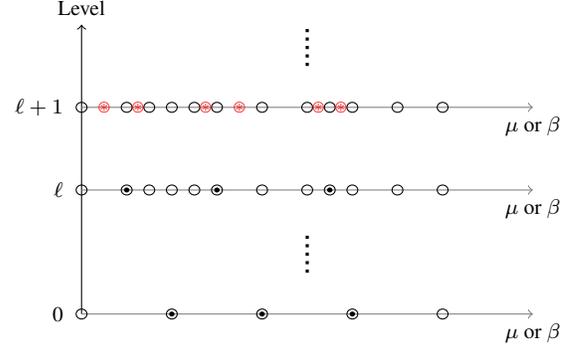


Figure 1: The multigrid dictionary refinement procedure with $\eta = 1$. (○) atoms in the dictionary; (●) activated atoms; (⊕) new atoms

Algorithm 1: Dictionary refinement (DICREF)

input : A vector $\mathbf{d} \in \mathbb{R}^N$ of sorted frequencies or damping factors, an index set Ω of activated atoms, the number of atoms $\eta \in \mathbb{N}$ to add at each side of an activated one
output: Updated vector $\mathbf{d}_{\text{updated}}$
for $i = 1 : \text{numel}(\Omega)$ **do**
 $\mathbf{d}_{i,1} = \text{linspace}(\mathbf{d}(\Omega(i) - 1), \mathbf{d}(\Omega(i)), \eta)$
 $\mathbf{d}_{i,2} = \text{linspace}(\mathbf{d}(\Omega(i)), \mathbf{d}(\Omega(i) + 1), \eta)$
 $\mathbf{d}_i = [\mathbf{d}_{i,1}^\top, \mathbf{d}_{i,2}^\top(2 : \eta)]^\top$
end
 $\mathbf{d}_{\text{updated}} = \text{union}(\mathbf{d}_1, \dots, \mathbf{d}_{\text{numel}(\Omega)})$
return $\mathbf{d}_{\text{updated}}$

method. Then we refine the dictionary by inserting atoms inbetween pairs of $\mathbf{Q}_r(\ell)$, in the neighborhood of each activated atom, and we apply again the SOMP method to restore $\mathbf{X}_r(\ell + 1)$ with respect to the refined dictionary $\mathbf{Q}_r(\ell + 1)$. This process is repeated until the desired level of resolution is reached. This procedure is applied for both frequencies and damping factors. Algorithm 1 presents the one-step dictionary refinement (DICREF), from level ℓ to $\ell + 1$, where, for a and b reals, $\text{linspace}(a, b, \eta)$ generates a set of η equispaced points in the interval $[a, b]$. The difference between the present framework and that in [19] is the following. In [19] the multigrid algorithm refines jointly R 2-D grids, which leads to expensive computations when $R \geq 2$, without convergence guarantees. The present mutigrid scheme refines linear grids, which leads to low computational complexity with convergence guarantees as will be shown in the next section.

Finding the convergence conditions of the new multigrid strategy in the general case (multiple tones) is not easy and depends on the selected sparse approximation algorithm. By contrast, it is possible to show that, under mild conditions, the convergence may be guaranteed in the single tone case. This issue is discussed in the next section. In section 5, we make use of an alternative representation of the data model in the case of multiple tones and we propose a method allowing one to retrieve the signal tones separately.

4 Single R-D Mode Estimation

In the previous section, we have shown how the R-D modal retrieval problem may be tackled using a sparse approximation algorithm by estimating the set of parameters in each dimension $r = 1, \dots, R$. Here, we give the sufficient conditions for convergence of the multigrid dictionary refinement for $F = 1$. Without loss of generality, we set $R = 1$. For notation simplicity, we omit reference to the dimension index r .

According to (3), the 1-D modal signal containing a single mode can be written as follows:

$$y(m) = c_1 a_1^{m-1} = c_1 e^{(\alpha_1 + j2\pi\nu_1)(m-1)}, m = 1, \dots, M. \quad (14)$$

Let \mathbf{Q} be a normalized modal dictionary $\mathbf{Q} = [\mathbf{q}_1, \dots, \mathbf{q}_N]$, with

$$\mathbf{q}_n = \frac{1}{\sqrt{\sum_m |q_n|^2 m}} [1, q_n, \dots, q_n^{M-1}]^T, \quad (15)$$

$q_n = \exp(\beta_n + j2\pi\mu_n)$, $\mu_n \in [0, 1)$, $\beta_n \in (\beta_{\min}, 0]$, for $n = 1, \dots, N$. The single tone sparse approximation of \mathbf{y} with respect to \mathbf{Q} is the solution of the criterion:

$$\min_{\mathbf{x}} J(\mathbf{x}) = \|\mathbf{y} - \mathbf{Q}\mathbf{x}\|^2 \quad \text{s.t.} \quad \|\mathbf{x}\|_0 = 1. \quad (16)$$

The optimal solution is given by

$$x_n^* = \mathbf{q}_n^H \mathbf{y}, \quad \mathbf{x}_{\{1, \dots, N\} \setminus n}^* = \mathbf{0}, \quad J(\mathbf{x}^*) = \|\mathbf{y}\|^2 - \mathbf{y}^H \mathbf{q}_n \mathbf{q}_n^H \mathbf{y} \quad (17)$$

where n is the selected column number in \mathbf{Q} . Finally, the minimum $J(\mathbf{x}^*)$ is reached for an atom \mathbf{q}_n that maximizes $J'(\mathbf{q}_n) = \mathbf{y}^H \mathbf{q}_n \mathbf{q}_n^H \mathbf{y} = |\mathbf{q}_n^H \mathbf{y}|^2$, $n = 1, \dots, N$.

Remark 1 At this point we notice that, for 1-D single-tone harmonic signals, maximizing $J'(\mathbf{q}_n)$ leads to the well-known beamforming method (or more precisely to the periodogram in the single snapshot case). Hence, an estimate of the frequency ν_1 may be obtained from the peak of $J'(\mathbf{q}_n)$. Here, as in [31], we use the sparse approximation framework. In our case, it allows to estimate the frequency and then the damping factor from a unified point of view and using a unique algorithm based on SOMP and DICREF.

4.1 Estimating the Frequency: The Harmonic Dictionary

First, we estimate frequency ν_1 using a harmonic dictionary (i.e. assuming $\beta_n = 0, \forall n$). In this case, we have:

$$J'(\mu_n) = \frac{|c_1|^2}{M} \left| \frac{1 - e^{\alpha_1 M + j2\pi(\nu_1 - \mu_n)M}}{1 - e^{\alpha_1 + j2\pi(\nu_1 - \mu_n)}} \right|^2. \quad (18)$$

The following theorem gives a sufficient condition for the multigrid dictionary refinement scheme to converge to the global maximum of J' .

Theorem 1 Let $y(m)$ be a single tone ($F = 1$) noiseless signal of length M and $\mathbf{Q}(\ell = 0) = [\mathbf{q}_1 \ \mathbf{q}_2 \ \dots \ \mathbf{q}_{N(0)}]^T$ be the initial harmonic dictionary in which the columns are sorted in increasing order of $\mu_n(0), n = 1, 2, \dots, N(0)$ and covering the frequency interval $[0, 1)$: $\mu_1(0) = 0$ and $\mu_{N(0)}(0) = 1 - 1/M$. Then the refinement scheme is convergent (i.e. $\exists n \in \{1, \dots, N(\ell)\}$ s.t. $\lim_{\ell \rightarrow \infty} \mu_n(\ell) = \nu_1$) if the following condition is satisfied:

$$\max_{n \in \{1, \dots, N(0)-1\}} |\mu_{n+1}(0) - \mu_n(0)| < 2\zeta_M \quad (19)$$

where ζ_M is a constant depending only on M .

Proof It is easy to check that the global maximum of $J'(\mu_n)$ is reached for $\mu_n = \nu_1, \forall \alpha_1$. Figure 2 shows the shape of $J'(\mu_n)$ for $\beta_n = 0$. For $\alpha_1 = 0$, $J'(\mu_n)$ reduces to a Fejér kernel of order M which has exactly one local maximum in the interval $[\nu_1 + k/M, \nu_1 + (k+1)/M]$, $k \neq 0$. Let J'_1 be the amplitude of the first sidelobe and $\nu_1 + \zeta_M$ be the value of μ_n such that $J'(\mu_n = \nu_1 + \zeta_M) = J'_1$ in the interval $[\nu_1, \nu_1 + 1/M]$ (we assume² that $M > 2$). For the dictionary refinement strategy to converge to the global maximum, it is sufficient to the sparse approximation algorithm to select, at a given level ℓ , an atom whose frequency satisfies $|\mu_{n^*}(\ell) - \nu_1| < \zeta_M < 1/M$, where $\mu_{n^*}(\ell) = \arg \max_n J'(\mu_n)$. Indeed, if $\mu_{n^*}(\ell) \in (\nu_1 - \zeta_M, \nu_1 + \zeta_M)$ then adding two atoms whose frequencies are located on both sides of $\mu_{n^*}(\ell)$ will lead to the selection, at level $\ell + 1$, of an atom that satisfies $|\mu_{n^*}(\ell + 1) - \nu_1| \leq |\mu_{n^*}(\ell) - \nu_1|$: the distance between the selected atom and the true frequency is a monotonically decreasing sequence. Finally, the convergence is guaranteed if the initial dictionary contains an atom n such that $|\mu_n(0) - \nu_1| < \zeta_M$, which is satisfied if

$$\max_{n \in \{1, \dots, N(0)-1\}} |\mu_{n+1}(0) - \mu_n(0)| < 2\zeta_M. \quad (20)$$

given the fact that the sequence $\{\mu_n(0)\}$ covers the interval $[0, 1)$. For $\alpha_1 < 0$, the main lobe of $J'(\mu_n)$ becomes broader and ζ_M larger than for $\alpha_1 = 0$. Consequently, condition (20) is also sufficient for $\alpha_1 < 0$. ■

Corollary 1 In the single tone case, the harmonic dictionary refinement is convergent if the initial frequency grid ($\ell = 0$) is the Fourier grid.

Proof Fourier bins are obtained for $N = M$ and $\mu_n(0) = (n-1)/M$. Since $\zeta_M > 1/2M$, the proof is straightforward. ■

It is important to note that condition (20) is sufficient but not necessary. Moreover, this condition is established when adding a single atom on both sides of the selected one (i.e. $\eta = 1$ in Algorithm 1). When $\eta \gg 1$, the condition may be relaxed and the rate of convergence is expected to be higher.

²The case of $M \leq 2$ is not of practical interest but the theorem is still valid by setting $\zeta_M = \frac{1}{2}$ because $J'(\mu_n)$ is a monotonically decreasing function in the interval $[\nu_1, \min\{\frac{1}{2}, \frac{1}{2} + \nu_1\}]$.

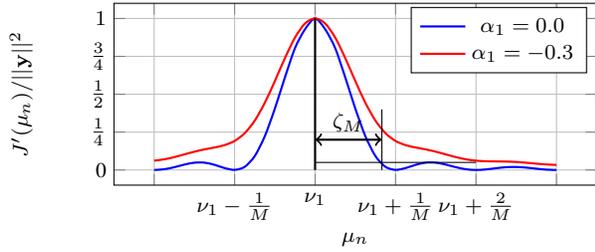


Figure 2: $J'(\mu_n)$ in the single mode case with $\beta_n = 0$

4.2 Estimating the Damping Factor: The Modal Dictionary

Assume that the previous sparse approximation using a harmonic dictionary has converged to select an atom with $\mu_n = \nu_1$. Now, we have to estimate the damping factor α_1 . We form a modal dictionary using the damping factor grid and the frequency ν_1 , i.e. $q_n = \exp(\beta_n + j2\pi\nu_1)$. Consequently,

$$J'(\beta_n) = \frac{|c_1|^2(1 - e^{2\beta_n})}{1 - e^{2\beta_n M}} \left(\frac{1 - e^{(\alpha_1 + \beta_n)M}}{1 - e^{(\alpha_1 + \beta_n)}} \right)^2. \quad (21)$$

Theorem 2 Let $y(m)$ be a single tone ($F = 1$) noiseless signal of length M and $\mathbf{Q}(0) = [\mathbf{q}_1 \ \mathbf{q}_2 \ \dots \ \mathbf{q}_{N(0)}]^\top$ be the initial modal dictionary formed using the frequency ν_1 , i.e., $q_n = \exp(\beta_n(0) + j2\pi\nu_1)$, where ν_1 is the frequency of signal \mathbf{y} . The columns are sorted in increasing order of $\beta_n(0)$, $n = 1, 2, \dots, N$ and covering the damping factor interval $(\beta_{\min}, 0]$. Then the refinement scheme is convergent (i.e. $\exists n$ s.t. $\lim_{\ell \rightarrow \infty} \beta_n(\ell) = \alpha_1$) if $\alpha_1 \in (\beta_{\min}, 0]$.

Proof Let $g(\beta_n)$ be the derivative of $J'(\beta_n)$ in (21) with respect to β_n . It is easy to check that $g(\beta_n) > 0$ for $\beta_n < \alpha_1$, $g(\beta_n) < 0$ for $\beta_n > \alpha_1$, and $g(\beta_n) = 0$ when $\beta_n = \alpha_1$. In other words, $J'(\beta_n)$ is monotonically increasing before the maximum reached at α_1 and monotonically decreasing after α_1 . Therefore, the multigrid algorithm converges to α_1 if $\beta_{\min} < \alpha_1$. ■

As a consequence of Theorem 2, the initial modal dictionary can be formed using only two points in the damping factor grid: $\beta_1(0) = \beta_{\min}$ and $\beta_2(0) = 0$.

We can now state that the multigrid algorithm based on two sparse approximations (for frequency and then damping factor) converges in the single tone case under some conditions. Note that in the noisy case when the SNR is sufficiently high, the convergence analysis is still valid as in the noiseless case, and the proposed multigrid sparse scheme for single tone converges to the global maximum of the Fejér kernel. The extension to the single tone R -D modal retrieval problem is straightforward and can be performed according to the formulation presented in Section 3.2. The details of this approach (*STSM: Single Tone Sparse Method*) are presented in Algorithm 2. The algorithm takes as input a noisy single tone R -D signal, and a couple of integers η_ν and η_α that correspond respectively to the number of frequency and damping factor atoms to be added on both sides of the corresponding

Algorithm 2: Single tone sparse method (STSM) based on a multigrid refinement

input : A tensor $\mathcal{Y} \in \mathbb{C}^{M_1 \times \dots \times M_R}$, $(\eta_\nu, \eta_\alpha) \in \mathbb{N} \times \mathbb{N}$
output: Parameters of the single R -D mode: a_1, \dots, a_R

initialization: $(k_\nu, k_\alpha) = (0, 0)$

initialize $\mathbf{d}_\nu^{(0)}$ and $\mathbf{d}_\alpha^{(0)}$ using ζ

for $r = 1 : R$ **do**

while halting criterion false **do**

$k_\nu = k_\nu + 1$

$\Omega_\nu^{(k_\nu)} = \text{SOMP}(\mathbf{Q}(\mathbf{d}_\nu^{(k_\nu)}), 0, \mathbf{Y}_{(r)}, \text{Iter} = 1)$

$\mathbf{d}_\nu^{(k_\nu+1)} = \text{DICREF}(\mathbf{d}_\nu^{(k_\nu)}, \Omega_\nu^{(k_\nu)}, \eta_\nu)$

end

while halting criterion false **do**

$k_\alpha = k_\alpha + 1$

$\Omega_\alpha^{(k_\alpha)} = \text{SOMP}(\mathbf{Q}(\mathbf{d}_\nu^{(k_\nu)}(\Omega_\nu^{(k_\nu)}), \mathbf{d}_\alpha^{(k_\alpha)}), \mathbf{Y}_{(r)})$

$\mathbf{d}_\alpha^{(k_\alpha+1)} = \text{DICREF}(\mathbf{d}_\alpha^{(k_\alpha)}, \Omega_\alpha^{(k_\alpha)}, \eta_\alpha)$

end

$a_r = \exp(\mathbf{d}_\alpha^{(k_\alpha)}(\Omega_\alpha^{(k_\alpha)}) + 2\pi \mathbf{d}_\nu^{(k_\nu)}(\Omega_\nu^{(k_\nu)}))$

end

return a_1, \dots, a_R

selected ones. Next, for each dimension $r = 1, \dots, R$, we run two tasks to estimate the frequency and then the damping factor: in each step we apply SOMP combined to DICREF using corresponding dictionaries and taking into account the convergence conditions discussed previously. Then parameters of a_r , i.e., ν_r and α_r , are given by the corresponding selected atoms.

5 Multiple R -D Modes Estimation

In the multiple tones case, sparse approximation algorithms yield suboptimal solutions when the coherence of the dictionary is high [32]. This is a crucial point because the refinement procedure will increase the coherence with increasing ℓ , which may prevent convergence even in the noiseless case. In the following, we present a low complexity algorithm that is accurate and robust in the presence of noise. The idea is to begin by an initialization step where F single tone modal signals of order $R - 1$ are extracted from the *multiple tones* R -D signal. Then an iterative technique is proposed to improve this decomposition and estimate more accurately the underlying parameters.

It is assumed that the frequencies are distinct in at least one dimension with $M_r > F$. Then dimensions are permuted such that the dimension with distinct frequencies becomes the first one ($r = 1$).

5.1 From Multiple Tones to Multiple Single-Tone Signals

According to (5), \mathcal{Y} can be written as

$$\mathcal{Y} = \mathcal{I}_{R+1, F} \underset{r=1}{\overset{R}{\bullet}} \mathbf{A}_r \underset{R+1}{\bullet} \mathbf{c}^\top \quad (22)$$

$$= \mathcal{S} \underset{1}{\bullet} \mathbf{A}_1 \quad (23)$$

where $\mathcal{I}_{R+1,F}$ is the diagonal tensor of order $R+1$ and size $F \times F \times \dots \times F$, containing ones on its diagonal, and

$$\mathcal{S} = \mathcal{I}_{R+1,F} \underset{r=2}{\bullet} \mathbf{A}_r \underset{R+1}{\bullet} \mathbf{c}^\top \quad (24)$$

is a complex tensor of order R and size $F \times M_2 \times \dots \times M_R$. Similar expressions are evoked in, among others, [11]. The new tensor \mathcal{S} can also be written as the concatenation of F tensors along the first dimension

$$\mathcal{S} = \mathcal{S}_1 \sqcup_1 \mathcal{S}_2 \sqcup_1 \dots \sqcup_1 \mathcal{S}_F \quad (25)$$

where each $\mathcal{S}_f, f = 1, \dots, F$ is a modal $(R-1)$ -D signal of size $1 \times M_2 \times \dots \times M_R$ containing a single $(R-1)$ -D tone:

$$\mathcal{S}_f = c_f \mathbf{a}_{f,2} \otimes \mathbf{a}_{f,3} \otimes \dots \otimes \mathbf{a}_{f,R}. \quad (26)$$

The singular value decomposition (SVD) of $\tilde{\mathbf{Y}}_{(1)}$ yields

$$\tilde{\mathbf{Y}}_{(1)} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^H \quad (27)$$

where matrices \mathbf{U} and \mathbf{V} contain respectively the left and right singular vectors of $\tilde{\mathbf{Y}}_{(1)}$, and $\mathbf{\Sigma}$ is a diagonal matrix containing the singular values $\sigma_i, i = 1, \dots, \min\{M_1, M_1'\}$ sorted in a decreasing order. As the number of components in \mathcal{Y} is equal to F , then an approximation of $\mathbf{Y}_{(1)}$, denoted by $\hat{\mathbf{Y}}_{(1)}$, can be obtained using the first F principal components of the SVD:

$$\hat{\mathbf{Y}}_{(1)} = \mathbf{U}_F \mathbf{\Sigma}_F \mathbf{V}_F^H \quad (28)$$

where \mathbf{U}_F (resp. \mathbf{V}_F) stands for the matrix formed with the first F columns of \mathbf{U} (resp. \mathbf{V}) and $\mathbf{\Sigma}_F = \text{diag}(\sigma_1, \dots, \sigma_F)$. It can be established from (7) and (28) that \mathbf{A}_1 and \mathbf{U}_F span the same subspace, and thus there exists an unknown nonsingular matrix \mathbf{T} that satisfies

$$\mathbf{A}_1 = \mathbf{U}_F \mathbf{T}. \quad (29)$$

Denote by $\underline{\mathbf{M}}$ (resp. $\overline{\mathbf{M}}$) the matrix obtained from \mathbf{M} by deleting the first (resp. last) row. By harnessing the Vandermonde structure of \mathbf{A}_1 , there exists a diagonal matrix \mathbf{D} such that $\underline{\mathbf{A}}_1 = \overline{\mathbf{A}}_1 \mathbf{D}$. Since $\underline{\mathbf{A}}_1 = \underline{\mathbf{U}}_F \mathbf{T}$ and $\overline{\mathbf{A}}_1 = \overline{\mathbf{U}}_F \mathbf{T}$, then $\underline{\mathbf{U}}_F \mathbf{T} = \overline{\mathbf{U}}_F \mathbf{T} \mathbf{D}$, which proves that matrix \mathbf{T} can be estimated by the eigenvectors of $\underline{\mathbf{U}}_F^\dagger \overline{\mathbf{U}}_F$.

Thereby \mathcal{S} can be estimated from the noisy data and $\hat{\mathbf{A}}_1$ using equation (23) as follows

$$\hat{\mathcal{S}} = \tilde{\mathcal{Y}} \underset{1}{\bullet} \hat{\mathbf{A}}_1^\dagger, \quad (30)$$

then $\hat{\mathcal{S}}_f, f = 1, \dots, F$ are extracted from $\hat{\mathcal{S}}$ according to (25). Each $\mathcal{Y}_f = c_f \mathbf{a}_{f,1} \otimes \dots \otimes \mathbf{a}_{f,R}$ can be estimated by $\tilde{\mathcal{Y}}_f^{(0)} = \hat{\mathcal{S}}_f \underset{1}{\bullet} \hat{\mathbf{a}}_{f,1}$. The sparse multigrid algorithm for single tone (STSM) can be applied on each $\tilde{\mathcal{Y}}_f^{(0)}, f = 1, \dots, F$ to estimate the parameters. However, we propose in the following to improve the separated components using an iterative technique.

5.2 Improving the Estimation Accuracy

It is clear from (30) that, in the noisy case, the error in estimating \mathcal{S} (due to the estimation of \mathbf{A}_1) will propagate when estimating the parameters $a_{f,2}, \dots, a_{f,R}$. Hence, we propose to improve iteratively the mode estimates. The following procedure is used to update estimates at each iteration $i = 0, \dots, K$

1. apply STSM to estimate $\mathbf{a}_{f,2}, \dots, \mathbf{a}_{f,R}, f = 1, \dots, F$

$$\{\hat{a}_{f,2}, \dots, \hat{a}_{f,R}\} = \text{STSM}(\tilde{\mathcal{Y}}_f^{(i)}, \eta_\nu, \eta_\alpha, r = 2, \dots, R) \quad (31)$$

2. estimate $c_f \mathbf{a}_{f,1}, f = 1, \dots, F$ by least squares using the already estimated $\mathbf{a}_{f,2}, \dots, \mathbf{a}_{f,R}, f = 1, \dots, F$

$$\widehat{c_f \mathbf{a}_{f,1}} = \tilde{\mathbf{Y}}_{f(1)}^{(i)} ((\hat{\mathbf{a}}_{f,R} \boxtimes \dots \boxtimes \hat{\mathbf{a}}_{f,2})^\top)^\dagger \quad (32)$$

3. compute $\hat{\mathcal{Y}}_f^{(i)}$

$$\hat{\mathcal{Y}}_f^{(i)} = \widehat{c_f \mathbf{a}_{f,1}} \otimes \hat{\mathbf{a}}_{f,2} \otimes \dots \otimes \hat{\mathbf{a}}_{f,R} \quad (33)$$

where $\tilde{\mathcal{Y}}_f^{(i)} = \hat{\mathcal{Y}}_f^{(i-1)} + \mathcal{R}_{f-1}^{(i)}, \mathcal{R}_{(f)}^{(i)} = \mathcal{R}_{f-1}^{(i)} + \hat{\mathcal{Y}}_f^{(i-1)} - \hat{\mathcal{Y}}_f^{(i)}, f = 1, \dots, F, \mathcal{R}_0^{(i)} \stackrel{\text{def}}{=} \mathcal{R}_F^{(i-1)}$, and $\mathcal{R}_F^{(0)} = \tilde{\mathcal{Y}} - \sum_{f=1}^F \hat{\mathcal{Y}}_f^{(0)}$. This iterative scheme will be analyzed in the next section.

Finally, the algorithm we propose (*MTM: Multiple Tones Method*) is summarized in Algorithm 3. Note that no association step of R -D modes is required. The initialization step consists in initializing: i) $\hat{\mathbf{A}}_1$ and $\hat{\mathcal{S}}$ using (29), (30) and (25), ii) the estimated single tones $\tilde{\mathcal{Y}}_f^{(0)}, f = 1, \dots, F$.

Note that the columns of $\hat{\mathbf{A}}_1$ are iteratively updated without extracting the related modes, whereas the modes of the other dimensions are extracted at each iteration using (31). Solely after the last iteration ($i = K$), the parameters of the first dimension are extracted using STSM algorithm. K denotes the maximum number of iterations, which is fixed to 2 in the simulations since no improvement was observed for $K > 2$.

5.3 Analysis of the Algorithm

Following the separation step described in (27)–(30), we can state that the algorithm yields the expected solution when the SNR is sufficiently high. We want now to prove that the second stage (next iterations), in addition to estimating the parameters from the single tones, is also improving the estimation accuracy. The general idea is inspired from greedy forward/backward sparse approximation, where the solution is refined by adding/removing atoms to/from the set of activated atoms. The improvement of the estimates is stated by the following theorem.

Theorem 3 *Assuming that the noise \mathcal{E} is sufficiently small such that the ordering of the singular values in $\mathbf{\Sigma}$ in (27) is the same as the ordering of the corresponding singular values*

Algorithm 3: Multiple tones method (MTM)

input : A tensor $\tilde{\mathbf{Y}} \in \mathbb{C}^{M_1 \times \dots \times M_R}$, $(\eta_\nu, \eta_\alpha) \in \mathbb{N} \times \mathbb{N}$
output: Parameters of the multiple R -D modes : $\{a_{f,r}\}_{f=1,r=1}^{F,R}$
initialization:
1. Compute $\hat{\mathbf{A}}_1$ and $\hat{\mathbf{S}}_f$, $f = 1, \dots, F$ using (29), (30) and (25)
2. $\tilde{\mathbf{Y}}_f^{(0)} = \hat{\mathbf{S}}_f \bullet \hat{\mathbf{a}}_{f,1}$, $f = 1, \dots, F$

For $f = 1, \dots, F$, compute $\hat{\mathbf{Y}}_f^{(0)}$ using (31), (32) and (33)
 $\mathcal{R}_F^{(0)} \stackrel{\text{def}}{=} \mathcal{R}_0^{(1)} = \tilde{\mathbf{Y}} - \sum_{f=1}^F \hat{\mathbf{Y}}_f^{(0)}$
for $i = 1 : K$ **do**
 for $f = 1 : F$ **do**
 $\hat{\mathbf{Y}}_f^{(i)} = \hat{\mathbf{Y}}_f^{(i-1)} + \mathcal{R}_{f-1}^{(i)}$
 compute $\hat{\mathbf{Y}}_f^{(i)}$ using (31), (32) and (33)
 $\mathcal{R}_f^{(i)} = \tilde{\mathbf{Y}}_f^{(i)} - \hat{\mathbf{Y}}_f^{(i)}$, if $f = F$, then $\mathcal{R}_0^{(i+1)} \stackrel{\text{def}}{=} \mathcal{R}_F^{(i)}$
 end
end
For $f = 1, \dots, F$, extract $a_{f,1}$ using
 $a_{f,1} = \text{STSM}(\hat{\mathbf{Y}}_f^{(K)} + \mathcal{R}_F^{(K)}, \eta_\nu, \eta_\alpha, r = 1)$
return $\{\hat{a}_{f,r}\}_{f=1,r=1}^{F,R}$

when $\mathcal{E} = 0$. Using the procedure expressed by (31), (32) and (33) to estimate \mathbf{Y}_f at iteration $i = 0, \dots, K$

$$\hat{\mathbf{Y}}_f^{(i)} = \arg \min_{\mathcal{X} \in \mathcal{H}} \|\tilde{\mathbf{Y}}_f^{(i)} - \mathcal{X}\| \quad (34)$$

where $\mathcal{H} = \{\mathcal{X} \in \mathbb{C}^{M_1 \times \dots \times M_R} | \mathcal{X} = \mathbf{b}_1 \otimes \mathbf{b}_2 \otimes \dots \otimes \mathbf{b}_R, \mathbf{b}_r \in \mathcal{P} \text{ for } r \neq 1\}$ with $\mathcal{P} = \{\mathbf{v} \in \mathbb{C}^{M_r} | \mathbf{v} = [1, v, \dots, v^{M_r-1}]^T, v = \exp(\beta + j\omega), \beta \in \mathbb{R}^-, \omega \in [0, 2\pi)\}$. Then, at each iteration i , the residual is decreased:

$$\|\tilde{\mathbf{Y}} - \hat{\mathbf{Y}}^{(i)}\| \leq \|\tilde{\mathbf{Y}} - \hat{\mathbf{Y}}^{(i-1)}\| \quad (35)$$

where $\hat{\mathbf{Y}}^{(i)} = \sum_{f=1}^F \hat{\mathbf{Y}}_f^{(i)}$.

Proof See A. ■

5.4 Identifiability

Based on the assumptions under which Algorithm 3 is operating, the identifiability condition can be stated as $F < M_1$ and $\min\{M_2, \dots, M_R\} \geq 2$. In [33], the condition is $M_r \geq 4$, $r = 1, \dots, R$, and $F \leq \lfloor \frac{M_1}{2} \rfloor \prod_{r=1}^R \lceil \frac{M_r}{2} \rceil$.

We note that, when $M_r \geq 4$, $r = 1, \dots, R$, the number of identifiable modes is slightly smaller than in [33], but the proposed algorithm is able to outperform the conventional methods in terms of computational complexity and accuracy. In addition, another advantage of the proposed algorithm is clear when the number of samples in one or more dimensions is less than 4 (i.e. $M_r < 4$), where identifiability in [33] is not satisfied. This latter case (i.e. $\exists r, M_r < 4$) can be encountered in signal processing applications when the size of one or more diversities (dimensions in our formulation problem) is less than 4.

6 Cramér-Rao Lower Bounds for R -D Cisoids in Noise

In this section, we derive the expressions of the CRLB for the parameters of R -D damped exponentials in Gaussian white noise. We then give the CRLB in the cases of single damped and undamped R -D cisoids. We consider the R -D sinusoidal model given in (3). Let

$$\boldsymbol{\theta} = [\omega_{1,1} \dots \omega_{1,R} \quad \omega_{2,1} \dots \omega_{F,R} \quad \alpha_{1,1} \dots \alpha_{1,R} \\ \alpha_{2,1} \dots \alpha_{F,R} \quad \lambda_1 \dots \lambda_F \quad \phi_1 \dots \phi_F]^T$$

be the unknown parameter vector. The aim here is to derive the CRLB of the parameters in $\boldsymbol{\theta}$.

The joint probability density function (pdf) of $\tilde{\mathbf{y}}$ is

$$p(\tilde{\mathbf{y}}; \boldsymbol{\theta}) = \frac{1}{(\sigma^2 \pi)^M} \exp \left\{ -\frac{1}{\sigma^2} (\tilde{\mathbf{y}} - \boldsymbol{\mu}(\boldsymbol{\theta}))^H (\tilde{\mathbf{y}} - \boldsymbol{\mu}(\boldsymbol{\theta})) \right\} \quad (36)$$

where $\boldsymbol{\mu}(\boldsymbol{\theta})$ is the noise-free part of \mathbf{y} and

$$\tilde{\mathbf{y}} = [\tilde{y}(1, \dots, 1, 1), \dots, \tilde{y}(1, \dots, 1, M_R), \tilde{y}(1, \dots, 2, 1), \\ \dots, \tilde{y}(1, \dots, 2, M_R), \dots, \tilde{y}(M_1, \dots, M_R)]^T. \quad (37)$$

The i -th entry of $\boldsymbol{\mu}(\boldsymbol{\theta})$ can be written as:

$$\boldsymbol{\mu}(\boldsymbol{\theta})_i = \sum_{f=1}^F c_f \prod_{r=1}^R a_{f,r}^{t_{i,r}}, \quad (38)$$

for $i = 1, \dots, M$, where

$$t_{i,r} = \left\lfloor \frac{i-1}{\prod_{\ell=r+1}^R M_\ell} \right\rfloor \text{ mod } M_r, \quad (39)$$

and $\lfloor \cdot \rfloor$ is the floor function. In the following, we derive the expressions of the CRLB in the general case ($F > 1$) and then we deduce the result corresponding to a single R -D modal signal ($F = 1$).

6.1 Derivation of the CRLB

Given the joint pdf in (36), the (k, l) entry of the Fisher information matrix is [34, 35]:

$$[\mathbf{F}(\boldsymbol{\theta})]_{kl} = \frac{2}{\sigma^2} \text{Re} \left\{ \left[\frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial \theta_k} \right]^H \frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial \theta_l} \right\}. \quad (40)$$

After some lengthy calculations, the $M \times (2RF + 2F)$ matrix $\partial \boldsymbol{\mu}(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}$ can be expressed as

$$\frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \underbrace{[j\mathbf{Z}'\boldsymbol{\Phi} \quad \mathbf{Z}'\boldsymbol{\Phi} \quad \mathbf{Z}\boldsymbol{\phi} \quad j\mathbf{Z}\boldsymbol{\phi}]}_{\mathbf{v}} \cdot \underbrace{\text{blkdiag}(\boldsymbol{\Lambda}, \boldsymbol{\Lambda}, \mathbf{I}_F, \boldsymbol{\lambda})}_{\mathbf{s}} \quad (41)$$

where

$$\mathbf{Z}' = [\mathbf{Z}'_1, \dots, \mathbf{Z}'_F] \in \mathbb{C}^{M \times RF}, \text{ with } \mathbf{Z}'_f(i, l) = t_{i,l} \prod_{r=1}^R a_{f,r}^{t_{i,r}}, \quad (42)$$

$$\mathbf{\Lambda} = \text{blkdiag}(\lambda_1 \mathbf{I}_R, \dots, \lambda_F \mathbf{I}_R) \in \mathbb{R}^{RF \times RF}, \quad (43)$$

$$\mathbf{\Phi} = \text{blkdiag}(e^{j\phi_1} \mathbf{I}_R, \dots, e^{j\phi_F} \mathbf{I}_R) \in \mathbb{C}^{RF \times RF}, \quad (44)$$

$$\mathbf{Z} = [\mathbf{z}_1, \dots, \mathbf{z}_F] \in \mathbb{C}^{M \times F}, \text{ with } \mathbf{z}_f(i) = \prod_{r=1}^R a_{f,r}^{t_{i,r}}, \quad (45)$$

$$\boldsymbol{\lambda} = \text{diag}([\lambda_1, \dots, \lambda_F]) \in \mathbb{R}^{F \times F}, \quad (46)$$

$$\boldsymbol{\phi} = \text{diag}([e^{j\phi_1}, \dots, e^{j\phi_F}]) \in \mathbb{C}^{F \times F}. \quad (47)$$

Finally, the inverse of the Fisher information matrix is

$$\mathbf{F}^{-1}(\boldsymbol{\theta}) = \frac{\sigma^2}{2} \mathbf{S}^{-1} [\text{Re}\{\mathbf{V}^H \mathbf{V}\}]^{-1} \mathbf{S}^{-1} = \frac{\sigma^2}{2} \mathbf{S}^{-1} \mathbf{W} \mathbf{S}^{-1} \quad (48)$$

where $\text{Re}\{\cdot\}$ stands for the real part. The CRLB of θ_k is given by $[\mathbf{F}^{-1}(\boldsymbol{\theta})]_{kk}$. More explicitly, for $f = 1, \dots, F$ and $r = 1, \dots, R$:

$$\text{CRLB}(\omega_{f,r}) = \frac{2\sigma^2 \mathbf{W}_{R(f-1)+r, R(f-1)+r}}{\lambda_f^2} \quad (49)$$

$$\text{CRLB}(\alpha_{f,r}) = \frac{2\sigma^2 \mathbf{W}_{RF+R(f-1)+r, RF+R(f-1)+r}}{\lambda_f^2} \quad (50)$$

$$\text{CRLB}(\lambda_f) = 2\sigma^2 \mathbf{W}_{2RF+f, 2RF+f} \quad (51)$$

$$\text{CRLB}(\phi_f) = \frac{2\sigma^2 \mathbf{W}_{2RF+F+f, 2RF+F+f}}{\lambda_f^2} \quad (52)$$

Theorem 4 For the general R -D exponential process, the CRLB's for $f = 1, \dots, F$ and $r = 1, \dots, R$ satisfy

$$\text{CRLB}(\omega_{f,r}) = \text{CRLB}(\alpha_{f,r}) \quad (53)$$

$$\text{CRLB}(\lambda_f) = \lambda^2 \text{CRLB}(\phi_f) \quad (54)$$

Proof It is based on the special block structure of matrix $\text{Re}\{\mathbf{V}^H \mathbf{V}\}$ (see for instance [34]).

6.2 Single Mode Case

In this section, the CRLB's will be simplified in the case of a single R -D modal signal ($F = 1$) to obtain more precise details on their parameter dependency. For the sake of simplicity, the subscripts denoting the mode $f = 1$ will be omitted. First, assume that $|a_r| = \exp(\alpha_r) < 1$. We shall express the products $\mathbf{Z}'^H \mathbf{Z}'$, $\mathbf{Z}^H \mathbf{Z}$ and $\mathbf{Z}'^H \mathbf{Z}$. After some calculations, we

get:

$$[\mathbf{Z}'^H \mathbf{Z}']_{nk} = \prod_{\substack{r=1 \\ r \neq n, k}}^R \left(\frac{1 - |a_r|^{2M_r}}{1 - |a_r|^2} \right) \times \begin{cases} \sum_{m=0}^{M_n-1} m |a_n|^{2m} \sum_{m=0}^{M_k-1} m |a_k|^{2m}, & \text{if } n \neq k \\ \sum_{m=0}^{M_n-1} m^2 |a_n|^{2m}, & \text{if } n = k \end{cases} \quad (55)$$

$$\mathbf{Z}^H \mathbf{Z} = \prod_{r=1}^R \left(\frac{1 - |a_r|^{2M_r}}{1 - |a_r|^2} \right) \quad (56)$$

$$[\mathbf{Z}'^H \mathbf{Z}]_n = \prod_{\substack{r=1 \\ r \neq n}}^R \left(\frac{1 - |a_r|^{2M_r}}{1 - |a_r|^2} \right) \times \sum_{m=0}^{M_n-1} m |a_n|^{2m}. \quad (57)$$

Denoting $M^{(\alpha)} = \prod_{r=1}^R (1 - |a_r|^{2M_r}) / (1 - |a_r|^2)$, $q_1(n) = \sum_{m=0}^{M_n-1} m |a_n|^{2m} / \sum_{m=0}^{M_n-1} |a_n|^{2m}$ and $q_2(n) = \sum_{m=0}^{M_n-1} m^2 |a_n|^{2m} / \sum_{m=0}^{M_n-1} |a_n|^{2m}$, we then obtain:

$$[\mathbf{P}]_{nk} = M^{(\alpha)} \times \begin{cases} q_1(n)q_1(k), & \text{if } n \neq k \\ q_2(n), & \text{if } n = k \end{cases} \quad (58)$$

$$\mathbf{G} = M^{(\alpha)} \quad (59)$$

$$[\mathbf{Q}]_n = M^{(\alpha)} q_1(n), \quad (60)$$

and

$$\text{Re}\{\mathbf{V}^H \mathbf{V}\} = \begin{bmatrix} \mathbf{P} & 0 & 0 & \mathbf{Q} \\ 0 & \mathbf{P} & \mathbf{Q} & 0 \\ 0 & \mathbf{Q}^T & \mathbf{G} & 0 \\ \mathbf{Q}^T & 0 & 0 & \mathbf{G} \end{bmatrix}. \quad (61)$$

The inversion of $\text{Re}\{\mathbf{V}^H \mathbf{V}\}$ yields the following expressions of the CRLB's:

$$\begin{aligned} \text{CRLB}(\omega_r) &= \text{CRLB}(\alpha_r) = \frac{\sigma^2}{2\lambda^2 M^{(\alpha)}} \\ &\times \frac{(1 - |a_r|^2)^2 (1 - |a_r|^{2M_r})^2}{-M_r^2 |a_r|^{2M_r} (1 - |a_r|^2)^2 + |a_r|^2 (1 - |a_r|^{2M_r})^2}, \end{aligned} \quad (62)$$

$$\frac{\text{CRLB}(\lambda)}{\lambda^2} = \text{CRLB}(\phi) = \frac{\sigma^2}{2\lambda^2 M^{(\alpha)}} \left(1 + \sum_{r=1}^R \frac{q_1^2(r)}{q_2(r) - q_1^2(r)} \right). \quad (63)$$

Finally, for a single R -D purely harmonic signal ($\alpha_r = 0, \forall r$), we have $M^{(\alpha)} = \prod_{r=1}^R M_r = M$ and taking the limit of the CRLB's when $\alpha_r \rightarrow 0$ leads to:

$$\lim_{\alpha_r \rightarrow 0} \text{CRLB}(\omega_r) = \frac{6\sigma^2}{\lambda^2 M (M_r^2 - 1)} \quad (64)$$

$$\lim_{\alpha_r \rightarrow 0} \frac{\text{CRLB}(\lambda)}{\lambda^2} = \frac{\sigma^2}{2\lambda^2 M} \left(1 + 3 \sum_{r=1}^R \frac{M_r - 1}{M_r + 1} \right). \quad (65)$$

Hence, for the undamped case, our result in (64) is consistent with [11].

7 Simulation Results

Numerical simulations have been carried out to assess the performances of the proposed method for 2-D and 3-D modal signals in the presence of white Gaussian noise. The performances are measured by the total root-mean square error (RMSE) on estimated parameters and the computational time. The total RMSE is defined as $\text{RMSE}_{\text{total}} = \sqrt{\frac{1}{RF} \mathbb{E}_p \left\{ \sum_{r=1}^R \sum_{f=1}^F (\xi_{f,r} - \hat{\xi}_{f,r})^2 \right\}}$ where $\hat{\xi}_{f,r}$ is an estimate of $\xi_{f,r}$, and \mathbb{E}_p is the average on p Monte-Carlo trials. In our simulations, $\xi_{f,r}$ can be either a frequency or a damping factor.

7.1 RMSE for 2-D and 3-D Signals

Experiment 1 To show the effectiveness of the multigrid scheme, this experiment presents the results obtained on Signal #1 (see Table 1) with different multigrid levels and different initial grids. Signal #1 is a single tone 2-D modal signal of size 10×10 . The number of multigrid levels is fixed to $L = 2$, i.e., $\ell = 0, 1, 2$. Then the results are presented as a function of the number of atoms in the initial dictionaries $N(0)$ and the number of atoms η_ν or η_α added at each level ℓ . The results we obtain for the first step, i.e., for the harmonic estimation, are presented in Figure 3. We can observe that the frequency RMSE obtained with the R -D sparse algorithm can reach the CRLB using a uniform initial harmonic dictionary of 10 atoms and $\eta_\nu = 31$ (Figure 3.a). Figure 3.b shows that the frequency RMSE is improved at low SNR if the initial dictionary contains 20 atoms, and reaches the optimal estimates with $\eta_\nu = 21$. Figure 4 shows the damping factor RMSE obtained by R -D sparse algorithm using different settings of the initial damping factor dictionary and η_α . We can observe that the damping factor RMSE depends on the number of atoms in the dictionary, the more atoms the better. At low SNR, the RMSE also depends β_{\min} . Therefore, it is better to choose β_{\min} with small absolute value if we have a prior knowledge of the interval of damping factors in the signal. In general, the estimation error is of order $\frac{1}{N(0)\eta^2}$. For instance, in the frequency step estimation, we recommend to chose $N(0)$ to be greater than or equal to $\frac{3}{2}M_r$ if we want a good accuracy at lower SNR levels. Otherwise, we can set $N(0) = M_r$. Once $N(0)$ is set, η can be chosen with respect to the desired accuracy. Let ε be the desired estimation error, then $\varepsilon = \frac{1}{N(0)\eta^2}$ and we can set $\eta = \frac{1}{\sqrt{\varepsilon N(0)}}$.

In the rest of this section, the proposed algorithms are compared with 2-D ESPRIT [6], Tensor-ESPRIT [10], PUMA [12] and TPUMA [11]. If the R -D signal contains one tone then Algorithm 2 (STSM) is used, otherwise Algorithm 3 (MTM) is used. Thus, to facilitate notation, both proposed algorithms, Algorithm 2 and Algorithm 3, will be called R -D sparse. For the proposed method, the initial grid used to build the harmonic dictionary is the same for all dimensions; it contains 50 frequency points uniformly distributed over the interval $[0, 1)$ and 10 damping factors $\beta \in [-0.05, 0]$. To

Table 1: 2-D and 3-D parameters of Signal #1 through #4

Signal	$\nu_{f,1}$	$\alpha_{f,1}$	$\nu_{f,2}$	$\alpha_{f,2}$	$\nu_{f,3}$	$\alpha_{f,3}$	c_f
#1	0.22	-0.011	0.34	-0.015	-	-	1
#2	0.40	-0.01	0.1	-0.01	0.1	-0.01	1
	0.20	-0.01	0.3	-0.15	0.25	-0.01	1
#3	0.28	-0.01	0.31	-0.01	0.22	-0.01	1
	0.12	-0.01	0.45	-0.015	0.11	-0.01	1
	0.20	-0.01	0.31	-0.01	0.11	-0.01	1
#4	0.30	-0.01	0.1	-0.01	0.1	-0.01	1
	0.13	-0.01	0.45	-0.015	0.4	-0.01	1
	0.20	-0.01	0.31	-0.01	0.1	-0.01	1
	0.42	-0.012	0.22	-0.01	0.32	-0.01	1

simulate a random dictionary, at each run, the frequency grid is perturbed by a small random quantity. As a consequence of experiment 1, we use the following settings $(L, \eta_\mu, \eta_\beta) = (2, 21, 11)$. The number of iterations in Algorithm 3 is set to $K = 2$ because no improvement was observed for $K > 2$.

Since the proposed method is applied directly on data without using spatial smoothing, i.e., it does not require the construction of a large matrix or an augmented order tensor, then a relevant comparison will be with algorithms that do not use spatial smoothing. Thereby, in the next experiments, the proposed algorithm is compared to PUMA [12] and TPUMA [11], which are algorithms that do not require spatial smoothing. We also report comparisons with 2-D ESPRIT [6] and Tensor-ESPRIT [10], which need spatial smoothing.

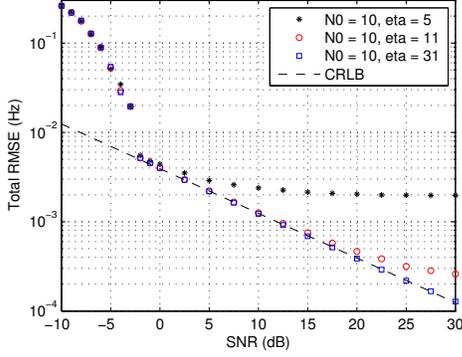
7.1.1 Single tone R -D modal signal

Experiment 2 This experiment tends to show the efficiency of the proposed algorithm in estimating parameters of single tone R -D modal signals. We use the same signal as before (Signal #1). Our R -D sparse algorithm is compared to 2-D ESPRIT [6] and PUMA [12]. For each level of noise, 1000 Monte-Carlo trials are performed. Figure 5 shows the obtained results. We can observe that: i) the proposed algorithm and PUMA reach the CRLB and outperform 2-D ESPRIT, ii) R -D sparse outperform PUMA in SNR less than 3 dB.

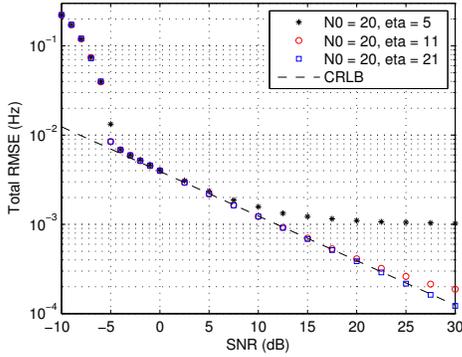
7.1.2 Multiple tones R -D modal signals

Several configurations are studied in the case of multiple tones to compare the proposed algorithm with Tensor-ESPRIT [10] and TPUMA [11]. These configurations (Experiments 3–4) are summarized in Table 2, in which the number of modes and the distance between frequencies in different dimensions are varied. Δ_{Fr} denotes the Rayleigh frequency resolution limit, which has the same value in all dimensions because $M_1 = M_2 = M_3$. In Experiment 5 we examine the case when the size of only one dimension is larger than 4, i.e., the identifiability condition of [33] is not satisfied. The parameters of the used signals are given in Table 1.

Experiment 3 In this experiment, we simulate a 3-D signal (Signal #2) of size $8 \times 8 \times 8$ and containing two modes whose



(a) $N(0) = 10$



(b) $N(0) = 20$

Figure 3: Frequency RMSE using R -D sparse algorithm with different η_ν . 2-D signal containing a single tone (Signal #1). $(M_1, M_2) = (10, 10)$. 1000 Monte-Carlo trials. (a) The initial harmonic dictionary contains $N(0) = 10$ atoms, (b) $N(0) = 20$ atoms.

Table 2: Different configurations for experiments 3 and 4

	F	Dim. 1	Dim. 2	Dim. 3
Exp. 3	2	$\Delta\nu > \Delta_{Fr}$	$\Delta\nu > \Delta_{Fr}$	$\Delta\nu > \Delta_{Fr}$
Exp. 4	3	$\Delta\nu < \Delta_{Fr}$	\exists identical modes	\exists identical modes

frequencies in each dimension are well separated. Parameters of the signal are given in Table 1. Figure 6 shows the obtained results. Here, the proposed method performs as TPUMA. Tensor-ESPRIT yields a slightly higher RMSE.

Experiment 4 3-D signal of size $10 \times 10 \times 10$ containing three 3-D modes (Signal #3). Note that there exists identical modes in two dimensions and frequencies in the first dimension are separated by less than $1/M_1$. The results are shown on Figure 7. In this experiment, the proposed R -D sparse approach performs better than TPUMA and Tensor-ESPRIT. Observe also that Tensor-ESPRIT outperforms TPUMA in this configuration (close frequencies and identical modes in dimensions 2 and 3).

Experiment 5 Results on Signal #4 of size $10 \times 3 \times 3$ containing 4 modes are given in Figure 8. We observe that the

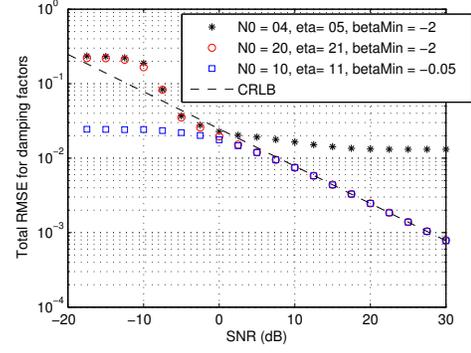


Figure 4: Damping factor RMSE using R -D sparse algorithm with different η_α , β_{\min} and $N(0)$. 2-D signal containing a single tone (Signal #1). $(M_1, M_2) = (10, 10)$. 1000 Monte-Carlo trials.

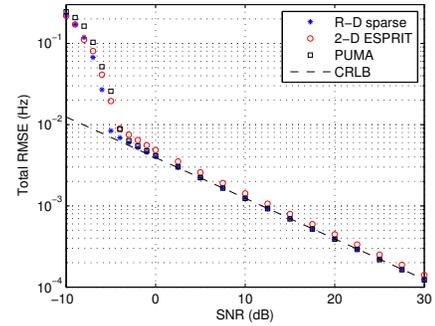


Figure 5: Frequency total root-mean square error for a 2-D signal containing a single tone (Signal #1). $(M_1, M_2) = (10, 10)$. 1000 Monte-Carlo trials.

proposed method outperforms TPUMA mainly at low SNR levels.

7.2 Numerical Complexity

It is known that in the case of 1-D signals of size M , OMP costs $O(NFM)$ in terms of multiplications [36]; F is the sparsity (number of components) and N is the number of atoms in the dictionary. For a M -measurements R -D signal, the complexity of the STSM algorithm over a set of L multigrid levels is $O(MNLR)$, assuming that the number of dictionary atoms is maintained constant (equal to N) over all levels. Regarding the approach proposed in Algorithm 3, the main operations are the call of STSM and the update of $\widehat{c}_f \widehat{\mathbf{a}}_{f,1} = \bar{\mathbf{Y}}_{f(1)}^{(i)} ((\hat{\mathbf{a}}_{f,R} \boxtimes \dots \boxtimes \hat{\mathbf{a}}_{f,2})^T)^\dagger$ which has a complexity of $O(M)$ since $((\hat{\mathbf{a}}_{f,R} \boxtimes \dots \boxtimes \hat{\mathbf{a}}_{f,2})^T)^\dagger$ is a row vector of length $\prod_{r=2}^R M_r$ and $\bar{\mathbf{Y}}_{f(1)}^{(i)}$ is a matrix of size $M_1 \times \prod_{r=2}^R M_r$. Therefore, the whole complexity of the proposed algorithm is $O((NL(F(R-1)K+1) + FK)M)$, which is linear in the number of measurements M . The complexity of the Tensor-ESPRIT algorithm with spatial smoothing is mainly related to that of the SVD which is at least $O(k_t F(R+1)PM)$ where k_t is a constant depending on the implementation of the SVD algorithm. Here $P = \prod_{r=1}^R P_r$

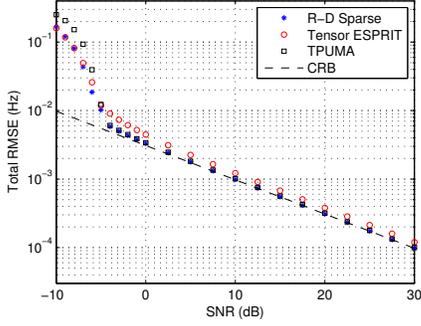


Figure 6: Frequency total root-mean square error for a 3-D signal containing two 3-D modes (Signal #2). $(M_1, M_2, M_3) = (8, 8, 8)$. 1000 Monte-Carlo.

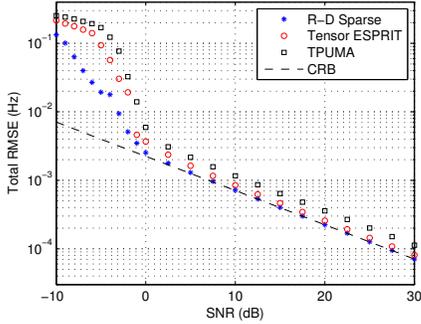


Figure 7: Frequency total root-mean square error for a 3-D signal containing 3 modes with identical modes in two dimensions (Signal #3), with close modes in the first dimension $(0.28, 0.12, 0.2)$. $(M_1, M_2, M_3) = (10, 10, 10)$. 200 Monte-Carlo.

where $\{P_r\}_{r=1}^R$ are design parameters used to get smoothed measurements (see [10]). The accuracy of the estimates provided by ESPRIT depends on these parameters. Since the optimal value for P_r is a fraction of M_r (e.g. [37, 38, 39]), the complexity of the SVD step is, in fact, $O(M^2)$. The complexities of PUMA and TPUMA algorithms are $O(M^3)$ and $O(k_t M(R+F-1) + \sum_{r=1}^R O(K(F+1)M_r^3))$, respectively. Compared to PUMA and TPUMA, the proposed algorithm has an attractive computational complexity for large size signals. Figure 9 shows the CPU time results of the proposed, Tensor-ESPRIT and TPUMA algorithms versus M_1 for a 3-D damped signal containing two modes with $M_2 = M_3 = 4$. We observe that the proposed method involves a low computational complexity compared to TPUMA and Tensor-ESPRIT when M_1 is large.

8 Conclusion

We presented an efficient sparse estimation approach for the analysis of multidimensional (R -D) damped or undamped modal signals. The idea consists in exploiting the simultaneous sparse approximation principle to separate this joint estimation problem into R multiple measurements problems. To be able to handle large size signals and yield accurate estimates, a multigrid dictionary refinement scheme is associated

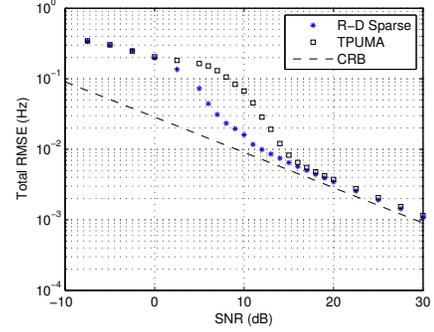


Figure 8: Frequency total root-mean square error for a 3-D signal containing 4 modes with $(M_1, M_2, M_3) = (10, 3, 3)$ (Signal #4). 200 Monte-Carlo.

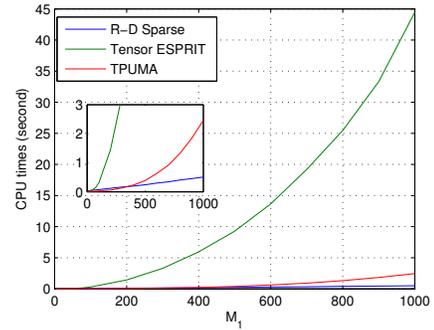


Figure 9: Average CPU time for a single run under $M_2 = M_3 = 4$ and $F = 2$.

with the simultaneous orthogonal matching pursuit (SOMP) algorithm. We gave the convergence proof of the the refinement procedure in the single tone case. Then, for the general multiple tones R -D case, the signal tensor model is decomposed in order to handle each tone separately in an iterative scheme so that the pairing of the R -D parameters is automatically achieved. Also, the CRLB of the R -D modal signal parameters were derived. The tests performed on simulated signals showed that the proposed algorithm attains the CRLB and outperforms state-of-the-art subspace algorithms. We also have shown that the complexity of the algorithm is linear with respect to the number of measurements, which allows the processing of large size signals. Finally, it is worth mentioning that this approach can be straightforwardly applied to other multidimensional array processing problems.

A Proof of Theorem 3

We begin the proof by introducing the following lemma.

Lemma 1 Consider $\tilde{\mathcal{Y}} = \mathcal{Y} + \Delta\mathcal{Y}$, where $\tilde{\mathcal{Y}}$ is the perturbed version of the data tensor \mathcal{Y} and $\Delta\mathcal{Y}$ is the perturbation. Assuming that $\Delta\mathcal{Y}$ is sufficiently small such that the ordering of the F singular values in Σ in (27) is the same as the ordering of the corresponding singular values when $\Delta\mathcal{Y} = 0$. Then the perturbation $\Delta\mathcal{Y}_f$ contains a linear combination of

all $\mathcal{Y}_f, f = 1, \dots, F$:

$$\Delta \mathcal{Y}_f = \mathcal{D}_f + \sum_{i=1}^F v_{f,i} \mathcal{Y}_i$$

where $\mathbf{v}_f^\top = [v_{f,1}, \dots, v_{f,F}] = \Delta \mathbf{A}_1^\dagger(f, :) \mathbf{A}_1$ and $\mathcal{D}_f = \Delta \mathcal{Y}_1 \bullet \mathbf{a}_{f,1} \mathbf{A}_1^\dagger(f, :) + \mathcal{Y}_{s,f} \bullet \Delta \mathbf{a}_{f,1}$.

Proof From (30) $\mathcal{S} = \mathcal{Y}_1 \bullet \mathbf{A}_1^\dagger$, we differentiate and obtain

$$\Delta \mathcal{S} = \Delta \mathcal{Y}_1 \bullet \mathbf{A}_1^\dagger + \mathcal{Y}_1 \bullet \Delta \mathbf{A}_1^\dagger$$

Then, we get $\Delta \mathcal{S}_f = \Delta \mathcal{Y}_1 \bullet \mathbf{A}_1^\dagger(f, :) + \sum_{p=1}^F v_{f,p} \mathcal{S}_p$. \mathcal{Y}_f is estimated using $\mathcal{Y}_f = \mathcal{S}_f \bullet \mathbf{a}_{f,1}$, we differentiate and obtain $\Delta \mathcal{Y}_f = \sum_{p=1}^F v_{f,p} \mathcal{Y}_p + \mathcal{S}_f \bullet \Delta \mathbf{a}_{f,1} + \Delta \mathcal{Y}_1 \bullet \mathbf{a}_{f,1} \mathbf{A}_1^\dagger(f, :)$. ■

Using the previous lemma

$$\begin{aligned} \bar{\mathcal{Y}}_f^{(0)} &= \mathcal{S}_f \bullet (\mathbf{a}_{f,1} + v_{f,f} \mathbf{a}_{f,1} + \Delta \mathbf{a}_f) + \\ &\quad \sum_{p=1, p \neq f}^F v_{f,p} \mathcal{Y}_p + \Delta \mathcal{Y}_1 \bullet \mathbf{a}_{f,1} \mathbf{A}_1^\dagger(f, :) \end{aligned}$$

Therefore, $\mathbf{a}_{f,2}, \dots, \mathbf{a}_{f,R}, f = 1, \dots, F$ can be estimated using STSM algorithm since

$$\begin{aligned} \bar{\mathbf{Y}}_{f(r)}^{(0)} &= c_f \mathbf{a}_{f,r} (\mathbf{a}_{f,R} \boxtimes \dots \boxtimes \mathbf{a}_{f,r+1} \boxtimes \mathbf{a}_{f,r-1} \boxtimes \\ &\quad \dots \boxtimes (\mathbf{a}_{f,1} + v_{f,f} \mathbf{a}_{f,1} + \Delta \mathbf{a}_{f,1})) \\ &\quad + \left(\sum_{p=1, p \neq f}^F v_{f,p} \mathcal{Y}_p + \Delta \mathcal{Y}_1 \bullet \mathbf{a}_{f,1} \mathbf{A}_1^\dagger(f, :) \right)_{(r)}. \end{aligned}$$

Since $\bar{\mathbf{Y}}_{f(1)}^{(0)}$ has the following form

$$\begin{aligned} \bar{\mathbf{Y}}_{f(1)}^{(0)} &= c_f (\mathbf{a}_{f,1} + v_{f,f} \mathbf{a}_{f,1} + \Delta \mathbf{a}_{f,1}) (\mathbf{a}_{f,R} \boxtimes \dots \boxtimes \mathbf{a}_{f,2}) \\ &\quad + \left(\sum_{p=1, p \neq f}^F v_{f,p} \mathcal{Y}_p + \Delta \mathcal{Y}_1 \bullet \mathbf{a}_{f,1} \mathbf{A}_1^\dagger(f, :) \right)_{(1)}, \end{aligned}$$

we estimate $c_f \mathbf{a}_{f,1}$ by least squares once $\mathbf{a}_{f,2}, \dots, \mathbf{a}_{f,R}$ are estimated using STSM

$$\begin{aligned} \widehat{c_f \mathbf{a}_{f,1}} &= \min_{\mathbf{a}} \|\bar{\mathcal{Y}}_f^{(0)} - \mathbf{a} \otimes \hat{\mathbf{a}}_{f,2} \otimes \dots \otimes \hat{\mathbf{a}}_{f,R}\| \\ &= \bar{\mathbf{Y}}_{f(1)}^{(0)} \left((\hat{\mathbf{a}}_{f,R} \boxtimes \dots \boxtimes \hat{\mathbf{a}}_{f,2})^\top \right)^\dagger \end{aligned}$$

So, we put $\hat{\mathcal{Y}}_f^{(0)} = \widehat{c_f \mathbf{a}_{f,1}} \otimes \hat{\mathbf{a}}_{f,2} \otimes \dots \otimes \hat{\mathbf{a}}_{f,R}$ and $\mathcal{R}_f = \bar{\mathcal{Y}}_f^{(0)} - \hat{\mathcal{Y}}_f^{(0)}$. Therefore, the procedure to estimate \mathcal{Y}_f at iteration $i = 0, \dots, K$ can be summarized in (31), (32) and (33). Note that this procedure is optimal because STSM and the least squares are optimal when they are used to estimate $\mathbf{a}_{f,2}, \dots, \mathbf{a}_{f,R}, f = 1, \dots, F$ and $c_f \mathbf{a}_{f,1}, f = 1, \dots, F$, respectively.

Now we present the technique for improving the estimation of \mathcal{Y}_f . Let $\mathcal{R}_F^{(0)} = \mathcal{R}_0^{(1)} = \tilde{\mathcal{Y}} - \sum_{f=1}^F \hat{\mathcal{Y}}_f^{(0)}$ and

$$\hat{\mathcal{Y}}_f^{(1)} = \arg \min_{\mathcal{X} \in \mathcal{H}} \|\hat{\mathcal{Y}}_f^{(0)} + \mathcal{R}_{f-1}^{(1)} - \mathcal{X}\| \quad (66)$$

where $\mathcal{R}_f^{(1)} = \hat{\mathcal{Y}}_f^{(0)} + \mathcal{R}_{f-1}^{(1)} - \hat{\mathcal{Y}}_f^{(1)}, f = 1, \dots, F$, and $\hat{\mathcal{Y}}_f^{(i)}$ is an improved estimate of \mathcal{Y}_f . We follow the same procedure as described in equations (31), (32) and (33) to calculate $\hat{\mathcal{Y}}_f^{(1)}$.

We can state that there is improvement in the estimation of \mathcal{Y}_f if

$$\left\| \tilde{\mathcal{Y}} - \sum_{f=1}^F \hat{\mathcal{Y}}_f^{(1)} \right\| \leq \left\| \tilde{\mathcal{Y}} - \sum_{f=1}^F \hat{\mathcal{Y}}_f^{(0)} \right\| \quad (67)$$

We have $\|\mathcal{R}_0^{(1)}\| = \|\bar{\mathcal{Y}}_1^{(0)} - \hat{\mathcal{Y}}_1^{(0)} + \sum_{f=2}^F \mathcal{R}_f + \mathcal{V}\|$ where $\mathcal{V} = \tilde{\mathcal{Y}} - \bar{\mathcal{Y}}$ and $\bar{\mathcal{Y}} = \sum_{f=1}^F \bar{\mathcal{Y}}_f^{(0)}$. It can be verified that

$$\|\mathcal{R}_f^{(1)}\| = \left\| \left(\bar{\mathcal{Y}}_f^{(0)} + \sum_{p=1}^{f-1} (\bar{\mathcal{Y}}_p^{(0)} - \hat{\mathcal{Y}}_p^{(1)}) + \sum_{p=f+1}^F \mathcal{R}_p + \mathcal{V} \right) - \hat{\mathcal{Y}}_f^{(1)} \right\|$$

$$\|\mathcal{R}_{f-1}^{(1)}\| = \left\| \left(\bar{\mathcal{Y}}_f^{(0)} + \sum_{p=1}^{f-1} (\bar{\mathcal{Y}}_p^{(0)} - \hat{\mathcal{Y}}_p^{(1)}) + \sum_{p=f+1}^F \mathcal{R}_p + \mathcal{V} \right) - \hat{\mathcal{Y}}_f^{(0)} \right\|$$

However, from equation (66), $\hat{\mathcal{Y}}_f^{(1)}$ is the minimizer with respect to $\mathcal{X} \in \mathcal{H}$ of

$$\left\| \left(\bar{\mathcal{Y}}_f^{(0)} + \sum_{p=1}^{f-1} (\bar{\mathcal{Y}}_p^{(0)} - \hat{\mathcal{Y}}_p^{(1)}) + \sum_{p=f+1}^F \mathcal{R}_p \right) - \mathcal{X} \right\|$$

Therefore, $\|\mathcal{R}_f^{(1)}\| \leq \|\mathcal{R}_{f-1}^{(1)}\|, f = 1, \dots, F$. As consequence, $\|\mathcal{R}_F^{(1)}\| \leq \|\mathcal{R}_F^{(0)}\|$, which we are seeking in expression (67). Similarly, we can prove that $\|\mathcal{R}_F^{(i)}\| \leq \|\mathcal{R}_F^{(i-1)}\|, i > 1$, using the general forms of $\mathcal{R}_f^{(i)}$ and $\mathcal{R}_{f-1}^{(i)}$

$$\mathcal{R}_f^{(i)} = \left(\bar{\mathcal{Y}}_f^{(0)} + \sum_{p=1}^{f-1} (\bar{\mathcal{Y}}_p^{(0)} - \hat{\mathcal{Y}}_p^{(i)}) + \sum_{p=f+1}^{f-1} (\bar{\mathcal{Y}}_p^{(0)} - \hat{\mathcal{Y}}_p^{(i-1)}) + \mathcal{V} \right) - \hat{\mathcal{Y}}_f^{(i)}$$

$$\mathcal{R}_{f-1}^{(i)} = \left(\bar{\mathcal{Y}}_f^{(0)} + \sum_{p=1}^{f-1} (\bar{\mathcal{Y}}_p^{(0)} - \hat{\mathcal{Y}}_p^{(i)}) + \sum_{p=f+1}^{f-1} (\bar{\mathcal{Y}}_p^{(0)} - \hat{\mathcal{Y}}_p^{(i-1)}) + \mathcal{V} \right) - \hat{\mathcal{Y}}_f^{(i-1)}$$

which we are seeking in (35).

Acknowledgment

The authors would like to thank PhD Weize Sun for providing them with the TPUMA algorithm code.

References

References

- [1] P. Stoica and R. Moses, *Introduction to Spectral Analysis*. Upper Saddle River, NJ: Prentice Hall, 1997.
- [2] A. B. Gershman and N. D. Sidiropoulos, *Space-time processing for MIMO communications*. Wiley Online Library, 2005.
- [3] D. Nion and S. Sidiropoulos, "Tensor algebra and multidimensional retrieval in signal processing for MIMO radar," *IEEE Trans. Signal Process.*, vol. 58, no. 1, pp. 5693–5705, 2010.
- [4] J. Sacchini, W. Steedly, and R. Moses, "Two-dimensional Prony modeling and parameter estimation," *IEEE Trans. Signal Process.*, vol. 41, no. 11, pp. 3127–3137, 1993.
- [5] Y. Hua, "Estimating two-dimensional frequencies by matrix enhancement and matrix pencil," *IEEE Trans. Signal Process.*, vol. 40, no. 9, pp. 2267–2280, 1992.
- [6] S. Rouquette and M. Najim, "Estimation of frequencies and damping factors by two-dimensional ESPRIT type methods," *IEEE Trans. Signal Process.*, vol. 49, no. 1, pp. 237–245, 2001.
- [7] K. Mokios, N. Sidiropoulos, M. Pesavento, and C. Mecklenbrauker, "On 3-D harmonic retrieval for wireless channel sounding," in *Proc. IEEE ICASSP*, Montreal, Canada, May 2004, pp. ii89–ii92.
- [8] J. Liu and X. Liu, "An eigenvector-based approach for multidimensional frequency estimation with improved identifiability," *IEEE Trans. Signal Process.*, vol. 54, no. 12, pp. 4543–4556, December 2006.
- [9] J. Liu, X. Liu, and X. Ma, "Multidimensional frequency estimation with finite snapshots in the presence of identical frequencies," *IEEE Trans. Signal Process.*, vol. 55, pp. 5179–5194, 2007.
- [10] M. Haardt, F. Roemer, and G. Del Galdo, "Higher-order SVD-based subspace estimation to improve the parameter estimation accuracy in multidimensional harmonic retrieval problems," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 3198–3213, 2008.
- [11] W. Sun and H. C. So, "Accurate and computationally efficient tensor-based subspace approach for multi-dimensional harmonic retrieval," *IEEE Trans. Signal Process.*, vol. 60, no. 10, pp. 5077–5088, October 2012.
- [12] H. So, F. Chan, W. Lau, and C. Chan, "An efficient approach for two-dimensional parameter estimation of a single-tone," *IEEE Trans. Signal Process.*, vol. 58, no. 4, pp. 1999–2009, April 2010.
- [13] L. Huang, Y. Wu, H. So, Y. Zhang, and L. Huang, "Multidimensional sinusoidal frequency estimation using subspace and projection separation approaches," *IEEE Trans. Signal Process.*, vol. 60, no. 10, pp. 5536–5543, 2012.
- [14] C. Lin and W. Fang, "Efficient multidimensional harmonic retrieval: A hierarchical signal separation framework," *IEEE Signal Process. Lett.*, vol. 20, no. 5, pp. 427–430, May 2013.
- [15] S. Sahnoun, E.-H. Djermoune, and D. Brie, "Sparse multigrid modal estimation: Initial grid selection," in *Proc. European Signal Process. Conf. (EUSIPCO-2012)*, 2012, pp. 450–454.
- [16] M. Goodwin and M. Vetterli, "Matching pursuit and atomic signal models based on recursive filter banks," *IEEE Trans. Signal Process.*, vol. 47, no. 7, pp. 1890–1902, 1999.
- [17] D. Malioutov, M. Cetin, and A. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 3010–3022, 2005.
- [18] P. Stoica, P. Babu, and J. Li, "SPICE: A sparse covariance-based estimation method for array processing," *IEEE Trans. Signal Process.*, vol. 59, no. 2, pp. 629–638, 2011.
- [19] S. Sahnoun, E.-H. Djermoune, C. Soussen, and D. Brie, "Sparse multidimensional modal analysis using a multigrid dictionary refinement," *EURASIP J. Adv. Signal Process.*, March 2012.
- [20] S. Sahnoun, E.-H. Djermoune, and D. Brie, "Sparse modal estimation of 2-D NMR signals," in *Proc. IEEE ICASSP*, Vancouver, Canada, May 2013, pp. 8751–8755.
- [21] J. Sward, S. I. Adalbjornsson, and A. Jakobsson, "High resolution sparse estimation of exponentially decaying signals," in *Proc. IEEE ICASSP*. IEEE, 2014, pp. 7203–7207.
- [22] S. I. Adalbjörnsson, J. Swärd, and A. Jakobsson, "High resolution sparse estimation of exponentially decaying two-dimensional signals," in *Proc. European Signal Process. Conf. (EUSIPCO)*. EURASIP, 2014.
- [23] G. Tang, B. N. Bhaskar, P. Shah, and B. Recht, "Compressed sensing off the grid," *Information Theory, IEEE Transactions on*, vol. 59, no. 11, pp. 7465–7490, 2013.
- [24] Z. Tan, "Discrete and continuous sparse recovery methods and their applications," Ph.D. dissertation, Washington University in St. Louis, 2015.
- [25] J. Tropp, A. Gilbert, and M. Strauss, "Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit," *Signal Process.*, vol. 86, pp. 572–588, 2006.
- [26] R. Boyer, "Deterministic asymptotic Cramér–Rao bound for the multidimensional harmonic model," *Signal Processing*, vol. 88, no. 12, pp. 2869–2877, 2008.
- [27] M. Clark and L. Scharf, "Two-dimensional modal analysis based on maximum likelihood," *IEEE Trans. Signal Process.*, vol. 42, no. 6, pp. 1443–1451, 1994.
- [28] P. Comon, "Tensors: a brief introduction," *IEEE Sig. Proc. Magazine*, vol. 31, no. 3, May 2014, special issue on BSS. hal-00923279.
- [29] T. Kolda and B. Bader, "Tensor decompositions and applications," *SIAM Review*, vol. 51, no. 3, pp. 455–500, 2009.
- [30] A. Rakotomamonjy, "Surveying and comparing simultaneous sparse approximation (or group-Lasso) algorithms," *Signal Process.*, vol. 91, no. 7, pp. 1505–1526, 2011.
- [31] M. F. Duarte and R. G. Baraniuk, "Spectral compressive sensing," *Applied and Computational Harmonic Analysis*, vol. 35, no. 1, pp. 111–129, 2013.
- [32] J. Tropp, "Greed is good: algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory*, vol. 50, pp. 2231–2242, Oct 2004.
- [33] T. Jiang, N. D. Sidiropoulos, and J. M. Ten Berge, "Almost-sure identifiability of multidimensional harmonic retrieval," *IEEE Trans. Signal Process.*, vol. 49, no. 9, pp. 1849–1859, 2001.
- [34] Y.-X. Yao and S. Pandit, "Cramér–Rao lower bounds for a damped sinusoidal process," *IEEE Trans. Signal Process.*, vol. 43, no. 4, pp. 878–885, 1995.

- [35] S. Kay, *Fundamentals of statistical signal processing. Estimation theory*. Englewood Cliffs, New Jersey: Prentice Hall International Editions, 1993.
- [36] J. Tropp and S. Wright, "Computational methods for sparse solution of linear inverse problems," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 948–958, 2010.
- [37] A. Lemma, A.-J. van der Even, and E. Deprettere, "Analysis of joint angle-frequency estimation using ESPRIT," *IEEE Trans. Signal Process.*, vol. 51, pp. 1264–1283, May 2003.
- [38] E.-H. Djermoune and M. Tomczak, "Perturbation analysis of subspace-based methods in estimating a damped complex exponential," *IEEE Trans. Signal Process.*, vol. 57, no. 11, pp. 4558–4563, 2009.
- [39] Y. Hua and T. Sarkar, "Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 38, pp. 814–824, 1990.