Frequency estimation from arbitrary time samples

Kaushik Mahata and Md Mashud Hyder

***Abstract*—We consider the problem of estimating the line spectrum of a signal from finitely many time domain samples. We present a gridless algorithm for solving the total variation minimization approach associated with this problem. Unlike the related previous results, our method does not require the sampling instants to lie on an uniform grid. The resulting algorithm is a semidefinite program, structurally similar to some of the existing methods. One key observation made in our analysis also allows us to develop a gridless version of the SPICE algorithm. The simulation results demonstrate the superiority of these in performance compared to other related methods.**

***Index Terms*—Prolate Spheroidal Wave Functions, atomic norm, total variation, line spectrum, sparse recovery.**

I. Introduction

A. Problem statement and the main contribution

Let $\mathbb{R}$ and $\mathbb{C}$ be the set of all real and complex numbers, respectively. We view a vector $\boldsymbol{v}$ in $\mathbb{C}^n$ (or $\mathbb{R}^n$) as a map from $\{1,2,\ldots,n\}$ to $\mathbb{C}$ (or $\mathbb{R}$). The k-th component of $\boldsymbol{v}$ maps k to $\mathbb{C}$, and is denoted by $\boldsymbol{v}(k)$. We consider the following problem:

Given $F>0$, a data vector $\boldsymbol{y}\in\mathbb{C}^M$, and a set of M sampling instants $\{\bar{t}_j\}_{j=1}^M$, $\bar{t}_j<\bar{t}_{j+1}$, find out K numbers $\{\bar{f}_k\}_{k=1}^K$ in $[-F/2,F/2]$ and $\{a_k\}_{k=1}^K$ in $\mathbb{C}$ such that

$$\boldsymbol{y}(j)=\sum_{k=1}^{K} a_k \mathrm{e}^{\mathrm{i}2\pi\bar{f}_k\bar{t}_j},\ j=1,2,\ldots,M. \tag{1}$$

Without any loss of generality we assume $\bar{t}_1=0$. In Appendix A we show that there always exists a solution to (1) with $K=M$. In practice, we often prefer K as small as possible. However, a solution with $K<M$ may not exist. For example, if $M=2$ and $\boldsymbol{y}\in\mathbb{C}^2$ is such that $|\boldsymbol{y}(1)|\neq|\boldsymbol{y}(2)|$, then there is no solution to (1) for $K=1$. By scaling

$$f_k=\bar{f}_k/F,\qquad t_j=F\bar{t}_j,$$

so that $f_k\in[-1/2,1/2]$, we can rewrite (1) as

$$\boldsymbol{y}(j)=\sum_{k=1}^{K} a_k \mathrm{e}^{\mathrm{i}2\pi f_k t_j},\ j=1,2,\ldots,M. \tag{2}$$

Note that t_j and f_k are dimensionless quantities. In the following we focus on finding K as small as possible, and $\{a_k,f_k\}_{k=1}^K$ satisfying (2). From the solution one may readily find $\{\bar{f}_k\}_{k=1}^K$.

We view this as a sparse function recovery problem and present a gridless algorithm for total variation minimization approach (TVMA) [1], [2]. TVMA views (2) as

$$\boldsymbol{y}(j)=\int_{-1/2}^{1/2} \mathrm{e}^{\mathrm{i}2\pi f t_j}\ \mathrm{d}s(f),\ j=1,2,\ldots,M, \tag{3}$$

This work was supported by the Australian Research Council.

The authors are with the University of Newcastle, Callaghan, NSW 2308, Australia (email: Kaushik.Mahata@newcastle.edu.au, md.hyder@uon.edu.au).

where

$$s(f)=\sum_{k=1}^{K} a_k u(f-f_k), \tag{4}$$

with u denoting the unit step function, *i.e.*, $u(f)=1$ if $f\geq 0$, and $u(f)=0$ if $f<0$. Given $\boldsymbol{y}$ there are uncountably infinitely many s consistent with (3). TVMA picks the one with the smallest total variation. Our main result is to show that TVMA is equivalent to solving a semidefinite program given in Theorem 1, which uses the Prolate Spheroidal Wave Functions (PSWFs) [3], [4]. Let

$$c=\pi t_M,$$

and $\mathbb{L}^2$ be the set of all square integrable functions on $[-1/2,1/2]$. PSWFs used here are the eigenfunctions of the linear map $\mathcal{E}:\mathbb{L}^2\to\mathbb{L}^2$ such that for any $r\in\mathbb{L}^2$

$$(\mathcal{E}r)(\tau)=\int_{-1}^{1}\mathrm{e}^{\mathrm{i}c\xi\tau}\ r(\xi)\ \mathrm{d}\xi,\qquad \forall\tau\in[-1,1]. \tag{5}$$

Thus for any j, the j th PSWF φ_j satisfies

$$\mathcal{E}\varphi_j=\lambda_j\varphi_j. \tag{6}$$

Here λ_j is the j-th eigenvalue of $\mathcal{E}$ in order: $|\lambda_j|>|\lambda_{j+1}|,\forall j$ where $|\lambda_j|$ denotes the absolute value of λ_j. In particular, $|\lambda_j|$ falls off to zero rapidly with increasing j beyond $2c/\pi$ [5]. Let ϵ be the working precision of the underlying computational platform. In the sequel d is such that $|\lambda_j|<\epsilon$ for all $j>2d$. Also A^* denotes the conjugate transpose of A, while $A^\intercal$ is the transpose of A. For a matrix A, $A\succeq 0$ means that A is positive semidefinite. $\mathbf{I}_M$ denotes an $M\times M$ identity matrix and $\mathbf{0}_{d\times 1}$ be a $d\times 1$ zero vector. Our main result is:

Theorem 1. *Let $\boldsymbol{e}$ be the first column of $\mathbf{I}_M$, $\mathbf{J}_1=[\ \mathbf{I}_d\quad \mathbf{0}_{d\times 1}\ ]$, $\mathbf{J}_2=[\ \mathbf{0}_{d\times 1}\quad \mathbf{I}_d\ ]$, $\theta_0=c/d$. Let $\boldsymbol{\Phi}\in\mathbb{R}^{(2d+1)\times(2d+1)}$ and $\boldsymbol{h}_{kl}\in\mathbb{R}^{2d+1}$ be defined as*

$$\boldsymbol{\Phi}_{kj}=\varphi_{j-1}((k-d-1)/d),\quad \boldsymbol{h}_{kl}(j)=\varphi_{j-1}((t_k-t_l)/t_M).$$

Then TVMA is equivalent of solving

$$\begin{aligned}
&\underset{w,\nu_0,\ \nu_1,\ldots,\nu_d}{\text{minimize}}\quad (w+\boldsymbol{e}^*\mathbf{Q}\boldsymbol{e})/2\\
&\text{subject to}\ \begin{bmatrix} w & \boldsymbol{y}^*\\ \boldsymbol{y} & \mathbf{Q}\end{bmatrix}\succeq 0,\\
&\qquad \mathbf{Q}_{jl}=\boldsymbol{h}_{jl}^\intercal\boldsymbol{\Phi}^{-1}[\ \nu_d^*\ \cdots\ \nu_1^*\ \nu_0\ \nu_1\ \cdots\ \nu_d\ ]^\intercal,\\
&\qquad \mathbf{T}:=\begin{bmatrix}\nu_0 & \nu_1^* & \cdots & \nu_d^*\\ \nu_1 & \nu_0 & \ddots & \vdots\\ \vdots & \ddots & \ddots & \nu_1^*\\ \nu_d & \cdots & \nu_1 & \nu_0\end{bmatrix}\succeq 0,\\
&\qquad \mathcal{W}(\mathbf{T}):=\tan^2(\theta_0/2)(\mathbf{J}_1+\mathbf{J}_2)\ \mathbf{T}\ (\mathbf{J}_1+\mathbf{J}_2)^*\\
&\qquad\qquad -(\mathbf{J}_1-\mathbf{J}_2)\ \mathbf{T}\ (\mathbf{J}_1-\mathbf{J}_2)^*\succeq 0,\\
&\qquad w,\nu_0\in\mathbb{R},\ \nu_1,\ldots,\nu_d\in\mathbb{C}.
\end{aligned}$$

For evaluating PSWFs we can use the algorithm outlined in [6, Section 4], which is similar to the classical Bouwkamp method [7]. The optimization problem outlined above is solved in the variables $\{\nu_k\}_{k=0}^{d}$ and w. Among these, w and ν_0 are real-valued, and the rest are complex-valued. $\mathbf{Q}$ is linear in $\{\nu_k\}_{k=0}^{d}$. In Appendix D we show that $\mathbf{Q}$ is Hermitian. Hence the above is a semidefinite programming (SDP) problem with two linear matrix inequality constraints. Note that $\boldsymbol{e}^*\mathbf{Q}\boldsymbol{e} = \mathbf{Q}_{11}$.

The algorithm presented in Theorem 1 is an approximate algorithm, where the approximation error can be reduced below some target level by choosing d properly. The rapid rate of decay of $|\lambda_j|$ with j beyond $2c/\pi$ allows us to maintain the approximation error below the precision of the computational platform (or the noise floor when the data are noisy) without having to increase d much above c/π. The typical complexity of this SDP is $O(d^3)$, and the worst case complexity is $O(d^6)$.

The optimal value $\mathbf{T}_*$ of $\mathbf{T}$ leads to the solution (2). The rank of $\mathbf{T}_*$ is K. $\mathbf{T}_*$ admits a Vandermonde decomposition

$$\mathbf{T}_* = \sum_{k=1}^{K} |a_k| \; \boldsymbol{\omega}(2\pi f_k t_M/d) \; \boldsymbol{\omega}^*(2\pi f_k t_M/d), \qquad (7)$$

where $\boldsymbol{\omega}(\theta) := [\ 1 \ \ \mathrm{e}^{-\mathrm{i}\theta} \ \ \cdots \ \ \mathrm{e}^{-\mathrm{i}d\theta}\ ]^*$. To find $\{a_k, f_k\}_{k=1}^{K}$, we write the individual elements of (7):

$$\nu_{l*} = \sum_{k=1}^{K} |a_k| z_k^l, \quad z_k = \mathrm{e}^{\mathrm{i}2\pi f_k t_M/d}, \qquad (8)$$

where ν_{l*} denotes the optimal value of ν_l. When $\mathbf{T}_*$ is singular, (8) admits unique solutions for $\{a_k, f_k\}_{k=1}^{K}$, and these can be computed from $\{\nu_{l*}\}_{l=0}^{d}$ via Proney's method [8], [9, Appendix A]. Almost always $\mathbf{T}_*$ is singular to the working precision. In some rare cases the TVMA solution is non-unique. Then $\mathbf{T}_*$ is non-singular. Previous authors on TVMA have discussed some procedures of computing a desirable solution from a non-singular $\mathbf{T}_*$. For instance, if we adopt the strategy in [9] then we need to apply Proney's method on $\mathbf{T}_* - \delta\mathbf{I}_{d+1}$, where δ is the smallest eigenvalue of $\mathbf{T}_*$.

B. Contributions relative to the literature

Line spectral analysis problem with regularly time-spaced samples is extensively researched [8]. Common classical methods like periodogram, MUSIC, etc developed for uniformly sampled data need modifications for irregular case [10]. However, as we show later, these modified algorithms exhibit significantly worse performance than more recent algorithms like the iterative adaptive approach (IAA) [11], [12] and the sparse iterative covariance-based estimation (SPICE) [13], [14]. IAA is a nonparametric algorithm, while SPICE is a semiparametric method. Among these SPICE being a convex optimization approach, is more popular as the global convergence is not guaranteed for IAA. SPICE can also estimate the noise variance. Apart from SPICE and IAA, there are other sparse recovery methods [15]–[17]. These methods discretize the continuous frequency domain into finite set of grid points, and apply some discrete domain sparse recovery methods. The discretize approach has several drawbacks. To achieve high resolution one requires a very fine grid consisting of a large number of points, resulting large computational complexity. Also, it is difficult to predict how the reconstruction algorithm will behave when true frequencies do not lie on the grid (which is true almost always). Several grid refinement methods have been proposed to deal with these problems [15], [18]–[21].

Recently several gridless continuous-domain methods have appeared [1], [2], [9], [22]–[26]. The pioneering contribution [1] shows how the spike train estimation problem from linear measurements can be solved exactly by solving a semidefinite program. In addition, the authors of [1] derive the conditions under which this super resolution recovery is possible. This has motivated several variants [2], [9], [22], [23], [25], [26]. These methods can allow missing data [9], [26], provide confidence bounds on the parameter estimates in presence of noise [9], [22], and give a concrete way of choosing the regularization hyper-parameter in the atomic norm regularization problem [22]. Authors of [9] have also shown how a gridless SPICE [13] can be formulated. Further connection with maximum likelihood and SPICE is explored in [27].

All above gridless methods need the sampling instants to lie on a uniform grid. In other words, t_j in (2) must be integer valued. This may not hold in many practical applications due to experimental conditions. For instance, in astrophysics specific observational conditions make the sampling instants rather arbitrary [28], [29]. Seismic data processing is another example [30]. Due to different physical conditions involved in seismic surveys, sensors locations are often placed off the regular grid. Therefore, t_j in (2) can be any real number. To the best of our knowledge Theorem 1 is the first gridless spectral estimation method for such cases. The complexity of the resulting algorithm in Theorem 1 is of the same order as those derived for regular sampling grid. We noted that d is slightly above $c/\pi = t_M$, which is the time-bandwidth product for the problem. Hence the complexity of the proposed algorithm is of the order of the cube of the time-bandwidth product, which is similar to the previous methods.

Like basis pursuit denoising (BPDN) [31] it is straightforward to extend Theorem 1 to handle noisy data, and the related theory readily extends to gridless implementation of SPICE for arbitrary sampling situation. In this way the BPDN variation of our algorithm can use the output of gridless SPICE as initializer. The benefit of this hybrid strategy helps solving the frequency splitting problem[1] of SPICE.

II. Theory

A. Minimization of total variation

Given $f \in [-1/2, 1/2]$, let $\boldsymbol{\gamma}(f)$ be such that

$$\boldsymbol{\gamma}(f) = [\ \mathrm{e}^{\mathrm{i}2\pi f t_1} \ \ \mathrm{e}^{\mathrm{i}2\pi f t_2} \ \ \cdots \ \ \mathrm{e}^{\mathrm{i}2\pi f t_M}\ ]^{\mathsf{T}}.$$

Let $\alpha_k = |a_k|$, $\phi_k = a_k/\alpha_k$, so that $|\phi_k| = 1$. Write (2) as

$$\boldsymbol{y} = \sum_{k=1}^{K} \boldsymbol{\gamma}(f_k)\phi_k\alpha_k, \qquad (9)$$

[1]Many grid-based sparse recovery methods often produce multiple frequency components around the actual frequency locations. This phenomenon has been termed as the frequency splitting problem [9].

meaning that our problem is to express $\boldsymbol{y}$ as a positive linear combination of K distinct atoms in the dictionary

$$\mathcal{A} = \{\boldsymbol{\gamma}(f)\phi : \ f \in [-1/2, 1/2], \ \phi \in \mathbb{C}, \ |\phi| = 1\}.$$

Finding the atomic decomposition (9) with smallest K is very hard. TVMA has been advocated as the most suitable convex relaxation [32], where one evaluates the atomic one-norm of $\boldsymbol{y}$ with respect to $\mathcal{A}$ [1], [2]:

$$\begin{aligned} ||\boldsymbol{y}||_{\mathcal{A}} = \inf & \sum_k \alpha_k, \\ \text{st} \ \ & \boldsymbol{y} = \sum_k \boldsymbol{\gamma}(f_k)\phi_k\alpha_k, \\ & f_k \in [-1/2, 1/2], \quad \alpha_k > 0, \quad |\phi_k| = 1. \end{aligned} \tag{10}$$

The following result is the first step towards a tractable finite dimensional characterization of (10). It is a bit more general than some analogous results presented before. The generalization is in the sense that Lemma 1 holds for any general function $\boldsymbol{\gamma}$ whose first component is unimodular.

Lemma 1. *Let $\mathbb{H}$ be the set of all $M \times M$ Hermitian matrices, and $\boldsymbol{\Gamma} : [-1/2, 1/2] \to \mathbb{H}$ be such that $\boldsymbol{\Gamma}(f) = \boldsymbol{\gamma}(f)\boldsymbol{\gamma}^*(f)$. Let $\mathbb{K}$ be closed conic hull of the set $\{\boldsymbol{\Gamma}(f) \in \mathbb{H} : f \in [-1/2, 1/2]\}$. Then $||\boldsymbol{y}||_{\mathcal{A}}$ is the optimum value of*

$$\begin{aligned} \underset{w\in\mathbb{R}, \ \mathbf{Q}\in\mathbb{K}}{\text{minimize}} \ & (w + \boldsymbol{e}^*\mathbf{Q}\boldsymbol{e})/2 \\ \text{subject to} \ & \begin{bmatrix} w & \boldsymbol{y}^* \\ \boldsymbol{y} & \mathbf{Q} \end{bmatrix} \succeq 0. \end{aligned} \tag{11}$$

In addition if an atomic decomposition $\boldsymbol{y} = \sum_k \boldsymbol{\gamma}(\mathring{f}_k)\mathring{\phi}_k\mathring{\alpha}_k$ is such that $||\boldsymbol{y}||_{\mathcal{A}} = \sum_k \mathring{\alpha}_k$, then $w = \sum_k \mathring{\alpha}_k$, $\mathbf{Q} = \sum_k \mathring{\alpha}_k\boldsymbol{\Gamma}(\mathring{f}_k)$ is a solution to (11).

Conversely, if $\mathbf{Q} = \sum_k \mathring{\alpha}_k\boldsymbol{\Gamma}(\mathring{f}_k)$ is a solution to (11) for some $\mathring{\alpha}_k > 0$, then the corresponding optimum value of w is $\sum_k \mathring{\alpha}_k = ||\boldsymbol{y}||_{\mathcal{A}}$, and there are unimodular complex numbers $\mathring{\phi}_k$ such that $\boldsymbol{y} = \sum_k \boldsymbol{\gamma}(\mathring{f}_k)\mathring{\phi}_k\mathring{\alpha}_k$.

The proof appears in Appendix B. Lemma 1 allows us to compute a solution to (10) by solving a more convenient approach (11). However, a tractable algorithm for solving (11) needs a finite parameterization $\mathbb{K}$. This is addressed next. We start with a generic characterization of $\mathbb{K}$ [33]: $\mathbf{Q} \in \mathbb{K}$ if and only if there is positive measure μ on $[-1/2, 1/2]$ such that

$$\mathbf{Q} = \int_{-1/2}^{1/2} \boldsymbol{\Gamma}(f) \ \mathrm{d}\mu(f). \tag{12}$$

This characterization makes (11) an optimization problem in μ. But it is not numerically tractable to 'parameterize' (11) by μ. Nevertheless, let us define $\hat{\mu}(t)$ such that

$$\hat{\mu}(t) = \int_{-1/2}^{1/2} \mathrm{e}^{\mathrm{i}2\pi ft} \ \mathrm{d}\mu(f).$$

Since μ is supported on $[-1/2, 1/2]$, the corresponding time domain signal $\hat{\mu}(t)$ is $[-1/2, 1/2]$ bandlimited. Now from (12) and the definition of $\boldsymbol{\Gamma}$ note that

$$[\mathbf{Q}]_{jl} = \int_{-1/2}^{1/2} \mathrm{e}^{\mathrm{i}2\pi f(t_j - t_l)} \ \mathrm{d}\mu(f) = \hat{\mu}(t_j - t_l). \tag{13}$$

In other words, the elements of $\mathbf{Q}$ are the samples of a $[-1/2, 1/2]$ bandlimited signal, and these observations are all confined within a time window $[-t_M, t_M]$. This is a classical result that the space of $[-1/2, 1/2]$ bandlimited signals captured within a time window $[-t_M, t_M]$ is of a finite dimension, and therefore can be parameterized via a finite number of parameters. The results in [3], [5] establish that the most powerful basis of this space are the PSWFs. The following lemma uses these observations to accomplish our goal, where PSWFs play a central role.

Lemma 2. *$\mathbf{Q} \in \mathbb{K}$ if and only if there exists $\nu_0 \in \mathbb{R}$ and $\nu_k \in \mathbb{C}, \ k = 1, \ldots, d$ such that $\mathbf{T} \succeq 0$, $\mathcal{W}(\mathbf{T}) \succeq 0$ satisfying*

$$\mathbf{Q}_{jl} = \boldsymbol{h}_{jl}^{\mathsf{T}}\boldsymbol{\Phi}^{-1}[\ \nu_d^* \ \ \cdots \ \ \nu_1^* \ \ \nu_0 \ \ \nu_1 \ \ \cdots \ \ \nu_d \ ]^{\mathsf{T}},$$

where $\boldsymbol{h}_{jl}, \boldsymbol{\Phi}, \mathbf{T}$ and $\mathcal{W}(\mathbf{T})$ are defined in the statement of Theorem 1. In addition, if there is a singular measure

$$\mu(f) = \sum_{k=1}^{K} \alpha_k u(f - f_k), \ K \leq d, \ |\alpha_k| > 0, \tag{14}$$

satisfying (12), then it is unique, and $\mathbf{T}$ is singular with the Vandermonde decomposition

$$\mathbf{T} = \sum_{k=1}^{K} \alpha_k \ \boldsymbol{\omega}(2\pi f_k t_M/d) \ \boldsymbol{\omega}^*(2\pi f_k t_M/d).$$

The proof appears in Appendix C. Combining Lemma 1 and Lemma 2 we note that Theorem 1 gives an algorithm for solving (10). This yields the optimal solution $\mathbf{T}_*$. TVMA gets a sparse solution if $\mathbf{T}_*$ is rank deficient, and in that case its Vandermonde decomposition gives $\{\alpha_k\}_{k=1}^K$ and $\{f_k\}_{k=1}^K$, where K is the rank of $\mathbf{T}_*$. However, (12) and (14) imply that the optimal solution $\mathbf{Q}_* = \sum_{k=1}^K \alpha_k\boldsymbol{\Gamma}(f_k)$. Then by Lemma 2 we know that there are unimodular complex numbers $\{\phi_k\}_{k=1}^K$ satisfying atomic decomposition $\boldsymbol{y} = \sum_k \boldsymbol{\gamma}(f_k)\phi_k\alpha_k$. To find $\{\phi_k\}_{k=1}^K$ we solve this linear system. This also holds while working with noisy data where we estimate $\boldsymbol{y}$ from the data, see (15) below.

B. TVMA in noisy environment and gridless SPICE

In practice, we observe a noise corrupted version $\hat{\boldsymbol{y}} = \boldsymbol{y} + \boldsymbol{n}$ of $\boldsymbol{y}$, where $\boldsymbol{n}$ is additive noise with covariance matrix $\sigma^2\mathbf{I}$. We employ basis pursuit denoising (BPDN) [31] to handle noisy data which estimates $\boldsymbol{y}$ as well. This is done by adding a regularization term in the cost function:

$$\begin{aligned} \underset{w\in\mathbb{R}, \ \mathbf{Q}\in\mathbb{K}, \boldsymbol{y}}{\text{minimize}} \ & (w + \boldsymbol{e}^*\mathbf{Q}\boldsymbol{e}) + \frac{1}{\varrho}\|\hat{\boldsymbol{y}} - \boldsymbol{y}\|_2^2 \\ \text{subject to} \ & \begin{bmatrix} w & \boldsymbol{y}^* \\ \boldsymbol{y} & \mathbf{Q} \end{bmatrix} \succeq 0 \end{aligned} \tag{15}$$

where $\|\boldsymbol{x}\|_2$ denotes the 2-norm of vector $\boldsymbol{x}$ and ϱ is a noise variance dependent parameter to be chosen by the user. For best results one should set $\varrho = \sigma\sqrt{M\ln(t_M)}$ [9], [22]. However, σ is unknown in general. In our implementation we

use SPICE [13], [14] to obtain an estimate of σ, and use that in (15). SPICE proposes to

$$\begin{aligned} \underset{w\in\mathbb{R},\ \boldsymbol{\theta},\delta_1,\ldots,\delta_M}{\text{minimize}} \quad & w\hat{\boldsymbol{y}}^*\hat{\boldsymbol{y}} + \mathrm{Tr}(\mathbf{R}(\boldsymbol{\theta})) + \sum_{k=1}^{M}\delta_k \\ \text{subject to} \quad & \begin{bmatrix} w & \hat{\boldsymbol{y}}^* \\ \hat{\boldsymbol{y}} & \mathbf{R}(\boldsymbol{\theta})+\mathbf{D} \end{bmatrix} \succeq 0, \\ & \mathbf{D} = \mathrm{diag}\{\delta_1,\delta_2,\ldots,\delta_M\} \succeq 0, \end{aligned} \tag{16}$$

where $\mathbf{R}$ models the covariance matrix of $\boldsymbol{y}$ and parameterized linearly by a parameter vector $\boldsymbol{\theta}$. Here $\mathrm{diag}\{\boldsymbol{x}\}$ denotes a diagonal matrix where the vector $\boldsymbol{x}$ is in its diagonal. In its original formulation SPICE is a grid based approach where the user chooses a suitably large integer N, and sets up a frequency grid $\{\tilde{f}_k\}_{k=0}^{N-1}$ such that $\tilde{f}_k = -1/2 + k/N$, and parameterizes $\mathbf{R}$ in terms of $\boldsymbol{\theta}\in\mathbb{R}^N$ as

$$\mathbf{R}(\boldsymbol{\theta}) = \sum_{k=0}^{N-1}\boldsymbol{\theta}(k)\boldsymbol{\gamma}(\tilde{f}_k)\boldsymbol{\gamma}^*(\tilde{f}_k). \tag{17}$$

This parameterization is motivated by (9), which implies that the covariance matrix of $\boldsymbol{y}$ is $\sum_{k=1}^{K}\alpha_k^2\boldsymbol{\gamma}(f_k)\boldsymbol{\gamma}^*(f_k)$. Since $\{f_k\}_{k=1}^K$ are unknown at the beginning a grid can be used as in (17). However, like some other cases [9], we can avoid the grid. This is because $\mathbf{R}\in\mathbb{K}$ regardless of $\{f_k\}_{k=1}^K$ and $\{\alpha_k\}_{k=1}^K$. This allows us to formulate SPICE using Lemma 2:

$$\begin{aligned} \underset{w\in\mathbb{R},\ \mathbf{R}\in\mathbb{K},\ \delta_1,\delta_2,\ldots,\delta_M}{\text{minimize}} \quad & w\hat{\boldsymbol{y}}^*\hat{\boldsymbol{y}} + \mathrm{Tr}(\mathbf{R}) + \sum_{k=1}^{M}\delta_k \\ \text{subject to} \quad & \begin{bmatrix} w & \hat{\boldsymbol{y}}^* \\ \hat{\boldsymbol{y}} & \mathbf{R}+\mathbf{D} \end{bmatrix} \succeq 0 \\ & \mathbf{D} = \mathrm{diag}\{\delta_1,\delta_2,\ldots,\delta_M\} \succeq 0. \end{aligned} \tag{18}$$

While solving (18) we use Lemma 2 to parameterize $\mathbf{R}$ in terms of $\{\bar{\nu}_k\}_{k=0}^d$ as

$$\mathbf{R}_{jl} = \boldsymbol{h}_{jl}^{\mathsf{T}}\boldsymbol{\Phi}^{-1}[\ \bar{\nu}_d^* \ \cdots\ \bar{\nu}_1^*\ \ \bar{\nu}_0\ \ \bar{\nu}_1\ \ \cdots\ \ \bar{\nu}_d\ ]^{\mathsf{T}},$$

where the Toeplitz matrix

$$\bar{\mathbf{T}} := \begin{bmatrix} \bar{\nu}_0 & \bar{\nu}_1^* & \cdots & \bar{\nu}_d^* \\ \bar{\nu}_1 & \bar{\nu}_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \bar{\nu}_1^* \\ \bar{\nu}_d & \cdots & \bar{\nu}_1 & \bar{\nu}_0 \end{bmatrix} \succeq 0,$$

and $\mathcal{W}(\bar{\mathbf{T}})\succeq 0$. We call the resulting gridless SPICE algorithm TV motivated SPICE (TVSP). Upon convergence of TVSP (18), we estimate the noise standard deviation as $\frac{1}{M}\sum_{i=1}^{M}\sqrt{\delta_i}$, and use in (15). We call this approach as TVSP aided Denoising (TVSPDN). TVSP can also supply the initial estimates of $\{\nu_k\}_{k=0}^d$ in Theorem 1. The solution to (18) yields the optimal $\{\bar{\nu}_{l*}\}_{l=0}^d$ and $\bar{\mathbf{T}}_*$. Typically $\bar{\mathbf{T}}_*$ is singular, and its rank estimates K. Since $\mathbf{R}_* = \sum_{k=1}^{K}|a|_k^2\boldsymbol{\gamma}(f_k)\boldsymbol{\gamma}^*(f_k)$, by Lemma 2 we can write individual elements of $\bar{\mathbf{T}}_*$ as

$$\bar{\nu}_{l*} = \sum_{k=1}^{K}|a_k|^2 z_k^l, \quad z_k = \mathrm{e}^{\mathrm{i}2\pi f_k t_M/d}. \tag{19}$$

Notice that (19) is very similar to (8) with a subtle difference. In (8) $\{|a_k|\}_{k=1}^K$ appear, while (19) has $\{|a_k|^2\}_{k=1}^K$. Nevertheless, we can estimate $\{|a_k|, f_k\}_{k=1}^K$ from $\{\bar{\nu}_{l*}\}_{l=0}^d$ via Proney's method. We can use these estimates in (8) to calculate the initial values of $\{\nu_k\}_{k=0}^d$ to be supplied for solving (15). Supplying initial values close to the final solution results a quick convergence.

III. Simulation results

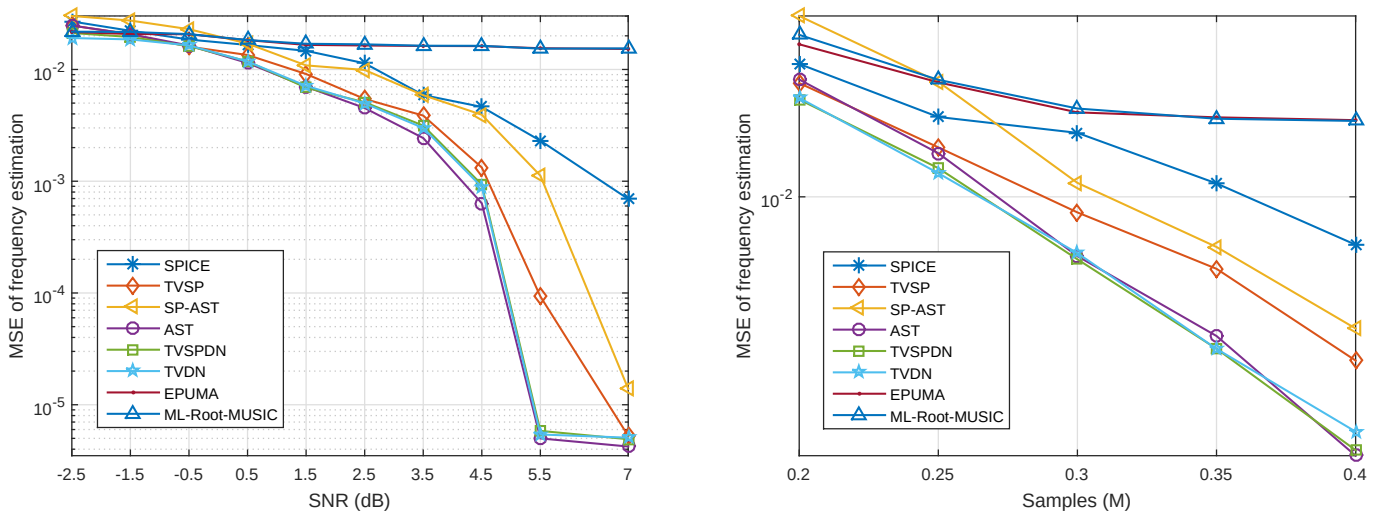
In the simulations we compare SPICE [13], TVSP, TVSPDN, atomic-norm soft thresholding (AST) [9], [22], maximum likelihood (ML) [34] and enhanced principle-singular-vector utilization for model analysis (EPUMA) [35]. ML solves a non-convex problem, and its performance depends heavily on the initialization. We often initialize ML by the root-MUSIC [36] estimate. Both EPUMA and root-MUSIC need to know K, and the estimates of the signal autocorrelation at integer valued instants $0, 1, 2, 3, \ldots$. We apply the interpolation methodology in [10] to estimate the signal autocorrelation at integers from irregularly sampled data. For SPICE we discretize the frequency grid into $N = 1000$ points. In the simulations, different versions of algorithms are denoted by

- TVSP: Proposed gridless SPICE.
- TVSPDN: Proposed gridless TVMA where noise variance is estimated by TVSP.
- TVDN: Proposed gridless TVMA where we use the actual value of noise variance.
- SP-AST: Atomic-norm soft thresholding (AST) [9], [22] where noise variance is estimated by TVSP.
- AST: Atomic-norm soft thresholding (AST) [9], [22] where we use the actual value of noise variance.
- SPICE: Grid based SPICE [13].
- EPUMA: EPUMA [35].
- ML-Root-MUSIC: ML initialized by root-MUSIC [36].

In the simulations we maintain $t_1 = 0$ and $t_M = 80$, while $M < t_M$. We consider two schemes of generating $\{t_j\}_{j=2}^{M-1}$:

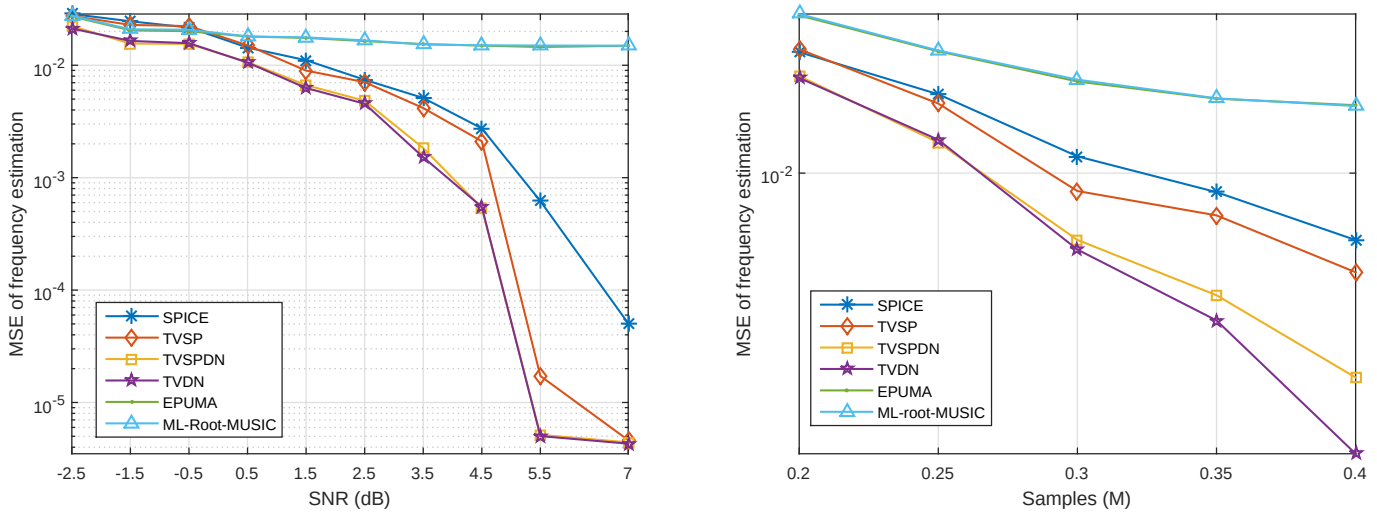
1) *Random Sampling on regular grid:* $\{t_j\}_{j=2}^{M-1}$ are drawn uniformly at random from the set $\{1, 2, \cdots, 79\}$, and we maintain $t_i \neq t_j, \forall i \neq j$. In this scheme every t_i is integer valued.
2) *Arbitrary sampling:* Every t_j is drawn uniformly at random from the interval $(0, 80)$, where $t_i \neq t_j, \forall i \neq j$.

The performance of the algorithms are measured in terms of mean squared error (MSE) of frequency estimation. To compute frequency estimation accuracy, we need to obtain an estimate of K. As noted in [9], [13], SPICE being non-parametric, generates many noise peaks at low signal-to-noise ratio (SNR). This makes estimation of K difficult. On the other hand root-MUSIC and PUMA assumes K is known. Therefore, for a fair comparison, we follow the procedure proposed in [14], [37]. We assume K is known. For SPICE, the K largest peaks of the spectra are used to estimate MSE, while it corresponds to the largest K components for gridless algorithms. The SNR is defined as $10\log_{10}(1/\sigma^2)$ [9], [38]. The following results are based on 1000 independent Monte-Carlo simulations.



(a) (b)

Fig. 1: MSE of estimating three frequencies with random sampling on regular grid. (a) MSE of frequency estimation as a function of SNR. $M = 24$. (b) MSE of frequency estimation as a function of number of samples. SNR=1.5 dB.



(a) (b)

Fig. 2: MSE of estimating three frequencies with arbitrary sampling. (a) MSE of frequency estimation as a function of SNR. $M = 24$. (b) MSE of frequency estimation as a function of number of samples. SNR=1.5 dB.

A. Random sampling on regular grid

We take $K = 3, f_1 = 0.05 + w_1, f_2 = 0.351 + w_2, f_3 = 0.362 + w_3, a_1 = 0.5\mathrm{e}^{\mathrm{i}\phi_1}, a_2 = \mathrm{e}^{\mathrm{i}\phi_2}, a_3 = \mathrm{e}^{\mathrm{i}\phi_3}$, where ϕ_1, ϕ_2, ϕ_3 are mutually independent identically and uniformly distributed in $[0, 2\pi)$. In addition, w_1, w_2 and w_3 are mutually independent, identically and uniformly distributed in $[-10^{-3}, 10^{-3}]$. In Figure-1(a) we plot the MSE as a function of SNR for $M = 24$. In Figure 1(b) we plot MSE as a function of M for 1.5 dB SNR. TVDN, AST and TVSPDN outperform other methods. TVSP performs similarly to these when M is smaller. There is a performance difference between AST and SP-AST. One may conclude that atomic soft thresholding requires to know the noise variance exactly to perform at its best. However, SP-AST cost function has a regularization term like the second term in (15), and the performance of SP-AST may improve by tuning the weight of regularization term. Note that EPUMA or ML-Root-MUSIC exhibit poor performance. We have seen that the performance of ML would improve and be the best if it is initialized by the output of, e.g., TVSPDN or TVSP, see below in Section III-C.

B. Arbitrary sampling

Figure 2 shows the simulations results for arbitrary sampling scenario. As before, We take $K = 3$, and $\{a_k, f_k\}_{k=1}^{K}$

are taken as in Section III-A. TVDN (TVMA with perfect knowledge of σ^2) outperforms TVSPDN when M is increased. Otherwise TVSPDN outperforms other algorithms. TVSP is somewhat better than SPICE. As before PUMA and ML-root-MUSIC are not as good as the others. AST does not appear in Figure 2 as it cannot handle arbitrarily sampled data.

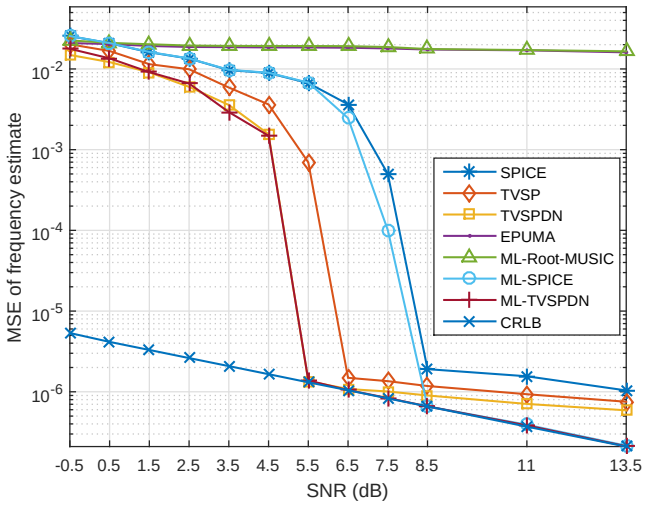


Fig. 3: The CRLB of MSE of frequency estimation for different SNR is compared with the MSE of different algorithms for arbitrary sampling case. The arbitrary sampled signal has three frequencies with $M = 24$.

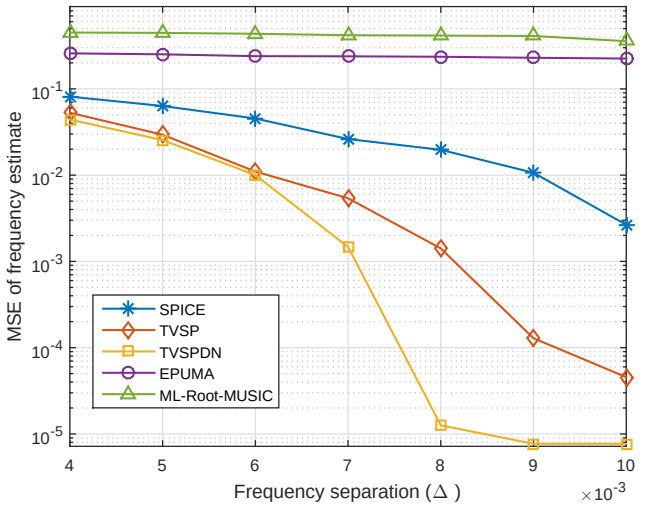


Fig. 4: MSE of frequency estimation of two frequencies separated by Δ. The arbitrary sampled signal has two frequencies with $M = 24$ and SNR=1.5 dB..

C. Cramér-Rao lower bound (CRLB)

Figure 3 compares the MSE of different algorithms with the associated CRB. Here $K = 3$, $f_1 = 0.496109, f_2 = 0.3509603, f_3 = 0.362684$, $a_1 = 0.5\mathrm{e}^{\mathrm{i}0.39}, a_2 = \mathrm{e}^{\mathrm{i}0.28}, a_2 = \mathrm{e}^{\mathrm{i}0.15}$. We generate $M = 24$ arbitrary samples with $t_1 = 0$ and $t_M = 80$. However, unlike in above, here the sampling instants are kept fixed, and only the noise realization varies from one simulation to the next. We compare the numerical MSE of different algorithms with the analytical CRB [8] in Figure 3, where several variants of ML algorithms are considered. ML-SPICE refers to the ML implementation initialized by the output of TVSP, while ML-TVSPDN refers to the ML implementation initialized by the output of TVSPDN. ML-SPICE and ML-TVSPDN are able to achieve CRB at high SNR. Note that at large SNR the performance of TVSP and TVSPDN saturate. This saturation of level depends on d. Recall that we set a desired precision level ϵ, and choose d such that $|\lambda_j| < \epsilon$ for all $j > 2d$. The systematic error due to this approximation dominates when σ^2 is small causing the performance saturation. This saturation level can be lowered by increasing d in expense of added computational complexity.

D. Estimation of two closely spaced frequencies

Figure-4 demonstrates the capability of estimating two closely spaced frequencies from an arbitrary sampled signal by different algorithms. We set $K = 2, a_1 = a_2 = 1, M = 24$ and SNR=1.5 dB. The frequencies are taken as $f_1 = 0.351 + w_1, f_2 = f_1 + \Delta$ where w_1 is uniformly distributed in $[-10^{-3}, 10^{-3}]$. The MSE results in Figure 4 plotted as a function of Δ clearly demonstrate the superiority of the gridless methods in resolving closely spaced frequencies.

IV. Conclusions

We have proposed a gridless algorithm for estimating the sparse line spectrum of a signal from arbitrarily sampled data. In particular, we have given a gridless semidefinite programming algorithm for computing the TVMA estimate. This algorithm can be seen as the extension of some available results [2], [9], [22], [27], [39]–[41] which solve the problem when the sampling instants lie on the uniform sampling grid. The numerical simulation study demonstrates the utility of the proposed method. It produces satisfactory results with small number of samples and demonstrates good resolution performance.

To derive our algorithm we have presented few new results. Lemma 1 gives a generic characterization of the TVMA in terms of the cone $\mathbb{K}$. This characterization might be useful in other applications. The finite dimensional characterization of $\mathbb{K}$ in Lemma 2 is an useful result as well. For instance, we have used it in deriving the gridless version of SPICE. Finally Theorem 3 used to prove Lemma 2 is a generalization of the classical Caratheodory-Fejer characterization of the classical moment problem, which might be useful in other applications.

Appendix

A. Proof of solvability of (1) for $K = M$

We show that there always exists a solution to (1) for $K = M$. We let $M_1 = \max(M, t_M)$. It suffices to show that there are $\{a_k\}_{k=1}^{M}$ satisfying (1) with $K = M$ if we take

$$\bar{f}_k = (k-1)F/M_1 - F/2, \quad k = 0, 1, \ldots, M-1.$$

This means $f_k = k/M_1 - 1/2$. Verify that $|f_k| \leq 1/2$ for all k. It is sufficient to show that the $\mathbf{B} \in \mathbb{C}^{M\times M}$ defined as

$$\mathbf{B}_{jk} = \mathrm{e}^{\mathrm{i}2\pi \bar{f}_k \bar{t}_j} = \mathrm{e}^{\mathrm{i}2\pi f_k t_j} = \mathrm{e}^{\mathrm{i}2\pi t_j (k/M_1 - 1/2)} = z_j^k \mathrm{e}^{-\mathrm{i}\pi t_j}$$

is non-singular, because then $\mathbf{B}^{-1}\boldsymbol{y}$ gives $\{a_k\}_{k=1}^M$. Here we write $z_j = \mathrm{e}^{\mathrm{i}2\pi t_j/t_M}$ for short. Now the above implies

$$\mathbf{B} = \mathbf{OV}, \quad \mathbf{O} = \mathrm{diag}\{\mathrm{e}^{-\mathrm{i}\pi t_1}, \ldots, \mathrm{e}^{-\mathrm{i}\pi t_M}\},$$

with $\mathbf{V}$ being a Vandermonde matrix with $\mathbf{V}_{jk} = z_j^k$. Since all t_j are distinct, and $t_j < t_M \leq M_1$, we note that $\{z_j\}_{j=1}^M$ are all distinct. Hence $\mathbf{V}$ is non-singular. Now since $\mathbf{O}$ is unitary, $\mathbf{B} = \mathbf{OV}$ is non-singular.

B. Proof of Lemma 1

We need the following result, which has been also used in several other papers on atomic norm minimization based methods, e.g., [9].

Proposition 1. *Suppose that the LMI*

$$\left[\begin{array}{cc} w & \boldsymbol{y}^* \\ \boldsymbol{y} & \mathbf{Q} \end{array}\right] \succeq 0. \tag{20}$$

holds. Then for every $\bar{\mathbf{Q}}$ satisfying $\mathbf{Q} = \bar{\mathbf{Q}}\bar{\mathbf{Q}}^*$ there exists a corresponding vector $\bar{\boldsymbol{y}}$ such that $\boldsymbol{y} = \bar{\mathbf{Q}}\bar{\boldsymbol{y}}$. In addition, the minimum value of w subject to the LMI (20) is

$$\boldsymbol{y}^*\mathbf{Q}^\dagger\boldsymbol{y} = \min\{\bar{\boldsymbol{y}}^*\bar{\boldsymbol{y}} \ : \ \boldsymbol{y} = \bar{\mathbf{Q}}\bar{\boldsymbol{y}}\}.$$

Let $w_*, \mathbf{Q}_*$ be the solutions to (11). Since $\mathbf{Q}_* \in \mathbb{K}$ we note by definition of $\mathbb{K}$ that there are strictly positive numbers $\mathring{\alpha}_k$ and frequencies $\mathring{f}_k \in [-1/2, 1/2]$ such that

$$\mathbf{Q}_* = \sum_{k=1}^P \mathring{\alpha}_k \mathbf{\Gamma}(\mathring{f}_k) = \sum_{k=1}^P [\boldsymbol{\gamma}(\mathring{f}_k)\sqrt{\mathring{\alpha}_k}][\boldsymbol{\gamma}(\mathring{f}_k)\sqrt{\mathring{\alpha}_k}]^*.$$

Here we don't require the above decomposition to be unique, and the value of P can be more than M. The above equation can be re-written as $\mathbf{Q}_* = \bar{\mathbf{Q}}\bar{\mathbf{Q}}^*$, with

$$\bar{\mathbf{Q}} = [\ \boldsymbol{\gamma}(f_1)\sqrt{\mathring{\alpha}_1} \quad \cdots \quad \boldsymbol{\gamma}(f_P)\sqrt{\mathring{\alpha}_P}\ ]$$

Since $w_*, \mathbf{Q}_*$ are the solutions to (11), the linear matrix inequality (LMI)

$$\left[\begin{array}{cc} w_* & \boldsymbol{y}^* \\ \boldsymbol{y} & \mathbf{Q}_* \end{array}\right] \succeq 0$$

must hold. Hence Proposition 1 ensures the existence of complex numbers $\{\beta_k\}_{k=1}^P$ such that

$$\boldsymbol{y} = \sum_{k=1}^P \boldsymbol{\gamma}(\mathring{f}_k)\beta_k\sqrt{\mathring{\alpha}_k}, \qquad w_* = \sum_{k=1}^P |\beta_k|^2. \tag{21}$$

Next we show $|\beta_k| = \sqrt{\mathring{\alpha}_k}$ by contradiction. Suppose $|\beta_k| \neq \sqrt{\mathring{\alpha}_k}$. Take

$$\hat{w} = \sum_{k=1}^P |\beta_k|\sqrt{\mathring{\alpha}_k}, \quad \hat{\mathbf{Q}} = \sum_{k=1}^P |\beta_k|\sqrt{\mathring{\alpha}_k}\ \boldsymbol{\gamma}(\mathring{f}_k)\boldsymbol{\gamma}^*(\mathring{f}_k),$$

so that $\hat{\mathbf{Q}} \in \mathbb{K}$. Verify that

$$\left[\begin{array}{cc} \hat{w} & \boldsymbol{y}^* \\ \boldsymbol{y} & \hat{\mathbf{Q}} \end{array}\right] = \sum_{k=1}^P \left[\begin{array}{c} 1 \\ \boldsymbol{\gamma}(\mathring{f}_k)\frac{\beta_k}{|\beta_k|} \end{array}\right] |\beta_k|\sqrt{\mathring{\alpha}_k} \left[\begin{array}{c} 1 \\ \boldsymbol{\gamma}(\mathring{f}_k)\frac{\beta_k}{|\beta_k|} \end{array}\right]^*$$

is non-negative definite, and thereby $\hat{w}, \hat{\mathbf{Q}}$ belong to the feasible set of the optimization problem (11). In addition,

$$w_* - \hat{w} + \boldsymbol{e}^*(\mathbf{Q}_* - \hat{\mathbf{Q}})\boldsymbol{e} = \sum_{k=1}^P \{\mathring{\alpha}_k + |\beta_k|^2 - 2|\beta_k|\sqrt{\mathring{\alpha}_k}\} > 0,$$

but that leads to a contradiction since w_*,$\mathbf{Q}_*$ are the solutions to (11). Hence $|\beta_k| = \sqrt{\mathring{\alpha}_k}$, and

$$w_* = \boldsymbol{e}^*\mathbf{Q}_*\boldsymbol{e} = (w_* + \boldsymbol{e}^*\mathbf{Q}_*\boldsymbol{e})/2 = \sum_k \mathring{\alpha}_k.$$

Also we can write $\beta_k = \mathring{\alpha}_k\mathring{\phi}_k$ where $|\mathring{\phi}_k| = 1$. Hence the first equation in (21) gives us an atomic decomposition of $\boldsymbol{y}$:

$$\boldsymbol{y} = \sum_{k=1}^P \boldsymbol{\gamma}(\mathring{f}_k)\mathring{\phi}_k\mathring{\alpha}_k, \tag{22}$$

and therefore using (10) we infer that

$$||\boldsymbol{y}||_{\mathcal{A}} \leq \sum_{k=1}^P \mathring{\alpha}_k = (w_* + \boldsymbol{e}^*\mathbf{Q}_*\boldsymbol{e})/2. \tag{23}$$

Now consider a solution to (10). Such a solution constitute some $\check{P}$ positive numbers $\{\check{\alpha}_k\}_{k=1}^{\check{P}}$, associated frequencies $\{\check{f}_k\}_{k=1}^{\check{P}}$ with each $\check{f}_k \in [-1/2, 1/2]$, and unimodular complex numbers $\{\check{\phi}_k\}_{k=1}^{\check{P}}$ such that

$$||\boldsymbol{y}||_{\mathcal{A}} = \sum_{k=1}^{\check{P}} \check{\alpha}_k, \tag{24}$$

and in addition the atomic decomposition

$$\boldsymbol{y} = \sum_{k=1}^P \boldsymbol{\gamma}(\check{f}_k)\check{\phi}_k\check{\alpha}_k \tag{25}$$

holds. Take

$$\check{w} = \sum_{k=1}^{\check{P}} \check{\alpha}_k, \quad \hat{\mathbf{Q}} = \sum_{k=1}^P \check{\alpha}_k\ \boldsymbol{\gamma}(\check{f}_k)\boldsymbol{\gamma}^*(\check{f}_k),$$

so that $\check{\mathbf{Q}} \in \mathbb{K}$, and in addition

$$\left[\begin{array}{cc} \check{w} & \boldsymbol{y}^* \\ \boldsymbol{y} & \check{\mathbf{Q}} \end{array}\right] = \sum_{k=1}^{\check{P}} \left[\begin{array}{c} 1 \\ \boldsymbol{\gamma}(\check{f}_k)\check{\phi}_k \end{array}\right] \check{\alpha}_k \left[\begin{array}{c} 1 \\ \boldsymbol{\gamma}(\check{f}_k)\check{\phi}_k \end{array}\right]^*$$

is non-negative definite. Thus $\check{w}, \check{\mathbf{Q}}$ belong to the feasible set of the optimization problem (11). Since $w_*, \mathbf{Q}_*$ are the solutions to (11) we conclude that

$$(w_* + \boldsymbol{e}^*\mathbf{Q}_*\boldsymbol{e})/2 = \sum_k \mathring{\alpha}_k \leq (\check{w} + \boldsymbol{e}^*\check{\mathbf{Q}}\boldsymbol{e})/2 = \sum_k \check{\alpha}_k = ||\boldsymbol{y}||_{\mathcal{A}}.$$

This observation and (23) imply that $||\boldsymbol{y}||_{\mathcal{A}} = (w_* + \boldsymbol{e}^*\mathbf{Q}_*\boldsymbol{e})/2$. Consequently, it follows that

- The atomic decomposition (22) obtained from the solution of (11) gives an optimal solution to (10); and
- An atomic decomposition (25) achieving the optimality criterion (24) gives an optimal solution $\check{w}$ and $\check{\mathbf{Q}}$ to (11).

C. Proof of Lemma 2

1) Some preliminary results involving PSWFs: The PSWFs $\{\varphi_j\}_{j=0}^{\infty}$ are real-valued, and form a complete orthonormal basis of $\mathbb{L}^2$. In particular, $\varphi_j(-\xi) = \varphi_j(\xi)$ for an even valued j; and $\varphi_j(-\xi) = -\varphi_j(\xi)$ for an odd j. These functions form a T-system on $[-1, 1]$, *i.e.* for any n and some distinct points $u_0, u_1, \cdots, u_n$ in $[-1, 1]$ the $n \times n$ matrix $\mathbf{\Phi}$ with $\mathbf{\Phi}_{jk} = \varphi_{k-1}(u_{j-1})$ is nonsingular.

Fix $\tau \in [-1, 1]$. Define $q_\tau \in \mathbb{L}^2$ such that $q_\tau(\xi) = \exp(\mathrm{i}c\xi\tau)$. Using (5) with standard inner product in $\mathbb{L}^2$ we have $(\mathcal{E}r)(\tau) = \langle q_\tau, r\rangle$. Expand q_τ in the ortho-basis $\{\varphi_j\}_{j=0}^{\infty}$. For $\xi \in [-1, 1]$

$$\begin{aligned}\mathrm{e}^{\mathrm{i}c\xi\tau} = q_\tau(\xi) &= \sum_{j=0}^{\infty} \varphi_j(\xi)\ \langle q_\tau, \varphi_j\rangle \\ &= \sum_{j=0}^{\infty} \varphi_j(\xi)\ \{(\mathcal{E}\varphi_j)(\tau)\} = \sum_{j=0}^{\infty} \lambda_j\ \varphi_j(\xi)\ \varphi_j(\tau), \quad (26)\end{aligned}$$

where we use (6) in the last equality. Although (26) is infinite series, only a finite number terms in the expansions have some noticeable magnitude. The number of terms needed can be found using the following result [5], [42].

Theorem 2. *Given positive number ϵ there are*

$$n = 2c/\pi + \log\{2\pi/(\epsilon^2 c) - 1\}\log(c)/\pi^2 + O\{\log(c)\}$$

eigenvalues of $\mathcal{E}$ that are greater in magnitude than ϵ.

We refer the readers to [3], [5], [42] for numerical illustration of the decay phenomenon. In the implementation of our method we set ϵ somewhat smaller than the precision of the optimization routine used to solve the problem in Theorem 1. Our application requires n odd, i.e. $n = 2d + 1$ for an appropriate d. In particular,

$$2d > 2c/\pi = 2t_M \quad (27)$$

for practical values of ϵ. Take any $t \in [-t_M, t_M]$ and $f \in [-1/2, 1/2]$. By setting $\tau = t/t_M$, $\xi = 2f$ and using $c = \pi t_M$ in (26) we have

$$\mathrm{e}^{\mathrm{i}2\pi ft} = \sum_{j=0}^{2d} \lambda_j\ \varphi_j(2f)\ \varphi_j(t/t_M). \quad (28)$$

without any loss of precision. Let us define $\boldsymbol{\psi}_f, \boldsymbol{\phi}_f \in \mathbb{C}^{2d+1}$ such that

$$\boldsymbol{\phi}_f(k) = \mathrm{e}^{\mathrm{i}2\pi f t_M(k-d-1)/d}, \qquad \boldsymbol{\psi}_f(j) = \lambda_{j-1}\varphi_{j-1}(2f).$$

In particular, (28) gives $\boldsymbol{\phi}_f = \mathbf{\Phi}\boldsymbol{\psi}_f,\ \forall f \in [-1/2, 1/2]$, where $\mathbf{\Phi} \in \mathbb{R}^{(2d+1)\times(2d+1)}$ is defined as

$$\mathbf{\Phi}_{kj} = \varphi_{j-1}((k - d - 1)/d).$$

Since $\{\varphi_j\}_{j=0}^{\infty}$ forms a T-system $\mathbf{\Phi}$ is nonsingular.

2) The main proof: Let $\boldsymbol{h}_{jl}$ is such that $\boldsymbol{h}_{kl}(j) = \varphi_{j-1}((t_k - t_l)/t_M)$. Using (28), the definition of $\mathbf{\Gamma}$, and (12) it follows that $\mathbf{Q} \in \mathbb{K}$ if and only if

$$\mathbf{Q}_{jl} = \int_{-1/2}^{1/2} \mathrm{e}^{\mathrm{i}2\pi f(t_j - t_l)}\ \mathrm{d}\mu(f) = \int_{-1/2}^{1/2} \boldsymbol{h}_{jl}^{\mathsf{T}}\boldsymbol{\psi}(f)\ \mathrm{d}\mu(f)$$

for some positive measure μ. Substituting $\boldsymbol{\psi}_f = \mathbf{\Phi}^{-1}\boldsymbol{\phi}_f$ this means $\mathbf{Q} \in \mathbb{K}$ if and only if

$$\mathbf{Q}_{jl} = \boldsymbol{h}_{jl}^{\mathsf{T}}\mathbf{\Phi}^{-1}[\ \nu_{-d}\ \ \nu_{-d+1}\ \ \cdots\ \ \nu_d\ ]^{\mathsf{T}}, \quad (29)$$

where $\{\nu_k\}_{k=-d}^{d}$ are such that

$$[\ \nu_{-d}\ \ \nu_{-d+1}\ \ \cdots\ \ \nu_d\ ]^{\mathsf{T}} = \int_{-1/2}^{1/2} \boldsymbol{\phi}_f\ \mathrm{d}\mu(f) \quad (30)$$

for some positive measure μ. Substituting $\theta = 2\pi f t_M/d$, and denoting $\bar{\boldsymbol{\mu}}(\theta) = \mu(d\theta/(2\pi t_M))$ we can rewrite (30) as

$$\nu_k = \int_{-\theta_0}^{\theta_0} \mathrm{e}^{\mathrm{i}k\theta}\ \mathrm{d}\bar{\mu}(\theta). \quad (31)$$

where $\theta_0 = \pi t_M/d < \pi$ according to (27). Hence we can rephrase: $\mathbf{Q} \in \mathbb{K}$ if and only if (29) and (31) holds for some positive measure $\bar{\boldsymbol{\mu}}$ on $[-\theta_0, \theta_0]$. That $\nu_{-k} = \nu_k^*$ is an obvious requirement. It remains to prove the following result.

Theorem 3. *There is a positive measure $\bar{\mu}$ on $[-\theta_0, \theta_0],\ \theta_0 < \pi$ satisfying (31) if and only if $\mathbf{T} \succeq 0$ and $\mathcal{W}(\mathbf{T}) \succeq 0$. In addition $\bar{\mu}$ is unique if and only of $\mathbf{T}$ is singular with rank $r \leq d$, and in that $\bar{\mu}$ is of the form*

$$\bar{\mu}(\theta) = \sum_{k=1}^{r} w_k u(\theta - \theta_k),\ w_k > 0,\ \theta_k \in [-\theta_0, \theta_0]$$

for all $k = 1, 2, \ldots, r$.

3) Proof of Theorem 3: It is a classical result [33] that $\bar{\mu}$ satisfying (30) is unique (and in that case a degenerate positive measure) if and only if $\mathbf{T}$ is singular, non-negative definite. The following therefore, gives a proof of the first part of the Theorem 3, which essentially extends the classical theory (valid when $\theta_0 = \pi$) in order to account for any $\theta_0 < \pi$. We show that this can be ensured by imposing an additional constraint $\mathcal{W}(\mathbf{T}) \succeq 0$. Denoting

$$\boldsymbol{\omega}(\theta) := [\ 1\ \ \mathrm{e}^{-\mathrm{i}\theta}\ \ \cdots\ \ \mathrm{e}^{-\mathrm{i}d\theta}\ ]^*,$$

let $\mathbb{V} \subset \mathbb{H}$ be the conic hull of $\{\boldsymbol{\omega}(\theta)\boldsymbol{\omega}^*(\theta) : \theta \in [-\theta_0, \theta_0]\}$. It is sufficient to show that $\mathbf{T} \in \mathbb{V}$ if and only if $\mathbf{T} \succeq 0$ and $\mathcal{W}(\mathbf{T}) \succeq 0$.

Proposition 2. *A complex vector $\boldsymbol{u} = z\boldsymbol{\omega}(\theta)$ for some real-valued θ and some complex number z if and only if*

$$(\mathbf{J}_2 - \mathbf{J}_1)\boldsymbol{u} = \mathrm{i}\tan(\theta/2)(\mathbf{J}_1 + \mathbf{J}_2)\boldsymbol{u} \quad (32)$$

Proof: Note that

$$\mathrm{e}^{\mathrm{i}\theta} = \frac{\mathrm{e}^{\mathrm{i}\theta/2}}{\mathrm{e}^{-\mathrm{i}\theta/2}} = \frac{\cos(\theta/2) + \mathrm{i}\sin(\theta/2)}{\cos(\theta/2) - \mathrm{i}\sin(\theta/2)} = \frac{1 + \mathrm{i}\tan(\theta/2)}{1 - \mathrm{i}\tan(\theta/2)} \quad (33)$$

Now $\boldsymbol{u} = z\boldsymbol{\omega}(\theta)$ for some real-valued θ and some complex number z if and only if $\mathbf{J}_2\boldsymbol{u} = \mathbf{J}_1\boldsymbol{u}\mathrm{e}^{\mathrm{i}\theta}$, which using (33) is equivalent to (32). ∎

Lemma 3. *If $\mathbf{T}$ be a $(d+1)\times(d+1)$ non-negative definite Toeplitz matrix such that $\mathcal{W}(\mathbf{T}) \succeq 0$, then $\mathbf{T} \in \mathbb{V}$.*

Proof: Since $\mathbf{T}$ is Toeplitz non-negative definite matrix it admits a decomposition

$$\mathbf{T} = \sum_{k=1}^{r} w_k \boldsymbol{\omega}(\theta_k)\boldsymbol{\omega}^*(\theta_k) \tag{34}$$

where $w_k > 0$ for all k, and r is the rank of $\mathbf{T}$. The above decomposition is unique only if $r \leq d$. The following series of arguments holds regardless of whether $r \leq d$ or not. Using Proposition 2 and (34) in the definition of $\mathcal{W}(\mathbf{T})$ we get

$$\mathcal{W}(\mathbf{T}) = \sum_{k=1}^{r} \bar{w}_k (\mathbf{J}_1+\mathbf{J}_2)\boldsymbol{\omega}(\theta_k)\boldsymbol{\omega}^*(\theta_k)(\mathbf{J}_1+\mathbf{J}_2)^{\intercal}, \tag{35}$$

where

$$\bar{w}_k = w_k\{\tan^2(\theta_0/2) - \tan^2(\theta_k/2)\}. \tag{36}$$

It is sufficient to show for any decomposition of $\mathbf{T}$ of the form (34) $\theta_k \in [-\theta_0, \theta_0]$, $k = 1, 2, \ldots, r$, and thereby $\mathbf{T} \in \mathbb{V}$. We prove this by contradiction. Suppose that there is at least one element in the set $\{\theta_1, \ldots, \theta_r\}$ which does not belong to $[-\theta_0, \theta_0]$. Without any loss of generality suppose that $|\theta_1| > \theta_0$, so that

$$\tan^2(\theta_0/2) - \tan^2(\theta_1/2) < 0.$$

Consider the economy size QL factorization

$$(\mathbf{J}_2+\mathbf{J}_2)[\ \boldsymbol{\omega}(\theta_1) \quad \cdots \quad \boldsymbol{\omega}(\theta_r)\ ] = \mathbf{O}\mathbf{L},$$

where $\mathbf{L}$ is an $r\times r$ lower triangular matrix, and $\mathbf{O}$ is a $N\times r$ matrix such that $\mathbf{O}^*\mathbf{O} = \mathbf{I}$. Denote the first column of $\mathbf{O}$ by $\mathbf{q}$. Then

$$\begin{aligned}\mathbf{q}^*\{\mathcal{W}(\mathbf{T})\}\mathbf{q} &= \mathbf{q}^*\mathbf{O}\mathbf{L}\mathrm{diag}\{\bar{w}_1, \ldots, \bar{w}_r\}\mathbf{L}^*\mathbf{O}^*\mathbf{q} \\ &= [\mathbf{L}]_{11}^2 w_1\{\tan^2(\theta_0/2) - \tan^2(\theta_1/2)\} < 0. \end{aligned} \tag{37}$$

But this contradicts the assumption that $\mathcal{W}(\mathbf{T}) \succeq 0$, and the proof is complete. ∎

Lemma 4. *If $\mathbf{T} \in \mathbb{V}$ then $\mathbf{T}$ is a Toeplitz nonnegative definite matrix such that $\mathcal{W}(\mathbf{T}) \succeq 0$.*

Proof: Since $\mathbf{T} \in \mathbb{V}$ there are $\theta_1, \theta_2, \ldots, \theta_r$ with each $\theta_k \in [-\theta_0, \theta_0]$ and positive numbers $\{w_k\}_{k=1}^r$ such that (34) holds. Since each $\boldsymbol{\omega}(\theta_k)\boldsymbol{\omega}^*(\theta_k)$ are positive semidefinite Toeplitz matrices, their conic combination $\mathbf{T}$ is a non-negative definite Toeplitz matrix. Using Proposition 2 equation (34) implies (35). Since each $\theta_k \in [-\theta_0, \theta_0]$, we know $\bar{w}_k \geq 0$ for all k. Then it is readily verified from (35) that $\mathcal{W}(\mathbf{T})$ is non-negative definite. ∎

Combining Lemma 3 and 4 we get Theorem 3.

D. Hermitian property of $\mathbf{Q}$ *in Theorem 1*

By definition $\boldsymbol{\Phi} = [\ \boldsymbol{p}_0 \ \ \boldsymbol{p}_1 \ \ \cdots \ \ \boldsymbol{p}_{2d}\ ]$, where we denote

$$\boldsymbol{p}_j = [\ \varphi_j(-1) \quad \varphi_j(-1+\tfrac{1}{d}) \quad \cdots \quad \varphi_j(1-\tfrac{1}{d}) \quad \varphi_j(1)\ ]^{\intercal}.$$

In addition, $\boldsymbol{v}^{(\updownarrow)}$ denotes the vector $\boldsymbol{v}$ flipped upside down. Let $\boldsymbol{\nu} = [\ \nu_d^* \ \cdots \ \nu_1^* \ \ \nu_0 \ \ \nu_1 \ \cdots \ \nu_d\ ]^{\intercal}$, and $\boldsymbol{q} = \boldsymbol{\Phi}^{-1}\boldsymbol{\nu}$. We first show that $\boldsymbol{q}(k)$ is real-valued when k is odd and $\boldsymbol{q}(k)$ is purely imaginary when k is even.When j is even, φ_j is an even function, and thus, $\boldsymbol{p}_j^{(\updownarrow)} = \boldsymbol{p}_j$. When j is odd, φ_j is an odd function, and thus, $\boldsymbol{p}_j^{(\updownarrow)} = -\boldsymbol{p}_j$. Hence using $\boldsymbol{\nu} = \boldsymbol{\Phi}\boldsymbol{q}$,

$$\boldsymbol{\nu} + \boldsymbol{\nu}^{(\updownarrow)} = \sum_{j=0}^{2d}\{\boldsymbol{p}_j + \boldsymbol{p}_j^{(\updownarrow)}\}\boldsymbol{q}(j+1) = 2\sum_{j=0}^{d}\boldsymbol{p}_{2j}\boldsymbol{q}(2j+1)$$

$$\boldsymbol{\nu} - \boldsymbol{\nu}^{(\updownarrow)} = \sum_{j=0}^{2d}\{\boldsymbol{p}_j - \boldsymbol{p}_j^{(\updownarrow)}\}\boldsymbol{q}(j+1) = 2\sum_{j=1}^{d}\boldsymbol{p}_{2j-1}\boldsymbol{q}(2j).$$

The PSWFs φ_j are all real-valued. Hence $\boldsymbol{p}_j$ is real-valued. Also, $\boldsymbol{\nu} + \boldsymbol{\nu}^{(\updownarrow)}$ is real-valued by construction of $\boldsymbol{\nu}$. On the other hand, Φ is a non-singular matrix. Hence the first of above equalities imply that $\mathrm{Im}\{\boldsymbol{q}(2j+1)\} = 0$. Similarly, $\boldsymbol{\nu} - \boldsymbol{\nu}^{(\updownarrow)}$ being imaginary, we get $\mathrm{Re}\{\boldsymbol{q}(2j)\} = 0$. Since $\boldsymbol{q} = \boldsymbol{\Phi}^{-1}\boldsymbol{\nu}$, by definition of $\mathbf{Q}$ in Theorem 1 we note that

$$\mathbf{Q}_{kl} = \sum_{j=0}^{2d}\varphi_j((t_k - t_l)/t_M)\boldsymbol{q}(j+1).$$

Since $\mathrm{Re}\{\boldsymbol{q}(2j)\} = 0$ and $\mathrm{Im}\{\boldsymbol{q}(2j+1)\} = 0$, we have

$$\mathrm{Re}(\mathbf{Q}_{kl}) = \sum_{j=0}^{d}\varphi_{2j}((t_l - t_k)/t_M)\boldsymbol{q}(2j+1)$$

$$\mathrm{Im}(\mathbf{Q}_{kl}) = -\mathrm{i}\sum_{j=1}^{d}\varphi_{2j-1}((t_l - t_k)/t_M)\boldsymbol{q}(2j).$$

Since φ_{2j} is an even function, and φ_{2j-1} is an odd function, the above imply that $\mathbf{Q}_{lk} = \mathbf{Q}_{kl}^*$, implying $\mathbf{Q} = \mathbf{Q}^*$.

REFERENCES

[1] E. J. Candés and C. Fernandez-Granda, "Towards a mathematical theory of super-resolution," vol. 67, p. 906956, 2014.

[2] G. Tang, B. Bhaskar, P. Shah, and B. Recht, "Compressed sensing off the grid," *Information Theory, IEEE Transactions on*, vol. 59, no. 11, pp. 7465–7490, Nov 2013.

[3] D. Slepian and H. O. Pollak, "Prolate spheroidal wave functions, Fourier analysis and uncertainty-I," *Bell System Technical Journal*, vol. 40, pp. 43–63, 1961.

[4] H. J. Landau and H. O. Pollak, "Prolate spheroidal wave functions, Fourier analysis and uncertainty-II," *Bell System Technical Journal*, vol. 40, pp. 65–84, 1961.

[5] ——, "Prolate spheroidal wave functions, Fourier analysis and uncertainty-III: The dimension of the space of essentially time and bandlimited signals," *Bell System Technical Journal*, vol. 41, pp. 1295–1336, 1962.

[6] H. Xiao, V. Rokhlin, and N. Yarvin, "Prolate spheroidal wave functions, quadrature and interpolation," *Inverse problems*, vol. 17, pp. 805–838, 2001.

[7] C. J. Bouwkamp, "On spheroidal wave functions of order zero," *Journal of mathematical physics*, vol. 26, pp. 79–92, 1957.

[8] P. Stoica and R. Moses, *Spectral Analysis of Signals*. Prentice-Hall, Upper Saddle River, NJ, 2005.

[9] Z. Yang and L. Xie, "On gridless sparse methods for line spectral estimation from complete and incomplete data," *Signal Processing, IEEE Transactions on*, vol. 63, no. 12, pp. 3139–3153, June 2015.

[10] P. Stoica and N. Sandgren, "Spectral analysis of irregularly-sampled data: Paralleling the regularly-sampled data approaches," *Digital Signal Processing*, vol. 16, no. 6, pp. 712 – 734, 2006. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S105120040600131X

[11] T. Yardibi, J. Li, P. Stoica, M. Xue, and A. Baggeroer, "Source localization and sensing: A nonparametric iterative adaptive approach based on weighted least squares," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 46, no. 1, pp. 425–443, Jan 2010.

[12] P. Stoica, J. Li, and J. Ling, "Missing data recovery via a nonparametric iterative adaptive approach," *Signal Processing Letters, IEEE*, vol. 16, no. 4, pp. 241–244, April 2009.

[13] P. Stoica, P. Babu, and J. Li, "SPICE: A sparse covariance-based estimation method for array processing," *Signal Processing, IEEE Transactions on*, vol. 59, no. 2, pp. 629–638, Feb 2011.

[14] P. Stoica and P. Babu, "SPICE and LIKES: Two hyperparameter-free methods for sparse-parameter estimation," *Signal Processing*, vol. 92, no. 7, pp. 1580 – 1590, 2012.

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

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

[17] M. F. Duarte and R. G. Baraniuk, "Spectral compressive sensing," *Applied and Computational Harmonic Analysis*, vol. 35, no. 1, pp. 111 – 129, 2013.

[18] L. Hu, Z. Shi, J. Zhou, and Q. Fu, "Compressed sensing of complex sinusoids: An approach based on dictionary refinement," *IEEE Transactions on Signal Processing*, vol. 60, no. 7, pp. 3809–3822, 2012.

[19] Z. Yang, C. Zhang, and L. Xie, "Robustly stable signal recovery in compressed sensing with structured matrix perturbation," *IEEE Transactions on Signal Processing*, vol. 60, no. 9, pp. 4658–4671, 2012.

[20] C. Austin, J. Ash, and R. Moses, "Dynamic dictionary algorithms for model order and parameter estimation," *IEEE Transactions on Signal Processing*, vol. 61, no. 20, pp. 5117–5130, 2013.

[21] P. Stoica and P. Babu, "Sparse estimation of spectral lines: Grid selection problems and their solutions," *IEEE Transactions on Signal Processing*, vol. 60, pp. 962–967, 2012.

[22] B. Bhaskar, G. Tang, and B. Recht, "Atomic norm denoising with applications to line spectral estimation," *Signal Processing, IEEE Transactions on*, vol. 61, no. 23, pp. 5987–5999, Dec 2013.

[23] Y. Chen and Y. Chi, "Robust spectral compressed sensing via structured matrix completion," *Information Theory, IEEE Transactions on*, vol. 60, no. 10, pp. 6576–6601, Oct 2014.

[24] J. Azais, Y. D. Castro, and F. Gamboa, "Spike detection from inaccurate samplings," *Applied and Computational Harmonic Analysis*, vol. 38, no. 2, pp. 177 – 195, 2015.

[25] G. Tang, B. Bhaskar, and B. Recht, "Near minimax line spectral estimation," in *Information Sciences and Systems (CISS), 2013 47th Annual Conference on*, March 2013, pp. 1–6.

[26] Z. Yang, L. Xie, and C. Zhang, "A discretization-free sparse and parametric approach for linear array signal processing," *Signal Processing, IEEE Transactions on*, vol. 62, no. 19, pp. 4959–4973, Oct 2014.

[27] P. Stoica, G. Tang, Z. Yang, and D. Zachariah, "Gridless compressive-sensing methods for frequency estimation," in *EUSIPCO*, Lisbon, Portugal, 2014.

[28] G. Foster, "The cleanest fourier spectrum," *The Astronomical Journal*, vol. 109, pp. 1889–1902, April 1995.

[29] D. H. Roberts, J. Lehar, and J. W. Dreher, "Time series analysis with CLEAN-I derivation of a spectrum," *The Astronomical Journal*, vol. 93, pp. 968–989, April 1987.

[30] O. López, R. Kumar, O. Yilmaz, and F. J. Herrmann, "Off-the-grid low-rank matrix recovery and seismic data reconstruction," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 4, pp. 658–671, June 2016.

[31] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, pp. 33–61, 1999.

[32] V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. S. Willsky, "The convex geometry of linear inverse problems," *Foundations of Computational Mathematics*, vol. 12, no. 6, pp. 805–849, 2012. [Online]. Available: http://dx.doi.org/10.1007/s10208-012-9135-7

[33] M. G. Krein and A. A. Nudelman, *The Markov moment problem and extremal problems: Ideas and problems of P. L. Chebyshev and A. A. Markov and their further development*, ser. Translations of Mathematical Monographs. American Mathematical Society, Providence, R.I., 1977, vol. 50, translated from the Russian by D. Louvish.

[34] P. Stoica and N. Arye, "Music, maximum likelihood, and cramer-rao bound," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 5, pp. 720–741, May 1989.

[35] C. Qian, L. Huang, N. Sidiropoulos, and H. C. So, "Enhanced puma for direction-of-arrival estimation and its performance analysis," *IEEE Transactions on Signal Processing*, vol. PP, no. 99, pp. 1–1, 2016.

[36] B. D. Rao and K. V. S. Hari, "Performance analysis of root-music," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 12, pp. 1939–1949, Dec 1989.

[37] Z. Yang, L. Xie, and C. Zhang, "A discretization-free sparse and parametric approach for linear array signal processing," *Signal Processing, IEEE Transactions on*, vol. 62, no. 19, pp. 4959–4973, Oct 2014.

[38] P. Stoica, P. Babu, and J. Li, "New method of sparse parameter estimation in separable models and its use for spectral analysis of irregularly sampled data," *IEEE Transactions on Signal Processing*, vol. 59, no. 1, pp. 35–47, Jan 2011.

[39] G. Tang, B. Bhaskar, P. Shah, and B. Recht, "Compressed sensing off the grid," *Information Theory, IEEE Transactions on*, vol. 59, no. 11, pp. 7465–7490, Nov 2013.

[40] E. J. Candès and C. Fernandez-Granda, "Towards a mathematical theory of super-resolution," *Communications on Pure and Applied Mathematics*, vol. 67, no. 6, pp. 906–956, 2014. [Online]. Available: http://dx.doi.org/10.1002/cpa.21455

[41] B. Bhaskar and B. Recht, "Atomic norm denoising with applications to line spectral estimation," in *Communication, Control, and Computing (Allerton), 2011 49th Annual Allerton Conference on*, Sept 2011, pp. 261–268.

[42] H. J. Landau and H. Widom, "Eigenvalue distribution of time and frequency limiting," *Journal of Mathematical Analysis and applications*, vol. Appl. 77, pp. 469–81, 1980.