HAL open science

A New Frequency Estimation Method for Equally and Unequally Spaced Data

Fredrik Andersson, Marcus Carlsson, Jean-Yves Tourneret, Herwig Wendt

To cite this version:

Fredrik Andersson, Marcus Carlsson, Jean-Yves Tourneret, Herwig Wendt. A New Frequency Estimation Method for Equally and Unequally Spaced Data. IEEE Transactions on Signal Processing, 2014, 62 (21), pp.5761-5774. 10.1109/TSP.2014.2358961 . hal-03266929

HAL Id: hal-03266929

https://hal.science/hal-03266929v1

Submitted on 11 Jan 2022

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.

Open Archive TOULOUSE Archive Ouverte (OATAO)
OATAO is an open access repository that collects the work of Toulouse researchers and makes it freely available over the web where possible.

This is an author-deposited version published in : http://oatao.univ-toulouse.fr/
Eprints ID : 12147

To link to this article : doi: 10.1007/s00162-014-0330-9
URL : http://dx.doi.org/10.1007/s00162-014-0330-9

To cite this version : Andersson, Fredrik and Carlsson, Marcus and Tourneret, Jean-Yves and Wendt, Herwig A New Frequency Estimation Method for Equally and Unequally Spaced Data. (2014) IEEE Transactions on Signal Processing, vol. 62 (n° 21). pp. 5761-5774. ISSN 1053-587X

Any correspondance concerning this service should be sent to the repository administrator: staff-oatao@listes-diff.inp-toulouse.fr

A New Frequency Estimation Method for Equally and Unequally Spaced Data

Fredrik Andersson, Marcus Carlsson, Jean-Yves Tourneret, and Herwig Wendt,

***Abstract*—Spectral estimation is an important classical problem that has received considerable attention in the signal processing literature. In this contribution, we propose a novel method for estimating the parameters of sums of complex exponentials embedded in additive noise from regularly or irregularly spaced samples. The method relies on Kronecker's theorem for Hankel operators, which enables us to formulate the nonlinear least squares problem associated with the spectral estimation problem in terms of a rank constraint on an appropriate Hankel matrix. This matrix is generated by sequences approximating the underlying sum of complex exponentials. Unequally spaced sampling is accounted for through a proper choice of interpolation matrices. The resulting optimization problem is then cast in a form that is suitable for using the alternating direction method of multipliers (ADMM). The method can easily include either a nuclear norm or a finite rank constraint for limiting the number of complex exponentials. The usage of a finite rank constraint makes, in contrast to the nuclear norm constraint, the method heuristic in the sense that the problem is non-convex and convergence to a global minimum can not be guaranteed. However, we provide a large set of numerical experiments that indicate that usage of the finite rank constraint nevertheless makes the method converge to minima close to the global minimum for reasonably high signal to noise ratios, hence essentially yielding maximum-likelihood parameter estimates. Moreover, the method does not seem to be particularly sensitive to initialization and performs substantially better than standard subspace-based methods.**

***Index Terms*—Alternating direction method of multipliers, frequency estimation, Hankel matrix, irregular sampling, Kronecker's theorem, missing data, spectral estimation.**

I. Introduction

SPECTRAL estimation is a classical problem that appears in an immense variety of problems encountered in signal and image processing applications. These applications include astronomy, radar, communications, economics, medical imaging, spectroscopy, to name but a few, (e.g., see [1] and the references therein). One important class of spectral estimation problems arises for signals that can be modeled as sums of complex exponentials (or sinusoids) in noise, i.e., for signals that admit the parametric model

$$f(x) = \sum_{p=1}^{P} a_p e^{\zeta_p x} + e(x), \quad a_p, \zeta_p \in \mathbb{C}, \tag{1}$$

where $\zeta_p = 2\pi(\gamma_p + \mathrm{i}\nu_p)$ is a complex frequency mode with frequency parameter ν_p and damping parameter γ_p, a_p is a complex amplitude and $e(x)$ is an additive noise term. Note that the model (1) includes the case of exponentially damped signals defined by $\gamma_p < 0$, with important applications notably in spectroscopy (see for instance [2], [3] and references therein). The main goal of frequency estimation is to estimate the $2P$ values of the parameter vector $\boldsymbol{\zeta} = (\zeta_1, \ldots, \zeta_P)^T$ from a vector of values $f(x_k)$ defined by (1) sampled at positions $(x_k)_{k=1}^K$, where K is the number of samples. In the present contribution, the approach to this problem in addition has the capacity to treat the case when f is sampled at *unequally spaced* positions $(x_k)_{k=1}^K$, a frequently appearing situation in many areas, for instance in spectroscopic data analysis [2], [3] and synthetic aperture radar [4], [5]. See also [6] for more examples. Note that once estimates of $\boldsymbol{\zeta}$ have been obtained, the vector of complex amplitudes $\boldsymbol{a} = (a_1, \ldots, a_P)^T$ can be determined by simple linear regression, hence the focus lies on the estimation of the modes $\boldsymbol{\zeta}$.

The estimation problem associated with (1) has been studied extensively for the case of regular sampling. Many estimation methods exist, e.g., see [7]–[12] and [1] for an overview. Some of these methods can be adapted to the unequally spaced samples, cf., [13] for a review. One prominent class of estimation techniques for the model (1) is given by subspace methods, including the popular MUSIC [14], ESPRIT [15] and min-norm [16] methods. These methods are constructed on the basis of the sample covariance of f and rely on the assumption that the noise e is white. Subspace methods can provide high-resolution estimates (i.e., below Nyquist sampling frequency resolution and equal to numerical precision in the noise free case). However, they are known to be statistically suboptimal [17]–[20] and cannot be applied to unequally spaced data samples.

As a second important class, maximum likelihood (ML) estimation has received considerable attention, cf., e.g., [1],[19], [20], for an overview, [21], [22] for asymptotic properties, [23] for an example of ML estimation for non Gaussian noise

F. Andersson and M. Carlsson are with the Centre for Mathematical Sciences, Lund University, 22100 Lund, Sweden (e-mail: fa@maths.lth.se; marcus.carlsson@math.lu.se).

J.-Y. Tourneret is with the Signal and Communication Group of the IRIT Laboratory, University of Toulouse, ENSEEIHT, 31071 Toulouse Cedex 7, France (e-mail: jean-yves.tourneret@irit.fr).

H. Wendt is with the CNRS and with the Signal and Communication Group, IRIT Laboratory, CNRS UMR 5505, University of Toulouse, ENSEEIHT, BP 7122, 31071 Toulouse Cedex 7, France (e-mail: herwig.wendt@irit.fr).

Digital Object Identifier 10.1109/TSP.2014.2358961

and [24] for ML estimation for 2D data. For Gaussian noise, ML spectral estimation consists of solving the nonlinear least squares (NLS) problem [25] associated with (1), i.e., of minimizing the cost function

$$\Gamma(\boldsymbol{a}, \boldsymbol{\zeta}) = \sum_{k=1}^{K} \left| f(x_k) - \sum_{p=1}^{P} a_p e^{\zeta_p x_k} \right|^2 \tag{2}$$

with respect to the parameter vectors $\boldsymbol{a}$ and $\boldsymbol{\zeta}$. ML estimation has been shown to be statistically optimal asymptotically and can be used for unequally spaced data. A major practical shortcoming arises, however, from the fact that it is extremely difficult to find the global minimizer of (2) because the function $\Gamma(\boldsymbol{a}, \boldsymbol{\zeta})$ is strongly multimodal, with many local minima and a very narrow valley at the global minimum (cf., e.g., [1] and references therein). As a result, performance strongly depends on the choice of initial values for the minimization procedure, limiting the practical usefulness of the NLS formulation.

As an alternative, iterative ("greedy" type) strategies [2], [26], [27] as well as the use of l_1 norm penalty [28]–[30] have been proposed. Most of these methods can be adopted for the unequally spaced data case. However, they do require the use of a discrete (finite resolution) parameter grid and are hence suboptimal, see [31].

A related problem concerns the estimation of the complex frequencies ζ_p for irregularly sampled data without making use of the model (1). The Lomb-Scargle method [32]–[35] is for instance targeted towards this problem and enables the periodogram for unequally spaced data to be computed. Its frequency resolution is by nature constrained to the Nyquist limit.

In this work, we develop a novel high-resolution method for the estimation of $\boldsymbol{\zeta}$ in (1) which seeks to minimize the ML-cost function (2) for a fixed choice of P. The proposed method relies on *Kronecker's theorem* for Hankel matrices and the *alternating direction method of multipliers* (ADMM). Kronecker's theorem [36]–[38] essentially states that if a function f is uniformly sampled, then the *Hankel matrix* that is generated by f has rank P if and only if f coincides at the sample points with a function that is a linear combination of P exponential functions. This theorem has been used for approximating functions by sums of exponentials in the alternating projections method [39] and the con-eigenvalue approach [40]. A main contribution of this paper is to show that the Kronecker theorem can be combined with the ADMM to solve the spectral estimation problem. This work is a generalization of the one presented in [41] which dealt with the particular case of regularly sampled data. A key step is that the model (1) is imposed *implicitly* by constraining the rank of the Hankel matrix generated by an approximation (denoted by g) to equal P. Moreover, we accommodate for the unequally spaced sampling through the use of appropriate interpolation matrices. As a consequence, unlike the classical NLS formulation, the minimum of (2) is not resolved directly in the parameter space $\{\boldsymbol{a}, \boldsymbol{\zeta}\}$, but in the space of vectors g which generate Hankel matrices of rank P. The frequency modes $\boldsymbol{\zeta}$ are encoded in the Hankel matrix and can be extracted by considering its con-eigenvectors. We note that this formulation is related to structured sparse matrix estimation [42]–[45] and superresolution problems [46], [47]. The optimization problem yielding the solution Hankel matrix is cast in a form that can be effectively solved by the ADMM, see [48] and references therein. This iterative technique has recently received considerable attention due to its robustness, scalability and versatility with performance comparable to dedicated problem-specific solvers.

In addition to the already mentioned areas of application, the proposed Hankel ADMM frequency estimation method can be used in the multi-dimensional case (for images notably). Furthermore, if available, prior information on the reliability (e.g., the noise variance) of data samples can be incorporated using appropriate weighting. We will first derive the general one dimensional algorithm for unequally spaced samples, then consider regular sampling and missing data as special cases that yield simpler algorithms. Finally the case of multi-dimensional data will be addressed. The procedure yields a remarkably simple and easy to implement algorithm that can be summarized in only a few lines of code.

Due to the rank constraint on the Hankel matrices, the problem considered here is nonconvex and thus there is no convergence guarantee. However, numerical experiments indicate that our Hankel ADMM formulation yields excellent parameter estimates for the model (1) without requiring specific initialization. At reasonably high signal to noise ratio (SNR), the method finds the global minimum and hence the ML solution to the problem. At low SNR values, it converges to a non-optimal solution. Our numerical experiments indicate that even in the latter case, approximations of the function f are remarkably precise. It is interesting to note that a similar algorithm could be obtained for the *convex* problem obtained by replacing the rank constraint by the nuclear norm, as proposed, e.g., in [49] for system identification. Proximity problems with either rank or nuclear norm constraints have been vastly studied during the recent years, see, e.g., [50]. Advantages and disadvantages of using either of these two approaches will be inherited in the problem considered here. For this reason, we do not aim at an exhaustive study of the advantages/disadvantages of the two penalty approaches. Our numerical simulations, however, indicate that the use of the finite rank constraint yields better estimates than the use of the nuclear norm, and we will therefore focus mainly on this approach.

The remainder of this work is organized as follows. In Section II, we recall Kronecker's theorem and its consequences for the estimation problem considered here. In Section III, we derive the optimization problem associated with the approximation of functions by sums of complex exponentials using Kronecker's theorem for unequally spaced data. We then formulate the ADMM based procedure for obtaining the Hankel matrix solution of this problem. The strategy for extracting the frequency modes from the solution Hankel matrix is summarized in Section III-C. In Section IV, we develop modified versions of our algorithm for the practically important special cases of real-valued data and equally spaced data with missing samples. Concerning the real valued case, the primary reason for the alternative formulation is that the model order is reduced by one half as compared to the version for complex data. In Section V, we evaluate the performance of the proposed method by means of numerical simulations. Comparisons to standard approaches such as ESPRIT and the Lomb-Scargle

method as well as to the iterative method in [2] are used for illustrating the performance of the method. Systematic comparisons with the theoretical Cramér-Rao bounds are also provided for quantifying optimal estimation performance (in terms of estimation variance). Section VI develops and illustrates the 2D case. Conclusions and future work are reported in Section VII.

II. Hankel Matrices and Complex Exponentials

A. Kronecker's Theorem for Complex Exponentials

In the remainder of the paper, we assume the problem to be normalized so that f in (1) is a function defined on the interval $[-\frac{1}{2}, \frac{1}{2}]$. Suppose for the moment that we are given sampled values of f at the equally spaced nodes x^{eq} on the interval $|x| \leq \frac{1}{2}$ defined by

$$x_n^{\mathrm{eq}} = \frac{n}{2N}, \quad -N \leq n \leq N. \tag{3}$$

We denote by $\boldsymbol{f} = (f_1, \ldots, f_{2N+1})$ the vector whose elements are the sampled values $f(x_n^{\mathrm{eq}})$, i.e.,

$$f_n = f\left(x_{n-(N+1)}^{\mathrm{eq}}\right), \quad 1 \leq n \leq 2N+1. \tag{4}$$

For completeness, we recall that a Hankel matrix $\boldsymbol{A}$ has constant anti-diagonals terms, i.e., $\boldsymbol{A}(j,k) = \boldsymbol{A}(j',k')$ if $j+k = j'+k'$. It can thus be generated elementwise from a vector $\boldsymbol{g} = (g_1, \ldots, g_{2N+1})$, such that $\boldsymbol{A}(j,k) = \boldsymbol{g}_{j+k-1}$, $1 \leq j,k \leq N+1$. The Hankel matrix generated by the vector $\boldsymbol{g}$ will be denoted by $H(\boldsymbol{g})$, and we denote the set of Hankel matrices by $\mathcal{H}$.

Kronecker's theorem [36]–[38] states that if the Hankel matrix $\boldsymbol{A} = H(\boldsymbol{g})$ is of rank P, with the exception of degenerate cases, there exists $\{\zeta_p\}_{p=1}^P$ and $\{a_p\}_{p=1}^P$ in $\mathbb{C}$ such that $\boldsymbol{g}$ is obtained by sampling at the equally spaced nodes x^{eq} the function

$$g(x) = \sum_{p=1}^{P} a_p e^{\zeta_p x}. \tag{5}$$

The converse holds as well.

It follows that the best approximation (in the l^2 sense) of $\boldsymbol{f}$ by a linear combination of P complex exponentials is given by the vector $\boldsymbol{g}$ which satisfies $\mathrm{rank}[H(\boldsymbol{g})] = P$ and minimizes the l^2 norm of the difference $\boldsymbol{f} - \boldsymbol{g}$. In other words, $\boldsymbol{g}$ can be obtained as the solution of the optimization problem

$$\begin{aligned} &\underset{\boldsymbol{g} \in \mathbb{C}^{2N+1}}{\mathrm{minimize}} \frac{1}{2} \sum_{n=1}^{2N+1} |f_n - g_n|^2 \\ &\text{subject to} \quad \mathrm{rank}[H(\boldsymbol{g})] = P. \end{aligned} \tag{6}$$

III. Approximation and Frequency Estimation for Unequally Spaced Data

The problem considered here is to reconstruct the function g of the form (5) from knowledge of f as in (1) at the *unequally spaced* sample nodes x_m for $m = 1, \ldots, M$. The optimization problem (6) does not apply directly to this situation because Kronecker's theorem requires the function f to be uniformly sampled. For this reason, we will work simultaneously with the two sets of nodes; $(x_m)_{m=1}^M$ and the equally spaced nodes $(x_n^{\mathrm{eq}})_{n=-N}^N$ defined in (3). Given a function f or g, represented by its sample points $\boldsymbol{f}$ or $\boldsymbol{g}$ via (4), we will use interpolation to evaluate the function at $(x_m)_{m=1}^M$.

We assume that the sampling at the equally spaced grid is sufficiently dense so that we can approximately recover the exponentials we are interested in by interpolation (essentially the Nyquist sampling rate, but with proper adjustments for it being on a finite interval [51]). We propose to recover the function at the points $(x_m)_{m=1}^M$ by discrete convolution with a suitably chosen function φ. The corresponding interpolation matrix $\boldsymbol{\mathcal{K}}$ that maps samples from the equally spaced grid $(x_n^{\mathrm{eq}})_{n=-N}^N$ to samples from the unequally spaced grid $(x_m)_{m=1}^M$ is then given by

$$\boldsymbol{\mathcal{K}}(n,m) = \varphi(2N(x_n^{\mathrm{eq}} - x_m)) \tag{7}$$

and $g(x_m) = \left(\boldsymbol{\mathcal{K}}\boldsymbol{g}\right)_m$, where $(\boldsymbol{\mathcal{K}}\boldsymbol{g})_m$ is the interpolated value of g at x_m. Note that an exact reproduction of all functions of the form (5) is impossible, even with bounds on the values of ζ_p. The function φ is preferably a compactly supported function that is reminiscent of the sinc-function used in the Whittaker-Shannon interpolation formula. One can choose φ to satisfy the *Strang-Fix* conditions whereby $\boldsymbol{\mathcal{K}}$ becomes capable of reproducing polynomials of orders $\leq m$ [52]. Another possibility is Key's interpolating function [53], [54]. A third alternative is to use the Lanczos function

$$\varphi_{\mathrm{Lanczos}}(x) = \begin{cases} \mathrm{sinc}(x)\mathrm{sinc}(x/a) & \text{if } 0 \leq |x| < a \\ 0 & \text{if } |x| \geq a \end{cases}$$

where the number of lobes is controlled by the integer parameter $a > 1$. However, in this case $\boldsymbol{\mathcal{K}}$ does not interpolate constant functions correctly. A simple (but not optimal) remedy that is commonly used to account for this deficiency is to normalize the matrix $\boldsymbol{\mathcal{K}}$ by the sums of its rows. This trick also accounts for boundary problems where the finite support of φ falls outside the equally spaced nodes x^{eq}. The Lanczos function is very popular due to its flexibility in controlling the number of lobes and its simplicity, whereas functions satisfying, e.g., the Strang-Fix conditions quickly become complicated as m increases. In our simulations, we have used $\varphi_{\mathrm{Lanczos}}$ with $a = 9$ and normalized $\boldsymbol{\mathcal{K}}$.

A. Problem Formulation for Unequally Spaced Data

With these definitions, the function g that defines the best approximation of f by a linear combination of P complex exponentials from the sampled values $f(x_m)$ can be obtained as the solution of the optimization problem

$$\begin{aligned} &\underset{\boldsymbol{g} \in \mathbb{C}^{2N+1}}{\mathrm{minimize}} \quad \frac{1}{2} \sum_{m=1}^{M} \mu_m \left| \left(\boldsymbol{\mathcal{K}}\boldsymbol{g}\right)_m - f(x_m) \right|^2 \\ &\text{subject to} \quad \mathrm{rank}\big[H(\boldsymbol{g})\big] = P \end{aligned} \tag{8}$$

where $\mu_m \geq 0$ are weights that can be chosen depending on the specific application. We will usually work with the uniform case $\mu_m \equiv 1$, but there can be various reasons for choosing other weights. For example, if we want the l^2-norm in (8) to be as close as possible to the L^2-norm of the corresponding function $f - g$, i.e., $\int_{-1/2}^{1/2} |f(x) - g(x)|^2 dx$, then one should let μ_m

be inversely proportional to the local density of samples near x_m. Another situation could be if data is more reliable in certain regions or sample points. In this case, one is clearly interested in giving higher weights to these measurements. Lastly, these weights are suitable for dealing with the case of missing data on a regular grid, as explained in detail in Section IV-B.

B. ADMM Based Approximate Solution

Let $\mathcal{R}_P$ be an indicator function for $(2N+1)\times(2N+1)$ matrices $\boldsymbol{A} = (A_{i,j})_{j,k=1,\ldots,2N+1}$ defined by

$$\mathcal{R}_P(\boldsymbol{A}) = \begin{cases} 0 & \text{if rank}[\boldsymbol{A}] \leq P \\ \infty & \text{otherwise.} \end{cases}$$

The approximation/reconstruction problem (8) can now be formulated as the following optimization problem

$$\begin{aligned} \underset{\boldsymbol{A},\boldsymbol{g}}{\text{minimize}} \quad & \mathcal{R}_P(\boldsymbol{A}) + \frac{N}{2M}\sum_{m=1}^{M}\mu_m\Big|\big(\mathcal{K}\boldsymbol{g}\big)_m - f(x_m)\Big|^2 \\ \text{subject to} \quad & \boldsymbol{A}_{j,k} = \boldsymbol{g}_{j+k-1} \text{for} j,k = 1,\ldots,2N+1. \end{aligned} \tag{9}$$

The optimization problem (9) has a cost function that consists of two terms, each depending only on one of the variables $\boldsymbol{A}$ and $\boldsymbol{g}$, along with a linear constraint. The problem formulation is therefore well suited to be solved using ADMM. ADMM is an iterative technique in which a solution to a large global problem is found by coordinating solutions to smaller subproblems. It can be seen as an attempt to merge the benefits of dual decomposition and augmented Lagrangian methods for constrained optimization. For an overview of the ADMM method see [48]. For a convex cost function, the ADMM is guaranteed to converge to the unique solution (if the parameter ρ of the augmented Lagrangian (10) below tends to infinity during the iterations). For non-convex problems, the algorithm can get stuck at local minima, although it can work quite well in practice also in these situations, cf. [48, Chapter 9], see also, e.g., [55]. Since the rank constraint $\mathcal{R}_P$ is nonconvex, the convergence of the ADMM scheme associated with (9) is not guaranteed. However, our numerical simulations indicate that the method works substantially better than established high resolution techniques like ESPRIT. Moreover, it can be applied to cases where high resolution methods such as MUSIC or ESPRIT are non-applicable, e.g., to unequally spaced or missing data.

Note that a convex problem could be obtained by replacing the $\mathcal{R}_P$ term by the nuclear norm (as in [49], at the cost of biased solutions. The numerical experiments reported in Section V-F indicate that the solutions obtained when using $\mathcal{R}_P$ are systematically better than those obtained using the nuclear norm.

Given matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, we write $\langle \boldsymbol{A},\boldsymbol{B}\rangle_{\mathtt{Re}} = \mathtt{Re}\left\{\sum_{j,k}\boldsymbol{A}(j,k)\overline{\boldsymbol{B}(j,k)}\right\}$ for the scalar product, and we use $\|\boldsymbol{A}\|_2^2$ for the Frobenius norm $\langle \boldsymbol{A},\boldsymbol{A}\rangle_{\mathtt{Re}}$. Note that the linear constraint in (9) can be written as $\boldsymbol{A} = H(\boldsymbol{g})$. Consequently, the Lagrange multiplier is naturally represented as a $(2N+1)\times(2N+1)$-matrix $\boldsymbol{\Lambda}$. The augmented Lagrangian associated with (9) then takes the form

$$\begin{aligned} L(\boldsymbol{A},\boldsymbol{g},\boldsymbol{\Lambda}) = \mathcal{R}_P(\boldsymbol{A}) &+ \frac{N}{2M}\sum_{m=1}^{M}\mu_m\Big|(\mathcal{K}\boldsymbol{g})_m - f(x_m)\Big|^2 \\ &+ \langle \boldsymbol{\Lambda}, \boldsymbol{A} - H(\boldsymbol{g})\rangle_{\mathtt{Re}} + \frac{\rho}{2}\|\boldsymbol{A} - H(\boldsymbol{g})\|_2^2. \end{aligned} \tag{10}$$

The factor $\frac{N}{2M}$ in (9) has the effect that the ratio of the two norms becomes independent of the sampling sizes M and N. Hence, the size of ρ (which needs to be chosen in the algorithm) does not become affected by the sample densities. The corresponding ADMM iterative steps are

$$\boldsymbol{A}^{q+1} = \underset{\boldsymbol{A}}{\text{argmin}}\ L(\boldsymbol{A},\boldsymbol{g}^q,\boldsymbol{\Lambda}^q), \tag{11}$$

$$\boldsymbol{g}^{q+1} = \underset{\boldsymbol{g}}{\text{argmin}}\ L(\boldsymbol{A}^{q+1},\boldsymbol{g},\boldsymbol{\Lambda}^q), \tag{12}$$

$$\boldsymbol{\Lambda}^{q+1} = \boldsymbol{\Lambda}^q + \rho\big(\boldsymbol{A}^{q+1} - H(\boldsymbol{g}^{q+1})\big). \tag{13}$$

1) Minimization With Respect to (w.r.t.) the Objective Variables $\boldsymbol{A}$: To compute the solution of the minimization (11), we drop terms of the augmented Lagrangian $L(\boldsymbol{A},\boldsymbol{g},\boldsymbol{\Lambda})$ defined in (10) that are independent of $\boldsymbol{A}$ and obtain

$$\begin{aligned} \boldsymbol{A}^{q+1} &= \underset{\boldsymbol{A}}{\text{argmin}}\ \mathcal{R}_P(\boldsymbol{A}) + \langle \boldsymbol{\Lambda}^q, \boldsymbol{A}\rangle_{\mathtt{Re}} + \frac{\rho}{2}\|\boldsymbol{A} - H(\boldsymbol{g}^q)\|_2^2 \\ &= \underset{\boldsymbol{A}}{\text{argmin}}\ \mathcal{R}_P(\boldsymbol{A}) + \frac{\rho}{2}\Big\|\boldsymbol{A} - \boldsymbol{B}^q\Big\|_2^2 \end{aligned}$$

where $\boldsymbol{B}^q$ is the following matrix

$$\boldsymbol{B}^q = H(\boldsymbol{g}^q) - \frac{\boldsymbol{\Lambda}^q}{\rho}.$$

It follows that $\boldsymbol{A}^{q+1}$ is equal to the best rank P approximation of the matrix $\boldsymbol{B}^q$. Let $\boldsymbol{B}^q = \boldsymbol{U}^q\boldsymbol{\Sigma}^q(\boldsymbol{V}^q)^*$ be the singular value decomposition of $\boldsymbol{B}^q$. It then follows by the Eckart-Young theorem that

$$\boldsymbol{A}^{q+1} = \boldsymbol{U}^q\widetilde{\boldsymbol{\Sigma}}^q(\boldsymbol{V}^q)^* \tag{14}$$

where $\widetilde{\boldsymbol{\Sigma}}^q$ is the diagonal matrix satisfying

$$\widetilde{\boldsymbol{\Sigma}}^q_{l,l} = \begin{cases} \boldsymbol{\Sigma}^q_{l,l} & \text{if } l \leq P \\ 0 & \text{otherwise.} \end{cases}$$

2) Minimization w.r.t. the Objective Variables $\boldsymbol{g}$: The second step (12) in the ADMM iteration is solved by the least squares (LS) method. More precisely, after removing terms (10) that do not depend on $\boldsymbol{g}$, incorporating the term $\langle \boldsymbol{\Lambda}, H(\boldsymbol{g})\rangle$ as in the previous section and finally multiplying everything by 2, we see that $\boldsymbol{g}^{q+1}$ is given by the minimum of

$$\frac{N}{M}\sum_{m=1}^{M}\mu_m\Big|\big(\mathcal{K}\boldsymbol{g}\big)_m - f(x_m)\Big|^2 + \rho\left\|\boldsymbol{A}^{q+1} + \frac{\boldsymbol{\Lambda}^q}{\rho} - H(\boldsymbol{g})\right\|_2^2.$$

To simplify this expression, we introduce the weights

$$\boldsymbol{w}_n = \sum_{j+k-1=n} 1 = \begin{cases} n & \text{if } n \leq N+1 \\ N+2-n & \text{otherwise} \end{cases}$$

where it is understood that the summation takes place for $1 \leq j,k \leq N+1$. Note that $\boldsymbol{w}_n$ is the number of times that $\boldsymbol{g}_n$ appears in the matrix $H(\boldsymbol{g})$. We also introduce the vector

$$\boldsymbol{b} = \left(\sum_{j+k-1=n}\boldsymbol{A}^{q+1}_{j,k} + \frac{\boldsymbol{\Lambda}^q_{j,k}}{\rho}\right)_{n=1}^{2N+1} \tag{15}$$

which is obtained by summing the terms of $\boldsymbol{A}^{q+1} + \frac{\boldsymbol{\Lambda}^q}{\rho}$ anti-diagonally. It is easy to check that the term $\rho \left\| \boldsymbol{A}^{q+1} + \frac{\boldsymbol{\Lambda}^q}{\rho} - H(\boldsymbol{g}) \right\|_2^2$ differs only by a constant from

$$\rho \sum_{n=1}^{2N+1} \boldsymbol{w}_n \left| \boldsymbol{g}_n - \boldsymbol{b}_n / \boldsymbol{w}_n \right|^2 .$$

Summing up we have that

$$\boldsymbol{g}^{q+1} = \operatorname*{argmin}_{\boldsymbol{g}} \sum_{m=1}^{M} \left| \sqrt{\frac{N\mu_m}{M}} (\boldsymbol{\mathcal{K}}\boldsymbol{g})_m - \sqrt{\frac{N\mu_m}{M}} f(x_m) \right|^2 + \\ + \sum_{n=1}^{2N+1} \left| \left(\sqrt{\rho} \operatorname{diag}(\sqrt{\boldsymbol{w}}) \boldsymbol{g} - \sqrt{\rho} \operatorname{diag}(\sqrt{\boldsymbol{w}}^{-1}) \boldsymbol{b} \right)_n \right|^2 \quad (16)$$

with $\sqrt{\boldsymbol{w}}^{-1} = (1/\sqrt{w_1}, \ldots, 1/\sqrt{w_{2N+1}})$. The standard LS estimation theory leads to

$$\boldsymbol{g}^{q+1} = \left(\frac{N}{M} \boldsymbol{\mathcal{K}}^* \operatorname{diag}(\boldsymbol{\mu}) \boldsymbol{\mathcal{K}} + \rho \operatorname{diag}(\boldsymbol{w}) \right)^{-1} \\ \left(\frac{N}{M} \boldsymbol{\mathcal{K}}^* \operatorname{diag}(\boldsymbol{\mu}) \big(f(x_m) \big)_{m=1}^{M} + \rho \boldsymbol{b} \right) \quad (17)$$

where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_M)$. This completes the derivation of the ADMM algorithm for solving (9). The most time consuming operation in each iteration step is SVD computation of $\boldsymbol{B}^q$. The time complexity of the SVD operation is $\mathcal{O}(N^3)$ and hence the time complexity of the total algorithm is $\mathcal{O}(\kappa N^3)$, where κ is the amount of iterations. For comparison, the ESPRIT method requires the computation of one single SVD.

3) Stopping Criterion: Classical stopping criteria are based on the primal and dual residuals, which are given by

$$\boldsymbol{R}^q = \boldsymbol{A}^q - H(\boldsymbol{g}^q), \qquad \boldsymbol{S}^q = \rho H\big(\boldsymbol{g}^q - \boldsymbol{g}^{q-1}\big).$$

The ADMM iterations terminate when the stopping criteria

$$\|\boldsymbol{R}^q\|_2^2 \le \epsilon^{\text{prim}} \quad \text{and} \quad \|\boldsymbol{S}^q\|_2^2 \le \epsilon^{\text{dual}}$$

are satisfied, where $\epsilon^{\text{prim}} > 0$ and $\epsilon^{\text{dual}} > 0$ can be chosen using absolute and relative tolerances $\epsilon^{\text{abs}} > 0$ and $\epsilon^{\text{rel}} > 0$ [48], i.e.,

$$\epsilon^{\text{prim}} = (2N+1)^2 \epsilon^{\text{abs}} + \epsilon^{\text{rel}} \max\{\|\boldsymbol{A}^q\|_2^2, \|H(\boldsymbol{g}^q)\|_2^2\} \\ \epsilon^{\text{dual}} = (2N+1)^2 \epsilon^{\text{abs}} + \epsilon^{\text{rel}} \|\boldsymbol{\Lambda}^q\|_2^2.$$

We denote by $\hat{\boldsymbol{g}}$ and $\hat{\boldsymbol{A}}$ the values obtained at convergence. Due to the term $\rho\|\boldsymbol{A} - H(\boldsymbol{g})\|^2$ in (10), one usually does not have $\boldsymbol{A} = H(\hat{\boldsymbol{g}})$. In other words, neither $\hat{\boldsymbol{g}}$ nor the vector obtained by averaging the anti-diagonals terms of $\boldsymbol{A}$, i.e.,

$$\boldsymbol{f}_{\text{approx}} = \left(\frac{1}{\boldsymbol{w}_n} \sum_{j+k-1=n} \hat{\boldsymbol{A}}(j,k) \right)_{n=1}^{2N+1}$$

will be equal to a sum of P exponential functions. One way to overcome this problem is to let the values of ρ gradually increase to ∞ as the algorithm gets closer to convergence. We will not pursue this further here, thus keeping ρ fixed. In this case, we have observed that $\boldsymbol{f}_{\text{approx}}$ is a more suitable choice than $\hat{\boldsymbol{g}}$. This is to be expected since (16) forces the iterates of $\boldsymbol{g}$ to be closer to the usually noisy signal f.

C. Frequency and Amplitude Estimation

Once the solution $\boldsymbol{f}_{\text{approx}}$ has been found, the parameter vector $\boldsymbol{\zeta}$ can be obtained from the singular value decomposition of $H(\boldsymbol{f}_{\text{approx}})$. We outline a method for this below; assuming for simplicity that $\boldsymbol{f}_{\text{approx}}$ exactly equals a sum of P exponential functions. Since the rank of $H(\boldsymbol{f}_{\text{approx}})$ then equals P, its singular value decomposition can be written

$$H(\boldsymbol{f}_{\text{approx}}) = \sum_{p=1}^{P} \sigma_p \boldsymbol{u}_p \boldsymbol{v}_p^* \quad (18)$$

where $\boldsymbol{u}_p$ are the singular (column)-vectors corresponding to the non-zero singular values $(\sigma_p)_{p=1}^{P}$. Since $\boldsymbol{f}_{\text{approx}}$ is (a sampled version of) a sum of P exponentials, each $\boldsymbol{u}_p$ will also be a sum of the same exponentials. Let

$$\boldsymbol{U} = \boldsymbol{U}_P = (\boldsymbol{u}_1, \ldots, \boldsymbol{u}_P). \quad (19)$$

By (5), it is possible to write $\boldsymbol{U} = \boldsymbol{E}\boldsymbol{G}$, where $\boldsymbol{E}$ is the $(2N+1) \times P$ Vandermonde matrix generated by sampling the functions $e^{\zeta_p x}$, i.e.,

$$\boldsymbol{E}(n,p) = e^{\zeta_p x_{n-N-1}^{\text{eq}}} = e^{\zeta_p (n-N-1)/2N}$$

and $\boldsymbol{G}$ is some (invertible) $P \times P$ matrix. Denote by $\boldsymbol{U}^f$ (resp. $\boldsymbol{U}^l$) the matrix $\boldsymbol{U}$ whose first row (resp. last row) has been dropped. Clearly, we have $\boldsymbol{U}^f = \boldsymbol{E}^f \boldsymbol{G}$, $\boldsymbol{U}^l = \boldsymbol{E}^l \boldsymbol{G}$ and the Vandermonde structure of $\boldsymbol{E}$ leads to $\boldsymbol{E}^f = \boldsymbol{E}^l \operatorname{diag}(e^{\zeta_0/2N}, \ldots, e^{\zeta_{P-1}/2N})$. It follows that

$$\boldsymbol{U}^{l\dagger} \boldsymbol{U}^f = \boldsymbol{G}^{-1} \operatorname{diag}(e^{\zeta_0/2N}, \ldots, e^{\zeta_{P-1}/2N}) \boldsymbol{G},$$

where $\dagger$ denotes the pseudo inverse. Therefore, we can compute the modes $\boldsymbol{\zeta}$ by computing the eigenvalues of $\boldsymbol{U}^{l\dagger}\boldsymbol{U}^f$, i.e., as $\boldsymbol{\zeta} = 2N \log \left\{ \operatorname{eig} \left[\boldsymbol{U}^{l\dagger} \boldsymbol{U}^f \right] \right\}$. This method can be computed in $\mathcal{O}(NP^2 + P^3)$ time. Note that in the case of regular sampling, the direct application of the above scheme to the samples $f(x^{\text{eq}})$ instead of $\boldsymbol{f}_{\text{approx}}$ yields the ESPRIT method as described in [1].

Once the frequency modes $\boldsymbol{\zeta}$ have been computed, the amplitudes $\boldsymbol{a}$ are obtained by solving the linear LS problem

$$\Pi \boldsymbol{a} = \big(f(x_m) \big)_{m=1}^{M}, \quad \Pi = \begin{pmatrix} e^{x_1 \zeta_0} & \ldots & e^{x_1 \zeta_{P-1}} \\ \vdots & & \vdots \\ e^{x_M \zeta_0} & \ldots & e^{x_M \zeta_{P-1}} \end{pmatrix}.$$

The complete algorithm in the case of unequally spaced data with uniform weights $\mu_m \equiv 1$, is summarized in the MATLAB function described in Table I.

D. Model Order Estimation

As for the ESPRIT method, the number of modes, or the model order, to be estimated is an input parameter to the respective algorithms. We briefly discuss how to adapt one pop-

TABLE I

MAIN FUNCTION (IN MATLAB SYNTAX) FOR THE FREQUENCY ESTIMATION PROBLEM (1) (IN THE CASE WITH UNIFORM WEIGHTS $\mu_m \equiv 1$) FOR DATA SAMPLES `fus` ACQUIRED AT UNEQUALLY SPACED NODES `xus` (TOP), AND AUXILIARY FUNCTIONS `form_hank`, `sum_hank` AND `param_est` USED FOR HANKEL MATRIX MANIPULATION AND PARAMETER ESTIMATION (BOTTOM). THE MATRIX $\mathcal{K}$ IS COMPUTED FROM (7) WITH A SUITABLE CHOICE OF φ

```
function [fa,zeta,a]=expo_est_us(xus,fus,P,N,K,rho,niter)
  xeq=[-N:N]'/(2*N); M=length(xus); g=K\fus;
  Lambda=zeros(N+1);w=[1:N+1,N:-1:1]';
  for iter=1:niter,
    B=(form_hank(g)-Lambda/rho);
    [u,s,v]=svds(B,P);A=u*s*v';
    g=(N/M*K'*K+rho*diag(w))\
        (N/M*K'*fus+sum_hank(rho*A+Lambda));
    Lambda=Lambda+rho*(A-form_hank(g));
  end;
  fa=sum_hank(A)./w;
  [zeta,a] = param_est(fa,P,xeq,N)
```

```
function H=form_hank(f)            % form Hankel matrix
  H=hankel(f(1:(end+1)/2),f((end+1)/2:end));
function f=sum_hank(H)             % sum anti-diagonals
  N=size(H,1)-1; f=zeros(2*N+1,1); H=flipud(H);
  for n=-N:N, f(n+N+1,1)=sum(diag(H,n)); end;
function [zeta,a] = param_est(g,P,xeq,N) % estimation
  [u,s,v]=svds(form_hank(g),P);S=zeros(2*N+1,P);
  zeta=2*N*log(eig(u(1:end-1,1:P)\u(2:end,1:P)));
  for p=1:P,S(:,p)=exp(zeta(p)*xeq);end;a=S\g;
```

ular method, namely the ESTER method introduced in [56], for unequally spaced data. For any p, let

$$E(p) = \boldsymbol{U}_p^f - \boldsymbol{U}_p^l {\boldsymbol{U}_p^l}^\dagger \boldsymbol{U}_p^f, \tag{20}$$

using the notation of Section III-C for $\boldsymbol{U}_p^f$ and $\boldsymbol{U}_p^l$, respectively. In the absence of noise, the ESTER method relies on the fact that if the function that is generating the Hankel matrix in (18) is a linear combination of P exponential functions, then the matrix $E(p)$ will vanish for $p = P$, but not for other values. For this reason, the ESTER method uses the quantity $J(p) = \|E(p)\|^{-2}$ as a measure of the model order.

Since the ESPRIT method will not work for unequally spaced data or missing data, the original ESTER approach can not be applied to these cases. However, we propose a modification of the ESTER method to adapt to the methods described in this paper. For each p, we denote by $\boldsymbol{g}_p$ the computed approximate solution to (8). We now define $\boldsymbol{U}_{p,r} = (\boldsymbol{u}_1^p, \ldots, \boldsymbol{u}_r^p)$ where $\boldsymbol{u}^p$ are the (left) singular vectors of $H(\boldsymbol{g}_p)$. To replace (20), one can verify that in the same spirit as done for the original ESTER, the matrix

$$E_{\texttt{Hadmm}}(p) = \boldsymbol{U}_{p+1,p}^f - \boldsymbol{U}_{p,p}^l {\boldsymbol{U}_{p,p}^l}^\dagger \boldsymbol{U}_{p+1,p}^f \tag{21}$$

equals zero (in the absence of noise and approximation errors) when $p = P$ if the original data was a sum of P exponential functions, but not for other values of p. As a motivation to the form (21), consider the case with no noise present and where the data sampling is sufficient to correctly recover the P exponential functions when $p \geq P$. The $P+1$-th singular value of the Hankel matrix generated by $\boldsymbol{g}_{P+1}$ is zero, and thus the $\boldsymbol{U}_{P+1,P}$ and $\boldsymbol{U}_{P,P}$ are generated by the same exponential functions, implying that $E_{\texttt{Hadmm}}(P) = 0$. Consequently, we propose to estimate the model order P by looking for the maximum value of $J(p) = \|E_{\texttt{Hadmm}}(p)\|^{-2}$.

E. Structure of the Iterates

Hankel matrices belong to the class of complex symmetric matrices satisfying $\boldsymbol{A} = \boldsymbol{A}^T$. For these matrices, one can choose the matrices $\boldsymbol{U}$ and $\boldsymbol{V}$ appearing in the singular value decomposition of $\boldsymbol{A} = \boldsymbol{U}\widetilde{\boldsymbol{\Sigma}}\boldsymbol{V}^*$ such that $\boldsymbol{U} = \overline{\boldsymbol{V}}$ (where the bar denotes complex conjugation). This gives rise to the so called Takagi factorization of $\boldsymbol{A}$,

$$\boldsymbol{A} = \sum_{n=0}^{N} s_n \boldsymbol{u}_n \boldsymbol{u}_n^T$$

cf., [57]. This property can be used to simplify the expressions (14) and (18). Moreover, it can also be used to show that the iterates $\boldsymbol{A}^q$, $\boldsymbol{B}^q$ and $\boldsymbol{\Lambda}^q$ are complex symmetric matrices.

IV. REAL-VALUED AND EQUALLY SPACED DATA

In this section, we derive algorithms for the practically important special cases of real data, and of uniformly sampled data, including situations where samples are missing.

A. Real-Valued Data

In the case of real valued data, the frequency modes appear in conjugate pairs $(\zeta_p, \overline{\zeta_p})$, which leads to the representation

$$f(x) = \sum_{p=1}^{P} \tilde{a}_p \left(e^{\zeta_p x + \mathrm{i}\varphi_p} + e^{\overline{\zeta_p x + \mathrm{i}\varphi_p}} \right) \tag{22}$$

where $\tilde{\boldsymbol{a}} \in \mathbb{R}^P$ and $\boldsymbol{\varphi} \in \mathbb{R}^P$ are vectors containing real amplitudes and phase shifts. The application of the algorithm derived in the previous section would thus require the estimation of $2P$ complex exponentials with frequency modes $\boldsymbol{\zeta}$ and $\overline{\boldsymbol{\zeta}}$ in (22). However, we can modify the problem formulation such that the approximation is obtained as the *real part* of the linear combination of P *complex* exponentials

$$\sum_{p=1}^{P} a_p e^{\zeta_p x}, a_p = \tilde{a}_p e^{\mathrm{i}\varphi_p} \in \mathbb{C}. \tag{23}$$

This formulation is interesting since it allows fewer parameters (half as many as in the complex case) to be estimated. The corresponding optimization problem reads (cf. equation (9))

$$\begin{aligned} \underset{\boldsymbol{A},\boldsymbol{g}}{\text{minimize}} \quad & \mathcal{R}_P(\boldsymbol{A}) + \frac{N}{2M} \sum_{m=1}^{M} \mu_m \Big| \texttt{Re}(\mathcal{K}\boldsymbol{g}) - f(x_m) \Big|^2 \\ \text{subject to} \quad & \boldsymbol{A} = H(\boldsymbol{g}). \end{aligned} \tag{24}$$

We solve the above problem as in Section III, but updating the real vector $\tilde{\boldsymbol{g}} = (\texttt{Re}(\boldsymbol{g}), \texttt{Im}(\boldsymbol{g}))^T$ instead of $\boldsymbol{g}$. Clearly, the first minimization step yielding $\boldsymbol{A}^{q+1}$ remains unaltered and is given by (14). Likewise, the update of $\boldsymbol{\Lambda}^{q+1}$ given by (13) is unchanged. However, the projection algorithm corresponding to (17) needs to be modified. Since $\mathcal{K}$ is real valued we have $\texttt{Re}(\mathcal{K}\boldsymbol{g}) = [\mathcal{K} \quad \boldsymbol{0}]\tilde{\boldsymbol{g}}$, where $\boldsymbol{0}$ denotes the zero matrix and

TABLE II
FUNCTION (IN MATLAB SYNTAX) FOR THE APPROXIMATION WITH WEIGHTS `mu` AND FREQUENCY ESTIMATION USING `P` COMPLEX EXPONENTIALS FOR EQUALLY SPACED DATA `f`

```
function [g,zeta,a]=expo_est_eq_w(f,mu,P,rho,niter)
  N=(length(f)-1)/2;M=length(mu(mu==1));g=f;
  Lambda=zeros(N+1);w=[1:N+1,N:-1:1]';
  for iter=1:niter,
    B=form_hank(g)-Lambda/rho;[u,s,v]=svds(B,P);A=u*s*v';
    g=(N/M*mu.*f+sum_hank(rho*A+Lambda))./(N/M*mu+rho*w);
    Lambda=Lambda+rho*(A-form_hank(g));
  end;
  fa=sum_hank(A)./w;
  [zeta,a] = param_est(fa,P,[-N:N]'/2/N,N)
```

$[\boldsymbol{\mathcal{K}} \quad \mathbf{0}]$ the concatenation of $\boldsymbol{\mathcal{K}}$ and $\mathbf{0}$. Following the calculations yielding (16), we obtain

$$\begin{aligned}\tilde{\boldsymbol{g}}^{q+1} = \underset{\tilde{\boldsymbol{g}}}{\operatorname{argmin}} &\left\| \sqrt{\frac{N\mu_m}{M}}\,[\boldsymbol{\mathcal{K}} \quad \mathbf{0}]\,\tilde{\boldsymbol{g}} - \sqrt{\frac{N\mu_m}{M}}\,(f(x_m))_m \right\|^2 \\ &+ \left\| \sqrt{\rho} \begin{bmatrix} \operatorname{diag}(\sqrt{\boldsymbol{w}}) & \mathbf{0} \\ \mathbf{0} & \operatorname{diag}(\sqrt{\boldsymbol{w}}) \end{bmatrix} \tilde{\boldsymbol{g}} \right. \\ &\left. - \sqrt{\rho} \begin{pmatrix} \operatorname{diag}(\sqrt{\boldsymbol{w}}^{-1})\mathtt{Re}(\boldsymbol{b}) \\ \operatorname{diag}(\sqrt{\boldsymbol{w}}^{-1})\mathtt{Im}(\boldsymbol{b}) \end{pmatrix} \right\|^2\end{aligned}$$

where $\boldsymbol{b}$ is unchanged and has been defined in (15). The step corresponding to (17) then has to be modified as follows

$$\begin{aligned}\tilde{\boldsymbol{g}}^{q+1} = &\begin{bmatrix} \frac{N}{M}\boldsymbol{\mathcal{K}}^*\operatorname{diag}(\boldsymbol{\mu})\boldsymbol{\mathcal{K}} + \rho\operatorname{diag}(\boldsymbol{w}) & \mathbf{0} \\ \mathbf{0} & \rho\operatorname{diag}(\boldsymbol{w}) \end{bmatrix}^{-1} \\ &\begin{pmatrix} \frac{N}{M}\boldsymbol{\mathcal{K}}^*\operatorname{diag}(\boldsymbol{\mu})\big(f(x_m)\big)_m + \rho\mathtt{Re}(\boldsymbol{b}) \\ \rho\mathtt{Im}(\boldsymbol{b}) \end{pmatrix}. \end{aligned} \quad (25)$$

The ADMM algorithm for solving (24) consists of iterating the steps (13), (14) and (25) until the stopping criteria are satisfied.

B. Equally Spaced Data With Missing Points

We now consider the case when $\{x_m\}_{m=1}^M$ coincide with $\{x_n^{\mathrm{eq}}\}_{n=-N}^N$, but where samples are missing at indices $j \in \mathcal{J}$, i.e., $\{x_m\}_{m=1}^M \subset \{x_n^{\mathrm{eq}}\}_{n=-N}^N$. This case is potentially interesting for the numerous signal and image processing applications requiring missing data to be handled (see [6] for examples and more details). This situation can be handled by simply replacing $\{x_m\}_{m=1}^M$ by the equally spaced grid $\{x_n^{\mathrm{eq}}\}_{n=-N}^N$ and instead defining the following weights

$$\mu(j) = \mu_{\mathcal{J}}(j) = \begin{cases} 0 & \text{if } j \in \mathcal{J} \\ 1 & \text{otherwise.} \end{cases}$$

In this case, we obtain a simple algorithm, since we do not need to worry about interpolation, allowing the matrix $\boldsymbol{\mathcal{K}}$ to be set to identity. Moreover, M equals the number of ones in μ, and (17) simplifies to

$$\boldsymbol{g}^{q+1} = \left(\frac{N}{M}\operatorname{diag}(\boldsymbol{\mu}) + \rho\operatorname{diag}(\boldsymbol{w})\right)^{-1} \left(\frac{N}{M}\operatorname{diag}(\boldsymbol{\mu})\boldsymbol{f} + \rho\boldsymbol{b}\right).$$

Note that the matrix to be inverted is diagonal, making this step faster to execute than (17) and (25). The corresponding ADMM algorithm is summarized in the simple MATLAB function of Table II.

V. RESULTS

A. Numerical Simulations

We illustrate the performance of the proposed algorithms by constructing functions f_0 with $P = 4$ and $P = 9$ exponentials with parameters given in Table III. We set $N = 128$ (corresponding to $2N + 1 = 257$ equally spaced modes), $\rho = 0.025$ and $\mathtt{niter} = 200$. Estimation is performed based on $M = 157$ noisy data samples $f(x_m) = f_0(x_m) + \varepsilon_m$ where $f_0(x_m) = \sum_{p=1}^P a_p e^{\zeta_p x_m}$ as in (1). We use $N_{MC} = 200$ independent realizations of white Gaussian circular noise ε_m for various signal to noise ratios (SNR) defined by

$$SNR_f = 10\log_{10}\Big(\sum_{m=1}^M |f_0(x_m)|^2 \Big/ \sum_{m=1}^M |\varepsilon_m|^2\Big).$$

We use two different criteria for quantifying the estimation error. The quality of the approximation of the function f_0 by the estimate $\hat{f}_0$ constructed using the estimates $\hat{\boldsymbol{\zeta}}$ and $\hat{\boldsymbol{a}}$ is assessed by the mean squared deviation of $\hat{f}_0$ from f_0 evaluated at the equally spaced locations x^{eq}. For better readability across varying SNR values, we use the normalized root mean squared error (nRMSE) defined as the mean squared deviation normalized by the noise energy,

$$\begin{aligned}&\mathrm{nRMSE} \\ &\quad = 10\log_{10}\left(\frac{\frac{1}{2N+1}\sum_{n=-N}^N \langle|\hat{f}_0(x_n^{\mathrm{eq}}) - f_0(x_n^{\mathrm{eq}})|^2\rangle}{\mathbb{E}[|\varepsilon|^2]}\right)\end{aligned}$$

where $\langle\cdot\rangle$ stands for average over the independent realizations and $\mathbb{E}[\cdot]$ for the expectation. Furthermore, the estimation errors for the frequency and damping parameters $\boldsymbol{\theta} \in \{\boldsymbol{\nu}, \boldsymbol{\gamma}\}$ are quantified using their bias, standard deviation and root mean squared error (RMSE) defined as

$$\begin{aligned}\hat{b}(\boldsymbol{\theta}_p) &= \langle \boldsymbol{\theta}_p - \boldsymbol{\theta}_p^{(\mathrm{true})} \rangle \\ \hat{s}(\boldsymbol{\theta}_p) &= \sqrt{\frac{N_{MC}}{N_{MC}-1}\langle(\boldsymbol{\theta}_p - \langle\boldsymbol{\theta}_p\rangle)^2\rangle} \\ \hat{r}(\boldsymbol{\theta}_p) &= \sqrt{\langle(\boldsymbol{\theta}_p - \boldsymbol{\theta}_p^{(\mathrm{true})})^2\rangle}\end{aligned}$$

where $\boldsymbol{\theta}^{(\mathrm{true})}$ denotes the true value of the parameter vector $\boldsymbol{\theta}$. The standard deviations and RMSEs of the estimators are compared with the squared roots of the Cramér-Rao lower bounds (CRB) for the parameters associated with the estimation problem (available in [1], [3]). Note that the problem of associating an estimate $\hat{\zeta}_p$ with the true parameter ζ_p is not trivial. Here, we adopt the simple rule of ordering both true values and estimates w.r.t. the frequency ν_p and comparing the ordered set of estimates to the ordered set of true values.

Different sets of unequally spaced sampling nodes (x_m) result in different estimation problems and hence different CRBs. We therefore draw one single random set of sampling nodes (x_m) on the interval $(-0.5, 0.5)$ and keep it fixed throughout the numerical experiments. The functions f_0 and the samples $f_0(x_m)$ at fixed random positions used in the numerical experiments are plotted in the top rows of Figs. 3(a) and 3(b), respectively. Our algorithms are initialized with $\boldsymbol{g} = \boldsymbol{\mathcal{K}}^*(f(x_m))_{m=1}^M$

TABLE III
FREQUENCY, DAMPING AND AMPLITUDE PARAMETERS USED IN THE NUMERICAL EXPERIMENTS WITH $P = 4$ (TOP) AND $P = 9$ (BOTTOM) COMPLEX EXPONENTIALS

p	1	2	3	4	5	6	7	8	9
ν_p	-7.68	39.68	40.96	99.84					
γ_p	-0.274	-0.150	0.133	-0.221					
$\lvert a_p \rvert$	0.4	1.2	1.0	0.9					
$\angle a_p$	-0.93	-1.55	-0.83	0.07					
ν_p	-92.16	-19.20	-7.68	3.71	11.90	14.98	39.68	40.96	99.84
γ_p	0.177	-0.327	-0.274	-0.097	-0.116	0.026	-0.150	0.133	-0.221
$\lvert a_p \rvert$	1.0	0.4	1.5	0.7	0.6	1.2	1.0	0.9	0.9
$\angle a_p$	0.42	-0.93	-0.95	0.40	0.02	-1.55	-0.83	0.09	0.07

and with all elements of the matrix $\mathbf{\Lambda}$ set to zero. We assume the number of modes P to be known; the performance of the model order estimation algorithm of Section III-D will be investigated separately in Section V-E.

B. Periodogram and Lomb-Scargle Periodogram

We begin with an illustration of the performance of the Lomb-Scargle periodogram for spectral density estimation for unequally spaced data and compare it with the proposed method. It is applied here to the real part of the function composed of $P = 4$ complex exponentials for 200 realizations of white Gaussian noise and $SNR_f = 10$ dB (we set $\gamma_p = 0$—hence no damping—in order to avoid ambiguities due to time-varying amplitudes). Averaged periodograms are plotted as black solid lines in Fig. 1 for equally spaced data (top row, standard periodogram) and unequally spaced data (bottom row, Lomb-Scargle periodogram) together with the true frequency locations (indicated by blue vertical lines). The averages of the frequency estimates obtained with the proposed method are indicated by red vertical lines. Furthermore, averages of the standard periodograms of the reconstructions $\hat{f}_0$ of f_0 at equally spaced locations, obtained by using the estimates $\hat{\boldsymbol{\zeta}}$ and $\hat{\boldsymbol{a}}$, are plotted as thin solid green lines. As expected, due to its non-parametric nature, the periodogram fails to unambiguously resolve the two closely spaced frequencies both in the equally and unequally spaced case. Moreover, the Lomb-Scargle periodogram fails to identify the lowest frequency component for this example of unequally spaced data. Indeed, the power of the spectral density estimate is far below the artefact peaks appearing over the entire frequency range. Note that it is even difficult to conclude from the Lomb-Scargle periodogram that the unequally spaced data have a line spectrum. In contrast, the proposed method provides good estimates of the frequency content for both equally and unequally spaced data.

C. Unequally Spaced Data

The objective of this Section is to quantify the estimation error of the proposed method. Fig. 2 shows the RMSEs $\hat{r}(\boldsymbol{\nu})$ (top row) and $\hat{r}(\boldsymbol{\gamma})$ (bottom row) as a function of SNR for $P = 4$ and $P = 9$ exponentials (left and right column, respectively), together with the corresponding square roots of the CRBs for unbiased estimators (dashed lines). The bias $\hat{b}(\boldsymbol{\theta})$ of the estimators is found to be several orders of magnitude below the standard deviations $\hat{s}(\boldsymbol{\theta})$ and is not further discussed here. For sufficiently high SNR values ($SNR_f \geq 5$ dB for $P = 4$ and $SNR_f \geq 10$ dB for $P = 9$, respectively), the RMSEs of the estimators are close to the CRBs. This indicates that the proposed procedure yields the global minimum of the NLS problem (2) and hence provides the maximum likelihood estimator for the white Gaussian noise case, despite the problem being non convex. For lower SNR values, the performance of the proposed estimation strategy decreases, indicating that it returns a non-optimal solution for (2). However, this does not imply that the approximation $\hat{f}_0$ found by the algorithm is a bad estimate of the true function f_0. In order to illustrate this point, Fig. 3 shows the means and standard deviations of the approximations $\hat{f}_0$ for $SNR_f = 0$ dB and $SNR_f = 5$ dB together with the true functions f_0. The means and standard deviations of the unequally spaces samples $f(x_m)$ are also displayed in this figure. Despite the fact that the procedure may converge to a non-optimal solution only for these SNR values, the results indicate that these solutions correspond to functions $\hat{f}_0$ that well approximate the true function f_0. This is further illustrated in Fig. 4, in which the normalized mean squared errors are plotted as a function of SNR_f. At $SNR_f = 0$ dB, the mean squared error increases only by 2 dB ($P = 4$) and 4 dB ($P = 9$) w.r.t. the values obtained for high SNR.

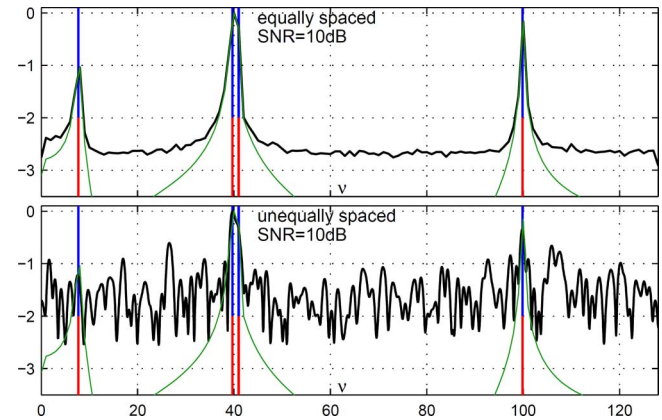

Fig. 1. Spectral density estimation for $P = 4$ undamped cosines with frequencies ν_p in white Gaussian noise for equally (top) and unequally (bottom) spaced data: true frequencies (blue solid vertical lines); standard (top) and Lomb-Scargle (bottom) periodogram (black solid lines); frequency estimates obtained with the proposed method (red solid vertical lines) and standard periodograms of the reconstructions from these frequency estimates (thin green solid lines).

D. Equally Spaced Data With Missing Samples

We consider three scenarios for equally spaced data: first, no missing samples (i.e., $\mathcal{J} = \{\emptyset\}$); second, 100 samples missing at fixed random positions; third, two blocks of 50 samples each missing (at positions 21 and 188). We compare performance with those of the ESPRIT method with missing data samples replaced by zeros. By ESPRIT, we mean the version outlined in [1] (also reproduced in Section III-C). Upon request during the review, we also include results obtained with the iterative method proposed in [2], termed dIAA, and with the popular forward backward ESPRIT, also known as ESPRIT-CS (see e.g., [1], [58]). Note that the latter utilizes a symmetry which is not present in the case of damped signals.

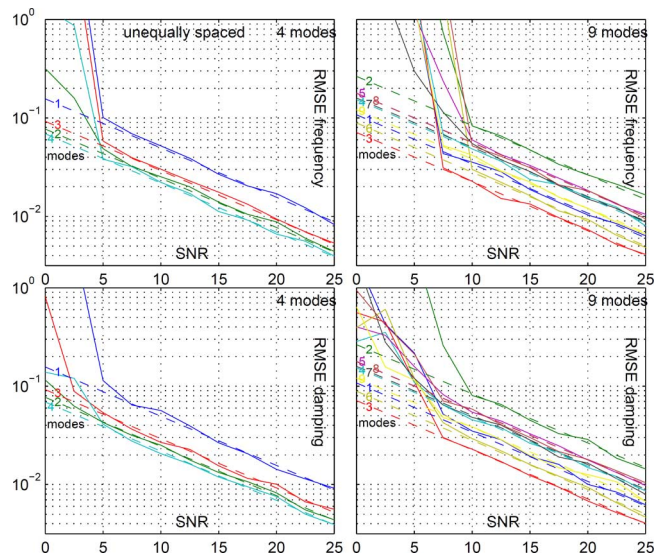

Fig. 2. RMSEs of $\hat{\boldsymbol{\nu}}$ (top row) and $\hat{\boldsymbol{\gamma}}$ (bottom row) for unequally spaced data for $P = 4$ (left column) and $P = 9$ exponentials (right column), respectively; each mode is plotted in a different color. The dashed lines indicate the CRBs for the standard deviations. Note that for $P = 9$ modes, the CRBs of modes number 5 and 8 are so close that they appear as one single line.

In Fig. 5, RMSEs of the different estimators are plotted for these three scenarios (from top to bottom row, respectively) for $P = 4$ exponentials for the proposed Hankel ADMM based approach (left column), ESPRIT and ESPRIT-CS (center column) and dIAA (right column) together with the corresponding square roots of the CRBs for unbiased estimators. The results obtained with the proposed estimation algorithm are similar to those obtained for unequally spaced data: at sufficiently high SNR, the estimation errors are close to the CRBs for all scenarios, indicating that the procedure converges to the global minimizer of the NLS problem (2). The ESPRIT method has estimation error slightly larger than the CRBs when there are no samples missing and yields large errors in the missing data case, and ESPRIT-CS does not work well for any of the three scenarios due to the damping in the signal. The performance of dIAA is largely determined by the resolution of the frequency-damping parameter grid employed by the method. We use the original parameter setup of [2] with $n_\nu = 1000$ and $n_\gamma = 101$ frequency and damping grid points, respectively, and $\kappa = 10$ iterations. The theoretical RMSEs for the grid points closest to the modes are indicated by triangles in Fig. 5 (note that for modes 2 and 3, these values essentially coincide and collapse in the plot). For sufficiently large SNR values, the method succeeds in selecting the grid points closest to modes 3 and 4 and a grid point reasonably close to mode 2 for all three missing sample scenarios. Yet, it fails to identify mode 1 for any SNR value (the RMSE value for this mode is far larger than the range plotted in Fig. 5 and hence not visible). The computational cost of dIAA is $\mathcal{O}(\kappa(K^3 + K^2 n_\nu n_\gamma))$, where K is the number of available data samples. Since $n_\nu n_\gamma \gg K$ and $K \sim N$, the dIAA has a worse time complexity than the proposed method. Hence, despite performing significantly worse, the computational cost of dIAA is about one order of magnitude larger than that of the proposed method for this

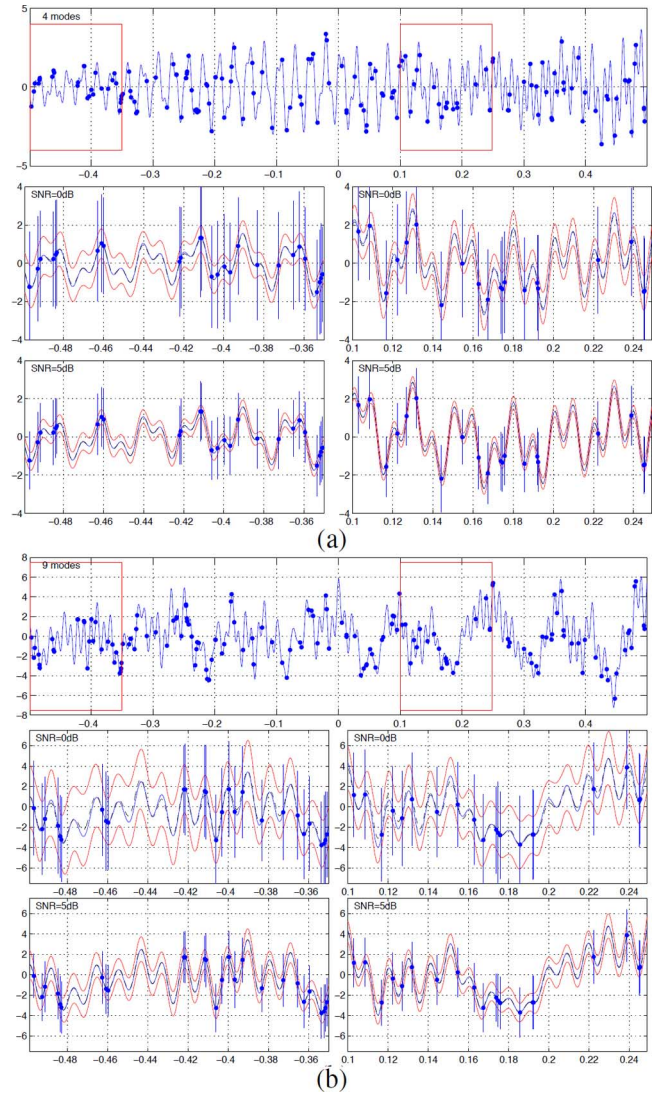

Fig. 3. Top rows: real part of true noise-free function (blue solid) and unequally spaced samples (blue discs) for $P = 4$ (subfigure (a)) and $P = 9$ mode (subfigure (b)) case. Second and third rows: Zooms on approximations in the region highlighted in the top rows for $SNR_f = 0$ dB and $SNR_f = 5$ dB (second and third rows, respectively). True function (blue solid), averaged approximation (black dashed) with 1.96 standard deviation error tube (red solid), averaged sample values (blue discs) with 1.96 standard deviation error bars (blue solid).

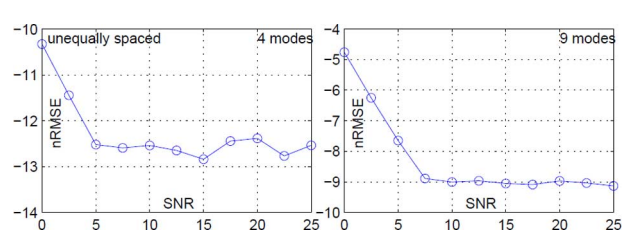

Fig. 4. nRMSE (in dB) of approximations obtained for unequally spaced data for $P = 4$ (left column) and $P = 9$ exponentials (right column).

example. Similar conclusions are drawn from Fig. 6, plotting the normalized mean squared errors of the approximations: nRMSE values for ESPRIT are consistently roughly 1 dB above those of the proposed method for the regular sampling case and are large for the missing data cases. The dIAA method has large approximation error in all of the three scenarios. We

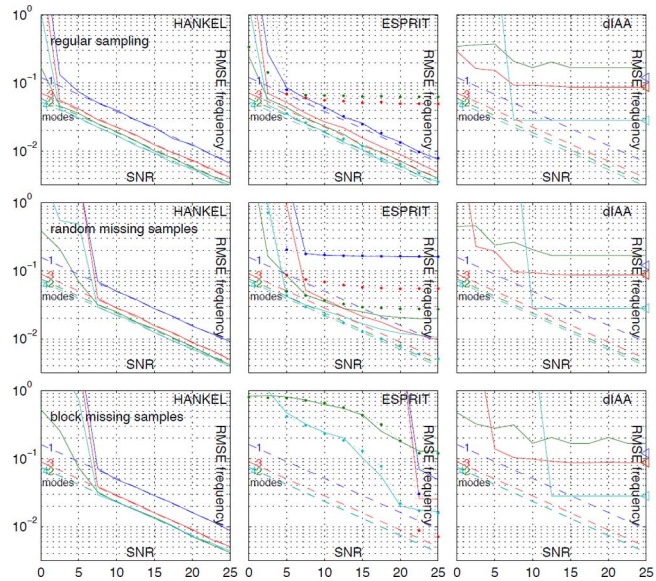

Fig. 5. RMSEs of $\hat{\boldsymbol{\nu}}$ for $P = 4$ exponentials obtained with the proposed method (left column), ESPRIT (center column: standard (solid lines), ESPRIT-CS (dotted)) and dIAA (right column), respectively; each mode is plotted in a different color. The dashed lines show the square roots of the CRBs. Top row: no missing data. Second row: 100 samples missing at random locations. Bottom row: 2 blocks of 50 missing samples.

have noted that similar results (not reported here for space reasons) are obtained for ESPRIT when applied to interpolated data $\boldsymbol{\mathcal{K}}^{-1}\boldsymbol{f}$ instead of missing data replaced by zeros, i.e., when mimicking the unequally spaced algorithm rather than using the weighted algorithm. The performance is further illustrated in Fig. 7, where averages and standard deviations of the approximations $\hat{f}_0$ are plotted together with the true function f_0. Averages and standard deviations of the available samples for $SNR_f = 5$ dB are also displayed for the case of two blocks of missing samples. The Hankel ADMM procedure yields good approximations of the true function f_0, including in the zones where no data is available. In contrast, both the ESPRIT and dIAA method yield only rough approximations with large errors. Similar results have been obtained for $P = 9$ exponentials and lead to the same conclusions.

E. Model Order Estimation

This section is testing the modified ESTER criterion on the two cases $P = 4$ and $P = 9$. The left panel of Fig. 8 shows error bar plots for the criterion ($J(p) = \|E_{\mathtt{Hadmm}}(p)\|^{-2}$) in blue ($P = 4$) and red ($P = 9$) for 100 independent noise realizations. For each realization $J(p)$ was normalized to have a maximum value of 1, to account for the relative certainty in each simulation. The right panel of Fig. 8 shows histograms of the outcomes of using the modified ESTER criterion for model order selection for the two cases $P = 4$ and $P = 9$, respectively. Clearly, the criterion yields excellent model order estimates for large SNR values. For SNR values below 7.5 dB (10 dB) for $P = 4$ ($P = 9$), the performance of the method degrades. This is consistent with the above results indicating that for low SNR values, the error of the Hankel ADMM based estimates from which the ESTER criterion is computed also increases significantly, see Fig. 2.

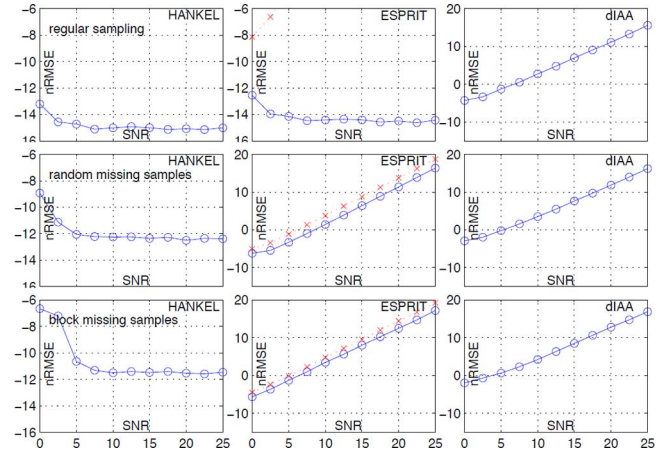

Fig. 6. nRMSE (in dB) of approximations obtained for equally spaced data with the proposed Hankel ADMM approach (left column) , ESPRIT (center column; standard (blue circles) and ESPRIT-CS (red crosses)) and dIAA (right column). Top row: no missing data. Second row: 100 samples missing at random locations. Bottom row: 2 blocks of 50 samples missing.

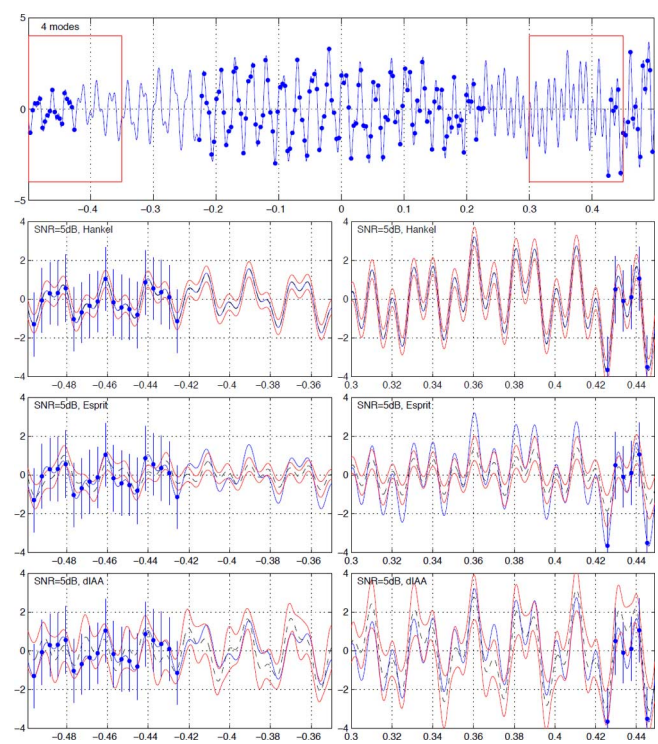

Fig. 7. Top row: real part of true noise-free function (blue solid) and equally spaced samples (blue discs) for $P = 4$ exponentials and missing data. Second to fourth row: Zooms on approximations for the regions highlighted in the top rows for $SNR_f = 5$ dB. The second row is obtained with the proposed Hankel ADMM based approach, the third row corresponds to ESPRIT, the fourth row to dIAA. True function (blue solid), averaged approximation (black dashed) with 1.96 standard deviation error tube (red solid), averaged sample values (blue discs) with 1.96 standard deviation error bars (blue solid).

F. Convex Relaxation Using Nuclear Norm

This section investigates the potential use of the nuclear norm replacing $\mathcal{R}_P$ in the Hankel ADMM based procedure. We consider the representative example of unequally spaced samples with $SNR_f = 10$ dB for $P = 4$ exponentials. Fig. 9 (top left) plots the normalized mean squared errors obtained with the proposed method as a function of the augmented parameter ρ (blue,

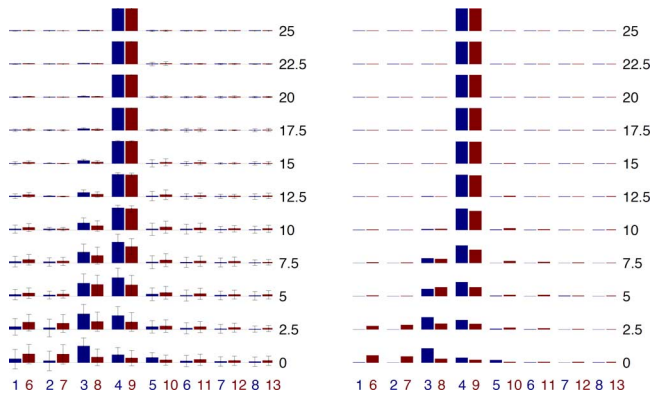

Fig. 8. Left: Error bar plot of the ESTER model criterion for the two unequally spaced cases ($P = 4$ and $P = 9$ in blue and red, respectively). Right: Histogram of model selection outcome using the criterion. The y-axis gives the SNR in dB.

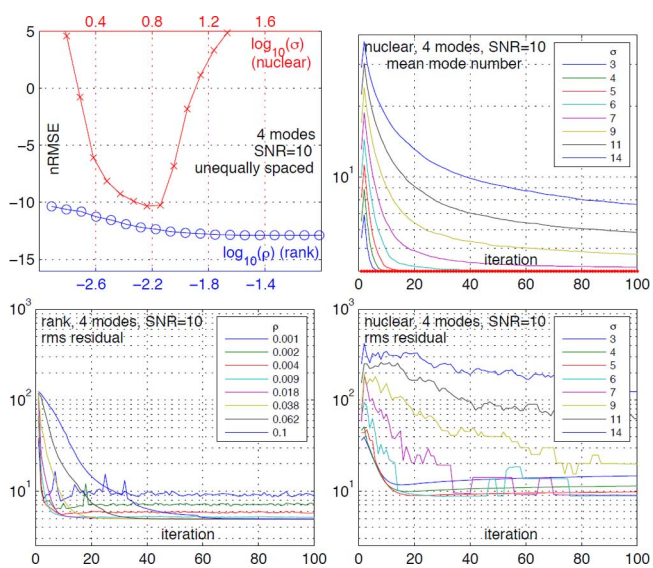

Fig. 9. Top left: nRMSE of approximations obtained with rank constraint (blue, circles) and nuclear norm (red, crosses) as a function of augmented parameter ρ and nuclear parameter σ, respectively. Top right: averaged number of complex exponentials $\hat{P}$ selected by nuclear norm as a function of iteration number for various values of the nuclear parameter σ. Bottom row: root mean squared distance between the true function and its estimate versus iteration number q for rank constraint (left, for various values ρ) and for nuclear norm (right, for various values of σ, ρ fixed).

circles) and with nuclear norm as a function of the nuclear norm penalty parameter σ (red, circles; the augmented parameter ρ is fixed).

Clearly, the proposed method yields solutions that are consistently better than those obtained with the nuclear norm for which nRMSE values are more than 2.5 dB higher. This can also be observed in Table IV, where RMSEs of parameters obtained with the rank constraint are compared to those obtained with the nuclear norm with optimal parameter σ. Moreover, Fig. 9 (top left) indicates that the proposed method is very robust with respect to the precise choice of the tuning parameter ρ. In contrast, the nuclear norm alternative is very sensitive to the choice of σ: there exists only a small zone around the "optimal" value

TABLE IV
RMSEs of Estimators $\hat{\boldsymbol{\nu}}$ and $\hat{\boldsymbol{\gamma}}$ for the Proposed Method With Non-Convex Rank Constraint and With Nuclear Norm With Optimal Parameter σ ($SNR_f = 10$ dB)

RMSE $\times 10^2$	ν_1	ν_2	ν_3	ν_4	γ_1	γ_2	γ_3	γ_4
rank	4.43	2.60	3.02	1.95	5.16	2.38	2.83	2.14
nuclear norm	4.75	4.27	5.19	3.09	16.22	6.19	4.08	6.38

of σ for which the performance is reasonable. This renders the selection of σ difficult while it is at the same time critical since it has a strong impact on the model order selected by the algorithm (cf., Fig. 9, top right).

We finally illustrate the convergence of the Hankel ADMM based approach. Root mean squared residuals between the approximations $\hat{f}$ and the true function f_0 are plotted in Fig. 9 (bottom) w.r.t. iteration number q for rank constraint (bottom left) and nuclear norm (bottom right). For reasonable choices of ρ, the proposed method converges in about $15 - 20$ iterations.

VI. The Two-Dimensional Case

The proposed method is fairly straightforward to generalize to higher dimensions. We will briefly discuss the two-dimensional (2D) case and illustrate it with an example. As in the one-dimensional case, there is a relationship between *block-Hankel* matrices and sums of exponentials. A block Hankel matrix is generated by a sampled function of two variables. Assume that

$$f(x_1, x_2) = \sum_{p=1}^{P} a_p e^{\zeta_{p,1} x_1 + \zeta_{p,2} x_2}$$

where the complex frequencies $(\zeta_{p,1}, \zeta_{p,2})$ are assumed to be distinct. Consider two sets of equally spaced nodes

$$(x_1^{\mathrm{eq}})_{n_1} = \frac{n_1 - N_1 - 1}{2N_1}, \quad (x_2^{\mathrm{eq}})_{n_2} = \frac{n_2 - N_2 - 1}{2N_2}$$

and let $\boldsymbol{f}_{n_1,n_2} = f\Big((x_1^{\mathrm{eq}})_{n_1}, (x_2^{\mathrm{eq}})_{n_2}\Big)$ for $1 \le n_1 \le 2N_1 + 1, 1 \le n_2 \le 2N_2 + 1$ define a 2D sampling of a function as in (4). A block-Hankel matrix can be generated by $\boldsymbol{f}$ as follows

$$H(\boldsymbol{f}) = \begin{pmatrix} H_1\boldsymbol{f} & H_2\boldsymbol{s}f & \cdots & H_{N_1+1}\boldsymbol{f} \\ H_2\boldsymbol{f} & \iddots & \iddots & H_{N_1+2}\boldsymbol{f} \\ \vdots & \iddots & \iddots & \vdots \\ H_{N_1+1}\boldsymbol{f} & H_{N_1+2}\boldsymbol{f} & \cdots & H_{2N_1+1}\boldsymbol{f} \end{pmatrix}$$

where

$$H_{n_1}\boldsymbol{f} = \begin{pmatrix} \boldsymbol{f}_{n_1,1} & \boldsymbol{f}_{n_1,2} & \cdots & \boldsymbol{f}_{n_1,N_2+1} \\ \boldsymbol{f}_{n_1,2} & \iddots & \iddots & \boldsymbol{f}_{n_1,N_2+2} \\ \vdots & \iddots & \iddots & \vdots \\ \boldsymbol{f}_{n_1,N_2+1} & \boldsymbol{f}_{n_1,N_2+2}\boldsymbol{f} & \cdots & \boldsymbol{f}_{n_1,2N_2+1} \end{pmatrix}.$$

For the minimization counterparts of (6) and (8) in 2D we simply replace the Hankel matrix by a block-Hankel matrix. The ADMM based methods for obtaining approximate solutions based on (9) is easy to generalize using block-Hankel matrices instead of Hankel matrices. We briefly describe the necessary changes in the main algorithm of Table I. The operation `form_hank` now corresponds to the

construction of the block-Hankel matrices above and the operation `sum_hank` computes sums over matrices of size $(2N_1+1)(2N_2+1) \times (2N_1+1)(2N_2+1)$ according to the Hankel indexing above. For the interpolation matrix $\boldsymbol{\mathcal{K}}$, the definition (7) is replaced by a tensor product of one-dimensional interpolating functions φ (see, e.g., [59], [60]). The only difficult part in the generalization of the main algorithm in Table I concerns the computation of the modes $(\zeta_{p,1}, \zeta_{p,2})_{p=1}^{P}$. If $P \leq N_1$ and $P \leq N_2$, it is possible to find the modes using each independent variable separately. For the general case, one can obtain the modes by finding the roots of a system of two polynomials in two variables, as suggested in [61]. This last approach will be used in this paper.

We conclude this section with a simple numerical illustration of the method. We are interested in recovering the function

$$f(x,y) = 1/2\cos(3x+4y) + 2/3\cos(5x-6y)$$

from 20 quasi-random samples in the box $[-1,1] \times [-1,1]$. P equals 4 in this example since two exponential functions are associated with each cosine. An example of function f and the corresponding samples are shown in Fig. 10. The top panel of the figure shows the function f as a transparent surface plot, along with the sampled data indicated by red and blue dots. The color of the dots is related to their sign (blue for positive and red for negative sign, respectively). Moreover, the location in the xy-plane is depicted with black dots, and blue and red lines that connect the projection onto the xy-plane with the sampled points on the graph of f. To further illustrate the 3d-shape, curves connecting each data point with its two closest neighbors (in the xy-plane) are depicted. The bottom left panel of Fig. 10 shows the same setup but viewed from above.

The bottom middle and right panels show the correct f and its reconstruction obtained by the proposed algorithm (viewed as images), using $\rho = 0.01$ and $\texttt{niter} = 200$. The actual and reconstructed images are indistinguishable by eye. The average point-wise error in the reconstruction is 0.0025 (and the maximum amplitude of f is 1). The estimated values of $(\zeta_{p,1}, \zeta_{p,2})$ are given in Table V. This example illustrates the fact that the proposed method can estimate sums of exponentials in two dimensions.

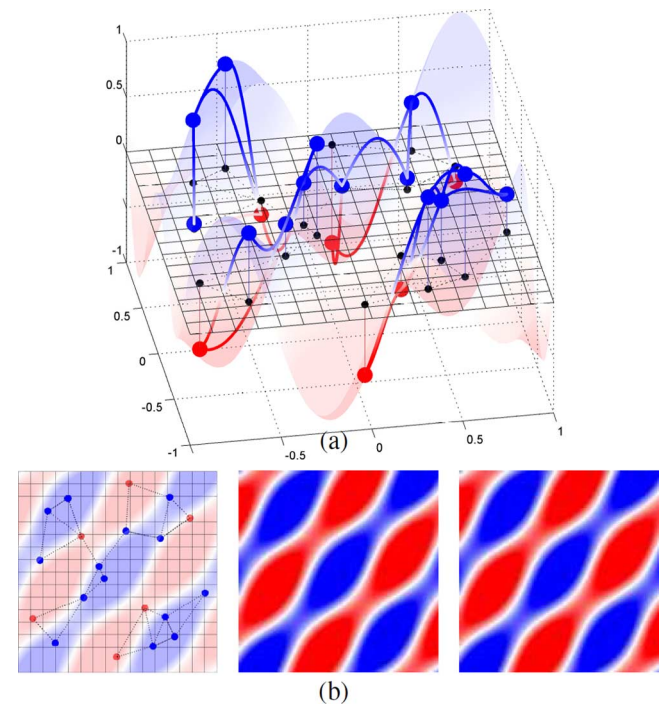

Fig. 10. Top panel: 20 quasi-random data samples (red and blue points) of a sum of two 2D cosine functions. For illustration purpose, the underlying function is shown as a surface and each data sample and its two closest neighbors are connected by lines o this surface. Bottom panels from left to right: The quasi-random sampling seen from above; the original underlying function; the reconstruction obtained by the proposed algorithm given the 20 unequally spaced samples.

TABLE V
ESTIMATED VALUES FOR THE 2D ESTIMATION PROBLEM

p	1	2	3	4
$\mathsf{Re}(\zeta_{p,1})$	+0.0001	+0.0001	-0.0052	-0.0052
$\mathsf{Im}(\zeta_{p,1})$	+4.9952	-4.9952	+3.0129	-3.0129
$\mathsf{Re}(\zeta_{p,2})$	-0.0012	-0.0012	+0.0056	+0.0056
$\mathsf{Im}(\zeta_{p,2})$	-5.9929	+5.9929	+3.9946	-3.9946

VII. CONCLUSIONS

We have developed a parametric frequency estimation procedure based on approximations with sums of P complex exponentials. Kronecker's theorem is used to cast the approximation problem in terms of generating functions for Hankel matrices of rank P. The corresponding optimization problem is solved by an ADMM type procedure (using interpolation matrices to accommodate for irregular sampling) allowing an appropriate Hankel matrix to be estimated. The parameters of the complex exponentials associated with the proposed frequency estimation model are then estimated from the solution Hankel matrix. This is in contrast with other methods, such as classical NLS, subspace methods, or greedy approaches.

The proposed Hankel ADMM based procedure can be applied to irregularly sampled data, as well as to equally spaced samples, including situations with missing data. If available, prior information on the reliability (e.g., noise variance) of data samples can be incorporated using appropriate weighting.

The optimization problem studied in this work is non convex due to the rank constraint on the Hankel matrices. However, numerical simulations indicate that the method performs well both for the unequally spaced and regular sampling cases. For sufficiently high SNR values, the procedure yields ML parameter estimates, without the need of initializing at a "good" starting point as is crucial for classical NLS in order to avoid local minima. The proposed estimation strategy also consistently outperforms a convex variant in which the rank constraint is replaced by the nuclear norm. The proposed algorithm is practically attractive due to its simplicity and ease of implementation.

The proposed method is also relevant in other contexts, including, e.g., image restoration, source detection and seismic data processing. Future work will also include applications in the context of direction of arrival estimation, extending recent work of the authors on the problem in [62]. In the multidimensional case, the proposed method can be relevant, e.g., in

synthetic aperture radar (SAR) [4], [5] where, under suitable assumptions, data acquisition gives rise to unequally spaced Fourier data (e.g., on polar grids), and sampling constraints often induce missing data situations.

Acknowledgment

We are grateful to the authors of [2] which have kindly provided us the original code for dIAA. Part of this work was conducted within the Labex CIMI during visits of F. Andersson and M. Carlsson at University of Toulouse.

References

[1] P. Stoica and R. Moses, *Spectral Analysis of Signals*. Englewood Cliffs, NJ, USA: Prentice-Hall, 2005.

[2] E. Gudmundson, P. Stoica, J. Li, A. Jakobsson, M. Rowe, J. Smith, and J. Ling, "Spectral estimation of irregularly sampled exponentially decaying signals with applications to RF spectroscopy," *J. Magn. Reson.*, vol. 203, no. 1, pp. 167–176, 2010.

[3] E. Gudmundson, P. Wirfalt, A. Jakobsson, and M. Jansson, "An ESPRIT-based parameter estimator for spectroscopic data," in *Proc. IEEE Statist. Signal Process. Workshop (SSP)*, Ann Arbor, MI, USA, Aug. 2012, pp. 77–80.

[4] M. Cheney, "A mathematical tutorial on synthetic aperture radar," *SIAM Rev.*, vol. 43, no. 2, pp. 301–312, 2001.

[5] F. Andersson, R. Moses, and F. Natterer, "Fast Fourier methods for synthetic aperture radar imaging," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 48, no. 1, pp. 215–229, 2012.

[6] R. J. A. Little and D. B. Rubin, *Statistical analysis with missing data*, 2nd ed. New York, NY, USA: Wiley, 2002.

[7] R. Kumaresan and D. Tufts, "Estimating the parameters of exponentially damped sinusoids and pole-zero modeling in noise," *IEEE Trans. Signal Process.*, vol. 30, no. 6, pp. 833–840, 1982.

[8] D. Tufts and R. Kumaresan, "Estimation of frequencies of multiple sinusoids: Making linear prediction perform like maximum likelihood," *Proc. IEEE*, vol. 70, no. 9, pp. 975–989, 1982.

[9] P. Stoica, R. Moses, B. Friedlander, and T. Söderström, "Maximum likelihood estimation of the parameters of multiple sinusoids from noisy measurements," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 3, pp. 378–392, 1989.

[10] M. Macleod, "Fast nearly ML estimation of the parameters of real or complex single tones or resolved multiple tones," *IEEE Trans. Signal Process.*, vol. 46, no. 1, pp. 141–148, 1998.

[11] Y. Li, K. J. R. Liu, and J. Razavilar, "A parameter estimation scheme for damped sinusoidal signals based on low-rank Hankel approximation," *IEEE Trans. Signal Process.*, vol. 45, no. 2, pp. 481–486, 1998.

[12] B. Quinn and E. Hannan, *The Estimation and Tracking of Frequency*, ser. Statistical And Probabilistic Mathematics. Cambridge, U.K.: Cambridge Univ. Press, 2001.

[13] P. Babu and P. Stoica, "Spectral analysis of nonuniformly sampled data—a review," *Digit. Signal Process.*, vol. 20, no. 2, pp. 359–378, 2010.

[14] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. 34, no. 3, pp. 276–280, 1986.

[15] R. Roy and T. Kailath, "ESPRIT-estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 7, pp. 984–995, 1989.

[16] R. Kumaresan and D. W. Tufts, "Estimating the angles of arrival of multiple plane waves," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 19, no. 1, pp. 134–139, 1983.

[17] H. Clergeot, S. Tressens, and A. Ouamri, "Performance of high resolution frequencies estimation methods compared to the Cramer-Rao bounds," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 11, pp. 1703–1720, 1989.

[18] P. Stoica and A. Nehorai, "Music, maximum likelihood and Cramér-Rao bound," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 5, pp. 720–741, 1989.

[19] P. Stoica and A. Nehorai, "Music, maximum likelihood and Cramér-Rao bound: further results and comparisons," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 12, pp. 2140–2150, 1990.

[20] P. Stoica and T. Söderström, "Statistical analysis of MUSIC and subspace rotation estimates of sinusoidal frequencies," *IEEE Trans. Signal Process.*, vol. 39, no. 8, pp. 1836–1847, 1991.

[21] C. Rao and L. Zhao, "Asymptotic behavior of maximum likelihood estimates of superimposed exponential signals," *IEEE Trans. Signal Process.*, vol. 41, no. 3, pp. 1461–1464, 1993.

[22] T.-H. Li and K.-S. Song, "On asymptotic normality of nonlinear least squares for sinusoidal parameter estimation," *IEEE Trans. Signal Process.*, vol. 56, no. 9, pp. 4511–4515, 2008.

[23] T.-H. Li and K.-S. Song, "Estimation of the parameters of sinusoidal signals in non-Gaussian noise," *IEEE Trans. Signal Process.*, vol. 57, no. 1, pp. 62–72, 2009.

[24] C. Rao, L. Zhao, and B. Zhou, "Maximum likelihood estimation of 2-d superimposed exponential signals," *IEEE Trans. Signal Process.*, vol. 42, no. 7, pp. 1795–1802, 1994.

[25] Y. Bresler and A. Macovski, "Exact maximum likelihood parameter estimation of superimposed exponential signals in noise," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 34, no. 5, pp. 1081–1089, 1986.

[26] G.-O. Glentis and A. Jakobsson, "Efficient implementation of iterative adaptive approach spectral estimation techniques," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4154–4167, 2011.

[27] G.-O. Glentis and A. Jakobsson, "Superfast approximative implementation of the IAA spectral estimate," *IEEE Trans. Signal Process.*, vol. 60, no. 1, pp. 472–478, 2012.

[28] S. Bourguignon, H. Carfantan, and J. Idier, "A sparsity-based method for the estimation of spectral lines from irregularly sampled data," *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 4, pp. 575–585, 2007.

[29] P. Stoica, P. Babu, and L. Jian, "New method of sparse parameter estimation in separable models and its use for spectral analysis of irregularly sampled data," *IEEE Trans. Signal Process.*, vol. 59, no. 1, pp. 35–47, 2011.

[30] M. Duarte and R. Baraniuk, "Spectral compressive sensing," *Appl. Comput. Harmon. Anal.*, vol. 35, no. 1, pp. 111–129, 2013.

[31] P. R. Stoica and P. Babu, "Sparse estimation of spectral lines: Grid selection problems and their solutions," *IEEE Trans. Signal Process.*, vol. 60, no. 2, pp. 962–967, 2012.

[32] N. Lomb, "Least-squares frequency analysis of unequally spaced data," *Astrophys. Space Sci.*, vol. 39, no. 2, pp. 447–462, 1976.

[33] J. Scargle, "Studies in astronomical time series analysis. II-statistical aspects of spectral analysis of unevenly spaced data," *Astrophys. J.*, vol. 263, pp. 835–853, 1982.

[34] W. Press and G. Rybicki, "Fast algorithm for spectral analysis of unevenly sampled data," *Astrophys. J.*, vol. 338, no. 1, pp. 277–280, 1989.

[35] M. Schulz and K. Stattegger, "Spectrum: Spectral analysis of unevenly spaced paleoclimatic time series," *Comput. Geosci.*, vol. 23, no. 9, pp. 929–945, 1997.

[36] V. Adamjan, D. Arov, and M. Krein, "Infinite Hankel matrices and generalized problems of Carathéodory-Fejér and F. Riesz," *Funkcional. Anal. i Prilozen.*, vol. 2, no. 1, pp. 1–19, 1968.

[37] F. Andersson, M. Carlsson, and M. d. Hoop, "Sparse approximation of functions using sums of exponentials and AAK theory," *J. Approx. Theory*, vol. 163, no. 2, pp. 213–248, Feb. 2011.

[38] R. Rochberg, "Toeplitz and Hankel operators on the Paley-Wiener space," *Integral Eq. Operator Theory*, vol. 10, no. 2, pp. 187–235, 1987.

[39] F. Andersson and M. Carlsson, "Alternating projections on non-tangential manifolds," *Constr. Approx.*, vol. 38, no. 3, pp. 489–525, 2013.

[40] G. Beylkin and L. Monzon, "On approximation of functions by exponential sums," *Appl. Comput. Harmon. Anal.*, vol. 19, no. 1, pp. 17–48, Jul. 2005.

[41] F. Andersson, M. Carlsson, J.-Y. Tourneret, and H. Wendt, "Frequency estimation based on Hankel matrices and the alternating direction method of multipliers," presented at the Eur. Signal Process. Conf. (EUSIPCO), Marrakech, Morocco, Sep. 2013.

[42] I. Markovksy, *Low Rank Approximation: Algorithms, Implementation, Applications*. New York, NY, USA: Springer, 2012.

[43] L. Condat and A. Hirabayashi, "Robust spike train recovery from noisy data by structured low rank approximation," presented at the Int. Conf. Sampl. Theory Appl. (SAMPTA), Bremen, Germany, Jul. 2013.

[44] L. Condat, J. Boulanger, N. Pustelnik, S. Sahnoun, and L. Sengmanivong, "A 2-d spectral analysis method to estimate the modulation parameters in structured illumination microscopy," presented at the IEEE Int. Symp. Biomed. Imag. (ISBI), Beijing, China, Apr. 2014.

[45] L. Condat and A. Hirabayashi, "Cadzow denoising upgraded: A new projection method for the recovery of Dirac pulses from noisy linear measurements," *Sampl. Theory Signal Image Process.*, 2014 [Online]. Available: http://hal.archives-ouvertes.fr/docs/00/84/79/84/PDF/double.pdf, to be published

[46] E. Candès and C. Frenandez-Granda, "Towards a mathematical theory of super-resolution," *Comm. Pure Appl. Math*, 2013 [Online]. Available: http://dx.doi.org/10.1002/cpa.21455

[47] *Super-Resolution Imaging*, ser. Digital Imaging and Computer Vision, P. Milanfar, Ed. Princeton, NJ, USA: Princeton Univ. Press, 2010.

[48] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2011.

[49] M. Fazel, T. K. Pong, D. Sun, and P. Tseng, "Hankel matrix rank minimization with applications to system identification and realization," *SIAM J. Matrix Anal. Appl.*, vol. 34, no. 3, pp. 946–977, 2013.

[50] C. Jian-Feng, E. Candès, and S. Zuowei, "A singular value thresholding algorithm for matrix completion," *SIAM J. Optim.*, vol. 20, no. 4, pp. 1956–1982, 2010.

[51] D. Slepian and H. O. Pollak, "Prolate spheroidal wave functions, Fourier analysis and uncertainty," *Bell Syst. Tech. J.*, vol. 40, no. 1, pp. 43–63, 1961.

[52] G. Strang and G. Fix, "A Fourier analysis of the finite element variational method," in *Constructive Aspects of Functional Analysis*. New York, NY, USA: Springer, 2011, pp. 793–840.

[53] R. Keys, "Cubic convolution interpolation for digital image processing," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 29, no. 6, pp. 1153–1160, 1981.

[54] I. German, "Short kernel fifth-order interpolation," *IEEE Trans. Signal Process.*, vol. 45, no. 5, pp. 1355–1359, 1997.

[55] R. Charpentier, "Nonconvex splitting for regularized low-rank + sparse decomposition," *IEEE Trans. Signal Process.*, vol. 60, no. 11, pp. 35–47, 2012.

[56] R. Badeau, B. David, and G. Richard, "Selecting the modeling order for the esprit high resolution method: An alternative approach," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2004, vol. 2, pp. 1025–1028.

[57] R. Horn and C. Johnson, *Topics in matrix analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1994.

[58] M. Haardt and J. A. Nossek, "Unitary esprit: How to obtain increased estimation accuracy with a reduced computational burden," *IEEE Trans. Signal Process.*, vol. 43, no. 5, pp. 1232–1242, 1995.

[59] L. D. Lathauwer, B. D. Moor, and J. Vandewalle, "On the best rank-1 and rank-(r 1, r 2,..., rn) approximation of higher-order tensors," *SIAM J. Matrix Anal. Appl.*, vol. 21, no. 4, pp. 1324–1342, 2000.

[60] J.-M. Papy, L. D. Lathauwer, and S. V. Huffel, "Exponential data fitting using multilinear algebra: the decimative case," *J. Chemometrics*, vol. 23, pp. 341–351, 2009.

[61] F. Andersson, M. Carlsson, and M. V. D. Hoop, "Nonlinear approximation of functions in two dimensions by sums of exponential functions," *Appl. Comput. Harmon. Anal.*, vol. 29, no. 2, pp. 156–181, 2010.

[62] F. Andersson, M. Carlsson, J.-Y. Tourneret, and H. Wendt, "On an iterative method for direction of arrival estimation using multiple frequencies," presented at the IEEE Int. Workshop Comput. Adv. Multi-Sensor Adapt. Process. (CAMSAP), Saint-Martin, France, Dec. 2013.

Fredrik Andersson received his M.Sc. in engineering physics (2000) and Ph.D. in mathematics (2005) from Lund University, Sweden. He is currently working as associate professor at the Centre for Mathematical Sciences at Lund University. His research interests include fast algorithms, inverse problems and approximation theory. During 2013, he was participating as a scientific expert in the CIMI thematic trimester on Image Processing at the University of Toulouse.

Marcus Carlsson received his M.Sc. degree in Mathematics in 2002, and his Ph.D. degree in Mathematics in 2007, both from Lund University, Lund, Sweden. From 2007 to 2010, he was a Research Associate Professor with the Department of Mathematics and with the Geo-Mathematical Imaging Group at Purdue University, Indiana, USA. In 2011, he worked as Research Professor at the Department of Mathematics at Universidad de Santiago de Chile. Since 2012, he has held a tenure track research position at the Centre for Mathematical Sciences at Lund University.

Jean-Yves Tourneret (SM'08) received the Ingénieur degree in electrical engineering from the Ecole Nationale Supérieure d'Electronique, d'Electrotechnique, d'Informatique, d'Hydraulique et des Télécommunications (ENSEEIHT) de Toulouse in 1989 and the Ph.D. degree from the National Polytechnic Institute from Toulouse in 1992. He is currently a professor in the University of Toulouse (ENSEEIHT) and a member of the IRIT laboratory (UMR 5505 of the CNRS). His research activities are centered around statistical signal and image processing with a particular interest to Bayesian and Markov chain Monte Carlo (MCMC) methods. He has been involved in the organization of several conferences including the European conference on signal processing EUSIPCO'02 (program chair), the international conference ICASSP'06 (plenaries), the statistical signal processing workshop SSP'12 (international liaisons), the International Workshop on Computational Advances in Multi-Sensor Adaptive Processing CAMSAP 2013 (local arrangement), and the statistical signal processing workshop SSP'2014 (special sessions). He has been the general chair of the CIMI workshop on optimization and statistics in image processing hold in Toulouse in 2013. He has been a member of different technical committees including the Signal Processing Theory and Methods (SPTM) committee of the IEEE Signal Processing Society (2001–2007, 2010–present). He has been serving as an associate editor for the IEEE Transactions on Signal Processing (2008–2011) and for the *EURASIP Journal on Signal Processing* (since July 2013).

Herwig Wendt (M'13) received the M.S. degree in electrical engineering and telecommunications from Vienna University of Technology, Austria, in 2005. In September 2008, he received the Ph.D. degree in Physics and Signal processing from Ecole Normale Supérieure de Lyon. From October 2008 to December 2011, he was a Postdoctoral Research Associate with the Department of Mathematics and with the Geomathematical Imaging Group, Purdue University, West Lafayette, Indiana. Since 2012, he is a permanent researcher of the Centre National de Recherche Scientifique (CNRS) within the Signal and Communicatons Group of the Institut de Recherche en Informatique de Toulouse (IRIT) of the University of Toulouse.