GRIDLESS COMPRESSIVE-SENSING METHODS FOR FREQUENCY ESTIMATION: POINTS OF TANGENCY AND LINKS TO BASICS

Petre Stoica, Gongguo Tang†, Zai Yang‡, Dave Zachariah**

* Dept. Information Technology
Uppsala University
Uppsala, Sweden

† Dept. Electrical Eng. & CSci.
Colorado School of Mines
Golden, USA

‡ Centre for E-City, School of EEE
Nanyang Technological University
Singapore

ABSTRACT

The gridless compressive-sensing methods form the most recent class of approaches that have been proposed for estimating the frequencies of sinusoidal signals from noisy measurements. In this paper we review these methods with the main goal of providing new insights into the relationships between them and their links to the basic approach of nonlinear least squares (NLS). We show that a convex relaxation of penalized NLS leads to the atomic-norm minimization method. This method in turn can be approximated by a gridless version of the SPICE method, for which the dual problem is shown to be equivalent to the global matched filter method.

Index Terms— frequency estimation, sparse signal processing, covariance estimation

1. INTRODUCTION

We will make frequent use of the following abbreviations and notations.

Abbreviations: ANM = Atomic norm minimization; FOCUSS = Focal underdetermined system solver; GMF = Global matched filter; IAA = Iterative adaptive approach; i.i.d. = independent and identically distributed; ML = Maximum likelihood; NLS = Nonlinear least squares; SDP = Semidefinite program; SNR = Signal to noise ratio; SPICE = Sparse iterative covariance-based estimation; GL-SPICE = Gridless SPICE; s.t. = subject to; w.r.t. = with respect to.

Notation: $\|\cdot\|_1$, $\|\cdot\|_2$ and $\|\cdot\|_F$ are the ℓ_1, ℓ_2 and Frobenius norms, respectively. x, $\mathbf{x}$ and $\mathbf{X}$ denote scalars, vectors and matrices. $\mathbf{x}^\top$ and $\mathbf{x}^*$ are the transpose and conjugate transpose. $\mathbf{A} \succeq \mathbf{B}$ means $(\mathbf{A}-\mathbf{B})$ is positive semi-definite. $\mathbf{A}^{1/2}$ is a square root of $\mathbf{A} \succeq \mathbf{0}$. $\mathrm{tr}\{\mathbf{A}\}$ is trace of $\mathbf{A}$. $\mathrm{Re}\{x\}$ denotes the real part of x and $|x|$ is the magnitude of x.

We consider the problem of estimating the frequencies of a number of sinusoidal signals from noisy measurements. Let

$$\mathbf{a}(\omega) = [1\ e^{i\omega}\ \cdots\ e^{i(N-1)\omega}]^\top/\sqrt{N}, \quad \omega \in [-\pi, \pi) \tag{1}$$

denote the vector whose entries are the N uniform samples of a sinusoid with frequency ω and unit energy. Using this notation we can describe the measurement vector $\mathbf{y} \in \mathbb{C}^{N\times 1}$ as:

$$\mathbf{y} = \mathbf{A}(\boldsymbol{\omega})\mathbf{s} + \mathbf{e}, \tag{2}$$

where

$$\begin{aligned} \mathbf{A}(\boldsymbol{\omega}) &= [\mathbf{a}(\omega_1)\ \cdots\ \mathbf{a}(\omega_n)] \\ \boldsymbol{\omega} &= [\omega_1\ \cdots\ \omega_n]^\top \\ \mathbf{s} &= [s_1\ \cdots\ s_n]^\top, \end{aligned} \tag{3}$$

(ω_k, s_k) are the frequency and amplitude of the k-th sinusoidal component, $\mathbf{e} \in \mathbb{C}^{N\times 1}$ is the noise vector, and n is the assumed number of components; we let n_0 denote the true value of n. The problem is to estimate $\boldsymbol{\omega}$ from $\mathbf{y}$; once $\boldsymbol{\omega}$ is estimated we can obtain an estimate of $\mathbf{s}$ by solving a simple least-squares problem.

There are a host of methods that can be used for frequency estimation (see, e.g., the book [1]). A relatively recent subset of high-resolution methods make use of ideas from the literature on sparse estimation and are usually designated using the term "compressive sensing approaches" owing to their ability to perform well based on a limited number of observations/measurements. Examples of such methods include FOCUSS [2], GMF [3] and SPICE [4]; a related competitive method is IAA [5] which, however, is not directly based on sparse estimation ideas. For a recent review of some these methods the reader can consult the paper [6].

Similarly to the non-parametric method of the periodogram (e.g. [1]), the compressive sensing methods referred to above rely on a frequency grid and thus their estimates are confined to the set of grid points. Such a quantization of the frequency variable induces an estimation bias that becomes statistically significant as N or SNR increases. To overcome this bias issue a recent trend in the literature on frequency estimation has been to propose so-called gridless compressive-sensing methods that treat the frequencies as continuous (as opposed to quantized) parameters (these methods are sometimes said to perform compressive sensing 'off the grid').

In this paper we will review the following gridless compressive-sensing methods for frequency estimation: ANM, GL-SPICE, and GMF (see [7], [8] and [9]). Our derivations of these methods are simpler than the original derivations in the cited papers, and their motivations provided here are consequently clearer. In addition, we show how these methods are related to one another as well as to the basic approach of NLS.

2. BASIC NLS AND A PENALIZED FORM

This method, which coincides with the ML in the case of normal i.i.d. noise, consists of solving the following minimization problem:

$$\min_{\boldsymbol{\omega},\mathbf{s}} \|\mathbf{y}-\mathbf{A}(\boldsymbol{\omega})\mathbf{s}\|_2^2. \tag{4}$$

The difficulty of this (multimodal) problem limits the practical usage of NLS, despite the statistically sound character of this method. On top of this hardness we have the additional problem of determining an appropriate value of n (i.e. estimating n_0). A possible way of tackling the latter problem is to consider the following penalized form of (4):

$$\min_{\boldsymbol{\omega},\mathbf{s}} \lambda\|\mathbf{s}\|_1+\|\mathbf{y}-\mathbf{A}(\boldsymbol{\omega})\mathbf{s}\|_2^2, \tag{5}$$

where $\lambda \geq 0$ is a weight and the $\|\cdot\|_1$ norm penalty is used to enforce sparsity on $\mathbf{s}$. The idea behind (5) can be explained as follows: using a priori knowledge we can usually choose n such that it is greater than n_0; this means that the amplitude vector $\mathbf{s}$ has $n-n_0$ entries equal to zero. Then there is a range of values of the penalty's weight λ in (5) for which these zero entries of $\mathbf{s}$ can be reliably identified (and hence n_0 determined) from the estimates provided by (5) . Note that in the case of grid-based methods (see, e.g., [2–6]), the $\{\omega_k\}$ in (5) would be pre-specified (viz., they will be the grid points) and n would be much larger than N (to keep the quantization bias small). In sharp contrast to this, $\{\omega_k\}$ in (5) are variables, and typically $n \leq N$ (an assumption that we make in what follows).

Finding an appropriate value for λ in (5) and deciding which elements of the estimated $\mathbf{s}$ are insignificant (i.e. theoretically zero) is a problem in itself that is beyond the scope of this paper. Our main focus here is on methods for solving (5), at least approximately. Clearly, (5) is not easier that the basic NLS problem in (4). However, a reparameterization of (5), followed by a certain relaxation, leads to a feasible problem as explained in the next section.

3. ANM AS A RELAXATION OF NLS

Consider the following minimization problem:

$$\begin{aligned}\min_{x,\mathbf{z},\mathbf{p},\boldsymbol{\omega}} \;& \lambda\left(x+\sum_{k=1}^{n}p_k\right)+\|\mathbf{y}-\mathbf{z}\|_2^2\\ \text{s.t.}\;& \begin{bmatrix} x & \mathbf{z}^* \\ \mathbf{z} & \mathbf{A}(\boldsymbol{\omega})\mathbf{P}\mathbf{A}^*(\boldsymbol{\omega})\end{bmatrix}\succeq \mathbf{0}\end{aligned} \tag{6}$$

where

$$\mathbf{P}=\operatorname{diag}(p_1,\ldots,p_n) \quad (p_k>0) \tag{7}$$

and

$$\mathbf{p}=[p_1\ \cdots\ p_n]^\top \tag{8}$$

(note that, for the sake of convenience, we assume that all $p_k > 0$; the case in which some $p_k = 0$ can be treated similarly by reducing the value of n accordingly). The above problem is an augmented version of (5) in the sense that *the minimization of* (6) *w.r.t. the auxiliary variables x and $\mathbf{p}$ leads to* (5). A simple proof of this fact is presented in what follows.

First we note that the constraint in (6) is equivalent to (see, e.g., [10]):

$$\mathbf{z}=\mathbf{A}(\boldsymbol{\omega})\mathbf{s} \;\text{ for some } \mathbf{s}\in\mathbb{C}^{n\times 1} \tag{9}$$

and

$$x\mathbf{A}(\boldsymbol{\omega})\mathbf{P}\mathbf{A}^*(\boldsymbol{\omega})\succeq \mathbf{z}\mathbf{z}^*. \tag{10}$$

Because $\mathbf{A}(\boldsymbol{\omega})$ has full column-rank (under the assumption that $n \leq N$ and $\omega_k \neq \omega_p$ for $k \neq p$), it follows from (9) and (10) that

$$x\mathbf{P}\succeq \mathbf{s}\mathbf{s}^* \Leftrightarrow x\mathbf{I}\succeq \mathbf{P}^{-1/2}\mathbf{s}\mathbf{s}^*\mathbf{P}^{-1/2} \Leftrightarrow x\geq \mathbf{s}^*\mathbf{P}^{-1}\mathbf{s}. \tag{11}$$

Therefore, for fixed $\mathbf{s}$ and $\mathbf{p}$, the minimizing x and the vectors $\mathbf{z}$ in the feasible set are given by $x=\mathbf{s}^*\mathbf{P}^{-1}\mathbf{s}$ and (9), respectively. Inserting these expressions in (6) leads to the problem:

$$\min_{\mathbf{p},\boldsymbol{\omega},\mathbf{s}} \lambda\sum_{k=1}^{n}\left(p_k+\frac{|s_k|^2}{p_k}\right)+\|\mathbf{y}-\mathbf{A}(\boldsymbol{\omega})\mathbf{s}\|_2^2. \tag{12}$$

The minimization of (12) w.r.t. $\{p_k\}$ yields

$$p_k=|s_k|,\ k=1,\ldots,n \tag{13}$$

and the following concentrated problem:

$$\min_{\boldsymbol{\omega},\mathbf{s}} \tilde{\lambda}\|\mathbf{s}\|_1+\|\mathbf{y}-\mathbf{A}(\boldsymbol{\omega})\mathbf{s}\|_2^2 \quad (\tilde{\lambda}=2\lambda) \tag{14}$$

which coincides with (5). With this observation the proof of the fact that the solution of (5) can be obtained from that of (6) is concluded.

Continuing our discussion on (6), observe that the matrix

$$\mathbf{R}=\mathbf{A}(\boldsymbol{\omega})\mathbf{P}\mathbf{A}^*(\boldsymbol{\omega}) \tag{15}$$

is Toeplitz, and also that

$$\operatorname{tr}\{\mathbf{R}\}=\sum_{k=1}^{n}p_k\|\mathbf{a}(\omega_k)\|_2^2=\sum_{k=1}^{n}p_k. \tag{16}$$

Let

$$\mathbf{R}=\begin{bmatrix} r_0 & r_1 & \cdots & r_{N-1}\\ r_1^* & \ddots & & \vdots\\ \vdots & & \ddots & \vdots\\ r_{N-1}^* & \cdots & \cdots & r_0\end{bmatrix};\ \mathbf{r}=[r_0,\cdots,r_{N-1}]^\top. \tag{17}$$

According to the Carathéodory theorem on Toeplitz matrices (see, e.g., [1]) there is a one-to-one mapping between the parameterization of $\mathbf{R}$ via $\boldsymbol{\omega}$ and $\mathbf{p}$ and that via $\mathbf{r}$ (under the constraints that $\mathbf{R} \succeq \mathbf{0}$ and $\text{rank}(\mathbf{R}) = n$ whenever $n \leq N$). Exploiting this fact we can re-write (6) as follows:

$$\begin{aligned} &\min_{x,\mathbf{z},\mathbf{r}} \lambda(x + \text{tr}\{\mathbf{R}\}) + \|\mathbf{y} - \mathbf{z}\|_2^2 \\ &\text{s.t.} \begin{bmatrix} x & \mathbf{z}^* \\ \mathbf{z} & \mathbf{R} \end{bmatrix} \succeq \mathbf{0}, \quad \text{rank}(\mathbf{R}) = n \end{aligned} \tag{18}$$

and $\mathbf{R}$ as given in (17).

The following *convex relaxation* of (18) is obtained by omitting the rank constraint:

$$\begin{aligned} &\min_{x,\mathbf{z},\mathbf{r}} \lambda(x + \text{tr}\{\mathbf{R}\}) + \|\mathbf{y} - \mathbf{z}\|_2^2 \\ &\text{s.t.} \begin{bmatrix} x & \mathbf{z}^* \\ \mathbf{z} & \mathbf{R} \end{bmatrix} \succeq \mathbf{0} \end{aligned} \tag{19}$$

The term $\text{tr}\{\mathbf{R}\}$ in (19), which also equals the nuclear norm of $\mathbf{R}$, is known to penalize the rank of $\mathbf{R}$. This simple observation motivates the use of the relaxed problem in (19) without the need for any additional modifications.

The frequency estimation method based on solving (19) is nothing but the ANM approach proposed in [7, 8] (and references therein). Note that once $\mathbf{r}$ is estimated, we can obtain frequency estimates using the Carathéodory decomposition in (15) and standard subspace-based techniques (see, e.g., [1]). Also, note that solving (19) as an SDP can be computationally rather expensive for large values of N. A more efficient algorithm than the off-the-shelves SDP software for solving (19) has been suggested in [8].

Finally we remark on the case in which we have a good guess of n_0 and hence would be interested in solving the NLS problem (4) for a small value of $n \approx n_0$. In such a case we would select λ in (19) to be small in an attempt to approach the basic NLS solution; however, the rank-penalizing term $\text{tr}\{\mathbf{R}\}$ in (19) should still receive a significant weight to compensate for the omission of the rank-constraint when passing from (18) to (19); consequently in this case the penalty term in (18) should be $\lambda_1 x + \lambda_2 \text{tr}\{\mathbf{R}\}$, where λ_1 can be arbitrarily small but λ_2 should be appropriately selected (alternatively, the rank constraint should be approximately enforced in a different way).

The above derivation of ANM as a relaxation of a re-parameterized NLS problem makes an interesting connection between ANM and the basic method of NLS. In the next section we relate ANM to the GL-SPICE approach recently introduced in [8].

4. GL-SPICE AS AN APPROXIMATION OF ANM

As we will see GL-SPICE, like ANM, is basically a covariances estimation method (first the covariance $\{r_k\}$ are estimated and then they are used to obtain frequency estimates). However, whereas ANM estimates the covariance matrix of the noise-free signal only, GL-SPICE considers the total covariance matrix of signal and noise. Assuming that the elements of $\mathbf{e}$ are i.i.d. with zero mean and variance denoted by σ, the latter covariance matrix can be written as:

$$\boldsymbol{\Gamma} = \mathbf{R} + \sigma\mathbf{I} = \begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{N-1} \\ \gamma_1^* & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ \gamma_{N-1}^* & \cdots & \cdots & \gamma_0 \end{bmatrix} \tag{20}$$

where $\mathbf{R}$ is as defined in (15). If we attempt to estimate the vector

$$\boldsymbol{\gamma} = [\gamma_0 \ \dots \gamma_{N-1}]^\top \tag{21}$$

via the minimization of the following least-squares criterion

$$\|\mathbf{y}\mathbf{y}^* - \boldsymbol{\Gamma}\|_F^2 \tag{22}$$

then the result will be the standard sample covariances (as can be readily verified). Presumably, a statistically more sound criterion is the following weighted version of (22) (see, e.g., [11]):

$$\|\boldsymbol{\Gamma}^{-1/2}(\mathbf{y}\mathbf{y}^* - \boldsymbol{\Gamma})\|_F^2 = \text{const.} + \text{tr}\{\boldsymbol{\Gamma}\} + \|\mathbf{y}\|_2^2 \mathbf{y}^* \boldsymbol{\Gamma}^{-1} \mathbf{y}. \tag{23}$$

The minimization of (23) by using a grid-based parameterization of $\boldsymbol{\Gamma}$ was considered in [4] (also see the references therein) and the so-obtained method was termed SPICE. More recently, the paper [8] used the Toeplitz parameterization in (20) of $\boldsymbol{\Gamma}$ and suggested estimating $\boldsymbol{\gamma}$ also by minimizing the SPICE criterion in (23):

$$\min_{\boldsymbol{\gamma}} \ \text{tr}\{\boldsymbol{\Gamma}\} + \|\mathbf{y}\|_2^2 \mathbf{y}^* \boldsymbol{\Gamma}^{-1} \mathbf{y} \quad \text{s.t. } \boldsymbol{\Gamma} \succeq \mathbf{0}. \tag{24}$$

The so-obtained method is called GL-SPICE. Note that the rank constraint on the $\mathbf{R}$ matrix in $\boldsymbol{\Gamma}$ is ignored in (24), exactly as it was in the ANM problem in (19).

Let us re-write (24) as an SDP:

$$\min_{x,\boldsymbol{\gamma}} \ \|\mathbf{y}\|_2^2 x + \text{tr}\{\boldsymbol{\Gamma}\} \quad \text{s.t.} \begin{bmatrix} x & \mathbf{y}^* \\ \mathbf{y} & \boldsymbol{\Gamma} \end{bmatrix} \succeq \mathbf{0}. \tag{25}$$

Similar comments, to those following (19), about algorithms for efficiently solving the above SDP and on how to obtain frequency estimates from the solution $\boldsymbol{\gamma}$ of (25), can be made here as well. More importantly, observe that (25) can be viewed as an *approximation of the ANM problem* in (19) corresponding to setting $\mathbf{z} = \mathbf{y}$, which is a natural choice given the fact that now $\boldsymbol{\Gamma}$ is the covariance matrix of the noisy data vector $\mathbf{y}$. Interestingly, the weight λ in (19) becomes irrelevant in (25), an observation that provides further motivation to the finding in [4] that the SPICE criterion of (24) does not necessarily require a careful differential weighting of its two terms. This fact constitutes an advantage of the SPICE approach over many other methods (ANM included) which require a fine-tuning of the penalty's weight (such as λ in (19)), or else they may fail to provide satisfactory performance.

5. GMF AS A DUAL OF GL-SPICE

The GMF method (see, e.g., [3, 9]; and also [7]) consists of solving the following semi-infinite maximization problem:

$$\max_{\mathbf{b}\in\mathbb{C}^{N\times 1}} \ \mathrm{Re}\{\mathbf{b}^*\mathbf{y}\} \quad \text{s.t. } |\mathbf{a}^*(\omega)\mathbf{b}|^2 \leq \rho \quad \forall\omega \in [-\pi, \pi), \ \rho = 1/\|y\|_2^2. \tag{26}$$

Note that there exist an infinite number of constraints in (26), which is why (26) was called a semi-infinite problem. Also, the specific value of ρ in (26) was chosen only for the convenience of the following analysis; any other positive value of ρ would work just as well. Finally, once the solution $\mathbf{b}$ of (26) is found, the frequency estimates are obtained as those values of ω at which $|\mathbf{a}^*(\omega)\mathbf{b}|^2$ is equal (or close) to ρ. The intuitive motivation for GMF is that the optimal matched filter $\mathbf{b}$ determined from (26) is likely to have maximum gains at the frequencies of the sinusoidal components in $\mathbf{y}$ and small gains otherwise as the total gain of the filter is limited.

At first sight the GMF problem above has nothing in common with the problem solved by either ANM or GL-SPICE. However, we will show in the following that in fact (26) is *equivalent to the dual of the* GL-SPICE *optimization problem* in (25). To that end, we prove in the Appendix that the dual problem for (25) is the following SDP:

$$\max_{\mathbf{Z}\in\mathbb{C}^{(N+1)\times(N+1)}} \ \mathrm{Re}\{\mathbf{b}^*\mathbf{y}\} \quad \text{s.t. } \mathbf{Z} = \begin{bmatrix} 1 & -\mathbf{b}^* \\ -\mathbf{b} & \mathbf{C} \end{bmatrix} \succeq \mathbf{0}$$
$$\mathrm{tr}\{\mathbf{C}\} = \rho N, \ \mathrm{tr}_k\{\mathbf{C}\} = 0, k = 1, \ldots, N-1, \tag{27}$$

where $\mathrm{tr}_k\{\mathbf{C}\}$ denotes the sum of the elements on the kth subdiagonal of $\mathbf{C}$. Next, we note that for a matrix $\mathbf{C}$ satisfying the conditions in (27) it holds that:

$$\begin{aligned} N\mathbf{a}^*(\omega)\mathbf{C}\mathbf{a}(\omega) &= N\mathrm{tr}\{\mathbf{C}\mathbf{a}(\omega)\mathbf{a}^*(\omega)\} \\ &= \mathrm{tr}\left\{\mathbf{C}(\mathbf{I} + \sum_{k=1}^{N-1}\mathbf{J}_k e^{ik\omega} + \sum_{k=1}^{N-1}\mathbf{J}_k^* e^{-ik\omega})\right\} \\ &= \rho N \end{aligned} \tag{28}$$

where

$$\mathbf{J}_k = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{I}_{N-k} & \mathbf{0} \end{bmatrix} \quad k = 1, \ldots, N-1. \tag{29}$$

It follows from this observation that, for $\mathbf{Z} \succeq \mathbf{0}$ (as in (27)),

$$\rho - |\mathbf{a}^*(\omega)\mathbf{b}|^2 = \mathbf{a}^*(\omega)(\mathbf{C} - \mathbf{b}\mathbf{b}^*)\mathbf{a}(\omega) \geq \mathbf{0} \quad \forall\omega \in [-\pi, \pi) \tag{30}$$

and thus the conditions in (27) are sufficient for the constraint in (26) to hold. The conditions in (27) are also necessary for the constraint in (26). To see this note that it follows from standard results on positive trigonometric polynomials (see, e.g., [12]) that there exists a matrix $\mathbf{Q} \succeq \mathbf{0}$ such that, under the condition in (26),

$$\rho - |\mathbf{a}^*(\omega)\mathbf{b}|^2 = \mathbf{a}^*(\omega)\mathbf{Q}\mathbf{a}(\omega). \tag{31}$$

Let $\mathbf{C}$ be defined by the equality

$$\mathbf{C} = \mathbf{Q} + \mathbf{b}\mathbf{b}^*. \tag{32}$$

Then we have that

$$\mathbf{a}^*(\omega)\mathbf{C}\mathbf{a}(\omega) = \rho \tag{33}$$

which, in view of the calculation in (28), can hold true if and only if $\mathbf{C}$ satisfies the constraint in (27). Note also that $\mathbf{C} - \mathbf{b}\mathbf{b}^* \succeq \mathbf{0}$. Therefore the constraint in (26) implies those in (27), which concludes the proof that (26) and (27) are *equivalent problems*. More concretely, we have shown that: for every $(\mathbf{b}, \mathbf{C})$ satisfying (27), $\mathbf{b}$ satisfies (26) as well; and, conversely, for every $\mathbf{b}$ that satisfies (26) there exists a $\mathbf{C}$ such that $(\mathbf{b}, \mathbf{C})$ satisfies (27). Among other things, the shown equivalence between (26) and (27) can be exploited to find the solution of the semi-infinite optimization problem in (26) by solving the SDP in (27).

6. CONCLUSIONS

This paper has presented simple derivations of and new insights into the frequency estimation methods of ANM, GL-SPICE and GMF, as well as the relationships between them and their link to the basic approach of NLS. It will hopefully stimulate the interest in these recently proposed techniques and, in particular, in a more detailed practical study of them to determine their possible niche in this competitive area of signal processing that has numerous applications from biomedicine and spectroscopy to radar and communications. Preliminary numerical experience with these methods, as well as a limited comparison with a few of the existing methods for frequency estimation (such as IAA), have been reported in [3, 7–9]. Finally, the hope is that the discussion here can also inspire novel versions of the methods discussed, with improved performance.

7. APPENDIX: THE GL-SPICE DUAL PROBLEM

The Lagrangian associated with (25) is given by (see, e.g., [13])

$$L(x, \gamma, \mathbf{Z}) = x + N\rho\gamma_0 - \mathrm{tr}\left\{\mathbf{Z}\begin{bmatrix} x & \mathbf{y}^* \\ \mathbf{y} & \mathbf{\Gamma} \end{bmatrix}\right\} \tag{34}$$

(here $\mathbf{Z}$ is the multiplier matrix). Using the following partition of $\mathbf{Z}$,

$$\mathbf{Z} = \begin{bmatrix} z & -\mathbf{b}^* \\ -\mathbf{b} & \mathbf{C} \end{bmatrix} \tag{35}$$

we can write (34) in the following equivalent form:

$$\begin{aligned} L(x,\boldsymbol{\gamma},\mathbf{Z}) &= x + N\rho\gamma_0 - xz + \mathbf{b}^*\mathbf{y} + \mathbf{y}^*\mathbf{b} - \text{tr}\{\mathbf{C}\boldsymbol{\Gamma}\} \\ &= (1-z)x + 2\text{Re}\{\mathbf{b}^*\mathbf{y}\} - \text{tr}\{\mathbf{C}\boldsymbol{\Gamma}\} + N\rho\gamma_0 \end{aligned} \tag{36}$$

which is an affine function of x and $\boldsymbol{\gamma}$. Making use of the matrices $\{\mathbf{J}_k\}$ defined in (29), we can write:

$$\begin{aligned} \text{tr}\{\mathbf{C}\boldsymbol{\Gamma}\} &= \text{tr}\left\{\mathbf{C}(\gamma_0\mathbf{I} + \sum_{k=1}^{N-1}\mathbf{J}_k^*\gamma_k + \sum_{k=1}^{N-1}\mathbf{J}_k\gamma_k^*)\right\} \\ &= \gamma_0\text{tr}\{\mathbf{C}\} + \sum_{k=1}^{N-1}\text{tr}\{\mathbf{C}\mathbf{J}_k^*\}\gamma_k + \sum_{k=1}^{N-1}\text{tr}\{\mathbf{C}\mathbf{J}_k\}\gamma_k^*. \end{aligned} \tag{37}$$

Inserting (37) in (36) and observing that $\text{tr}\{\mathbf{C}\mathbf{J}_k^*\} = \text{tr}_k\{\mathbf{C}\}$, lead to the following expression for the Lagrangian:

$$\begin{aligned} L(x,\boldsymbol{\gamma},\mathbf{Z}) =& (1-z)x + \gamma_0(N\rho - \text{tr}\{\mathbf{C}\}) \\ & - 2\sum_{k=1}^{N-1}\text{Re}\{\text{tr}_k\{\mathbf{C}\}\gamma_k^*\} + 2\text{Re}\{\mathbf{b}^*\mathbf{y}\} \end{aligned} \tag{38}$$

The infimum of this function w.r.t. the (unconstrained) x and $\boldsymbol{\gamma}$ is given by:

$$\inf_{x,\boldsymbol{\gamma}} L(x,\boldsymbol{\gamma},\mathbf{Z}) = 2\text{Re}\{\mathbf{b}^*\mathbf{y}\} \tag{39}$$

if

$$\begin{cases} z = 1 \\ \text{tr}\{\mathbf{C}\} = \rho N \\ \text{tr}_k\{\mathbf{C}\} = 0 \end{cases}. \tag{40}$$

Else the infimum is $-\infty$. This observation concludes the proof of the fact that the GL-SPICE dual problem is given by (27).

REFERENCES

[1] P. Stoica and R. Moses, *Spectral Analysis of Signals*, Prentice Hall, 2005.

[2] I.F. Gorodnitsky and B.D. Rao, "Sparse signal reconstruction from limited data using FOCUSS: A re-weighted minimum norm algorithm," *IEEE Trans. Signal Processing*, vol. 45, no. 3, pp. 600–616, 1997.

[3] J-J. Fuchs, "On the application of the global matched filter to DOA estimation with uniform circular arrays," *IEEE Trans. Signal Processing*, vol. 49, no. 4, pp. 702–709, 2001.

[4] 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.

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

[6] P. Stoica, D. Zachariah, and J. Li, "Weighted SPICE: A unifying approach for hyperparameter-free sparse estimation," Submitted.

[7] G Tang, B.N. Bhaskar, P. Shah, and B. Recht, "Compressed sensing off the grid," arxiv.org/abs/1207.6053.

[8] Z. Yang and L. Xie, "On gridless sparse methods for line spectral estimation from complete and incomplete data," Submitted.

[9] A. Panahi and M. Viberg, "Gridless compressive sensing," Submitted.

[10] E. Kreindler and A. Jameson, "Conditions for nonnegativeness of partitioned matrices," *IEEE Trans. Automatic Control*, vol. 17, no. 1, pp. 147–148, 1972.

[11] B. Ottersten, P. Stoica, and R. Roy, "Covariance matching estimation techniques for array signal processing applications," *Digital Signal Processing*, vol. 8, no. 3, pp. 185–210, 1998.

[12] B. Dumitrescu, I. Tabus, and P. Stoica, "On the parameterization of positive real sequences and MA parameter estimation," *IEEE Trans. Signal Processing*, vol. 49, no. 11, pp. 2630–2639, 2001.

[13] S.P. Boyd and L. Vandenberghe, *Convex optimization*, Cambridge University Press, 2004.