Semidefinite representations of gauge functions for structured low-rank matrix decomposition

Hsiao-Han Chao*             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.

1 Introduction

The notion of atomic norm introduced in [CRPW12] gives a unified description of convex penalty functions that extend the $\ell_1$-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 non-empty set $C$ is defined as the gauge of its convex hull, i.e., the convex function

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

This function is convex, nonnegative, positively homogeneous, and zero if $x = 0$. 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 $\ell_1$-norm and the matrix trace norm. The $\ell_1$-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 $k$th 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 $\mathbb{C}^{n \times m}$ is the atomic norm for $C = \{ vv^H \mid \|v\| = \|w\| = 1 \}$, where $w^H$ is the conjugate

*Electrical Engineering Department, University of California, Los Angeles. Email: suzihchao@ucla.edu, vandenbe@ucla.edu. Research partially supported by NSF Grants 1128817 and 1509789.
transpose and $\| \cdot \|$ denotes the Euclidean norm. Many other examples are discussed in [CRPW12, BTR12, TBSR13].

The atomic norm associated with the set

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

where $j = \sqrt{-1}$, has been studied extensively in recent research in signal processing [dCG12, GHL15, CFG14, BTR12, TBSR13, YX14, LC14, YX15, CC15]. It is known that the atomic norm for this set is the optimal value of the semidefinite program (SDP)

$$\begin{align*}
\text{minimize} & \quad (\text{tr } V + w)/2 \\
\text{subject to} & \quad \begin{bmatrix} V & x \\ x^H & w \end{bmatrix} \succeq 0 \\
& \quad V \text{ is Toeplitz,}
\end{align*}$$

(3)

with variables $w$ and $V \in \mathbb{H}^n$ (the $n \times n$ Hermitian matrices). This result can be proved via convex duality and semidefinite characterizations of bounded trigonometric polynomials [dCG12], or directly by referring to Carathéodory’s decomposition of positive semidefinite Toeplitz matrices [TBSR13]. More generally, one can consider the atomic norm of the set of matrices

$$C = \{ vw^H \in \mathbb{C}^{n \times m} \mid v \in C_e, \ |w| = 1 \}. $$

The atomic norm for this set, evaluated at a matrix $X \in \mathbb{C}^{n \times m}$, is the optimal value of the SDP

$$\begin{align*}
\text{minimize} & \quad (\text{tr } V + \text{tr } W)/2 \\
\text{subject to} & \quad \begin{bmatrix} V & X \\ X^H & W \end{bmatrix} \succeq 0 \\
& \quad V \text{ is Toeplitz,}
\end{align*}$$

(4)

with variables $V \in \mathbb{H}^n$ and $W \in \mathbb{H}^m$; see [YX14, LC14, FG15]. Further extensions, that place restrictions on the parameter $\omega$ in the definition (2), can be found in [MCKX14, MCKX15].

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

$$C_e = \{ a \mid (\lambda G - F)a = 0, \ \lambda \in \mathbb{C}, \ |a| = 1 \}$$

(5)

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

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

We generalize (5) in three ways and derive semidefinite representations of the corresponding atomic norms. The first generalization is to replace $\lambda G - F$ with an arbitrary matrix pencil. Second, we allow $\mathbb{C}$ 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 $\|Ea\| \leq 1$ where $E$ is not necessarily full column rank. Specific examples of these extensions, with different choices of $F$, $G$, and $\mathbb{C}$, 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 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 $\Phi$ and $\Psi$ are Hermitian $2 \times 2$ matrices with $\det \Phi < 0$. We define

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

where

$$\mathcal{C} = \{ (\mu, \nu) \in \mathbb{C}^2 \mid (\mu, \nu) \neq 0, \ q_\Phi(\mu, \nu) = 0, \ q_\Psi(\mu, \nu) \leq 0 \}.$$  \hspace{1cm} (7)

Here $q_\Phi$, $q_\Psi$ are the quadratic forms defined by $\Phi$ and $\Psi$:

$$q_\Phi(\mu, \nu) = \begin{bmatrix} \mu \\ \nu \end{bmatrix}^H \Phi \begin{bmatrix} \mu \\ \nu \end{bmatrix}, \quad q_\Psi(\mu, \nu) = \begin{bmatrix} \mu \\ \nu \end{bmatrix}^H \Psi \begin{bmatrix} \mu \\ \nu \end{bmatrix}. \hspace{1cm} (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 $A$ as

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

If $\Phi_{11} = 0$ and $\Psi_{11} \leq 0$, then the pair $(1, 0)$ is also in $\mathcal{C}$ and the set $A$ in (9) is the union of the right-hand side of (9) and the nullspace of $G$. Examples of sets $A$ 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 $a \in A$, i.e., the set

$$\text{conv} \{ aa^H \mid a \in A \} = \{ \sum_{k=1}^m a_k a_k^H \mid a \in A \}. \hspace{1cm} (10)

2.1 Conic decomposition

The key decomposition result (Theorem 1) is known under various forms in system theory, signal processing, and moment theory [KS66, KN77, GS84]. 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 1** Let $A$ be defined by (9) and (7), where $F, G \in \mathbb{C}^{p \times n}$ and $\Phi, \Psi \in \mathbb{H}^2$ with $\det \Phi < 0$. If $X \in \mathbb{H}^n$ is a positive semidefinite matrix of rank $r \geq 1$ that satisfies

$$\Phi_{11} FXF^H + \Phi_{21} FXG^H + \Phi_{12} GXF^H + \Phi_{22} GXG^H = 0 \hspace{1cm} (11)$$

$$\Psi_{11} FXF^H + \Psi_{21} FXG^H + \Psi_{12} GXF^H + \Psi_{22} GXG^H \preceq 0, \hspace{1cm} (12)$$

then $X \in A$.\]
then $X$ can be decomposed as

$$X = \sum_{k=1}^{r} a_k a_k^H,$$

with linearly independent vectors $a_1, \ldots, a_r \in A$.

Proof. We start from any factorization $X = YY^H$ where $Y \in \mathbb{C}^{n \times r}$ has rank $r$. It follows from Lemma 2 in appendix B applied to the matrices $U = FY$ and $V = GY$, that there exist a matrix $W \in \mathbb{C}^{p \times r}$, a unitary matrix $Q \in \mathbb{C}^{r \times r}$, and two vectors $\mu, \nu \in \mathbb{C}^r$ such that

$$FYQ = W \text{diag}(\mu), \quad GYQ = W \text{diag}(\nu), \quad (\mu_i, \nu_i) \in \mathcal{C}, \quad i = 1, \ldots, r.$$  \quad (14)

Choosing $a_k$ equal to the $k$th column of $YQ$ gives the decomposition (13). \hfill \Box

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 2 in the appendix is constructive and gives a simple algorithm, based on singular value and Schur decompositions, for computing the matrices $W$, $Q$ and the vectors $\mu$, $\nu$. In the following three sections we illustrate the decomposition in Theorem 1 with different choices of $F$, $G$, $\Phi$, $\Psi$.

2.2 Trigonometric polynomials

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

$$F = \begin{bmatrix} 0 & I_{n-1} \end{bmatrix}, \quad G = \begin{bmatrix} I_{n-1} & 0 \end{bmatrix}, \quad \Phi = \Phi_u = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \quad \Psi = 0. \quad (15)$$

A nonzero pair $(\mu, \nu)$ satisfies $q_{\Phi}(\mu, \nu) = |\mu|^2 - |\nu|^2 = 0$ only if $\mu$ and $\nu$ are nonzero and $\lambda = \mu/\nu$ is on the unit circle. The condition $(\lambda G - F)a = 0$ in the definition of $A$ gives a recursion

$$\lambda a_1 = a_2, \quad \lambda a_2 = a_3, \quad \ldots, \quad \lambda a_{n-1} = a_n.$$  

Defining $\exp(j\omega) = \lambda$, we find that $A$ contains the vectors

$$a = c(1, e^{j\omega}, e^{j2\omega}, \ldots, e^{j(n-1)\omega}),$$  \quad (16)

for all $\omega \in [0, 2\pi)$ and $c \in \mathbb{C}$. The matrix constraints (11)–(12) reduce to $FXF^H = GXG^H$, i.e., $X$ is a Toeplitz matrix. Theorem 1 therefore states that every $n \times n$ positive semidefinite Toeplitz matrix can be decomposed as

$$X = \sum_{k=1}^{r} |c_k|^2 \begin{bmatrix} 1 \\ e^{j\omega_k} \\ e^{j2\omega_k} \\ \vdots \\ e^{j(n-1)\omega_k} \end{bmatrix} \begin{bmatrix} 1 \\ e^{j\omega_k} \\ e^{j2\omega_k} \\ \vdots \\ e^{j(n-1)\omega_k} \end{bmatrix}^H,$$  \quad (17)

with $c_k \neq 0$ and distinct $\omega_1, \ldots, \omega_r$. This is often called the Carathéodory parameterization of positive semidefinite Toeplitz matrices [SM97, page 170].
For this example, the algorithm outlined in the proof of Theorem 1 and Lemma 2 reduces to the following. Compute a factorization $X = YY^H$ where $Y \in \mathbb{C}^{n \times r}$ with rows $y_k^H, k = 1, \ldots, n$. Then find a unitary $r \times r$ matrix $\Lambda$ that satisfies

$$
\begin{bmatrix}
y_2^H \\
\vdots \\
y_n^H
\end{bmatrix} =
\begin{bmatrix}
y_1^H \\
\vdots \\
y_{n-1}^H
\end{bmatrix} \Lambda,
$$

and compute a Schur decomposition $\Lambda = Q \operatorname{diag}(\lambda) Q^H$. The eigenvalues give $\lambda_k = \exp(j\omega_k)$, $k = 1, \ldots, r$, and the columns of $YQ$ are the vectors $a_k$.

**Restricted complex exponentials** Define $F$, $G$, $\Phi$ as in (15), and

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

with $\alpha \in [0, 2\pi)$ and $\beta \in [0, \pi)$. The elements $a \in A$ have the same general form (16), with the added constraint that $\cos \beta \leq \cos(\omega - \alpha)$. Since we can restrict $\omega$ to the interval $[\alpha - \pi, \alpha + \pi]$, this is equivalent to $|\omega - \alpha| \leq \beta$. The constraints (11)–(12) specify that $X$ is Toeplitz and satisfies the matrix inequality

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

(18)

The theorem states that a positive semidefinite Toeplitz matrix of rank $r$ satisfies (18) if and only if it can be decomposed as (17) with nonzero $c_k$ and $|\omega_k - \alpha| \leq \beta$ for $k = 1, \ldots, r$.

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

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

and

$$
\Phi = \Phi_r = \begin{bmatrix} 0 & j \\
-j & 0 \end{bmatrix}, \\
\Psi = \Phi_u = \begin{bmatrix} 1 & 0 \\
0 & -1 \end{bmatrix}.
$$

A nonzero pair $(\mu, \nu)$ satisfies $q_\Phi(\mu, \nu) = j(\bar{\mu}\nu - \mu\bar{\nu}) = 0$ and $q_\Phi(\mu, \nu) = |\mu|^2 - |\nu|^2 \leq 0$ only if $\nu \neq 0$ and $\lambda = \mu/\nu$ is real with $|\lambda| \leq 1$. The condition $(\lambda G - F)a = 0$ gives a recursion

$$
\lambda a_1 = a_2, \\
2\lambda a_2 = a_1 + a_3, \\
\ldots, \\
2\lambda a_{n-1} = a_{n-2} + a_n.
$$

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 $G = [2I_{n-1} \ 0]$, the condition $(\lambda G - F)a = 0$ reduces to

$$
2\lambda a_1 = a_2, \\
2\lambda a_2 = a_1 + a_3, \\
\ldots, \\
2\lambda a_{n-1} = a_{n-2} + a_n.
$$

5
If we write \( \lambda = \cos \omega \), the solutions are the vectors

\[
a = c \left( 1, \sin 2\omega, \sin 3\omega, \ldots, \sin n\omega \right),
\]

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

**Trigonometric vector polynomials** We take \( p = (k-1)l, n = kl \), and replace \( F \) and \( G \) in (15) with

\[
F = \begin{bmatrix}
0 & I & 0 & \cdots & 0 \\
0 & 0 & I & \cdots & 0 \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & I
\end{bmatrix}, \quad G = \begin{bmatrix}
I & 0 & \cdots & 0 & 0 \\
0 & I & \cdots & 0 & 0 \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
0 & 0 & \cdots & I & 0
\end{bmatrix},
\]

and blocks of size \( l \times l \). Then \( \mathcal{A} \) contains the vectors of the form

\[
a = (1, e^{j\omega}, e^{j2\omega}, \ldots, e^{j(k-1)\omega}) \otimes c,
\]

for all \( c \in \mathbb{C}^l \) and \( \omega \in [0, 2\pi) \), where \( \otimes \) denotes Kronecker product.

**2.3 Polynomials**

**Real powers** Next, define \( F, G \) as in (15), and

\[
\Phi = \Phi_r = \begin{bmatrix} 0 & j \\
-j & 0 \end{bmatrix}, \quad \Psi = 0.
\]

An array \((\mu, \nu)\) satisfies \( q_\Phi(\mu, \nu) = 0 \) if and only if \( \mu \nu \) is real. If \((\mu, \nu) \neq 0\), we either have \( \nu = 0 \) and \( \mu \) arbitrary, or \( \nu \neq 0 \) and \( \lambda = \mu/\nu \) real. The set \( \mathcal{A} \) therefore contains the vectors

\[
a = c (1, \lambda, \lambda^2, \ldots, \lambda^{n-1}), \quad a = c (0, 0, \ldots, 0, 1)
\]

for all \( \lambda \in \mathbb{R} \) and \( c \). The matrix constraints (11)–(12) reduce to \( FXF^H = GXF^H \), i.e., \( X \) is a symmetric (real) Hankel matrix. Hence, a real symmetric positive semidefinite Hankel matrix of rank \( r \) can be decomposed in one of two forms

\[
X = \sum_{k=1}^{r} c_k^2 \begin{bmatrix}
1 & 1 & 1 & \cdots \\
\lambda_k & \lambda_k & \lambda_k & \cdots \\
\vdots & \vdots & \vdots & \ddots \\
\lambda_k^{n-2} & \lambda_k^{n-2} & \lambda_k^{n-2} & \cdots & \lambda_k^{n-1}
\end{bmatrix} + |c_r|^2 \begin{bmatrix} 0 \\
0 \\
\vdots \\
1
\end{bmatrix},
\]

with distinct real \( \lambda_k \) and nonzero \( c_k \).

**Restricted polynomials** If \( \Psi = 0 \) in (19) is replaced by

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

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

\[
2FF^H - (\alpha + \beta)(XF^H + GX^H) + 2\alpha\beta XX^H \preceq 0.
\]
Orthogonal polynomials  Let \( p_0(\lambda), p_1(\lambda), p_2(\lambda), \ldots \) be a sequence of real polynomials on \( \mathbb{R} \), 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

\[
\langle f(\lambda), \lambda g(\lambda) \rangle = \langle \lambda f(\lambda), g(\lambda) \rangle \tag{20}\]

(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 recursion

\[
\beta_{i+1}p_{i+1}(\lambda) = (\lambda - \alpha_i)p_i(\lambda) - \beta_ip_{i-1}(\lambda), \tag{21}\]

where \( p_{-1}(\lambda) = 0 \) and \( p_0(\lambda) = 1/d_0 \) where \( d_0^2 = \langle 1, 1 \rangle \). This can be seen as follows \([GK83]\).

Suppose \( p_0, \ldots, p_{n-1} \) 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 \( p_0(\lambda), \ldots, p_{i+1}(\lambda) \), and therefore

\[
\lambda \begin{bmatrix} p_0(\lambda) \\ p_1(\lambda) \\ \vdots \\ p_{n-2}(\lambda) \end{bmatrix} = \begin{bmatrix} J & \beta_{n-1}e_{n-1} \\ \vdots & \ddots & \vdots \\ \beta_1 & \cdots & \beta_{n-2} \\ \alpha_0 & \beta_1 & \cdots & 0 \\ \alpha_1 & \beta_2 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \end{bmatrix} \begin{bmatrix} p_0(\lambda) \\ p_1(\lambda) \\ \vdots \\ p_{n-1}(\lambda) \end{bmatrix} \tag{22}\]

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 (22), we find that

\[
H = JG + \beta_{n-1}e_{n-1}g^T
\]

where

\[
H_{ij} = \langle \lambda p_{i-1}(\lambda), p_{j-1}(\lambda) \rangle, \quad G_{ij} = \langle p_{i-1}(\lambda), p_{j-1}(\lambda) \rangle, \quad g_j = \langle p_{n-1}(\lambda), p_{j-1}(\lambda) \rangle,
\]

for \( i, j = 1, \ldots, n-1 \). The polynomials are orthonormal for the inner product if and only if \( G = I \) and \( g = 0 \). The inner product satisfies the property (20) if and only if \( H \) is symmetric. Hence if the polynomials are orthonormal for an inner product that satisfies (20), then \( J \) is a symmetric tridiagonal matrix. If we use the notation

\[
J = \begin{bmatrix} \alpha_0 & \beta_1 & 0 & \cdots & 0 & 0 \\ \beta_1 & \alpha_1 & \beta_2 & \cdots & 0 & 0 \\ 0 & \beta_2 & \alpha_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \alpha_{n-3} & \beta_{n-2} \\ 0 & 0 & 0 & \cdots & \beta_{n-2} & \alpha_{n-2} \end{bmatrix}, \tag{23}\]

the recursion (21) follows. Conversely, if the three-term recursion holds, and we define the inner product by setting \( G = I, g = 0 \), then \( H \) is symmetric and the inner product satisfies (20).

Now consider (4) and (7), with \( p = n - 1 \) and

\[
\Phi = \Phi_r, \quad \Psi = 0, \quad G = \begin{bmatrix} I_{n-1} & 0 \end{bmatrix}, \quad F = \begin{bmatrix} J & \beta_{n-1}e_{n-1} \end{bmatrix},
\]

where \( J \) is the Jacobi matrix (23) of a system of orthogonal polynomials. Then \( (\mu, \nu) \in \mathbb{C} \) if and only if either \( \nu \neq 0 \) and \( \lambda = \mu/\nu \in \mathbb{R} \), or \( \nu = 0 \). The set contains the vectors \( a \) of the form

\[
a = c(p_0(\lambda), p_1(\lambda), p_2(\lambda), \ldots, p_{n-1}(\lambda)), \quad a = c(0, 0, \ldots, 0, 1)
\]

for all \( \lambda \in \mathbb{R} \).
2.4 Rational functions

As a final example, we consider the controllability pencil of a linear system:

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

(24)

where \( A \in \mathbb{C}^{n_s \times n_s} \) and \( B \in \mathbb{C}^{n_s \times m} \). With this choice, \( A \) contains the vectors \( a = (x, u) \) that satisfy the equality \((\mu I - \nu A)x = \nu Bu\) 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 \(\lambda\) that satisfy \((\lambda, 1) \in \mathcal{C}\), then \( A \) contains the vectors

\[
a = \begin{bmatrix} (\lambda I - A)^{-1}Bu \\ u \end{bmatrix}
\]

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

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

\[
G = \begin{bmatrix} E & 0 \end{bmatrix}, \quad F = \begin{bmatrix} A & B \end{bmatrix},
\]

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

\[
a = \begin{bmatrix} (\lambda E - A)^{-1}Bu \\ u \end{bmatrix}
\]

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

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 \([\text{Roc70, section 15}]\), \([\text{KN77, chapter 1}]\). 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 \( \text{conv} \ C \) of a set \( C \) can be expressed as

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

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 \([\text{CRPW12}]\).
3.1 Symmetric matrices

Let $F, G, \Phi, \Psi$ be defined as in Theorem 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

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

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

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

$$= \inf \left\{ \sum_{k=1}^{r} \|a_k\|^2 \mid X = \sum_{k=1}^{r} a_k a_k^H, a_k \in \mathcal{A}, k = 1, \ldots, r \right\}. \quad (25)$$

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

The expressions $\sum_k \theta_k$ and $\sum \|a_k\|^2$ in these minimizations take only two possible values: $\text{tr} X$ if $X$ can be decomposed as in (25) and (26), and $+\infty$ otherwise. Theorem 1 tells us that a decomposition exists if only if $X$ is positive semidefinite and satisfies the two constraints (11), (12). Therefore

$$g(X) = \begin{cases} \text{tr} X & X \succeq 0, \quad (11), (12) \\ +\infty & \text{otherwise} \end{cases}. \quad (27)$$

Now consider an optimization problem in which we minimize the sum of a function $f: \mathbb{H}^n \to \mathbb{R}$ and the gauge defined in (26) and (27),

$$\text{minimize} \quad f(X) + g(X). \quad (28)$$

If we substitute the definition (26), this can be written as

$$\text{minimize} \quad f(X) + \sum_{k=1}^{r} \|a_k\|^2$$

subject to

$$X = \sum_{k=1}^{r} a_k a_k^H$$

$$a_k \in \mathcal{A}, k = 1, \ldots, r. \quad (29)$$

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 (28) acts as a regularization term that promotes a structured low rank property in $X$. If we substitute the expression (27) we obtain the equivalent formulation

$$\text{minimize} \quad f(X) + \text{tr} X$$

subject to

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

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

$$X \succeq 0. \quad (30)$$

This problem is convex if $f$ is convex.

A useful generalization of (26) is the gauge of the convex hull of

$$C = \{aa^H \mid a \in \mathcal{A}, \|Ea\| \leq 1\}$$
where $E$ may have rank less than $n$. The gauge of $\text{conv} \ C$ is
\[
g(X) = \inf \left\{ \sum_{k=1}^{r} \theta_k \mid X = \sum_{k=1}^{r} \theta_k a_k a_k^H, \theta_k \geq 0, a_k \in \mathcal{A}, \|Ea_k\| \leq 1, k = 1, \ldots, r \right\}. \tag{31}
\]
The variables $\theta_k$ 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 (31) are given, but not their norms or the coefficients $\theta_k$. If $0 < \|Ea_k\| < 1$, we can decrease $\theta_k$ by scaling $a_k$ until $\|Ea_k\| = 1$. If $Ea_k = 0$, $\theta_k$ 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:
\[
g(X) = \inf \left\{ \sum_{k=1}^{r} \|Ea_k\|^2 \mid X = \sum_{k=1}^{r} a_k a_k^H, a_k \in \mathcal{A}, k = 1, \ldots, r \right\}. \tag{32}
\]
Therefore $g(X) = \sum_k \|Ea_k\|^2 = \text{tr}(EXE^H)$ if $X$ can be decomposed as in (32) and $+\infty$ otherwise. Using Theorem 1 we can express this result as
\[
g(X) = \begin{cases} \text{tr}(EXE^H) & X \succeq 0, (11), (12) \\ +\infty & \text{otherwise.} \end{cases} \tag{33}
\]
Minimizing $f(X) + g(X)$ is equivalent to the optimization problem
\[
\begin{align*}
& \text{minimize} & & f(X) + \sum_{k=1}^{r} \|Ea_k\|^2 \\
& \text{subject to} & & X = \sum_{k=1}^{r} a_k a_k^H \\
& & & a_k \in \mathcal{A}, k = 1, \ldots, r,
\end{align*} \tag{34}
\]
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 (29). By choosing different $E$ we assign different weights to the vectors $a_k$. Using the expression (33), the problem (34) can be written as
\[
\begin{align*}
& \text{minimize} & & f(X) + \text{tr}(EXE^H) \\
& \text{subject to} & & \Phi_{11} FXF^H + \Phi_{21} FXG^H + \Phi_{12} GXF^H + \Phi_{22} GXG^H = 0 \\
& & & \Psi_{11} FXF^H + \Psi_{21} FXG^H + \Psi_{12} GXF^H + \Psi_{22} GXG^H \preceq 0 \\
& & & X \succeq 0.
\end{align*} \tag{35}
\]
Example Parametric line spectrum estimation is concerned with fitting signal models of the form
\[
y(t) = \sum_{k=1}^{r} c_k e^{j\omega_k t} + v(t), \tag{36}
\]
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 $\mathbb{E}|v(t)|^2 = \sigma^2$, then the covariance matrix of $y(t)$ of order $n$ is given by
\[
\begin{bmatrix}
  r_0 & r_{-1} & \cdots & r_{-n+1} \\
r_1 & r_0 & \cdots & r_{-n+2} \\
\vdots & \vdots & \ddots & \vdots \\
r_{n-1} & r_{n-2} & \cdots & r_0
\end{bmatrix}
= \sigma^2 I + \sum_{k=1}^{r} |c_k|^2 
\begin{bmatrix}
  1 \\
  e^{j\omega_k} \\
\vdots \\
  e^{j(n-1)\omega_k}
\end{bmatrix}
\begin{bmatrix}
  1 \\
  e^{j\omega_k} \\
\vdots \\
  e^{j(n-1)\omega_k}
\end{bmatrix}^H, \tag{37}
\]
\[
10
\]
where \( r_k = \mathbf{E}(y(t)\bar{y}(t-k)) \) \[SM97, \text{section 4.1}\] \[PM96, \text{section 12.5}\]. 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

\[
\begin{align*}
\text{minimize} & \quad f(R) + n \sum_{k=1}^{r} |c_k|^2 \\
\text{subject to} & \quad R = \sigma^2 I + \sum_{k=1}^{r} |c_k|^2 \left[ \begin{array}{c}
1 \\
e^{j\omega_k} \\
\vdots \\
e^{j(n-1)\omega_k}
\end{array} \right] \left[ \begin{array}{c}
1 \\
e^{j\omega_k} \\
\vdots \\
e^{j(n-1)\omega_k}
\end{array} \right]^H,
\end{align*}
\]

with variables \( R \in \mathbb{H}^n, \sigma^2, |c_k|, \omega_k, \) 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

\[
\begin{align*}
\text{minimize} & \quad f(X + tI) + \text{tr} \ X \\
\text{subject to} & \quad X \succeq 0, \ t \geq 0 \\
& \quad X \text{ is Toeplitz}.
\end{align*}
\]

A numerical example is given in section 5.

3.2 Non-symmetric matrices

We define \( F, G, E, \Phi, \Psi \), and \( A \) as in the previous section, but add the assumption that the matrices \( F, G, \) and \( E \) are block-diagonal:

\[
G = \left[ \begin{array}{cc}
G_1 & 0 \\
0 & G_2
\end{array} \right], \quad F = \left[ \begin{array}{cc}
F_1 & 0 \\
0 & F_2
\end{array} \right], \quad E = \left[ \begin{array}{cc}
E_1 & 0 \\
0 & E_2
\end{array} \right].
\]

(38)

Here \( F_1, G_1 \in \mathbb{C}^{p_1 \times n_1} \) and \( F_2, G_2 \in \mathbb{C}^{p_2 \times n_2} \) (possibly with \( p_1 \) or \( p_2 \) equal to zero). The matrices \( E_1 \) and \( E_2 \) have \( n_1 \) and \( n_2 \) columns, respectively. In this section we discuss the function

\[
h(Y) = \frac{1}{2} \inf_{V,W} g\left( \begin{array}{cc}
V & Y \\
Y^H & W
\end{array} \right)
\]

of \( Y \in \mathbb{C}^{n_1 \times n_2} \), where \( g \) is the function defined in (32) and (33). Using (32) we can write \( h(Y) \) as

\[
h(Y) = \inf \left\{ \frac{1}{2} \sum_{k=1}^{r} (||E_1 v_k||^2 + ||E_2 w_k||^2) : Y = \sum_{k=1}^{r} v_k w_k^H, \ (v_k, w_k) \in A \right\},
\]

(39)

while the equivalent characterization (33) shows that \( h(Y) \) is the optimal value of the SDP

\[
\begin{align*}
\text{minimize} & \quad \left( \text{tr}(E_1 V E_1^H) + \text{tr}(E_2 W E_2^H) \right) / 2 \\
\text{subject to} & \quad \Phi_{11} FXF^H + \Phi_{21} FXG^H + \Phi_{12} GXF^H + \Phi_{22} GXG^H = 0 \\
& \quad \Psi_{11} FXF^H + \Psi_{21} FXG^H + \Psi_{12} GXF^H + \Psi_{22} GXG^H \preceq 0 \\
& \quad X = \left[ \begin{array}{cc}
V & Y \\
Y^H & W
\end{array} \right] \succeq 0,
\end{align*}
\]

(40)
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 rectangular matrix. If we take $F$ and $G$ to have zero row dimensions (equivalently, define $\mathcal{A} = \mathbb{C}^{n_1} \times \mathbb{C}^{n_2}$ and omit the first two constraints in (40)) and choose identity matrices for $E_1$ and $E_2$, then $h(Y) = \|Y\|_*$, the trace norm of $Y$.

The block-diagonal structure of $F$ and $G$ implies that if $(v, w) \in \mathcal{A}$, then $(\alpha v, \beta w) \in \mathcal{A}$ for all $\alpha, \beta$. This observation leads to a number of useful equivalent expressions for (39). First, we note that $h(Y)$ can be written as

$$h(Y) = \inf \left\{ \sum_{k=1}^{r} \| E_1 v_k \| \| E_2 w_k \| \mid Y = \sum_{k=1}^{r} v_k w_k^H, (v_k, w_k) \in \mathcal{A} \right\}. \quad (41)$$

This follows from the fact $\| E_1 v_k \|^2 + \| E_2 w_k \|^2 \geq 2 \| E_1 v_k \| \| E_2 w_k \|$, with equality if $\| E_1 v_k \| = \| E_2 w_k \|$. If the decomposition of $Y$ in (39) 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 $\tilde{v}_k = \| E_2 w_k \|^{1/2} \| E_1 v_k \|^{1/2} v_k$, $\tilde{w}_k = \| E_1 v_k \|^{1/2} \| E_2 w_k \|^{1/2} w_k$ 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

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

as $\alpha$ 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 (39) and (41) give the same result.

From (41) we obtain two other useful expressions:

$$h(Y) = \inf \left\{ \sum_{k=1}^{r} \| E_1 v_k \| \mid Y = \sum_{k=1}^{r} v_k w_k^H, (v_k, w_k) \in \mathcal{A}, \| E_2 w_k \| \leq 1 \right\} \quad (42)$$

$$= \inf \left\{ \sum_{k=1}^{r} \| E_2 w_k \| \mid Y = \sum_{k=1}^{r} v_k w_k^H, (v_k, w_k) \in \mathcal{A}, \| E_1 v_k \| \leq 1 \right\}. \quad (43)$$

This again follows from the property that the two components of elements $(v_k, w_k)$ in $\mathcal{A}$ can be scaled independently. At the optimal decomposition in (42), all terms in the decomposition satisfy $E_2 w_k = 0$ or $\| E_2 w_k \| = 1$. In (43), all terms satisfy $E_1 v_k = 0$ or $\| E_1 v_k \| = 1$.

A final interpretation of $h$ is

$$h(Y) = \inf \left\{ \sum_{k=1}^{r} \theta_k \mid Y = \sum_{k=1}^{r} \theta_k v_k w_k^H, \theta_k \geq 0, (v_k, w_k) \in \mathcal{A}, \| E_1 v_k \| \leq 1, \| E_2 w_k \| \leq 1 \right\}. \quad (44)$$

The equivalence with (41) follows from the fact that if the optimal decomposition of $Y$ in (44) 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 one. (If $0 <
\[ \|E_1 v_k\| < 1 \text{ we can decrease } \theta_k \text{ by scaling } v_k \text{ until } \|E_1 v_k\| = 1, \text{ and similarly for } w_k. \]

The expression (44) shows that \( h(Y) \) is the gauge of the convex hull of the set
\[ \{vw^H \in \mathbb{C}^{n_1 \times n_2} \mid (v, w) \in \mathcal{A}, \|E_1 v\| \leq 1, \|E_2 w\| \leq 1 \}. \] (45)

The SDP representation of \( h \) in (40) allows us to reformulate problems
\[
\begin{align*}
\text{minimize} & \quad f(Y) + h(Y), \\
\text{subject to} & \quad Y = r \sum_{k=1}^{r} v_k w_k^H.
\end{align*}
\] (46)

where \( f \) is convex and \( h \) is the gauge (39)–(44), as a convex problem. Minimizing \( f(Y) + h(Y) \) is equivalent to
\[
\begin{align*}
\text{minimize} & \quad f(Y) + \sum_{k=1}^{r} \|E_1 v_k\| \|E_2 w_k\| \\
\text{subject to} & \quad Y = \sum_{k=1}^{r} v_k w_k^H \\
& \quad (v_k, w_k) \in \mathcal{A}, \ k = 1, \ldots, r.
\end{align*}
\] (47)

Alternatively, one can replace the second term in the objective with \( \sum_k \|E_2 w_k\| \) and add constraints \( \|E_1 v_k\| \leq 1, \) as in
\[
\begin{align*}
\text{minimize} & \quad f(Y) + \sum_{k=1}^{r} \|E_2 w_k\| \\
\text{subject to} & \quad Y = \sum_{k=1}^{r} v_k w_k^H \\
& \quad (v_k, w_k) \in \mathcal{A}, \ k = 1, \ldots, r \\
& \quad \|E_1 v_k\| \leq 1, \ k = 1, \ldots, r.
\end{align*}
\] (48)

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 (47) and (48) can be reformulated as
\[
\begin{align*}
\text{minimize} & \quad f(Y) + \frac{\text{tr}(E_1 V E_1^H) + \text{tr}(E_2 W E_2^H)}{2} \\
\text{subject to} & \quad \Phi_{11} FXF^H + \Phi_{21} F X G^H + \Phi_{12} G X F^H + \Phi_{22} G X G^H = 0 \\
& \quad \Psi_{11} FXF^H + \Psi_{21} F X G^H + \Psi_{12} G X F^H + \Psi_{22} G X G^H \preceq 0 \\
& \quad X = \begin{bmatrix} V & Y \\ Y^H & W \end{bmatrix} \succeq 0.
\end{align*}
\] (49)

**Example: column structure** When \( p_2 = 0 \), the matrices \( F \) and \( G \) in (38) have the form \( F = \begin{bmatrix} F_1 & 0 \end{bmatrix} \) and \( G = \begin{bmatrix} G_1 & 0 \end{bmatrix} \). This means that \( \mathcal{A} = \mathcal{A}_1 \times \mathbb{C}^{n_2} \) where
\[ \mathcal{A}_1 = \{v \in \mathbb{C}^{n_1} \mid (\mu G_1 - \nu F_1)v = 0, (\mu, \nu) \in \mathcal{C} \}. \]

There are no restrictions on the \( w \)-component of elements \((v, w) \in \mathcal{A} \). Problem (47) simplifies to
\[
\begin{align*}
\text{minimize} & \quad f(Y) + \sum_{k=1}^{r} \|E_1 v_k\| \|E_2 w_k\| \\
\text{subject to} & \quad Y = \sum_{k=1}^{r} v_k w_k^H \\
& \quad v_k \in \mathcal{A}_1, \ k = 1, \ldots, r.
\end{align*}
\] (50)
and the equivalent semidefinite formulation [49] to

\[
\begin{align*}
\text{minimize} & \quad f(Y) + (\text{tr}(E_1VE_1^H) + \text{tr}(E_2WE_2^H))/2 \\
\text{subject to} & \quad \Phi_{11}F_1VF_1^H + \Phi_{21}F_1VG_1^H + \Phi_{12}G_1VF_1^H + \Phi_{22}G_1VG_1^H = 0 \\
& \quad \Psi_{11}F_1VF_1^H + \Psi_{21}F_1VG_1^H + \Psi_{12}G_1VF_1^H + \Psi_{22}G_1VG_1^H \preceq 0 \\
& \quad \begin{bmatrix} V & Y \\ Y^H & W \end{bmatrix} \succeq 0.
\end{align*}
\]

As an example, we again consider the signal model (36). A natural idea for estimating the parameters \(\omega_k\) and \(c_k\) is to solve a nonlinear least squares problem

\[
\begin{align*}
\text{minimize} & \quad \sum_{t=0}^{n-1} |y_m(t) - \sum_{k=1}^r c_k e^{j\omega_k t}|^2,
\end{align*}
\]

where \(y_m(t)\) is the observed signal. This problem is not convex and difficult to solve iteratively without a good starting point [SM97, page 148]. However, suppose that, instead of fixing \(r\), we impose a penalty on \(\sum_k |c_k|\), and consider the optimization problem

\[
\begin{align*}
\text{minimize} & \quad \gamma \|y - y_m\|^2 + \sum_{k=1}^r |c_k| \\
\text{subject to} & \quad y = \sum_{k=1}^r c_k \begin{bmatrix} 1 \\ e^{j\omega_k} \\ \vdots \\ e^{j(n-1)\omega_k} \end{bmatrix},
\end{align*}
\]

The optimization variables are \(y\) and the parameters \(c_k, \omega_k, r\) in the decomposition of \(y\). The vector \(y_m\) has elements \(y_m(0), \ldots, y_m(n-1)\). This is a special case of [48] with \(f(y) = \gamma \|y - y_m\|^2\), \(n_1 = n, n_2 = 1, E_1 = \frac{1}{\sqrt{n}} I, E_2 = 1, F_1 = \begin{bmatrix} 0 & I_{n_1-1} \end{bmatrix}, G_1 = \begin{bmatrix} I_{n_1-1} & 0 \end{bmatrix}\), and \(\Phi = \Phi_u, \Psi = 0\), so that \(A_1\) is the set of all multiples of the vectors \((1, e^{i\omega}, \ldots, e^{j(n-1)\omega})\). The problem is therefore equivalent to the convex problem

\[
\begin{align*}
\text{minimize} & \quad \gamma \|y - y_m\|^2 + \text{tr}(V)/(2n) + w/2 \\
\text{subject to} & \quad \begin{bmatrix} V & y \\ y^H & w \end{bmatrix} \succeq 0 \\
& \quad V \text{ is Toeplitz.}
\end{align*}
\]

A related numerical example will be given in section 5.2.

**Example: joint column and row structure** To illustrate the general problem [47], we consider a variation on the previous example. Suppose we arrange the observations in an \(n \times m\) Hankel matrix

\[
Y_m = \begin{bmatrix} 
\begin{array}{cccc}
  y_m(0) & y_m(1) & \cdots & y_m(m-1) \\
  y_m(1) & y_m(2) & \cdots & y_m(m) \\
  \vdots & \vdots & \ddots & \vdots \\
  y_m(n-1) & y_m(n) & \cdots & y_m(m+n-1)
\end{array}
\end{bmatrix},
\]

14
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

$$\begin{align*}
\text{minimize} & \quad \gamma \| Y - Y_m \|_F^2 + \sum_{k=1}^{r} |c_k| \\
\text{subject to} & \quad Y = \sum_{k=1}^{r} c_k \begin{bmatrix}
1 \\
e^{j\omega_k} \\
\vdots \\
e^{j(n-1)\omega_k}
\end{bmatrix} \begin{bmatrix}
1 \\
e^{-j\omega_k} \\
\vdots \\
e^{-j(m-1)\omega_k}
\end{bmatrix}.
\end{align*}$$

(52)

This is an instance of (47) with $n_1 = n$, $n_2 = m$, $E_1 = (1/\sqrt{m})I$, $E_2 = (1/\sqrt{m})I$, and

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

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

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

The convex formulation is

$$\begin{align*}
\text{minimize} & \quad \gamma \| Y - Y_m \|_F^2 + (\text{tr} V)/(2n) + (\text{tr} W)/(2m) \\
\text{subject to} & \quad \begin{bmatrix} V & Y \\ Y^H & W \end{bmatrix} \succeq 0 \\
& \quad \begin{bmatrix} F_1 & 0 \\ 0 & F_2 \end{bmatrix} \begin{bmatrix} V & Y \\ Y^H & W \end{bmatrix} \begin{bmatrix} F_1 & 0 \\ 0 & F_2 \end{bmatrix}^T = \begin{bmatrix} G_1 & 0 \\ 0 & G_2 \end{bmatrix} \begin{bmatrix} V & Y \\ Y^H & W \end{bmatrix} \begin{bmatrix} G_1 & 0 \\ 0 & G_2 \end{bmatrix}^T.
\end{align*}$$

An 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 (28) and (46).

We assume that the subset of the complex plane represented by $C$ in (7) is one-dimensional, i.e., $C$ 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 $C$ represents a line or circle), or it is not redundant and then there exist elements of $C$ 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 (rank$(\mu G - \nu F) = p$) for all nonzero $(\mu, \nu)$. 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 (33). The conjugate is defined as

$$g^*(Z) = \sup_{X} (\text{tr}(XZ) - g(X)),$$
i.e., the optimal value of the SDP

\[
\begin{align*}
\text{maximize} & \quad \text{tr } ((Z - E^H E)X) \\
\text{subject to} & \quad X \succeq 0 \\
& \quad \Phi_{11}FXF^H + \Phi_{21}FXG^H + \Phi_{12}GXF^H + \Phi_{22}GXG^H = 0 \\
& \quad \Psi_{11}FXF^H + \Psi_{21}FXG^H + \Psi_{12}GXF^H + \Psi_{22}GXG^H \preceq 0.
\end{align*}
\]

(53)

The dual of this problem is

\[
\begin{align*}
\text{minimize} & \quad 0 \\
\text{subject to} & \quad Z - \begin{bmatrix} F \\ G \end{bmatrix}^H (\Phi \otimes P + \Psi \otimes Q) \begin{bmatrix} F \\ G \end{bmatrix} \preceq E^H E \\
& \quad Q \succeq 0,
\end{align*}
\]

(54)

with variables \(P, Q \in \mathbb{H}^p\). It is shown in appendix C that strong duality holds (under the assumptions listed at the top of section 4).

If strong duality holds, then \(g^*(Z)\) is the optimal value of (54), i.e., equal to zero if there exist \(P, Q\) that satisfy the constraints in (54), and \(+\infty\) otherwise. We now show that this can be expressed as

\[
g^*(Z) = \begin{cases} 
0 & a^HZa \leq \|Ea\|^2 \text{ for all } a \in \mathcal{A} \\
+\infty & \text{otherwise}.
\end{cases}
\]

(55)

Suppose \(P\) and \(Q\) are feasible in (54). Consider any \(a \in \mathcal{A}\) and \((\mu, \nu) \in \mathcal{C}\) with \(\mu Ga = \nu Fa\). Define \(y = (1/\nu)Ga\) if \(\nu \neq 0\) and \(y = (1/\mu)Fa\) otherwise. Then

\[
a^HZa - \|Ea\|^2 \leq \begin{bmatrix} Fa \\ Ga \end{bmatrix}^H (\Phi \otimes P + \Psi \otimes Q) \begin{bmatrix} Fa \\ Ga \end{bmatrix} = \begin{bmatrix} \mu y \\ \nu y \end{bmatrix}^H (\Phi \otimes P + \Psi \otimes Q) \begin{bmatrix} \mu y \\ \nu y \end{bmatrix} = (y^H P y)q_\Phi(\mu, \nu) + (y^H Q y)q_\Psi(\mu, \nu) \leq 0.
\]

The last line follows from \(Q \succeq 0\) and \(q_\Phi(\mu, \nu) = 0, q_\Psi(\mu, \nu) \leq 0\). Conversely, if problem (54) is infeasible, then the optimal value is \(+\infty\) and, since strong duality holds, there exist matrices \(X\) that are feasible for (53) with \(\text{tr } ((Z - E^H E)X) > 0\). Applying Theorem 1 we see that there exist \(a_1, \ldots, a_r \in \mathcal{A}\) with

\[
\sum_{k=1}^{r} (a_k^HZa_k - \|Ea_k\|^2) > 0.
\]

Therefore \(a_k^HZa_k > \|Ea_k\|^2\) for at least one \(a_k\).

The interpretation of the conjugate gives useful insight in problem (28), where \(g\) is defined in (33). The dual problem is

\[
\begin{align*}
\text{maximize} & \quad -f^*(Z) - g^*(-Z). \\
\text{subject to} & \quad -Z - \begin{bmatrix} F \\ G \end{bmatrix}^H (\Phi \otimes P + \Psi \otimes Q) \begin{bmatrix} F \\ G \end{bmatrix} \preceq E^H E \\
& \quad Q \succeq 0,
\end{align*}
\]

(56)
with variables \( Z, P, Q \), and using the expression (55) we can put the constraints in this problem more succinctly as

\[
\begin{align*}
\text{maximize} & \quad -f^*(Z) \\
\text{subject to} & \quad \|Ea\|^2 + a^HZa \geq 0 \quad \text{for all} \quad a \in A.
\end{align*}
\]

This last form leads to an interesting set of optimality conditions. Suppose \( X \) and \( Z \) are feasible for (34) and (57), respectively. Then

\[
f(X) + \sum_{k=1}^r \|Ea_k\|^2 \geq -f^*(Z) + \text{tr}(XZ) + \sum_{k=1}^r \|Ea_k\|^2 + a_k^HZa_k
\]

\[
\geq -f^*(Z).
\]

The first inequality follows by definition of \( f^*(Z) = \sup_X (\text{tr}(ZX) - f(X)) \), and the second and third line from primal and dual feasibility. If \( X \) and \( Z \) are optimal and strong duality holds, then

\[
f(X) + \sum_k \|Ea_k\|^2 = -f^*(Z). \quad \text{This is only possible if} \quad f(X) + f^*(Z) = \text{tr}(XZ) \quad \text{and}
\]

\[
\|Ea_k\|^2 + a_k^HZa_k = 0, \quad k = 1, \ldots, r.
\]

Hence only the vectors \( a \in A \) at which the inequality in (57) is active, can be used to form an optimal \( X = \sum_k a_k a_k^H \).

**Example: Generalized Kalman-Yakubovich-Popov lemma** When specialized to the controllability pencil (24), the equivalence between the constraints in (57) and (56) is known as the (generalized) Kalman-Yakubovich-Popov lemma [Kal63, Yak62, Pop62, Sch06, IH05].

We assume that \( A \) has no eigenvalues \( \lambda \) with \( (\lambda, 1) \in \mathbb{C} \), and that the pair \((A, B)\) is controllable, so the pencil satisfies the rank condition that \( \text{rank}(\lambda F - G) = n_s \) for all \( \lambda \). The dual problem (57) becomes

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

where

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

and \( M = EH^TE \). The function \( F \) is called the *Popov function* with central matrix \( M + Z \) [IOW99, HSK99].

### 4.2 Non-symmetric matrix gauge

Next we consider the conjugate of the gauge defined in (39)–(42). We have

\[
h^*(Z) = \sup_Y (\text{tr}(Z^TY) - h(Y))
\]
where \( h(Y) \) is the optimal value of (40). Therefore \( h^*(Z) \) is the optimal value of the SDP

\[
\begin{align*}
\text{maximize} & \quad \frac{1}{2} \text{tr}( \begin{pmatrix} -E_1^H E_1 & Z \\ Z^H & -E_2^H E_2 \end{pmatrix} X ) \\
\text{subject to} & \quad \Phi_{11} F X F^H + \Phi_{21} F X G^H + \Phi_{12} G X F^H + \Phi_{22} G X G^H = 0 \\
& \quad \Psi_{11} F X F^H + \Psi_{21} F X G^H + \Psi_{12} G X F^H + \Psi_{22} G X G^H \preceq 0 \\
& \quad X \succeq 0.
\end{align*}
\]

(58)

The dual of this problem is

\[
\begin{align*}
\text{minimize} & \quad 0 \\
\text{subject to} & \quad \begin{bmatrix} 0 & Z \\ Z^H & 0 \end{bmatrix} - \begin{bmatrix} F \\ G \end{bmatrix}^H (\Phi \otimes P + \Psi \otimes Q) \begin{bmatrix} F \\ G \end{bmatrix} \preceq \begin{bmatrix} E_1^H E_1 & 0 \\ 0 & E_2^H E_2 \end{bmatrix}
\end{align*}
\]

(59)

As in the previous section, it follows from appendix C that strong duality holds. Therefore \( h^*(Z) \) is equal to the optimal value of (59), i.e., zero if there exists \( P \) and \( Q \) that satisfy the constraints of this problem, and \( +\infty \) otherwise. This will now be shown to be equivalent to

\[
\begin{align*}
\begin{cases}
0 & \text{Re}(v^H Z w) \leq (\|E_1 v\|^2 + \|E_2 w\|^2)/2 \quad \text{for all } (v, w) \in \mathcal{A} \\
+\infty & \text{otherwise}
\end{cases}
\end{align*}
\]

(60)

To see this, first assume \( P \) and \( Q \) are feasible in (59), and \( a = (v, w) \in \mathcal{A} \) satisfies \((\mu G - \nu F) a = 0\) with \((\mu, \nu) \in \mathcal{C}\). Then

\[
v^H Z w + w^H Z^H v - \|E_1 v\|^2 - \|E_2 w\|^2 \leq \begin{bmatrix} F a \\ G a \end{bmatrix}^H (\Phi \otimes P + \Psi \otimes Q) \begin{bmatrix} F a \\ G a \end{bmatrix} = (y^H P y) q_\Phi(\mu, \nu) + (y^H Q y) q_\Psi(\mu, \nu) \leq 0,
\]

where we defined \( y = (1/\nu) G a \) if \( \nu \neq 0 \) and \( y = (1/\mu) F a \) otherwise. Conversely, if problem (59) is infeasible, then (58) is unbounded above, so there exists a feasible \( X \) with positive objective value. If we decompose \( X \) as in Theorem 1, with \( a_k = (v_k, w_k) \), we find that

\[
\begin{align*}
0 & < \text{tr}\left( \begin{bmatrix} -E_1^H E_1 & Z \\ Z^H & -E_2^H E_2 \end{bmatrix} \sum_{k=1}^r \begin{bmatrix} v_k \\ w_k \end{bmatrix} \begin{bmatrix} v_k \\ w_k \end{bmatrix}^H \right) \\
& = \sum_{k=1}^r (v_k^H Z w_k + w_k^H Z^H v_k - \|E_1 v_k\|^2 - \|E_2 w_k\|^2)
\end{align*}
\]

so at least one term in the sum is positive. The second expression for \( h^*(Z) \) in (60) follows from the block diagonal structure of \( F \) and \( G \).

The interpretation of the conjugate \( h^* \) can be applied to interpret the dual of (46), i.e.,

\[
\text{maximize} \quad -f^*(Z) - h^*(-Z).
\]
Substituting the expression (59) for $h^*(-Z)$, one can write this as

$$\begin{align*}
\text{maximize} & \quad -f^*(Z) \\
\text{subject to} & \quad \begin{bmatrix} 0 & -Z \\ -Z^H & 0 \end{bmatrix} - \begin{bmatrix} F \\ G \end{bmatrix} \begin{bmatrix} \Phi \otimes P + \Psi \otimes Q \end{bmatrix} \begin{bmatrix} F \\ G \end{bmatrix} \preceq \begin{bmatrix} E_1^H E_1 & 0 \\ 0 & E_2^H E_2 \end{bmatrix} \\
Q & \succeq 0,
\end{align*}$$

with variables $Z, P, Q$. Substituting (60) we obtain

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

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 $(v_k, w_k) \in A$ at which $\text{Re} \left( v_k^H Z w_k \right) = \|E_1 v_k\| \|E_2 w_k\|$.

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

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

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

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

for all $u \in \mathbb{C}^m$ and all $(\lambda, 1) \in \mathcal{C}$, plus the vectors $(0, u)$ if $(0, 1) \in \mathcal{C}$. Since $v$ is arbitrary and $E_1 = I$, the inequality in (60) reduces to $\|Z w\|_2 \leq \|E_2 w\|$ for all $w \in \mathcal{A}_2$. If $Z$ is partitioned as $Z = \begin{bmatrix} C & D \end{bmatrix}$, this is equivalent to a bound on the transfer function

$$\|D + C(\lambda I - A)^{-1} B\|_2 \leq 1 \quad \text{for all } (\lambda, 1) \in \mathcal{C}, \quad \|D\|_2 \leq 1 \quad \text{if } (1, 0) \in \mathcal{C}. \quad (61)$$

5 **Examples**

The formulations in section 3 will now be illustrated with a few examples from signal processing. The convex optimization problems in the examples were solved with CVX [GB14].

5.1 **Line spectrum estimation by Toeplitz covariance fitting**

In this example we fit a covariance matrix of the form (37) 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 (67), with $r = 3$, and frequencies $\omega_k$ and magnitudes $|c_k|$ shown in figure IV. The noise is Gaussian white noise with variance $\sigma^2 = 64$. The sample covariance matrix is constructed as

$$R_m = \frac{1}{N - n + 1} Y Y^H$$

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 parameters we solve the optimization problem

\[
\minimize \gamma \| R - R_m \|_2 + \sum_{k=1}^r |c_k|^2 \\
\text{subject to} \quad R = \sigma^2 I + \sum_{k=1}^r |c_k|^2 \begin{bmatrix} 1 \\ e^{j\omega_k} \\ \vdots \\ e^{j(n-1)\omega_k} \end{bmatrix} \begin{bmatrix} 1 \\ e^{j\omega_k} \\ \vdots \\ e^{j(n-1)\omega_k} \end{bmatrix}^H,
\]

with variables $|c_k|^2$, $\omega_k$, $r$, and $R$. The norm $\| \cdot \|_2$ in the objective is the spectral norm. The regularization parameter $\gamma$ is set to 0.25. This problem is equivalent to the convex problem

\[
\minimize \gamma \| tI + X - R_m \|_2 + \left( \frac{1}{n} \right) \text{tr} X \\
\text{subject to} \quad X \succeq 0, \quad t \geq 0 \quad FXF^H - GXG^H = 0
\]

with variables $X$ and $t$, and $F$ and $G$ defined in (15). As can be seen from Figure 1, the recovered parameters $\omega_k$ and $|c_k|$ are quite accurate, despite the very low signal-to-noise ratio. The estimated noise variance $t$ is 79.6.

The semidefinite optimization approach allows us to fit a covariance matrix with the structure prescribed in (37) 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 [Geo06].

### 5.2 Line spectrum estimation by penalty approximation

This example is a variation on problem (51). We take $n = 50$ consecutive measurements of the signal defined in (36). There are three sinusoids with frequencies and magnitudes shown in figure 3.
Figure 2: The input data for the example in section 5.2. The red dashed lines show the exact, noise-free signal. The blue and black circles show the exact signal corrupted by Gaussian white noise, plus a few larger errors in 20 positions. The green circles show the reconstructed signal.

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 the problem

$$\begin{align*}
\text{minimize} & \quad \gamma \sum_{i=1}^{n} \phi(y_i - y_{m,i}) + \sum_{k=1}^{r} |c_k| \\
\text{subject to} & \quad y = \sum_{k=1}^{r} c_k \begin{bmatrix}
1 \\
e^{j\omega_k} \\
\vdots \\
e^{j(n-1)\omega_k}
\end{bmatrix} \\
& \quad |\omega_k| \leq \omega_c, \quad k = 1, \ldots, r,
\end{align*}$$

(63)

where $\phi$ is the Huber penalty, $\gamma = 0.071$, and $\omega_c = \pi/6$. The variables in this problem are the $n$-vector $y$, and the parameters $r$, $c_k$, $\omega_k$ in the decomposition of $y$. The problem is equivalent to the convex problem

$$\begin{align*}
\text{minimize} & \quad \gamma \sum_{i=1}^{n} \phi(y_i - y_{m,i}) + (\text{tr} V)/(2n) + w/2 \\
\text{subject to} & \quad \begin{bmatrix}
V & y \\
yH & w
\end{bmatrix} \succeq 0 \\
& \quad FV F^H - GV G^H = 0 \\
& \quad -FVG^H - GXF^H + 2(\cos \omega_c)GVG^H \preceq 0
\end{align*}$$

(64)

with $F$ and $G$ defined in (15). The variables are the $n$-vector $y$, the Hermitian $n \times n$ matrix $V$, and the scalar $w$. The results are shown in Figure 3. The second figure shows the result of a simple
implementation of the matrix pencil method with a $30 \times 21$ Hankel matrix constructed from the measurements \cite{HS88}. The comparison shows the importance of the prior frequency constraints in the formulation \cite{63}.

It is interesting to note that problem \cite{63} can be equivalently formulated as

\[
\text{minimize } \gamma \sum_{i=1}^{n} \phi(y_i - y_{m,i}) + \sum_{k=1}^{r} |c_k| \\
\text{subject to } \begin{bmatrix} y_1 & y_2 & \cdots & y_{n_2} \\ y_2 & y_3 & \cdots & y_{n_2-1} \\ \vdots & \vdots & & \vdots \\ y_{n_1} & y_{n_1-1} & \cdots & y_{n_1+n_2-1} \end{bmatrix} = \sum_{k=1}^{r} c_k \begin{bmatrix} 1 \\ e^{j\omega_k} \\ \vdots \\ e^{j(n_1-1)\omega_k} \end{bmatrix} \begin{bmatrix} 1 \\ e^{-j\omega_k} \\ \vdots \\ e^{-j(n_2-1)\omega_k} \end{bmatrix}^H \\
|\omega_k| \leq \omega_c, \quad k = 1, \ldots, r,
\]

where $n_1 + n_2 - 1 = n$. This problem is equivalent to

\[
\text{minimize } \gamma \sum_{i=1}^{m} \phi(y_i - y_{m,i}) + (\text{tr} V)/(2n_1) + (\text{tr} W)/(2n_2) \\
\text{subject to } X = \begin{bmatrix} V & Y \\ Y^H & W \end{bmatrix} \succeq 0 \\
FXF^T = GXG^T \\
-FXG^T - GXF^T + 2\cos\omega_c GXG^T \preceq 0
\]

where $G$ and $F$ are block diagonal with blocks

\[
G_1 = \begin{bmatrix} I_{n_1-1} & 0 \end{bmatrix}, \quad F_1 = \begin{bmatrix} 0 & I_{n_1-1} \end{bmatrix}, \quad G_2 = \begin{bmatrix} 0 & I_{n_2-1} \end{bmatrix}, \quad F_2 = \begin{bmatrix} I_{n_2-1} & 0 \end{bmatrix}.
\]

The variables in \cite{66} are the matrices $V$, $Y$, $W$. The elements $y_i$ in the objective are the elements in the first row and last column of the matrix variable $Y$. The two SDP \cite{64} and \cite{66} give the same result $y$, but may have different numerical properties (in terms of accuracy or complexity).
5.3 Direction of arrival estimation

This example illustrates the use of frequency interval constraints in direction of arrival estimation. We consider the example described in [CV16, section 3.1]:

\[
\begin{align*}
\minimize & \quad \sum_{j=1}^{3} r_j \sum_{k=1}^{r_j} |x_{jk}| \\
\subjectto & \quad y_j = \sum_{k=1}^{r_j} x_{jk} \begin{bmatrix}
1 \\
e^{j\pi \sin \theta_{jk}} \\
\vdots \\
e^{j(n-1)\pi \sin \theta_{jk}}
\end{bmatrix} \\
\theta_{jk} & \in \Theta_j, \quad k = 1, \ldots, r_j, \quad j = 1, 2, 3 \\
(y_1 + y_2)I_1 &= b_1, \quad (y_2 + y_3)I_2 = b_2.
\end{align*}
\] (67)

The vectors \(b_1\) and \(b_2\) contain the outputs of two subsets of the elements in a linear array of \(n\) non-isotropic antennas. Elements in the first group, indexed by the index set \(I_1\), measure input signals arriving from angles in \(\Theta_1 \cup \Theta_2 = [-\pi/2, -\pi/6] \cup [-\pi/6, \pi/6]\). Elements in the second group, indexed by the index set \(I_2\), measure input signals arriving from \(\Theta_2 \cup \Theta_3 = [-\pi/6, \pi/6] \cup [\pi/6, \pi/2]\).

The convex formulation of this problem can be found in [CV16].

Figure 4 shows the results of an instance with \(n = 500\) elements in the array, but using only a total of 40 randomly selected measurements (\(|I_1| = |I_2| = 20\)). 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 \(|c_{jk}|\) are shown with blue lines. The right-hand plot shows the solution if we omit the interval constraints in (67).

Figure 5 shows the success rate as a function of the number \(|I_1| + |I_2|\) of available measurements, for an example with \(n = 50\) elements, and the same angles as in [CV16] and figure 4. Each data point is the average of 100 trials, with different, randomly generated coefficients, and different random selections of the two sensor groups. We observe that solving the optimization problem
with the interval constraints has a higher rate of exact recovery. For example, with 30 available measurements, including the interval constraints gave the exact answer in all instances, whereas the method without the interval constraints was successful in only about 25% of the instances.

5.4 Direction of arrival from multiple measurement vectors

This example demonstrates the advantage of using multiple measurement vectors (or snapshots), as pointed out in [LC14][YX14]. 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, and assume that the directions of arrival and the source magnitudes remain constant while the measurements are taken. The problem is formulated as

$$\begin{align*}
\text{minimize} \quad & \sum_{k=1}^{r} \|c_k\| \\
\text{subject to} \quad & Y = \sum_{k=1}^{r} \begin{bmatrix} 1 \\ e^{j\alpha \sin \theta_k} \\ \vdots \\ e^{j(n-1)\alpha \sin \theta_k} \end{bmatrix} c_k^H \\
& Y_I = B \\
& |\theta_k| \leq \theta_c, \quad k = 1, \ldots, r,
\end{align*}$$

(68)

with variables $Y \in \mathbb{C}^{n \times m}$, $c_k \in \mathbb{C}^m$, $\omega_k$, and $r$. Here $\alpha = 2\pi d/\lambda_c$, where $d$ is the distance between the grid points and $\lambda_c$ is the signal wavelength, and $\theta_c$ is a given cutoff angle. The columns of the $K \times m$ vector $B$ are the measurement vectors. The matrix $Y_I$ is the submatrix of $Y$ containing the rows indexed by $I$. The problem can be interpreted as identifying a continuous form of group
The convex formulation is

\[
\begin{align*}
\text{minimize} & \quad \frac{\text{tr} V}{2n} + \frac{\text{tr} W}{2} \\
\text{subject to} & \quad \begin{bmatrix} V & Y \\ Y^H & W \end{bmatrix} \succeq 0 \\
& \quad FV F^H - GV G^H = 0 \\
& \quad -FV G^H - GV F^H + 2 \cos \omega_c GV G^H \preceq 0 \\
& \quad Y_I = B
\end{align*}
\]

with \( F \) and \( G \) defined in (15) and \( \omega_c = \alpha \sin \theta_c \).

Figure 6 shows an example with \( n = 30, K = 7, \alpha = 2, \) and \( \theta_c = \pi/4 \). We show the solution for \( m = 1, m = 15, m = 30 \). The blue lines show the values of \( \omega_k \) and \( \|c_k\|/\sqrt{m} \) computed by solving problem (68).

## 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 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 [GHNV03,LP04,RV06,LV07,HV14], should be useful in the development of fast solvers for the SDPs discussed in this paper.

## A Subsets of the complex plane

In this appendix we explain the notation used in equation (7) to describe subsets of the closed complex plane. Recall that we use the notation

\[
q_\Theta(\mu, \nu) = \begin{bmatrix} \mu \\ \nu \end{bmatrix}^H \begin{bmatrix} \Theta_{11} & \Theta_{12} \\ \Theta_{21} & \Theta_{22} \end{bmatrix} \begin{bmatrix} \mu \\ \nu \end{bmatrix}
\]

for the quadratic form defined by a Hermitian \( 2 \times 2 \) matrix \( \Theta \).

### Lines and circles

If \( \Phi \) is a \( 2 \times 2 \) 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. Three important special cases are

\[
\Phi_u = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \quad \Phi_1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \Phi_r = \begin{bmatrix} 0 & j \\ -j & 0 \end{bmatrix},
\]

25
Figure 6: From top to bottom are shown the results of recovery with 1, 15, 30 measurement vectors in the DOA estimation problem of section 5.4. The figures on the right show the magnitude of the trigonometric polynomials obtained from the dual optimal solutions. The red dots represent the true directions of arrival (and magnitudes).
\[\angle \lambda \quad \Psi \quad \text{Assumptions}\]

\[
\begin{array}{ccc}
[a - b, a + b] & \begin{bmatrix} 0 & -e^{ja} \\ -e^{-ja} & 2\cos b \end{bmatrix} & 0 \leq b \leq \pi \\
[a, 2\pi - a] & \begin{bmatrix} 0 & 1 \\ 1 & -2\cos a \end{bmatrix} & 0 \leq a \leq \pi \\
\end{array}
\]

Table 1: Common choices of \(\Psi\) with \(\Phi = \Phi_u\) (\(\lambda\) on the unit circle).

\[
\begin{array}{ccc}
\text{Im} \lambda & \Psi & \text{Assumptions} \\
[a, b] & \begin{bmatrix} 2 & -j(a + b) \\ j(a + b) & 2ab \end{bmatrix} & a \leq b \\
[-\infty, -a] \cup [a, \infty] & \begin{bmatrix} -1 & 0 \\ 0 & a^2 \end{bmatrix} & a \geq 0 \\
\end{array}
\]

Table 2: Common choices of \(\Psi\) with \(\Phi = \Phi_i\) (\(\lambda\) imaginary).

for the unit circle, imaginary axis, and real axis, respectively. Curves defined by two different matrices \(\Phi, \tilde{\Phi}\) can be mapped to one another by applying a nonsingular congruence transformation \(\tilde{\Phi} = R\Phi R^H\).

When \(\Phi_{11} = 0\), we include the point \(\lambda = \infty\) in the solution set of (69). Alternatively, one can define points in the closed complex plane as directions \((\mu, \nu) \neq 0\). If \(\nu \neq 0\), the pair \((\mu, \nu)\) 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

\[q_{\Phi}(\mu, \nu) = \begin{bmatrix} \mu \\ \nu \end{bmatrix}^H \Phi \begin{bmatrix} \mu \\ \nu \end{bmatrix} = 0,
\]

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 \(\Phi\) and \(\tilde{\Phi}\).

**Segments of lines and circles** The second type of set we encounter is defined by a quadratic equality and inequality

\[q_{\Phi}(\lambda, 1) = 0, \quad q_{\Psi}(\lambda, 1) \leq 0. \quad (70)\]

We assume that \(\det \Phi < 0\). If the inequality is redundant (e.g., \(\Psi = 0\)) the solution set of (70) 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 points \((\mu, \nu)\) that satisfy

\[q_{\Phi}(\mu, \nu) = 0, \quad q_{\Psi}(\mu, \nu) \leq 0, \quad (\mu, \nu) \neq 0. \quad (71)\]

For easy reference, we list the most common combinations of \(\Phi\) and \(\Psi\) in tables [13][14][15][16][17].

As for circles and lines, we can apply a congruence transformation to reduce (70) to a simple canonical case. We mention two examples. Iwasaki and Hara [16][17] lemma 2 show that for every
| $\lambda$                      | $\Psi$                                                                | Assumptions |
|-------------------------------|-----------------------------------------------------------------------|-------------|
| $[a, b]$                      | \[
\begin{bmatrix}
2 & -(a + b) \\
-(a + b) & 2ab
\end{bmatrix}
\]                                                     | $a \leq b$  |
| $[-\infty, a] \cup [b, \infty]$ | \[
\begin{bmatrix}
-2 & a + b \\
a + b & -2ab
\end{bmatrix}
\]                                                     | $a \leq b$  |
| $[a, \infty]$                 | \[
\begin{bmatrix}
0 & -1 \\
-1 & 2a
\end{bmatrix}
\]                                                       |             |
| $[-\infty, a]$                | \[
\begin{bmatrix}
0 & 1 \\
1 & -2a
\end{bmatrix}
\]                                                       |             |

Table 3: Common choices of $\Psi$ with $\Phi = \Phi_r$ ($\lambda$ real).

$\Phi$, $\Psi$ with $\det \Phi < 0$, there exists a nonsingular $R$ such that

$$
\Phi = R^H \Phi_1 R, \quad \Psi = R^H \begin{bmatrix} \alpha & \beta \\ \beta & \gamma \end{bmatrix} R
$$

with $\alpha, \beta, \gamma$ real, and $\alpha \geq \gamma$. To see this, we first apply a congruence transformation $\Phi = R^H \Phi_1 R_1$ to transform $\Phi$ to $\Phi_1$. Define

$$
R_1^{-H} \Psi R_1^{-1} = \begin{bmatrix} x & \beta + jz \\ \beta - jz & y \end{bmatrix}
$$

with real $x, y, z, \beta$, and consider the eigenvalue decomposition

$$
\begin{bmatrix} x & jz \\ -jz & y \end{bmatrix} = Q \begin{bmatrix} \alpha & 0 \\ 0 & \gamma \end{bmatrix} Q^H,
$$

with eigenvalues sorted as $\alpha \geq \gamma$. Since the 2, 1 element of the matrix on the left-hand side of (73) is purely imaginary, the columns of $Q$ can be normalized to be of the form

$$
Q = \begin{bmatrix} u & jv \\ jv & u \end{bmatrix}
$$

with $u$ and $v$ real, and $u^2 + v^2 = 1$. This implies that $Q \Phi_1 Q^H = Q^H \Phi_1 Q = \Phi_1$ and

$$
Q^H \begin{bmatrix} x & \beta + jz \\ \beta - jz & y \end{bmatrix} Q = Q^H \begin{bmatrix} x & jz \\ -jz & y \end{bmatrix} Q + \begin{bmatrix} 0 & \beta \\ \beta & 0 \end{bmatrix} = \begin{bmatrix} \alpha & \beta \\ \beta & \gamma \end{bmatrix}.
$$

The transformation (72) now follows by taking $R = Q^H R_1$.

Applying the congruence defined by $R$, we can reduce the conditions (71) to an equivalent system

$$
\begin{bmatrix} \mu' \\ \nu' \end{bmatrix}^H \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \mu' \\ \nu' \end{bmatrix} = 0, \quad \begin{bmatrix} \mu' \\ \nu' \end{bmatrix}^H \begin{bmatrix} \alpha & 0 \\ 0 & \gamma \end{bmatrix} \begin{bmatrix} \mu' \\ \nu' \end{bmatrix} \leq 0, \quad (\mu', \nu') \neq 0,
$$

(74)

where $(\mu', \nu') = R(\mu, \nu)$. In non-homogeneous coordinates,

$$
\text{Re } \lambda' = 0, \quad \alpha |\lambda'|^2 + \gamma \leq 0.
$$

(75)
Keeping in mind that $\alpha \geq \gamma$, we can distinguish four cases. If $0 < \gamma \leq \alpha$ the solution set of (75) is empty. If $\gamma = 0 < \alpha$ the solution set is a singleton $\{0\}$. If $\gamma < 0 < \alpha$, the solution set of (75) 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 entire imaginary axis.

Another useful canonical form of (70) is obtained by transforming the solution set to a subset of the unit circle. If we define

$$T = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} R, \quad \epsilon = \frac{1}{2}(\alpha + \gamma), \quad \delta = \frac{1}{2}(\alpha - \gamma), \quad \eta = \beta,$$

then it follows from from (72) that

$$\Phi = T^H \Phi_u T, \quad \Psi = T^H \begin{bmatrix} \epsilon + \eta & -\delta \\ -\delta & \epsilon - \eta \end{bmatrix} T.$$

The coefficients $\epsilon, \delta, \eta$ are real, with $\delta \geq 0$. The congruence defined by $T$ therefore transforms the conditions (71) to an equivalent system

$$\begin{bmatrix} \mu' \\ \nu' \end{bmatrix}^H \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} \mu' \\ \nu' \end{bmatrix} = 0, \quad \begin{bmatrix} \mu' \\ \nu' \end{bmatrix}^H \begin{bmatrix} 0 & -\delta \\ -\delta & 2\epsilon \end{bmatrix} \begin{bmatrix} \mu' \\ \nu' \end{bmatrix} \leq 0,$$

where $(\mu', \nu') = T(\mu, \nu)$. In non-homogeneous coordinates, this is

$$|\lambda'|^2 = 1, \quad \delta \Re \lambda' \geq \epsilon.$$

The solution set is empty if $\epsilon > \delta$. It is the unit circle if $\epsilon \leq -\delta$. It is the singleton $\{1\}$ if $\epsilon = \delta > 0$. It is a segment of the unit circle if $-\delta < \epsilon < \delta$.

## B Matrix factorization results

This appendix contains a self-contained proof of Lemma 2, needed in the proof of Theorem 1, and some other matrix factorization results that have appeared in papers on the Kalman-Yakubovich-Popov (KYP) lemma [Ran96, IMF00, BV02, BV03, PV11]. We include the proofs because their constructive character is important for the result in Theorem 1.

Lemma 1 is based on [Ran96, lemma 3] and [IH05, lemma 5]. Lemma 2 can be found in [PV11 corollary 1].

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

- If $UU^H = VV^H$, then $U = VA$ for some unitary matrix $A \in \mathbb{C}^{r \times r}$.
- If $UU^H = VV^H$ and $UV^H + VU^H = 0$, then $U = VA$ for some unitary and skew-Hermitian matrix $A \in \mathbb{C}^{r \times r}$.
- If $UU^H \preceq VV^H$ and $UV^H + VU^H = 0$, then $U = VA$ for some skew-Hermitian matrix $A \in \mathbb{C}^{r \times r}$ with $\|A\|_2 \leq 1$. 

29
Proof. If $UU^H = VV^H$, then $U$ and $V$ have singular value decompositions of the form

$$U = P\Sigma Q_u^H, \quad V = P\Sigma Q_v^H,$$

with unitary matrices $P \in \mathbb{C}^{p \times p}$, diagonal $\Sigma \in \mathbb{C}^{p \times r}$, and unitary $Q_u, Q_v \in \mathbb{C}^{r \times r}$. The unitary matrix $\Lambda = Q_v Q_u^H$ satisfies $U = V\Lambda$.

To show the second part of the lemma, we substitute the singular value decompositions of $U$ and $V$ in the equation $UV^H + V^HU = 0$:

$$\Sigma (Q_u^H Q_v + Q_v^H Q_u) \Sigma^T = 0.$$

We define $\tilde{\Lambda} = Q_u^H Q_v$ (a unitary $r \times r$ matrix) and write this as

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

with $\Sigma_1$ positive diagonal of size $q \times q$, where $q = \text{rank}(U) = \text{rank}(V)$, and $\tilde{\Lambda}_{11}$ 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 $Q \in \mathbb{C}^{q \times q}$, and $\Delta$ a diagonal and purely imaginary matrix. Moreover $\Delta \Delta^H \preceq I$ because $\tilde{\Lambda}_{11}$ is a submatrix of the unitary matrix $\tilde{\Lambda}$.

Partition $Q$ and $\Delta$ as

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

with $\Delta_1 \Delta_1^H \preceq I$ and $\Delta_2 \Delta_2^H = I$. Since $\tilde{\Lambda}$ is unitary, we have

$$\tilde{\Lambda}_{12} \tilde{\Lambda}_{12}^H = I - \tilde{\Lambda}_{11} \tilde{\Lambda}_{11}^H = Q_1 Q_1^H + Q_2 Q_2^H - Q_1 \Delta_1 \Delta_1^H Q_1^H - Q_2 \Delta_2 \Delta_2^H Q_2^H = Q_1 (I - \Delta_1 \Delta_1^H) Q_1^H,$$

and, by the first part of the lemma, $\tilde{\Lambda}_{12} = Q_1 (I - \Delta_1 \Delta_1^H)^{1/2} W$ for some unitary matrix $W$. Therefore the matrix

$$\begin{bmatrix}
\tilde{\Lambda}_{11} & \tilde{\Lambda}_{12} \\
-W^H \Delta_1 W & \Delta_1^H W
\end{bmatrix} = \begin{bmatrix}
Q_1 & Q_2 \\
0 & 0
\end{bmatrix} \begin{bmatrix}
\Delta_1 & 0 \\
0 & \Delta_2 \\
0 & 0
\end{bmatrix} \begin{bmatrix}
(I - \Delta_1 \Delta_1^H)^{1/2} \\
0 \\
W^H
\end{bmatrix} \begin{bmatrix}
Q_1^H \\
Q_2^H \\
0 \end{bmatrix}$$

is skew-Hermitian (from the expression on the left-hand side and the fact that $\tilde{\Lambda}_{11}$ is skew-Hermitian and $\Delta_1$ is purely imaginary) and unitary (the right-hand side is a product of three unitary matrices).

If we now define

$$\Lambda = Q_v \begin{bmatrix}
\tilde{\Lambda}_{11} & \tilde{\Lambda}_{12} \\
-W^H \Delta_1 W & \Delta_1^H W
\end{bmatrix} Q_v^H,$$
then Λ is unitary and skew-Hermitian, and

\[
U = P \left[ \begin{array}{cc}
\Sigma_1 & 0 \\
0 & 0
\end{array} \right] \left[ \begin{array}{cc}
\tilde{\Lambda}_{11} & \tilde{\Lambda}_{12} \\
\tilde{\Lambda}_{21} & \tilde{\Lambda}_{22}
\end{array} \right] Q_v^H \\
= P \left[ \begin{array}{cc}
\Sigma_1 & 0 \\
0 & 0
\end{array} \right] \left[ \begin{array}{cc}
\tilde{\Lambda}_{11} & \tilde{\Lambda}_{12} \\
-\tilde{\Lambda}_{12}^H W^H \Delta_1^H W
\end{array} \right] Q_v^H \\
= P \left[ \begin{array}{cc}
\Sigma_1 & 0 \\
0 & 0
\end{array} \right] Q_v^H \Lambda \\
= V \Lambda.
\]

This proves part two of the lemma.

Assume \( UU^H \preceq VV^H \) and \( VV^H - UU^H \) has rank \( s \). We use any factorization \( VV^H - UU^H = \tilde{U}\tilde{U}^H \) with \( \tilde{U} \in \mathbb{C}^{p \times s} \) and write \( UU^H \preceq VV^H \) and \( UV^H + VU^H = 0 \) as

\[
\left[ \begin{array}{cc}
U & \tilde{U}
\end{array} \right] \left[ \begin{array}{cc}
U & \tilde{U}
\end{array} \right]^H = \left[ \begin{array}{cc}
V & 0
\end{array} \right] \left[ \begin{array}{cc}
V & 0
\end{array} \right]^H
\]

and

\[
\left[ \begin{array}{cc}
U & \tilde{U}
\end{array} \right] \left[ \begin{array}{cc}
V & 0
\end{array} \right]^H + \left[ \begin{array}{cc}
V & 0
\end{array} \right] \left[ \begin{array}{cc}
U & \tilde{U}
\end{array} \right]^H = 0.
\]

It follows from part 2 that

\[
\left[ \begin{array}{cc}
U & \tilde{U}
\end{array} \right] = \left[ \begin{array}{cc}
V & 0
\end{array} \right] \left[ \begin{array}{cc}
\tilde{\Lambda}_{11} & \tilde{\Lambda}_{12} \\
\tilde{\Lambda}_{21} & \tilde{\Lambda}_{22}
\end{array} \right]
\]

with \( \tilde{\Lambda} \) unitary and skew-Hermitian. The subblock \( \Lambda = \tilde{\Lambda}_{11} \) satisfies \( U = V\Lambda \), \( \Lambda + \Lambda^H = 0 \) and \( \Lambda^H \Lambda \preceq I \).

\[\blacksquare\]

**Lemma 2** Let \( \Phi, \Psi \in \mathbb{H}^2 \) with \( \det \Phi < 0 \). If \( U, V \in \mathbb{C}^{p \times r} \) satisfy

\[
\Phi_{11}UU^H + \Phi_{21}UV^H + \Phi_{12}VU^H + \Phi_{22}VV^H = 0, \tag{76}
\]

\[
\Psi_{11}UU^H + \Psi_{21}UV^H + \Psi_{12}VU^H + \Psi_{22}VV^H \preceq 0, \tag{77}
\]

then there exist a matrix \( W \in \mathbb{C}^{p \times r} \), a unitary matrix \( Q \in \mathbb{C}^{r \times r} \), and vectors \( \mu, \nu \in \mathbb{C}^r \) such that

\[
U = W \text{diag}(\mu)Q^H, \quad V = W \text{diag}(\nu)Q^H, \tag{78}
\]

and

\[
q_\Phi(\mu_i, \nu_i) = 0, \quad q_\Psi(\mu_i, \nu_i) \leq 0, \quad (\mu_i, \nu_i) \neq 0, \quad i = 1, \ldots, r. \tag{79}
\]

**Proof.** Suppose \( U \) and \( V \) are \( p \times r \) matrices that satisfy (76) and (77). As explained in appendix [A], there exists a nonsingular \( R \) such that

\[
\Phi = R^H \left[ \begin{array}{cc}
0 & 1 \\
1 & 0
\end{array} \right] R, \quad \Psi = R^H \left[ \begin{array}{cc}
\alpha & \beta \\
\beta & \gamma
\end{array} \right] R
\]

with \( \beta \) real and \( \gamma \leq \alpha \). Define \( S = R_{11}U + R_{12}V \) and \( T = R_{21}U + R_{22}V \). From (76) and (77),

\[
\left[ \begin{array}{cc}
S & T
\end{array} \right] \left[ \begin{array}{cc}
0 & I \\
I & 0
\end{array} \right] \left[ \begin{array}{cc}
S^H \\
T^H
\end{array} \right] = \left[ \begin{array}{cc}
U & V
\end{array} \right] \left[ \begin{array}{cc}
\Phi_{11}I & \Phi_{21}I \\
\Phi_{12}I & \Phi_{22}I
\end{array} \right] \left[ \begin{array}{cc}
U^H \\
V^H
\end{array} \right] = 0
\]
and
\[
\begin{bmatrix}
S & T
\end{bmatrix}
\begin{bmatrix}
\alpha I & \beta I \\
\beta I & \gamma I
\end{bmatrix}
\begin{bmatrix}
S^H \\
T^H
\end{bmatrix}
= \begin{bmatrix}
U & V
\end{bmatrix}
\begin{bmatrix}
\Psi_{11}I & \Psi_{21}I \\
\Psi_{12}I & \Psi_{22}I
\end{bmatrix}
\begin{bmatrix}
U^H \\
V^H
\end{bmatrix} \leq 0.
\]
Therefore
\[ST^H + TS^H = 0, \quad \alpha SS^H + \gamma TT^H \leq 0. \tag{80}\]

We show that this implies that
\[S = W \text{diag}(s)Q^H, \quad T = W \text{diag}(t)Q^H, \tag{81}\]
for some \(W \in \mathbb{C}^{p \times r}\), unitary \(Q \in \mathbb{C}^{r \times r}\), and vectors \(s, t \in \mathbb{C}^r\) that satisfy
\[s_i t_i + \bar{s}_i \bar{t}_i = 0, \quad \alpha |s_i|^2 + \gamma |t_i|^2 \leq 0, \quad (s_i, t_i) \neq 0, \quad i = 1, \ldots, r. \tag{82}\]

The result is trivial if \(S\) and \(T\) are zero, since in that case we can choose \(W\) zero, and arbitrary \(Q, s, t\). If at least one of the two matrices is nonzero, then the inequality in (80), combined with \(\alpha \geq \gamma\), implies that \(\gamma \leq 0\). Therefore there are three cases to consider.

- If \(\alpha \leq 0\), we write the equality in (80) as
  \[(S + T)(S + T)^H = (S - T)(S - T)^H.\]
  From Lemma 4 this implies that \(S + T = (S - T)\Lambda\) with \(\Lambda\) unitary. Let \(\Lambda = Q\text{diag}(\rho)Q^H\) be the Schur decomposition of \(\Lambda\), with \(|\rho_i| = 1\) for \(i = 1, \ldots, r\). Define
  \[W = (S - T)Q, \quad s = \frac{1}{2}(\rho + 1), \quad t = \frac{1}{2}(\rho - 1).\]

- If \(\gamma = 0 < \alpha\), then \(S = 0\), and we can take \(Q = I\)
  \[W = T, \quad s = 0, \quad t = 1.\]

- If \(\gamma < 0 < \alpha\), then from Lemma 4 we have \(S = (-\gamma/\alpha)^{1/2}TA\) for some skew-Hermitian \(A\) with \(A^H A \leq I\). This matrix has a Schur decomposition \(\Lambda = Q\text{diag}(\rho)Q^H\) with \(|\rho_i| \leq 1\) for \(j = 1, \ldots, r\). Define
  \[W = TQ, \quad s = (-\gamma/\alpha)^{1/2}p, \quad t = 1.\]

The factorizations of \(U\) and \(V\) now follow from
\[
\begin{bmatrix}
U \\
V
\end{bmatrix} = (R^{-1} \otimes I)
\begin{bmatrix}
S \\
T
\end{bmatrix}
= (R^{-1} \otimes I)
\begin{bmatrix}
W \text{diag}(s) \\
W \text{diag}(t)
\end{bmatrix}Q^H
= \begin{bmatrix}
W \text{diag}(\mu) \\
W \text{diag}(\nu)
\end{bmatrix}Q^H
\]
where \(\mu\) and \(\nu\) are defined as
\[
\begin{bmatrix}
\mu_i \\
\nu_i
\end{bmatrix} = R^{-1}
\begin{bmatrix}
s_i \\
t_i
\end{bmatrix}, \quad i = 1, \ldots, r.
\]
These pairs \((\mu_i, \nu_i)\) are nonzero and satisfy
\[
\begin{bmatrix}
\mu_i \\
\nu_i
\end{bmatrix}^H \Phi
\begin{bmatrix}
\mu_i \\
\nu_i
\end{bmatrix} = \begin{bmatrix}
s_i \\
t_i
\end{bmatrix}^H \begin{bmatrix}
0 & 1 \\
1 & 0
\end{bmatrix} \begin{bmatrix}
s_i \\
t_i
\end{bmatrix} = \bar{s}_i t_i + s_i \bar{t}_i = 0
\]
and
\[
\begin{bmatrix}
\mu_i \\
\nu_i
\end{bmatrix}^H \Psi
\begin{bmatrix}
\mu_i \\
\nu_i
\end{bmatrix} = \begin{bmatrix}
s_i \\
t_i
\end{bmatrix}^H \begin{bmatrix}
\alpha & \beta \\
\beta & \gamma
\end{bmatrix} \begin{bmatrix}
s_i \\
t_i
\end{bmatrix} = \alpha |s_i|^2 + \beta \bar{s}_i t_i + s_i \bar{t}_i + \gamma |t_i|^2 \leq 0.
\]

\[\square\]

32
C Strict feasibility

In this appendix we discuss strict feasibility of the constraints \( X \geq 0, (11), (12) \) in Theorem 1.

We assume that the set \( C \) 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 \( C \) with \( q_\Psi(\mu, \nu) < 0 \). We will distinguish these two cases.

- **Line or circle.** If the inequality \( q_\Psi(\mu, \nu) \leq 0 \) in the definition is redundant, we have

\[
C = \{ (\mu, \nu) \in C^2 \mid (\mu, \nu) \neq 0, \ q_\Phi(\mu, \nu) = 0 \},
\]

and \( C \) is a line or circle in homogeneous coordinates. In this case we understand by strict feasibility of \( X \) that

\[
X \succ 0, \quad \Phi_{11} F X F^H + \Phi_{21} F X G^H + \Phi_{12} G X F^H + \Phi_{22} G X G^H = 0. \tag{83}
\]

We also define \( C^\circ = C \).

- **Segment of line or circle.** In the second case, \( C \) is a proper one-dimensional subset of the line or circle defined by \( q_\Psi(\mu, \nu) = 0 \). In this case we define strict feasibility of \( X \) as

\[
\Psi_{11} F X F^H + \Psi_{21} F X G^H + \Psi_{12} G X F^H + \Psi_{22} G X G^H \prec 0. \tag{84}
\]

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

The conditions on \( F \) and \( G \) that guarantee strict feasibility will be expressed in terms of the Kronecker structure of the matrix pencil \( \lambda G - F \) [Gam05|Van79]. For every matrix pencil there exist nonsingular matrices \( P \) and \( Q \) such that

\[
P(\lambda G - F)Q = \begin{bmatrix}
\eta_1(\lambda)^T & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 \\
0 & \eta_2(\lambda)^T & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \eta_1(\lambda)^T & 0 & 0 & 0 & \cdots & 0 \\
0 & 0 & \cdots & 0 & \lambda B - A & 0 & 0 & \cdots & 0 \\
0 & 0 & \cdots & 0 & 0 & L_{\epsilon_1}(\lambda) & 0 & \cdots & 0 \\
0 & 0 & \cdots & 0 & 0 & 0 & L_{\epsilon_2}(\lambda) & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\
0 & 0 & \cdots & 0 & 0 & 0 & \cdots & L_{\epsilon_r}(\lambda)
\end{bmatrix}, \tag{85}
\]

where \( L_{\epsilon}(\lambda) \) is the \( \epsilon \times (\epsilon + 1) \) pencil

\[
L_{\epsilon}(\lambda) = \begin{bmatrix}
\lambda & -1 & 0 & \cdots & 0 & 0 \\
0 & \lambda & -1 & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & \cdots & -1 & 0 & 0 \\
0 & 0 & \cdots & \lambda & -1
\end{bmatrix}.
\]
and $\lambda B - A$ is a regular pencil, \textit{i.e.}, it is square and $\det(\lambda B - A)$ is not identically zero. The generalized eigenvalues of $\lambda B - A$ are sometimes referred to as the generalized eigenvalues of the pencil $\lambda G - F$ [1099, page 16]. The parameters $\epsilon_1, \ldots, \epsilon_r$ are the \textit{right Kronecker indices} of the pencil and the parameters $\eta_1, \ldots, \eta_l$ are the \textit{left Kronecker indices}. The \textit{normal rank} of the pencil is equal to $p - l$, where $p$ is the row dimension of $F$ and $G$.

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 (85).

2. The generalized eigenvalues of the pencil $\lambda G - F$ (defined as the generalized eigenvalues of $\lambda B - A$) are nondefective, \textit{i.e.}, their algebraic multiplicity is equal to the geometric multiplicity, and lie in $C^\circ$. (More accurately, if $\lambda$ is a finite generalized eigenvalue, then $(\lambda, 1) \in C^\circ$. If it is an infinite generalized eigenvalue, then $(1, 0) \in C^\circ$).

A sufficient but more easily verified condition is that $\text{rank}(\mu G - \nu F) = p$ for all $(\mu, \nu) \neq 0$, \textit{i.e.}, $l = 0$ and the block $\lambda B - A$ in (85) is not present.

\textbf{Proof.} Without loss of generality we can assume that the pencil is in the Kronecker canonical form ($P = I, Q = I$ in (85)) and that $\Phi = \Phi_u$, so the equality constraint in (83) is

$$ FXF^T = GXG^T. $$

(86)

We first show that the two 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 (85). Suppose $l \geq 1$ and consider the $k$th diagonal block $X_{kk}$ with $1 \leq k \leq l$. The $k$th diagonal block of the pencil is

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

The $k$th diagonal block of (86) is $F_k X_{kk} F_k^H = G_k X_{kk} G_k^H$ or

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

This is impossible since $X_{kk} \succ 0$. Hence, if (86) 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 $\mu/\nu$ 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, \textit{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 (86) is

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

(87)

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 $q_{\Phi}(\mu, \nu) = |\mu|^2 - |\nu|^2 = 0$, \textit{i.e.}, the generalized eigenvalues are on the unit circle. In addition, if the inequality in (84) holds, then

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

34
and from this, $qq(\mu, \nu)(u^H X_{11} u) < 0$. This is only possible if $qq(\mu, \nu) < 0$. We conclude that if $\det(\mu B - \nu A) = 0$ for nonzero $(\mu, \nu)$, then $(\mu, \nu) \in \mathbb{C}^o$.

Next we show that the generalized eigenvalues of the pencil $\lambda B - A$ are nondefective. Since $\mathbb{C}^o$ 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.,

$$\lambda B - A = \begin{bmatrix}
(\lambda - \rho_1)I - J_{s_1} & 0 & \cdots & 0 \\
0 & (\lambda - \rho_2)I - J_{s_2} & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & (\lambda - \rho_t)I - J_{s_t}
\end{bmatrix},$$

where $\rho_1, \ldots, \rho_t$ are the generalized eigenvalues (which satisfy $|\rho_i| = 1$), and $J_s$ is the $s \times s$ matrix

$$J_s = \begin{bmatrix}
0 & 1 & 0 & \cdots & 0 & 0 \\
0 & 0 & 1 & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & 0 & 1 \\
0 & 0 & 0 & \cdots & 0 & 0
\end{bmatrix}.$$

Then (87) implies that

$$(\rho_i - J_{s_i})X_{11,i}(\rho_i - J_{s_i})^H = X_{11,i}$$

where $X_{11,i}$ is the $i$th diagonal block of $X_{11}$, if we partition $X_{11}$ as a $t \times t$ block matrix with $i, j$ block of size of $s_i \times s_j$. Expanding this gives

$$|\rho_i|^2 X_{11,i} - \rho_i X_{11,i} J_{s_i}^T - \rho_i J_{s_i} X_{11,i} + J_{s_i} X_{11,i} J_{s_i}^T = X_{11,i}.$$  

Since $|\rho_i| = 1$ this simplifies to

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

The last row of the second matrix on the left-hand side and the last row of the matrix on the right-hand side are zero. Therefore the last row of the first matrix on the left is zero. However the element in column $s_i - 1$ is the last diagonal element of the positive definite matrix $X_{11,i}$. Hence, we have a contradiction unless $s_i = 1$, i.e., the generalized eigenvalue $\rho_i$ is nondefective. We conclude that the two conditions are necessary.

It remains to show that the conditions are sufficient. If the two conditions hold, then $\lambda G - F$ has the Kronecker canonical form

$$\lambda G - F = \begin{bmatrix}
\lambda - \rho_1 & \cdots & 0 & 0 & \cdots & 0 \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
0 & \cdots & \lambda - \rho_t & 0 & \cdots & 0 \\
0 & \cdots & 0 & L_{\epsilon_1}(\lambda) & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\
0 & \cdots & 0 & 0 & \cdots & L_{\epsilon_r}(\lambda)
\end{bmatrix}.$$
with \( \rho_i \in \mathcal{C}^\circ \) for \( i = 1, \ldots, t \). Define a block diagonal matrix

\[
X = \begin{bmatrix}
1 & \cdots & 0 & 0 & \cdots & 0 \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
0 & \cdots & 1 & 0 & \cdots & 0 \\
0 & \cdots & 0 & X_{11} & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & \cdots & 0 & 0 & \cdots & X_{rr}
\end{bmatrix}
\]

with diagonal blocks

\[
X_{kk} = \sum_{i=1}^{\epsilon_k+1} \begin{bmatrix}
\lambda_{ki}^1 \\
\lambda_{ki}^2 \\
\vdots \\
\lambda_{ki}^{\epsilon_k} \\
\end{bmatrix} \begin{bmatrix}
1 \\
\lambda_{ki}^1 \\
\lambda_{ki}^2 \\
\vdots \\
\lambda_{ki}^{\epsilon_k}
\end{bmatrix}^H
\]

for \( k = 1, \ldots, r \), where \( \lambda_{k1}, \ldots, \lambda_{k,\epsilon_k+1} \) are distinct elements of \( \mathcal{C}^\circ \). This matrix \( X \) is strictly feasible. \( \square \)

References

[BTR12] B. N. Bhaskar, G. Tang, and B. Recht. Atomic norm denoising with applications to line spectral estimation. 2012. arXiv:cs/0412009.

[BV02] 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, 2002.

[BV03] V. Balakrishnan and L. Vandenberghe. Semidefinite programming duality and linear time-invariant systems. IEEE Transactions on Automatic Control, 48:30–41, 2003.

[CC15] Y. Chi and Y. Chen. Compressive two-dimensional harmonic retrieval via atomic norm minimization. IEEE Transactions on Signal Processing, 63(4):1030–1042, Feb 2015.

[CFG14] E. J. Candès and C. Fernandez-Granda. Towards a mathematical theory of super-resolution. Communications of Pure and Applied Mathematics, 67(6):906–956, 2014.

[CRPW12] V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. S. Willsky. The convex geometry of linear inverse problems. Foundations of Computational Mathematics, 12:805–849, 2012.

[CV16] 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.

[dCG12] Y. de Castro and F. Gamboa. Exact reconstruction using Beurling minimal extrapolation. Journal of Mathematical Analysis and Applications, 395(1):336 – 354, 2012.
[dGHL15] Y de Castro, F Gamboa, D Henrion, and J.-B Lasserre. Exact solutions to Super Resolution on semi-algebraic domains in higher dimensions. Feb 2015. arXiv:cs/0412009, 2015.

[FG15] C. Fernandez-Granda. Super-resolution of point sources via convex programming. 2015. arXiv:1507.07034.

[Gan05] F. R. Gantmacher. Applications of the Theory of Matrices. Dover Publications, 2005. Originally published in 1959 by Interscience Publishers, Inc., New York.

[GB14] M. Grant and S. Boyd. CVX: Matlab software for disciplined convex programming, version 2.1. http://cvxr.com/cvx, March 2014.

[Geo06] T. T. Georgiou. Decomposition of Toeplitz matrices via convex optimization. IEEE Signal Processing Letters, 13(9):537–540, 2006.

[GHNV03] Y. Genin, Y. Hachez, Yu. Nesterov, and P. Van Dooren. Optimization problems over positive pseudopolynomial matrices. SIAM Journal on Matrix Analysis and Applications, 25(1):57–79, 2003.

[GK83] G. H. Golub and J. Kautsky. Calculation of Gauss quadratures with multiple free and fixed knots. Numerische Mathematik, 41:147–163, 1983.

[GS84] U. Grenander and G. Szegö. Toeplitz Forms and Their Applications. Chelsea, New York, 1984. First published in 1958.

[HS88] Y. Hua and T. K. Sakar. Matrix pencil method and its performance. IEEE International Conference on Acoustics, Speech, and Signal Processing, 4:2476–2479, April 1988.

[HSK99] B. Hassibi, A. H. Sayed, and T. Kailath. Indefinite-Quadratic Estimation and Control. A Unified Approach to $H^2$ and $H^\infty$ Theories. Society for Industrial and Applied Mathematics, 1999.

[HV14] A. Hansson and L. Vandenberghe. Sampling method for semidefinite programmes with non-negative Popov function constraints. International Journal of Control, 87(2):330–345, 2014.

[IH03] T. Iwasaki and S. Hara. Generalization of Kalman-Yakubović-Popov lemma for restricted frequency inequalities. In Proceedings of the American Control Conference, pages 3828–3833, 2003.

[IH05] T. Iwasaki and S. Hara. Generalized KYP lemma: unified frequency domain inequalities with design applications. IEEE Transactions on Automatic Control, 50(1):41–59, 2005.

[IMF00] T. Iwasaki, G. Meinsma, and M. Fu. Generalized $S$-procedure and finite-frequency KYP lemma. Mathematical Problems in Engineering, 6:305–320, 2000.

[IOW99] V. Ionescu, C. Oară, and M. Weiss. Generalized Riccati Theory and Robust Control. John Wiley and Sons, 1999.
[TBSR13] G. Tang, B. N. Bhaskar, P. Shah, and B. Recht. Compressed sensing off the grid. *IEEE Transactions on Information Theory*, 59(11):7465–7490, 2013.

[Van79] P. Van Dooren. The computation of Kronecker’s canonical form of a singular pencil. *Linear Algebra and Its Applications*, 27:103–140, 1979.

[Yak62] V. A. Yakubovich. The solution of certain matrix inequalities in automatic control theory. *Soviet Math. Dokl.*, 3:620–623, 1962.

[YX14] Z. Yang and L. Xie. Exact joint sparse frequency recovery via optimization methods. 2014. arXiv:cs/1405.6585.

[YX15] Z. Yang and L. Xie. On gridless sparse methods for line spectral estimation from complete and incomplete data. *IEEE Transactions on Signal Processing*, 63(12):3139–3153, June 2015.