

# Spectrum estimation using frequency shifting and decimation

ISSN 1751-9675

Received on 27th June 2019

Revised 18th November 2019

Accepted on 12th December 2019

E-First on 30th January 2020

doi: 10.1049/iet-spr.2019.0306

www.ietdl.org

Raymundo Albert<sup>1,2</sup> ✉, Cecilia Galarza<sup>1,2</sup><sup>1</sup>Facultad de Ingeniería, Universidad de Buenos Aires, Paseo Colón 850, Buenos Aires, Argentina<sup>2</sup>Centro de Simulación Computacional, CONICET, Godoy Cruz 2390, Buenos Aires, Argentina

✉ E-mail: araymundo@fi.uba.ar

**Abstract:** Parametric spectral estimation techniques are widely used to estimate the parameters of sums of complex sinusoids corrupted by noise. In this work, the authors show that the numerical stability of the estimated frequencies not only depends on the size of the amplitudes associated with the real frequencies, but also to the distance among frequencies. Therefore, for closely spaced frequencies, the estimates are vulnerable to large deviate from their true values. To overcome this problem, they propose a strategy to artificially increase the frequency separation by downsampling the baseband equivalent of the noisy signal before applying a spectral estimation technique. This methodology significantly improves the estimation performance especially in the low signal to noise ratio regime. The performance of the technique is assessed in terms of the root mean square error and it is compared to results obtained in previous publications.

## 1 Introduction

Problems dealing with a mixture of closely spaced damped sinusoidal signals in a noisy environment are regularly observed in engineering. This is a problem of major significance in applications such as seismic exploration, communication, radar, and sonar, to name but a few. In those cases, the signal of relevance contains unknown parameters such as amplitude, phase, and frequency of each mode. This path led in the past to developing parametric spectral estimation techniques, in particular, the family of subspace methods. These algorithms generate frequency component estimates of a given signal based on the decomposition of an observation vector space into two subspaces: one associated with the signal and the other with the noise [1, 2]. A major concern when developing these techniques was to cope with noise and model uncertainties, in particular, the unknown order of the model that corresponds to the dimension of the signal subspace. Different strategies for obtaining this magnitude have been worked in the past [3, 4]. Lately, novel optimisation techniques have been proposed to estimate the model order. This method, which relies on Kronecker's theorem for Hankel operators, was used to formulate an appropriate non-linear least squares problem in terms of a rank constraint on the Hankel matrix associated with the observed signal [5].

Given the model order, the frequencies making up the observed signal are estimated on a second step. A largely used technique, known as ESPRIT (Estimation of Signal Parameters by Rotational Invariance Technique) [6], exploits the structure of the signal space and it formulates a generalised eigenvalue problem to obtain the frequency estimates. An alternative approach, known as the Matrix Pencil Method (MPM), has its roots in the Vandermonde decomposition of the Hankel matrix built from the observed signal [7]. A more recent approach based on the Hankel matrix associated with the observation was proposed in [8], where the parameter estimation is performed by minimising the nuclear norm of the Hankel matrix.

A different path was followed in [9], where the authors considered sums of undamped complex exponentials and they exploited a sparse description for them. In this case, using a previously defined set of discrete frequencies, compressed sensing techniques are used to estimate the unknown parameters. An improvement to this approach was presented in [10, 11]. There, the solution is obtained by minimising the atomic norm of the

estimated signal. It was shown that, in this case, the exact recovery of the observed signal is guaranteed when the frequencies are adequately separated [12]. Unfortunately, these methods are not suitable for a sum of damped complex exponentials.

The numerical conditioning of the estimates when recovering a sum of undamped complex exponentials was analysed in [13]. In this work, an upper bound of the condition number of the related Vandermonde matrix was provided, and a decimative approach for a single cluster of frequencies was proposed. As it was pointed out in [14] the performance of spectrum estimation methods such as ESPRIT, MPM, etc., degrades for closely spaced complex exponentials. To overcome this problem, the frequency separation was artificially increased by decimating the signal and then applying ESPRIT. In [15], a decimation approach is applied, and a total least squares problem is solved. However, in [14, 15], the downsampling factor is chosen to make sure that no aliasing is introduced. To overcome this restriction, the original signal was oversampled.

On a different angle, the authors of [16] analyse the spectral content of geophysical time series by frequency shifting and later filtering the observed signal. By doing so, they were able to identify the resonant frequencies contained in the observed signal.

Nevertheless, accurate estimation of spectral parameters for sums of damped complex exponentials is performed via the solution of generalised eigenvalue problems. These eigenvalue problems are usually ill-conditioned and non-square. Solving generalised eigenvalue problems of these characteristics is a challenge from an algorithmic perspective. This difficulty turns to be harder when the number of frequencies in the original mixture is very large and the frequencies are clustered in small regions of the complex plane [17, 18]. Our focus is to analyse the behaviour of the frequency estimation step when the observed data is subject to small perturbations. In particular, we study the numerical stability of the generalised eigenvalue problem that is built from the Hankel matrix associated to the observed signal. A related problem has been addressed before [19, 20]. In particular, it has been observed that the sensitivity of each eigenvalue is inversely proportional to the inner product of the left and right generalised eigenvectors weighted by the Hankel matrix. This inner product is proportional to the amplitude of the damped oscillation corresponding to the analysed eigenvalue.

In this paper, we extend these results and we show that the first-order approximation of the perturbed eigenvalues not only depends

on the amplitudes associated with the eigenvalues, but also on the distance among eigenvalues. Therefore, the eigenvalues that are close to each other are prone to exhibit large deviations from their actual values, even when the observed signal is only lightly perturbed. Similar results were obtained for a different context by the authors of [21].

Based on the analytical results, we propose a strategy to pre-processed the observed signal using multi-rate processing. This has the ability to accommodate the complex frequencies for more stable computation. The idea is validated with numerical experiments and compared to the results obtained in previous publications.

The paper is organised as follows. In Section 2, we make a brief introduction to subspace methods, in particular, ESPRIT and MPM. In Section 3, we study the numerical stability of the non-square generalised eigenvalue problem and we obtain a new bound for the first-order derivative of the generalised eigenvalue. A method to obtain better numerical stability and improved frequency estimation is presented in Section 4. Finally, numerical experiments are presented in Section 5 and conclusions are given Section 6.

Throughout the paper, we use the standard notation: the lower case ( $\mathbf{v}$ ) for scalars, boldface lower case ( $\mathbf{v}$ ) for vectors, uppercase bold face  $\mathbf{A}$  for matrices. The  $k$ th entry of a vector  $\mathbf{v}$  is denoted  $(\mathbf{v})_k$ . Given a matrix  $\mathbf{A}$ , we denote its transpose, Hermitian, and Moore–Penrose pseudo-inverse as  $\mathbf{A}^T$ ,  $\mathbf{A}^H$ ,  $\mathbf{A}^\dagger$ , respectively.  $[\mathbf{A}]_{ij}$  refers to the element in row  $i$  and column  $j$  of a matrix  $\mathbf{A}$ .  $\|\mathbf{A}\|_F$  refers to the Frobenius norm of the matrix  $\mathbf{A}$ ,  $\|\mathbf{A}\|$  is reserved for the induced 2-norm. Finally, the notation  $\mathbf{I}_m$  is used for the  $m \times m$  identity matrix.

## 2 Spectrum estimation methods

### 2.1 Background

Spectrum estimation techniques are usually treated as a two-step procedure. In the first step, the model order is estimated and the observed signal is cleaned up from the perturbing noise. In the second step, the values for the complex frequencies and their amplitudes are obtained [2]. This paper is focused on the analysis of this second step and how sensitive the estimated complex frequencies are when the observed signal is subject to small perturbations. To begin with, the description of the traditional techniques, consider a noiseless discrete-time signal composed by a sum of damped complex exponential as follows:

$$x_k = \sum_{i=1}^n c_i z_i^k \quad k = 0, 1, \dots, \quad (1)$$

where  $n$  is the number of exponential terms,  $z_i \in \mathbb{C}$  is a complex resonant frequency, and  $c_i \in \mathbb{C}$  the amplitude associated to it. We assume that  $z_i$  are all mutually distinct and  $c_i \neq 0$  for all  $i$ . In the first approach, we assume that  $n$  is known. This assumption is revised in Section 2.2.

Given  $m \geq n > 0$ , define the following  $((m+1) \times n)$ -complex Hankel matrix

$$\mathbf{H} = \begin{bmatrix} x_0 & x_1 & \cdots & x_{n-1} \\ x_1 & x_2 & \cdots & x_n \\ \vdots & \vdots & \ddots & \vdots \\ x_m & x_{m+1} & \cdots & x_{m+n-1} \end{bmatrix} \quad (2)$$

and the complex vector  $\mathbf{x}_n = [x_n \cdots x_{m+n}]^T$ . The solution to the linear system of equations

$$\mathbf{H}\mathbf{q} = -\mathbf{x}_n, \quad (3)$$

is a complex vector  $\mathbf{q} = [q_0 \cdots q_{n-1}]^T$  whose entries are the coefficients of the polynomial

$q(z) = \prod_{i=1}^n (z - z_i) = z^n + \sum_{k=0}^{n-1} q_k z^k$ . This simple setup is the starting point for classical Prony's method [22]. Unfortunately,  $\mathbf{H}$  is an ill-conditioned matrix, and solving (3) for  $\mathbf{q}$  could be very sensitive to noise in the data. Moreover, obtaining  $z_i$  as the roots of  $q(z)$  is also a challenging task to perform when (1) is perturbed by noise.

More robust approaches have been proposed by exploiting the Vandermonde decomposition of  $\mathbf{H}$ . Let us define the  $((l+1) \times n)$ -Vandermonde matrix as

$$\mathbf{Z}_l = \begin{bmatrix} 1 & \cdots & 1 \\ z_1 & \cdots & z_n \\ \vdots & & \vdots \\ z_1^l & \cdots & z_n^l \end{bmatrix}. \quad (4)$$

Then, the Vandermonde decomposition of  $\mathbf{H}$  is

$$\mathbf{H} = \mathbf{Z}_m \mathbf{D} \mathbf{Z}_{n-1}^T, \quad (5)$$

where  $\mathbf{D} = \text{diag}(c_1, \dots, c_n)$ . Consider the singular value decomposition of  $\mathbf{H} = \mathbf{P} \mathbf{S} \mathbf{Q}^H$ , where the columns of  $\mathbf{P}$  span the range of  $\mathbf{H}$  and they are identified with the signal subspace in  $\mathbb{C}^{m+1}$ . Since  $z_i$  are distinct and  $m \geq n$ ,  $\text{rank}(\mathbf{Z}_m) = \text{rank}(\mathbf{Z}_{n-1}) = n$ , and from (5), we see that the columns of  $\mathbf{Z}_m$  also span the signal subspace. Then, there exists an invertible matrix  $\mathbf{G} \in \mathbb{R}^{n \times n}$  such that

$$\mathbf{Z}_m = \mathbf{P} \mathbf{G}. \quad (6)$$

Define now the following  $(m \times n)$  complex matrices:

$$\begin{aligned} \mathbf{Z}_{m,l} &= \begin{bmatrix} \mathbf{I}_{m-1} \\ \mathbf{0}_{1 \times m-1} \end{bmatrix}^T \mathbf{Z}_m, & \mathbf{Z}_{m,f} &= \begin{bmatrix} \mathbf{0}_{1 \times m-1} \\ \mathbf{I}_{m-1} \end{bmatrix}^T \mathbf{Z}_m, \\ \mathbf{P}_l &= \begin{bmatrix} \mathbf{I}_{m-1} \\ \mathbf{0}_{1 \times m-1} \end{bmatrix}^T \mathbf{P}, & \mathbf{P}_f &= \begin{bmatrix} \mathbf{0}_{1 \times m-1} \\ \mathbf{I}_{m-1} \end{bmatrix}^T \mathbf{P}. \end{aligned} \quad (7)$$

Matrices  $\mathbf{Z}_{m,f}$  and  $\mathbf{Z}_{m,l}$  are obtained from  $\mathbf{Z}_m$  by deleting its first and last rows, respectively. Analogously, we define  $\mathbf{P}_f$  and  $\mathbf{P}_l$  from  $\mathbf{P}$ . Using these definitions, it is easy to verify the rotational invariance property that says that

$$\mathbf{Z}_{m,f} = \mathbf{Z}_{m,l} \mathbf{Z}, \quad (8)$$

where  $\mathbf{Z} = \text{diag}(z_1, \dots, z_n)$ . By replacing (6) in (7), we have that  $\mathbf{Z}_{m,l} = \mathbf{P}_l \mathbf{G}$  and  $\mathbf{Z}_{m,f} = \mathbf{P}_f \mathbf{G}$ . Therefore, we obtain from (8) that

$$\mathbf{P}_f \mathbf{G} = \mathbf{P}_l \mathbf{G} \mathbf{Z}. \quad (9)$$

This equation leads to a generalised eigenvalue problem that obtains the unknown frequencies  $z_1, \dots, z_n$  and characterises the algorithm known as ESPRIT [6].

A related technique exploits the relationship between  $\mathbf{H}_f$  and  $\mathbf{H}_l$ , which are defined in accordance with (7) using the Hankel matrix  $\mathbf{H}$ . Notice that  $\mathbf{Z}_{m,l} = \mathbf{Z}_{m-1}$ . Then, using (5) and (8), we obtain

$$\mathbf{H}_l = \mathbf{Z}_{m-1} \mathbf{D} \mathbf{Z}_{n-1}^T, \quad \mathbf{H}_f = \mathbf{Z}_{m-1} \mathbf{Z} \mathbf{D} \mathbf{Z}_{n-1}^T. \quad (10)$$

Now for any  $z \in \mathbb{C}$ , we have that

$$\mathbf{H}_f - z \mathbf{H}_l = \mathbf{Z}_{m-1} (\mathbf{Z} - z \mathbf{I}) \mathbf{D} \mathbf{Z}_{n-1}^T. \quad (11)$$

Since  $\mathbf{Z}_{m-1}$ ,  $\mathbf{D}$ , and  $\mathbf{Z}_{n-1}$  are all full rank matrices, the unknown diagonal entries of  $\mathbf{Z}$  are the solutions to the generalised eigenvalue problem

$$\mathbf{H}_f \mathbf{v} = z \mathbf{H} \mathbf{v}. \quad (12)$$

Equation (12) is known as the Matrix Pencil Method [7] because the set  $\{\mathbf{H}_f - z\mathbf{H}; z \in \mathbb{C}\}$  constitutes a matrix pencil. Notice that in the noiseless case, when the model order  $n$  is known, (9) and (12) are equivalent, and MPM and ESPRIT obtain the same result.

General matrix pencils  $\{\mathbf{A} - \lambda\mathbf{B}; \lambda \in \mathbb{C}\}$  do not always have eigenstructures  $(\lambda, \mathbf{x})$  that satisfy  $\mathbf{A}\mathbf{x} = \lambda\mathbf{B}\mathbf{x}$ . However, the pencil defined in (9) or (12) will always have a solution as long as  $z_i$  are all distinct and  $c_i \neq 0, i = 1, \dots, n$ . This becomes clear by observing (11). Notice that for all  $z_i, i = 1, \dots, n, (\mathbf{Z} - z_i\mathbf{I})\mathbf{e}_i = 0$ , where  $\mathbf{e}_i \in \mathbb{R}^n$  is the  $i$ th unit vector. Let  $\mathbf{v}_i \in \mathbb{C}^n$  be the solution to an under-determined system of equations  $\mathbf{e}_i = \mathbf{D}\mathbf{Z}_{n-1}^T \mathbf{v}_i$ . Clearly,  $(z_i, \mathbf{v}_i)$  satisfies the matrix pencil in (12). Similarly, we find the set of left eigenvectors by solving  $\mathbf{e}_i^T = \mathbf{w}_i^H \mathbf{Z}_{m-1} \mathbf{D}$ .

## 2.2 Signals with noise

When working with noisy signals, we deal with the following model:

$$y_k = x_k + w_k, \quad k = 0, 1, \dots, \quad (13)$$

where  $x_k$  is given in (1) and  $w_k$  is an additive noise term. We assume that  $w_k$  is a circularly symmetric complex Gaussian process with identically distributed zero-mean uncorrelated real and imaginary parts whose variance is  $\sigma_w^2/2$ . Generally, the order model  $n$  is unknown and it needs to be estimated from the collected data before solving the estimation problem. Even when  $n$  is known, the observed data  $y_k$  needs to be processed to mitigate the effect of  $w_k$  on  $x_k$ . For that, several procedures have been proposed in the past to denoise  $y_k$  before proceeding to the spectral estimation step. When  $n$  is unknown, a first Hankel matrix  $\tilde{\mathbf{H}}$  is built using  $\tilde{n} > n$ .

A first simple approach searches for the dominant singular values of  $\tilde{\mathbf{H}}$  and defines  $n$  as their number [23]. Then, to perform the spectral estimation procedure,  $\tilde{\mathbf{H}}$  is replaced by its truncation to the first  $n$  singular directions. Unfortunately, the resulting approximation does not preserve the Hankel structure. However, Kronecker's theorem [24] states that there is a one-to-one correspondence between a linear combination of  $n$  complex exponentials and a Hankel matrix with rank  $n$ . Then, accuracy is lost when a simple truncation of the Hankel matrix is performed.

This problem was observed in [18, 25] where an iterative procedure that is performed in two steps was proposed. For each iteration, the first step consists of performing the matrix truncation to the first  $n$  singular directions of  $\tilde{\mathbf{H}}$ . In the second step, the truncated matrix is modified by forcing a Hankel structure on it. Both steps are repeated until a stopping criterion is satisfied. An alternative solution performs a search on  $\mathcal{H}$ , the space of Hankel matrices [26]. For a given  $n$ , it is possible to formulate the following optimisation problem:

$$\min_{\substack{\mathbf{H} \in \mathcal{H} \\ \text{rank}(\mathbf{H}) = n}} \|\tilde{\mathbf{H}} - \mathbf{H}\|_F, \quad (14)$$

In [27, 28], this problem was tackled by solving a total least square problem. Recently, a new denoising technique was proposed in [5] by revising Kronecker's theorem, which states that the Hankel matrix generated by a signal  $x_k$  has a rank  $n$ , if and only if  $x_k$  coincides with a function that is a linear combination of  $n$  exponential functions. Let  $\mathbf{H}(\tilde{\mathbf{y}})$  be the Hankel matrix obtained from the sequence  $\tilde{\mathbf{y}} = [\tilde{y}_0, \tilde{y}_1, \dots, \tilde{y}_{m+n-1}]$ . Now, the following optimisation problem returns an approximate signal  $\tilde{y}_k$  that satisfies:

$$\min_{\tilde{\mathbf{y}}} \sum_{k=0}^{m+n-1} |y_k - \tilde{y}_k|^2 \quad (15)$$

s. t.  $\text{rank}[\mathbf{H}(\tilde{\mathbf{y}})] = n$

Due to the constraint on the Hankel matrix, this is a non-convex problem that may be reformulated as

$$\min_{\mathbf{A}, \tilde{\mathbf{y}}} \mathcal{R}_n(\mathbf{A}) + \sum_{k=0}^{m+n-1} |y_k - \tilde{y}_k|^2 \quad (16)$$

s. t.  $\mathbf{A} = \mathbf{H}(\tilde{\mathbf{y}}),$

where  $\mathcal{R}_n(\mathbf{A})$  is an indicator function for matrices that are defined such that  $\mathcal{R}_n(\mathbf{A}) = \infty$  if  $\text{rank}(\mathbf{A}) > n$  and 0 otherwise. In [5], the authors propose to solve the problem using the Alternative Direction Method of Multiplier (ADMM) [29]. This is an efficient technique for solving non-convex optimisation problems. The formulation of ADDM associates an augmented Lagrangian associated with the original problem (16)

$$\mathcal{L}(\mathbf{A}, \tilde{\mathbf{y}}, \mathbf{\Lambda}) = \mathcal{R}_n(\mathbf{A}) + \sum_{k=0}^{m+n-1} |y_k - \tilde{y}_k|^2 + \langle \mathbf{\Lambda}, \mathbf{A} - \mathbf{H}(\tilde{\mathbf{y}}) \rangle + \frac{\rho}{2} \|\mathbf{A} - \mathbf{H}(\tilde{\mathbf{y}})\|_F^2,$$

where  $\mathbf{\Lambda} \in \mathbb{R}^{m \times n}$  is the Lagrange multiplier matrix and  $\rho$  is a constant penalty parameter. The solution is obtained performing the following steps until convergence:

$$\begin{aligned} \mathbf{A}^{i+1} &= \min_{\mathbf{A}} \mathcal{L}(\mathbf{A}, \tilde{\mathbf{y}}^i, \mathbf{\Lambda}^i), \\ \tilde{\mathbf{y}}^{i+1} &= \min_{\tilde{\mathbf{y}}} \mathcal{L}(\mathbf{A}^{i+1}, \tilde{\mathbf{y}}, \mathbf{\Lambda}^i), \\ \mathbf{\Lambda}^{i+1} &= \mathbf{\Lambda}^i + \rho(\mathbf{A}^{i+1} - \mathbf{H}(\tilde{\mathbf{y}}^{i+1})). \end{aligned} \quad (18)$$

The final estimate  $\tilde{\mathbf{y}}$  is obtained by averaging the anti-diagonal terms of the matrix  $\mathbf{A}$ .

We notice that the goal of these procedures is always to obtain an approximated sequence  $\tilde{\mathbf{y}}$  and its associated Hankel matrix  $\mathbf{A} = \mathbf{H}(\tilde{\mathbf{y}})$ . Then, for estimating  $z_i, i = 1, \dots, n$ , a second step is performed by solving (9) or (12).

Even when the first step that mitigates the effect of noise is successful, the computation of the generalised eigenvalues in (9) or (12) may be unstable. This problem is particularly acute when the number of oscillation modes is very large or the complex frequencies are close to each other in the plane.

## 3 Numerical stability

### 3.1 Derivatives of generalised eigenvalues

Suppose that  $\lambda_i$  is the  $i$ th generalised eigenvalue of the pair  $(\mathbf{A}, \mathbf{B}) \in \mathbb{C}^{m \times n}$  with associated right and left eigenvectors  $\mathbf{v}_i \in \mathbb{C}^n$  and  $\boldsymbol{\psi}_i \in \mathbb{C}^m$ , respectively. We assume further that  $\mathbf{A}$  and  $\mathbf{B}$  are full rank and  $\lambda_i$  are all distinct. By definition

$$\mathbf{A}\mathbf{v}_i = \lambda_i \mathbf{B}\mathbf{v}_i, \quad \boldsymbol{\psi}_i^H \mathbf{A} = \lambda_i \boldsymbol{\psi}_i^H \mathbf{B}. \quad (19)$$

Now, consider smooth variations of the matrices  $\mathbf{A}$  and  $\mathbf{B}$  with a real parameter  $\varepsilon$  in a neighbourhood of the origin, i.e.  $\varepsilon \in \mathcal{B}(0)$ . Assume that there exist differentiable functions  $v_i(\varepsilon), \boldsymbol{\psi}_i(\varepsilon), \lambda_i(\varepsilon) \dots \lambda_n(\varepsilon)$ , so that  $\forall \varepsilon \in \mathcal{B}(0)$

$$\begin{aligned} \mathbf{A}(\varepsilon)\mathbf{v}_i(\varepsilon) &= \lambda_i(\varepsilon)\mathbf{B}(\varepsilon)\mathbf{v}_i(\varepsilon), \\ \boldsymbol{\psi}_i^H(\varepsilon)\mathbf{A}(\varepsilon) &= \lambda_i(\varepsilon)\boldsymbol{\psi}_i^H(\varepsilon)\mathbf{B}(\varepsilon), \end{aligned} \quad (20)$$

where  $\lambda_i(0) = \lambda_i, \mathbf{v}_i(0) = \mathbf{v}_i, \boldsymbol{\psi}_i(0) = \boldsymbol{\psi}_i, i = 1, \dots, n$ .

*Theorem 1:* Given  $\mathbf{A}(\varepsilon), \mathbf{B}(\varepsilon) \in \mathbb{C}^{m \times n}$  that satisfy (20) for all  $\varepsilon \in \mathcal{B}(0)$ , we have that

$$\dot{\lambda}_i = \frac{d\lambda_i}{d\varepsilon} = \frac{\boldsymbol{\psi}_i^H [\dot{\mathbf{A}} - \lambda_i \dot{\mathbf{B}}] \mathbf{v}_i}{\boldsymbol{\psi}_i^H \mathbf{B} \mathbf{v}_i}. \quad (21)$$

Here,  $\dot{\mathbf{A}}$  and  $\dot{\mathbf{B}}$  are the first derivatives with respect to  $\varepsilon$  of  $\mathbf{A}(\varepsilon)$  and  $\mathbf{B}(\varepsilon)$ .

*Proof:* See the Appendix.  $\square$

The eigenvalue  $\lambda_i$  will be ill-posed, i.e. small perturbations of  $\mathbf{A}$  and  $\mathbf{B}$  produce large deviations of  $\lambda_i$  from its unperturbed value, if  $\boldsymbol{\psi}_i^H \mathbf{B} \mathbf{v}_i$  is small. This is the case when  $\boldsymbol{\psi}_i$  is close to perpendicular to  $\mathbf{B} \mathbf{v}_i$ . The product  $\boldsymbol{\psi}_i^H \mathbf{B} \mathbf{v}_i$  has been associated with the numerical condition of  $\lambda_i$  [30, ch. 7]

### 3.2 Perturbations on $\mathbf{H}$

When analysing (12), the perturbed matrices  $\mathbf{H}_f(\varepsilon)$  and  $\mathbf{H}_l(\varepsilon)$  are obtained from  $\mathbf{H}(\varepsilon)$ , where  $\varepsilon$  is a scalar parameter. In that case, we can formulate the following corollary:

*Corollary 1:* Let  $\mathbf{H}_f(\varepsilon)$  and  $\mathbf{H}_l(\varepsilon)$  be differentiable functions that represent perturbed versions of the pencil  $(\mathbf{H}_f, \mathbf{H}_l)$ . Assume that for each  $\varepsilon \in \mathcal{B}(0)$ , the eigenvalues  $z_i(\varepsilon)$ , and the left and right eigenvectors  $\mathbf{v}_i(\varepsilon)$  and  $\mathbf{w}_i(\varepsilon)$  exist and they satisfy (20). Then

$$|\dot{z}_i| = \left| \frac{dz_i}{d\varepsilon} \right| \leq E \frac{(1 + |z_i|)}{|c_i|} \frac{\varepsilon_i}{|P_i(z_i)|^2}, \quad i = 1, \dots, n \quad (22)$$

where  $E = \max(\|\dot{\mathbf{H}}_f\|, \|\dot{\mathbf{H}}_l\|)$ ,  $P_i(z) = \prod_{l=1, l \neq i}^n (z - z_l)$  is a polynomial in  $z$  of degree  $n - 1$ , and  $\varepsilon_i = \frac{1}{2\pi} \int_{-\pi}^{\pi} |P_i(e^{j\omega})|^2 d\omega$ .

*Proof:* Applying Theorem 1 to the pencil  $(\mathbf{H}_f, \mathbf{H}_l)$ , we obtain

$$\begin{aligned} |\dot{z}_i| &\leq \frac{\|\dot{\mathbf{H}}_f - z_i \dot{\mathbf{H}}_l\| \|\mathbf{w}_i\| \|\mathbf{v}_i\|}{|\mathbf{w}_i^H \mathbf{H}_l \mathbf{v}_i|} \\ &\leq \max(\|\dot{\mathbf{H}}_f\|, \|\dot{\mathbf{H}}_l\|) \frac{(1 + |z_i|) \|\mathbf{w}_i\| \|\mathbf{v}_i\|}{|\mathbf{w}_i^H \mathbf{H}_l \mathbf{v}_i|}, \end{aligned} \quad (23)$$

where  $\mathbf{w}_i$  and  $\mathbf{v}_i$  are left and right eigenvectors of the pencil  $(\mathbf{H}_f, \mathbf{H}_l)$ .

Recalling the decomposition in (10), we have that  $\mathbf{v}_i \in \mathbb{C}^n$  and  $\mathbf{w}_i \in \mathbb{C}^m$  satisfy

$$\mathbf{Z}_{m-1}^T \mathbf{v}_i = \mathbf{1}_i, \quad \mathbf{w}_i^H \mathbf{Z}_{m-1} = \mathbf{1}_i^T, \quad (24)$$

where  $\mathbf{1}_i$  is the  $i$ th unitary vector in  $\mathbb{R}^n$ . Moreover

$$\mathbf{w}_i^H \mathbf{H}_l \mathbf{v}_i = \mathbf{w}_i^H \mathbf{Z}_{m-1} \mathbf{D} \mathbf{Z}_{n-1}^T \mathbf{v}_i = c_i. \quad (25)$$

Since  $\mathbf{Z}_{m-1}^T = [\mathbf{Z}_{n-1}^T \quad \mathbf{M}^T]$ , with  $\mathbf{M} \in \mathbb{C}^{(m-n) \times n}$ , one may choose  $\mathbf{w}_i^T = [\mathbf{v}_i^T \quad \mathbf{0}_{1 \times (m-n)}]$  to satisfy (24). On the other hand, and since  $\mathbf{Z}_{n-1}$  is an invertible matrix, (24) implies that  $\mathbf{v}_i$  is the  $i$ th column of  $\mathbf{Z}_{n-1}^T$ . Then, using the result given in [31], we have that for  $k = 1, \dots, n$

$$\begin{aligned} (\mathbf{v}_i)_k &= (-1)^{k-1} \frac{(\alpha_i)_k}{\prod_{l=1, l \neq i}^n (z_l - z_i)}, \\ (\alpha_i)_k &= \sum_{l=1}^{n-1} z_{s_1} \cdots z_{s_{n-k}}, \end{aligned} \quad (26)$$

where  $s_1 \cdots s_{n-k}$  is a combination of  $n - k$  elements taken from  $\{1, \dots, i-1, i+1, \dots, n\}$ . Given a choice for the left eigenvector,  $\|\mathbf{w}_i\| = \|\mathbf{v}_i\|$ . Then

$$\|\mathbf{w}_i\| \|\mathbf{v}_i\| = \frac{\sum_{k=1}^n |(\alpha_i)_k|^2}{\prod_{l=1, l \neq i}^n |z_l - z_i|^2}. \quad (27)$$

Notice that  $(\alpha_i)_k$  is the  $k$ th coefficient of the polynomial  $P_i(z) = \prod_{l=1, l \neq i}^n (z - z_l)$ . Re-arranging terms, we have that

$$z^{-(n-1)} P_i(z) = \sum_{k=0}^{n-1} (\alpha_i)_{n-1-k} z^{-k}.$$

Then, the Parseval's theorem between the sequence  $\alpha_{n-1-k}$  and its discrete-time Fourier transform states that

$$\sum_{k=0}^{n-1} |(\alpha_i)_k|^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} |P_i(e^{j\omega})|^2 d\omega = \varepsilon_i, \quad (28)$$

where to compute the integral, we have evaluated  $P_i(z)$  in the unit circle,  $z = e^{j\omega}$ . Replacing this equation in (27), and its result together with (25) into (23), we obtain

$$|\dot{z}_i| \leq E \frac{(1 + |z_i|)}{|c_i|} \frac{\varepsilon_i}{\prod_{l=1, l \neq i}^n |z_l - z_i|^2}. \quad \square \quad (29)$$

Corollary 1 shows that when performing spectral estimation, the estimation of  $z_i$  is sensitive to  $|c_i|$  but also to  $\min_l |z_l - z_i|$ . The constant  $c_i$  represents the residue associated with  $z_i$  and it is related to the observed strength of the complex frequency. As one may suspect, small  $|c_i|$  lead to an associated eigenvalue that can be easily perturbed. On the other hand, when an eigenvalue is located inside a cluster of eigenvalues,  $z_i$  may take a large value. Then, the sensitivity to numerical errors of the estimated frequency  $z_i$  is affected not only by its strength in the mixture, but also by its proximity to the other complex frequencies. In the next section, we present a technique to cope with this problem.

## 4 Downsampling the base-band equivalent

Consider that  $y_k$  in (13) is obtained after uniformly sampling a continuous-time version

$$y(t) = \sum_{i=1}^n c_i e^{\xi_i t} + w(t),$$

where  $\xi_i = (\gamma_i + j2\pi\nu_i)$ . In this case,  $y_k = y(kT_s)$ , where  $T_s$  is the sampling period, and  $z_i = e^{\xi_i T_s}$ . Clearly, the location of  $z_i$  in the complex plane may be controlled by judiciously adjusting the sampling time. However, the selection of  $T_s$  responds to several constraints when designing the signal acquisition system, and adding additional requirements may not be feasible. Nevertheless, the location of  $z_i$  may also be changed by the decimation of the observed signal [32].

Suppose that the observed signal has its spectral content concentrated on  $\Upsilon = [\nu_{\min}, \nu_{\max}]$ , i.e. for  $i = 1, \dots, n$

$$z_i = e^{j\nu_i T_s} [\cos(2\pi\nu_i T_s) + j \sin(2\pi\nu_i T_s)], \quad \nu_i \in \Upsilon. \quad (30)$$

According to Section 3, one would like to 'zoomed in'  $\Upsilon$  before performing the estimation of  $z_i$  to improve the performance of the estimation algorithm. For that, we proceed as follows. Let  $\nu_c = ((\nu_{\max} - \nu_{\min})/2)$  and define

$$y_k^{\text{bb}} = y_k e^{-j\nu_c k T_s}.$$

The signal  $y_k^{\text{bb}}$  is known as the baseband equivalent of  $y_k$ . Consider a scalar  $Q$  such that  $\Upsilon Q \leq (2\pi/T_s)$ . Then, using Fig. 1, we construct  $y_k^{\text{bb}, D}$  after filtering and decimating  $y_k^{\text{bb}}$ . To avoid aliasing,  $y_k^{\text{bb}}$  is filtered before decimation with a linear phase finite impulse

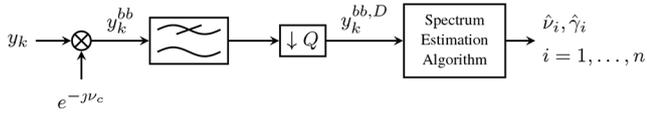


Fig. 1 Shift-and-zoom scheme for estimating complex modes

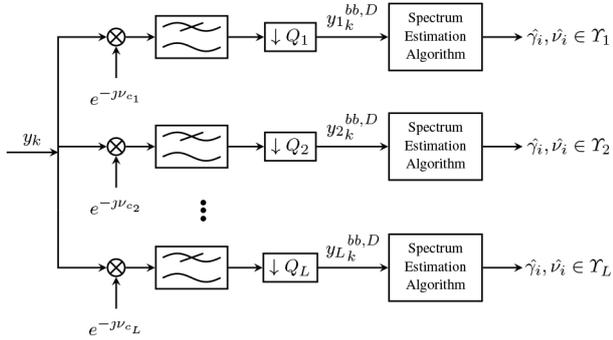


Fig. 2 Block diagram for processing the signal compound of  $L$  clusters

Table 1 Parameters for the example taken from [5]

$i$	1	2	3	4
$\nu_i$	-7.68	39.68	40.96	99.84
$\gamma_i$	-0.274	-0.150	0.133	-0.221
$ c_i $	0.4	1.2	1.0	0.9
$\angle c_i$	-0.93	-1.55	-0.83	0.07

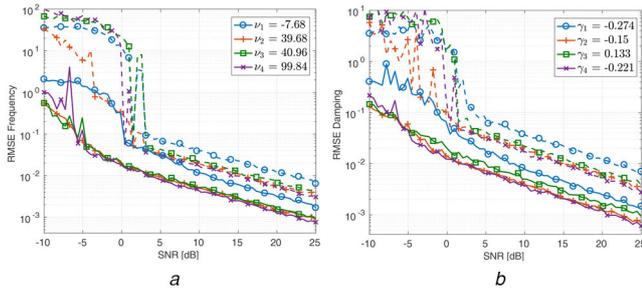


Fig. 3 RMSE as a function of SNR: dashed lines correspond to [5]; solid lines correspond to shift-and-zoom  
(a)  $\hat{\sigma}_\nu(\text{SNR})$ , (b)  $\hat{\sigma}_\gamma(\text{SNR})$

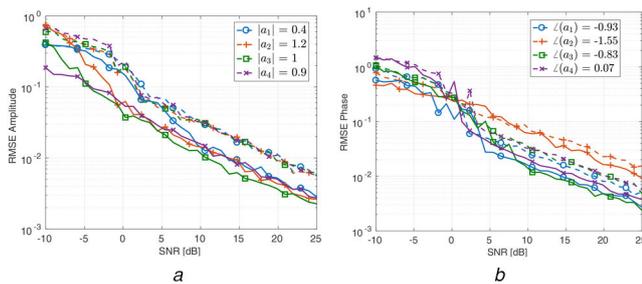


Fig. 4 RMSE as a function of SNR: dashed lines correspond to [5]; solid lines correspond to shift-and-zoom  
(a) Amplitude RMSE, (b) Phase RMSE

response filter [33]. Now, we apply the frequency estimation procedure to  $y_k^{\text{bb},D}$  and shift back the estimated modes. The final structure is shown in Fig. 1. We refer to this estimation strategy as a *shift-and-zoom* technique.

Using the decimated baseband signal  $y_k^{\text{bb},D}$ , we construct a Hankel matrix as in (2) and estimate the complex frequencies as in Section 2. Let  $z_i^{\text{bb},D}$  be the frequency estimates obtained from  $y_k^{\text{bb},D}$ . Clearly

$$z_i = (z_i^{\text{bb},D})^{1/Q} e^{j\nu_i T_s}.$$

When decimating or downsampling  $y_k^{\text{bb}}$  by a factor  $Q$ , the complex frequencies move in the complex plane. For  $Q > 1$ , the frequencies that were clustered in  $\Upsilon$  may spread apart, diminishing the value of  $z_i$ , and making the spectral estimation more accurate. It is important to remark that the low-pass filter included in Fig. 1 modifies the amplitudes related to the complex modes  $z_i$ , affecting the value of the derivative of  $z_i$  as seen in (22).

Now, suppose that  $\nu_i \in \cup_{l=1}^L \Upsilon_l$ , for all  $i = 1, \dots, n$ , where  $\Upsilon_l = [\nu_{l,\min}, \nu_{l,\max}]$  are disjoint intervals. In this case,  $z_i$  are separated in  $L$  different clusters in the complex plane. To improve performance, the procedure outlined above could be repeated for each cluster. In particular, since  $\Upsilon_l$  are intervals of unequal lengths, the decimation factor could be different for each  $l$ . The complete procedure is summarised in Fig. 2.

## 5 Numerical experiments

In this section, we evaluate the performance of the shift-and-zoom procedure outlined in the previous section. For that, we analyse two examples where sums of complex exponentials are simulated for different signal-to-noise ratios (SNRs). In both cases, we estimated the complex modes  $\xi_i = \gamma_i + j\nu_i$  and compared the results using the procedure presented in [5]. To appraise the performance of each algorithm, we repeat the experiments  $K$  times and we compute the root mean square errors (RMSEs) for the estimated frequencies and damping factors

$$\hat{\sigma}_\nu = \sqrt{\frac{1}{K} \sum_{k=1}^K \sum_{i=1}^n (\nu_i - \hat{\nu}_{ik})^2},$$

$$\hat{\sigma}_\gamma = \sqrt{\frac{1}{K} \sum_{k=1}^K \sum_{i=1}^n (\gamma_i - \hat{\gamma}_{ik})^2},$$
(31)

where  $\hat{\nu}_{ik}$  and  $\hat{\gamma}_{ik}$  are the estimations obtained after the  $k$ th Monte Carlo run.

### 5.1 Weak modes

As a first example, we consider the sum of four modes described in Table 1. This example is taken from [5]. We simulate  $y_k$  using a sampling interval  $T_s = 0.0039$  s.

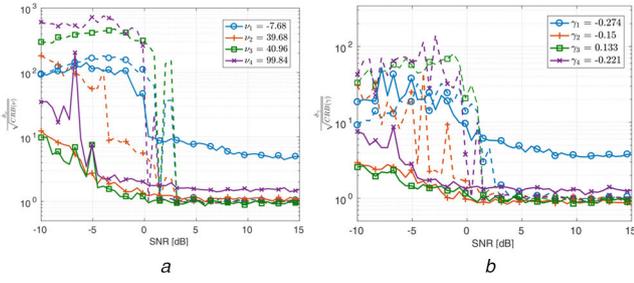
When applying the shift-and-zoom technique, we consider three disjoint frequency intervals,  $\Upsilon_1 = [0 \ 15.39]$  Hz,  $\Upsilon_2 = [32.62 \ 48.01]$  Hz, and  $\Upsilon_3 = [92.14 \ 107.54]$  Hz. Clearly,  $\nu_1 \in \Upsilon_1$ ,  $\nu_2, \nu_3 \in \Upsilon_2$ , and  $\nu_4 \in \Upsilon_3$ .

Here, we will assume that the cluster locations are known a priori. When this information is not available, a preliminary run of the estimation algorithm can be performed in order to obtain rough estimates of the frequency bands.

Notice that all three segments have the same bandwidth. Although this is not a necessary condition for the algorithm, it simplifies its implementation. Using this fact, we use the same low-pass filter and downsampling coefficient for all three frequency segments. In particular, we design an FIR linear phase filter using a rectangular window of 16 coefficients, with bandwidth 15.39 Hz. The downsampling coefficient is  $Q = 4$ . For  $K = 200$  we compute  $\hat{\sigma}_\nu$  and  $\hat{\sigma}_\gamma$  for each SNR. The results are shown in Fig. 3 as functions of the SNR. In Fig. 4, we also show the RMSEs of the amplitude and phase.

For high SNR, shift-and-zoom has a similar performance to [5]. However, for small SNR, the better conditioning of  $z_i$  becomes critical, resulting on remarkably better performance of shift-and-zoom when compared to a more traditional approach. However, it is important to notice that when we downsample the signal, there is a tradeoff between the number of samples remaining after decimation and how far apart the complex modes result.

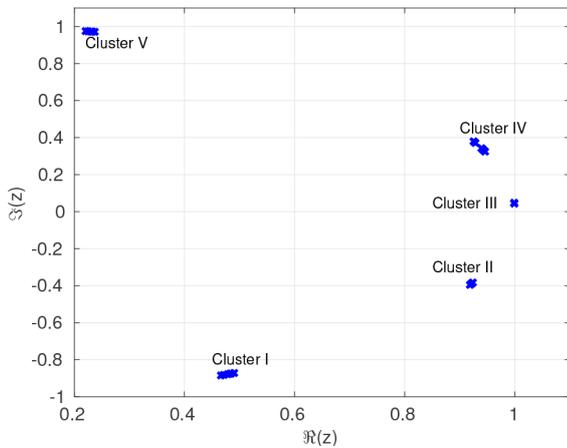
To evaluate our procedure, we compute the Cramér Rao bound (CRB) using the expressions in the Appendix [34]. The CRB is a lower bound on the variance of any unbiased estimator. Although we have not proved that the proposed estimators are unbiased, we



**Fig. 5** Ratio between RMSE and the CRB. Solid lines correspond to shift-and-zoom and dashed lines to [5] (a) Ratio between RMSE and the CRB for  $\nu$ , (b) Ratio between RMSE and the CRB for  $\gamma$

**Table 2** Residual RMSE as a function of the oversampling factor

Q	1	2	4	7	9	14	15
RMSE	15.27	15.26	7.82	11.25	12.09	14.54	15.82



**Fig. 6** Location of the modes described in Table 4 when  $T_s = 0.4883 \times 10^{-3}$

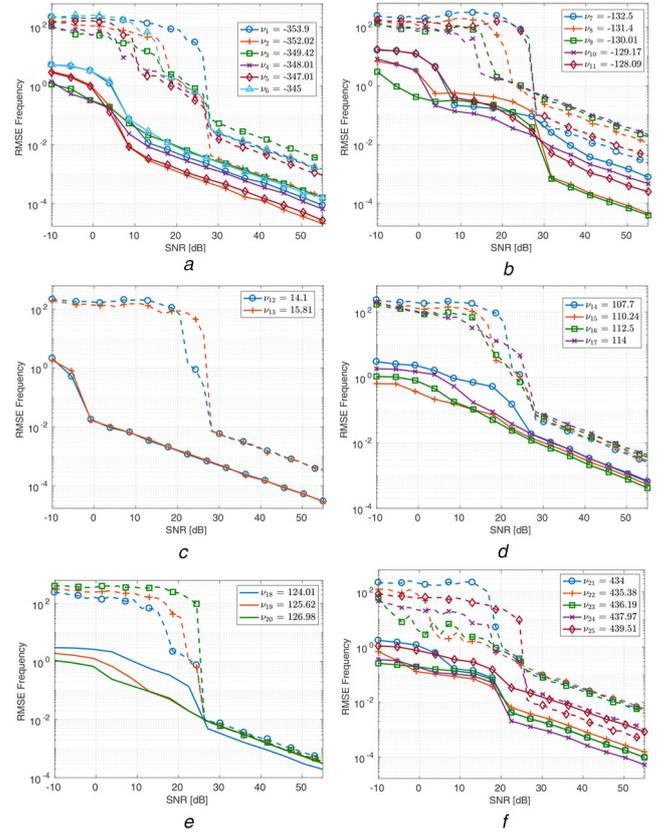
**Table 3** Frequency segments and decimation factors

	$\nu_{c_i}$ , Hz	$\nu_{i_{min}}$ , Hz	$\nu_{i_{max}}$ , Hz	$Q_i$
$Y_1$	-349.45	-390.41	-308.49	4
$Y_2$	-130.29	-191.73	-68.85	3
$Y_3$	14.95	4.71	25.19	4
$Y_4$	140.95	100.61	166.45	5
$Y_5$	436.75	416.27	457.23	6

will use this bound as an indicator of the margin for improvement in the estimation procedure. Notice that the CRB depends on  $z_i$  and not on  $\xi_i$ . Therefore, when downsampling  $y_k^{bb}$ , the CRB changes. In Fig. 5, we compared the ratios  $\hat{\sigma}_\nu / \sqrt{\text{CRB}(\nu)}$  and  $\hat{\sigma}_\gamma / \sqrt{\text{CRB}(\gamma)}$  for the estimations obtained through shift-and-zoom and through the procedure in [5].

Notice that the downsampling factor has an upper bound determined by the number of samples in the original signal. As was pointed out in [34], the CRB depends not only on the distance among frequencies but also on the number of samples. In fact, it is pointed out that the CRB depends on the product of the number of samples and the separation among frequencies.

The length of the observed signal is also significant when estimating the damping factors. This issue becomes critical when the estimated model order is larger than the actual one. In this case, erroneous phantom damping factors could affect the estimation of the actual parameters. A solution in [35] tackles this problem by using  $\gamma'_i = \gamma_i / (m + n - 1)$ .



**Fig. 7** RMSE for  $\hat{\nu}_p$  as a function of SNR for all clusters. Dashed line corresponds to [5]; solid lines correspond to shift-and-zoom (a) Cluster I, (b) Cluster II, (c) Cluster III, (d) Cluster IV (a), (e) Cluster IV (b), (f) Cluster VI

To analyse the effect of the downsampling factor  $Q$ , we have computed the residual RMSE between the estimated signal and the true signal for different values of  $Q$ . The results are shown in Table 2. The minimum RMSE is obtained for  $Q = 4$ . As we increase  $Q$ , the RMSE becomes higher because fewer samples are used. On the other hand, as  $Q$  decreases, even though a longer signal record is analysed, we lose the effect of frequency separation.

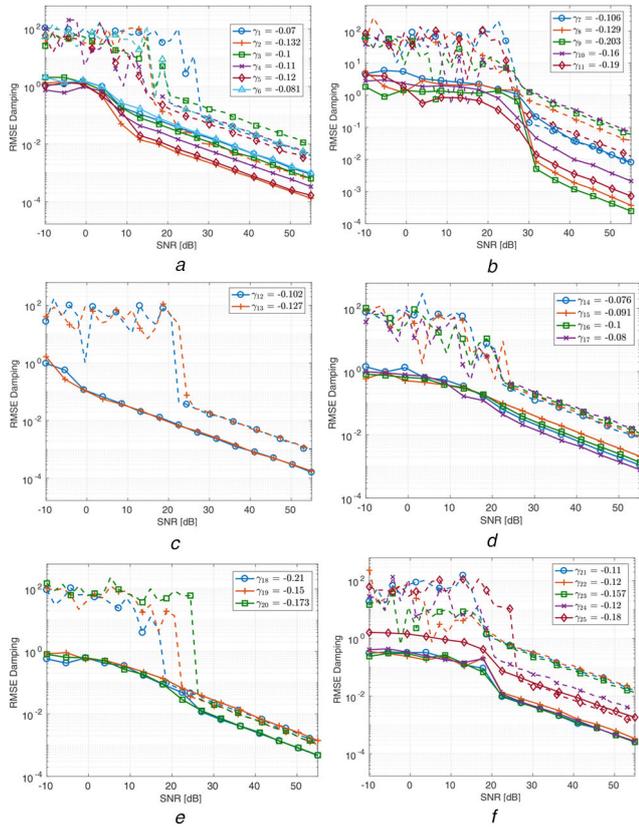
### 5.2 Clustered modes

In the second example, we consider the sum of 25 damped oscillations with different complex amplitudes each. This example that corresponds to data obtained from magnetic resonance spectroscopy (MRS) was presented in [17]. The sampling interval is  $T_s = 0.4883 \times 10^{-3}$ . The parameters are described in Table 4 where frequencies, damping factors, phases and amplitudes are provided for each complex mode. The authors grouped the 25 modes in 6 disjoint clusters. In Fig. 6, we have plotted the locations of  $z_i = e^{(\gamma_i + j2\pi\nu_i)T_s}$ ,  $i = 1, \dots, 25$ .

For the purpose of using the shift-and-zoom procedure, we consider  $L = 5$  different frequency intervals. Interval limits and decimation factors are found in Table 3.

We processed  $y_k$  with the shift-and-zoom algorithm and compared the estimations with those obtained with no preprocessing (shifting and decimation). Figs. 7 and 8 show the RMSE of the estimated frequencies and damping factors for the different clusters using both techniques. For a better presentation of the results, the Cluster IV is separated into two. Cluster IV (a) with four different frequencies and Cluster IV (b) with three frequencies.

At low SNR, shift-and-zoom performs better than a more traditional approach. Since shift-and-zoom increases the frequency separation, improving  $z_i$  makes the algorithm more resilient to noise disturbances. This observation is even more relevant when



**Fig. 8** RMSE for  $\hat{\gamma}_p$  as a function of SNR for all clusters. Dashed line corresponds to [5]; solid lines correspond to shift-and-zoom (a) Cluster I, (b) Cluster II, (c) Cluster III, (d) Cluster IV (a), (e) Cluster IV (b), (f) Cluster VI

the amplitude  $|c_i|$  is small. This is the case of modes  $\xi_{14}$  to  $\xi_{17}$  in Cluster IV.

In this example, it was critical to decimate  $y_k^{bb}$  instead of  $y_k$  directly. Should we have done the latter, the maximum decimation factor would have been  $Q = 2$  in order to avoid aliasing. Such value for  $Q$  results in a poor estimation of some modes, in particular, those in Cluster IV (a). By implementing parallel frequency estimation on each cluster, we could separate further the complex frequencies, improving the final performance. Using this approach, we transform a poorly conditioned problem into five different problems that are better conditioned.

## 6 Conclusion

We have analysed the performance of spectrum estimation techniques when the complex frequencies are close to each other. We have shown that the estimations become vulnerable to small perturbations in the observed data when the modes have small energy and/or the distance among them is small. Shift-and-zoom is a technique to cope with this problem when some a priori information is available. In those cases, we have shown that it outperforms Hankel-based direct approaches that do not benefit from the enhancement spacing among complex modes introduced in shift-and-zoom. In particular, we have shown that the proposed scheme is more efficient than traditional schemes when working under a low-SNR regime. When the SNR is high, the shift-and-zoom is as efficient as a more traditional method. In this case, the decrease in the number of samples due to the decimation step becomes relevant, and shift-and-zoom requires a longer data stream to converge to the CRB. Finally, with the shift-and-zoom technique, we overcome the problem of introducing aliasing when we make the downsampling. However, as we mentioned earlier there exists a trade-off between the number of samples and the decimation factor.

## 7 References

- [1] Bai, Z., Demmel, J., Dongarra, J., et al.: 'Templates for the solution of algebraic eigenvalue problems: a practical guide', *Software, Environments and Tools* (Society for Industrial and Applied Mathematics, USA, 2000)
- [2] Stoica, P., Moses, R.L.: 'Spectral analysis of signals' (Pearson Prentice-Hall, USA, 2005)
- [3] Stoica, P., Selen, Y.: 'Model-order selection: a review of information criterion rules', *IEEE Signal Process. Mag.*, 2004, **21**, (4), pp. 36–47
- [4] Gavish, M., Donoho, D.L.: 'The optimal hard threshold for singular values is  $4/\sqrt{3}$ ', *IEEE Trans. Inf. Theory*, 2014, **60**, (8), pp. 5040–5053
- [5] Andersson, F., Carlsson, M., Tourneret, J., et al.: 'A new frequency estimation method for equally and unequally spaced data', *IEEE Trans. Signal Process.*, 2014, **62**, (21), pp. 5761–5774
- [6] Roy, R., Kailath, T.: 'Esprit-estimation of signal parameters via rotational invariance techniques', *IEEE Trans. Acoust. Speech Signal Process.*, 1989, **37**, (7), pp. 984–995
- [7] Hua, Y., Sarkar, T.K.: 'Matrix pencil method for estimating parameter of exponentially damped/undamped sinusoids in noise', *IEEE Trans. Acoust. Speech Signal Process.*, 1990, **38**, (5), pp. 814–824
- [8] Ying, J., Cai, J., Guo, D., et al.: 'Vandermonde factorization of Hankel matrix for complex exponential signal recovery – application in fast NMR spectroscopy', *IEEE Trans. Signal Process.*, 2018, **66**, (21), pp. 5520–5533
- [9] Yang, Z., Li, J., Stoica, P., et al.: 'Chapter 11 – sparse methods for direction-of-arrival estimation', in Chellappa, R., Theodoridis, S. (Eds.): 'Academic press library in signal processing', vol. 7 (Academic Press, UK, 2018), pp. 509–581. Available at <http://www.sciencedirect.com/science/article/pii/B9780128118870000110>
- [10] Yang, Z., Xie, L.: 'On gridless sparse methods for line spectral estimation from complete and incomplete data', *IEEE Trans. Signal Process.*, 2015, **63**, (12), pp. 3139–3153
- [11] Yang, Z., Xie, L.: 'Exact joint sparse frequency recovery via optimization methods', *IEEE Trans. Signal Process.*, 2016, **64**, (19), pp. 5145–5157
- [12] Candès, E.J., Fernandez-Granda, C.: 'Towards a mathematical theory of superresolution', *Commun. Pure Appl. Math.*, 2014, **67**, (6), pp. 906–956. Available at <https://onlinelibrary.wiley.com/doi/abs/10.1002/cpa.21455>
- [13] Batenkov, D.: 'Stability and super-resolution of generalized spike recovery', *Appl. Comput. Harmon. Anal.*, 2018, **45**, (2), pp. 299–323. Available at <http://www.sciencedirect.com/science/article/pii/S1063520316300641>
- [14] Halder, B., Kailath, T.: 'Efficient estimation of closely spaced sinusoidal frequencies using subspace-based methods', *IEEE Signal Process. Lett.*, 1997, **4**, (2), pp. 49–51
- [15] Morren, G., Lemmerling, P., Huffel, S.V.: 'Decimative subspace-based parameter estimation techniques', *Signal Process.*, 2003, **83**, (5), pp. 1025–1033. Available at <http://www.sciencedirect.com/science/article/pii/S0165168402005042>
- [16] Bolt, B.A., Brillinger, D.R.: 'Estimation of uncertainties in eigenspectral estimates from decaying geophysical time series', *Geophys. J. Int.*, 1979, **59**, (3), pp. 593–603. Available at <https://doi.org/10.1111/j.1365-246X.1979.tb02574.x>
- [17] Cuyt, A., Tsai, M., Verhoye, M., et al.: 'Faint and clustered components in exponential analysis', *Appl. Math. Comput.*, 2018, **327**, pp. 93–103. Available at <http://www.sciencedirect.com/science/article/pii/S0096300317307804>
- [18] Li, Y., Liu, K.J.R., Razavilar, J.: 'A parameter estimation scheme for damped sinusoidal signals based on low-rank Hankel approximation', *IEEE Trans. Signal Process.*, 1997, **45**, (2), pp. 481–486
- [19] Golub, G., Milanfar, P., Varah, J.: 'A stable numerical method for inverting shape from moments', *SIAM J. Sci. Comput.*, 1999, **21**, (4), pp. 1222–1243. Available at <https://doi.org/10.1137/S1064827597328315>
- [20] Beckermann, B., Golub, G.H., Labahn, G.: 'On the numerical condition of a generalized Hankel eigenvalue problem', *Numer. Math.*, 2007, **106**, (1), pp. 41–68. Available at <https://doi.org/10.1007/s00211-006-0054-x>
- [21] Kravanja, P., Sakurai, T., Sugiura, H., et al.: 'A perturbation result for generalized eigenvalue problems and its application to error estimation in a quadrature method for computing zeros of analytic functions', *J. Comput. Appl. Math.*, 2003, **161**, (2), pp. 339–347. Available at <http://www.sciencedirect.com/science/article/pii/S0377042703007040>
- [22] Hildebrand, F.B.: 'Introduction to numerical analysis: second edition', *International Series in Pure and Applied Mathematics* (McGraw-Hill, USA, 1974)
- [23] Hua, Y., Sarkar, T.K.: 'On SVD for estimating generalized eigenvalue of singular matrix pencil in noise', *IEEE Trans. Signal Process.*, 1991, **39**, (4), pp. 892–900
- [24] Gantmacher, F.R.: 'The theory of matrices' (Chelsea Pub. Co., USA, 1960, 2nd edn.)
- [25] Lu, B., Wei, D., Evans, B.L., et al.: 'Improved matrix pencil methods'. Conf. record of Thirty-Second Asilomar Conf. on Signals, Systems and Computers (Cat. No. 98CH36284), Pacific Grove, CA, USA, 1998, vol. 2, pp. 1433–1437
- [26] Cadzow, J.A.: 'Signal enhancement a composite property mapping algorithm', *IEEE Trans. Acoust. Speech Signal Process.*, 1988, **36**, (1), pp. 49–62
- [27] Park, H., Zhang, L., Rosen, J.B.: 'Low rank approximation of a Hankel matrix by structured total least norm', *BIT Numer. Math.*, 1999, **39**, (4), pp. 757–779. Available at <https://doi.org/10.1023/A:1022347425533>
- [28] Dicle, C., Camps, O.I., Szaier, M.: 'The way they move: tracking multiple targets with similar appearance'. 2013 IEEE Int. Conf. on Computer Vision, Sydney, NSW, Australia, 2013, pp. 2301–2311
- [29] Boyd, S., Parikh, N., Chu, E., et al.: 'Distributed optimization and statistical learning via the alternating direction method of multipliers' (Now, USA, 2011). Available at <https://ieeexplore.ieee.org/document/8186925>

- [30] Golub, G.H., Van Loan, C.F.: ‘*Matrix computation*’ (The Johns Hopkins University Press, USA, 1996, 3rd edn.)
- [31] Rawashdeh, E.A.: ‘*A simple method for finding the inverse matrix of vandermonde matrix*’ (MATEMATICKI VESNIK, Serbia, 2018)
- [32] Vaidyanathan, P.P.: ‘*Multirate systems and filter banks*’ (Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1993)
- [33] Oppenheim, A.V., Schaffer, R.W.: ‘*Discrete-time signal processing*’ (Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1989)
- [34] Yao, Y.-X., Pandit, S.M.: ‘Cramer-Rao lower bounds for a damped sinusoidal process’, *IEEE Trans. Signal Process.*, 1995, **43**, (4), pp. 878–885
- [35] Hasan, T.: ‘Nonlinear time series regression for a class of amplitude modulated consinusoids’, *J. Time Ser. Anal.*, 1982, **3**, (2), pp. 109–122. Available at <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9892.1982.tb00333.x>

## 8 Appendix

### 8.1 Computation of $\hat{\lambda}_i$

Let  $\mathbf{V} = [\mathbf{v}_1, \dots, \mathbf{v}_n] \in \mathbb{C}^{n \times n}$  and  $\mathbf{W}^H = [\mathbf{w}_1^H, \dots, \mathbf{w}_n^H]^T \in \mathbb{C}^{n \times m}$  be two matrices containing the right and left eigenvectors, and  $\Lambda = \text{diag}(\lambda_i)$  the diagonal matrix of eigenvalues. Then, (19) can be rewritten as

$$\mathbf{A}\mathbf{V} = \mathbf{B}\mathbf{V}\Lambda \quad (32)$$

$$\mathbf{W}^H\mathbf{A} = \Lambda\mathbf{W}^H\mathbf{B} \quad (33)$$

Multiplying both sides of (32) by  $\mathbf{W}^H$ , we obtain  $\mathbf{W}^H\mathbf{A}\mathbf{V} = \mathbf{W}^H\mathbf{B}\mathbf{V}\Lambda = \Lambda\mathbf{W}^H\mathbf{B}\mathbf{V}$ , where the last equality comes from (33). Since the diagonal matrix  $\Lambda$  commutes with  $\mathbf{W}^H\mathbf{B}\mathbf{V}$ , this last matrix is diagonal and so is  $\mathbf{W}^H\mathbf{A}\mathbf{V}$ .

Now, considering the small perturbations defined in (20) and differentiating with respect to  $\varepsilon$ , we obtain

$$\dot{\mathbf{A}}\mathbf{V} + \mathbf{A}\dot{\mathbf{V}} = \dot{\mathbf{B}}\mathbf{V}\Lambda + \mathbf{B}\dot{\mathbf{V}}\Lambda + \mathbf{B}\mathbf{V}\dot{\Lambda} \quad (34)$$

Since  $\mathbf{V}$  spans  $\mathbb{C}^n$ , any vector in that space can be formulated as a linear combination of  $\mathbf{v}_i$ . Then,  $\dot{\mathbf{v}}_j = \sum_i r_{ij}\mathbf{v}_i$ . Let  $\mathbf{R} = [r_{ij}] \in \mathbb{C}^{n \times n}$  be such that  $\dot{\mathbf{V}} = \mathbf{V}\mathbf{R}$ . Then, replacing in (34) and left multiplying both sides of the equation by  $\mathbf{W}^H$ , we obtain

$$\mathbf{W}^H\dot{\mathbf{A}}\mathbf{V} + \mathbf{W}^H\mathbf{A}\mathbf{V}\mathbf{R} = \mathbf{W}^H\dot{\mathbf{B}}\mathbf{V}\Lambda + \mathbf{W}^H\mathbf{B}\mathbf{V}\mathbf{R}\Lambda + \mathbf{W}^H\mathbf{B}\mathbf{V}\dot{\Lambda} \quad (35)$$

Notice that  $\mathbf{W}^H\mathbf{A}\mathbf{V}\mathbf{R} - \mathbf{W}^H\mathbf{B}\mathbf{V}\mathbf{R}\Lambda = \mathbf{W}^H\mathbf{B}\mathbf{V}[\Lambda\mathbf{R} - \mathbf{R}\Lambda]$ . The diagonal elements of  $[\Lambda\mathbf{R} - \mathbf{R}\Lambda]$  are zeros, because  $\Lambda$  is a diagonal matrix. Moreover, since  $\mathbf{W}^H\mathbf{B}\mathbf{V}$  is also a diagonal matrix,  $\mathbf{W}^H\mathbf{B}\mathbf{V}[\Lambda\mathbf{R} - \mathbf{R}\Lambda]$  has zeroes in its diagonal too.

Equality (35) defines  $m \times n$  scalar equations. Evaluating those corresponding to the diagonal elements, we obtain

$$\mathbf{w}_k^H\mathbf{B}\mathbf{v}_k\dot{\lambda}_k = \mathbf{w}_k^H[\dot{\mathbf{A}} - \lambda_k\dot{\mathbf{B}}]\mathbf{v}_k$$

as in (21).

### 8.2 Parameters for numerical example from [17]

See Table 4.

### 8.3 Expression of the CRB

Following [34] the CRB for the frequencies and damping factors is

**Table 4** Signal contains 25 complex modes clustered in 5 regions

Cluster	$i$	$\gamma_i$	$\nu_i$	$ c_i $	$\angle c_i$	
I	1	-0.07	-353.90	0.77	0.15	
	2	-0.132	-352.02	6.2	0.0	
	3	-0.1	-349.42	0.98	0.3	
	4	-0.11	-348.01	5.4	0.9	
	5	-0.12	-347.01	6.1	0.7	
	6	-0.081	-345.00	0.95	0.2	
II	7	-0.106	-132.50	4.71	0.12	
	8	-0.129	-131.40	3.9	0.1	
	9	-0.203	-130.01	7.0	-0.234	
	10	-0.16	-129.17	5.43	0.2	
	11	-0.19	-128.09	4.4	-0.52	
III	12	-0.102	14.10	3	0.21	
	13	-0.127	15.81	3	-0.8	
IV	14	-0.076	107.70	0.39	-0.3	
	15	-0.091	110.24	0.37	-0.8	
	16	-0.1	112.50	0.36	0.1	
	17	-0.08	114.00	0.3	0.9	
	18	-0.21	124.01	3.2	-0.106	
	19	-0.15	125.62	5.53	0.2	
	20	-0.173	126.98	4.7	-0.3	
	V	21	-0.11	434.00	1	-0.15
		22	-0.12	435.38	5	0.26
		23	-0.157	436.19	6.1	-0.2
24		-0.12	437.97	5.1	0.0	
25		-0.18	439.51	6	-0.1	

$$\text{CRB}(\nu_i) = \frac{[\tilde{\mathbf{Q}}]_{ii}}{\text{SNR}_i}, \quad i = 1, \dots, M \quad (36)$$

$$\text{CRB}(\gamma_i) = \frac{[\tilde{\mathbf{Q}}]_{(i+M)(i+M)}}{\text{SNR}_i}, \quad i = 1, \dots, M, \quad (37)$$

where  $\text{SNR}_i = |c_i|^2 / \sigma_w^2$ , with  $\sigma_w^2$  the variance of the noise signal, and  $\tilde{\mathbf{Q}} = [2\text{Re}(\mathbf{Z}\mathbf{Z}^H)]^{-1}$ . Here

$$\mathbf{Z} = \begin{bmatrix} j\Theta\mathbf{T}\mathbf{Z}_{N-1} \\ -\Theta\mathbf{T}\mathbf{Z}_{N-1} \\ \Theta\mathbf{Z}_{N-1} \\ j\Theta\mathbf{Z}_{N-1} \end{bmatrix}, \quad \Theta = \text{diag}[e^{j\angle c_1}, \dots, e^{j\angle c_M}],$$

$$\mathbf{T} = \text{diag}[0 \ 1 \dots N-1],$$

$\mathbf{Z}_{N-1}$  is as defined in (4), and  $N$  is the length of the signal in (13). When the complex frequencies come close to each other,  $\mathbf{Z}_{N-1}\mathbf{Z}_{N-1}^H$  are near singular. Therefore, since the CRB is inversely proportional to the determinant of  $\mathbf{Z}_{N-1}\mathbf{Z}_{N-1}^H$ , the value of the CRB increases.