Sampling Patterns for
Off-the-Grid Spectral Estimation

Maxime Ferreira Da Costa and Wei Dai

Department of Electrical and Electronic Engineering, Imperial College London, United Kingdom

Email: {maxime.ferreira, wei.dai1}@imperial.ac.uk

Abstract—Recent results have shown that it is possible to recover signals having sparse line spectra from few temporal observations via the use of convex programming. However, the computational cost of such approaches remains the major flaw to their use in practical system. We propose a highly scalable algorithm relying on partial observation of the signal in the time domain. A deterministic sub-sampling pattern is introduced and its benefit on the computational complexity of the recovery are highlighted. Moreover, specific sub-Nyquist recovery guarantees are derived for this pattern.

I. INTRODUCTION

A. Partial line spectral estimation

Spectral estimation is probably one of the most common, yet fundamental task in signal processing. In one of its generalized form, it consists in estimating the fundamental frequencies of a spectrally sparse signal from noisy time domain observations, acquired through linear combinations of the output of a uniform sampler.

More accurately, the sampled ground truth signal $u_\star = u(f_\star, c_\star) \in \mathbb{C}^n$ would read $u_{\star,j} = \sum_{r=1}^{s_\star} \alpha_r^\star e^{i2\pi f_r^\star j}$ for all $j \in [\![0, n-1]\!]$, where $f_\star = \{f_r^\star\}_{r=1}^{s_\star}$ is the set of its *reduced* spectral components to reconstruct, and $\alpha_\star$ is the set of associated complex amplitudes. The number of spectral components $s_\star$ is assumed to be unknown. Alternatively, denoting by $\mathcal{M}(\mathbb{T})$ the set of complex Radeon measures defined over the unidimensional torus $\mathbb{T} = [0, 1)$, one can express the ground truth signal as the image of a sparse measure $\mu_\star = \sum_{r=1}^{s_\star} \alpha_r^\star \delta_{f_r^\star} \in \mathcal{M}(\mathbb{T})$ through the linear map $u_\star = \mathcal{F}_{n,1}(\mu_\star) = \mathcal{F}_n(\mu_\star)$ whereby $\mathcal{F}_{n,F_s}$ is the adjoint of the discrete time Fourier operator in dimension n at sampling frequency F_s. As a result, introducing the linear operator $\mathcal{L}_\mathbf{M} = \mathbf{M}\mathcal{F}_n$, the observation vector $y \in \mathbb{C}^m$ writes

$$y = \mathcal{L}_\mathbf{M}(\mu_\star) + \eta, \tag{1}$$

whereby $\mathbf{M} \in \mathbb{C}^{m \times n}$ is a known fat measurement matrix, and $\eta \in \mathbb{C}^m$ represents to system noise.

We empathize on the fact that we are interested in a *gridless reconstruction* of the reduced frequencies, in the sense that the set of frequencies $f_\star$ is *drawn continuously* in the unidimensional torus $\mathbb{T}$ and is not restricted to belong to some arbitrary discrete grid in the Fourier domain. As a direct consequence, estimating the ground truth frequencies $f_\star$ from the sole finite-dimensional observation vector y is an ill-posed problem, since, even in absence of noise, the set of consistent spectral measures μ verifying the consistency

constraint $y = \mathcal{L}_\mathbf{M}(\mu)$ is an infinite-dimensional affine subspace of $\mathcal{M}(\mathbb{T})$. Among all those consistent measures, the one of interest, in the presented sparse recovery context, will be the one of smallest support. This *optimal estimator* μ_0 can be represented in the noisy settings as the output of the abstract regularized program

$$\mu_0 = \arg \min_{\mu \in \mathcal{M}(\mathbb{T})} \|\mu\|_0, \text{ s.t. } \|y - \mathcal{L}_\mathbf{M}(\mu)\|_2^2 \le \tau, \tag{2}$$

where $\|\cdot\|_0$ is the (potentially infinite) cardinality counting pseudo-norm, and $\tau \in \mathbb{R}^+$ is a regularization parameter that can be adjusted according to the noise level [1]. In practice, the output of Program (2) cannot be computed by a direct approach, due to the combinatorial nature of the cost function and the infinite dimensionally of the search space $\mathcal{M}(\mathbb{T})$. A common alternative to this problems consists in relaxing the original program into a *convex* total mass minimization program [2], [3] of the form

$$\mu_1 = \arg \min_{\mu \in \mathcal{M}(\mathbb{T})} |\mu|(\mathbb{T}), \text{ s.t. } \|y - \mathcal{L}_\mathbf{M}(\mu)\|_2^2 \le \tau, \tag{3}$$

whereby $|\cdot|(\mathbb{T})$ denotes the total mass of any complex Radeon measure and is defined by,

$$\forall \mu \in \mathcal{M}(\mathbb{T}), \ |\mu|(\mathbb{T}) = \int_\mathbb{T} |\mathrm{d}\mu|.$$

B. Background on line spectral estimation and inverse problems

The convex relaxation approach (3) was originally introduced in the context of line spectral estimation with full measurement ($\mathbf{M} = \mathbf{I}_n$) in [4]. It is shown that the relaxation is tight (i.e. $\mu_\star = \mu_1$) in absence of noise whenever the spectral components $f_\star$ composing the ground truth signal have a warp-around distance $\Delta_\mathbb{T}(f_\star) \ge \frac{4}{n-1}$, and under provisio that the number of observations n is large enough. This bound was tighten later on to $\frac{2.52}{n-1}$ under a stronger provisio [5]. On the other hand, it was recently shown in [6] that no algorithms can *robustly* recover a spectral support from *noisy observations* if $f_\star$ do not satisfies the *necessary* condition $\Delta_\mathbb{T}(f_\star) \ge \frac{2}{n-1}$, which comforts the early work of Slepian on the asymptotic properties of the time and band limiting operators [7].

High probability recovery guarantees were given for the *partial* line spectral estimation problem whenever the subsampling matrix $\mathbf{M}$ is picked at random [8] or verifies the incoherence property [9]. Line spectral estimation is a canonical example of sparse inverse problems defined over the set of measures.

We refer the interested reader to [2], [10] for more generic considerations and guarantees on sparse measure reconstruction.

C. Contribution

This paper is divided into two parts. In Section II, we present a scalable algorithm to recover the sparse spectrum from partial measurements. In Section III, we study a specific deterministic sub-sampling pattern and discuss its low-complexity and sub-nyquist recovery guarantee. We show, in particular, that applying the algorithm proposed in Section II results in order of magnitude changes on the computation complexity of this sampling scheme.

D. Notations

The adjunction of $\mathbf{X}$ is denoted $\mathbf{X}^*$, whenever $\mathbf{X}$ is a vector, a matrix, or a linear operator. The transposition of a matrix or a vector $\mathbf{X}$ is written $\mathbf{X}^\mathsf{T}$. Unless stated differently, vectors of $\mathbb{C}^n$ are indexed in $[\![0, n-1]\!]$ so that every vector $u \in \mathbb{C}^n$ writes $u = [u_0, \ldots, u_{n-1}]^\mathsf{T}$, and $e_k \in \mathbb{C}^n$ denotes the kth vector of the canonical basis (starting at 0). The space of square complex matrices and the one of Hermitian matrices of dimension n are respectively denoted $\mathbb{M}_n(\mathbb{C})$ and $\mathbb{S}_n(\mathbb{C})$. The cone of positive Hermitian matrices of same dimension is denoted $\mathbb{S}_n^+(\mathbb{C})$. Vectorial spaces of matrices are all endowed with the Frobenius inner product denoted $\langle \cdot, \cdot \rangle$ and defined by $\langle A, B \rangle = \mathrm{tr}(A^*B)$, where $\mathrm{tr}(\cdot)$ is the trace operator. We denote by $\boldsymbol{\Theta}_k^n$ the elementary Toeplitz matrix equal to 1 on the kth upper diagonal and zero elsewhere. For every matrix $\mathbf{M} \in \mathbb{C}^{m \times n}$, $m \leq n$, we denote by $\mathcal{R}_\mathbf{M}^* \in (\mathbb{M}_m(\mathbb{C}) \to \mathbb{C}^m)$ the linear operator characterized for every matrix $S \in \mathbb{M}_m(\mathbb{C})$ by $\mathcal{R}_\mathbf{M}^*(\mathbf{S}) = \sum_{k=0}^{n-1} \langle \mathbf{M}\boldsymbol{\Theta}_k\mathbf{M}, \mathbf{S} \rangle e_k$. Finally, a selection matrix $\mathbf{C}_\mathcal{I} \in \{0,1\}^{m \times n}$ for a subset $\mathcal{I} \subseteq [\![0, n-1]\!]$ of cardinality m is a boolean matrix whose rows are equal to $\{e_k^*, k \in \mathcal{I}\}$.

II. LOW COMPLEXITY SPECTRAL RECOVERY

A. Dimensionality reduction for the partial line spectral estimation problem

In our setting of interest, it is assumed that one disposes of *partial* observations gathered as linear combination of the output of a uniform sampler through a fat measurement matrix $\mathbf{M} \in \mathbb{C}^{m \times n}$. We recall a result from [11] attesting that the Lagrange dual problem for the partial line spectral estimation problem (3) takes the form of the semidefinite program (SDP) of dimension $m + 1$.

Theorem 1 (Dimensionality reduction). *If the sub-sampling matrix $\mathbf{M} \in \mathbb{C}^{m \times n}$ is such that $e_0 \in \mathrm{range}(\mathbf{M}^*)$ and $\mathbf{M}$ is full rank, the Lagrange dual problem of Problem (3) writes*

$$c_\star = \arg\max_{c \in \mathbb{C}^m} \Re(y^\mathsf{T}c) - \frac{\tau}{2} \|c\|_2^2 \qquad (4)$$

$$\text{subject to} \quad \begin{bmatrix} \mathbf{S} & c \\ c^* & 1 \end{bmatrix} \succeq 0$$

$$\mathcal{R}_\mathbf{M}^*(\mathbf{S}) = e_0.$$

Moreover, the frequencies can be directly estimated by the output $c_\star$ of Program (4) by defining $q_\star = M^* c_\star \in \mathbb{C}^n$

and locating whenever the polynomial $Q(z) = \sum_{k=0}^{n-1} q_k z^{-k}$ reaches one in modulus around the unit circle. Computing the solution of the semidefinite program (4) using out of the box solvers based on interior points methods will result in a *worst case* complexity of $\mathcal{O}(m^7)$ operations. This section aims to provide a lower complexity alternative to those computationally expensive methods.

In the rest of the paper, we focus on "keep or discard" subsampling patterns which are characterized by boolean subsampling matrices $\mathbf{C}_\mathcal{I}$ introduced in Section I-D. Using the particular structure of the matrix $\mathbf{C}_\mathcal{I}$, one can show the existence of a skew-symmetric partition of the square $[\![1, m]\!]^2$ into m subset $\{J_k\}_{k=1}^m$ such that the linear operator $\mathcal{R}_{\mathbf{C}_\mathcal{I}}^*$ can be decoupled into m independent linear forms

$$\forall \mathbf{S} \in \mathbb{M}_m(\mathbb{C}), \quad \mathcal{R}_{\mathbf{C}_\mathcal{I}}^*(\mathbf{S}) = \sum_{k=1}^m \left(\sum_{(l,r) \in J_k} \mathbf{S}_{l,r} \right) e_k. \quad (5)$$

B. Fast dual computation using ADMM

In the same spirit than in [12], we derive the steps and update equations to approach the optimal solution via ADMM. Unlike the original work, we choose to perform ADMM on the dual space instead of the primal one, and adjust the update steps in order to take advantage of the particular structure of $\mathcal{R}_{\mathbf{C}_\mathcal{I}}^*$. The overall idea of this algorithm is to cut the augmented Lagrangian of the problem into a sum of separable sub-functions. Each iteration consists in performing independent local minimization on each of those quantities. The interested reader can find a detailed survey of this method in [13]. Before any further analysis, we reformulate Program (4) into the equivalent formulation

$$c_\star = \arg\min_{c \in \mathbb{C}^m} -\Re(y^\mathsf{T}c) + \frac{\tau}{2} \|c\|_2^2 \qquad (6)$$

$$\text{subject to} \quad \mathbf{Z} \succeq 0$$

$$\mathbf{Z} = \begin{bmatrix} \mathbf{S} & c \\ c^* & 1 \end{bmatrix}$$

$$\sum_{(i,j) \in J_k} \mathbf{S}_{i,j} = \delta_k, \quad k \in [\![1, m]\!],$$

which is more friendly for the ADMM decomposition.

1) Lagrangian separability: We denote by L the restricted Lagrangian of Problem (6), obtained by ignoring the semidefinite constraint $\mathbf{Z} \succeq 0$, and introduce the augmented Lagrangian L_+ as follows

$$L_+(\mathbf{Z}, \mathbf{S}, c, \boldsymbol{\Lambda}, \mu) = L(\mathbf{Z}, \mathbf{S}, c, \boldsymbol{\Lambda}, \mu) + \frac{\rho}{2} \left\| \mathbf{Z} - \begin{bmatrix} \mathbf{S} & c \\ c^* & 1 \end{bmatrix} \right\|_F^2$$

$$+ \frac{\rho}{2} \sum_{k=1}^m \left(\sum_{(i,j) \in J_k} \mathbf{S}_{i,j} - \delta_k \right)^2,$$

whereby the variables $\boldsymbol{\Lambda} \in \mathbb{S}_{m+1}(\mathbb{C})$ and $\mu \in \mathbb{C}^m$ denote respectively the Lagrange multipliers associated with the first and the second equality constraints of Problem (6). The regularizing parameter $\rho > 0$ is set to ensure a well conditioned differentiability and to fasten the convergence speed of the

alternating minimization towards the global optimum of the cost function L_+. For clarity and convenience, the following decompositions of the parameters $\mathbf{Z}$ and $\boldsymbol{\Lambda}$ are introduced

$$\mathbf{Z} = \begin{bmatrix} \mathbf{Z}_0 & z \\ z^* & \zeta \end{bmatrix} \qquad \boldsymbol{\Lambda} = \begin{bmatrix} \boldsymbol{\Lambda}_0 & \lambda \\ \lambda^* & \eta \end{bmatrix}.$$

Moreover, for any square matrix $\mathbf{A} \in \mathbb{M}_m(\mathbb{C})$, we let by $\mathbf{A}_{J_k} \in \mathbb{C}^{|J_k|}$ the vector constituted of the terms $\{\mathbf{A}_{i,j},\ (i,j) \in J_k\}$. The order in which the elements of J_k are extracted and placed in this vector has no importance, as long as, once chosen, it remains the same for every matrix $\mathbf{A}$. This allows to decompose the augmented Lagrangian into

$$L_+ (\mathbf{Z}, \mathbf{S}, c, \boldsymbol{\Lambda}, \mu) = L_c(z, c, \lambda) + L_\gamma(\zeta, \eta)$$
$$+ \sum_{k=1}^{m} L_k(\mathbf{Z}_{0,J_k}, \mathbf{S}_{J_k}, \boldsymbol{\Lambda}_{0,J_k}),$$

whereby each of the sub-functions reads

$$L_c(z, c, \lambda) = -\Re(y^\mathsf{T} c) + \frac{\tau}{2} \|c\|_2^2 + 2\langle \lambda, z - c \rangle$$
$$+ \rho \|z - c\|_2^2$$
$$L_\gamma(\zeta, \eta) = \langle \eta, \zeta - 1 \rangle + \frac{\rho}{2}(\zeta - 1)^2$$
$$L_k(\mathbf{Z}_{0,J_k}, \mathbf{S}_{J_k}, \boldsymbol{\Lambda}_{0,J_k}) = \langle \boldsymbol{\Lambda}_{0,J_k}, \mathbf{Z}_{0,J_k} - \mathbf{S}_{J_k} \rangle$$
$$+ \mu_k \left(\sum_{(i,j) \in J_k} \mathbf{S}_{i,j} - \delta_k \right),$$
$$+ \frac{\rho}{2} \|\mathbf{Z}_{0,J_k} - \mathbf{S}_{J_k}\|_2^2$$
$$+ \frac{\rho}{2} \left(\sum_{(i,j) \in J_k} \mathbf{S}_{i,j} - \delta_k \right)^2.$$

2) Update rules: The ADMM will consist in successively performing the following decoupled update steps:

$$c^{t+1} \leftarrow \arg\min_c L_c(z^t, c, \lambda^t)$$
$$\forall k \in [\![1,m]\!], \quad \mathbf{S}_{J_k}^{t+1} \leftarrow \arg\min_{\mathbf{S}_{J_k}} L_k(\mathbf{Z}_{0,J_k}^t, \mathbf{S}_{J_k}, \boldsymbol{\Lambda}_{0,J_k}^t)$$
$$\mathbf{S}_{j,i}^{t+1} \leftarrow \overline{\mathbf{S}_{i,j}^{t+1}}, \quad \forall (i,j) \in \bigcup_{k=1}^{m} J_k$$
$$\mathbf{Z}^{t+1} \leftarrow \arg\min_{\mathbf{Z} \succeq 0} L_+(\mathbf{Z}, \mathbf{S}^{t+1}, c^{t+1}, \boldsymbol{\Lambda}^t, \mu^t)$$
$$\boldsymbol{\Lambda}^{t+1} \leftarrow \boldsymbol{\Lambda}^t + \rho \left(\mathbf{Z}^{t+1} - \begin{bmatrix} \mathbf{S}^{t+1} & c^{t+1} \\ c^{t+1*} & 1 \end{bmatrix} \right)$$
$$\forall k \in [\![1,m]\!], \quad \mu^{t+1}(k) \leftarrow \mu^t(k) + \rho \left(\sum_{(i,j) \in J_k} \mathbf{S}_{i,j}^{t+1} - \delta_k \right).$$

The third update step is necessary to maintain the Hermitian structure of the matrix $\mathbf{S}^{t+1}$ at every iteration. The update steps for the variables c^{t+1} and $\{\mathbf{S}_{J_k}^{t+1}\}_{k=1}^m$ admit closed form expressions given by

$$c^{t+1} = \frac{1}{2\rho + \tau} \left(\bar{y} + 2\rho z^t + 2\lambda^t \right)$$

$$\forall k \in [\![1,m]\!], \quad \mathbf{S}_{J_k}^{t+1} = \left(\mathbf{Z}_0^t + \frac{1}{\rho} \boldsymbol{\Lambda}_0^t \right)_{J_k}$$
$$- \left(\sum_{(i,j) \in J_k} \left(\mathbf{Z}_0^t + \frac{\boldsymbol{\Lambda}_0^t}{\rho} \right)_{i,j} - \left(\delta_k - \frac{\mu_k^t}{\rho} \right) \right) j_{|J_k|}$$

whereby $\bar{y} \in \mathbb{C}^m$ denotes the conjugate of the observation vector y, and j_v is the all-one vector of $\mathbb{C}^v$ for all $v \in \mathbb{N}$. The update $\mathbf{Z}^{t+1}$ at the tth iteration can be interpreted as an orthogonal projection of $\mathbf{Y}^t$ onto $\mathbb{S}_{m+1}^+(\mathbb{C})$ for the Frobenius inner product. This projection can be computed by looking for the eigenpairs of $\mathbf{Y}^t$, and setting all negative eigenvalues to 0. More precisely, denoting $\mathbf{Y}^t = \mathbf{V}^t \mathbf{D}^t \mathbf{V}^{t*}$ an eigen-decomposition of $\mathbf{Y}^t$, one get $\mathbf{Z}^{t+1} = \mathbf{V}^t \mathbf{D}_+^t \mathbf{V}^{t*}$ where $\mathbf{D}_+^t$ is a diagonal matrix whose j^{th} diagonal entry $d_+^t[j]$ satisfies $d_+^t[j] = \max\{d^t[j], 0\}$.

C. Computational complexity

On the computational point of view, at each step of ADMM, the update c^{t+1} is a vector addition and performed in a linear time $\mathcal{O}(m)$. On every extractions $\mathbf{S}_{J_k}^{t+1}$ of $\mathbf{S}^{t+1}$, the update equation is assimilated to a vector averaging requiring $\mathcal{O}(|J_k|)$ operations when firstly calculating the common second term of the addition. Since $\bigcup_{k=1}^m J_k = \frac{m(m+1)}{2}$, we conclude that the global update of the matrix $\mathbf{S}^{t+1}$ is done in $\mathcal{O}(m^2)$. The update of $\mathbf{Z}^{t+1}$ requires the computation of its spectrum, which can be done in $\mathcal{O}(m^3)$ via power method. Finally updating the multipliers $\boldsymbol{\Lambda}^{t+1}$ and μ^{t+1} consists in simple matrix and vector additions, thus of order $\mathcal{O}(m^2)$.

To summarize, the projection is the most costly operation of the loop. Each step of ADMM method runs in $\mathcal{O}(m^3)$ operations, which is a significant improvement compared to interior point methods requiring around $\mathcal{O}(m^7)$ operations.

III. APPLICATION TO MULTIRATE SAMPLING SYSTEMS

A. Sampling model

A multirate sampling system (MRSS) is defined by a set $\mathbb{A}$ of p distinct grids (or samplers) $\mathcal{A}_j$, $j \in [\![1,p]\!]$. Each grid is assimilated to a triplet $\mathcal{A}_j = (f_j, \gamma_j, n_j)$, where $f_j \in \mathbb{R}^+$ is its sampling frequency, $\gamma_j \in \mathbb{R}$ is its processing delay, expressed in sample unit for normalization purposes, and $n_j \in \mathbb{N}$ the number of measurements acquired by the grid. We assume those intrinsic characteristics to be known. The output $y_j \in \mathbb{C}^{n_j}$ of the grid $\mathcal{A}_j$ sampling a spectrally sparse with *real* frequency set $\xi_\star = \{\xi_r^\star\}_{r=1}^s$ signal reads

$$\forall k \in [\![0, n_j - 1]\!], \quad y_j[k] = \sum_{r=1}^{s_\star} \alpha_r e^{i2\pi \frac{\xi_r^\star}{f_j}(k-\gamma_j)}. \quad (7)$$

Applications of the MRSS framework are numerous in signal processing and appears to be a possible extension of the classic uniform sampling scheme. The MRSS framework is also naturally fitted to describe sampling processes in distributed sensor networks: each node, with limited processing capabilities, samples at its own rate, a delayed version of a

complex signal. Collected data are then sent and merged at a higher level processing unit, performing a global estimation of the spectral distribution on a joint manner.

The frequency estimation problem naturally consists in finding the sparsest spectral density that jointly matches the p observation vectors y_j for all $j \in [\![1,p]\!]$. Equivalently to (3), this problem can be relaxed into a convex minimization program over the set of complex measures *over the real line*

$$\mu_1 = \arg \min_{\mu \in \mathcal{M}(\mathbb{R})} |\mu|(\mathbb{R}), \text{ s.t. } \|y_j - \mathcal{L}_j(\mu)\|_2^2 \leq \tau_j, \forall j, \quad (8)$$

where $\mathcal{L}_j = \mathcal{F}_{n,f_j} \circ \mathcal{M}_{\frac{\gamma_j}{f_j}}$ and defining by $\mathcal{M}_\tau$, $\tau \in \mathbb{R}$ the temporal shift (or spectral modulation) operator defined for all $h \in \mathcal{M}(\mathbb{T})$ by $\mathcal{M}_\tau(h)(\xi) = e^{-i2\pi\tau\xi}h(\xi)$ for all $\xi \in \mathbb{R}$.

Finally, we denote by $m = \sum_{j=1}^p n_j$ the total number of measurements gathered by the arrays. Considering the vector $y = [y_1^\mathsf{T}, \ldots, y_p^\mathsf{T}]^\mathsf{T} \in \mathbb{C}^m$ obtained by concatenating the partial measurements, one can rewrite the linear constraint of Program (8) as $y = \mathcal{L}_\mathbb{A}(\mu)$ where $\mathcal{L}_\mathbb{A}$ is the linear operator admitting the partial operators $\{\mathcal{L}_j\}_{j \in [\![1,p]\!]}$ for restrictions on the p subspaces induced by the construction of the observation vector y.

B. Common grid expansion and SDP formulation

Theorem 1 states that the dual problem can take the form of a low dimensional SDP whenever the observation operator can be written under the form $\mathcal{L}_\mathbf{M} = \mathbf{M}\mathcal{F}_n$ for some admissible matrix $\mathbf{M} \in \mathbb{C}^{m \times n}$. This remarkable property is due to the polynomial nature of the adjoint measurement operator $\mathcal{L}_\mathbf{M}^*$. However, in the MRSS context, the dual observation operator defined by $\mathcal{L}_\mathbb{A}^*(c) = \sum_{j=1}^m \mathcal{L}_j^*(c_j)$ does not take such polynomial form in the general case. A direct calculation reveals that $\mathcal{L}_\mathbb{A}^*(c)$ is instead an exponential polynomial for all $c \in \mathbb{C}^m$. Up to our knowledge, there is no welcoming algebraic characterization for optimization purposes of the dual feasible in those settings. Therefore, Theorem 1 cannot be directly transcribed in the MRSS framework.

To bridge this concern, we restrict our analysis to the case where the consistency operator admits a factorization of the form $\mathcal{L}_\mathbb{A} = \mathcal{L}_\mathbf{M} = \mathbf{M}\mathcal{F}_n$ for some $n \in \mathbb{N}$ and $\mathbf{M} \in \mathbb{C}^{m \times n}$. The following aims to provide an algebraic criterion on the parameters $\{(f_j, \gamma_j, n_j)\}$ of $\mathbb{A}$ for this hypothesis to hold. We will see that this extra hypothesis consists in supposing that the samples acquired by $\mathbb{A}$ can by virtually aligned at a higher rate on another grid $\mathcal{A}_+$. Such grid will be called common supporting grid for $\mathbb{A}$, and are defined as follows.

Definition 2. A grid $\mathcal{A}_+ = (f_+, \gamma_+, n_+)$ is said to be a *common supporting grid* for a set of sampling grids $\mathbb{A} = \{\mathcal{A}_j\}_{j \in [\![1,p]\!]}$ if and only if the set of samples acquired by the MRSS induced by $\mathbb{A}$ is a subset of the one acquired by $\mathcal{A}_+$. In formal terms, the definition is equivalent to,

$$\left\{ \frac{1}{f_j}(k_j - \gamma_j), j \in [\![1,p]\!], k_j \in [\![0, n_j - 1]\!] \right\}$$
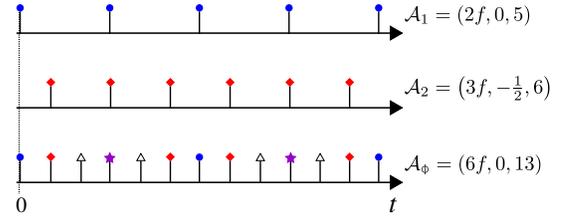$$\subseteq \left\{ \frac{1}{f_+}(k - \gamma_+), k \in [\![0, n_+ - 1]\!] \right\}. \quad (9)$$

Figure 1. A representation of a multirate sampling system $\mathbb{A}$ composed of two arrays $(\mathcal{A}_1, \mathcal{A}_2)$, and its associated minimal common grid $\mathcal{A}_\Phi$. Purple stars in the common grid correspond to time instant acquired multiple times by the system $\mathbb{A}$, and blank triangles to omitted samples. In this example, the dimension of the minimal common grid is $n_\Phi = 13$, The net number of observations of $\mathbb{A}$ is $m = 9$. Finally the equivalent observation set of the common grid is $\mathcal{I} = \{0, 1, 3, 5, 6, 7, 9, 11, 12\}$.

The set of common supporting grids of $\mathbb{A}$ is denoted by $\mathcal{C}(\mathbb{A})$. Moreover, a common supporting grid $\mathcal{A}_\Phi = (f_\Phi, \gamma_\Phi, n_\Phi)$ for $\mathbb{A}$ is said to be *minimal* if and only it satisfies the minimality condition, $\forall \mathcal{A}_+ \in \mathcal{C}(\mathbb{A})$, $n_\Phi \leq n_+$. Finally, the *equivalent observation set* of the minimal common grid $\mathcal{A}_\Phi$, denoted by $\mathcal{I}$, is the subset of $[\![0, n_\Phi - 1]\!]$ of cardinality m, formed by the k's for which the time instant $\frac{1}{f_\Phi}(k - \gamma_\Phi)$ is acquired by $\mathbb{A}$.

It is clear that if $\mathcal{C}(\mathbb{A})$ is not empty then the minimal common supporting grid for $\mathbb{A}$ exists and is unique. For ease of understanding, Figure 1 illustrates the notion of common supporting grid by showing a MRSS formed by two arrays and their minimal common grid. Proposition 3 states necessary and sufficient conditions in terms of the parameters of $\mathbb{A}$ such that the set $\mathcal{C}(\mathbb{A})$ is not empty. The interested reader is invited to refer to [11] for a complete proof of this proposition.

Proposition 3. *Given a set of p grids $\mathbb{A} = \{\mathcal{A}_j = (f_j, \gamma_j, n_j)\}_{j \in [\![1,p]\!]}$, the set $\mathcal{C}(\mathbb{A})$ is not empty if and only if there exist $f_+ \in \mathbb{R}^+$, $\gamma_+ \in \mathbb{R}$, a set of p positive integers $\{l_j\} \in \mathbb{N}^p$, and a set of p integers $\{a_j\} \in \mathbb{Z}^p$ satisfying $f_+ = l_j f_j$ and $\gamma_+ = l_j \gamma_j - a_j$ for all $j \in [\![1,p]\!]$. Moreover a common grid $\mathcal{A}_\Phi = (f_\Phi, \gamma_\Phi, n_\Phi)$ is minimal, if and only if*

$$\begin{cases} \gcd\left(\{a_j\}_{j \in [\![1,p]\!]} \cup \{l_j\}_{j \in [\![1,p]\!]}\right) = 1 \\ \gamma_\Phi = \max_{j \in [\![1,p]\!]}\{l_j \gamma_j\} \\ n_\Phi = \max_{j \in [\![1,p]\!]}\{l_j(n_j - 1) - a_j\}. \end{cases}$$

Although the conditions of Proposition 3 appear to be strong since one get $\mathcal{C}(\mathbb{A}) = \emptyset$ almost surely in the Lebesgue sense when the sampling frequencies and delays are drawn at random, assuming the existence of a common supporting grid for $\mathbb{A}$ is not meaningless in our context. By density, one can approximately align the system $\mathbb{A}$ on an arbitrary fine grid $\mathcal{A}_\varepsilon$, for any given maximal jitter $\varepsilon > 0$, and perform the proposed super-resolution on this common grid. The resulting error from this approximation can be interpreted as a "basis mismatch". The detailed analysis of this approach will not be covered in this work, however, similar approximations can be found in the literature for the analogue atomic norm minimization view of the super-resolution problem [1]. We claim that those results

extend in our settings and that the approximation error vanishes in the noiseless settings when going to the limit $\varepsilon \to 0$.

The next proposition concludes that the requested factorization of the linear observation operator $\mathcal{L}$ is possible whenever $\mathcal{C}(\mathbb{A}) \neq \emptyset$.

Proposition 4. *The set $\mathcal{C}(\mathbb{A})$ is not empty if and only if there exists a subset $\mathcal{I} \subseteq [0, n_\Phi - 1]$ of cardinality m such that $\mathcal{L}_\mathbb{A} = \mathbf{C}_\mathcal{I}\left(\mathcal{F}_{n_\Phi, f_\Phi} \circ \mathcal{M}_{\frac{\gamma_\Phi}{f_\Phi}}\right)$, whereby $\mathcal{A}_\Phi = (f_\Phi, \gamma_\Phi, n_\Phi)$ denotes the minimal grid of $\mathbb{A}$. Moreover Program (8) is equivalent to Program (3) for the consistency constraint $\mathcal{L}_\mathbb{A} = \mathcal{L}_{\mathbf{C}_\mathcal{I}}$.*

C. Dual certifiability and sub-Nyquist guarantees

In this section, sufficient conditions are presented to ensure the tightness of the relaxation (3) in absence of noise. In addition, mild conditions to ensure a sub-Nyquist recovery of the spectral spikes at a rate f_Φ from measurements taken at various lower rates $\{f_j\}_{j \in [\![1,p]\!]}$ are given. The proof of this result is available in [11] and relies on an extension of the polynomial construction methods presented in [4], [8].

Theorem 5. *Suppose that the sampled signal is noise free ($\eta = 0$), and let $\mathbb{A} = \{\mathcal{A}_j = (f_j, \gamma_j, n_j)\}_{j \in [\![1,p]\!]}$ be a set of sampling arrays. Suppose that $\mathcal{C}(\mathbb{A})$ is not empty, and denote by $\mathcal{A}_\Phi = (f_\Phi, \gamma_\Phi, n_\Phi)$ the minimal common supporting grid of $\mathbb{A}$. Assume that the system induced by $\mathbb{A}$ satisfies at least one of the two following separability conditions,*

- *Strong condition:*

$$\forall j \in [\![1,p]\!], \quad \Delta_\mathbb{T}\left(\frac{\xi_\star}{f_j}\right) \geq \frac{2.52}{n_j - 1} \text{ and } n_j > 2000,$$

- *Weak condition:*

$$\exists j \in [\![1,p]\!], \quad \Delta_\mathbb{T}\left(\frac{\xi_\star}{f_j}\right) \geq \frac{2.52}{n_j - 1} \text{ and } n_j > 2000$$
$$\text{and } m \geq (l_j + 1)\, s_\star,$$

then the output μ_1 of Program (3) is equal to $\mu_\star$.

The weak condition guarantees that the spectral components $\xi_\star$ can be recovered with an ambiguity modulo f_Φ when jointly resolving the MRSS, while individual estimations would guarantee to recover them with an ambiguity modulo $f_j \leq f_\Phi$. The weak condition require standard spectral separation from a single array $\mathcal{A}_j$, and sufficient measurements m of the time signal. The extra measurements $m - n_j$ corresponding to the other grids are *not uniformly aligned* with the sampler $\mathcal{A}_j$. Therefore the sampling system induced by $\mathbb{A}$ achieves sub-Nyquist spectral recovery of the spectral spikes, and pushes away the classic spectral range f_j from a factor $\frac{f_\Phi}{f_j} = l_j$. Nevertheless, this results relies on a construction of a dual polynomial certificate having a modulus close to unity on the aliasing frequencies induced by the zero forcing upscaling from f_j to f_Φ. Consequently, one can expect to obtain degraded performances in noisy environments when the sub-sampling factor l_j becomes large.

D. Complexity improvements

Recovering the frequencies in an MRSS using the original method proposed in [8] would requires to solve a SDP of dimension driven by the size n_Φ of the common grid. The actual value of n_Φ, fully determined of the observation pattern induced by $\mathbb{A}$, is given by $n_\Phi = \max_{j \in [\![1,p]\!]} \{l_j(n_j - 1) - a_j\}$, whereby the parameters $\{(a_j, l_j)\}_{j \in [\![1,p]\!]}$ are defined in Proposition 3. This is particularly disappointing since n_Φ grows at a speed driven by the product of the n_j's, whereas the *essential dimension* m of the problem is given by the number of net observations acquired by the grid $m = \sum_{j=1}^p n_j$. Suppose, for instance, a delay-only MRSS constituted of p grids $\mathcal{A}_j = \left(f, -\frac{1}{b_j}, n_0\right)$ sampling the time signal with the same frequency f and such that the delays $\{b_j\}_{j=1}^p$ forms a set of pairwise coprime integers. One would have $n_\Phi = \Omega(b^p n_0)$ while $m = pn_0$. Hence the ratio $\frac{m}{n_\Phi} = o\left(\frac{p}{b^p}\right)$ and tends to 0 exponentially fast with the number of samplers p of the system.

Under the common grid hypothesis, and using the equivalence guaranteed by Proposition 4, one could instead recover the frequency using the equivalent ADMM framework described in Section II which would requires $\mathcal{O}(m^3)$ operations. Hence the proposed method brings order of magnitude changes in the computational complexity of the line spectrum recovery, allowing a reconstruction of the spectral measure $\mu_\star$ up to *poly-logarithmic* orders in n_Φ.

REFERENCES

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

[2] V. Duval and G. Peyré, "Exact support recovery for sparse spikes deconvolution," *Foundations of Computational Mathematics*, vol. 15, no. 5, pp. 1315–1355, 2015.

[3] T. Bendory, S. Dekel, and A. Feuer, "Robust recovery of stream of pulses using convex optimization," *Journal of Mathematical Analysis and Applications*, vol. 442, no. 2, pp. 511 – 536, 2016.

[4] 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.

[5] C. Fernandez-Granda, "Super-resolution of point sources via convex programming," *Information and Inference: A Journal of the IMA*, vol. 5, no. 3, pp. 251–303, 2016.

[6] A. Moitra, "Super-resolution, Extremal Functions and the Condition Number of Vandermonde Matrices," Aug. 2014.

[7] D. Slepian, "Prolate spheroidal wave functions, fourier analysis, and uncertainty; V: the discrete case," *The Bell System Technical Journal*, vol. 57, no. 5, pp. 1371–1430, May 1978.

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

[9] R. Heckel and M. Soltanolkotabi, "Generalized line spectral estimation via convex optimization," *arXiv:1609.08198*, Sep. 2016.

[10] Y. De Castro and F. Gamboa, "Exact reconstruction using Beurling minimal extrapolation," *Journal of Mathematical Analysis and Applications*, vol. 395, no. 1, pp. 336–354, 2012.

[11] M. Ferreira Da Costa and W. Dai, "Exact dimensionality reduction for partial line spectra estimation problems," *arXiv:1609.03142*.

[12] G. Tang, B. N. Bhaskar, and B. Recht, "Near minimax line spectral estimation," *IEEE Transactions on Information Theory*, vol. 61, no. 1, pp. 499–512, Jan. 2015.

[13] 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, Jan. 2011.