SIAM J. Optim.
Vol. 27, No. 3, pp. 1362–1389

SEMIDEFINITE REPRESENTATIONS OF GAUGE FUNCTIONS FOR STRUCTURED LOW-RANK MATRIX DECOMPOSITION*

HSIAO-HAN CHAO† AND LIEVEN VANDENBERGHE†

Abstract. This paper presents generalizations of semidefinite programming formulations of 1-norm optimization problems over infinite dictionaries of vectors of complex exponentials, which were recently proposed for superresolution, gridless compressed sensing, and other applications in signal processing. Results related to the generalized Kalman–Yakubovich–Popov lemma in linear system theory provide simple, constructive proofs of the semidefinite representations of the penalty functions used in these applications. The connection leads to several extensions to gauge functions and atomic norms for sets of vectors parameterized via the nullspace of matrix pencils. The techniques are illustrated with examples of low-rank matrix approximation problems arising in spectral estimation and array processing.

Key words. semidefinite programming, atomic norm, Kalman–Yakubovich–Popov lemma, matrix pencils

AMS subject classifications. 90C22, 90C25, 93B60, 93B30

DOI. 10.1137/16M1071730

1. Introduction.

The notion of the atomic norm introduced in [5] gives a unified description of convex penalty functions that extend the ℓ_1-norm penalty, used to promote sparsity in the solution of an optimization problem, to various other types of structure. The atomic norm associated with a nonempty set C is defined as the gauge of its convex hull, i.e., the convex function

$$\begin{aligned} g(x) &= \inf \{t \geq 0 \mid x \in t \operatorname{conv} C\} \\ &= \inf \left\{ \sum_{k=1}^{r} \theta_k \mid x = \sum_{k=1}^{r} \theta_k a_k,\ \theta_k \geq 0,\ a_k \in C \right\}. \end{aligned} \tag{1}$$

This function is convex, nonnegative, and positively homogeneous. It is not necessarily a norm, but it is common to use the term "atomic norm" even when g is not a norm. When used as a regularization term in an optimization problem, the function $g(x)$ defined in (1) promotes the property that x can be expressed as a nonnegative linear combination of a small number of elements (or "atoms") of C.

The best-known examples of atomic norms are the vector ℓ_1-norm and the matrix trace norm. The ℓ_1-norm of a real or complex n-vector is the atomic norm associated with $C = \{se_k \mid |s| = 1,\ k = 1, \ldots, n\}$, where e_k is the kth unit vector of length n. The matrix trace norm (or nuclear norm) is the atomic norm for the set of rank-1 matrices with unit norm. Specifically, the trace norm on $\mathbf{C}^{n\times m}$ is the atomic norm for $C = \{vw^H \mid v \in \mathbf{C}^n,\ w \in \mathbf{C}^m,\ \|v\| = \|w\| = 1\}$, where w^H is the conjugate transpose and $\|\cdot\|$ denotes the Euclidean norm. Many other examples are discussed in [3, 5, 42].

The atomic norm associated with the set

$$C_{\mathrm{e}} = \{\gamma\,(1, e^{\mathrm{j}\omega}, \ldots, e^{\mathrm{j}(n-1)\omega}) \in \mathbf{C}^n \mid \omega \in [0, 2\pi),\ |\gamma| = 1/\sqrt{n}\}, \tag{2}$$

*Received by the editors April 21, 2016; accepted for publication (in revised form) February 7, 2017; published electronically July 13, 2017.
http://www.siam.org/journals/siopt/27-3/M107173.html
Funding: Research partially supported by NSF grants 1128817 and 1509789, and NIH grant 5R01GM107639.
†Electrical Engineering Department, University of California, Los Angeles, Los Angeles, CA 90095 (suzihchao@ucla.edu, vandenbe@ucla.edu).

where $j = \sqrt{-1}$, has been studied extensively in recent research in signal processing [3, 4, 8, 9, 10, 27, 42, 45]. It is known that the atomic norm for this set is the optimal value of the semidefinite program (SDP)

$$\begin{array}{ll} \text{minimize} & (\mathbf{tr}\, V + w)/2 \\ \text{subject to} & \begin{bmatrix} V & x \\ x^H & w \end{bmatrix} \succeq 0, \;\; V \text{ is Toeplitz}, \end{array} \tag{3}$$

with variables w and $V \in \mathbf{H}^n$ (the $n \times n$ Hermitian matrices). This result can be proved via convex duality and semidefinite characterizations of bounded trigonometric polynomials [3, 4, 8], or directly by referring to the classical Vandermonde decomposition of positive semidefinite Toeplitz matrices as a positive sum of the outer product of vectors in C_e [41, 42]. More generally, one can consider the atomic norm of the set of matrices $C = \{vw^H \in \mathbf{C}^{n\times m} \mid v \in C_\mathrm{e},\ \|w\| = 1\}$. The atomic norm for this set, evaluated at a matrix $X \in \mathbf{C}^{n\times m}$, is the optimal value of the SDP

$$\begin{array}{ll} \text{minimize} & (\mathbf{tr}\, V + \mathbf{tr}\, W)/2 \\ \text{subject to} & \begin{bmatrix} V & X \\ X^H & W \end{bmatrix} \succeq 0, \;\; V \text{ is Toeplitz}, \end{array} \tag{4}$$

with variables $V \in \mathbf{H}^n$ and $W \in \mathbf{H}^m$; see [10, 27, 45]. Further extensions, which place restrictions on the parameter ω in the definition (2), can be found in [30, 31].

In this paper we discuss extensions of the SDP representations (3) and (4) to a larger class of atomic norms and gauge functions. The starting point is the observation that C_e can be parameterized as

$$C_\mathrm{e} = \{a \mid (\lambda G - F)a = 0,\ \lambda \in \mathcal{C},\ \|a\| = 1\}, \tag{5}$$

where $\mathcal{C}$ is the unit circle in the complex plane, and F and G are the $(n-1) \times n$ matrices

$$F = \begin{bmatrix} 0 & I_{n-1} \end{bmatrix}, \qquad G = \begin{bmatrix} I_{n-1} & 0 \end{bmatrix}.$$

We generalize (5) in three ways and derive semidefinite representations of the corresponding atomic norms. The first generalization is to allow F and G to be arbitrary matrices of equal size, i.e., to replace $\lambda G - F$ with an arbitrary matrix pencil (a matrix polynomial of degree 1). Second, we allow $\mathcal{C}$ to be an arbitrary circle or line in the complex plane, or a segment of a line or a circle. Third, we replace the normalization $\|a\| = 1$ with a condition of the type $\|Ea\| \le 1$, where E is not necessarily full column rank. Specific examples of these extensions, with different choices of F, G, and $\mathcal{C}$, are discussed in sections 2.2–2.4.

We present direct, constructive proofs, based on elementary matrix algebra, of the semidefinite representations of the atomic norms. These results are the subject of sections 2 and 3, and Appendix B. In section 4 we derive the convex conjugates of the atomic norms and gauge functions, and discuss the relation between the dual SDP representations and the Kalman–Yakubovich–Popov (KYP) lemma from linear system theory. Appendix C contains a discussion of the properties of the matrix pencil $\lambda F - G$ that are needed to ensure strong duality in the dual problems. In section 5 the SDP formulations are illustrated with several applications in signal processing.

2. Positive semidefinite matrix factorization. Throughout the paper we assume that F and G are complex matrices of size $p \times n$, and Φ and Ψ are Hermitian 2×2 matrices with $\det \Phi < 0$. We define

$$\mathcal{A} = \{a \in \mathbf{C}^n \mid (\mu G - \nu F)a = 0,\ (\mu, \nu) \in \mathcal{C}\}, \tag{6}$$

where

$$\mathcal{C} = \left\{(\mu, \nu) \in \mathbf{C}^2 \mid (\mu, \nu) \neq 0,\ q_\Phi(\mu, \nu) = 0,\ q_\Psi(\mu, \nu) \leq 0\right\}. \tag{7}$$

Here q_Φ, q_Ψ are the quadratic forms defined by Φ and Ψ:

$$q_\Phi(\mu, \nu) = \begin{bmatrix} \mu \\ \nu \end{bmatrix}^H \Phi \begin{bmatrix} \mu \\ \nu \end{bmatrix}, \qquad q_\Psi(\mu, \nu) = \begin{bmatrix} \mu \\ \nu \end{bmatrix}^H \Psi \begin{bmatrix} \mu \\ \nu \end{bmatrix}. \tag{8}$$

The set $\mathcal{C}$ is a subset of a line or circle in the complex plane, expressed in homogeneous coordinates, as explained in Appendix A. If $\Phi_{11} \neq 0$ or $\Psi_{11} > 0$, then $\nu \neq 0$ for all elements $(\mu, \nu) \in \mathcal{C}$, and we can simplify the definition of $\mathcal{A}$ as

$$\mathcal{A} = \{a \in \mathbf{C}^n \mid (\lambda G - F)a = 0,\ (\lambda, 1) \in \mathcal{C}\}. \tag{9}$$

If $\Phi_{11} = 0$ and $\Psi_{11} \leq 0$, then the pair $(1, 0)$ is also in $\mathcal{C}$ and the set $\mathcal{A}$ in (6) is the union of the right-hand side of (9) and the nullspace of G. Examples of sets $\mathcal{A}$ are given in sections 2.2–2.4.

The purpose of this section is to discuss a semidefinite representation of the convex hull of the set of matrices aa^H with $a \in \mathcal{A}$:

$$\operatorname{conv}\left\{aa^H \mid a \in \mathcal{A}\right\} = \left\{\sum_{k=1}^r a_k a_k^H \mid a_k \in \mathcal{A},\ k = 1, \dots, r\right\}. \tag{10}$$

2.1. Conic decomposition. The key result (Theorem 2.1) is known in various forms in system theory, signal processing, and moment theory [17, 25, 26]. Our purpose is to give a simple semidefinite formulation that encompasses a wide variety of interesting special cases, and to present a constructive proof that can be implemented using the basic decompositions of numerical linear algebra (specifically, symmetric eigenvalue, singular value, and Schur decompositions).

THEOREM 2.1. *Let $\mathcal{A}$ be defined by* (6) *and* (7)*, where F, $G \in \mathbf{C}^{p\times n}$ and Φ, $\Psi \in \mathbf{H}^2$ with $\det \Phi < 0$. A matrix $X \in \mathbf{H}^n$ is positive semidefinite of rank $r \geq 1$ and satisfies*

$$\Phi_{11} FXF^H + \Phi_{21} FXG^H + \Phi_{12} GXF^H + \Phi_{22} GXG^H = 0, \tag{11}$$

$$\Psi_{11} FXF^H + \Psi_{21} FXG^H + \Psi_{12} GXF^H + \Psi_{22} GXG^H \preceq 0 \tag{12}$$

if and only if X can be decomposed as $X = \sum_{k=1}^r a_k a_k^H$ with linearly independent vectors $a_1, \dots, a_r \in \mathcal{A}$.

Proof. Sufficiency is readily proved by substituting $X = \sum_{k=1}^r a_k a_k^H$ in (11) and (12), and verifying that if $(\mu_k G - \nu_k F)a_k = 0$ with $(\mu_k, \nu_k) \neq 0$, then

$$\Phi_{11} FXF^H + \Phi_{21} FXG^H + \Phi_{12} GXF^H + \Phi_{22} GXG^H = \sum_{k=1}^r \alpha_k q_\Phi(\mu_k, \nu_k) y_k y_k^H,$$

$$\Psi_{11} FXF^H + \Psi_{21} FXG^H + \Psi_{12} GXF^H + \Psi_{22} GXG^H = \sum_{k=1}^r \alpha_k q_\Psi(\mu_k, \nu_k) y_k y_k^H,$$

where $\alpha_k = 1/|\nu_k|^2$, $y_k = Ga_k$ if $\nu_k \neq 0$, and $\alpha_k = 1/|\mu_k|^2$, $y_k = Fa_k$ if $\nu_k = 0$.

To show necessity, we start from any factorization $X = YY^H$ where $Y \in \mathbf{C}^{n\times r}$ has rank r. It follows from Lemma B.2 in Appendix B, applied to $U = FY$ and

$V = GY$, that there exist a matrix $W \in \mathbf{C}^{p\times r}$, a unitary matrix $Q \in \mathbf{C}^{r\times r}$, and two vectors $\mu, \nu \in \mathbf{C}^r$ such that

$$FYQ = W\,\mathbf{diag}(\mu), \qquad GYQ = W\,\mathbf{diag}(\nu), \qquad (\mu_i, \nu_i) \in \mathcal{C}, \quad i = 1, \ldots, r.$$

Choosing a_k equal to the kth column of YQ gives the decomposition of X. ◻

Viewed geometrically, the theorem says that (10) is the set of positive semidefinite matrices X that satisfy (11) and (12).

It is useful to note that the proof of Lemma B.2 in Appendix B is constructive and gives a simple algorithm, based on singular value and Schur decompositions, for computing the matrices W, Q and the vectors μ, ν.

2.2. Trigonometric polynomials. In this and the following two sections we illustrate the decomposition in Theorem 2.1 with different choices of F, G, Φ, and Ψ.

Complex exponentials. As a first example, we take $p = n - 1$,

$$F = \begin{bmatrix} 0 & I_{n-1} \end{bmatrix}, \quad G = \begin{bmatrix} I_{n-1} & 0 \end{bmatrix}, \quad \Phi = \Phi_{\mathrm{u}} \triangleq \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \quad \Psi = 0. \tag{13}$$

A nonzero pair (μ, ν) satisfies $q_\Phi(\mu, \nu) = |\mu|^2 - |\nu|^2 = 0$ only if μ and ν are nonzero and $\lambda = \mu/\nu$ is on the unit circle. The condition $(\lambda G - F)a = 0$ in the definition of $\mathcal{A}$ gives a recursion $\lambda a_1 = a_2$, $\lambda a_2 = a_3, \ldots, \lambda a_{n-1} = a_n$. Defining $\exp(\mathrm{j}\omega) = \lambda$, we find that $\mathcal{A}$ contains the vectors

$$a = c\,(1, e^{\mathrm{j}\omega}, e^{\mathrm{j}2\omega}, \ldots, e^{\mathrm{j}(n-1)\omega}) \tag{14}$$

for all $\omega \in [0, 2\pi)$ and $c \in \mathbf{C}$. The matrix constraints (11)–(12) reduce to $FXF^H = GXG^H$, i.e., X is a Toeplitz matrix. Theorem 2.1 therefore reduces to the well-known fact that every $n \times n$ positive semidefinite Toeplitz matrix can be decomposed as

$$X = \sum_{k=1}^{r} |c_k|^2 \begin{bmatrix} 1 \\ e^{\mathrm{j}\omega_k} \\ \vdots \\ e^{\mathrm{j}(n-1)\omega_k} \end{bmatrix} \begin{bmatrix} 1 \\ e^{\mathrm{j}\omega_k} \\ \vdots \\ e^{\mathrm{j}(n-1)\omega_k} \end{bmatrix}^H \tag{15}$$

with $c_k \neq 0$ and distinct $\omega_1, \ldots, \omega_r$ [40, page 170].

Restricted complex exponentials. Define F, G, and Φ as in (13), and

$$\Psi = \begin{bmatrix} 0 & -e^{\mathrm{j}\alpha} \\ -e^{-\mathrm{j}\alpha} & 2\cos\beta \end{bmatrix}$$

with $\alpha \in [0, 2\pi)$ and $\beta \in [0, \pi)$. The elements $a \in \mathcal{A}$ have the same general form (14), with the added constraint that $\cos\beta \le \cos(\omega - \alpha)$. Since we can restrict ω to the interval $[\alpha - \pi, \alpha + \pi]$, this is equivalent to $|\omega - \alpha| \le \beta$. The constraints (11)–(12) specify that X is Toeplitz and satisfies the matrix inequality

$$-e^{-\mathrm{j}\alpha} FXG^H - e^{\mathrm{j}\alpha} GXF^H + 2(\cos\beta) GXG^H \preceq 0. \tag{16}$$

The theorem states that a positive semidefinite Toeplitz matrix of rank r satisfies (16) if and only if it can be decomposed as (15) with nonzero c_k and $|\omega_k - \alpha| \le \beta$.

Real trigonometric functions. Next consider $p = n-1$,

$$G = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 2 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 2 & 0 \end{bmatrix}, \qquad F = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 & 0 & 0 \\ 1 & 0 & 1 & \cdots & 0 & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 & 1 \end{bmatrix},$$

and

$$\Phi = \Phi_{\mathrm{r}} \triangleq \begin{bmatrix} 0 & \mathrm{j} \\ -\mathrm{j} & 0 \end{bmatrix}, \qquad \Psi = \Phi_{\mathrm{u}} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

A nonzero pair (μ, ν) satisfies $q_\Phi(\mu,\nu) = \mathrm{j}(\bar{\mu}\nu - \mu\bar{\nu}) = 0$ and $q_\Psi(\mu,\nu) = |\mu|^2 - |\nu|^2 \le 0$ only if $\nu \neq 0$ and $\lambda = \mu/\nu$ is real with $|\lambda| \le 1$. The condition $(\lambda G - F)a = 0$ gives a recursion $\lambda a_1 = a_2$, $2\lambda a_k = a_{k-1} + a_{k+1}$ for $k = 2, \ldots, n-1$. If we write $\lambda = \cos\omega$, we recognize the recursion $2\cos\omega\cos k\omega = \cos{(k-1)\omega} + \cos{(k+1)\omega}$ and find that $\mathcal{A}$ contains the vectors

$$a = c\,(1,\, \cos\omega,\, \cos 2\omega, \ldots,\, \cos{(n-1)\omega})$$

for all $\omega \in [0, 2\pi)$ and all c. With the same F and $G = [\ 2I_{n-1} \quad 0\ ]$, the condition $(\lambda G - F)a = 0$ reduces to $2\lambda a_1 = a_2$, $2\lambda a_k = a_{k-1} + a_{k+1}$ for $k = 2, \ldots, n-1$. If we write $\lambda = \cos\omega$, the solutions are the vectors

$$a = c\,\left(1, \frac{\sin 2\omega}{\sin\omega}, \frac{\sin 3\omega}{\sin\omega}, \ldots, \frac{\sin n\omega}{\sin\omega}\right)$$

for all $\omega \in [0, 2\pi)$ and all c.

2.3. Polynomials.

Real powers. Next, define F, G as in (13), and $\Phi = \Phi_{\mathrm{r}}$, $\Psi = 0$. A pair (μ, ν) satisfies $q_\Phi(\mu,\nu) = 0$ if and only if $\bar{\mu}\nu$ is real. If $(\mu,\nu) \neq 0$, we either have $\nu = 0$ and μ arbitrary, or $\nu \neq 0$ and $\lambda = \mu/\nu$ real. The set $\mathcal{A}$ therefore contains the vectors

$$a = c\,(1, \lambda, \lambda^2, \ldots, \lambda^{n-1}), \qquad a = c\,(0, 0, \ldots, 0, 1)$$

for all $\lambda \in \mathbf{R}$ and c. The matrix constraints (11)–(12) reduce to $FXG^H = GXF^H$, i.e., X is a symmetric (real) Hankel matrix.

Restricted polynomials. If F, G are defined as in (13) and $\Phi = \Phi_{\mathrm{r}}$,

$$\Psi = \begin{bmatrix} 2 & -(\alpha+\beta) \\ -(\alpha+\beta) & 2\alpha\beta \end{bmatrix},$$

where $-\infty < \alpha < \beta < \infty$, then $\mathcal{A}$ contains all vectors $a = c(1, \lambda, \ldots, \lambda^{n-1})$ with $\lambda \in [\alpha, \beta]$. The matrix constraints require X to be a real symmetric Hankel matrix that satisfies

$$2FXF^H - (\alpha+\beta)(FXG^H + GXF^H) + 2\alpha\beta GXG^H \preceq 0.$$

Orthogonal polynomials. Let $p_0(\lambda)$, $p_1(\lambda)$, $p_2(\lambda)$, ... be a sequence of real polynomials on $\mathbf{R}$ with p_i of degree i. It is well known that the polynomials are orthonormal with respect to an inner product that satisfies the property

$$\langle f(\lambda), \lambda g(\lambda)\rangle = \langle \lambda f(\lambda), g(\lambda)\rangle \tag{17}$$

(for example, an inner product of the form $\langle f, g\rangle = \int f(\lambda)g(\lambda)w(\lambda)d\lambda$ with $w(\lambda) \geq 0$) if and only if the polynomials satisfy a three-term recurrence

$$\beta_{i+1}p_{i+1}(\lambda) = (\lambda - \alpha_i)p_i(\lambda) - \beta_i p_{i-1}(\lambda) \tag{18}$$

with $p_{-1}(\lambda) = 0$ and $p_0(\lambda) = 1/d_0$, where $d_0^2 = \langle 1, 1\rangle$. This can be seen as follows [15].

Suppose $p_0, \dots, p_{n-1}$ is any set of polynomials with p_i of degree i. Then $\lambda p_i(\lambda)$ can be expressed as a linear combination of the polynomials $p_0(\lambda), \dots, p_{i+1}(\lambda)$, and therefore

$$\lambda \begin{bmatrix} p_0(\lambda) \\ p_1(\lambda) \\ \vdots \\ p_{n-2}(\lambda) \end{bmatrix} = \begin{bmatrix} J & \beta_{n-1}e_{n-1} \end{bmatrix} \begin{bmatrix} p_0(\lambda) \\ p_1(\lambda) \\ \vdots \\ p_{n-1}(\lambda) \end{bmatrix} \tag{19}$$

for some lower-Hessenberg matrix J (i.e., satisfying $J_{ij} = 0$ for $j > i + 1$). Let $\langle \cdot, \cdot \rangle$ be an inner product on the space of polynomials of degree $n - 1$ or less. Taking inner products on both sides of (19), we find that

$$H = JG + \beta_{n-1}e_{n-1}g^T,$$

where $H_{ij} = \langle \lambda p_{i-1}(\lambda), p_{j-1}(\lambda)\rangle$, $G_{ij} = \langle p_{i-1}(\lambda), p_{j-1}(\lambda)\rangle$, and $g_j = \langle p_{n-1}(\lambda), p_{j-1}(\lambda)\rangle$ for $i, j = 1, \dots, n - 1$. The polynomials are orthonormal for the inner product if and only if $G = I$ and $g = 0$. The inner product satisfies the property (17) if and only if H is symmetric. Hence, if the polynomials are orthonormal for an inner product that satisfies (17), then J is a symmetric tridiagonal matrix. If we use the notation

$$J = \begin{bmatrix} \alpha_0 & \beta_1 & 0 & \cdots & 0 & 0 \\ \beta_1 & \alpha_1 & \beta_2 & \cdots & 0 & 0 \\ 0 & \beta_2 & \alpha_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \alpha_{n-3} & \beta_{n-2} \\ 0 & 0 & 0 & \cdots & \beta_{n-2} & \alpha_{n-2} \end{bmatrix}, \tag{20}$$

the recurrence (18) follows. Conversely, if the three-term recurrence holds, and we define the inner product by setting $G = I$, $g = 0$, then H is symmetric and the inner product satisfies (17).

Now consider (6) and (7), with $p = n - 1$ and

$$\Phi = \Phi_{\mathrm{r}}, \qquad \Psi = 0, \qquad G = \begin{bmatrix} I_{n-1} & 0 \end{bmatrix}, \qquad F = \begin{bmatrix} J & \beta_{n-1}e_{n-1} \end{bmatrix},$$

where J is the Jacobi matrix (20) of a system of orthogonal polynomials. Then $(\mu, \nu) \in \mathcal{C}$ if and only if either $\nu \neq 0$ and $\lambda = \mu/\nu \in \mathbf{R}$, or $\nu = 0$. The set contains the vectors a of the following form for all $\lambda \in \mathbf{R}$:

$$a = c\,(p_0(\lambda), p_1(\lambda), p_2(\lambda), \dots, p_{n-1}(\lambda)), \qquad a = c\,(0, 0, \dots, 0, 1).$$

2.4. Rational functions. As a final example, we consider the controllability pencil of a linear system

$$G = \begin{bmatrix} I & 0 \end{bmatrix}, \qquad F = \begin{bmatrix} A & B \end{bmatrix}, \tag{21}$$

where $A \in \mathbf{C}^{n_{\mathrm{s}} \times n_{\mathrm{s}}}$ and $B \in \mathbf{C}^{n_{\mathrm{s}} \times m}$. With this choice, $\mathcal{A}$ contains the vectors $a = (x, u)$ that satisfy the equality $(\mu I - \nu A)x = \nu B u$ for some $(\mu, \nu) \in \mathcal{C}$. Since $(\mu, \nu) \neq 0$, we either have $\nu = 0$ and $x = 0$, or $\nu \neq 0$ and $((\mu/\nu)I - A)x = Bu$. If A has no eigenvalues λ that satisfy $(\lambda, 1) \in \mathcal{C}$, then $\mathcal{A}$ contains the vectors

$$a = \begin{bmatrix} (\lambda I - A)^{-1} B u \\ u \end{bmatrix}$$

for all $(\lambda, 1) \in \mathcal{C}$ and all $u \in \mathbf{C}^m$. If $\mathcal{C}$ includes the point $(1, 0)$ at infinity, then $\mathcal{A}$ also contains the vectors $(0, u)$ for all $u \in \mathbf{C}^m$.

This can be extended to the controllability pencil of a descriptor system

$$G = \begin{bmatrix} E & 0 \end{bmatrix}, \qquad F = \begin{bmatrix} A & B \end{bmatrix},$$

where $E \in \mathbf{C}^{n_s \times n_s}$ is possibly singular. With this choice, $\mathcal{A}$ contains the vectors $a = (x, u)$ that satisfy the equality $(\mu E - \nu A)x = \nu B u$ for some $(\mu, \nu) \in \mathcal{C}$. If $\det(\mu E - \nu A) \neq 0$ for all $(\mu, \nu) \in \mathcal{C}$, then $\mathcal{A}$ contains all vectors

$$a = \begin{bmatrix} (\lambda E - A)^{-1} B u \\ u \end{bmatrix}$$

for $(\lambda, 1) \in \mathcal{C}$ and $u \in \mathbf{C}^m$. If $(1, 0) \in \mathcal{C}$, then $\mathcal{A}$ also contains $(0, u)$ for all $u \in \mathbf{C}^m$.

3. Semidefinite representation of gauges and atomic norms.

A function g is called a *gauge* if it is convex, positively homogeneous ($g(tx) = tg(x)$ for $t > 0$), nonnegative, and vanishes at the origin [36, section 15], [26, Chapter 1]. Examples are the *(Minkowski) gauges* of nonempty convex sets C, which are defined as

$$g(x) = \inf \{t \geq 0 \mid x \in tC\}.$$

Conversely, if g is a gauge, then it is the Minkowski gauge of the set $C = \{x \mid g(x) \leq 1\}$. A gauge is a norm if it is defined everywhere, positive except at the origin, and symmetric ($g(x) = g(-x)$).

The gauge of the convex hull $\operatorname{conv} C$ of a set C can be expressed as

$$g(x) = \inf \left\{ \sum_{k=1}^{r} \theta_k \mid x = \sum_{k=1}^{r} \theta_k x_k,\ \theta_k \geq 0,\ x_k \in C,\ k = 1, \ldots, r \right\}.$$

The minimum is over all possible decompositions of x as a nonnegative combination of a finite number of elements of C. The gauge of the convex hull of a compact set is also called the *atomic norm* associated with the set [5].

3.1. Symmetric matrices.

Let F, G, Φ, and Ψ be defined as in Theorem 2.1. We assume that the set $\mathcal{C}$ defined in (7) is not empty. In this section we discuss the gauge of the convex hull of the set

$$C = \{aa^H \in \mathbf{H}^n \mid a \in \mathcal{A},\ \|a\| = 1\},$$

where $\mathcal{A}$ is defined in (6). The gauge of the convex hull of C is the function

$$g(X) = \inf \left\{ \sum_{k=1}^{r} \theta_k \mid X = \sum_{k=1}^{r} \theta_k a_k a_k^H,\ \theta_k \geq 0,\ a_k \in \mathcal{A},\ \|a_k\| = 1 \right\} \tag{22}$$

$$= \inf \left\{ \sum_{k=1}^{r} \|a_k\|^2 \mid X = \sum_{k=1}^{r} a_k a_k^H,\ a_k \in \mathcal{A} \right\}. \tag{23}$$

The second expression follows from the fact that if $a \in \mathcal{A}$, then $\beta a \in \mathcal{A}$ for all β.

The expressions $\sum_k \theta_k$ and $\sum_k \|a_k\|^2$ in these minimizations take only two possible values: $\mathbf{tr}\, X$ if X can be decomposed as in (22) and (23), and $+\infty$ otherwise. Theorem 2.1 tells us that a decomposition exists if and only if X is positive semidefinite and satisfies the two constraints (11), (12). Therefore,

$$g(X) = \begin{cases} \mathbf{tr}\, X, & X \succeq 0,\ (11),\ (12), \\ +\infty & \text{otherwise}. \end{cases} \tag{24}$$

Now consider an optimization problem in which we minimize the sum of a function $f : \mathbf{H}^n \to \mathbf{R}$ and the gauge defined in (23) and (24):

$$\text{minimize} \quad f(X) + g(X). \tag{25}$$

If we substitute the definition (23), this can be written as

$$\begin{array}{ll} \text{minimize} & f(X) + \sum_{k=1}^{r} \|a_k\|^2 \\ \text{subject to} & X = \sum_{k=1}^{r} a_k a_k^H, \quad a_k \in \mathcal{A}, \quad k = 1, \ldots, r. \end{array} \tag{26}$$

The variables are X and the parameters $a_1, \ldots, a_r$, and r of the decomposition of X. This formulation shows that the function $g(X)$ in (25) acts as a regularization term that promotes a structured low-rank property in X. If we substitute the expression (24), we obtain the equivalent formulation

$$\begin{array}{ll} \text{minimize} & f(X) + \mathbf{tr}\, X \\ \text{subject to} & \Phi_{11} F X F^H + \Phi_{21} F X G^H + \Phi_{12} G X F^H + \Phi_{22} G X G^H = 0, \\ & \Psi_{11} F X F^H + \Psi_{21} F X G^H + \Psi_{12} G X F^H + \Psi_{22} G X G^H \preceq 0, \\ & X \succeq 0. \end{array} \tag{27}$$

This problem is convex if f is convex.

A useful generalization of (23) is the gauge of the convex hull of

$$C = \{ a a^H \mid a \in \mathcal{A},\ \|Ea\| \le 1 \},$$

where E may have rank less than n. The gauge of $\operatorname{conv} C$ is

$$g(X) = \inf \left\{ \sum_{k=1}^{r} \theta_k \mid X = \sum_{k=1}^{r} \theta_k a_k a_k^H,\ \theta_k \ge 0,\ a_k \in \mathcal{A},\ \|E a_k\| \le 1 \right\}. \tag{28}$$

The variables θ_k in this definition can be eliminated by making the following observation. Suppose that the directions of the vectors a_k in the decomposition of X in (28) are given, but not their norms or the coefficients θ_k. If $0 < \|Ea_k\| < 1$, then we can decrease θ_k by scaling a_k until $\|Ea_k\| = 1$. If $Ea_k = 0$, then θ_k can be made arbitrarily small by scaling a_k. Hence, we obtain the same result if we use $\sqrt{\theta_k} a_k$ as variables and write the infimum as

$$g(X) = \inf \left\{ \sum_{k=1}^{r} \|E a_k\|^2 \mid X = \sum_{k=1}^{r} a_k a_k^H,\ a_k \in \mathcal{A},\ k = 1, \ldots, r \right\}. \tag{29}$$

Therefore, $g(X) = \sum_k \|Ea_k\|^2 = \mathbf{tr}(EXE^H)$ if X can be decomposed as in (29) and $+\infty$ otherwise. Using Theorem 2.1, we can express this result as

$$g(X) = \begin{cases} \mathbf{tr}(EXE^H), & X \succeq 0,\ (11),\ (12), \\ +\infty & \text{otherwise}. \end{cases} \tag{30}$$

Minimizing $f(X) + g(X)$ is equivalent to the optimization problem

$$\begin{array}{ll} \text{minimize} & f(X) + \sum_{k=1}^{r} \|Ea_k\|^2 \\ \text{subject to} & X = \sum_{k=1}^{r} a_k a_k^H, \quad a_k \in \mathcal{A}, \quad k = 1, \ldots, r, \end{array} \tag{31}$$

with variables X and the parameters $a_1, \ldots, a_r$, r of the decomposition of X. When $E^H E = I$, this is the same as (26). By choosing different E we assign different weights to the vectors a_k. Using the expression (30), the problem (31) can be written as

$$\begin{array}{ll} \text{minimize} & f(X) + \mathbf{tr}\,(EXE^H) \\ \text{subject to} & \Phi_{11}FXF^H + \Phi_{21}FXG^H + \Phi_{12}GXF^H + \Phi_{22}GXG^H = 0, \\ & \Psi_{11}FXF^H + \Psi_{21}FXG^H + \Psi_{12}GXF^H + \Psi_{22}GXG^H \preceq 0, \\ & X \succeq 0. \end{array} \tag{32}$$

Example. Parametric line spectrum estimation is concerned with fitting signal models of the form

$$y(t) = \sum_{k=1}^{r} c_k e^{\mathrm{j}\omega_k t} + v(t), \tag{33}$$

where $v(t)$ is noise. If the phase angles of c_k are independent random variables, uniformly distributed on $[-\pi, \pi]$, and $v(t)$ is circular white noise with $\mathbf{E}\,|v(t)|^2 = \sigma^2$, then the covariance matrix of $y(t)$ of order n is given by

$$\begin{bmatrix} r_0 & r_{-1} & \cdots & r_{-n+1} \\ r_1 & r_0 & \cdots & r_{-n+2} \\ \vdots & \vdots & \ddots & \vdots \\ r_{n-1} & r_{n-2} & \cdots & r_0 \end{bmatrix} = \sigma^2 I + \sum_{k=1}^{r} |c_k|^2 \begin{bmatrix} 1 \\ e^{\mathrm{j}\omega_k} \\ \vdots \\ e^{\mathrm{j}(n-1)\omega_k} \end{bmatrix} \begin{bmatrix} 1 \\ e^{\mathrm{j}\omega_k} \\ \vdots \\ e^{\mathrm{j}(n-1)\omega_k} \end{bmatrix}^H, \tag{34}$$

where $r_k = \mathbf{E}\,(y(t)\overline{y(t-k)})$ [40, section 4.1], [34, section 12.5]. Classical methods, such as MUSIC and ESPRIT, are based on the eigenvalue decomposition of an estimated covariance matrix. With the formulation outlined in this section, one can solve related but more general covariance fitting problems, expressed as

$$\begin{array}{ll} \text{minimize} & f(R) + n \sum_{k=1}^{r} |c_k|^2 \\ \text{subject to} & R = \sigma^2 I + \sum_{k=1}^{r} |c_k|^2 \begin{bmatrix} 1 \\ e^{\mathrm{j}\omega_k} \\ \vdots \\ e^{\mathrm{j}(n-1)\omega_k} \end{bmatrix} \begin{bmatrix} 1 \\ e^{\mathrm{j}\omega_k} \\ \vdots \\ e^{\mathrm{j}(n-1)\omega_k} \end{bmatrix}^H \end{array}$$

with the variables $R \in \mathbf{H}^n$, σ^2, $|c_k|$, ω_k, and r, where f is a convex penalty or indicator function that measures the quality of the fit between R and the estimated covariance matrix. This is equivalent to the convex optimization problem

$$\begin{array}{ll} \text{minimize} & f(X + tI) + \mathbf{tr}\,X \\ \text{subject to} & X \succeq 0, \; t \geq 0, \; X \text{ is Toeplitz.} \end{array}$$

A numerical example is given in section 5.

3.2. Nonsymmetric matrices. We define F, G, E, Φ, Ψ, and $\mathcal{A}$ as in the previous section, but add the assumption that the matrices F, G, and E are block-diagonal:

$$G = \begin{bmatrix} G_1 & 0 \\ 0 & G_2 \end{bmatrix}, \qquad F = \begin{bmatrix} F_1 & 0 \\ 0 & F_2 \end{bmatrix}, \qquad E = \begin{bmatrix} E_1 & 0 \\ 0 & E_2 \end{bmatrix}. \tag{35}$$

Here F_1, $G_1 \in \mathbf{C}^{p_1 \times n_1}$ and F_2, $G_2 \in \mathbf{C}^{p_2 \times n_2}$ (possibly with $p_1 = 0$ or $p_2 = 0$). The matrices E_1 and E_2 have n_1 and n_2 columns, respectively. We discuss the function

$$h(Y) = \frac{1}{2} \inf_{V,W} g\left(\begin{bmatrix} V & Y \\ Y^H & W \end{bmatrix} \right)$$

of $Y \in \mathbf{C}^{n_1 \times n_2}$, where g is defined in (29) and (30). Using (29), we can write $h(Y)$ as

$$h(Y) = \inf \left\{ \frac{1}{2} \sum_{k=1}^{r} (\|E_1 v_k\|^2 + \|E_2 w_k\|^2) \mid Y = \sum_{k=1}^{r} v_k w_k^H,\ (v_k, w_k) \in \mathcal{A} \right\}, \tag{36}$$

while the characterization (30) shows that $h(Y)$ is the optimal value of the SDP:

$$\begin{array}{ll} \text{minimize} & \left(\mathbf{tr}(E_1 V E_1^H) + \mathbf{tr}(E_2 W E_2^H) \right)/2 \\ \text{subject to} & \Phi_{11} F X F^H + \Phi_{21} F X G^H + \Phi_{12} G X F^H + \Phi_{22} G X G^H = 0, \\ & \Psi_{11} F X F^H + \Psi_{21} F X G^H + \Psi_{12} G X F^H + \Psi_{22} G X G^H \preceq 0, \\ & X = \begin{bmatrix} V & Y \\ Y^H & W \end{bmatrix} \succeq 0 \end{array} \tag{37}$$

with V and W as variables. This can be seen as an extension of the well-known SDP formulation of the trace norm of a matrix. If we take F and G to have zero row dimensions (equivalently, define $\mathcal{A} = \mathbf{C}^{n_1} \times \mathbf{C}^{n_2}$ and omit the first two constraints in (37)) and choose $E_1 = I$, $E_2 = I$, then $h(Y) = \|Y\|_*$, the trace norm of Y.

The block-diagonal form of F and G implies that if $(v, w) \in \mathcal{A}$, then $(\alpha v, \beta w) \in \mathcal{A}$ for all α, β. This observation leads to a number of useful equivalent expressions for (36). First, we note that $h(Y)$ can be written as

$$h(Y) = \inf \left\{ \sum_{k=1}^{r} \|E_1 v_k\| \|E_2 w_k\| \mid Y = \sum_{k=1}^{r} v_k w_k^H,\ (v_k, w_k) \in \mathcal{A} \right\}. \tag{38}$$

This follows from the fact that $\|E_1 v_k\|^2 + \|E_2 w_k\|^2 \geq 2\|E_1 v_k\| \|E_2 w_k\|$, with equality if $\|E_1 v_k\| = \|E_2 w_k\|$. If the decomposition of Y in (36) involves a term $v_k w_k^H$ with $E_1 v_k$ and $E_2 w_k$ nonzero, then replacing v_k and w_k with $\tilde{v}_k = (\|E_2 w_k\| / \|E_1 v_k\|)^{1/2} v_k$ and $\tilde{w}_k = (\|E_1 v_k\| / \|E_2 w_k\|)^{1/2} w_k$ gives another valid decomposition with

$$\frac{1}{2} (\|E_1 \tilde{v}_k\|^2 + \|E_2 \tilde{w}_k\|^2) = \|E_1 v_k\| \|E_2 w_k\| \leq \frac{1}{2} (\|E_1 v_k\|^2 + \|E_2 w_k\|^2).$$

If $E_1 v_k = 0$ and $E_2 w_k \neq 0$, then replacing v_k and w_k with $\tilde{v}_k = \alpha v_k$, $\tilde{w}_k = (1/\alpha) w_k$ gives an equivalent decomposition with

$$\frac{1}{2} \left(\|E_1 \tilde{v}_k\|^2 + \|E_2 \tilde{w}_k\|^2 \right) = \frac{1}{2\alpha^2} \|E_2 w_k\|^2 \to 0$$

as α goes to infinity. The same argument applies when $E_1 v_k \neq 0$ and $E_2 w_k = 0$. In all cases, therefore, the two expressions (36) and (38) give the same result.

From (38), we obtain two other useful expressions:

$$h(Y) = \inf \left\{ \sum_{k=1}^{r} \|E_1 v_k\| \mid Y = \sum_{k=1}^{r} v_k w_k^H,\ (v_k, w_k) \in \mathcal{A},\ \|E_2 w_k\| \le 1 \right\} \tag{39}$$

$$= \inf \left\{ \sum_{k=1}^{r} \|E_2 w_k\| \mid Y = \sum_{k=1}^{r} v_k w_k^H,\ (v_k, w_k) \in \mathcal{A},\ \|E_1 v_k\| \le 1 \right\}. \tag{40}$$

This again follows from the property that the components v_k, w_k of elements (v_k, w_k) in $\mathcal{A}$ can be scaled independently. At the optimal decomposition in (39), all terms satisfy $E_2 w_k = 0$ or $\|E_2 w_k\| = 1$. In (40), all terms satisfy $E_1 v_k = 0$ or $\|E_1 v_k\| = 1$.

A final interpretation of h is

$$\begin{aligned} h(Y) = \inf \Bigg\{ & \sum_{k=1}^{r} \theta_k \mid Y = \sum_{k=1}^{r} \theta_k v_k w_k^H, \\ & \theta_k \ge 0,\ (v_k, w_k) \in \mathcal{A},\ \|E_1 v_k\| \le 1,\ \|E_2 w_k\| \le 1 \Bigg\}. \end{aligned} \tag{41}$$

The equivalence with (38) follows from the fact that if the optimal decomposition of $Y = \sum_{k=1}^{r} \theta_k v_k w_k^H$ involves the term $v_k w_k^H$, then the norms $\|E_1 v_k\|$ and $\|E_2 w_k\|$ will be either zero or 1. (If $0 < \|E_1 v_k\| < 1$, we can decrease θ_k by scaling v_k until $\|E_1 v_k\| = 1$, and similarly for w_k.) The expression (41) shows that $h(Y)$ is the gauge of the convex hull of the set $\{vw^H \in \mathbf{C}^{n_1 \times n_2} \mid (v, w) \in \mathcal{A},\ \|E_1 v\| \le 1,\ \|E_2 w\| \le 1\}$.

The SDP representation of h in (37) allows us to reformulate the problems

$$\text{minimize} \quad f(Y) + h(Y), \tag{42}$$

where f is convex and h is the gauge (36)–(41), as convex problems with SDP constraints. Minimizing $f(Y) + h(Y)$ is equivalent to

$$\begin{array}{ll} \text{minimize} & f(Y) + \sum_{k=1}^{r} \|E_1 v_k\| \|E_2 w_k\| \\ \text{subject to} & Y = \sum_{k=1}^{r} v_k w_k^H,\ \ (v_k, w_k) \in \mathcal{A},\ \ k = 1, \ldots, r. \end{array} \tag{43}$$

Alternatively, one can replace the second term in the objective with $\sum_k \|E_2 w_k\|$ and add constraints $\|E_1 v_k\| \le 1$, as in

$$\begin{array}{ll} \text{minimize} & f(Y) + \sum_{k=1}^{r} \|E_2 w_k\| \\ \text{subject to} & Y = \sum_{k=1}^{r} v_k w_k^H,\ \ (v_k, w_k) \in \mathcal{A},\ \ k = 1, \ldots, r, \\ & \|E_1 v_k\| \le 1,\ \ k = 1, \ldots, r, \end{array} \tag{44}$$

or vice versa. When E_1 and E_2 are identity matrices, we can interpret $h(Y)$ as a convex penalty that promotes a structured low-rank property of Y. The outer products $v_k w_k^H$ are constrained by the set $\mathcal{A}$; the penalty term in the objective is the sum of the norms $\|v_k w_k^H\|_2 = \|v_k\| \|w_k\|$. The matrices E_1 and E_2 can be chosen to assign a different weight to different terms $v_k w_k^H$.

Problems (43) and (44) can be reformulated as

$$
\begin{array}{ll}
\text{minimize} & f(Y) + (\mathbf{tr}(E_1 V E_1^H) + \mathbf{tr}(E_2 W E_2^H))/2 \\
\text{subject to} & \Phi_{11} F X F^H + \Phi_{21} F X G^H + \Phi_{12} G X F^H + \Phi_{22} G X G^H = 0, \\
& \Psi_{11} F X F^H + \Psi_{21} F X G^H + \Psi_{12} G X F^H + \Psi_{22} G X G^H \preceq 0, \\
& X = \begin{bmatrix} V & Y \\ Y^H & W \end{bmatrix} \succeq 0.
\end{array} \tag{45}
$$

Example: Column structure. When $p_2 = 0$, the matrices F and G in (35) have the form $F = [\ F_1 \quad 0\ ]$ and $G = [\ G_1 \quad 0\ ]$. This means that $\mathcal{A} = \mathcal{A}_1 \times \mathbf{C}^{n_2}$, where

$$
\mathcal{A}_1 = \{v \in \mathbf{C}^{n_1} \mid (\mu G_1 - \nu F_1) v = 0,\ (\mu, \nu) \in \mathcal{C}\}.
$$

There are no restrictions on the w-component in $(v, w) \in \mathcal{A}$. Problem (43) simplifies to

$$
\begin{array}{ll}
\text{minimize} & f(Y) + \sum_{k=1}^{r} \|E_1 v_k\| \|E_2 w_k\| \\
\text{subject to} & Y = \sum_{k=1}^{r} v_k w_k^H, \quad v_k \in \mathcal{A}_1, \quad k = 1, \ldots, r.
\end{array} \tag{46}
$$

The equivalent semidefinite formulation (45) simplifies to

$$
\begin{array}{ll}
\text{minimize} & f(Y) + (\mathbf{tr}(E_1 V E_1^H) + \mathbf{tr}(E_2 W E_2^H))/2 \\
\text{subject to} & \Phi_{11} F_1 V F_1^H + \Phi_{21} F_1 V G_1^H + \Phi_{12} G_1 V F_1^H + \Phi_{22} G_1 V G_1^H = 0, \\
& \Psi_{11} F_1 V F_1^H + \Psi_{21} F_1 V G_1^H + \Psi_{12} G_1 V F_1^H + \Psi_{22} G_1 V G_1^H \preceq 0, \\
& \begin{bmatrix} V & Y \\ Y^H & W \end{bmatrix} \succeq 0.
\end{array}
$$

This SDP formulation of (46) (with $E_1 = I$, $E_2 = I$) was studied in [6].

As an example, we again consider the signal model (33). A natural idea for estimating the parameters ω_k and c_k is to solve a nonlinear least squares problem

$$
\text{minimize} \quad \sum_{t=0}^{n-1} \left| y_{\mathrm{m}}(t) - \sum_{k=1}^{r} c_k e^{\mathrm{j}\omega_k t} \right|^2,
$$

where $y_{\mathrm{m}}(t)$ is the observed signal. This problem is not convex and is difficult to solve iteratively without a good starting point [40, page 148]. Instead of fixing r, we can also impose a penalty on $\sum_k |c_k|$ and consider the optimization problem

$$
\begin{array}{ll}
\text{minimize} & \gamma \|y - y_{\mathrm{m}}\|^2 + \sum_{k=1}^{r} |c_k| \\
\text{subject to} & y = \sum_{k=1}^{r} c_k (1,\ e^{\mathrm{j}\omega_k}, \ldots,\ e^{\mathrm{j}(n-1)\omega_k}).
\end{array} \tag{47}
$$

The optimization variables are y and the parameters c_k, ω_k, and r in the decomposition of y. The vector y_{m} has elements $y_{\mathrm{m}}(0), \ldots, y_{\mathrm{m}}(n-1)$. This is a special case of (44) with $f(y) = \gamma \|y - y_{\mathrm{m}}\|^2$, $n_1 = n$, $n_2 = 1$, $\Phi = \Phi_{\mathrm{u}}$, $\Psi = 0$, and

$$
E_1 = (1/\sqrt{n}) I, \qquad E_2 = 1, \qquad F_1 = \begin{bmatrix} 0 & I_{n_1 - 1} \end{bmatrix}, \qquad G_1 = \begin{bmatrix} I_{n_1 - 1} & 0 \end{bmatrix},
$$

so that $\mathcal{A}_1$ is the set of all multiples of vectors $(1, e^{j\omega}, \dots, e^{j(n-1)\omega})$. The problem is therefore equivalent to the convex problem

$$\begin{array}{ll} \text{minimize} & \gamma\|y - y_{\rm m}\|^2 + (\mathbf{tr}\, V)/(2n) + w/2 \\ \text{subject to} & \begin{bmatrix} V & y \\ y^H & w \end{bmatrix} \succeq 0, \quad V \text{ is Toeplitz.} \end{array}$$

A related numerical example will be given in section 5.2.

Example: Joint column and row structure. To illustrate the general problem (43), we consider a variation on the previous example. Suppose we arrange the observations in an $n \times m$ Hankel matrix,

$$Y_{\rm m} = \begin{bmatrix} y_{\rm m}(0) & y_{\rm m}(1) & \cdots & y_{\rm m}(m-1) \\ y_{\rm m}(1) & y_{\rm m}(2) & \cdots & y_{\rm m}(m) \\ \vdots & \vdots & & \vdots \\ y_{\rm m}(n-1) & y_{\rm m}(n) & \cdots & y_{\rm m}(m+n-2) \end{bmatrix},$$

and we fit to this matrix a matrix Y with the same Hankel structure and with elements $y(t) = \sum_{k=1}^{r} c_k \exp(j\omega_k t)$. We formulate the problem as

$$\begin{array}{ll} \text{minimize} & \gamma\|Y - Y_{\rm m}\|_F^2 + \sum_{k=1}^{r} |c_k| \\ \text{subject to} & Y = \sum_{k=1}^{r} c_k \begin{bmatrix} 1 \\ e^{j\omega_k} \\ \vdots \\ e^{j(n-1)\omega_k} \end{bmatrix} \begin{bmatrix} 1 \\ e^{-j\omega_k} \\ \vdots \\ e^{-j(m-1)\omega_k} \end{bmatrix}^H. \end{array} \tag{48}$$

This is an instance of (43) with $n_1 = n$, $n_2 = m$, $\Phi = \Phi_{\rm u}$, $\Psi = 0$, $E_1 = (1/\sqrt{n})I$, $E_2 = (1/\sqrt{m})I$, and

$$G_1 = \begin{bmatrix} I_{n-1} & 0 \end{bmatrix}, \quad F_1 = \begin{bmatrix} 0 & I_{n-1} \end{bmatrix}, \quad G_2 = \begin{bmatrix} 0 & I_{m-1} \end{bmatrix}, \quad F_2 = \begin{bmatrix} I_{m-1} & 0 \end{bmatrix}.$$

With these parameters, the set $\mathcal{A}$ contains the pairs (v, w) of the form

$$v = \alpha(1, e^{j\omega}, \dots, e^{j(n-1)\omega}), \qquad w = \beta(1, e^{-j\omega}, \dots, e^{-j(m-1)\omega}).$$

The convex formulation is

$$\begin{array}{ll} \text{min.} & \gamma\|Y - Y_{\rm m}\|_F^2 + (\mathbf{tr}\, V)/(2n) + (\mathbf{tr}\, W)/(2m) \\ \text{s.t.} & \begin{bmatrix} V & Y \\ Y^H & W \end{bmatrix} \succeq 0, \\ & \begin{bmatrix} F_1 & 0 \\ 0 & F_2 \end{bmatrix} \begin{bmatrix} V & Y \\ Y^H & W \end{bmatrix} \begin{bmatrix} F_1 & 0 \\ 0 & F_2 \end{bmatrix}^T = \begin{bmatrix} G_1 & 0 \\ 0 & G_2 \end{bmatrix} \begin{bmatrix} V & Y \\ Y^H & W \end{bmatrix} \begin{bmatrix} G_1 & 0 \\ 0 & G_2 \end{bmatrix}^T. \end{array}$$

A related example is discussed in section 5.2.

4. Duality. In this section, we derive the conjugates of the gauge functions defined in section 3 and show that they can be interpreted as indicator functions of sets of nonnegative or bounded generalized polynomials. This gives a useful interpretation of the dual problems for (25) and (42).

We assume that the subset of the complex plane represented by $\mathcal{C}$ in (7) is one-dimensional, i.e., $\mathcal{C}$ is not a singleton and not the empty set. Equivalently, the inequality $q_\Psi(\mu,\nu) \leq 0$ in the definition is either redundant (and $\mathcal{C}$ represents a line or circle), or it is not redundant and then there exist elements of $\mathcal{C}$ with $q_\Psi(\mu,\nu) < 0$. When stating and analyzing the dual problems, we will need to distinguish these two cases ($q_\Psi(\mu,\nu) \leq 0$ is redundant or not). For the sake of brevity, we only give the formulas for the case where the inequality is not redundant. The dual problems for the other case follow by setting $\Psi = 0$ and making obvious simplifications.

We also assume that $\mu G - \nu F$ has full row rank ($\mathbf{rank}(\mu G - \nu F) = p$) for all nonzero (μ,ν)). This condition will serve as a "constraint qualification" that guarantees strong duality.

4.1. Symmetric matrix gauge. We first consider the conjugate of the function g defined in (30). The conjugate is defined as $g^*(Z) = \sup_X (\mathbf{tr}(XZ) - g(X))$, i.e., the optimal value of the SDP

$$\begin{array}{ll} \text{maximize} & \mathbf{tr}\,((Z - E^H E)X) \\ \text{subject to} & X \succeq 0, \\ & \Phi_{11} F X F^H + \Phi_{21} F X G^H + \Phi_{12} G X F^H + \Phi_{22} G X G^H = 0, \\ & \Psi_{11} F X F^H + \Psi_{21} F X G^H + \Psi_{12} G X F^H + \Psi_{22} G X G^H \preceq 0. \end{array} \tag{49}$$

The dual of this problem is

$$\begin{array}{ll} \text{minimize} & 0 \\ \text{subject to} & Z - \begin{bmatrix} F \\ G \end{bmatrix}^H (\Phi \otimes P + \Psi \otimes Q) \begin{bmatrix} F \\ G \end{bmatrix} \preceq E^H E, \\ & Q \succeq 0 \end{array} \tag{50}$$

with variables $P, Q \in \mathbf{H}^p$. It is shown in Appendix C that strong duality holds under the assumptions listed at the beginning of section 4.

If strong duality holds, then $g^*(Z)$ is the optimal value of (50), i.e., equal to zero if there exist P, Q that satisfy the constraints in (50), and $+\infty$ otherwise. In other words, $g^*(Z)$ is the indicator function of the set described by the constraints in (50). To complete the picture, we now show that $g^*(Z)$ can be expressed as

$$g^*(Z) = \begin{cases} 0, & a^H Z a \leq \|Ea\|^2 \text{ for all } a \in \mathcal{A}, \\ +\infty & \text{otherwise.} \end{cases} \tag{51}$$

This expression of g^* follows directly from the definition of the conjugate and (29), since

$$g^*(Z) = \sup_X (\mathbf{tr}\,(XZ) - g(X)) = \sup_{a_1,\ldots,a_r \in \mathcal{A}} \sum_{k=1}^r \left(a_k^H Z a_k - \|Ea_k\|^2\right),$$

which is the same as (51). This is consistent with a property from gauge duality: the conjugate of the gauge of a set is the indicator of the unit level set of the polar gauge [11, Proposition 2.1]. It is also instructive to derive (51) from the dual SDP (50). Suppose P and Q are feasible in (50). Consider any $a \in \mathcal{A}$ and $(\mu,\nu) \in \mathcal{C}$ with $\mu G a = \nu F a$. Define $y = (1/\nu) G a$ if $\nu \neq 0$ and $y = (1/\mu) F a$ otherwise. Then

$$
\begin{aligned}
a^H Za - \|Ea\|^2 &\le \begin{bmatrix} Fa \\ Ga \end{bmatrix}^H (\Phi \otimes P + \Psi \otimes Q) \begin{bmatrix} Fa \\ Ga \end{bmatrix} \\
&= \begin{bmatrix} \mu y \\ \nu y \end{bmatrix}^H (\Phi \otimes P + \Psi \otimes Q) \begin{bmatrix} \mu y \\ \nu y \end{bmatrix} \\
&= (y^H P y) q_\Phi(\mu, \nu) + (y^H Q y) q_\Psi(\mu, \nu) \\
&\le 0.
\end{aligned}
$$

The last line follows from $Q \succeq 0$ and $q_\Phi(\mu, \nu) = 0$, $q_\Psi(\mu, \nu) \le 0$. Conversely, if problem (50) is infeasible, then the optimal value is $+\infty$ and, since strong duality holds, there exist matrices X that are feasible for (49) with $\mathbf{tr}((Z - E^H E)X) > 0$. Applying Theorem 2.1, we see that there exist $a_1, \ldots, a_r \in \mathcal{A}$ with $\sum_k (a_k^H Z a_k - \|Ea_k\|^2) > 0$. Therefore, $a_k^H Z a_k > \|Ea_k\|^2$ for at least one a_k.

The interpretation of the conjugate gives useful insight in problem (25), where g is defined in (30). The dual problem is

$$
\text{maximize} \quad -f^*(Z) - g^*(-Z).
$$

Expanding $g^*(-Z)$ using (50) gives the equivalent problem

$$
\begin{array}{ll}
\text{maximize} & -f^*(Z) \\
\text{subject to} & -Z - \begin{bmatrix} F \\ G \end{bmatrix}^H (\Phi \otimes P + \Psi \otimes Q) \begin{bmatrix} F \\ G \end{bmatrix} \preceq E^H E, \\
& Q \succeq 0
\end{array}
\tag{52}
$$

with variables Z, P, and Q, and using the expression (51) we can put the constraints in this problem more succinctly as

$$
\begin{array}{ll}
\text{maximize} & -f^*(Z) \\
\text{subject to} & \|Ea\|^2 + a^H Z a \ge 0 \quad \text{for all } a \in \mathcal{A}.
\end{array}
\tag{53}
$$

This last form leads to an interesting set of optimality conditions. Suppose X and Z are feasible for (31) and (53), respectively. Then

$$
\begin{aligned}
f(X) + \sum_{k=1}^r \|Ea_k\|^2 &\ge -f^*(Z) + \mathbf{tr}(XZ) + \sum_{k=1}^r \|Ea_k\|^2 \\
&= -f^*(Z) + \sum_{k=1}^r (\|Ea_k\|^2 + a_k^H Z a_k) \\
&\ge -f^*(Z).
\end{aligned}
$$

The first inequality follows by definition of $f^*(Z)$, and the second and third line follow from primal and dual feasibility. If X and Z are optimal and strong duality holds, then $f(X) + \sum_k \|Ea_k\|^2 = -f^*(Z)$. This is only possible if $f(X) + f^*(Z) = \mathbf{tr}(XZ)$ and $\|Ea_k\|^2 + a_k^H Z a_k = 0$ for $k = 1, \ldots, r$. Hence, only the vectors $a \in \mathcal{A}$, at which the inequality in (53) is active, can be used to form an optimal $X = \sum_k a_k a_k^H$.

Example: Generalized KYP lemma. When specialized to the controllability pencil (21), the equivalence between the constraints in (53) and (52) is known as the (generalized) KYP lemma [22, 24, 33, 39, 44].

We assume that A has no eigenvalues λ with $(\lambda, 1) \in \mathcal{C}$, and that the pair (A, B) is controllable, so the pencil satisfies the rank condition that $\mathbf{rank}(\lambda F - G) = n_{\mathrm{s}}$ for all λ. The dual problem (53) becomes

$$\begin{array}{ll} \text{maximize} & -f^*(Z) \\ \text{subject to} & \mathcal{F}(\lambda, Z) \succeq 0 \quad \text{for all } (\lambda, 1) \in \mathcal{C}, \\ & M_{22} + Z_{22} \succeq 0 \quad \text{if } (1,0) \in \mathcal{C}, \end{array}$$

where $M = E^H E$ and

$$\mathcal{F}(\lambda, Z) = \begin{bmatrix} (\lambda I - A)^{-1} B \\ I \end{bmatrix}^H \begin{bmatrix} M_{11} + Z_{11} & M_{12} + Z_{12} \\ M_{21} + Z_{21} & M_{22} + Z_{22} \end{bmatrix} \begin{bmatrix} (\lambda I - A)^{-1} B \\ I \end{bmatrix}.$$

The function $\mathcal{F}$ is called the *Popov function* with central matrix $M + Z$ [19, 21].

4.2. Nonsymmetric matrix gauge. Consider the conjugate of the gauge defined in (36)–(41). We have $h^*(Z) = \sup_Y (\mathrm{Re}\,(\mathbf{tr}\, Z^H Y) - h(Y))$, where $h(Y)$ is the optimal value of (37). Therefore, $h^*(Z)$ is the optimal value of the SDP

$$\begin{array}{ll} \text{maximize} & \frac{1}{2}\,\mathbf{tr}\left(\begin{bmatrix} -E_1^H E_1 & Z \\ Z^H & -E_2^H E_2 \end{bmatrix} X\right) \\ \text{subject to} & \Phi_{11} F X F^H + \Phi_{21} F X G^H + \Phi_{12} G X F^H + \Phi_{22} G X G^H = 0, \\ & \Psi_{11} F X F^H + \Psi_{21} F X G^H + \Psi_{12} G X F^H + \Psi_{22} G X G^H \preceq 0, \\ & X \succeq 0. \end{array} \tag{54}$$

The dual of this problem is

$$\begin{array}{ll} \text{min.} & 0 \\ \text{s.t.} & \begin{bmatrix} 0 & Z \\ Z^H & 0 \end{bmatrix} - \begin{bmatrix} F \\ G \end{bmatrix}^H (\Phi \otimes P + \Psi \otimes Q) \begin{bmatrix} F \\ G \end{bmatrix} \preceq \begin{bmatrix} E_1^H E_1 & 0 \\ 0 & E_2^H E_2 \end{bmatrix}, \\ & Q \succeq 0. \end{array} \tag{55}$$

As in the previous section, it follows from Appendix C that strong duality holds. Therefore, $h^*(Z)$ is equal to the optimal value of (55), i.e., zero if there exist P and Q that satisfy the constraints of this problem, and $+\infty$ otherwise. This will now be shown to be equivalent to

$$\begin{aligned} h^*(Z) &= \begin{cases} 0, & \mathrm{Re}\,(v^H Z w) \leq (\|E_1 v\|^2 + \|E_2 w\|^2)/2 \quad \text{for all } (v, w) \in \mathcal{A}, \\ +\infty & \text{otherwise,} \end{cases} \\ &= \begin{cases} 0, & \mathrm{Re}\,(v^H Z w) \leq \|E_1 v\| \|E_2 w\| \quad \text{for all } (v, w) \in \mathcal{A}, \\ +\infty & \text{otherwise.} \end{cases} \end{aligned} \tag{56}$$

To see this, first assume that P and Q are feasible in (55), and $a = (v, w) \in \mathcal{A}$ satisfies $(\mu G - \nu F) a = 0$ with $(\mu, \nu) \in \mathcal{C}$. Then

$$\begin{aligned} v^H Z w + w^H Z^H v - \|E_1 v\|^2 - \|E_2 w\|^2 &\leq \begin{bmatrix} Fa \\ Ga \end{bmatrix}^H (\Phi \otimes P + \Psi \otimes Q) \begin{bmatrix} Fa \\ Ga \end{bmatrix} \\ &= (y^H P y) q_\Phi(\mu, \nu) + (y^H Q y) q_\Psi(\mu, \nu) \\ &\leq 0, \end{aligned}$$

where $y = (1/\nu) G a$ if $\nu \neq 0$ and $y = (1/\mu) F a$ otherwise. Conversely, if problem (55) is infeasible, then (54) is unbounded above, so there exists a feasible X with positive objective value. If we decompose X as in Theorem 2.1, with $a_k = (v_k, w_k)$, we find that

$$0 < \mathbf{tr}\left(\begin{bmatrix} -E_1^H E_1 & Z \\ Z^H & -E_2^H E_2 \end{bmatrix} \sum_{k=1}^{r} \begin{bmatrix} v_k \\ w_k \end{bmatrix} \begin{bmatrix} v_k \\ w_k \end{bmatrix}^H\right)$$
$$= \sum_{k=1}^{r} \left(v_k^H Z w_k + w_k^H Z^H v_k - \|E_1 v_k\|^2 - \|E_2 w_k\|^2\right),$$

so at least one term in the sum is positive. The second expression for $h^*(Z)$ in (56) follows from the block-diagonal structure of F and G. Following similar arguments as in section 4.1, the expression (56) can also be derived directly from definition of the conjugate, (36), and (38), or via gauge duality.

The interpretation of the conjugate h^* can be applied to interpret the dual of (42)

$$\text{maximize} \quad -f^*(Z) - h^*(-Z).$$

Substituting the expression (56) for $h^*(-Z)$, we obtain

$$\begin{array}{ll} \text{maximize} & -f^*(Z) \\ \text{subject to} & \operatorname{Re}(v^H Z w) \leq \|E_1 v\| \|E_2 w\| \quad \text{for all } (v, w) \in \mathcal{A}. \end{array}$$

As in the previous section, the primal-dual optimality conditions provide a useful set of complementary slackness relations between primal optimal Y and dual optimal Z. The optimal Y can be decomposed as $Y = \sum_k v_k w_k^H$ with elements $(v_k, w_k) \in \mathcal{A}$ at which $\operatorname{Re}(v_k^H Z w_k) = \|E_1 v_k\| \|E_2 w_k\|$.

Example. Suppose $A \in \mathbf{C}^{n_s \times n_s}$, $B \in \mathbf{C}^{n_s \times m}$, $C \in \mathbf{C}^{l \times n_s}$, and $D \in \mathbf{C}^{l \times m}$ are matrices in a state-space model with (A, B) controllable, and A has no eigenvalues that satisfy $(\lambda, 1) \in \mathcal{C}$. We take $p_1 = 0$, $n_1 = l$, $p_2 = n_s$, $n_2 = n_s + m$,

$$G_2 = \begin{bmatrix} I & 0 \end{bmatrix}, \qquad F_2 = \begin{bmatrix} A & B \end{bmatrix}, \qquad E_1 = I, \qquad E_2 = \begin{bmatrix} 0 & I \end{bmatrix}.$$

With this choice of parameters, $\mathcal{A} = \mathbf{C}^l \times \mathcal{A}_2$, where $\mathcal{A}_2$ contains the vectors

$$w = \begin{bmatrix} (\lambda I - A)^{-1} B u \\ u \end{bmatrix}$$

for all $u \in \mathbf{C}^m$ and all $(\lambda, 1) \in \mathcal{C}$, plus the vectors $(0, u)$ if $(1, 0) \in \mathcal{C}$. Since v is arbitrary and $E_1 = I$, the inequality in (56) reduces to $\|Zw\| \leq \|E_2 w\|$ for all $w \in \mathcal{A}_2$. With $Z = \begin{bmatrix} C & D \end{bmatrix}$, this is equivalent to a bound on the transfer function

$$\|D + C(\lambda I - A)^{-1} B\|_2 \leq 1 \quad \text{for all } (\lambda, 1) \in \mathcal{C}, \qquad \|D\|_2 \leq 1 \quad \text{if } (1, 0) \in \mathcal{C}.$$

5. Examples.

The formulations in section 3 will now be illustrated with examples from signal processing. The optimization problems were solved with the software package CVX [16].

5.1. Line spectrum estimation by Toeplitz covariance fitting.

We fit a covariance matrix of the form (34) to an estimated covariance matrix R_m. The estimate R_m is constructed from $N = 150$ samples of the time series $y(t)$ defined in (33), with $r = 3$, and the frequencies ω_k and the magnitudes $|c_k|$ shown in Figure 1. The noise is Gaussian white noise with variance $\sigma^2 = 64$. The sample covariance matrix

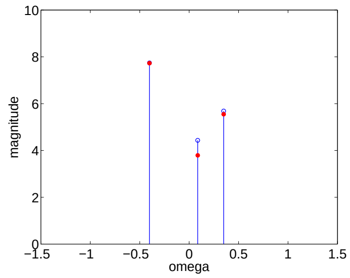

FIG. 1. *Line spectrum estimation by Toeplitz covariance fitting (section* 5.1*). The red dots represent the frequencies and magnitudes of the true model. The blue lines show the estimated parameters obtained by solving* (57)*.*

is constructed as $R_{\mathrm{m}} = YY^H/(N-n+1)$, where Y is the $n \times (N-n+1)$ Hankel matrix with $y(1), \ldots, y(N-n+1)$ in its first row. To estimate the model, we solve

$$
\begin{array}{ll}
\text{minimize} & \gamma \|R - R_{\mathrm{m}}\|_2 + \sum_{k=1}^{r} |c_k|^2 \\
\text{subject to} & R = \sigma^2 I + \sum_{k=1}^{r} |c_k|^2 \begin{bmatrix} 1 \\ e^{\mathrm{j}\omega_k} \\ \vdots \\ e^{\mathrm{j}(n-1)\omega_k} \end{bmatrix} \begin{bmatrix} 1 \\ e^{\mathrm{j}\omega_k} \\ \vdots \\ e^{\mathrm{j}(n-1)\omega_k} \end{bmatrix}^H
\end{array}
\tag{57}
$$

with variables σ^2, $|c_k|^2$, ω_k, r, and R. The norm $\|\cdot\|_2$ in the objective is the spectral norm. The regularization parameter γ is set to 0.25. As can be seen from Figure 1, the recovered parameters ω_k and $|c_k|$ are quite accurate, despite the very low signal-to-noise ratio. The estimated noise variance σ^2 is 79.6.

The semidefinite optimization approach allows us to fit a covariance matrix with the structure prescribed in (34) to a sample covariance matrix that may not be Toeplitz or positive semidefinite. The formulation can also be extended to applications where the noise $v(t)$ is modeled as a moving-average process, by combining it with the formulation in [14].

5.2. Line spectrum estimation by penalty approximation. This example is a variation on problem (47). We take $n = 50$ consecutive measurements of the signal defined in (33). There are three sinusoids with frequencies and magnitudes shown in Figure 3. The noise $v(t)$ is a superposition of white noise and a sparse corruption of 20 elements (see Figure 2). The model parameters are estimated by solving

$$
\begin{array}{ll}
\text{minimize} & \gamma \sum_{i=1}^{n} \phi(y_i - y_{\mathrm{m},i}) + \sum_{k=1}^{r} |c_k| \\
\text{subject to} & y = \sum_{k=1}^{r} c_k (1, e^{\mathrm{j}\omega_k}, \ldots, e^{\mathrm{j}(n-1)\omega_k}), \\
& |\omega_k| \le \omega_{\mathrm{c}}, \quad k = 1, \ldots, r,
\end{array}
\tag{58}
$$

where ϕ is the Huber penalty, $\gamma = 0.071$, and $\omega_{\mathrm{c}} = \pi/6$. The variables are y and the parameters r, c_k, and ω_k in the decomposition of y. Figure 3 shows the result, as well as the estimates obtained from a simple implementation (without filtering) of the matrix pencil method described in [20, 38], where we form a 30×21 Hankel matrix Y_{m} from the measurements and compute the generalized eigenvalues of $\lambda Y_{\mathrm{m}1} - Y_{\mathrm{m}2}$

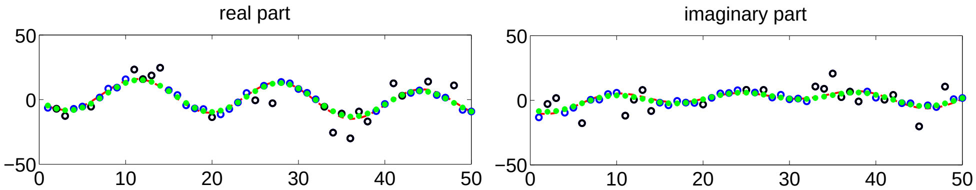

FIG. 2. *The data for the example in section* 5.2. *The red dashed lines show the exact, noise-free signal. The circles show the signal corrupted by Gaussian white noise (in blue), plus a few larger errors in* 20 *positions (in black). The green dots show the recovered signal y from* (58).

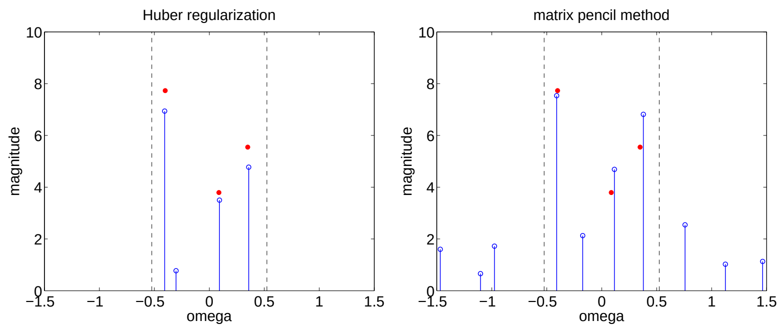

FIG. 3. *Line spectrum models estimated from the signal in Figure* 2 *by solving the optimization problem* (58) *(left) and using the matrix pencil method (right).*

as estimates of $e^{\mathrm{j}\omega_k}$ ($Y_{\mathrm{m}1}$ and $Y_{\mathrm{m}2}$ represent the matrix Y_{m} with the last and the first column removed, respectively). The comparison illustrates the usefulness of incorporating the prior frequency constraint and the Huber penalty in (58).

It is interesting to note that problem (58) can be equivalently formulated as

$$\begin{array}{ll} \text{min.} & \gamma \sum_{i=1}^{n} \phi(y_i - y_{\mathrm{m},i}) + \sum_{k=1}^{r} |c_k| \\ \text{s.t.} & \begin{bmatrix} y_1 & y_2 & \cdots & y_{n_2} \\ y_2 & y_3 & \cdots & y_{n_2+1} \\ \vdots & \vdots & & \vdots \\ y_{n_1} & y_{n_1+1} & \cdots & y_{n_1+n_2-1} \end{bmatrix} = \sum_{k=1}^{r} c_k \begin{bmatrix} 1 \\ e^{\mathrm{j}\omega_k} \\ \vdots \\ e^{\mathrm{j}(n_1-1)\omega_k} \end{bmatrix} \begin{bmatrix} 1 \\ e^{-\mathrm{j}\omega_k} \\ \vdots \\ e^{-\mathrm{j}(n_2-1)\omega_k} \end{bmatrix}^H, \\ & |\omega_k| \le \omega_{\mathrm{c}}, \quad k = 1, \ldots, r, \end{array} \tag{59}$$

where $n_1 + n_2 - 1 = n$. The SDPs equivalent to (58) and (59), respectively, give the same result y, but may have different numerical properties (in terms of accuracy or complexity).

5.3. Direction of arrival from multiple measurement vectors. This example demonstrates the advantage of using multiple measurement vectors (or snapshots) in direction-of-arrival (DOA) estimation, as pointed out in [27, 45]. Suppose we have K omnidirectional sensors placed at randomly chosen positions of a linear grid of length n. The measurements of the K sensors at one time instance form one measurement vector. We collect m of these measurement vectors, at m different times,

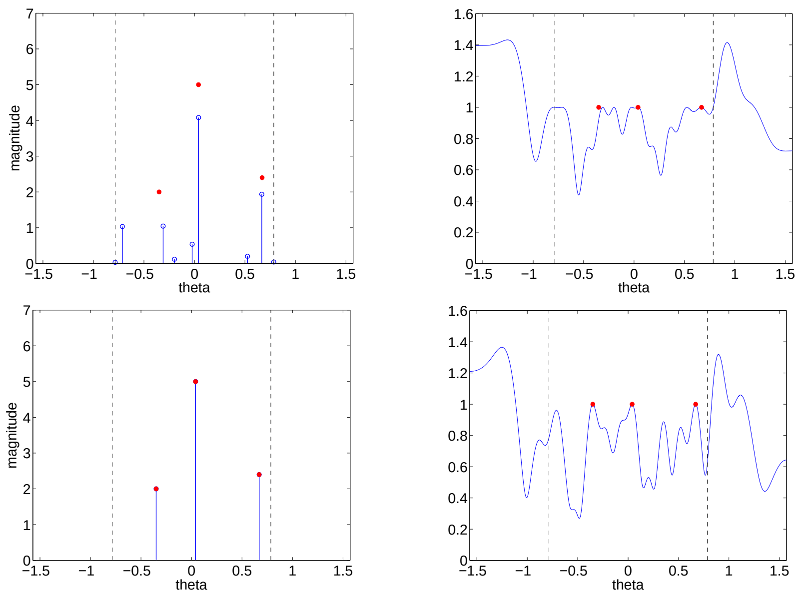

FIG. 4. *The results with* 15 *(top) and* 30 *(bottom) measurement vectors in the DOA estimation problem of section* 5.3*. The figures on the right show the magnitude of the trigonometric polynomials obtained from the dual solution. The red dots show the true directions of arrival (and magnitudes).*

and assume that the directions of arrival and the source magnitudes remain constant while the measurements are taken. The problem is formulated as

$$\begin{array}{ll} \text{minimize} & \sum_{k=1}^{r} \|c_k\| \\ \text{subject to} & Y = \sum_{k=1}^{r} \begin{bmatrix} 1 \\ e^{\mathrm{j}\alpha \sin\theta_k} \\ \vdots \\ e^{\mathrm{j}(n-1)\alpha \sin\theta_k} \end{bmatrix} c_k^H, \\ & |\theta_k| \le \theta_{\mathrm{c}}, \quad k = 1, \dots, r, \\ & Y_I = B, \end{array} \tag{60}$$

with variables $Y \in \mathbf{C}^{n \times m}$, $c_k \in \mathbf{C}^m$, θ_k, and r. Here $\alpha = 2\pi d/\lambda_{\mathrm{c}}$, where d is the distance between the grid points, λ_{c} is the signal wavelength, and θ_{c} is a given cutoff angle. The columns of the $K \times m$ matrix B are the measurement vectors. The matrix Y_I is the submatrix of Y containing the K rows indexed by $I \subset \{1, \dots, n\}$.

Figure 4 shows an instance with $n = 30$, $K = 7$, $\alpha = 2$, and $\theta_{\mathrm{c}} = \pi/4$. We show the solution for $m = 15$ and $m = 30$. The blue lines show the values of θ_k and $\|c_k\|/\sqrt{m}$ computed by solving problem (60). In an experiment of 150 trials with randomly chosen index sets I, the signal was recovered accurately in 67.3% of the trials for $m = 15$ and 85.3% for $m = 30$.

6. Conclusion. In this paper we developed semidefinite representations of a class of gauge functions and atomic norms for sets parameterized by linear matrix pencils. The formulations extend the semidefinite representation of the atomic norm associated with the trigonometric moment curve, which underlies recent results in continuous or "off-the-grid" compressed sensing. The main contribution is a self-contained constructive proof of the semidefinite representations, using techniques developed in the literature on the Kalman–Yakubovich–Popov (KYP) lemma. In addition to opening new possible areas of applications in system theory and control, the connection with the KYP lemma is important for numerical algorithms. Specialized techniques for solving SDPs derived from the KYP lemma, for example, by exploiting real symmetries and rank-one structure [13, 18, 28, 29, 37], should be useful in the development of fast solvers for the SDPs discussed in this paper.

Appendix A. Subsets of the complex plane. In this appendix we explain the notation used in (7) to describe subsets of the closed complex plane. Recall that we use the notation q_Φ, q_Ψ for the quadratic forms (8).

If Φ is a 2×2 Hermitian matrix with $\det \Phi < 0$, then the quadratic equation $q_\Phi(\lambda, 1) = 0$ defines a straight line (if $\Phi_{11} = 0$) or a circle (if $\Phi_{11} \neq 0$) in the complex plane. When $\Phi_{11} = 0$, we include the point $\lambda = \infty$ in the solution set of $q_\Phi(\lambda, 1) = 0$. Alternatively, one can define points in the closed complex plane as pairs $(\mu, \nu) \neq 0$. If $\nu \neq 0$, the pair (μ, ν) represents the complex number $\lambda = \mu/\nu$. If $\nu = 0$, it represents the point at infinity. Using this notation, a circle or line in the closed complex plane is defined as the nonzero solution set of a quadratic equation $q_\Phi(\mu, \nu) = 0$ with $\det \Phi < 0$. A congruence transformation $\tilde{\Phi} = R\Phi R^H$ corresponds to a linear transformation between the sets associated with the matrices Φ and $\tilde{\Phi}$.

The second type of set we encounter is defined by an equality and an inequality:

$$q_\Phi(\lambda, 1) = 0, \qquad q_\Psi(\lambda, 1) \leq 0. \tag{61}$$

We assume that $\det \Phi < 0$. If the inequality is redundant (e.g., $\Psi = 0$), then the solution set of (61) is the line or circle defined by the equality. Otherwise, it is an arc of a circle, a closed interval of a line, or the complement of an open interval of a line. It includes the point at infinity if $\Phi_{11} = 0$ and $\Psi_{11} \leq 0$. Alternatively, one can use homogeneous coordinates and consider sets of nonzero points (μ, ν) that satisfy $q_\Phi(\mu, \nu) = 0$ and $q_\Psi(\mu, \nu) \leq 0$. Some common combinations of Φ and Ψ are listed in [7].

As for circles and lines, we can apply a congruence transformation to reduce (61) to a simple canonical case. Iwasaki and Hara [22, Lemma 2] show that, for every Φ, Ψ with $\det \Phi < 0$, there exists a nonsingular R such that

$$\Phi = R^H \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} R, \qquad \Psi = R^H \begin{bmatrix} \alpha & \beta \\ \beta & \gamma \end{bmatrix} R \tag{62}$$

with α, β, γ real, and $\alpha \geq \gamma$ (see also [7]). Applying the congruence defined by R, we can reduce the conditions (61) to an equivalent system $\operatorname{Re} \lambda' = 0$, $\alpha|\lambda'|^2 + \gamma \leq 0$. Keeping in mind that $\alpha \geq \gamma$, we can distinguish four cases. If $0 < \gamma \leq \alpha$, the solution set is empty. If $\gamma = 0 < \alpha$, the solution set is a singleton $\{0\}$. If $\gamma < 0 < \alpha$, the solution set is the interval of the imaginary axis defined by $|\lambda'| \leq (-\gamma/\alpha)^{1/2}$. If $\gamma \leq \alpha \leq 0$, the inequality is redundant and the solution set is the imaginary axis.

Appendix B. Matrix factorization results. This appendix contains a self-contained proof of Lemma B.2, needed in the proof of Theorem 2.1, and some other matrix factorization results that have appeared in papers on the KYP lemma [1, 2, 23, 32, 35]. We include the proofs because their constructive character is important for the result in Theorem 2.1. Lemma B.1 is based on [35, Lemma 3] and [22, Lemma 5]. Lemma B.2 can be found in [32, Corollary 1].

LEMMA B.1. *Let U and V be two matrices in $\mathbf{C}^{p \times r}$.*

(a) *If* $UU^H = VV^H$, *then* $U = V\Lambda$ *for some unitary matrix* $\Lambda \in \mathbf{C}^{r\times r}$.
(b) *If* $UU^H = VV^H$ *and* $UV^H + VU^H = 0$, *then* $U = V\Lambda$ *for some unitary and skew-Hermitian matrix* $\Lambda \in \mathbf{C}^{r\times r}$.
(c) *If* $UU^H \preceq VV^H$ *and* $UV^H + VU^H = 0$, *then* $U = V\Lambda$ *for some skew-Hermitian matrix* $\Lambda \in \mathbf{C}^{r\times r}$ *with* $\|\Lambda\|_2 \leq 1$.

Proof. Part (a). If $UU^H = VV^H$, then U and V have singular value decompositions of the form

$$U = P\Sigma Q_u^H, \qquad V = P\Sigma Q_v^H \tag{63}$$

with unitary P, Q_u, Q_v. The unitary matrix $\Lambda = Q_v Q_u^H$ satisfies $U = V\Lambda$.

Part (b). If we substitute the singular value decompositions (63) in the equation $UV^H + VU^H = 0$, we obtain

$$\Sigma(Q_u^H Q_v + Q_v^H Q_u)\Sigma^T = 0. \tag{64}$$

If U and V, and therefore Σ, have full column rank, this implies that the matrix $\tilde{\Lambda} = Q_u^H Q_v$ is skew-Hermitian. The matrix $\Lambda = Q_v \tilde{\Lambda} Q_v^H = Q_v Q_u^H$ is skew-Hermitian and unitary, and satisfies $U = V\Lambda$. If U and V do not have full column rank, we modify $\tilde{\Lambda}$ as follows. We write (64) as

$$\begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \tilde{\Lambda}_{11} + \tilde{\Lambda}_{11}^H & \tilde{\Lambda}_{12} + \tilde{\Lambda}_{21}^H \\ \tilde{\Lambda}_{21} + \tilde{\Lambda}_{12}^H & \tilde{\Lambda}_{22} + \tilde{\Lambda}_{22}^H \end{bmatrix} \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix} = 0$$

with Σ_1 positive diagonal of size $q \times q$, where $q = \mathbf{rank}(U) = \mathbf{rank}(V)$, and $\tilde{\Lambda}_{11}$ is the $q \times q$ leading diagonal block of $\tilde{\Lambda}$. This shows that $\tilde{\Lambda}_{11} + \tilde{\Lambda}_{11}^H = 0$, so $\tilde{\Lambda}$ is unitary with a skew-Hermitian $1,1$ block. Since $\tilde{\Lambda}_{11}$ is skew-Hermitian, it has a Schur decomposition $\tilde{\Lambda}_{11} = Q\Delta Q^H$ with unitary $Q \in \mathbf{C}^{q\times q}$, and Δ is diagonal and purely imaginary. Moreover, $\Delta\Delta^H \preceq I$ because $\tilde{\Lambda}_{11}$ is a submatrix of the unitary matrix $\tilde{\Lambda}$. Partition Q and Δ as

$$\tilde{\Lambda}_{11} = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix} \begin{bmatrix} \Delta_1 & 0 \\ 0 & \Delta_2 \end{bmatrix} \begin{bmatrix} Q_1 & Q_2 \end{bmatrix}^H \tag{65}$$

with $\Delta_1\Delta_1^H \prec I$ and $\Delta_2\Delta_2^H = I$. Since $\tilde{\Lambda}$ is unitary, we have $\tilde{\Lambda}_{12}\tilde{\Lambda}_{12}^H = I - \tilde{\Lambda}_{11}\tilde{\Lambda}_{11}^H = Q_1(I - \Delta_1\Delta_1^H)Q_1^H$, and by part (a),

$$\tilde{\Lambda}_{12} = Q_1 \left(I - \Delta_1\Delta_1^H\right)^{1/2} \Omega \tag{66}$$

for some unitary Ω. Therefore the matrix

$$\begin{bmatrix} \tilde{\Lambda}_{11} & \tilde{\Lambda}_{12} \\ -\tilde{\Lambda}_{12}^H & \Omega^H\Delta_1^H\Omega \end{bmatrix} = \begin{bmatrix} Q_1 & Q_2 & 0 \\ 0 & 0 & \Omega^H \end{bmatrix} \begin{bmatrix} \Delta_1 & 0 & \Gamma \\ 0 & \Delta_2 & 0 \\ -\Gamma & 0 & \Delta_1^H \end{bmatrix} \begin{bmatrix} Q_1^H & 0 \\ Q_2^H & 0 \\ 0 & \Omega \end{bmatrix},$$

where $\Gamma = (I - \Delta_1\Delta_1^H)^{1/2}$, is skew-Hermitian (from the expression on the left-hand side and the fact that $\tilde{\Lambda}_{11}$ is skew-Hermitian and Δ_1 is purely imaginary) and unitary (the right-hand side is a product of three unitary matrices). If we now define

$$\Lambda = Q_v \begin{bmatrix} \tilde{\Lambda}_{11} & \tilde{\Lambda}_{12} \\ -\tilde{\Lambda}_{12}^H & \Omega^H\Delta_1^H\Omega \end{bmatrix} Q_v^H,$$

then Λ is unitary and skew-Hermitian, and

$$\begin{aligned} U &= P \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \tilde{\Lambda}_{11} & \tilde{\Lambda}_{12} \\ \tilde{\Lambda}_{21} & \tilde{\Lambda}_{22} \end{bmatrix} Q_v^H \\ &= P \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \tilde{\Lambda}_{11} & \tilde{\Lambda}_{12} \\ -\tilde{\Lambda}_{12}^H & \Omega^H \Delta_1^H \Omega \end{bmatrix} Q_v^H \\ &= P \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix} Q_v^H \Lambda \\ &= V\Lambda. \end{aligned}$$

Part (c). Assume $UU^H \preceq VV^H$ and $VV^H - UU^H$ has rank s. We factorize $VV^H - UU^H = \tilde{U}\tilde{U}^H$ with $\tilde{U} \in \mathbf{C}^{p\times s}$ and write $UU^H \preceq VV^H$ and $UV^H + VU^H = 0$ as

$$\begin{bmatrix} U & \tilde{U} \end{bmatrix} \begin{bmatrix} U & \tilde{U} \end{bmatrix}^H = \begin{bmatrix} V & 0 \end{bmatrix} \begin{bmatrix} V & 0 \end{bmatrix}^H \tag{67}$$

and

$$\begin{bmatrix} U & \tilde{U} \end{bmatrix} \begin{bmatrix} V & 0 \end{bmatrix}^H + \begin{bmatrix} V & 0 \end{bmatrix} \begin{bmatrix} U & \tilde{U} \end{bmatrix}^H = 0.$$

It follows from part (b) that there exists a unitary skew-Hermitian matrix $\tilde{\Lambda}$ for which

$$\begin{bmatrix} U & \tilde{U} \end{bmatrix} = \begin{bmatrix} V & 0 \end{bmatrix} \begin{bmatrix} \tilde{\Lambda}_{11} & \tilde{\Lambda}_{12} \\ \tilde{\Lambda}_{21} & \tilde{\Lambda}_{22} \end{bmatrix}.$$

The subblock $\Lambda = \tilde{\Lambda}_{11}$ satisfies $U = V\Lambda$, $\Lambda + \Lambda^H = 0$, and $\Lambda^H\Lambda \preceq I$. ∎

LEMMA B.2. *Let* Φ, $\Psi \in \mathbf{H}^2$ *with* $\det \Phi < 0$. *If* $U, V \in \mathbf{C}^{p\times r}$ *satisfy*

$$\Phi_{11}UU^H + \Phi_{21}UV^H + \Phi_{12}VU^H + \Phi_{22}VV^H = 0, \tag{68}$$

$$\Psi_{11}UU^H + \Psi_{21}UV^H + \Psi_{12}VU^H + \Psi_{22}VV^H \preceq 0, \tag{69}$$

then there exist a $W \in \mathbf{C}^{p\times r}$, *a unitary* $Q \in \mathbf{C}^{r\times r}$, *and vectors* $\mu, \nu \in \mathbf{C}^r$ *such that*

$$U = W\,\mathbf{diag}(\mu)Q^H, \qquad V = W\,\mathbf{diag}(\nu)Q^H,$$

and $q_\Phi(\mu_i, \nu_i) = 0$, $q_\Psi(\mu_i, \nu_i) \le 0$, $(\mu_i, \nu_i) \ne 0$ *for* $i = 1, \dots, r$.

Proof. Suppose U and V are $p \times r$ matrices that satisfy (68) and (69). As explained in Appendix A, there exists a nonsingular R such that

$$\Phi = R^H \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} R, \qquad \Psi = R^H \begin{bmatrix} \alpha & \beta \\ \beta & \gamma \end{bmatrix} R$$

with β real and $\gamma \le \alpha$. Define $S = R_{11}U + R_{12}V$ and $T = R_{21}U + R_{22}V$. Then

$$ST^H + TS^H = \begin{bmatrix} U & V \end{bmatrix} \begin{bmatrix} \Phi_{11}I & \Phi_{21}I \\ \Phi_{12}I & \Phi_{22}I \end{bmatrix} \begin{bmatrix} U^H \\ V^H \end{bmatrix} = 0 \tag{70}$$

and

$$\alpha SS^H + \gamma TT^H = \begin{bmatrix} U & V \end{bmatrix} \begin{bmatrix} \Psi_{11}I & \Psi_{21}I \\ \Psi_{12}I & \Psi_{22}I \end{bmatrix} \begin{bmatrix} U^H \\ V^H \end{bmatrix} \preceq 0. \tag{71}$$

We show that this implies that $S = W\,\mathbf{diag}(s)Q^H$, $T = W\,\mathbf{diag}(t)Q^H$ for some $W \in \mathbf{C}^{p\times r}$, unitary $Q \in \mathbf{C}^{r\times r}$, and vectors $s, t \in \mathbf{C}^r$ that satisfy

$$s_i\bar{t}_i + \bar{s}_i t_i = 0, \qquad \alpha|s_i|^2 + \gamma|t_i|^2 \leq 0, \qquad (s_i, t_i) \neq 0, \qquad i = 1, \dots, r.$$

The result is trivial if S and T are zero, since in that case we can choose $W = 0$ and arbitrary Q, s, t. If at least one of the two matrices is nonzero, then (71), combined with $\alpha \geq \gamma$, implies that $\gamma \leq 0$. Therefore, there are three cases to consider.

- If $\alpha \leq 0$, we write (70) as $(S+T)(S+T)^H = (S-T)(S-T)^H$. From Lemma B.1, $S+T = (S-T)\Lambda$ with Λ unitary. Let $\Lambda = Q\,\mathbf{diag}(\rho)Q^H$ be the Schur decomposition of Λ with $|\rho_i| = 1$ for $i = 1, \dots, r$. Define $W = (S-T)Q$, $s = (1/2)(\rho + \mathbf{1})$, and $t = (1/2)(\rho - \mathbf{1})$.
- If $\gamma = 0 < \alpha$, then $S = 0$, and we can take $Q = I$, $W = T$, $s = 0$, and $t = \mathbf{1}$.
- If $\gamma < 0 < \alpha$, then from Lemma B.1, $S = (-\gamma/\alpha)^{1/2}T\Lambda$ for some skew-Hermitian Λ with $\Lambda^H\Lambda \preceq I$. This matrix has a Schur decomposition $\Lambda = Q\,\mathbf{diag}(\rho)Q^H$ with $|\rho_i| \leq 1$ for $i = 1, \dots, r$. Define $W = TQ$, $s = (-\gamma/\alpha)^{1/2}\rho$, and $t = \mathbf{1}$.

The factorizations of U and V now follow from

$$\begin{bmatrix} U \\ V \end{bmatrix} = (R^{-1} \otimes I) \begin{bmatrix} S \\ T \end{bmatrix} = (R^{-1} \otimes I) \begin{bmatrix} W\,\mathbf{diag}(s) \\ W\,\mathbf{diag}(t) \end{bmatrix} Q^H = \begin{bmatrix} W\,\mathbf{diag}(\mu) \\ W\,\mathbf{diag}(\nu) \end{bmatrix} Q^H,$$

where μ and ν are defined as $(\mu_i, \nu_i) = R^{-1}(s_i, t_i)$ for $i = 1, \dots, r$. These pairs (μ_i, ν_i) are nonzero and satisfy $q_\Phi(\mu_i, \nu_i) = 0$ and $q_\Psi(\mu_i, \nu_i) \leq 0$. ∎

Appendix C. Strict feasibility. In this appendix, we discuss strict feasibility of the constraints $X \succeq 0$, (11), and (12) in Theorem 2.1. We assume that the set $\mathcal{C}$ defined in (7) is not empty and not a singleton. This means that if the inequality $q_\Psi(\mu, \nu) \leq 0$ in the definition is not redundant, then there exist points in $\mathcal{C}$ with $q_\Psi(\mu, \nu) < 0$. We will distinguish these two cases.

- *Line or circle.* If the inequality $q_\Psi(\mu, \nu) \leq 0$ is redundant, we have $\mathcal{C} = \{(\mu, \nu) \in \mathbf{C}^2 \mid (\mu, \nu) \neq 0,\ q_\Phi(\mu, \nu) = 0\}$, a line or circle in homogeneous coordinates. In this case we understand by strict feasibility of X that

  $$(72) \quad X \succ 0, \quad \Phi_{11}FXF^H + \Phi_{21}FXG^H + \Phi_{12}GXF^H + \Phi_{22}GXG^H = 0.$$

  We also define $\mathcal{C}^\circ = \mathcal{C}$.
- *Segment of line or circle.* In the second case, $\mathcal{C}$ is a proper one-dimensional subset of the line or circle defined by $q_\Phi(\mu, \nu) = 0$. In this case we define strict feasibility of X as

  $$(73) \qquad (72), \quad \Psi_{11}FXF^H + \Psi_{21}FXG^H + \Psi_{12}GXF^H + \Psi_{22}GXG^H \prec 0.$$

  We also define $\mathcal{C}^\circ = \{(\mu, \nu) \neq 0 \mid q_\Phi(\mu, \nu) = 0,\ q_\Psi(\mu, \nu) < 0\}$.

The conditions on F and G that guarantee strict feasibility will be expressed in terms of the Kronecker structure of the matrix pencil $\lambda G - F$ [12, 43]. For every matrix pencil, there exist nonsingular matrices P and Q such that

$$(74) \qquad P(\lambda G - F)Q = \mathbf{diag}\left(L_{\eta_1}(\lambda)^T, \dots, L_{\eta_l}(\lambda)^T, \lambda B - A, L_{\epsilon_1}(\lambda), \dots, L_{\epsilon_r}(\lambda)\right),$$

where **diag** represents the block-diagonal operator, $L_\epsilon(\lambda)$ is the $\epsilon \times (\epsilon+1)$ pencil

$$L_\epsilon(\lambda) = \begin{bmatrix} \lambda & -1 & 0 & \cdots & 0 & 0 \\ 0 & \lambda & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & 0 & 0 \\ 0 & 0 & 0 & \cdots & -1 & 0 \\ 0 & 0 & 0 & \cdots & \lambda & -1 \end{bmatrix},$$

and $\lambda B - A$ is a regular pencil, i.e., it is square and $\det(\lambda B - A)$ is not identically zero. The parameters $\epsilon_1,\ldots,\epsilon_r$, and $\eta_1,\ldots,\eta_l$ are the right and the left Kronecker indices of the pencil, respectively. The *normal rank* of the pencil is $p-l$, where p is the row dimension.

We show that there exists a strictly feasible X if and only if the following two conditions hold.

1. The normal rank of $\lambda G - F$ is p. This means that $l = 0$ in (74).
2. The generalized eigenvalues of $\lambda B - A$ are nondefective and lie in $\mathcal{C}^\circ$. (More accurately, if λ is a finite generalized eigenvalue, then $(\lambda, 1) \in \mathcal{C}^\circ$. If it is an infinite generalized eigenvalue, then $(1,0) \in \mathcal{C}^\circ$.)

A sufficient but more easily verified condition is that $\mathbf{rank}\,(\mu G - \nu F) = p$ for all $(\mu,\nu) \neq 0$, i.e., $l = 0$ and the block $\lambda B - A$ in (74) is not present.

Proof. Without loss of generality, we assume that the pencil is in the Kronecker canonical form ($P = I$, $Q = I$ in (74)) and $\Phi = \Phi_{\mathrm{u}}$, so the equality in (72) is

$$FXF^H = GXG^H. \tag{75}$$

We first show that the conditions are necessary. Assume X is strictly feasible. Partition X as an $(l+1+r)\times(l+1+r)$ block matrix, with block dimensions equal to the column dimensions of the $l+1+r$ block columns in (74). Suppose $l \geq 1$ and consider the kth diagonal block X_{kk} with $1 \leq k \leq l$. The kth diagonal block of the pencil is

$$\lambda G_k - F_k = L_{\eta_k}(\lambda)^T = \lambda \begin{bmatrix} I_{\eta_k} \\ 0_{1\times\eta_k} \end{bmatrix} - \begin{bmatrix} 0_{1\times\eta_k} \\ I_{\eta_k} \end{bmatrix}.$$

The kth diagonal block of (75) is $F_k X_{kk} F_k^H = G_k X_{kk} G_k^H$ or

$$\begin{bmatrix} 0_{1\times\eta_k} \\ I_{\eta_k} \end{bmatrix} X_{kk} \begin{bmatrix} 0_{\eta_k\times 1} & I_{\eta_k} \end{bmatrix} = \begin{bmatrix} I_{\eta_k} \\ 0_{1\times\eta_k} \end{bmatrix} X_{kk} \begin{bmatrix} I_{\eta_k} & 0_{\eta_k\times 1} \end{bmatrix}.$$

This is impossible, since $X_{kk} \succ 0$. Hence, if (75) holds with $X \succ 0$, then $l = 0$.

Next suppose $\det(\mu B - \nu A) = 0$ for some $(\mu,\nu) \neq 0$. If $\nu \neq 0$, then μ/ν is a finite generalized eigenvalue of the pencil $\lambda B - A$; if $\nu = 0$, then the pencil has a generalized eigenvalue at infinity. Let y be a corresponding left generalized eigenvector, i.e., $y^H(\mu B - \nu A) = 0$, while $y^H B$ and $y^H A$ are not both zero (since $y^H B = y^H A = 0$ would imply that the pencil $\lambda B - A$ is singular). Define $u^H = y^H B$ if $\nu \neq 0$ and $u^H = y^H A$ otherwise. This is a nonzero vector. The first diagonal block of (75) is

$$AX_{11}A^H = BX_{11}B^H. \tag{76}$$

From this it follows that $|\mu|^2 u^H X_{11} u = |\nu|^2 u^H X_{11} u$, and, since $X_{11} \succ 0$, we have $q_\Phi(\mu,\nu) = |\mu|^2 - |\nu|^2 = 0$, i.e., the generalized eigenvalues are on the unit circle. In addition, if the inequality in (73) holds, then

$$\Psi_{11} AX_{11}A^H + \Psi_{21} AX_{11}B^H + \Psi_{12} BX_{11}A^H + \Psi_{22} BX_{11}B^H \prec 0,$$

and from this, $q_\Psi(\mu,\nu)(u^H X_{11} u) < 0$. This is only possible if $q_\Psi(\mu,\nu) < 0$. We conclude that if $\det(\mu B - \nu A) = 0$ for nonzero (μ,ν), then $(\mu,\nu) \in \mathcal{C}^\circ$.

Next we show that the generalized eigenvalues of the pencil $\lambda B - A$ are nondefective. Since C° is the unit circle or a subset of the unit circle, there are no infinite generalized eigenvalues. Assume the pencil is in Weierstrass canonical form, i.e.,

$$\lambda B - A = \mathbf{diag}\left((\lambda - \rho_1)I - J_{s_1}, (\lambda - \rho_2)I - J_{s_2}, \ldots, (\lambda - \rho_t)I - J_{s_t}\right),$$

where $\rho_1,\ldots,\rho_t$ are the generalized eigenvalues (which satisfy $|\rho_i| = 1$), and J_s is the $s \times s$ matrix with 1s on the first superdiagonal and zeros elsewhere. Then (76) implies that $(\rho_i I + J_{s_i})X_{11,i}(\rho_i I + J_{s_i})^H = X_{11,i}$, where $X_{11,i}$ is the ith diagonal block of X_{11} if we partition X_{11} as a $t \times t$ block matrix with i,j block of size of $s_i \times s_j$. Since $|\rho_i| = 1$, this simplifies to

$$\rho_i X_{11,i} J_{s_i}^T + \bar{\rho}_i J_{s_i} X_{11,i} + J_{s_i} X_{11,i} J_{s_i}^T = 0.$$

The last rows of the second and third matrices are zero. Therefore, the last row of the first matrix is zero. However, the element in column $s_i - 1$ is the last diagonal element of the positive definite matrix $X_{11,i}$. This is a contradiction unless $s_i = 1$, i.e., the eigenvalue ρ_i is nondefective. We conclude that the two conditions are necessary.

It remains to show sufficiency. Suppose $\lambda G - F$ has the Kronecker canonical form

$$\lambda G - F = \mathbf{diag}\left(\lambda - \rho_1, \ldots, \lambda - \rho_t, L_{\epsilon_1}(\lambda), \ldots, L_{\epsilon_r}(\lambda)\right)$$

with $\rho_i \in \mathcal{C}^\circ$. Define $X = \mathbf{diag}\left(1, \ldots, 1, X_{11}, \ldots, X_{rr}\right)$ with diagonal blocks

$$X_{kk} = \sum_{i=1}^{\epsilon_k+1} \begin{bmatrix} 1 \\ \lambda_{ki} \\ \vdots \\ \lambda_{ki}^{\epsilon_k} \end{bmatrix} \begin{bmatrix} 1 \\ \lambda_{ki} \\ \vdots \\ \lambda_{ki}^{\epsilon_k} \end{bmatrix}^H, \qquad k = 1, \ldots, r,$$

where $\lambda_{k1},\ldots,\lambda_{k,\epsilon_k+1}$ are distinct and in $\mathcal{C}^\circ$. This matrix X is strictly feasible. ∎

Acknowledgments. We would like to thank the associate editor and the two referees for their very useful comments.

REFERENCES

[1] V. Balakrishnan and L. Vandenberghe, *Semidefinite Programming Duality and Linear Time-Invariant Systems*, Technical report TR-ECE-02-02, School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN, 2002.

[2] V. Balakrishnan and L. Vandenberghe, *Semidefinite programming duality and linear time-invariant systems*, IEEE Trans. Automat. Control, 48 (2003), pp. 30–41.

[3] B. N. Bhaskar, G. Tang, and B. Recht, *Atomic norm denoising with applications to line spectral estimation*, IEEE Trans. Signal Process., 61 (2013), pp. 5987–5999.

[4] E. J. Candès and C. Fernandez-Granda, *Towards a mathematical theory of super-resolution*, Comm. Pure Appl. Math., 67 (2014), pp. 906–956.

[5] V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. S. Willsky, *The convex geometry of linear inverse problems*, Found. Comput. Math., 12 (2012), pp. 805–849.

[6] H.-H. Chao and L. Vandenberghe, *Extensions of semidefinite programming methods for atomic decomposition*, in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2016.

[7] H.-H. Chao and L. Vandenberghe, *Semidefinite Representations of Gauge Functions for Structured Low-rank Matrix Decomposition*, preprint, arXiv:1604.02500, 2016.

[8] Y. de Castro and F. Gamboa, *Exact reconstruction using Beurling minimal extrapolation*, J. Math. Anal. Appl., 395 (2012), pp. 336–354.

[9] Y. de Castro, F. Gamboa, D. Henrion, and J. B. Lasserre, *Exact solutions to super resolution on semi-algebraic domains in higher dimensions*, IEEE Trans. Inform. Theory, 63 (2017), pp. 621–630.

[10] C. Fernandez-Granda, *Super-resolution of Point Sources via Convex Programming*, arXiv:1507.07034, 2015.

[11] M. P. Friedlander, I. Macêdo, and T. K. Pong, *Gauge optimization and duality*, SIAM J. Optim., 24 (2014), pp. 1999–2022.

[12] F. R. Gantmacher, *Applications of the Theory of Matrices*, Dover, New York, 2005. Originally published in 1959 by Interscience, New York.

[13] Y. Genin, Y. Hachez, Yu. Nesterov, and P. Van Dooren, *Optimization problems over positive pseudopolynomial matrices*, SIAM J. Matrix Anal. Appl., 25 (2003), pp. 57–79.

[14] T. T. Georgiou, *Decomposition of Toeplitz matrices via convex optimization*, IEEE Signal Process. Lett., 13 (2006), pp. 537–540.

[15] G. H. Golub and J. Kautsky, *Calculation of Gauss quadratures with multiple free and fixed knots*, Numer. Math., 41 (1983), pp. 147–163.

[16] M. Grant and S. Boyd, *CVX: MATLAB Software for Disciplined Convex Programming, version* 2.1. http://cvxr.com/cvx, Mar. 2014.

[17] U. Grenander and G. Szegö, *Toeplitz Forms and Their Applications*, Chelsea, New York, 1984. First published in 1958.

[18] A. Hansson and L. Vandenberghe, *Sampling method for semidefinite programmes with nonnegative Popov function constraints*, Internat. J. Control, 87 (2014), pp. 330–345.

[19] B. Hassibi, A. H. Sayed, and T. Kailath, *Indefinite-Quadratic Estimation and Control: A Unified Approach to H^2 and H^∞ Theories*, SIAM, Philadelphia, PA, 1999.

[20] Y. Hua and T. K. Sarkar, *Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise*, IEEE Trans. Acoust. Speech Signal Process., 38 (1990), pp. 814–824.

[21] V. Ionescu, C. Oară, and M. Weiss, *Generalized Riccati Theory and Robust Control*, John Wiley and Sons, New York, 1999.

[22] T. Iwasaki and S. Hara, *Generalized KYP lemma: Unified frequency domain inequalities with design applications*, IEEE Trans. Automat. Control, 50 (2005), pp. 41–59.

[23] T. Iwasaki, G. Meinsma, and M. Fu, *Generalized S-procedure and finite-frequency KYP lemma*, Math. Probl. Eng., 6 (2000), pp. 305–320.

[24] R. E. Kalman, *Lyapunov functions for the problem of Lur'e in automatic control*, Proc. Natl. Acad. Sci. USA, 49 (1963), pp. 201–205.

[25] S. Karlin and W. J. Studden, *Tchebycheff Systems: With Applications in Analysis and Statistics*, Interscience, New York, 1966.

[26] M. G. Krein and A. A. Nudelman, *The Markov Moment Problem and Extremal Problems*, American Mathematical Society, Providence, RI, 1977.

[27] Y. Li and Y. Chi, *Off-the-grid line spectrum denoising and estimation with multiple measurement vectors*, IEEE Trans. Signal Process., 64 (2016), pp. 1257–1269.

[28] 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.

[29] 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.

[30] K. V. Mishra, M. Cho, A. Kruger, and W. Xu, *Off-the-grid spectral compressed sensing with prior information*, in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2014.

[31] K. V. Mishra, M. Cho, A. Kruger, and W. Xu, *Spectral super-resolution with prior knowledge*, IEEE Trans. Signal Process., 63 (2015), pp. 5342–5357.

[32] G. Pipeleers and L. Vandenberghe, *Generalized KYP lemma with real data*, IEEE Trans. Automat. Control, 56 (2011), pp. 2942–2946.

[33] V. M. Popov, *Absolute stability of nonlinear systems of automatic control*, Autom. Remote Control, 22 (1962), pp. 857–875.

[34] J. G. Proakis and D. G. Manolakis, *Digital Signal Processing: Principles, Algorithms, and Applications*, 3rd ed., Prentice-Hall, Upper Saddle River, NJ, 1996.

[35] A. Rantzer, *On the Kalman–Yakubovich–Popov lemma*, Syst. Control Lett., 28 (1996), pp. 7–10.

[36] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.

[37] T. Roh and L. Vandenberghe, *Discrete transforms, semidefinite programming, and sum-of-squares representations of nonnegative polynomials*, SIAM J. Optim., 16 (2006), pp. 939–964.
[38] T. K. Sarkar and O. Pereira, *Using the matrix pencil method to estimate the parameters of a sum of complex exponentials*, IEEE Antennas Propag. Mag., 37 (1995), pp. 48–55.
[39] C. Scherer, *LMI relaxations in robust control*, Eur. J. Control, 12 (2006), pp. 3–29.
[40] P. Stoica and R. L. Moses, *Introduction to Spectral Analysis*, Prentice-Hall, Upper Saddle River, NJ, 1997.
[41] P. Stoica, G. Tang, Z. Yang, and D. Zachariah, *Gridless compressive-sensing methods for frequency estimation: Points of tangency and links to basics*, in Proceedings of the 22nd European Signal Processing Conference (EUSIPCO), 2014, pp. 1831–1835.
[42] G. Tang, B. N. Bhaskar, P. Shah, and B. Recht, *Compressed sensing off the grid*, IEEE Trans. Inform. Theory, 59 (2013), pp. 7465–7490.
[43] P. Van Dooren, *The computation of Kronecker's canonical form of a singular pencil*, Linear Algebra Appl., 27 (1979), pp. 103–140.
[44] V. A. Yakubovich, *The solution of certain matrix inequalities in automatic control theory*, Sov. Math. Dokl., 3 (1962), pp. 620–623.
[45] Z. Yang and L. Xie, *Exact joint sparse frequency recovery via optimization methods*, IEEE Tran. Signal Process., 64 (2016), pp. 5145–5157.