ON GRIDLESS SPARSE METHODS FOR MULTI-SNAPSHOT DOA ESTIMATION

Zai Yang[*†], *Member, IEEE, and Lihua Xie*[*], *Fellow, IEEE*

[*]School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798
[†]School of Automation, Nanjing University of Science and Technology, Nanjing 210094, China

ABSTRACT

The authors have recently proposed two kinds of gridless sparse methods for direction of arrival (DOA) estimation that exploit joint sparsity among snapshots and completely resolve the grid mismatch issue of previous grid-based sparse methods. One is based on covariance fitting from a statistical perspective and termed as the gridless SPICE (GL-SPICE, GLS); the other uses deterministic atomic norm optimization which extends the recent super-resolution and continuous compressed sensing framework from the single to the multi-snapshot case. In this paper, we unify the two techniques by interpreting GLS as atomic norm methods in various scenarios. As a byproduct, we are able to provide theoretical guarantees of GLS for DOA estimation in the case of limited snapshots.

***Index Terms*—** Atomic norm, DOA estimation, gridless sparse methods, gridless SPICE (GLS).

1. INTRODUCTION

In array processing, we are interested in estimation of directions of a few narrow-band sources from observed snapshots of a sensor array. In particular, for an N-element uniform linear array (ULA) with sensors spaced by half a wavelength the snapshots can be modeled as follows (see, e.g., [1]):

$$\boldsymbol{y}(t)=\sum_{k=1}^{K}\boldsymbol{a}\left(f_k\right)s_k(t)+\boldsymbol{e}(t),\quad t=1,\ldots,L, \tag{1}$$

where t indexes the snapshot and L is the number of snapshots. Each snapshot $\boldsymbol{y}(t)\in\mathbb{C}^N$ is a noisy superposition of K sinusoidal signals $\{\boldsymbol{a}\left(f_k\right)s_k(t)\}$, where $\boldsymbol{a}\left(f_k\right)=\left[1,e^{i2\pi f_k},\ldots,e^{i2\pi(N-1)f_k}\right]^T\in\mathbb{C}^N$ has uniform samples of a sinusoidal signal with frequency $f_k\in[0,1)$ and $s_k(t)\in\mathbb{C}$ is the complex amplitude at time t. $\boldsymbol{e}(t)$ denotes the noise. Note that there exists a one-to-one mapping between the frequencies $\{f_k\}$ and the source directions. For notational simplicity, we only consider estimation of $\{f_k\}$ in this paper.

There are a whole host of methods for estimation of the frequencies in (1). A relatively recent subset is inspired by the literature on sparse representation and compressed sensing (CS) and usually designated as "sparse methods" (see, e.g., [2, 3]). In these methods, however, the frequency estimates have to be confined in a set of grid points such that the observed snapshots can be sparsely represented under a finite discrete dictionary—a prerequisite for carrying out sparse recovery according to conventional wisdom. An estimation bias is thus induced by grid mismatch that becomes statistically significant when N or L or the signal-to-noise ratio (SNR) increases. To mitigate the problem several solutions have been proposed (see, e.g., [4, 5]).

The authors have recently proposed two kinds of *gridless sparse methods*, in which the frequencies are treated as continuous (as opposed to discretized/gridded) parameters, and completely resolve the grid mismatch issue of the existing *grid-based sparse methods* [6–10]. The first technique is termed as the gridless SPICE (GL-SPICE, GLS) [6, 7] [also referred to as the sparse and parametric approach (SPA)], that is from a statistical perspective and uses the covariance fitting criterion of the SPICE method [11]. The other uses deterministic atomic norm optimization inspired by the recent super-resolution and continuous CS framework introduced in [12, 13] for line spectral estimation (a.k.a. DOA estimation with a single snapshot). The papers [8,9] extend the atomic norm methods and theoretical results in [12, 13] to the multi-snapshot case by exploiting joint sparsity among the snapshots. A reweighted atomic norm method was also proposed in [10] for further enhancing sparsity and resolution. Note that both GLS and the atomic norm methods are applicable to the sparse linear array (SLA) case that corresponds to DOA estimation from partial entries of $\boldsymbol{y}(t)$ in (1).

Note that several existing results on gridless sparse methods have been focused on the single snapshot case (see, e.g., [14–16]). In the case of multiple snapshots, the atomic norm method as introduced in [8, 9] was also independently proposed in [17]. A different atomic norm method was presented in [18] which, however, is within the existing single snapshot framework and does not exploit the joint sparsity.

Motivated by the observation that GLS and the atomic norm methods have seemingly related SDP formulations, we investigate their relationship in this paper. In both cases of ULA and SLA we show that GLS is equivalent to atomic norm methods in various scenarios and therefore unify the two kinds of gridless sparse methods. As a byproduct, we provide theoretical guarantees of GLS for DOA estimation in the case of limited snapshots based on existing results on atomic norm methods.

Notations used in this paper are as follows. $\mathbb{R}$, $\mathbb{R}_+$ and $\mathbb{C}$ denote the sets of real numbers, nonnegative real numbers and complex numbers, respectively. Boldface letters are reserved for vectors and matrices. $\left\|\cdot\right\|_1$, $\left\|\cdot\right\|_2$ and $\left\|\cdot\right\|_{\text{F}}$ denote the ℓ_1, ℓ_2 and Frobenius norms, respectively. $\cdot^T$ and $\cdot^H$ are the matrix transpose and conjugate transpose. x_j is the jth entry of a vector $\boldsymbol{x}$, $\boldsymbol{A}_j$ is the jth row of a matrix $\boldsymbol{A}$, and $\boldsymbol{A}_{jk}$ is the jkth entry of $\boldsymbol{A}$. Unless otherwise stated, $\boldsymbol{A}_{\boldsymbol{\Omega}}$ is a submatrix of $\boldsymbol{A}$ and is composed of the rows indexed by a set $\boldsymbol{\Omega}$. For a vector $\boldsymbol{x}$, $\text{diag}\left(\boldsymbol{x}\right)$ is a diagonal matrix with $\boldsymbol{x}$ being its diagonal. $\boldsymbol{x}\succeq\boldsymbol{0}$ means $x_j\geq 0$ for all j. $\text{tr}\left(\cdot\right)$ denotes the matrix trace. $E\left[\cdot\right]$ denotes expectation. $\boldsymbol{A}\geq\boldsymbol{B}$ means that $\boldsymbol{A}-\boldsymbol{B}$ is positive semidefinite.

The research of the project was supported by Ministry of Education, Republic of Singapore, under grant AcRF TIER 1 RG78/15.

2. PRELIMINARY RESULTS

2.1. GLS

GLS estimates the frequencies via covariance fitting. Let $\boldsymbol{f} = [f_1, \ldots, f_K]^T \in [0,1)^K$, $\boldsymbol{A}(\boldsymbol{f}) = [\boldsymbol{a}(f_1), \ldots, \boldsymbol{a}(f_K)]$ and $\boldsymbol{Y} = [\boldsymbol{y}(1), \ldots, \boldsymbol{y}(L)]$. The data covariance is given by

$$\boldsymbol{R} = E\left\{\boldsymbol{y}(t)\boldsymbol{y}^H(t)\right\} = \boldsymbol{A}(\boldsymbol{f})\operatorname{diag}(\boldsymbol{p})\boldsymbol{A}^H(\boldsymbol{f}) + \operatorname{diag}(\boldsymbol{\sigma}) \quad (2)$$

under the assumption of uncorrelated sources, where the entries of $\boldsymbol{p} \in \mathbb{R}_+^K$ are powers of the sources and the entries of $\boldsymbol{\sigma} \in \mathbb{R}_+^N$, which are not necessarily equal, are noise variances. Denote the sample covariance $\widetilde{\boldsymbol{R}} = \frac{1}{L}\boldsymbol{Y}\boldsymbol{Y}^H$. According to whether $\widetilde{\boldsymbol{R}}^{-1}$ exists, GLS minimizes the covariance fitting criterion of the grid-based SPICE method [11, 19]:

$$h_1(\boldsymbol{\theta}, \boldsymbol{p}, \boldsymbol{\sigma}) = \left\|\boldsymbol{R}^{-\frac{1}{2}}\left(\widetilde{\boldsymbol{R}} - \boldsymbol{R}\right)\right\|_{\mathrm{F}}^2 \quad (3)$$

or

$$h_2(\boldsymbol{\theta}, \boldsymbol{p}, \boldsymbol{\sigma}) = \left\|\boldsymbol{R}^{-\frac{1}{2}}\left(\widetilde{\boldsymbol{R}} - \boldsymbol{R}\right)\widetilde{\boldsymbol{R}}^{-\frac{1}{2}}\right\|_{\mathrm{F}}^2. \quad (4)$$

The key step of GLS is to reparameterize $\boldsymbol{R}$ in a linear way and then rewrite the minimization of (3) or (4) as an SDP. In particular, there exists a (Hermitian) Toeplitz matrix $T(\boldsymbol{u})$ formed by $\boldsymbol{u} \in \mathbb{C}^N$ such that

$$T(\boldsymbol{u}) = \boldsymbol{A}(\boldsymbol{f})\operatorname{diag}(\boldsymbol{p})\boldsymbol{A}^H(\boldsymbol{f}). \quad (5)$$

The equation in (5) is known as the Vandermonde decomposition in which there exists a one-to-one mapping between $(\boldsymbol{f}, \boldsymbol{p})$ and $\boldsymbol{u}$ whenever $K \leq N-1$ and $T(\boldsymbol{u}) \geq \boldsymbol{0}$ (see, e.g., [1]). Therefore,

$$\boldsymbol{R} = T(\boldsymbol{u}) + \operatorname{diag}(\boldsymbol{\sigma}) \quad (6)$$

becomes a linear function of $(\boldsymbol{u}, \boldsymbol{\sigma})$. As a result, the GLS minimization problem with

$$h_1(\boldsymbol{\theta}, \boldsymbol{p}, \boldsymbol{\sigma}) = \operatorname{tr}\left(\widetilde{\boldsymbol{R}}\boldsymbol{R}^{-1}\widetilde{\boldsymbol{R}}\right) + \operatorname{tr}(\boldsymbol{R}) - 2\operatorname{tr}\left(\widetilde{\boldsymbol{R}}\right) \quad (7)$$

can be formulated as the following SDP:

$$\begin{aligned} &\min_{\boldsymbol{X}, \boldsymbol{u}, \{\boldsymbol{\sigma} \succeq \boldsymbol{0}\}} \operatorname{tr}(\boldsymbol{X}) + \operatorname{tr}(\boldsymbol{R}), \\ &\text{subject to } \begin{bmatrix} \boldsymbol{X} & \widetilde{\boldsymbol{R}} \\ \widetilde{\boldsymbol{R}} & \boldsymbol{R} \end{bmatrix} \geq \boldsymbol{0},\ T(\boldsymbol{u}) \geq \boldsymbol{0}, \end{aligned} \quad (8)$$

where $\boldsymbol{R}$ is given in (6). Once the SDP is solved, for example, using SDPT3 [20], the estimates of $\boldsymbol{f}$ and $\boldsymbol{p}$ can be retrieved based on the decomposition in (5) (see details in [6]). Like its grid-based version GLS does not require the noise level and is totally hyper-parameter free. It is noted that GLS can also deal with the SLA case (see Section 4 and [6]).

2.2. Atomic Norm Methods

For a multi-snapshot signal $\boldsymbol{Z} \in \mathbb{C}^{N \times L}$ the atomic norm is defined as [8, 9]

$$\begin{aligned} \|\boldsymbol{Z}\|_{\mathcal{A}} &\triangleq \inf_{f_k, \boldsymbol{s}_k}\left\{\sum_k \|\boldsymbol{s}_k\|_2 :\ \boldsymbol{Z} = \sum_k \boldsymbol{a}(f_k)\boldsymbol{s}_k\right\} \\ &= \min_{\boldsymbol{u}, \boldsymbol{X}} \frac{1}{2\sqrt{N}}\left[\operatorname{tr}(\boldsymbol{X}) + \operatorname{tr}(T(\boldsymbol{u}))\right], \\ &\quad \text{subject to } \begin{bmatrix} \boldsymbol{X} & \boldsymbol{Z}^H \\ \boldsymbol{Z} & T(\boldsymbol{u}) \end{bmatrix} \geq \boldsymbol{0}, \end{aligned} \quad (9)$$

where $\boldsymbol{s}_k \triangleq [s_k(1), \ldots, s_k(L)] \in \mathbb{C}^{1 \times L}$ is a row vector. The multi-snapshot atomic norm in (9) is an extension of the atomic norm (or total variation norm) in [12, 13] specified for a single snapshot. The frequencies are encoded in $T(\boldsymbol{u})$ and can be obtained based on the Vandermonde decomposition as in GLS. It is shown in [9, 12, 13] that if $\boldsymbol{Z}$ is composed of a few frequency components that are sufficiently separate, then the frequencies can be exactly recovered by computing $\|\boldsymbol{Z}\|_{\mathcal{A}}$ (or by minimizing $\|\boldsymbol{Z}\|_{\mathcal{A}}$ if only part of the rows of $\boldsymbol{Z}$ are observed).

To overcome the resolution limit of atomic norm for frequency estimation, the concept of *weighted atomic norm* is introduced in [10] by adaptively enhancing resolution. Given a weighting function $w(f) = \left(\frac{1}{N}\boldsymbol{a}^H(f)\boldsymbol{W}\boldsymbol{a}(f)\right)^{-\frac{1}{2}} \geq 0$ with $\boldsymbol{W} \in \mathbb{C}^{N \times N}$, which is used to specify the preference of the frequencies, the weighted atomic norm is defined as

$$\begin{aligned} \|\boldsymbol{Z}\|_{\mathcal{A}^w} &\triangleq \inf_{f_k, \boldsymbol{s}_k}\left\{\sum_k \frac{\|\boldsymbol{s}_k\|_2}{w(f_k)} : \boldsymbol{Z} = \sum_k \boldsymbol{a}(f_k)\boldsymbol{s}_k\right\} \\ &= \min_{\boldsymbol{u}, \boldsymbol{X}} \frac{1}{2\sqrt{N}}\left[\operatorname{tr}(\boldsymbol{X}) + \operatorname{tr}(\boldsymbol{W}T(\boldsymbol{u}))\right], \\ &\quad \text{subject to } \begin{bmatrix} \boldsymbol{X} & \boldsymbol{Z}^H \\ \boldsymbol{Z} & T(\boldsymbol{u}) \end{bmatrix} \geq \boldsymbol{0}. \end{aligned} \quad (10)$$

The atomic norm in (9) is obviously a special case of the weighted atomic norm in (10) with a constant weighting function (e.g., when $\boldsymbol{W}$ is an identity matrix).

The atomic norm methods in [8–10] carry out deterministic optimization which seeks a sparse candidate signal, measured by the (weighted) atomic norm, over a feasible domain defined by the measurements. As an example, in the ULA case the feasible domain is defined by $\left\{\boldsymbol{Z} : \|\boldsymbol{Z} - \boldsymbol{Y}\|_{\mathrm{F}}^2 \leq \eta\right\}$, where $\boldsymbol{Y}$ consists of sampled data and η is an upper bound of the noise energy. Therefore, unlike GLS the atomic norm methods require the noise level. Motivated by the observation that GLS and the (weighted) atomic norm have seemingly related SDP formulations, we explore their relationship in the ensuing two sections.

Remark 1 *GLS and the existing atomic norm methods are complementing each other rather than competing. In particular, GLS is preferable in the absence of the noise level or in the presence of heteroscedastic noise, whereas the atomic norm methods are favorite when the noise has known bounded energy or when the sources are highly correlated (note that GLS is derived based on the assumption of uncorrelated sources).*

3. GLS AS ATOMIC NORM METHODS: THE ULA CASE

3.1. Homoscedastic Noise and $L < N$

We first consider the case of homoscedastic noise where $\{\sigma_n\}$ are identical. According to [6] GLS can be simplified in this case since $\boldsymbol{R}$ in (6) is a Toeplitz matrix itself and can be reparameterized as

$$\boldsymbol{R} = T(\boldsymbol{u}) \quad (11)$$

for some $\boldsymbol{u} \in \mathbb{C}^N$ which is identical to the $\boldsymbol{u}$ in (6) except for the first entry. When $L < N$ the sample covariance $\widetilde{\boldsymbol{R}}$ is singular and the criterion h_1 is adopted. Therefore, the GLS minimization problem is given by (8):

$$\min_{\boldsymbol{X}, \boldsymbol{u}} \operatorname{tr}(\boldsymbol{X}) + \operatorname{tr}(T(\boldsymbol{u})),\ \text{subject to } \begin{bmatrix} \boldsymbol{X} & \widetilde{\boldsymbol{R}} \\ \widetilde{\boldsymbol{R}} & T(\boldsymbol{u}) \end{bmatrix} \geq \boldsymbol{0} \quad (12)$$

which exactly computes $\left\|\widetilde{\boldsymbol{R}}\right\|_{\mathcal{A}}$ (up to a scaling factor) by (9). Note that the dimension of the problem in (12) can be reduced. Let $\widetilde{\boldsymbol{Y}} = L^{-1}\boldsymbol{Y}\left(\boldsymbol{Y}^H\boldsymbol{Y}\right)^{\frac{1}{2}} \in \mathbb{C}^{N\times L}$. It follows that $\widetilde{\boldsymbol{R}}^2 = \widetilde{\boldsymbol{Y}}\widetilde{\boldsymbol{Y}}^H$. Hence, GLS is also equivalent to computing $\left\|\widetilde{\boldsymbol{Y}}\right\|_{\mathcal{A}}$ (recall (7)). Moreover, GLS exactly computes $\|\boldsymbol{Y}\|_{\mathcal{A}}$ in the single-snapshot case where $\boldsymbol{Y}^H\boldsymbol{Y}$ is a scalar.

To explain why GLS can estimate the frequencies, we consider the limiting noiseless case where $\boldsymbol{Y} = \sum_{k=1}^K \boldsymbol{a}(f_k)\boldsymbol{s}_k$. Then we have that

$$\widetilde{\boldsymbol{Y}} = \sum_{k=1}^{K} \boldsymbol{a}(f_k)\widetilde{\boldsymbol{s}}_k \tag{13}$$

and $\widetilde{\boldsymbol{s}}_k = L^{-1}\boldsymbol{s}_k\left(\boldsymbol{Y}^H\boldsymbol{Y}\right)^{\frac{1}{2}}$. Therefore, GLS modifies the source signals from $\{\boldsymbol{s}_k\}$ to $\{\widetilde{\boldsymbol{s}}_k\}$ and then computes the atomic norm. According to [9] we conclude that *GLS with a ULA estimates the frequencies with infinite precision in the limiting noiseless case when $L < N$ if the frequencies are mutually separated by at least $\frac{4}{N}$* (note that the frequency separation condition is sufficient but not necessary).

3.2. Homoscedastic Noise and $L \geq N$

When $L \geq N$ the criterion h_2 is adopted in GLS. By (11) and the identity that

$$h_2(\boldsymbol{f}, \boldsymbol{p}, \boldsymbol{\sigma}) = \operatorname{tr}\left(\boldsymbol{R}^{-1}\widetilde{\boldsymbol{R}}\right) + \operatorname{tr}\left(\widetilde{\boldsymbol{R}}^{-1}\boldsymbol{R}\right) - 2N \tag{14}$$

the GLS minimization problem is given by

$$\min_{\boldsymbol{u}} \operatorname{tr}\left(L^{-1}\boldsymbol{Y}^H T(\boldsymbol{u})^{-1}\boldsymbol{Y}\right) + \operatorname{tr}\left(\widetilde{\boldsymbol{R}}^{-1}T(\boldsymbol{u})\right) \tag{15}$$

with $T(\boldsymbol{u}) \geq \boldsymbol{0}$. According to (10), GLS is equivalent to computing $\|\boldsymbol{Y}\|_{\mathcal{A}^w}$ up to a scaling factor, where the weighting function of the weighted atomic norm $w(f) = \left(\frac{1}{N}\boldsymbol{a}^H(f)\widetilde{\boldsymbol{R}}^{-1}\boldsymbol{a}(f)\right)^{-\frac{1}{2}}$ is the square root of Capon's spectrum (see, e.g., [1]) which is a reasonable choice to specify the preference of the frequencies.

3.3. Heteroscedastic Noise and $L < N$

In the case of heteroscedastic noise where $\{\sigma_n\}$ are distinct, $\boldsymbol{R}$ is given by (6). When $L < N$ the GLS minimization problem is given by

$$\min_{\boldsymbol{u}, \{\boldsymbol{\sigma} \succeq \boldsymbol{0}\}} \operatorname{tr}\left(\widetilde{\boldsymbol{R}}\boldsymbol{R}^{-1}\widetilde{\boldsymbol{R}}\right) + \operatorname{tr}(\boldsymbol{R}), \text{ subject to } T(\boldsymbol{u}) \geq \boldsymbol{0}. \tag{16}$$

To relate the problem above to the atomic norm, we use the following identity (see, e.g., [7]):

$$\begin{aligned} &\boldsymbol{y}^H\boldsymbol{R}^{-1}\boldsymbol{y} \\ &= \min_{\boldsymbol{z}} \boldsymbol{z}^H T(\boldsymbol{u})^{-1}\boldsymbol{z} + (\boldsymbol{y}-\boldsymbol{z})^H \operatorname{diag}(\boldsymbol{\sigma})^{-1}(\boldsymbol{y}-\boldsymbol{z}). \end{aligned} \tag{17}$$

It follows that (the constraints $T(\boldsymbol{u}) \geq \boldsymbol{0}$ and $\boldsymbol{\sigma} \succeq \boldsymbol{0}$ are omitted for brevity)

$$\begin{aligned} (16) \Leftrightarrow & \min_{\boldsymbol{u},\boldsymbol{\sigma}} \operatorname{tr}\left(\widetilde{\boldsymbol{Y}}^H\boldsymbol{R}^{-1}\widetilde{\boldsymbol{Y}}\right) + \operatorname{tr}(\boldsymbol{R}) \\ \Leftrightarrow & \min_{\boldsymbol{u},\boldsymbol{\sigma},\boldsymbol{Z}} \operatorname{tr}\left(\boldsymbol{Z}^H T(\boldsymbol{u})^{-1}\boldsymbol{Z}\right) + \sum_{n=1}^N \frac{1}{\sigma_n}\left\|\left(\widetilde{\boldsymbol{Y}} - \boldsymbol{Z}\right)_n\right\|_2^2 \\ & + \operatorname{tr}(T(\boldsymbol{u})) + \sum_{n=1}^N \sigma_n \\ \Leftrightarrow & \min_{\boldsymbol{Z}} \sqrt{N}\|\boldsymbol{Z}\|_{\mathcal{A}} + \left\|\widetilde{\boldsymbol{Y}} - \boldsymbol{Z}\right\|_{2,1}, \end{aligned} \tag{18}$$

where $\left\|\widetilde{\boldsymbol{Y}} - \boldsymbol{Z}\right\|_{2,1} = \sum_{n=1}^N \left\|\left(\widetilde{\boldsymbol{Y}} - \boldsymbol{Z}\right)_n\right\|_2$. As a result, the GLS minimization problem is interpreted as an *atomic norm denoising* problem in which $\widetilde{\boldsymbol{Y}}$ consists of the modified snapshots and the $\ell_{2,1}$ norm is used for data fitting.

3.4. Heteroscedastic Noise and $L \geq N$

In this case GLS minimizes h_2 in (14) and $\boldsymbol{R}$ is given by (6). Then the GLS minimization problem is equivalent to

$$\begin{aligned} & \min_{\boldsymbol{u},\boldsymbol{\sigma}} \operatorname{tr}\left(L^{-1}\boldsymbol{Y}^H\boldsymbol{R}^{-1}\boldsymbol{Y}\right) + \operatorname{tr}\left(\widetilde{\boldsymbol{R}}^{-1}\boldsymbol{R}\right) \\ \Leftrightarrow & \min_{\boldsymbol{u},\boldsymbol{\sigma},\boldsymbol{Z}} \operatorname{tr}\left(\boldsymbol{Z}^H T(\boldsymbol{u})^{-1}\boldsymbol{Z}\right) + \sum_{n=1}^N \frac{1}{\sigma_n}\left\|\left(\frac{1}{\sqrt{L}}\boldsymbol{Y} - \boldsymbol{Z}\right)_n\right\|_2^2 \\ & + \operatorname{tr}\left(\widetilde{\boldsymbol{R}}^{-1}T(\boldsymbol{u})\right) + \sum_{n=1}^N \sigma_n\left(\widetilde{\boldsymbol{R}}^{-1}\right)_{nn} \\ \Leftrightarrow & \min_{\boldsymbol{Z}} \sqrt{N}\|\boldsymbol{Z}\|_{\mathcal{A}^w} + \sum_{n=1}^N \sqrt{\left(\widetilde{\boldsymbol{R}}^{-1}\right)_{nn}}\left\|\left(\frac{1}{\sqrt{L}}\boldsymbol{Y} - \boldsymbol{Z}\right)_n\right\|_2, \end{aligned} \tag{19}$$

where $w(f)$ remains the square root of Capon's spectrum and $\left(\widetilde{\boldsymbol{R}}^{-1}\right)_{nn}$ is known as an estimate of σ_n^{-1}. Therefore, GLS is interpreted as a *weighted atomic norm denoising* problem. It is worth noting that $\frac{1}{\sqrt{L}}\boldsymbol{Y}$ in (19) can be replaced by $\widetilde{\boldsymbol{R}}^{\frac{1}{2}}$ for reducing the problem dimension (recall (14)). It is similar in the case of homoscedastic noise in Section 3.2.

4. GLS AS ATOMIC NORM METHODS: THE SLA CASE

In the SLA case, we denote by $\boldsymbol{\Omega} \subset \{1, 2, \dots, N\}$ the set of indices of the sensors and let M be the number of sensors (cardinality of $\boldsymbol{\Omega}$). Then only the rows of $\boldsymbol{Y}$ indexed by $\boldsymbol{\Omega}$ are observed and comprises $\boldsymbol{Y}_{\boldsymbol{\Omega}} \in \mathbb{C}^{M\times L}$. It follows that $\boldsymbol{Y}_{\boldsymbol{\Omega}} = \boldsymbol{\Gamma}_{\boldsymbol{\Omega}}\boldsymbol{Y}$, where $\boldsymbol{\Gamma} \in \{0,1\}^{M\times N}$ has 1's only at the (j, Ω_j)th entry, $j = 1, \dots, M$. Moreover, we let $\widetilde{\boldsymbol{R}}_{\boldsymbol{\Omega}} = \frac{1}{L}\boldsymbol{Y}_{\boldsymbol{\Omega}}\boldsymbol{Y}_{\boldsymbol{\Omega}}^H = \boldsymbol{\Gamma}_{\boldsymbol{\Omega}}\widetilde{\boldsymbol{R}}\boldsymbol{\Gamma}_{\boldsymbol{\Omega}}^T$ and $\boldsymbol{R}_{\boldsymbol{\Omega}} = \boldsymbol{\Gamma}_{\boldsymbol{\Omega}}\boldsymbol{R}\boldsymbol{\Gamma}_{\boldsymbol{\Omega}}^T$ be the sample covariance and the data covariance, respectively.

4.1. Homoscedastic Noise and $L < M$

In this case, $\boldsymbol{R}$ is given by (11) and the GLS minimization problem based on the criterion h_1 is given by

$$\min_{\boldsymbol{u}} \operatorname{tr}\left(\widetilde{\boldsymbol{R}}_{\boldsymbol{\Omega}}\boldsymbol{R}_{\boldsymbol{\Omega}}^{-1}\widetilde{\boldsymbol{R}}_{\boldsymbol{\Omega}}\right) + \operatorname{tr}(\boldsymbol{R}_{\boldsymbol{\Omega}}). \tag{20}$$

First note that $\text{tr}\left(\boldsymbol{R}_{\boldsymbol{\Omega}}\right)=\frac{M}{N}\text{tr}\left(T\left(\boldsymbol{u}\right)\right)$. Let $\boldsymbol{B}=L^{-1}\boldsymbol{Y}_{\boldsymbol{\Omega}}\left(\boldsymbol{Y}_{\boldsymbol{\Omega}}^H\boldsymbol{Y}_{\boldsymbol{\Omega}}\right)^{\frac{1}{2}}\in\mathbb{C}^{M\times L}$. It follows that $\widetilde{\boldsymbol{R}}_{\boldsymbol{\Omega}}^2=\boldsymbol{B}\boldsymbol{B}^H$. We also use the following identity (see, e.g., [7]):

$$\boldsymbol{b}^H\boldsymbol{R}_{\boldsymbol{\Omega}}^{-1}\boldsymbol{b}=\min_{\boldsymbol{z}}\boldsymbol{z}^H\boldsymbol{R}^{-1}\boldsymbol{z},\ \text{subject to } \boldsymbol{z}_{\boldsymbol{\Omega}}=\boldsymbol{b}. \tag{21}$$

Now we are ready to show the following equivalences:

$$\begin{aligned}(20)&\Leftrightarrow\min_{\boldsymbol{u}}\text{tr}\left(\boldsymbol{B}^H\boldsymbol{R}_{\boldsymbol{\Omega}}^{-1}\boldsymbol{B}\right)+\frac{M}{N}\text{tr}\left(T\left(\boldsymbol{u}\right)\right)\\&\Leftrightarrow\min_{\boldsymbol{u},\boldsymbol{Z}}\text{tr}\left(\boldsymbol{Z}^HT\left(\boldsymbol{u}\right)^{-1}\boldsymbol{Z}\right)+\frac{M}{N}\text{tr}\left(T\left(\boldsymbol{u}\right)\right),\\&\qquad\text{subject to } \boldsymbol{Z}_{\boldsymbol{\Omega}}=\boldsymbol{B}\\&\Leftrightarrow\min_{\boldsymbol{Z}}\left\|\boldsymbol{Z}\right\|_{\mathcal{A}},\ \text{subject to } \boldsymbol{Z}_{\boldsymbol{\Omega}}=\boldsymbol{B}.\end{aligned}\tag{22}$$

Therefore, the GLS minimization problem is equivalent to an *atomic norm minimization* problem studied in [9]. Similarly to the ULA case, in the limiting noiseless case we have that $\boldsymbol{B}=\sum_{k=1}^K\boldsymbol{a}_{\boldsymbol{\Omega}}\left(f_k\right)\widetilde{\boldsymbol{s}}_k$, where $\widetilde{\boldsymbol{s}}_k=L^{-1}\boldsymbol{s}_k\left(\boldsymbol{Y}_{\boldsymbol{\Omega}}^H\boldsymbol{Y}_{\boldsymbol{\Omega}}\right)^{\frac{1}{2}}$. In general, we cannot conclude as in [9] that the frequencies can be exactly recovered by GLS since $\left(\boldsymbol{Y}_{\boldsymbol{\Omega}}^H\boldsymbol{Y}_{\boldsymbol{\Omega}}\right)^{\frac{1}{2}}$ depends on $\{\boldsymbol{s}_k\}$ and that the technical assumption that the signs of $\{\widetilde{\boldsymbol{s}}_k\}$ are independent does not hold. But the conclusion indeed holds in the single-snapshot case where $\left(\boldsymbol{Y}_{\boldsymbol{\Omega}}^H\boldsymbol{Y}_{\boldsymbol{\Omega}}\right)^{\frac{1}{2}}$ is a positive scalar. Therefore, *GLS with an SLA exactly estimates the frequencies with high probability in the limiting noiseless case when $L=1$ if $M\geq O\left(K\ln K\ln N\right)$ and the frequencies are mutually separated by at least $\frac{4}{N}$.*

4.2. Homoscedastic Noise and $L\geq M$

By (14) and (21) the GLS minimization problem is given by

$$\begin{aligned}&\min_{\boldsymbol{u}}\text{tr}\left(\boldsymbol{R}_{\boldsymbol{\Omega}}^{-1}\widetilde{\boldsymbol{R}}_{\boldsymbol{\Omega}}\right)+\text{tr}\left(\widetilde{\boldsymbol{R}}_{\boldsymbol{\Omega}}^{-1}\boldsymbol{R}_{\boldsymbol{\Omega}}\right)\\&\Leftrightarrow\min_{\boldsymbol{u}}\frac{1}{L}\text{tr}\left(\boldsymbol{Y}_{\boldsymbol{\Omega}}^H\boldsymbol{R}_{\boldsymbol{\Omega}}^{-1}\boldsymbol{Y}_{\boldsymbol{\Omega}}\right)+\text{tr}\left(\widetilde{\boldsymbol{R}}_{\boldsymbol{\Omega}}^{-1}\boldsymbol{\Gamma}_{\boldsymbol{\Omega}}T\left(\boldsymbol{u}\right)\boldsymbol{\Gamma}_{\boldsymbol{\Omega}}^T\right)\\&\Leftrightarrow\min_{\boldsymbol{u},\boldsymbol{Z}}\frac{1}{L}\text{tr}\left(\boldsymbol{Z}^HT\left(\boldsymbol{u}\right)^{-1}\boldsymbol{Z}\right)+\text{tr}\left(\boldsymbol{\Gamma}_{\boldsymbol{\Omega}}^T\widetilde{\boldsymbol{R}}_{\boldsymbol{\Omega}}^{-1}\boldsymbol{\Gamma}_{\boldsymbol{\Omega}}T\left(\boldsymbol{u}\right)\right),\\&\qquad\text{subject to } \boldsymbol{Z}_{\boldsymbol{\Omega}}=\boldsymbol{Y}_{\boldsymbol{\Omega}}\\&\Leftrightarrow\min_{\boldsymbol{Z}}\left\|\boldsymbol{Z}\right\|_{\mathcal{A}^w},\ \text{subject to } \boldsymbol{Z}_{\boldsymbol{\Omega}}=\boldsymbol{Y}_{\boldsymbol{\Omega}},\end{aligned}\tag{23}$$

where $w(f)=\left(\frac{1}{N}\boldsymbol{a}^H\left(f\right)\boldsymbol{\Gamma}_{\boldsymbol{\Omega}}^T\widetilde{\boldsymbol{R}}_{\boldsymbol{\Omega}}^{-1}\boldsymbol{\Gamma}_{\boldsymbol{\Omega}}\boldsymbol{a}\left(f\right)\right)^{-\frac{1}{2}}=\left(\frac{1}{N}\boldsymbol{a}_{\boldsymbol{\Omega}}^H\left(f\right)\widetilde{\boldsymbol{R}}_{\boldsymbol{\Omega}}^{-1}\boldsymbol{a}_{\boldsymbol{\Omega}}\left(f\right)\right)^{-\frac{1}{2}}$ is again the square root of Capon's spectrum.

4.3. Heteroscedastic Noise

In the case of heteroscedastic noise we can similarly reformulate the GLS minimization problems (when $L<M$ and $L\geq M$) as atomic norm methods by consecutively applying (17) and (21). In particular, when $L<M$ the atomic norm formulation is given by

$$\min_{\boldsymbol{Z}}\sqrt{M}\left\|\boldsymbol{Z}\right\|_{\mathcal{A}}+\left\|\boldsymbol{B}-\boldsymbol{Z}_{\boldsymbol{\Omega}}\right\|_{2,1}. \tag{24}$$

When $L\geq M$ it becomes

$$\min_{\boldsymbol{Z}}\sqrt{N}\left\|\boldsymbol{Z}\right\|_{\mathcal{A}^w}+\sum_{m=1}^M\sqrt{\left(\widetilde{\boldsymbol{R}}_{\boldsymbol{\Omega}}^{-1}\right)_{mm}}\left\|\left(\frac{1}{\sqrt{L}}\boldsymbol{Y}-\boldsymbol{Z}\right)_{\Omega_m}\right\|_2. \tag{25}$$

Note that $\boldsymbol{B}$ in (24) and $w(f)$ in (25) are defined as in the previous two subsections. We leave the detailed derivations to interested readers. Similarly to the ULA case, $\boldsymbol{Y}_{\boldsymbol{\Omega}}$ (or $\frac{1}{\sqrt{L}}\boldsymbol{Y}_{\boldsymbol{\Omega}}$) in (23) (or (25)) can be replaced by $\widetilde{\boldsymbol{R}}_{\boldsymbol{\Omega}}^{\frac{1}{2}}$ for reducing the problem dimension.

4.4. Remarks

Remark 2 *In the case of homoscedastic noise if we insist on splitting the signal and noise covariances in $\boldsymbol{R}$, then GLS can be reformulated as (weighted) atomic norm minimization problems similar to those in the case of heteroscedastic noise, whereas data fitting is measured by the Frobenious norm instead of the (weighted) $\ell_{2,1}$ norm.*

Remark 3 *In the case of highly correlated sources Capon's power spectrum, whose square root is used in GLS as the weighting function, can be biased and hence, the GLS power estimates can be biased as well. This explains the behavior of GLS reported in [6].*

5. CONCLUSION

We have shown the equivalence between GLS and atomic norm methods in this paper. Since GLS and the existing atomic norm methods have different pros and cons it will be of great interest in future studies to develop new gridless sparse methods that combine their merits, for example, being hyper-parameter free and highly robust to source correlations.

6. REFERENCES

[1] P. Stoica and R. L. Moses, *Spectral analysis of signals*. Pearson/Prentice Hall Upper Saddle River, NJ, 2005.

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

[3] M. Hyder and K. Mahata, "Direction-of-arrival estimation using a mixed $\ell_{2,0}$ norm approximation," *IEEE Transactions on Signal Processing*, vol. 58, no. 9, pp. 4646–4655, 2010.

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

[5] Z. Yang, L. Xie, and C. Zhang, "Off-grid direction of arrival estimation using sparse Bayesian inference," *IEEE Transactions on Signal Processing*, vol. 61, no. 1, pp. 38–43, 2013.

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

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

[8] Z. Yang and L. Xie, "Continuous compressed sensing with a single or multiple measurement vectors," in *IEEE Workshop on Statistical Signal Processing (SSP)*, 2014, pp. 308–311.

[9] Z. Yang and L. Xie, "Exact joint sparse frequency recovery via optimization methods," 2014. [Online]. Available: http://arxiv.org/abs/1405.6585

[10] Z. Yang and L. Xie, "Enhancing sparsity and resolution via reweighted atomic norm minimization," *IEEE Transactions on Signal Processing*, DOI: 10.1109/TSP.2015.2493987, 2015.

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

[12] 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.

[13] 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, 2013.

[14] 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, 2013.

[15] V. Duval and G. Peyré, "Exact support recovery for sparse spikes deconvolution," *Foundations of Computational Mathematics*, pp. 1–41, 2015.

[16] P. Stoica, G. Tang, Z. Yang, and D. Zachariah, "Gridless compressive-sensing methods for frequency estimation: Points of tangency and links to basics," in *22nd European Signal Processing Conference (EUSIPCO)*, 2014, pp. 1831–1835.

[17] Y. Chi, "Joint sparsity recovery for spectral compressed sensing," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2014, pp. 3938–3942.

[18] Z. Tan, Y. C. Eldar, and A. Nehorai, "Direction of arrival estimation using co-prime arrays: A super resolution viewpoint," *IEEE Transactions on Signal Processing*, vol. 62, no. 21, pp. 5565–5576, 2014.

[19] P. Stoica, D. Zachariah, and J. Li, "Weighted SPICE: A unifying approach for hyperparameter-free sparse estimation," *Digital Signal Processing*, vol. 33, pp. 1–12, 2014.

[20] K.-C. Toh, M. J. Todd, and R. H. Tütüncü, "SDPT3–a MATLAB software package for semidefinite programming, version 1.3," *Optimization Methods and Software*, vol. 11, no. 1-4, pp. 545–581, 1999.