

EXTENSIONS OF SEMIDEFINITE PROGRAMMING METHODS FOR ATOMIC DECOMPOSITION

Hsiao-Han Chao, Lieven Vandenbergh

Department of Electrical Engineering
University of California, Los Angeles

ABSTRACT

We present an extension of recent semidefinite programming formulations for atomic decomposition over continuous dictionaries, with applications to continuous or ‘gridless’ compressed sensing. The dictionary considered in this paper is defined in terms of a general matrix pencil and is parameterized by a complex variable that varies over a segment of a line or circle in the complex plane. The main result of the paper is the formulation as a convex semidefinite optimization problem, and a simple constructive proof of the equivalence. The techniques are illustrated with a direction of arrival estimation problem, and an example of low-rank structured matrix decomposition.

Index Terms— Sparse signal reconstruction, low-rank matrix completion, compressed sensing, semidefinite optimization, array processing.

1. INTRODUCTION

Few optimization problems have attracted as much interest in recent years as the problem of minimizing the sum of a convex function and an ℓ_1 -norm regularization term. A general form of problems of this type is

$$\begin{aligned} & \text{minimize} && f\left(\sum_{k=1}^r x_k a_k\right) + \sum_{k=1}^r |x_k| \\ & \text{subject to} && a_k \in \mathcal{D}, \quad k = 1, \dots, r, \end{aligned} \quad (1)$$

where f is a convex function and \mathcal{D} is a set (or *dictionary*) of vectors in \mathbf{C}^n or \mathbf{R}^n . The unknowns in problem (1) are the real or complex coefficients x_k , the vectors (or *atoms*) a_1, \dots, a_r selected from \mathcal{D} , and the number r of selected dictionary elements. If \mathcal{D} is a finite set, it can be represented by a matrix D with the elements of \mathcal{D} as its columns, and the problem can be written as

$$\text{minimize} \quad f(Dx) + \|x\|_1.$$

This includes as special cases the Lasso problem [1], basis pursuit [2], noisy basis pursuit [3, 4], and numerous other applications [5–8].

When reviewing the literature on ℓ_1 -norm methods in signal processing [5, 9–11], it is striking that many of the underlying applications involve signals in continuous domains (time, space, or frequency domain), and the ℓ_1 -norm problems arise after discretizing and truncating an infinite dictionary. The discretization is used when no exact method for the continuous problem is known, or when the discretized problem is believed to be easier to solve numerically by convex optimization techniques.

It was recently noted that certain problems of the form (1) with infinite dictionaries can be exactly solved by semidefinite optimization. In particular, the authors of [12–17] consider ℓ_1 -norm minimization with dictionaries of vectors of undamped complex exponentials,

$$\mathcal{D} = \left\{ \frac{1}{\sqrt{n}} (1, e^{j\omega}, e^{j2\omega}, \dots, e^{j(n-1)\omega}) \mid \omega \in [0, 2\pi] \right\}, \quad (2)$$

and use the fact that problem (1) is equivalent to the finite-dimensional convex optimization problem

$$\begin{aligned} & \text{minimize} && f(y) + (\text{tr } X + z)/2 \\ & \text{subject to} && \begin{bmatrix} X & y \\ y^H & z \end{bmatrix} \succeq 0 \\ & && X \text{ is Toeplitz.} \end{aligned}$$

The variables in this problem are $X \in \mathbf{H}^n$, $y \in \mathbf{C}^n$, $z \in \mathbf{R}$. (Here \mathbf{H}^n denotes the set of Hermitian $n \times n$ matrices.)

A useful matrix extension of problem (1) results from allowing the variables x_k to be vectors:

$$\begin{aligned} & \text{minimize} && f\left(\sum_{k=1}^r a_k x_k^H\right) + \sum_{k=1}^r \|x_k\|_2 \\ & \text{subject to} && a_k \in \mathcal{D}, \quad k = 1, \dots, r. \end{aligned} \quad (3)$$

For example, if \mathcal{D} is the set of unit-norm vectors in \mathbf{C}^n , the problem can be shown to be equivalent to

$$\text{minimize} \quad f(Y) + \|Y\|_*, \quad (4)$$

with variable $Y \in \mathbf{C}^{n \times m}$, where $\|Y\|_*$ denotes the *trace norm* or *nuclear norm* (sum of singular values). The equivalence can be seen by writing (4) in the equivalent form

$$\text{minimize} \quad f\left(\sum_{k=1}^r u_k v_k^H\right) + \sum_{k=1}^r (\|u_k\|_2^2 + \|v_k\|_2^2)/2$$

[18, §5.3], making a change of variables

$$t_k = \|u_k\|_2^2, \quad a_k = u_k/\sqrt{t_k}, \quad x_k = \sqrt{t_k}v_k,$$

and eliminating the scalars t_k by noting that $t_k + \|x_k\|_2^2/t_k$ is minimum for $t_k = \|x_k\|_2$.

The authors of [19,20] consider (3) with the dictionary of complex exponentials (2) and show that the problem is equivalent to the convex problem

$$\begin{aligned} & \text{minimize} && f(Y) + (\text{tr } X + \text{tr } Z)/2 \\ & \text{subject to} && \begin{bmatrix} X & Y \\ Y^H & Z \end{bmatrix} \succeq 0 \\ & && X \text{ is Toeplitz,} \end{aligned} \quad (5)$$

with variables $X \in \mathbf{H}^n$, $Y \in \mathbf{C}^{n \times m}$, $Z \in \mathbf{H}^m$. This extension includes problem (1) as a special case with $m = 1$. Applications of (1) or (3) with the dictionary (2) arise frequently in signal processing, for example, in the estimation of line spectra, point source localization, and sensor array processing [5,9,10,21].

In this paper we discuss extensions of the semidefinite programming formulations of (1) and (3) to a large class of dictionaries defined in terms of matrix pencils. In §2.1 we define the general dictionary and show that it includes as special cases the dictionary of complex exponentials (2) and subsets of the form

$$\mathcal{D} = \left\{ \frac{1}{\sqrt{n}}(1, e^{j\omega}, e^{j2\omega}, \dots, e^{j(n-1)\omega}) \mid |\omega - \alpha| \leq \beta \right\}, \quad (6)$$

in which we restrict the frequencies to intervals $[\alpha - \beta, \alpha + \beta]$. In §2.2 we give a semidefinite programming formulation of (3) for the matrix pencil dictionary, and outline a simple and constructive proof of the equivalence. Section 3 illustrates the use of constrained dictionaries with two examples. Section 4 gives the concluding remarks.

2. SEMIDEFINITE OPTIMIZATION FORMULATION

2.1. Dictionary

Define two $(n-1) \times n$ matrices

$$F = \begin{bmatrix} 0 & I_{n-1} \end{bmatrix}, \quad G = \begin{bmatrix} I_{n-1} & 0 \end{bmatrix} \quad (7)$$

(where I_{n-1} is the identity matrix of order $n-1$) and let \mathcal{C} be the unit circle in the complex plane. Then the set

$$\mathcal{D} = \{a \in \mathbf{C}^n \mid \|a\|_2 = 1, (\lambda G - F)a = 0, \lambda \in \mathcal{C}\} \quad (8)$$

contains all vectors of the form

$$a = \frac{c}{\sqrt{n}}(1, e^{j\omega}, e^{j2\omega}, \dots, e^{j(n-1)\omega}) \quad (9)$$

with $|c| = 1$. Up to a phase factor c , this is the dictionary of undamped complex exponentials in (2).

We generalize (8) in two ways. First, we allow F, G to be arbitrary matrices, so $\lambda G - F$ is a general matrix pencil. Second, we define \mathcal{C} to be a subset of the closed complex plane defined by a quadratic equality and inequality,

$$\mathcal{C} = \{\lambda \in \mathbf{C} \cup \{\infty\} \mid g_\Phi(\lambda, 1) = 0, g_\Psi(\lambda, 1) \leq 0\}, \quad (10)$$

where Φ and Ψ are 2×2 Hermitian matrices with $\det \Phi < 0$, and for a 2×2 Hermitian matrix Λ , we define

$$g_\Lambda(\mu, \nu) = \begin{bmatrix} \mu \\ \nu \end{bmatrix}^H \Lambda \begin{bmatrix} \mu \\ \nu \end{bmatrix}.$$

When $\Phi_{11} = 0$ and $\Psi_{11} \leq 0$ we admit the point $\lambda = \infty$ in the set (10). (When $\lambda = \infty$, the condition $(\lambda G - F)a = 0$ in (8) is interpreted as $Ga = 0$.) The solution set of the equality $g_\Phi(\lambda, 1) = 0$ is a circle or straight line in the complex plane. For the purposes of this paper the most important example is

$$\Phi = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \quad (11)$$

which parameterizes the unit circle. By adding the inequality in (10) we obtain segments of circles or lines [22]. For example, the dictionary of vectors (9) with $|\omega - \alpha| \leq \beta$ is defined by (11) and $\Psi_{11} = 0$, $\Psi_{12} = -e^{j\alpha}$, $\Psi_{22} = 2 \cos \beta$.

2.2. Main result

The following theorem gives a convex formulation of problem (3), with the dictionary \mathcal{D} defined in (8) and (10). To simplify the handling of the point at infinity, we define

$$\bar{\mathcal{C}} = \{(\mu, \nu) \neq 0 \mid g_\Phi(\mu, \nu) = 0, g_\Psi(\mu, \nu) \leq 0\}.$$

The pair (μ, ν) is identified with a finite $\lambda = \mu/\nu$ if $\nu \neq 0$ and with $\lambda = \infty$ otherwise.

Theorem 2.1 *Let F, G be matrices in $\mathbf{C}^{p \times n}$, and Φ, Ψ be Hermitian 2×2 matrices with $\det \Phi < 0$. Then the problem*

$$\begin{aligned} & \text{minimize} && f\left(\sum_{k=1}^r a_k x_k^H\right) + \sum_{k=1}^r \|x_k\|_2 \\ & \text{subject to} && \mu_k G a_k = \nu_k F a_k, \quad k = 1, \dots, r \\ & && \|a_k\|_2 = 1, \quad k = 1, \dots, r \\ & && (\mu_k, \nu_k) \in \bar{\mathcal{C}}, \quad k = 1, \dots, r \end{aligned} \quad (12)$$

with variables $a_k \in \mathbf{C}^n$, $x_k \in \mathbf{C}^m$, $(\mu_k, \nu_k) \in \mathbf{C}^2$, and r , is equivalent to the semidefinite optimization problem

$$\begin{aligned} & \text{min.} && f(Y) + (\text{tr } X + \text{tr } Z)/2 \\ & \text{s. t.} && \Phi_{11} F X F^H + \Phi_{21} F X G^H \\ & && \quad + \Phi_{12} G X F^H + \Phi_{22} G X G^H = 0 \\ & && \Psi_{11} F X F^H + \Psi_{21} F X G^H \\ & && \quad + \Psi_{12} G X F^H + \Psi_{22} G X G^H \preceq 0 \\ & && \begin{bmatrix} X & Y \\ Y^H & Z \end{bmatrix} \succeq 0, \end{aligned} \quad (13)$$

with variables $X \in \mathbf{H}^n$, $Y \in \mathbf{C}^{n \times m}$, and $Z \in \mathbf{H}^m$.

Problem (5) is a special case with F, G defined in (7), Φ defined in (11), and $\Psi = 0$. In this case, the matrix equality $FXF^H = GXG^H$ states that the upper-left $(n-1) \times (n-1)$ block of X equals its lower-right $(n-1) \times (n-1)$ block, or equivalently, X is Toeplitz.

Our proof of theorem 2.1 relies on the following result, which is an immediate consequence of Corollary 1 in [23].

Lemma 2.1 *Suppose X is positive semidefinite, has rank r , and satisfies the first two constraints in (13). Then X can be factorized as $X = UU^H$ where $U \in \mathbf{C}^{n \times r}$ and*

$$FU = W \text{diag}(\mu), \quad GU = W \text{diag}(\nu),$$

for some $W \in \mathbf{C}^{p \times r}$ and vectors $\mu, \nu \in \mathbf{C}^r$ that satisfy

$$(\mu_k, \nu_k) \in \bar{\mathcal{C}}, \quad k = 1, \dots, r.$$

We omit the proof of this lemma, but emphasize that the proof outlined in [23] gives a simple constructive algorithm, based on singular value and Schur decompositions, for computing the factorization of X .

Proof of Theorem 2.1. We first show that (13) is a relaxation of (12). Suppose $a_1, \dots, a_r, x_1, \dots, x_r$ are feasible in (12). Without loss of generality assume that $x_k \neq 0$ for all k . Define $t_k = \sqrt{\|x_k\|_2}$ and

$$\begin{bmatrix} X & Y \\ Y^H & Z \end{bmatrix} = \sum_{k=1}^r \begin{bmatrix} t_k a_k \\ (1/t_k)x_k \end{bmatrix} \begin{bmatrix} t_k a_k \\ (1/t_k)x_k \end{bmatrix}^H.$$

Then X, Y, Z are feasible in (13) with objective function equal to $f(\sum_k a_k x_k^H) + \sum_k \|x_k\|_2$.

To establish the converse, we assume X, Y, Z are feasible in (13) and show how to find a set of vectors a_k, x_k that are feasible in (12) with objective function less than or equal to the objective of (13). Suppose X has rank r . We can factorize X as in lemma 2.1 and define $a_k = u_k / \|u_k\|_2$ and $\theta_k = \|u_k\|_2^2$, where u_k is the k th column of U . This gives a factorization

$$X = \sum_{k=1}^r \theta_k a_k a_k^H \quad (14)$$

where $\theta_k > 0$ and the vectors a_k are linearly independent and satisfy the constraints in (12). The third constraint in (13) then implies that there exist vectors x_1, \dots, x_r such that

$$Y = \sum_{k=1}^r a_k x_k^H, \quad Z \succeq \sum_{k=1}^r \frac{1}{\theta_k} x_k x_k^H.$$

Substituting the expressions for X, Y, Z in the objective gives

$$\begin{aligned} f(Y) + \frac{1}{2}(\text{tr}(X) + \text{tr}(Z)) & \\ & \geq f\left(\sum_{k=1}^r a_k x_k^H\right) + \frac{1}{2} \sum_{k=1}^r \left(\theta_k + \frac{1}{\theta_k} \|x_k\|_2^2\right) \\ & \geq f\left(\sum_{k=1}^r a_k x_k^H\right) + \sum_{k=1}^r \|x_k\|_2. \end{aligned}$$

The last line follows from the arithmetic-geometric mean inequality. \square

3. EXAMPLES

3.1. Direction of arrival estimation

The first example illustrates the interval constraints in dictionaries of the form (6). We consider a uniform linear array of $n = 50$ sensors. The signal arriving at the array is a superposition of a small number of planar waves arriving from different directions in $[-\pi/2, \pi/2]$. We take $2d/\lambda_c = 1$, where d is the distance between the sensors and λ_c the signal wavelength [21, §6.2]. When all sensor measurements are available, the directions of arrival can be estimated by classical methods, such as MUSIC and ESPRIT [21, 24, 25]. In this example, however, we assume that only a randomly selected subset of 30 sensors is used. Moreover the sensors are not omnidirectional. The 30 sensors are randomly partitioned in two groups of size 15. Sensors of group 1 measure signals arriving from directions in $[-\pi/2, \pi/6]$; sensors in group 2 measure signals arriving from directions $[-\pi/6, \pi/2]$. The output of the remaining 20 sensors is not used. To simplify notation we will assume the measurements are noise-free.

The direction of arrival (DOA) estimation problem can be put in the framework discussed in this paper by defining three dictionaries

$$\mathcal{D}_j = \{(1, e^{j\omega}, e^{j2\omega}, \dots, e^{j(n-1)\omega}) \mid |\omega - \alpha_j| \leq \beta_j\},$$

for $j = 1, 2, 3$. Here $\omega = \pi \sin \theta$ is the spatial frequency associated with a direction of arrival θ . The three intervals $[\alpha_j - \beta_j, \alpha_j + \beta_j]$ are the images of the intervals $[-\pi/2, -\pi/6]$, $[-\pi/6, \pi/6]$, and $[\pi/6, \pi/2]$ after the transformation $\omega = \pi \sin \theta$. The DOA estimation problem can be formulated as an infinite-dimensional ‘basis-pursuit’ problem

$$\begin{aligned} \text{minimize} \quad & \sum_{j=1}^3 \sum_{k=1}^r |x_{jk}| \\ \text{subject to} \quad & y_j = \sum_{k=1}^{r_j} x_{jk} a_{jk}, \quad j = 1, 2, 3 \\ & a_{jk} \in \mathcal{D}_j, \quad k = 1, \dots, r_j, \quad j = 1, 2, 3 \\ & (y_1 + y_2)_{I_1} = b_1, \quad (y_2 + y_3)_{I_2} = b_2. \end{aligned} \quad (15)$$

Each of the variables y_j is a linear combination of r_j elements of the dictionary \mathcal{D}_j , i.e., an n -vector that represents a sum of r_j signals arriving from directions in the j th interval. In the last two constraints, $I_1 \subset \{1, 2, \dots, n\}$ is an index set containing the indices of the sensors in group 1, and I_2 is an index set for group 2. The vector b_1 contains the 15 measurements at the sensors of group 1, and b_2 contains the measurements for group 2. The variables in the problem are the vectors y_j, a_{jk} , the coefficients x_{jk} , and the number of elements r_j chosen from each dictionary.

Following the same idea as in theorem 2.1 we can cast the

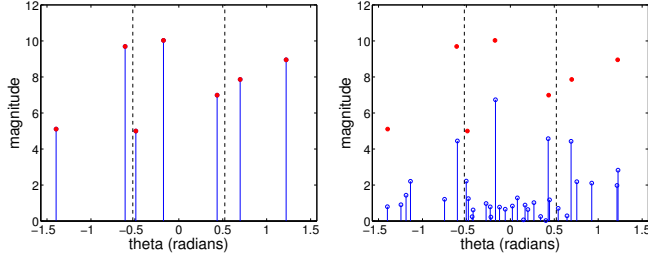


Fig. 1. DOA estimation with and without interval constraints.

problem as the SDP

$$\begin{aligned}
\min. \quad & \sum_{j=1}^3 (\text{tr}(X_j) + z_j)/2 \\
\text{s.t.} \quad & \begin{bmatrix} X_j & y_j \\ y_j^H & z_j \end{bmatrix} \succeq 0, \quad j = 1, 2, 3 \\
& X_1, X_2, X_3 \text{ are Toeplitz} \\
& -e^{-j\alpha_j} F X_j G^H - e^{j\alpha_j} G X_j F^H \\
& \quad + 2 \cos \beta_j G X_j G^H \preceq 0, \quad j = 1, 2, 3 \\
& (y_1 + y_2)_{I_1} = b_1, \quad (y_2 + y_3)_{I_2} = b_2
\end{aligned} \tag{16}$$

with variables $X_j \in \mathbf{H}^n$, $y_j \in \mathbf{C}^n$, $z_j \in \mathbf{R}$, for $j = 1, 2, 3$. The matrices F, G are defined in (7).

Figure 1 shows an example. The red dots show the angles and magnitudes of 7 signals used to compute the measurement vectors b_1, b_2 . The estimated angles and coefficients are shown with blue lines. The figure on the left-hand side shows that the angles and the coefficients are exactly recovered from the solution of (15) and (16). The right-hand plot shows the solution recovered from the SDP (16) if we omit the third set of constraints (which enforces the interval constraints). We also repeated the simulation with the same angles as in figure 1, different, randomly generated coefficients, and different random selections of the two sensor groups. The solution of the SDP (16) gave the exact answer in all instances, whereas the SDP without the interval constraints was successful in only about 27% of the instances.

3.2. Structured matrix decomposition

The second example is an application of theorem 2.1 with $m > 1$. We generate a 30×30 matrix $C = AB + N$ as a product of a 30×3 matrix A with entries $A_{ij} = \exp(j(i-1)\omega_j)$, for given values of $\omega_1, \omega_2, \omega_3$, and a randomly generated complex 3×30 matrix B with entries from a normal distribution, plus a Gaussian noise matrix N . The goal is to estimate the parameters ω_j and the matrix B from the noisy measurements C .

We compare two methods. In the first method we assume we are given a narrow interval that includes the parameters ω_j . In this case we define a dictionary

$$\mathcal{D} = \left\{ \frac{1}{\sqrt{n}} (1, e^{j\omega}, e^{j2\omega}, \dots, e^{j(n-1)\omega}) \mid |\omega - \alpha| \leq \beta \right\}$$

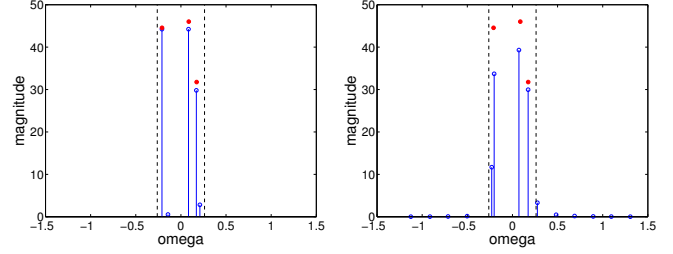


Fig. 2. Structured matrix decomposition of a matrix with rank 3, with and without interval constraint.

(with $n = 30$), and consider the optimization problem

$$\begin{aligned}
\text{minimize} \quad & \tau \left\| \sum_{k=1}^r a_k x_k^H - C \right\|_F + \sum_{k=1}^r \|x_k\|_2 \\
\text{subject to} \quad & a_k \in \mathcal{D}, \quad k = 1, \dots, r.
\end{aligned} \tag{17}$$

with τ a positive parameter. In the example, we use $\alpha = 0$, $\beta = \pi/12 = 0.2618$. Using theorem 2.1, the problem can be converted to the SDP

$$\begin{aligned}
\text{minimize} \quad & \tau \|Y - C\|_F + (\text{tr}(X) + \text{tr}(Z))/2 \\
\text{subject to} \quad & \begin{bmatrix} X & Y \\ Y^H & Z \end{bmatrix} \succeq 0 \\
& X \text{ is Toeplitz} \\
& -FXG^H - GXF^H + 2 \cos \beta GXG^H \preceq 0.
\end{aligned}$$

In the second method, we omit the third constraint (imposing the interval), *i.e.*, solve (17) with the dictionary (2). Figure 2 shows the two solutions, for independently tuned values of τ , with the estimates of ω_i on the horizontal axis, and the norms of the vectors x_k on the vertical axis. As can be seen, adding the interval constraints allowed the method to identify the parameters ω_i , and thus $\|x_k\|$, more accurately.

4. CONCLUSION

The proof technique and the factorization result in lemma 2.1 appear in the control literature on the generalized Kalman-Yakubovich-Popov lemma and its connections with SDP duality [23, 26–28]. In these applications, F and G are defined as

$$F = \begin{bmatrix} A & B \end{bmatrix}, \quad G = \begin{bmatrix} I & 0 \end{bmatrix},$$

and the vectors a in (8) take the form $a = ((\lambda I - A)^{-1} B u, u)$. The link is important for two reasons. First, it suggests a range of similar applications in system and control theory, statistics, and numerical analysis, with different choices for the matrices F, G , and the sets \mathcal{C} . Second, specialized techniques for solving SDPs derived from the Kalman-Yakubovich-Popov lemma, for example, by exploiting real symmetries and rank-one structure [29–32], will be useful in the development of fast solvers for SDPs of the form (13).

5. REFERENCES

- [1] R. Tibshirani, "Regression shrinkage and selection via the Lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.
- [2] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, pp. 33–61, 1998.
- [3] J. A. Tropp, "Just relax: Convex programming methods for identifying sparse signals in noise," *IEEE Transactions on Information Theory*, vol. 52, no. 3, pp. 1030–1051, 2006.
- [4] D. L. Donoho and M. Elad, "On the stability of the basis pursuit in the presence of noise," *Signal Processing*, vol. 86, no. 3, pp. 511 – 532, 2006.
- [5] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, 2006.
- [6] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [7] M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*, Springer, 2010.
- [8] T. Hastie, R. Tibshirani, and M. Wainwright, *Statistical Learning with Sparsity. The Lasso and Generalizations*, CRC Press, 2015.
- [9] J. F. Claerbout and F. Muir, "Robust modeling with erratic data," *Geophysics*, vol. 38, no. 5, pp. 826–844, 1973.
- [10] F. Santosa and W. H. Symes, "Linear inversion of band-limited reflection seismograms," *SIAM Journal on Scientific and Statistical Computing*, vol. 7, no. 4, pp. 1307–1330, 1986.
- [11] P. Shah, B. N. Bhaskar, G. Tang, and B. Recht, "Linear system identification via atomic norm regularization," in *Proceedings of the 51st IEEE Conference on Decision and Control*, 2012, pp. 6265–6270.
- [12] E. J. Candès and C. Fernandez-Granda, "Towards a mathematical theory of super-resolution," *Communications of 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 and B. Recht, "Atomic norm denoising with applications to line spectral estimation," in *Communication, Control, and Computing (Allerton), 2011 49th Annual Allerton Conference on*, Sept 2011, pp. 261–268.
- [15] E. J. Candès and C. Fernandez-Granda, "Super-resolution from noisy data," *Journal of Fourier Analysis and Applications*, vol. 19, pp. 1229–1254, 2013.
- [16] 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, June 2015.
- [17] C. Fernandez-Granda, "Super-resolution of point sources via convex programming," 2015, arXiv:1507.07034.
- [18] B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM Review*, vol. 52, no. 3, pp. 471–501, 2010.
- [19] Y. Li and Y. Chi, "Off-the-grid line spectrum denoising and estimation with multiple measurement vectors," 2014, arXiv:cs/0412009.
- [20] Z. Yang and L. Xie, "Exact joint sparse frequency recovery via optimization methods," 2014, arXiv:cs/1405.6585.
- [21] P. Stoica and R. L. Moses, *Introduction to Spectral Analysis*, Prentice Hall, London, 1997.
- [22] T. Iwasaki and S. Hara, "Generalized KYP lemma: unified frequency domain inequalities with design applications," *IEEE Transactions on Automatic Control*, vol. 50, no. 1, pp. 41–59, 2005.
- [23] G. Pipeleers and L. Vandenberghe, "Generalized KYP lemma with real data," *IEEE Transactions on Automatic Control*, vol. 56, no. 12, pp. 2942–2946, 2011.
- [24] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antennas and Propagation*, vol. 34, no. 3, pp. 276–280, Mar 1986.
- [25] A. Paulraj, R. Roy, and T. Kailath, "A subspace rotation approach to signal parameter estimation," *Proceedings of the IEEE*, vol. 74, no. 7, pp. 1044–1046, July 1986.
- [26] A. Rantzer, "On the Kalman-Yakubovich-Popov lemma," *Systems and Control Letters*, vol. 28, no. 1, pp. 7–10, 1996.
- [27] T. Iwasaki, G. Meinsma, and M. Fu, "Generalized S -procedure and finite-frequency KYP lemma," *Mathematical Problems in Engineering*, vol. 6, pp. 305–320, 2000.
- [28] V. Balakrishnan and L. Vandenberghe, "Semidefinite programming duality and linear time-invariant systems," *IEEE Transactions on Automatic Control*, vol. 48, pp. 30–41, 2003.
- [29] Y. Genin, Y. Hachez, Yu. Nesterov, and P. Van Dooren, "Optimization problems over positive pseudopolynomial matrices," *SIAM Journal on Matrix Analysis and Applications*, vol. 25, no. 1, pp. 57–79, 2003.
- [30] J. Löfberg and P. A. Parrilo, "From coefficients to samples: a new approach to SOS optimization," in *Proceedings of the 43rd IEEE Conference on Decision and Control*, 2004, pp. 3154–3159.
- [31] T. Roh and L. Vandenberghe, "Discrete transforms, semidefinite programming, and sum-of-squares representations of non-negative polynomials," *SIAM Journal on Optimization*, vol. 16, no. 4, pp. 939–964, 2006.
- [32] Z. Liu and L. Vandenberghe, "Low-rank structure in semidefinite programs derived from the KYP lemma," in *Proceedings of the 46th IEEE Conference on Decision and Control*, 2007, pp. 5652–5659.